HAL open science

Adaptive estimation of the hazard rate with multiplicative censoring

G Chagny, Fabienne Comte, Angelina Roche

To cite this version:

G Chagny, Fabienne Comte, Angelina Roche. Adaptive estimation of the hazard rate with multiplicative censoring. 2016. hal-01362694

HAL Id: hal-01362694

https://hal.science/hal-01362694v1

Preprint submitted on 9 Sep 2016

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

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

ADAPTIVE ESTIMATION OF THE HAZARD RATE WITH MULTIPLICATIVE CENSORING

G. CHAGNY[(A)], F. COMTE[(B)] AND A. ROCHE[(C)]

Abstract. We propose an adaptive estimation procedure of the hazard rate of a random variable X in the multiplicative censoring model, $Y = XU$, with $U \sim \mathcal{U}([0,1])$ independent of X. The variable X is not directly observed: an estimator is built from a sample $\{Y_1, ..., Y_n\}$ of copies of Y. It is obtained by minimisation of a contrast function over a class of general nested function spaces which can be generated e.g. by splines functions. The dimension of the space is selected by a penalised contrast criterion. The final estimator is proved to achieve the best bias-variance compromise and to reach the same convergence rate as the oracle estimator under conditions on the maximal dimension. The good behavior of the resulting estimator is illustrated over a simulation study.

(A) LMRS, UMR CNRS 6085, Université de Rouen Normandie, France. gaelle.chagny@univ-rouen.fr

(B) MAP5 UMR CNRS 8145, Université Paris Descartes, France. fabienne.comte@parisdescartes.fr

(C) CEREMADE UMR CNRS 7534, Université Paris-Dauphine, France. angelina.roche@dauphine.fr

Keywords. Adaptive procedure. Model selection. Hazard rate estimation. Multiplicative censoring model.

AMS Subject Classification 2010. 62G05; 62N01.

1. Introduction

In this paper, our aim is to estimate the hazard rate associated with a nonnegative random variable X, defined by

$$h = f_X/\bar{F}_X$$

where $\bar{F}_X(x) = 1 - F_X(x) = \mathbb{P}(X \geq x)$ (resp. f_X) is the survival function (resp. the density) of X. However, instead of having at our disposal an independent and identically distributed (*i.i.d.* in the sequel) sample $X_1, \ldots, X_n$ with the distribution of X, we assume that we observe $\{Y_1, \ldots, Y_n\}$ such that

$$Y_i = X_i U_i, \text{ for all } i = 1, \ldots, n, \tag{1}$$

where X_i is a non-negative unobserved random variable and U_i is also unobserved and follows the uniform distribution on the interval $[0,1]$. The quantity of interest, X_i, is supposed to be independent of U_i, for all $i = 1, \ldots, n$. The model $Y = XU$ is called a multiplicative censoring model by Vardi (1989), but we emphasise that this kind of censoring is very different from more standard right-censoring. It represents incomplete observations when the lifetime at hand is only known to belong to a random time interval; this happens in AIDS studies in particular.

Model (1) has been introduced by Vardi (1989) as a common model for different statistical problems such as inference for renewal processes, deconvolution with exponential noise and density estimation under decreasing constraint. Until now, only density and survival function estimation have been studied in this model. First, a maximum likelihood estimation procedure

Date: September 9, 2016.

has been introduced by Vardi (1989) and shown to be consistent. Then, Vardi and Zhang (1992) have proved its uniform consistency and asymptotic normality, but in both papers, it is assumed that two samples $\{Y_1, ..., Y_n\}$ and $\{X_1, ..., X_m\}$ are observed and that $m/(m+n)$ converges to a positive constant $c > 0$. More recently, Asgharian et al. (2012) have proposed a kernel density estimator and established conditions for strong uniform consistency. All previous results are not applicable in our context as we assume $m = 0$.

In our setting, where X is not directly observable, no estimation procedure of the hazard rate has been proposed, to our knowledge. Nonparametric hazard rate estimation has been developed in the context of direct or right-censored observations, mainly with quotient of functional estimators, built by kernel methods as in Patil (1993a,b), wavelet strategies as in Antoniadis et al. (1999), or projection and model selection techniques as in Brunel and Comte (2005). In the present paper, our references on the topic are two studies dealing with nonparametric estimation of hazard rate in the context of direct observations or right-censored data developed in Comte et al. (2011) and in Plancade (2011). We show how to generalise the method proposed in these papers to model (1); their specificity is to propose an adaptive regression estimator built by direct contrast minimisation (no quotient) and model selection. We do not provide exhaustive bibliography on the subject, but the interested reader is referred to the recent paper of Efromovich (2016) and references therein.

Concerning the specific model considered here, we obtain the following relationship between the density f_Y of Y and the density f_X from the link between the random variables given by (1),

$$f_Y(y) = \int_y^{+\infty} \frac{f_X(x)}{x} dx, \quad y > 0.$$

This formula indicates that estimating the density of X from the density of the observed variable Y is an inverse problem. Based on this observation, Andersen and Hansen (2001) have proposed an estimation procedure of the density f_X by a series expansion approach. Convergence rates for the mean integrated squared error are derived. Van Es et al. (2005) have proposed an estimation procedure of the density of $\log(X^2)$ in the non *i.i.d.* case, under a different assumption on the law of U. Abbaszadeh et al. (2012, 2013); Chesneau (2013); Chaubey et al. (2015) have proposed adaptive wavelet estimators of the density f_X, when the observations follow related – yet different – models, for instance with an additional bias on X (Abbaszadeh et al., 2012), in the non *i.i.d.* case (Chesneau, 2013) or under the assumption that X follows a mixing distribution (Chaubey et al., 2015). They obtain convergence rates for the $\mathbb{L}^2$-risk (or even the $\mathbb{L}^p$-risk Abbaszadeh et al. 2013) on $[0, 1]$. Brunel et al. (2015) have proposed an adaptive estimation procedure for both the density f_X and the survival function $\bar{F}_X$ in the case where X can take negative values. They also obtain rates of convergence, for both integrated and pointwise quadratic risk which are similar to the ones obtained by Andersen and Hansen (2001), though under different regularity assumptions on the functions to estimate. These rates are proved to be optimal in the minimax sense. Comte and Dion (2016) have proposed an adaptive estimation procedure for the density function in a different context, the noise is supposed to be uniform over an interval $[1-a, 1+a]$ ($a > 0$). None of the previous works considers hazard rate estimation, while this function is widely used in survival analysis.

In this paper, we provide a projection strategy for the estimation of the hazard rate function h, following the ideas developed by Comte et al. (2011) and by Plancade (2011). To this end, we take into account the specific model (1) and propose an original minimum contrast estimator. We first build a collection of projection estimators over linear models, and then choose an estimate in the collection, by using model selection. In Section 2, we detail the estimation procedure for a fixed model and justify the choice of our contrast. We give theoretical results in Section 3.

In Section 4, we define the empirical criterion for choosing the model dimension and provide theoretical results (oracle-inequality and rates of convergence) for the selected estimator. Finally, in Section 5, we study the numerical behavior of the proposed estimator. Section 6 is devoted to the proofs.

2. Estimation procedure

2.1. **Notations.** We estimate the target function h on a compact subset $A = [0, \mathbf{a}]$ of $[0, +\infty[$. Let $(\mathbb{L}^2(A), \|.\|, \langle .,. \rangle)$ be the space of square integrable functions on A, equipped with its classical Hilbert structure: $\langle f, g \rangle = \int_A f(t)g(t)dt$ and $\|f\| = \sqrt{\langle f, f \rangle}$, for all $f,\ g \in \mathbb{L}^2(A)$. We also introduce $\|\cdot\|_{\bar{F}_X}$, a reference semi-norm that naturally appears in our estimation problem, given by $\|t\|^2_{\bar{F}_X} := \langle t, t \rangle_{\bar{F}_X}$, $\langle s, t \rangle_{\bar{F}_X} := \int_A s(x)t(x)\bar{F}_X(x)dx$, for $s, t \in \mathbb{L}^2(A)$. It satisfies $\|t\|_{\bar{F}_X} \leq \|t\|$. We also denote $\|f\|_{\infty,I} := \sup_{x\in I} |f(x)|$, and $\|f\|_{p,I}$ the classical $\mathbb{L}^p-$norm of a function f on an interval $I \subset \mathbb{R}$.

We consider a collection $(S_m)_{m\in\{1,\dots,N_n\}}$ of linear subspaces such that

$$S_m = \mathrm{Span}\{\varphi_j,\ j \in \mathbb{J}_m\},$$

where $\mathbb{J}_m \subset \mathbb{N}\backslash\{0\}$, $N_n \geq 1$, $\{\varphi_j,\ j \in \mathbb{J}_m\}$ is a basis of the subspace, and φ_j has support in A. We denote by D_m the dimension of S_m, which means that $D_m = |\mathbb{J}_m|$, where $|B|$ denotes the cardinality of a set B. The following properties are required for the models.

($\mathcal{M}_1$): The models are nested: $S_m \subset S_{m'}$ if $m \leq m'$.

($\mathcal{M}_2$): The basis $\{\varphi_j,\ j \in \mathbb{J}_m\}$ is a Riesz basis:

$$\exists\ \mathfrak{d}_1, \mathfrak{d}_2 > 0,\ \ \forall (\alpha_j)_{j\in\mathbb{J}_m} \in \mathbb{R}^{D_m},\quad \mathfrak{d}_1 \sum_{j\in\mathbb{J}_m} \alpha_j^2 \leq \left\| \sum_{j\in\mathbb{J}_m} \alpha_j\varphi_j \right\|^2 \leq \mathfrak{d}_2 \sum_{j\in\mathbb{J}_m} \alpha_j^2.$$

($\mathcal{M}_3$): The functions φ_j, $j \in \mathbb{J}_m$ are differentiable on A and

$$\exists \Phi_0, \Phi_1 > 0,\quad \left\| \sum_{j\in\mathbb{J}_m} \varphi_j^2 \right\|_{\infty,A} \leq \Phi_0^2 D_m,\quad \left\| \sum_{j\in\mathbb{J}_m} (\varphi_j')^2 \right\|_{\infty,A} \leq \Phi_1^2 D_m^3,$$

and $\forall (\alpha_j)_{j\in\mathbb{J}_m} \in \mathbb{R}^{D_m}$, $\| \sum_{j\in\mathbb{J}_m} \alpha_j\varphi_j' \|^2 \leq \Phi_2^2 (\sum_{j\in\mathbb{J}_m} \alpha_j^2) D_m^2$.

($\mathcal{M}_4$): One of the two following properties is satisfied.

- **($\mathcal{M}_{4,1}$):** The basis is bounded: $\exists c_0 > 0,\ \forall j \in \mathbb{J}_m,\ \|\varphi_j\|_{\infty,A} \leq c_0$.
- **($\mathcal{M}_{4,2}$):** The two following assertions are fulfilled.
  - (i) $\exists c_1 > 0,\ \forall j \in \mathbb{J}_m,\ |\{k \in \mathbb{J}_m,\ \varphi_j\varphi_k \neq 0\}| \leq c_1$,
  - (ii) $\exists c_2 > 0,\ \sup_{j\in\mathbb{J}_m} \|\varphi_j\|^2_{\infty,A} \leq c_2 D_m$.

($\mathcal{M}_5$): There exists $c_3 > 0$ such that for any $f \in \mathbb{L}^2(A)$, $\sum_{j\in\mathbb{J}_m} \langle f, \varphi_j \rangle^2 \leq c_3 \|f_m\|^2$, with f_m the orthogonal projection of f onto S_m.

Assumption ($\mathcal{M}_1$) implies that we have $S_m \subset S_{N_n} =: \mathcal{S}_n$ for all $m = 1, \dots, N_n$. Assumption ($\mathcal{M}_2$) generalises the case of orthogonal bases. The differentiability of the basis (Assumption ($\mathcal{M}_3$)) is required for our estimation strategy (see (6) below). Clearly, it excludes the classical histograms and piecewise polynomial basis. The inequalities for the sup-norm of the sum of the functions of the basis and their derivatives are classical. A basis which satisfies ($\mathcal{M}_{4,2}$) (i) and (ii) is said to be localised. We have in mind at least two types of models that are convenient for this set of assumptions.

[T] - Trigonometric models. Let $\mathbb{J}_m = \{1, \dots, 2m+1 = D_m\}$, $\varphi_1(x) = \mathbf{a}^{-1/2}\mathbf{1}_A(x)$, $\varphi_{2j}(x) = \mathbf{a}^{-1/2}\mathbf{1}_A(x)\sqrt{2}\cos(2\pi jx/\mathbf{a})$ and $\varphi_{2j+1}(x) = \mathbf{a}^{-1/2}\mathbf{1}_A(x)\sqrt{2}\sin(2\pi jx/\mathbf{a})$.

We obtain that Assumptions $(\mathcal{M}_l)$, $l = 1, \dots, 4$ are satisfied, with $\mathfrak{d}_1 = \mathfrak{d}_2 = 1$ in $(\mathcal{M}_2)$ (orthonormal basis), $\Phi_0^2 = 1/\mathbf{a}$, $\Phi_1^2 = \pi^2/\mathbf{a}^3$, $\Phi_2^2 = 4\pi^2/\mathbf{a}^2$ in $(\mathcal{M}_3)$, $c_0 = \sqrt{2/\mathbf{a}}$ in $(\mathcal{M}_{4,1})$, and $c_3 = 1$ in $(\mathcal{M}_5)$ (see Section 6.1).

[BS] - B-spline models. Let $r \in \mathbb{N}\backslash\{0\}$, $\mathbb{J}_m = \{-r+1, 2, \dots, 2^m - 1\}$, in such a way that $D_m = 2^m - r + 1$ and

$$\varphi_j(x) := \varphi_{j,m}(x) = \frac{2^{m/2}}{\sqrt{\mathbf{a}}} N_r\left(\frac{2^m}{\mathbf{a}}x - j\right), \quad N_r(x) = r[0, 1, \dots, r](\cdot - x)_+^{r-1} = \mathbf{1}_{[0,1]}^{\star r}(x), \tag{2}$$

with the usual difference notation, see de Boor (1978); DeVore and Lorentz (1993), and where $\star$ is the convolution product. Assumptions $(\mathcal{M}_l)$, $l = 1, \dots, 4$ are satisfied, with $\Phi_0^2 = r/\mathbf{a}$, $\Phi_1^2 = 4(r-1)^4 r/\mathbf{a}^3$, and $\Phi_2^2 = (2r-1)\|N_r'\|_{2,\mathbb{R}}^2/\mathbf{a}^2$ in $(\mathcal{M}_3)$, $c_1 = 2r - 1$, $c_2 = \mathbf{a}^{-1}$ in $(\mathcal{M}_{4,2})$, and $c_3 = \|N_r\|_{2,\mathbb{R}}^4(2r-1)^2/\mathfrak{d}_1$ in $(\mathcal{M}_5)$.

Details of the computation of the constants can be found in Section 6.1, as well as additional technical properties. The following constraint is also required on the maximal model dimension D_{N_n} for a constant $K > 0$:

$$D_{N_n} \leq K\sqrt{n/\ln^3(n)}. \tag{3}$$

2.2. **Definition of the regression contrast.** The key of our estimation strategy relies on the following equalities, obtained from (1):

$$yf_Y'(y) = -f_X(y), \tag{4}$$

and

$$\bar{F}_Y(y) + yf_Y(y) = \bar{F}_X(y), \; y \in \mathbb{R}_+, \tag{5}$$

see Section 6.1 for details. These relations lead to consider the following contrast, for any differentiable function $t \in \mathbb{L}^2(A)$:

$$\gamma_n(t) = \|t\|_n^2 - 2\tilde{\nu}_n(t), \tag{6}$$

with $\|.\|_n$ an empirical semi-norm and $\tilde{\nu}_n$ an empirical process defined by

$$\|t\|_n^2 = \tfrac{1}{n}\textstyle\sum_{i=1}^n \left[\int_A t^2(x)\mathbf{1}_{\{Y_i \geq x\}}dx + Y_i t^2(Y_i)\mathbf{1}_{\{Y_i \in A\}}\right],$$
$$\tilde{\nu}_n(t) = \tfrac{1}{n}\textstyle\sum_{i=1}^n \left[t(Y_i) + Y_i t'(Y_i)\right]\mathbf{1}_{\{Y_i \in A\}} - \mathbf{a}t(\mathbf{a})\widehat{f}_Y(\mathbf{a}),$$

where $\widehat{f}_Y$ is an estimator of the density f_Y. Its properties will be specified later (see Section 2.3). Using (4) and (5), we prove in Section 6.2 the following result.

Lemma 2.1. *Let $t \in \mathbb{L}^2(A)$ be a bounded and continuously differentiable function on $\mathbb{R}_+$. Then*

$$\mathbb{E}\left[\gamma_n(t)\right] = \|t - h\|_{\bar{F}_X}^2 - \|h\|_{\bar{F}_X}^2 + 2\mathbf{a}t(\mathbf{a})\left(\mathbb{E}\left[\widehat{f}_Y(\mathbf{a})\right] - f_Y(\mathbf{a})\right). \tag{7}$$

The unusual term $\mathbf{a}t(\mathbf{a})\widehat{f}_Y(\mathbf{a})$ involved in the contrast permits to counterbalance the term $\mathbf{a}t(\mathbf{a})f_Y(\mathbf{a})$ which appears when we calculate $\mathbb{E}[\tilde{\nu}_n(t)]$, as (7) proves it. In practice (see Section 5), the value of $\widehat{f}_Y(\mathbf{a})$ is very small. Moreover, it follows from (5) that $\lim_{\mathbf{a}\to+\infty}\mathbf{a}f_Y(\mathbf{a}) = \lim_{\mathbf{a}\to+\infty}[\bar{F}_X(\mathbf{a}) - \bar{F}_Y(\mathbf{a})] = 0$. The term is thus intrinsically small when $\mathbf{a}$ gets large. This is why the last righ-hand-side term of (7) can be regarded as negligible.

Then, it follows from Lemma 2.1 that minimizing the contrast over S_m will amount, for n large enough, to find the function t in S_m that minimises the specific distance between t and h, $\|t - h\|_{\bar{F}_X}$. Under (H_2) specified below (see Section 2.3), the norm weighted by $\bar{F}_X$ is equivalent

to the $\mathbb{L}^2-$norm and the resulting estimate will approximate h_m, the orthogonal projection of h on S_m. Consequently, for m large enough, it should be a good approximation of h.

2.3. **Definition of hazard rate estimators.** Concerning the random variables appearing in the model, the following assumptions will be required.

(H_1): f_X is bounded on A.
(H_2): $\inf_{x\in A}\bar{F}_Y(x) =: \bar{F}_0 > 0$, $f_Y \in \mathbb{L}^2(A)$ and $\mathbb{E}[Y_1^2] < +\infty$.

Since A is bounded (recall that $A = [0, \mathbf{a}]$), the first part of Assumption (H_2) can hold (it does not hold on $\mathbb{R}_+$). It means also that the survival function $\bar{F}_X$ is also lower bounded, thanks to (5). Moreover, under (H_1) and (H_2), $\|h\|_A^2 \leq \|f_X\|_{\infty,A}^2/(\bar{F}_0)^2$ (which implies that $h \in \mathbb{L}^2(A)$) and the norms $\|.\|$ and $\|.\|_{\bar{F}_X}$ introduced in Section 2.1 are equivalent:

$$\forall t \in \mathbb{L}^2(A),\ \bar{F}_0^{-1}\|t\|^2 \leq \|t\|_{\bar{F}_X}^2 \leq \|t\|^2. \tag{8}$$

This explains why the quantity $\bar{F}_0$ plays an important role in the estimation procedure.

Now, we need to complete the definition of the estimators. It is easily seen that if $t = \sum_{j\in\mathbb{J}_m}\alpha_j\varphi_j$ is a minimiser of γ_n over S_m, it verifies

$$\frac{\partial\gamma_n(\sum_{j\in\mathbb{J}_m}\alpha_j\varphi_j)}{\partial\alpha_{j_0}} = 0,\ \ j_0 \in \mathbb{J}_m.$$

Let us denote by $\vec{\alpha} = {}^t(\alpha_j)_{j\in\mathbb{J}_m} \in \mathbb{R}^{D_m}$, the vector of the coefficients of t. The latter conditions are equivalent to

$$\widehat{\Psi}_m\widehat{\vec{\alpha}} = \widehat{\vec{b}},\ \ \text{with}\ \ \widehat{\vec{b}} = {}^t\left(\frac{1}{n}\sum_{i=1}^n(\varphi_j(Y_i) + Y_i\varphi_j'(Y_i))\mathbf{1}_{\{Y_i\in A\}} - \mathbf{a}\widehat{f}_Y(\mathbf{a})\varphi_j(\mathbf{a})\right)_{j\in\mathbb{J}_m},$$

and

$$\widehat{\Psi}_m := (\Psi_{j,k})_{j,k\in\mathbb{J}_m} = \left(\frac{1}{n}\sum_{i=1}^n\left\{\int_A\varphi_j(x)\varphi_k(x)\mathbf{1}_{Y_i\geq x}dx + Y_i\varphi_j(Y_i)\varphi_k(Y_i)\mathbf{1}_{\{Y_i\in A\}}\right\}\right)_{j,k\in\mathbb{J}_m}.$$

Thus, a unique minimiser of γ_n exists in S_m if and only if the matrix $\widehat{\Psi}_m$ is invertible. This is the case on the following set $\widehat{\Delta}_{\rho_1}^m$, defined for $0 < \rho_1 < 1$ and for $m \in \{1, \ldots, N_n\}$,

$$\widehat{\Delta}_{\rho_1}^m = \left\{\min\mathrm{Sp}(\widehat{\Psi}_m) \geq (1-\rho_1)\mathfrak{d}_1\widehat{\bar{F}}_0\right\},$$

where $\mathrm{Sp}(M)$ denotes the set of eigenvalues of a matrix M, $\widehat{\bar{F}}_0$ is an estimator of $\bar{F}_0$ and $\mathfrak{d}_1$ is defined in Property ($\mathcal{M}_2$). Our estimator is thus given by

$$\widehat{h}_m = \sum_{j\in\mathbb{J}_m}\widehat{\alpha}_j\varphi_j,\ \text{where}\ \widehat{\vec{\alpha}} = {}^t(\widehat{\alpha}_j)_{j\in\mathbb{J}_m} = \begin{cases}\widehat{\Psi}_m^{-1}\widehat{\vec{b}} & \text{on } \widehat{\Delta}_{\rho_1}^m\\ 0 & \text{on } (\widehat{\Delta}_{\rho_1}^m)^c\end{cases}.$$

Its definition depends on the estimator $\widehat{\bar{F}}_0$ of $\bar{F}_0$, the value of the survival function at the point $\mathbf{a}$ (note that $\bar{F}_0 > 0$ under (H_2)), and the estimator $\widehat{f}_Y(\mathbf{a})$ of $f_Y(\mathbf{a})$, the value of the density f_Y at the same point $\mathbf{a}$. In the sequel, we propose to choose

$$\widehat{\bar{F}}_0 = \frac{1}{n}\sum_{i=1}^n\mathbf{1}_{\{Y_i\geq\mathbf{a}\}} + \frac{1}{\sqrt{n}}, \tag{9}$$

and a kernel density estimator with form

$$\widehat{f}_Y(\mathbf{a}) = \frac{1}{n\widehat{\mathfrak{w}}} \sum_{i=1}^{n} K\left(\frac{\mathbf{a} - Y_i}{\widehat{\mathfrak{w}}}\right), \tag{10}$$

for a kernel $K : \mathbb{R} \to \mathbb{R}_+$ with a well-chosen data-driven bandwidth $\widehat{\mathfrak{w}} > 1/\sqrt{n}$ (see details below for the bandwidth choice). Alternative possibilities can be considered, they are also discussed below, Section 2.2 for $\widehat{f}_Y(\mathbf{a})$ and Section 6.1 for $\widehat{\bar{F}}_0$.

3. Upper-bound on the risk of the estimator

With previous definitions and assumptions, we prove the following result.

Proposition 3.1. *Assume (H_1) and (H_2). Then the estimator $\widehat{h}_m$ satisfies*

$$\mathbb{E}\left[\|\widehat{h}_m - h\|^2\right] \leq C_1\left(\|h_m - h\|^2 + V(m)\right) + C_2 \frac{\mathbf{a}^2\Phi_0^2}{\mathfrak{d}_1\bar{F}_0} D_m \mathbb{E}\left[\left(\widehat{f}_Y(\mathbf{a}) - f_Y(\mathbf{a})\right)^2\right] + \frac{C_3}{n}$$

with C_1, C_2 and C_3 some constants which do not depend on m and n, and

$$V(m) = \frac{1}{\mathfrak{d}_1\bar{F}_0}\left((1 + \mathbf{a}\|h\|^2)\Phi_0^2\frac{D_m}{n} + \mathbf{a}^2\Phi_1^2\frac{D_m^3}{n}\right).$$

The proof of Proposition 3.1 and all other results are deferred to Section 6.

The upper bound given in Proposition 3.1 contains a squared-bias term $\|h - h_m\|^2$ which is decreasing when m increases, and a variance term $V(m)$ which, on the contrary, increases with m. The last term C_3/n is a negligible residual term. It remains one term, which involves the pointwise risk of $\widehat{f}_Y$ at the point $\mathbf{a}$. Corollary 3.1 below proves that the convergence rate of the risk is not deteriorated by this additional term, under reasonable smoothness assumptions.

We introduce $\mathcal{H}^\beta(I, L)$ the Hölder ball with radius $L > 0$ and smoothness $\beta > 0$ for functions defined on an interval I of $\mathbb{R}$:

$$\begin{aligned}\mathcal{H}^\beta(I, L) \quad = \quad & \{f : I \to \mathbb{R}, \text{ such that } f \text{ is } \ell = \lfloor\beta\rfloor - \text{times differentiable} \\ & \text{and } |f^{(\ell)}(x) - f^{(\ell)}(y)| \leq L|x - y|^{\beta-\ell} \text{ for any } x, y \in I\Big\},\end{aligned}$$

where $\lfloor\beta\rfloor$ is the largest integer less than β.

We also choose $\widehat{\mathfrak{w}}$ in the definition (10) with a Goldenshluger-Lepski type method over a collection of possible bandwidths larger than $1/\sqrt{n}$, as described in Comte (2015) or Rebelles (2015). This permits to derive the following convergence rate.

Corollary 3.1. *Consider models* [T] *or* [BS] *and let the assumptions of Proposition 3.1 hold. If the hazard rate h belongs to the ball $\mathcal{H}^\beta([0, \mathbf{a} + \varepsilon], L)$ (for some fixed $\varepsilon > 0$), then there exists an index $m(\beta)$ in the collection $\mathcal{M}_n$ and a choice $\widehat{\mathfrak{w}} = \mathfrak{w}(\beta)$, such that*

$$\sup_{h \in \mathcal{H}^\beta([0, \mathbf{a}+\varepsilon], L)} \mathbb{E}\left[\|\widehat{h}_{m(\beta)} - h\|^2\right] \leq Cn^{-2\beta/(2\beta+3)},$$

for a constant C, which does not depend on m neither on n.

This result requires some comments. The convergence rate $n^{-2\beta/(2\beta+3)}$ is the optimal minimax one obtained for the estimation of the density of X in the same model by Brunel et al. (2015). This rate is obtained by computing the optimal bias-variance trade-off $\min_{m \in \mathcal{M}_n} \|h_m - h\|^2 + V(m)$ that appears in the right-hand-side of the result of Proposition 3.1. This proves that

the additional term $D_m\mathbb{E}[(\widehat{f}_Y(\mathbf{a})-f_Y(\mathbf{a}))^2]$ of the inequality does not degrade the rate, at the price of a slightly stronger smoothness condition. Indeed, the assumption that h belongs to $\mathcal{H}^\beta([0,\mathbf{a}+\varepsilon],L)$ permits to state that f_Y belongs to $\mathcal{H}^{\beta+1}([\mathbf{a}-\varepsilon',\mathbf{a}+\varepsilon'],L')$ for an $L'>0$ and a $\varepsilon'\in]0,\mathbf{a}[$ (see Proposition 6.1). This smoothness result ensures that the kernel estimate $\widehat{f}_Y$ has a pointwise risk upper-bounded by $C(n/\ln(n))^{-2(\beta+1)/(2\beta+3)}$ (C a constant), for a "hand-chosen" bandwidth $\widehat{\mathfrak{w}}\asymp n^{-1/(2\beta+1)}$, see Rebelles (2015) for example. Thus, as D_m is set to $D_{m_{\text{opt}}}\asymp n^{1/(2\beta+3)}$, we get $D_{m_{\text{opt}}}\mathbb{E}[(\widehat{f}_Y(\mathbf{a})-f_Y(\mathbf{a}))^2]\leq n^{-2\beta/(2\beta+3)}$. Notice also that any other estimator for $\widehat{f}_Y(\mathbf{a})$ that is adaptive (that is built without the knowledge of β) and has a pointwise risk upper-bounded as previously can be used in the procedure.

4. Model selection

It follows from Proposition 3.1 that m has to be chosen in order to make a good squared-bias/variance compromise, and the aim of the present section is to provide a criterion to select an estimator among the collection $(\widehat{h}_m)_{m\in\{1,\dots,N_n\}}$.

We follow the general model selection scheme developed by Birgé and Massart (1998) (see also Massart 2007). The idea is to estimate the two terms involved in the compromise. For the squared-bias, we have $\|h-h_m\|^2=\|h\|^2-\|h_m\|^2$ by Pythagoras Theorem. Then, as $\|h\|^2$ is a constant (though unknown), we only need to estimate $-\|h_m\|^2$ and we replace it by $-\|\widehat{h}_m\|_n^2=\gamma_n(\widehat{h}_m)$. For the variance, we mainly use the bound $V(m)$, up to a multiplicative constant κ. The theoretical study gives a value for this constant, but in an elaborate problem like hazard rate regression, this value is too large. In practice, it is calibrated on preliminary simulations, or by using specific numerical methods as described in Baudry et al. (2012).

Thus, the dimension is selected as follows

$$\widehat{m}\in\arg\min_{m\leq N_n,D_m\geq\ln(n)}\left\{\gamma_n(\widehat{h}_m)+\widehat{\text{pen}}(m)\right\},\tag{11}$$

with

$$\widehat{\text{pen}}(m):=\frac{\kappa}{\mathfrak{d}_1\widehat{\bar{F}}_0n}\left(\Phi_0^2D_m+(\mathbf{a}\Phi_0^2+\mathbf{a}^2\Phi_1^2)D_m^3\right)$$

where $\kappa>0$ is a universal constant. We denote by $\text{pen}(m)$ the theoretical version of $\widehat{\text{pen}}(m)$,

$$\text{pen}(m):=\frac{\kappa}{\mathfrak{d}_1\bar{F}_0n}\left(\Phi_0^2D_m+(\mathbf{a}\Phi_0^2+\mathbf{a}^2\Phi_1^2)D_m^3\right).$$

Then we obtain the following result.

Theorem 4.1. *Assume (H_1) and (H_2). Then there exists $\kappa_0>0$ and $n_1(h,\rho_0)\in\mathbb{N}\backslash\{0\}$ depending on $\|h\|^2$ and ρ_0, such that, for any $\kappa\geq\kappa_0$, and $n\geq n_1(h,\rho_0)$,*

$$\begin{aligned}\mathbb{E}\left[\left\|\widehat{h}_{\widehat{m}}-h\right\|^2\right] &\leq C_1'\min_{m=1,\dots,N_n,\ D_m\geq\ln(n)}\left\{\inf_{t\in S_m}\|h-t\|^2+\text{pen}(m)\right\}\\ &\quad+C_2'\mathbb{E}\left[(D_{\widehat{m}}\vee D_m)\left(\widehat{f}_Y(\mathbf{a})-f_Y(\mathbf{a})\right)^2\right]+\frac{C_3'}{n},\end{aligned}$$

where C_1',C_2',C_3' are positive constants.

The proof of Theorem 4.1 is given in Section 6. Notice that we restrict the model collection to spaces S_m with dimension larger than $\ln(n)$ for technical purposes.

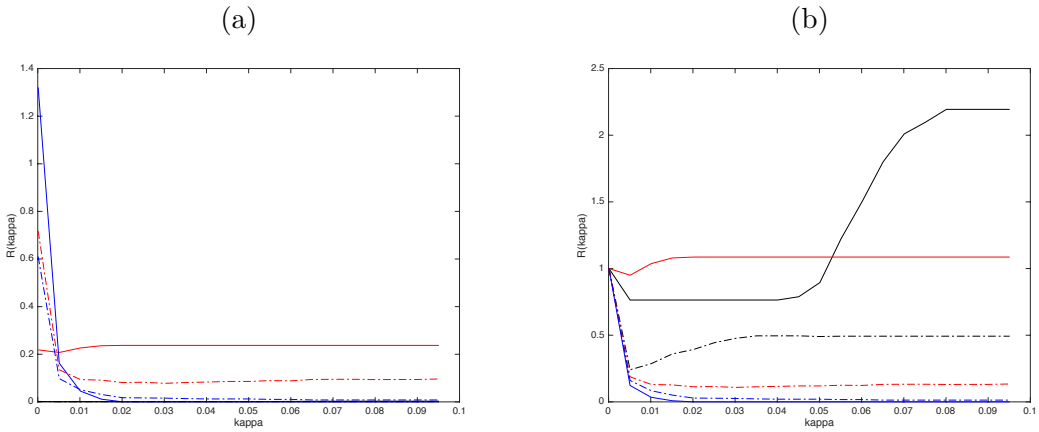

FIGURE 1. Plot of the mean risk $\widehat{R}$ defined by (12) (Figure (a)) and the mean risk renormalised $\widehat{R}/\widehat{R}(\kappa_{\min})$ (Figures (b)) as a function of κ ($\kappa_{\min} = 10^{-4}$) for the estimator calculated on the basis [T]. Black curves: Gamma distribution, red curves: rescaled Beta distribution, blue curves: exponential distribution, solid lines: $n = 2000$, dotted lines: $n = 200$.

The risk bound stated in Theorem 4.1 shows that the selected $\widehat{m}$ leads to an estimator that automatically performs the squared-bias variance compromise through the term

$$T_1 := \min_{m=1,\dots,N_n,\ D_m \geq \ln(n)} \left\{ \inf_{t \in S_m} \|h - t\|^2 + \mathrm{pen}(m) \right\}$$

up to the term

$$T_2 := \mathbb{E}\left[(D_{\widehat{m}} \vee D_m) \left(\widehat{f}_Y(\mathbf{a}) - f_Y(\mathbf{a}) \right)^2 \right]$$

which is more difficult to evaluate in this setting. We discussed in the previous section the fact that, if $D_{\widehat{m}}$ has the order of the dimension minimizing T_1, then the best possible rate can be obtained, related to the regularity assumptions. The residual term C_3'/n is negligible.

5. Numerical illustration

5.1. **Calibration of the constant appearing in the penalty term.** In order to calibrate the constant κ appearing in the penalty term, we classically calculate the risk of the estimator as a function of κ for different simulated frameworks. Indeed, this constant is universal and does not depend on the simulated samples. It depends only on the choice of the projection basis. We investigate the case of the Fourier basis and the B-spline basis, described in Section 2.1. Hence, we fix κ once and for all for these two kinds of models. Such a method has been used for the same purpose for instance in Bertin et al. (2016). The different models we consider are the following: Gamma distribution $X \sim \Gamma(8, 4)$, Rescaled Beta distribution $X = 5X'$, $X' \sim \mathrm{Beta}(3, 3)$, Exponential distribution $X \sim \mathcal{E}(2)$ the exponential distribution with mean 2. To avoid overfitting, the densities used for calibration are different from the ones used to study the performances of the estimation procedure, in section 5.2.

For each distribution, we consider two sample sizes ($n = 200$ and $n = 2000$). Hence, we plot 6 risk curves. The considered risk is the one associated with the $\mathbb{L}^2$-loss on $[0, \mathbf{a}]$ that we

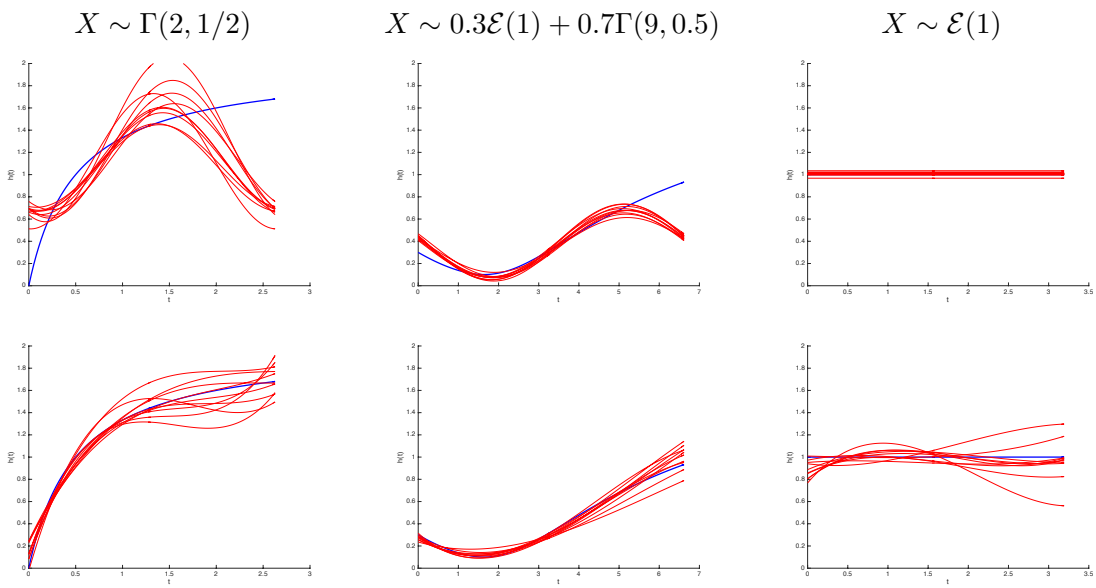

FIGURE 2. Plot of 10 estimators (thin lines), $n = 2000$. The bold line represents the function h to estimate. First line: Fourier basis, second line: spline basis.

	$X \sim \Gamma(2, 1/2)$		$X \sim 0.3\mathcal{E}(1) + 0.7\Gamma(9, 0.5)$		$X \sim \mathcal{E}(1)$	
	$n = 200$	$n = 2000$	$n = 200$	$n = 2000$	$n = 200$	$n = 2000$
Fourier basis	0.25	0.095	0.022	0.011	0.04	0.015
	(0.23)	(0.018)	(0.012)	(0.003)	(0.08)	(0.025)
spline basis	0.17	0.06	0.013	0.0015	0.11	0.034
	(0.22)	(0.05)	(0.016)	(0.0020)	(0.16)	(0.030)

TABLE 1. Mean risk with standard deviations in parenthesis, over 50 Monte Carlo replications.

approximate by the Monte Carlo method over $N = 50$ samples

$$\widehat{R}(\kappa) := \frac{1}{N} \sum_{l=1}^{N} \left\| \widehat{h}^{(l)}_{\widehat{m}(\kappa)} - h \right\|^2, \tag{12}$$

where $\widehat{h}^{(l)}_{\widehat{m}(\kappa)}$ is the estimator calculated from the l-th sample with $\widehat{m}(\kappa)$ selected from the criterion (11) with the given value of κ. The $\mathbb{L}^2$-distance $\|\cdot\|$ is approximated with the trapezoidal rule.

The results are presented in Figure 1 for the Fourier basis [T]. We choose a value of κ which seems to give correct results for all the models, namely $\kappa = 0.03$ here. An identical procedure has been applied for the splines basis [BS] of degree $r = 4$, which leads us to the choice $\kappa = 0.002$.

5.2. **Estimation procedures.** We considered three distributions for the simulations: a Gamma density $X \sim \Gamma(2, 1/2)$, a mixed Gamma density $X \sim 0.3\mathcal{E}(1) + 0.7\Gamma(9, 0.5)$ and a simple exponential density $X \sim \mathcal{E}(1)$. In Figure 2, we see that the estimation procedure, and in particular the model selection criterion, is very stable. Moreover, as could be expected, the spline estimators seem to have better performances than the one defined on the Fourier basis, except for the case where X follows the exponential distribution: the hazard rate is constant in this case and hence belongs to S_m, for all m, if the Fourier basis is chosen. For the other cases, the

good behavior of the spline basis can be explained by its local feature (see property $(\mathcal{M}_{4,2})$). In particular, it permits to avoid the usual side effects of the Fourier basis, which are observed when the hazard rate is not periodic. The results of Table 1 confirm these observations and indicate that the spline basis seems to be more stable and more efficient than the Fourier basis in most cases.

6. Proofs

6.1. **Preliminary remarks, specific notations.** We first state some intermediate equalities which will be useful in the sequel.

About the semi-norms involved in the problem. Three Hilbert semi-norms are involved in our problem: the classical Hilbert norm on the set A, denoted by $\|.\|$, the semi-norm $\|.\|_{\bar{F}_X}$ weighted by $\bar{F}_X$ on A and the empirical semi-norm $\|.\|_n$ involved in the contrast (see Section 2.2). The "empirical semi-scalar product" associated to $\|.\|_n$ is defined by

$$\langle t,s\rangle_n = \frac{1}{n}\sum_{i=1}^{n}\left(\int_A t(x)s(x)\mathbf{1}_{\{Y_i\leq x\}}dx + Y_i t(Y_i)s(Y_i)\mathbf{1}_{\{Y_i\in A\}}\right),$$

for all $s,t\in S_m$. Note that, from (5), $\mathbb{E}[\langle s,t\rangle_n] = \langle s,t\rangle_{\bar{F}_X}$. Finally, if

$$B_m = \{t\in S_m,\ \|t\|_{\bar{F}_X}\leq 1\} \tag{13}$$

is the unit ball of $(S_m,\|.\|_{\bar{F}_X})$ and if (H_2) is satisfied, we have, using (8),

$$B_m \subset \{t\in S_m,\ \|t\|\leq \bar{F}_0^{-1/2}\}. \tag{14}$$

About the models. We first justify that the models spanned by the trigonometric basis and the B-spline basis, as defined in Section 2.1 satisfy properties $(\mathcal{M}_l)$, $l=1,\dots,5$. For the Fourier basis, it is straightforward. The basis is orthogonal, which means that equality holds in $(\mathcal{M}_2)$ and $(\mathcal{M}_5)$ ($\mathfrak{d}_1=\mathfrak{d}_2=c_3=1$). Notice only that the second inequality of Assumption $(\mathcal{M}_3)$ comes from the fact that $\varphi'_{2k}(x) = -2\pi k\varphi_{2k+1}(x)/\mathbf{a}$ and $\varphi'_{2k+1}(x) = 2\pi k\varphi_{2k}(x)/\mathbf{a}$, for all $k\geq 1$. For the third one, we have

$$\begin{aligned}
\left\|\sum_{j\in\mathbb{J}_m}\alpha_j\varphi'_j\right\|^2 &= \left\|\sum_{k=1}^{m}\alpha_{2k}\varphi'_{2k}+\alpha_{2k+1}\varphi'_{2k+1}\right\|^2,\\
&= \left\|\sum_{k=1}^{m}\frac{2k\pi}{\mathbf{a}}\left(-\alpha_{2k}\varphi_{2k+1}+\alpha_{2k+1}\varphi_{2k}\right)\right\|^2,\\
&= \sum_{k=1}^{m}\frac{4k^2\pi^2}{\mathbf{a}^2}\left(\alpha_{2k}^2+\alpha_{2k+1}^2\right)\leq\frac{4\pi^2}{\mathbf{a}^2}D_m^2\left(\sum_{j\in\mathbb{J}_m}\alpha_j^2\right),
\end{aligned}$$

which ends to prove $(\mathcal{M}_3)$, with $\Phi_2^2 = 4\pi^2/\mathbf{a}^2$. Let us prove the properties for the B-spline models. We begin with the following lemma dealing with the basic property of the B-spline N_r.

Lemma 6.1. *Let $r\in\mathbb{N}\backslash\{0\}$, and $N_r = \mathbf{1}_{[0,1]}^{\star r}$. Then,*

(i) *N_r has support $[0;r]$, and $N_r(x) = \sum_{j=0}^{k}(-1)^j\binom{r}{j}(x-j)^{r-1}/(r-1)!$ if $x\in]k-1,k]$, $k=1,\dots,r$.*
(ii) *$\|N_r\|_{\infty,[0;r]}\leq 1$.*
(iii) *$\|N'_r\|_{\infty,[0;r]}\leq 2(r-1)^2$.*

Proof of Lemma 6.1. Property (i) can be found in Formula (9) of Killmann and von Collani (2001). Property (ii) is straightforward. Let us focus on (iii). Let $k \in [0;r]$ be fixed. On the interval $[k-1,k]$, N_r is a polynomial function with degree $r-1$ (see (i)), and $P_{r,k} : x \mapsto N_r(x/2+k-1/2)$, is a polynomial over $[-1,1]$ with the same degree. Thus, thanks to the Markov inequality (Theorem 1.4 in Chapter 4 of DeVore and Lorentz 1993), $\|P'_{r,k}\|_{\infty,[-1,1]} \leq (r-1)^2\|P_{r,k}\|_{\infty,[-1,1]}$. This means that

$$\|N'_r\|_{\infty,[k-1,k]} = 2\|P'_{r,k}\|_{\infty,[-1,1]} \leq 2(r-1)^2\|P_{r,k}\|_{\infty,[-1,1]} = 2(r-1)^2\|N_r\|_{\infty,[k-1,k]} \leq 2(r-1)^2.$$

This inequality is valid for any k, which ends the proof of Lemma 6.1. □

It follows from Lemma 6.1 that φ_j defined by (2) has support $[2^{-m}\mathbf{a}j, 2^{-m}\mathbf{a}(j+r)[$. Assumption $(\mathcal{M}_1)$ for the spline basis follows from the dyadic normalisation in the definition of φ_j (the models define a multiresolution analysis, see Härdle et al. 1998). From Proposition 6.2 of Härdle et al. (1998), $(N_r(\cdot - j))_{j\in\mathbb{Z}}$ is a Riesz basis: let $\mathfrak{d}_1, \mathfrak{d}_2 > 0$, such that for any finite subset $\Lambda \subset \mathbb{Z}$, and real numbers $(\alpha_j)_{j\in\Lambda}$,

$$\mathfrak{d}_1 \sum_{j\in\Lambda} \alpha_j^2 \leq \left\| \sum_{j\in\Lambda} \alpha_j N_r(\cdot - j) \right\|_{2,\mathbb{R}}^2 \leq \mathfrak{d}_2 \sum_{j\in\Lambda} \alpha_j^2.$$

Since $\|\sum_{j\in\mathbb{J}_m} \alpha_j\varphi_j\|^2 = \|\sum_{j\in\mathbb{J}_m} \alpha_j N_r(\cdot - j)\|_{2,\mathbb{R}}^2$ (by setting the change of variable $y = 2^m x/\mathbf{a}$ in the integrals that define the norms), we obtain Assumption $(\mathcal{M}_2)$. Let us now prove Assumption $(\mathcal{M}_3)$. Let $x \in A$ be fixed. There exists a unique $k \in \{0,\dots,2^m-1\}$ such that $x \in [k2^{-m}\mathbf{a}, (k+1)2^{-m}\mathbf{a}[$. For such a x, $\varphi_j(x) \neq 0$ only if

$$[2^{-m}\mathbf{a}j, 2^{-m}\mathbf{a}(j+r)[\cap[k2^{-m}\mathbf{a}, (k+1)2^{-m}\mathbf{a}[\neq \emptyset$$

that is if $j \in \{k-r+1,\dots,k\} = A_k$, and we have $|\varphi_j(x)| \leq 2^{m/2}\mathbf{a}^{-1/2}\|N_r\|_{\infty,[0,r]}$ and $|\varphi'_j(x)| \leq 2^{3m/2}\mathbf{a}^{-3/2}\|N'_r\|_{\infty,[0,r]}$. Since A_k has cardinality r, we deduce from (ii) and (iii) of Lemma 6.1

$$\sum_{j\in\mathbb{Z}} \varphi_j^2(x) \leq r \times 2^m\mathbf{a}^{-1}, \quad \sum_{j\in\mathbb{Z}} (\varphi'_j)^2(x) \leq r \times 2^{3m}\mathbf{a}^{-3}4(r-1)^4,$$

which ends to prove the second inequality of $(\mathcal{M}_3)$. For the third one,

$$\begin{aligned}
\left\| \sum_{j\in\mathbb{J}_m} \alpha_j\varphi'_j \right\|^2 &= \sum_{j,k\in\mathbb{J}_m} \alpha_j\alpha_k \int_A \varphi'_j(x)\varphi'_k(x)dx, \\
&= \sum_{j,k\in\mathbb{J}_m} \alpha_j\alpha_k \int_A \frac{2^{3m}}{\mathbf{a}^3} N'_r\left(\frac{2^m x}{\mathbf{a}} - j\right) N'_r\left(\frac{2^m x}{\mathbf{a}} - k\right) dx, \\
&= \frac{2^{2m}}{\mathbf{a}^2} \sum_{j,k\in\mathbb{J}_m} \alpha_j\alpha_k \int_{\mathbb{R}} N'_r(u-j)N'_r(u-k)du, \\
&= \frac{2^{2m}}{\mathbf{a}^2} {}^t\vec{\alpha}U_m\vec{\alpha},
\end{aligned}$$

where U_m is the positive symmetric matrix of size D_m with coefficients $[U_m]_{j,k} = \langle N'_r(\cdot - j), N'_r(\cdot - k)\rangle_{\mathbb{R}}$ (usual scalar product of $\mathbb{L}^2(\mathbb{R})$). If $\rho(U_m)$ is the spectral radius of U_m, we have $\rho(U_m) = \max_{\vec{\alpha}\neq 0} {}^t\vec{\alpha}U_m\vec{\alpha}/\|\vec{\alpha}\|_{\ell^2}^2$, where $\|\cdot\|_{\ell^2}$ is the classical Euclidean norm on $\mathbb{R}^{D_m}$. This leads

to

$$\left\| \sum_{j\in\mathbb{J}_m} \alpha_j \varphi_j' \right\|^2 \quad \leq \quad \frac{2^{2m}}{\mathbf{a}^2} \left(\sum_{j\in\mathbb{J}_m} \alpha_j^2 \right) \rho(U_m).$$

Now, U_m is a Toeplitz matrix: $[U_m]_{j+l,k+l} = [U_m]_{j,k}$ for any indices $j,k \in \mathbb{J}_m$ and l such that $j+l, k+l \in \mathbb{J}_m$, which also satisfies $[U_m]_{j,k} = 0$ if $|j-k| \geq r$. Applying a result of Gray (2006), we thus have

$$\rho(U_m) \leq \sum_{k\in\mathbb{Z}} [U_m]_{0,k} = \sum_{|k|\leq r} |\langle N_r', N_r'(\cdot - k)\rangle_{\mathbb{R}}| \leq \sum_{|k|\leq r} \|N_r'\|_{2,\mathbb{R}}^2 = (2r-1)\|N_r'\|_{2,\mathbb{R}}^2.$$

This ends to prove $(\mathcal{M}_3)$ with $\Phi_2^2 = (2r-1)\|N_r'\|_{2,\mathbb{R}}^2/\mathbf{a}^2$.

For $(\mathcal{M}_4)$, notice that $\varphi_j(x)\varphi_k(x) = 0$ as soon as $k \geq j+r$ or $k+r \leq j$ for any $x \in \mathbb{R}$. Thus, $(\mathcal{M}_{4,2})$ (i) holds, with $c_1 = 2r-1$. Moreover, thanks to (ii) of Lemma 6.1, for any $j \in \mathbb{J}_m$, $\|\varphi_j\|_{\infty,A}^2 \leq 2^m \mathbf{a}^{-1} \|N_r\|_{\infty,[0,r]}^2 \leq D_m \mathbf{a}^{-1}$, which is $(\mathcal{M}_{4,2})$ (ii) with $c_2 = \mathbf{a}^{-1}$. Let us finally prove $(\mathcal{M}_5)$. Let $f \in \mathbb{L}^2(A)$. We write $f_m = \sum_{j\in\mathbb{J}_m} \alpha_j \varphi_j$ with $\vec{\alpha} = (\alpha_j)_{j\in\mathbb{J}_m} \in \mathbb{R}^{D_m}$. Since $f - f_m$ is orthogonal to S_m,

$$\sum_{j\in\mathbb{J}_m} \langle f, \varphi_j\rangle^2 = \sum_{j\in\mathbb{J}_m} \langle f_m, \varphi_j\rangle^2 = \sum_{j\in\mathbb{J}_m} \left(\sum_{k\in\mathbb{J}_m} \alpha_k \langle \varphi_k, \varphi_j\rangle \right)^2 = \|{}^t\vec{\alpha} T_m\|_{\ell^2}^2,$$

where T_m the positive symmetric matrix with size D_m and coefficients $[T_m]_{j,k} = \langle \varphi_j, \varphi_k\rangle$. Since T_m is a symmetric matrix, we have

$$\sum_{j\in\mathbb{J}_m} \langle f, \varphi_j\rangle^2 \leq \rho^2(T_m)\|\vec{\alpha}\|_{\ell^2}^2 \leq \rho^2(T_m)\mathfrak{d}_1^{-1}\|f_m\|^2,$$

thanks to $(\mathcal{M}_2)$. Now, T_m is also a Toeplitz matrix, which satisfies $[T_m]_{j,k} = 0$ if $|j-k| \geq r$, thanks to $(\mathcal{M}_{4,2})$. Applying the same result of Gray (2006) as above, we thus have

$$\rho(T_m) \leq \sum_{k\in\mathbb{Z}} [T_m]_{0,k} = \sum_{|k|\leq r} |\langle \varphi_0, \varphi_k\rangle| \leq \sum_{|k|\leq r} \|\varphi_0\|\|\varphi_k\| = (2r-1)\|N_r\|_{2,\mathbb{R}}^2.$$

Property $(\mathcal{M}_5)$ follows, with $c_3 = (2r-1)^2\|N_r\|_{2,\mathbb{R}}^4/\mathfrak{d}_1$.

Properties $(\mathcal{M}_l)$, $l = 2,3$, permit to establish the following lemma that will be used several times in the proofs.

Lemma 6.2. *Assume (H_2), and Properties $(\mathcal{M}_l)$, $l = 1,\ldots,4$. Then, for any $m \in \mathcal{M}_n$,*

$$\sup_{t\in B_m} \|t\|_{\infty,A}^2 \leq \Phi_0^2 D_m/(\mathfrak{d}_1 \bar{F}_0). \tag{15}$$

Moreover, if $\nu : S_m \to \mathbb{R}$ is a linear map, then

$$\sup_{t\in B_m} \nu^2(t) \quad \leq \quad \sum_{j\in\mathbb{J}_n} \nu^2(\varphi_j)/(\mathfrak{d}_1 \bar{F}_0). \tag{16}$$

Proof of Lemma 6.2. The key of the proof is to start from (14) and to remark that it leads to

$$\begin{aligned} B_m &\subset \left\{ t \in S_m,\ t = \sum_{j\in\mathbb{J}_m} \alpha_j\varphi_j,\ (\alpha_j)_j \in \mathbb{R}^{D_m},\ \left\| \sum_{j\in\mathbb{J}_m} \alpha_j\varphi_j \right\|^2 \leq \bar{F}_0^{-1} \right\}, \\ &\subset \left\{ t \in S_m,\ t = \sum_{j\in\mathbb{J}_m} \alpha_j\varphi_j,\ (\alpha_j)_j \in \mathbb{R}^{D_m},\ \sum_{j\in\mathbb{J}_m} \alpha_j^2 \leq (\mathfrak{d}_1\bar{F}_0)^{-1} \right\}, \end{aligned}$$

with Assumption $(\mathcal{M}_2)$. Thus, if $t \in B_m$, $t = \sum_{j\in\mathbb{J}_m} \alpha_j\varphi_j$, with $\vec{\alpha} = (\alpha_j)_{j\in\mathbb{J}_m} \in \mathbb{R}^{D_m}$, we have

$$\sup_{t\in B_m} \|t\|_{\infty,A}^2 \leq \sup_{\substack{\vec{\alpha}=(\alpha_j)_{j\in\mathbb{J}_m}\in\mathbb{R}^{D_m} \\ {}^t\vec{\alpha}\vec{\alpha}\leq(\mathfrak{d}_1\bar{F}_0)^{-1}}} \sup_{x\in A} \left(\sum_{j\in\mathbb{J}_m} \alpha_j\varphi_j(x) \right)^2.$$

Then, by the Cauchy-Schwarz Inequality and Assumption $(\mathcal{M}_3)$,

$$\sup_{t\in B_m} \|t\|_{\infty,A}^2 \leq (\mathfrak{d}_1\bar{F}_0)^{-1} \sup_{x\in A} \sum_{j\in\mathbb{J}_m} \varphi_j(x)^2 \leq (\mathfrak{d}_1\bar{F}_0)^{-1}\Phi_0^2 D_m.$$

The reasoning is the same to obtain (16).

About the statistical framework. We establish explicit expression for the distribution of the variable $Y = XU$ in the multiplicative censoring setting we consider. First recall that we have the following formula for the density of Y

$$f_Y(y) = \int_y^{+\infty} \frac{f_X(x)}{x} dx,\ y > 0. \tag{17}$$

Therefore we get (4). On the other hand,

$$\begin{aligned} \bar{F}_Y(y) &:= \mathbb{P}(Y_1 \geq y) = \int_y^{+\infty} f_Y(z)dz = \int_y^{+\infty}\int_z^{+\infty} \frac{f_X(x)}{x} dxdz \\ &= \int \left(\int_y^x dz \right) \frac{f_X(x)}{x} \mathbb{1}_{y\leq x} dx = \int_y^{+\infty} (x-y)\frac{f_X(x)}{x} dx \\ &= \int_y^{+\infty} f_X(x)dx - y\int_y^{+\infty} \frac{f_X(x)}{x} dx \\ &= \bar{F}_X(y) - yf_Y(y). \end{aligned}$$

So, (5) follows. □

About the estimator $\widehat{\bar{F}}_0$. The following properties of $\widehat{\bar{F}}_0$ defined by (9) are simple but crucial for the proof of our main results.

Lemma 6.3. *We have* $\widehat{\bar{F}}_0 \geq 1/\sqrt{n}$ *and, for all* $\rho_0 \in (0,1)$,

$$\mathbb{P}(|\widehat{\bar{F}}_0 - \bar{F}_0| \geq \rho_0\bar{F}_0) \leq c/n^4.$$

Any other estimator of $\bar{F}_0$ which satisfies such properties could be used in the paper. Also, $1/\sqrt{n}$ in (9) could be replaced by a quantity $\mu_n > 0$ such that $\lim_{n\to\infty} \mu_n = 0$ and $\mu_n \geq n^{-\alpha}$, for a real number $\alpha > 0$.

Proof of Lemma 6.3. The first remark is trivial with definition (9). We then compute

$$\begin{aligned}\mathbb{P}(|\widehat{\bar{F}}_0 - \bar{F}_0| \geq \rho_0 \bar{F}_0) &\leq \mathbb{P}\left(\left|n^{-1}\sum_{i=1}^n \mathbf{1}_{\{Y_i \geq \mathbf{a}\}} - \mathbb{P}(Y_i \geq \mathbf{a})\right| > \rho_0 \bar{F}_0 - \frac{1}{\sqrt{n}}\right),\\ &\leq \mathbb{P}\left(\left|n^{-1}\sum_{i=1}^n \mathbf{1}_{\{Y_i \geq \mathbf{a}\}} - \mathbb{P}(Y_i \geq \mathbf{a})\right| > \rho_0 \bar{F}_0/2\right),\end{aligned}$$

as soon as $n^{-1/2} \leq \rho_0 \bar{F}_0/2$. Then, from the Hoeffding Inequality (Hoeffding, 1963), there exists $\widetilde{c} > 0$ such that, for all $n \geq 4/(\rho_0^2 \bar{F}_0)$,

$$\mathbb{P}(|\widehat{\bar{F}_0} - \bar{F}_0| \geq \rho_0 \bar{F}_0) \leq \widetilde{c}/n^4,$$

which proves Lemma 6.3 with $c = \max\left\{\widetilde{c}, \left(\frac{4}{\rho_0^2 \bar{F}_0}\right)^4\right\}$.

6.2. **Proof of Lemma 2.1.** First note that from (5),

$$\mathbb{E}\left[\|t\|_n^2\right] = \int_A (t^2(x)\bar{F}_Y(x) + xt^2(x)f_Y(x))dx = \int_A t^2(x)\bar{F}_X(x)dx = \|t\|_{\bar{F}_X}^2, \tag{18}$$

and also that, with (4),

$$\begin{aligned}\mathbb{E}\left[(t(Y_1) + Y_1 t'(Y_1))\mathbf{1}_{Y_1 \in A}\right] &= \int_A (t(x) + xt'(x))f_Y(x)dx = \int_A (xt(x))' f_Y(x)dx\\ &= [xt(x)f_Y(x)]_0^{\mathbf{a}} - \int_0^{\mathbf{a}} xt(x)f_Y'(x)dx = \mathbf{a}t(\mathbf{a})f_Y(\mathbf{a}) + \int_A t(x)f_X(x)dx,\end{aligned}$$

since $\lim_{x\to 0} xt(x)f_Y(x) = 0$. Indeed, Formula (5) implies that $\lim_{x\to 0}(xf_Y(x)) = \bar{F}_X(0) - \bar{F}_Y(0) = 0$. Using the definition of h, we also have

$$\int_A t(x)f_X(x)dx = \int_A t(x)h(x)\bar{F}_X(x)dx = \langle h, t\rangle_{\bar{F}_X}.$$

This leads to

$$\mathbb{E}\left[\tilde{\nu}_n(t)\right] = \langle h, t\rangle_{\bar{F}_X} + \mathbf{a}t(\mathbf{a})\left(f_Y(\mathbf{a}) - \mathbb{E}\left[\widehat{f}_Y(\mathbf{a})\right]\right). \tag{19}$$

The result follows by gathering (18) and (19).

6.3. **Proof of Proposition 3.1.** Let us define, for $0 < \rho_0, \rho_1 < 1$ such that $\rho_1 - 2\rho_0 + \rho_0\rho_1$, the following sets,

$$\Omega_{\rho_0} := \left\{\left|\frac{\|t\|_n^2}{\|t\|_{\bar{F}_X}^2} - 1\right| \leq \rho_0, \text{ for all } t \in \mathcal{S}_n\right\}$$

and

$$\Omega_{\rho_0,\rho_1} := \left\{(1-\rho_0)\bar{F}_0 \leq \widehat{\bar{F}}_0 \leq \frac{1-\rho_0}{1-\rho_1}\bar{F}_0\right\},$$

and decompose the risk as follows

$$\mathbb{E}\left[\left\|\widehat{h}_m - h\right\|_{\bar{F}_X}^2\right] = \mathbb{E}\left[\left\|\widehat{h}_m - h\right\|_{\bar{F}_X}^2 \mathbf{1}_{\Omega_{\rho_0}\cap\Omega_{\rho_0,\rho_1}}\right] + \mathbb{E}\left[\left\|\widehat{h}_m - h\right\|_{\bar{F}_X}^2 \mathbf{1}_{(\Omega_{\rho_0}\cap\Omega_{\rho_0,\rho_1})^c}\right]. \tag{20}$$

We now compute upper-bounds for each of the two terms of the right-hand-side of (20).

Upper-bound for $\mathbb{E}[\|\widehat{h}_m - h\|^2 \mathbf{1}_{\Omega_{\rho_0}\cap\Omega_{\rho_0,\rho_1}}]$. First we have

$$\Omega_{\rho_0} \subset \bigcap_{m=1}^{N_n} \Delta_{\rho_0}^m, \tag{21}$$

where

$$\Delta_{\rho_0}^m = \left\{\min \mathrm{Sp}(\widehat{\Psi}_m) \geq (1-\rho_0)\mathfrak{d}_1 \bar{F}_0\right\}. \tag{22}$$

Indeed,

$$\begin{aligned}
\Omega_{\rho_0} &\subset \left\{\forall t \in \mathcal{S}_n,\ \|t\|_n^2 \geq (1-\rho_0)\|t\|_{\bar{F}_X}^2\right\} \subset \bigcap_{m=1}^{N_n} \left\{\forall t \in S_m,\ \|t\|_n^2 \geq (1-\rho_0)\bar{F}_0\|t\|^2\right\} \\
&= \bigcap_{m=1}^{N_n} \left\{\forall \vec{\alpha} \in \mathbb{R}^{D_m},\ {}^t\vec{\alpha}\widehat{\Psi}_m\vec{\alpha} \geq (1-\rho_0)\bar{F}_0 \left\|\sum_{j\in\mathbb{J}_m} \alpha_j\varphi_j\right\|^2\right\}, \\
&\subset \bigcap_{m=1}^{N_n} \left\{\forall \vec{\alpha} \in \mathbb{R}^{D_m},\ {}^t\vec{\alpha}\widehat{\Psi}_m\vec{\alpha} \geq (1-\rho_0)\bar{F}_0\mathfrak{d}_1 \sum_{j\in\mathbb{J}_m} \alpha_j^2\right\},
\end{aligned}$$

by using (8) and $(\mathcal{M}_2)$. This proves Inclusion (21), which permits to deduce

$$\Omega_{\rho_0} \cap \Omega_{\rho_0,\rho_1} \subset \left(\bigcap_{m=1}^{N_n} \Delta_{\rho_0}^m\right) \cap \Omega_{\rho_0,\rho_1} \subset \left(\bigcap_{m=1}^{N_n} \widehat{\Delta}_{\rho_1}^m\right) \cap \Omega_{\rho_0,\rho_1},$$

since $(1-\rho_1)\widehat{\bar{F}}_0 \leq (1-\rho_0)\bar{F}_0$ on Ω_{ρ_0,ρ_1}.
Therefore, on the set $\Omega_{\rho_0} \cap \Omega_{\rho_0,\rho_1}$, for all $m = 1, ..., N_n$, $\widehat{h}_m$ is the minimiser of the contrast γ_n on S_m. In particular, we have $\gamma_n(\widehat{h}_m) \leq \gamma_n(h_m)$, which means

$$\left\|\widehat{h}_m\right\|_n^2 - \|h_m\|_n^2 \leq 2\widetilde{\nu}_n(\widehat{h}_m - h_m).$$

on the set $\Omega_{\rho_0} \cap \Omega_{\rho_0,\rho_1}$ and thus

$$\left\|\widehat{h}_m - h_m\right\|_n^2 \leq 2\|h_m\|_n^2 + 2\widetilde{\nu}_n(\widehat{h}_m - h_m) - 2\langle \widehat{h}_m, h_m\rangle_n,$$

since $\|\widehat{h}_m\|_n^2 = \|\widehat{h}_m - h_m\|_n^2 - \|h_m\|_n^2 + 2\langle \widehat{h}_m, h_m\rangle_n$. The key of the proof is the control of the deviations of the supremum of a centred empirical process over a unit ball. Since the process $\tilde{\nu}_n$ is not centred, we set

$$\nu_n(t) := \widetilde{\nu}_n(t) - \langle t, h\rangle_n - \mathbf{a}t(\mathbf{a})\left(f_Y(\mathbf{a}) - \widehat{f}_Y(\mathbf{a})\right), \tag{23}$$

with $\langle t, h\rangle_n$ defined by (13). The process ν_n is centered, thanks to (19) and the fact that $\mathbb{E}[\langle t, h\rangle_n] = \langle t, h\rangle_{\bar{F}_X}$. Moreover, $\|h_m\|_n^2 - \langle \widehat{h}_m, h_m\rangle_n = \langle h_m - \widehat{h}_m, h_m - h\rangle_n + \langle h_m - \widehat{h}_m, h\rangle_n$. This leads, on $\Omega_{\rho_0} \cap \Omega_{\rho_0,\rho_1}$, to

$$\left\|\widehat{h}_m - h_m\right\|_n^2 \leq 2\langle h_m - \widehat{h}_m, h_m - h\rangle_n + 2\nu_n\left(\widehat{h}_m - h_m\right) + 2\mathbf{a}\left(\widehat{h}_m - h_m\right)(\mathbf{a})\left(f_Y(\mathbf{a}) - \widehat{f}_Y(\mathbf{a})\right).$$

Then, thanks to the Cauchy-Schwarz inequality, and the fact that, for all $x, y \in \mathbb{R}$ and $\theta > 0$, $2xy \leq \theta x^2 + \theta^{-1}y^2$,

$$2\langle h_m - \widehat{h}_m, h_m - h\rangle_n \leq \theta\left\|\widehat{h}_m - h_m\right\|_n^2 + \theta^{-1}\|h_m - h\|_n^2.$$

Applying the last property twice, we obtain, for any $\delta, \eta > 0$, $0 < \theta < 1$,

$$(1-\theta)\left\|\widehat{h}_m - h_m\right\|_n^2 \mathbf{1}_{\Omega_{\rho_0}\cap\Omega_{\rho_0,\rho_1}} \leq \Bigg((\delta+\eta)\left\|\widehat{h}_m - h_m\right\|_{\bar{F}_X}^2 + \theta^{-1}\|h_m - h\|_n^2 + \delta^{-1}\sup_{t\in B_m}\nu_n^2(t) + \eta^{-1}\mathbf{a}^2\sup_{t\in B_m} t^2(\mathbf{a})\left(f_Y(\mathbf{a}) - \widehat{f}_Y(\mathbf{a})\right)^2\Bigg)\mathbf{1}_{\Omega_{\rho_0}\cap\Omega_{\rho_0,\rho_1}}.$$

where B_m is defined by (13). Since $\Omega_{\rho_0} \subset \Delta_{\rho_0}^m$, then, for any $t \in S_m$, $\|t\|_n^2 \geq (1-\rho_0)\|t\|_{\bar{F}_X}^2$. Taking also into account that $\|\widehat{h}_m - h\|_{\bar{F}_X}^2 \leq 2\|\widehat{h}_m - h_m\|_{\bar{F}_X}^2 + 2\|h_m - h\|_{\bar{F}_X}^2$, and that $\mathbb{E}\left[\|t\|_n^2\right] = \|t\|_{\bar{F}_X}^2$ $(t \in S_m)$, we have, as soon as $(1-\rho_0)(1-\theta) - \delta - \eta > 0$,

$$\mathbb{E}\left[\left\|\widehat{h}_m - h\right\|_{\bar{F}_X}^2 \mathbf{1}_{\Omega_{\rho_0}\cap\Omega_{\rho_0,\rho_1}}\right] \leq C_1\|h_m - h\|_{\bar{F}_X}^2 + C_2\mathbb{E}\left[\sup_{t\in B_m}\nu_n^2(t)\right] + C_3\mathbf{a}^2\sup_{t\in B_m} t^2(\mathbf{a})\mathbb{E}\left[\left(f_Y(\mathbf{a}) - \widehat{f}_Y(\mathbf{a})\right)^2\right], \tag{24}$$

where

$$C_1 = 2\left(\frac{\theta^{-1}}{(1-\rho_0)(1-\theta)-\delta-\eta} + 1\right), \quad C_2 = \frac{2\delta^{-1}}{(1-\rho_0)(1-\theta)-\delta-\eta},$$
$$\text{and } C_3 = \frac{2\eta^{-1}}{(1-\rho_0)(1-\theta)-\delta-\eta}. \tag{25}$$

An upper bound for the term with constant C_3 is given by (15) of Lemma 6.2: $\sup_{t\in B_m} t^2(\mathbf{a}) \leq \Phi_0^2/(\mathfrak{d}_1\bar{F}_0)D_m$. To bound the empirical process, we establish the following lemma below (see Section 6.6.1).

Lemma 6.4. *Assume that (H_1) and (H_2) are satisfied. Then,*

$$\mathbb{E}\left[\sup_{t\in B_m}\nu_n^2(t)\right] \leq 4\left(\frac{(1+\mathbf{a}\|h\|^2)\Phi_0^2}{\mathfrak{d}_1\bar{F}_0}\frac{D_m}{n} + \frac{\mathbf{a}^2\Phi_1^2}{2\mathfrak{d}_1\bar{F}_0}\frac{D_m^3}{n} + \frac{c_3\|h\|^2}{\mathfrak{d}_1\bar{F}_0}\frac{1}{n}\right).$$

By gathering this and the result (15) of Lemma 6.2 in (24), we obtain

$$\mathbb{E}\left[\left\|\widehat{h}_m - h\right\|_{\bar{F}_X}^2 \mathbf{1}_{\Omega_{\rho_0}\cap\Omega_{\rho_0,\rho_1}}\right] \leq \tilde{C}_1(\|h_m - h\|^2 + V(m)) + \tilde{C}_2\frac{\mathbf{a}^2\Phi_0^2}{\mathfrak{d}_1\bar{F}_0}D_m\mathbb{E}\left[\left(\widehat{f}_Y(\mathbf{a}) - f_Y(\mathbf{a})\right)^2\right] + \frac{\tilde{C}_3}{n} \tag{26}$$

with $\tilde{C}_1$, $\tilde{C}_2$ and $\tilde{C}_3$ some constants which do not depend on m and n, and $V(m)$ defined as in the statement of Proposition 3.1.

Upper-bound for $\mathbb{E}[\|\widehat{h}_m - h\|^2\mathbf{1}_{(\Omega_{\rho_0}\cap\Omega_{\rho_0,\rho_1})^c}]$. First,

$$\left\|\widehat{h}_m - h\right\|_{\bar{F}_X}^2 \leq 2\left\|\widehat{h}_m\right\|_{\bar{F}_X}^2 + 2\|h\|_{\bar{F}_X}^2 \leq 2\left\|\widehat{h}_m\right\|^2 + 2\|h\|_{\bar{F}_X}^2. \tag{27}$$

Now, by definition of $\widehat{h}_m$, and thanks to $(\mathcal{M}_2)$,

$$\begin{aligned}
\left\|\widehat{h}_m\right\| &= \left\|\sum_{j\in\mathbb{J}_m}\widehat{\alpha}_j\varphi_j\right\|\mathbf{1}_{\widehat{\Delta}^m_{\rho_1}} \leq \sqrt{\mathfrak{d}_2}\left(\sum_{j\in\mathbb{J}_m}\widehat{\alpha}_j^2\right)^{1/2}\mathbf{1}_{\widehat{\Delta}^m_{\rho_1}} = \sqrt{\mathfrak{d}_2}\|\widehat{\vec{\alpha}}\|_{\ell^2}\mathbf{1}_{\widehat{\Delta}^m_{\rho_1}},\\
&= \sqrt{\mathfrak{d}_2}\|\widehat{\Psi}_m^{-1}\widehat{\vec{b}}\|_{\ell^2}\mathbf{1}_{\widehat{\Delta}^m_{\rho_1}}\\
&\leq \sqrt{\mathfrak{d}_2}\rho(\widehat{\Psi}_m^{-1})\|\widehat{\vec{b}}\|_{\ell^2}\mathbf{1}_{\widehat{\Delta}^m_{\rho_1}} \leq \sqrt{\mathfrak{d}_2}\max\{\mathrm{Sp}(\widehat{\Psi}_m^{-1})\}\|\widehat{\vec{b}}\|_{\ell^2}\mathbf{1}_{\widehat{\Delta}^m_{\rho_1}}\\
&\leq \sqrt{\mathfrak{d}_2}\left(\min\{\mathrm{Sp}(\widehat{\Psi}_m)\}\right)^{-1}\|\widehat{\vec{b}}\|_{\ell^2}\mathbf{1}_{\widehat{\Delta}^m_{\rho_1}} \leq \sqrt{\mathfrak{d}_2}\frac{1}{(1-\rho_1)\mathfrak{d}_1\widehat{\bar{F}}_0}\|\widehat{\vec{b}}\|_{\ell^2},\\
&\leq \sqrt{\mathfrak{d}_2 n}\frac{1}{(1-\rho_1)\mathfrak{d}_1}\|\widehat{\vec{b}}\|_{\ell^2}, \qquad (28)
\end{aligned}$$

where we recall that $\|\cdot\|_{\ell^2}$ is the Euclidean norm on $\mathbb{R}^{D_m}$ and $\rho(\cdot)$ the spectral radius of a matrix.

Hence, as $\widehat{\vec{b}} = {}^t\left(\frac{1}{n}\sum_{i=1}^n(\varphi_j(Y_i)+Y_i\varphi_j'(Y_i)) - \widehat{f}_Y(\mathbf{a})\varphi_j(\mathbf{a})\right)_{j\in\mathbb{J}_m}$, we have

$$\|\widehat{\vec{b}}\|_{\ell^2} \leq (3T_1+3T_2+3T_3)^{1/2}$$

where

$$T_1 = \sum_{j\in\mathbb{J}_m}\left(\frac{1}{n}\sum_{i=1}^n\varphi_j(Y_i)\right)^2,\ T_2 = \sum_{j\in\mathbb{J}_m}\left(\frac{1}{n}\sum_{i=1}^n Y_i\varphi_j'(Y_i)\right)^2,\ \text{and } T_3 = \widehat{f}_Y^2(\mathbf{a})\sum_{j\in\mathbb{J}_m}(\varphi_j(\mathbf{a}))^2.$$

We first bound T_1, using the Cauchy-Schwarz Inequality and the first part of $(\mathcal{M}_3)$,

$$T_1 \leq \frac{1}{n}\sum_{j\in\mathbb{J}_m}\sum_{i=1}^n\varphi_j^2(Y_i) \leq \Phi_0^2 D_m.$$

Similarly with the second part of $(\mathcal{M}_3)$

$$T_2 \leq \frac{1}{n}\sum_{j\in\mathbb{J}_m}\sum_{i=1}^n Y_i^2(\varphi_j')^2(Y_i) \leq \mathbf{a}^2\Phi_1^2 D_m^3.$$

Moreover, thanks to the first part of $(\mathcal{M}_3)$ and the definition of $\widehat{f}_Y$, $\widehat{\mathfrak{w}} > 1/\sqrt{n}$,

$$T_3 \leq \widehat{f}_Y^2(\mathbf{a})\Phi_0^2 D_m \leq n\|K\|^2_{\infty,\mathbb{R}}\Phi_0^2 D_m$$

This implies that

$$\|\widehat{\vec{b}}\|_{\ell^2} \leq \sqrt{3}\left((1+n\|K\|^2_{\infty,\mathbb{R}})\Phi_0^2 D_m + \mathbf{a}^2\Phi_1^2 D_m^3\right)^{1/2}.$$

and, gathering equations (27) and (28),

$$\begin{aligned}
\mathbb{E}\left[\|\widehat{h}_m - h\|^2_{\bar{F}_X}\mathbf{1}_{\Omega^c_{\rho_0}}\right] &\leq \left(\frac{6\mathfrak{d}_2 n}{(1-\rho_1)^2\mathfrak{d}_1^2}\left((1+n\|K\|^2_{\infty,\mathbb{R}})\Phi_0^2 D_m + \mathbf{a}^2\Phi_1^2 D_m^3\right) + 2\|h\|^2_{\bar{F}_X}\right)\\
&\quad\times\mathbb{P}\left((\Omega_{\rho_0}\cap\Omega_{\rho_0,\rho_1})^c\right). \qquad (29)
\end{aligned}$$

It remains to bound the probability of the event $(\Omega_{\rho_0}\cap\Omega_{\rho_0,\rho_1})^c$. Notice first that

$$\mathbb{P}\left((\Omega_{\rho_0}\cap\Omega_{\rho_0,\rho_1})^c\right) \leq \mathbb{P}\left(\Omega^c_{\rho_0}\right) + \mathbb{P}\left(\Omega^c_{\rho_0,\rho_1}\right).$$

The aim of the following lemma (proved in Section 6.6.2) is to bound $\mathbb{P}(\Omega^c_{\rho_0})$.

Lemma 6.5. *Assume that (H_1) and (H_2) are satisfied. Then the following inequality holds*

$$\mathbb{P}\left(\Omega_{\rho_0}^c\right) \leq 4\exp\left(-c^*\frac{n}{D_{N_n}^2}\right) \leq 4\exp\left(-\frac{c^*}{K^2}\ln^3(n)\right), \tag{30}$$

with c^* a constant depending on $\mathbf{a}$, ρ_0, $\bar{F}_0$, $\mathbb{E}[Y^2]$, and on the constants involved in properties $(\mathcal{M}_l)$, $l = 2, 3, 4, 5$.

For $\mathbb{P}(\Omega_{\rho_0,\rho_1}^c)$, as soon as $\rho_1 - 2\rho_0 + \rho_0\rho_1 > 0$, we have

$$\begin{aligned}\Omega_{\rho_0,\rho_1}^c &= \left\{(1-\rho_0)\bar{F}_0 > \widehat{\bar{F}}_0\right\} \cup \left\{\widehat{\bar{F}}_0 > \frac{1-\rho_0}{1-\rho_1}\bar{F}_0\right\} \\ &\subset \left\{\bar{F}_0 - \widehat{\bar{F}}_0 > \rho_0\bar{F}_0\right\} \cup \left\{\widehat{\bar{F}}_0 - \bar{F}_0 > \rho_0\bar{F}_0\right\} \subset \left\{\left|\widehat{\bar{F}}_0 - \bar{F}_0\right| > \rho_0\bar{F}_0\right\}.\end{aligned}$$

Thus, thanks to Lemma 6.3,

$$\mathbb{P}(\Omega_{\rho_0,\rho_1}^c) \leq cn^{-4}. \tag{31}$$

Gathering this bound and Lemma 6.5 in (29) leads to

$$\mathbb{E}\left[\|\widehat{h}_m - h\|_{\bar{F}_X}^2 \mathbf{1}_{\Omega_{\rho_0}^c \cup \Omega_{\rho_0,\rho_1}^c}\right] \leq c\left(n\left((1+n\|K\|_{\infty,\mathbb{R}}^2\right)D_m + D_m^3\right)\left(1/n^4 + \exp\left(-\frac{c^*}{K^2}\ln^3(n)\right)\right).$$

where c is a constant. Thanks to the fact that $D_m \leq K\sqrt{n/(\ln(n)^3)}$, we obtain, for a constant $C > 0$,

$$\mathbb{E}\left[\left\|\widehat{h}_m - h\right\|_{\bar{F}_X}^2 \mathbf{1}_{(\Omega_{\rho_0}\cap\Omega_{\rho_0,\rho_1})^c}\right] \leq \frac{C}{n}, \tag{32}$$

The proof of Proposition 3.1 is completed by gathering (26) and (32), and keeping in mind that $\|.\|_{\bar{F}_X}$ and $\|.\|$ are equivalent (under (H_2)).

6.4. **Proof of Corollary 3.1.** If h belongs to the Hölder ball $\mathcal{H}^\beta([0, \mathbf{a}+\varepsilon], L)$, then h belongs also to the Besov ball $\mathcal{B}_{\infty,\infty}^\beta([0, \mathbf{a}+\varepsilon], L)$ (see *e.g.* DeVore and Lorentz 1993 p.54 for the definition of such spaces, and Barron et al. 1999 p.404, or DeVore and Lorentz 1993 (7.13) p.46). This permits to deduce that h also belongs to $\mathcal{B}_{2,\infty}^\beta([0, \mathbf{a}+\varepsilon], L')$ for another $L' > 0$ (Massart 2007, p.144). But, for such a ball, Lemma 12 of Barron et al. (1999) for [T] and DeVore and Lorentz 1993 for [BS], established that $\|h_m - h\| \leq D_m^{-\beta}$. Thus, Proposition 3.1 leads to

$$\mathbb{E}\left[\|\widehat{h}_m - h\|^2\right] \leq C_1\left(D_m^{-2\beta} + V(m)\right) + \tilde{C}_2 D_m \mathbb{E}\left[\left(\widehat{f}_Y(\mathbf{a}) - f_Y(\mathbf{a})\right)^2\right] + \frac{C_3}{n},$$

for three constants $C_1, \tilde{C}_2, C_3$. The computation of $m(\beta) \in \arg\min\{D_m^{-2\beta} + V(m),\ m \in \mathcal{M}_n\}$ leads to $m(\beta) = Cn^{1/(2\beta+3)}$. Now, we deduce from the smoothness assumption on h, a smoothness property for f_Y (the proof is deferred to Section 6.7).

Proposition 6.1. *If $h \in \mathcal{H}^\beta([0, \mathbf{a}+\varepsilon], L)$ and (H_2) is verified, then there exist $L' > 0$ and $0 < \varepsilon' < \min(\mathbf{a}, \varepsilon)$ such that $f_Y \in \mathcal{H}^{\beta+1}([\mathbf{a}-\varepsilon', \mathbf{a}+\varepsilon'], L')$.*

Since $\widehat{f}_Y$ is a kernel estimate with well-chosen bandwidth, Rebelles (2015) derives that

$$\mathbb{E}\left[\left(\widehat{f}_Y(\mathbf{a}) - f_Y(\mathbf{a})\right)^2\right] \leq C\left(\frac{n}{\ln(n)}\right)^{-\frac{2(\beta+1)}{2(\beta+1)+1}},$$

for a constant C. Thus,

$$\mathbb{E}\left[\|\widehat{h}_{m(\beta)} - h\|^2\right] \leq C_1 n^{-\frac{2\beta}{2\beta+3}} + \tilde{C}_2 D_{m(\beta)} C\left(\frac{n}{\ln(n)}\right)^{-\frac{2\beta+2}{2\beta+3}} + \frac{C_3}{n},$$

and Corollary 3.1 follows from the order of magnitude of $D_{m(\beta)}$.

6.5. **Proof of Theorem 4.1.** As previously, we decompose the risk

$$\mathbb{E}\left[\|\widehat{h}_{\widehat{m}}-h\|^2\right]=\mathbb{E}\left[\|\widehat{h}_{\widehat{m}}-h\|^2\mathbf{1}_{\Omega_{\rho_0}\cap\Omega_{\rho_0,\rho_1}}\right]+\mathbb{E}\left[\|\widehat{h}_{\widehat{m}}-h\|^2\mathbf{1}_{(\Omega_{\rho_0}\cap\Omega_{\rho_0,\rho_1})^c}\right].$$

We first focus on the first term and we get that, on $\Omega_{\rho_0}\cap\Omega_{\rho_0,\rho_1}$, for all $m=1,...,N_n$, $\widehat{h}_m$ do minimise the contrast γ_n on S_m. Therefore, by definition of $\widehat{m}$, we have, on $\Omega_{\rho_0}\cap\Omega_{\rho_0,\rho_1}$, for all $m=1,...,N_n$,

$$\gamma_n(\widehat{h}_{\widehat{m}})+\widehat{\text{pen}}(\widehat{m})\leq\gamma_n(\widehat{h}_m)+\widehat{\text{pen}}(m)\leq\gamma_n(h_m)+\widehat{\text{pen}}(m), \tag{33}$$

where we recall that h_m is the orthogonal projection of h on S_m. Now, with the same algebra as in the proof of Proposition 3.1, we get that, for $(1-\rho_0)(1-\theta)-\delta-\eta>0$,

$$\begin{aligned}\mathbb{E}\left[\left\|\widehat{h}_{\widehat{m}}-h\right\|^2_{\bar{F}_X}\mathbf{1}_{\Omega_{\rho_0}\cap\Omega_{\rho_0,\rho_1}}\right] &\leq C_1\|h_m-h\|^2_{\bar{F}_X}+C_4\text{pen}(m)\mathbf{1}_{\Omega_{\rho_0}\cap\Omega_{\rho_0,\rho_1}}\\ &\quad+C_2\mathbb{E}\left[\sup_{t\in B_{m\vee\widehat{m}}}\nu_n^2(t)-p(\widehat{m},m)\right]\\ &\quad+C_3\mathbf{a}^2\mathbb{E}\left[\sup_{t\in B_{m\vee\widehat{m}}}t^2(\mathbf{a})\left(f_Y(\mathbf{a})-\widehat{f}_Y(\mathbf{a})\right)^2\right]\\ &\quad+(C_2p(\widehat{m},m)-\widehat{\text{pen}}(\widehat{m}))\mathbf{1}_{\Omega_{\rho_0}\cap\Omega_{\rho_0,\rho_1}},\end{aligned} \tag{34}$$

where $B_{m\vee\widehat{m}}=\{t\in S_{m\vee\widehat{m}},\|t\|_{\bar{F}_X}\leq 1\}$ (see (13)), C_1,C_2,C_3 are given by (25), C_4 is a constant which does not depend on n,m and $p(m',m)$ is defined in the following Lemma.

Lemma 6.6. *Assume that (H_1) and (H_2) hold. There exists $\bar{\kappa}_0$ such that for $\kappa\geq\bar{\kappa}_0$*

$$\sum_{m'=1}^{N_n}\mathbb{E}\left[\left(\sup_{t\in B_{m\vee m'}}\nu_n^2(t)-p(m,m')\right)_+\right]\leq\frac{C}{n},$$

where

$$p(m,m')=\bar{\kappa}_0(\mathfrak{d}_1\bar{F}_0)^{-1}\left(\Phi_0^2(1+\mathbf{a}\|h\|^2)\frac{D_{m\vee m'}}{n}+\mathbf{a}^2\Phi_1^2\frac{D^3_{m\vee m'}}{n}\right),$$

and $C>0$ is a a constant.

We have $\|h\|^2\leq D^2_{m\vee\widehat{m}}$ for $n\geq n_0(h)$, since $D_{m\vee\widehat{m}}\geq\ln(n)$. By definition of Ω_{ρ_0,ρ_1}, this leads to

$$\begin{aligned}C_2p(m,\widehat{m})\mathbf{1}_{\Omega_{\rho_0,\rho_1}} &\leq C_2\bar{\kappa}_0\frac{1-\rho_0}{1-\rho_1}(\widehat{\bar{F}}_0\mathfrak{d}_1)^{-1}\left(\Phi_0^2\frac{D_{m\vee\widehat{m}}}{n}+(\mathbf{a}\Phi_0^2+\mathbf{a}^2\Phi_1^2)\frac{D^3_{m\vee\widehat{m}}}{n}\right)\mathbf{1}_{\Omega_{\rho_0,\rho_1}}\\ &\leq(\widehat{\text{pen}}(m)+\widehat{\text{pen}}(\widehat{m}))\,\mathbf{1}_{\Omega_{\rho_0,\rho_1}}\end{aligned}$$

as soon as $\kappa\geq C_2\bar{\kappa}_0\frac{1-\rho_0}{1-\rho_1}$, since $D_{m\vee\widehat{m}}\leq D_m+D_{\widehat{m}}$. This and Inequality (34) imply that

$$\begin{aligned}\mathbb{E}\left[\|\widehat{h}_{\widehat{m}}-h\|^2_{\bar{F}_X}\mathbf{1}_{\Omega_{\rho_0}\cap\Omega_{\rho_0,\rho_1}}\right] &\leq C_1\|h_m-h\|^2_{\bar{F}_X}+C_4\text{pen}(m)+C_2\mathbb{E}\left[\left(\sup_{t\in B_{m\vee\widehat{m}}}\nu_n^2(t)-p(m,\widehat{m})\right)_+\right]\\ &\quad+\mathbb{E}\left[\widehat{\text{pen}}(m)\mathbf{1}_{\Omega_{\rho_0,\rho_1}}\right]+C_4\text{pen}(m).\end{aligned} \tag{35}$$

Moreover,

$$\begin{aligned}\widehat{\mathrm{pen}}(m)\mathbf{1}_{\Omega_{\rho_0,\rho_1}} &\leq \kappa(\mathfrak{d}_1\widehat{\bar{F}}_0)^{-1}\left(\Phi_0^2\frac{D_m}{n}+(\mathbf{a}\Phi_0^2+\mathbf{a}^2\Phi_1^2)\frac{D_m^3}{n}\right)\mathbf{1}_{\Omega_{\rho_0,\rho_1}}\\ &\leq \frac{\kappa(\mathfrak{d}_1\bar{F}_0)^{-1}}{1-\rho_0}\left(\Phi_0^2\frac{D_m}{n}+(\mathbf{a}\Phi_0^2+\mathbf{a}^2\Phi_1^2)\frac{D_m^3}{n}\right)\\ &\leq \frac{1}{1-\rho_0}\mathrm{pen}(m).\end{aligned}$$

Hence, applying Lemma 6.6, we obtain

$$\begin{aligned}\mathbb{E}\left[\|\widehat{h}_{\widehat{m}}-h\|^2_{\bar{F}_X}\mathbf{1}_{\Omega_{\rho_0}\cap\Omega_{\rho_1}}\right] &\leq \left\{C_1\|h_m-h\|^2_{\bar{F}_X}+C_4'\mathrm{pen}(m)\right\}\\ &\quad +C_3\mathbf{a}^2\mathbb{E}\left[\sup_{t\in B_{m\vee\widehat{m}}}t^2(\mathbf{a})\left(f_Y(\mathbf{a})-\widehat{f}_Y(\mathbf{a})\right)^2\right]+\frac{C_5}{n}.\end{aligned}$$

Moreover, on the complementary event, we follow the same line as in the proof of Proposition 3.1, as all the bounds remain true for D_m replaced by N_n, the maximal dimension, and we get

$$\mathbb{E}\left[\|\widehat{h}_{\widehat{m}}-h\|^2_{\bar{F}_X}\mathbf{1}_{(\Omega_{\rho_0}\cap\Omega_{\rho_1})^c}\right]\leq\frac{c}{n},$$

for some constant c. Gathering the last two bounds gives an upper bound on $\mathbb{E}[\|\widehat{h}_{\widehat{m}}-h\|^2_{\bar{F}_X}]$. Recalling that the norm $\|.\|_{\bar{F}_X}$ is equivalent to $\|.\|$ (see Section 6.1), Theorem 4.1 is proved.

6.6. **Proof of technical lemmas.**

6.6.1. *Proof of Lemma 6.4.* The process ν_n, defined by $\nu_n(t)=\widetilde{\nu}_n(t)-\langle t,h\rangle_n-\mathbf{a}t(\mathbf{a})(f_Y(\mathbf{a})-\mathbb{E}[\widehat{f}_Y(\mathbf{a})])$, $t\in\mathbb{L}^2(A)$ is centred. Thus, we split it into 4 centred processes as follows: $\nu_n=\sum_{l=1}^3\nu_n^{(l)}$, with, for $t\in\mathbb{L}^2(A)$,

$$\begin{aligned}\nu_n^{(1)}(t) &:= \frac{1}{n}\sum_{i=1}^n t(Y_i)\mathbf{1}_{\{Y_i\in A\}}-\mathbb{E}[t(Y_i)\mathbf{1}_{\{Y_i\in A\}}],\\ \nu_n^{(2)}(t) &:= \frac{1}{n}\sum_{i=1}^n Y_it'(Y_i)\mathbf{1}_{\{Y_i\in A\}}-\mathbb{E}[Y_it'(Y_i)\mathbf{1}_{\{Y_i\in A\}}],\\ \nu_n^{(3)}(t) &:= -\langle t,h\rangle_n+\mathbb{E}[\langle t,h\rangle_n].\end{aligned}$$

The process $\nu_n^{(3)}$ is also written $\nu_n^{(3)}=\nu_n^{(3,1)}+\nu_n^{(3,2)}$, with

$$\begin{aligned}\nu_n^{(3,1)}(t) &:= \frac{1}{n}\sum_{i=1}^n\int_A t(x)h(x)\mathbf{1}_{\{Y_i\geq x\}}dx-\mathbb{E}\left[\int_A t(x)h(x)\mathbf{1}_{\{Y_i\geq x\}}dx\right],\\ \nu_n^{(3,2)}(t) &:= \frac{1}{n}\sum_{i=1}^n Y_it(Y_i)h(Y_i)\mathbf{1}_{\{Y_i\in A\}}-\mathbb{E}[Y_it(Y_i)h(Y_i)\mathbf{1}_{\{Y_i\in A\}}],\end{aligned}$$

Therefore, it remains to bound the four terms of the right-hand-side of the following inequality:

$$\mathbb{E}\left[\sup_{t\in B_m}\nu_n^2(t)\right]\leq 4\sum_{l\in\{1,2,(3,1),(3,2)\}}\mathbb{E}\left[\sup_{t\in B_m}\left(\nu_n^{(l)}(t)\right)^2\right].$$

For $l = 1, 2, (3,1), (3,2)$, we first apply Property (16) of Lemma 6.2:

$$\sup_{t\in B_m}\left(\nu_n^{(l)}(t)\right)^2 \le \frac{1}{\mathfrak{d}_1\bar{F}_0}\sum_{j\in\mathbb{J}_m}\left(\nu_n^{(l)}(\varphi_j)\right)^2.$$

Let us then begin with $\nu_n^{(1)}$. For any $j \in \mathbb{J}_m$, $\mathbb{E}[(\nu_n^{(1)}(\varphi_j))^2] = \mathrm{Var}(\varphi_j(Y_1))/n \le \mathbb{E}[\varphi_j^2(Y)]/n \le \Phi_0^2 D_m/n$ thanks to $(\mathcal{M}_3)$. Thus,

$$\mathbb{E}\left[\sup_{t\in B_m}\left(\nu_n^{(1)}(t)\right)^2\right] \le \frac{\Phi_0^2}{\mathfrak{d}_1\bar{F}_0}\frac{D_m}{n} \tag{36}$$

Similarly, we have $\mathbb{E}[(\nu_n^{(2)}(\varphi_j))^2] \le n^{-1}\mathbb{E}[Y_1^2(\varphi_j'(Y_1))^2]$. Hence, by $(\mathcal{M}_3)$ $\mathbb{E}[\sup_{t\in B_m}(\nu_n^{(2)}(t))^2] \le (\Phi_1^2/(\mathfrak{d}_1\bar{F}_0))\mathbb{E}[Y_1^2\mathbb{1}_A(Y_1)]D_m^3/n$. Now, Equation (5) implies that, for all y, $yf_Y(y) \le 1$. Thus $\mathbb{E}[Y_1^2\mathbb{1}_A(Y_1)] = \int_0^{\mathbf{a}} y^2 f_Y(y)dy \le \int_0^{\mathbf{a}} ydy = \mathbf{a}^2/2$ and

$$\mathbb{E}\left[\sup_{t\in B_m}\left(\nu_n^{(2)}(t)\right)^2\right] \le \frac{\mathbf{a}^2\Phi_1^2}{2\mathfrak{d}_1\bar{F}_0}\frac{D_m^3}{n}. \tag{37}$$

For $\nu_n^{(3,1)}$, we first have $\mathbb{E}[(\nu_n^{(3,1)}(\varphi_j))^2] \le \mathbb{E}[(\langle\varphi_j, h\mathbf{1}_{Y_1\ge\cdot}\rangle)^2]$. With $(\mathcal{M}_5)$, this leads to

$$\mathbb{E}\left[\sup_{t\in B_m}\left(\nu_n^{(3,1)}(t)\right)^2\right] \le \frac{1}{\mathfrak{d}_1\bar{F}_0}\frac{1}{n}\mathbb{E}\left[\sum_{j\in\mathbb{J}_m}(\langle\varphi_j, h\mathbf{1}_{Y_1\ge\cdot}\rangle)^2\right] \le \frac{c_3}{\mathfrak{d}_1\bar{F}_0}\frac{\|h_m\|^2}{n} \le \frac{c_3}{\mathfrak{d}_1\bar{F}_0}\frac{\|h\|^2}{n}. \tag{38}$$

Then for $\nu_n^{(3,2)}$, we have $\mathbb{E}[(\nu_n^{(3,2)}(\varphi_j))^2] \le n^{-1}\mathbb{E}[Y_1^2\varphi_j^2(Y_1)h^2(Y_1)\mathbf{1}_{Y_1\in A}]$, thus

$$\mathbb{E}\left[\sup_{t\in B_m}\left(\nu_n^{(3,2)}(t)\right)^2\right] \le \frac{\mathbf{a}\Phi_0^2\|h\|^2}{\mathfrak{d}_1\bar{F}_0}\frac{D_m}{n}. \tag{39}$$

where we use again $(\mathcal{M}_3)$ and that $0 \le yf_Y(y) \le 1$, $y \in A$. Lemma 6.4 is proved by gathering (36), (37), (38), and (39).

6.6.2. *Proof of Lemma 6.5.* Thanks to the definition of Ω_{ρ_0}, and then thanks to (H_2),

$$\begin{aligned}\Omega_{\rho_0}^c &= \left\{\exists t\in\mathcal{S}_n, \left|\frac{\|t\|_n^2}{\|t\|_{\bar{F}_X}^2} - 1\right| > \rho_0\right\} \subset \left\{\exists t\in\mathcal{S}_n, \left|\|t\|_n^2 - \|t\|_{\bar{F}_X}^2\right| > \rho_0\bar{F}_0\|t\|^2\right\},\\ &= \left\{\inf_{t\in\mathcal{S}_n\setminus\{0\}}\left|\left\|\frac{t}{\|t\|}\right\|_n^2 - \left\|\frac{t}{\|t\|}\right\|_{\bar{F}_X}^2\right| > \rho_0\bar{F}_0\right\} = \left\{\inf_{t\in\mathcal{S}_n,\|t\|=1}\left|\eta_n(t^2)\right| > \rho_0\bar{F}_0\right\}\end{aligned}$$

with

$$\eta_n(t) = \frac{1}{n}\sum_{i=1}^n\left(\int_A t(x)\mathbf{1}_{\{Y_i\ge x\}}dx + Y_i t(Y_i)\mathbf{1}_{\{Y_i\in A\}}\right) - \int_A t(x)\bar{F}_X(x)dx,\ t\in L^2(A).$$

If $t \in \mathcal{S}_n$, it can be written $t = \sum_{j\in\mathbb{J}_{N_n}}\alpha_j\varphi_j$, with $\vec{\alpha} = (\alpha_j)_{j\in\mathbb{J}_{N_n}} \in \mathbb{R}^{D_{N_n}}$, in such a way that $\eta_n(t^2) = {}^t\vec{\alpha}\Phi\vec{\alpha}$, with $\Phi = (\Phi_{j,k})_{j,k\in\mathbb{J}_{N_n}}$ the matrix with coefficients

$$\Phi_{j,k} = \frac{1}{n}\sum_{i=1}^n\left\{\int_A\varphi_j(x)\varphi_k(x)\mathbf{1}_{Y_i\ge x}dx + Y_i\varphi_j(Y_i)\varphi_k(Y_i)\mathbf{1}_{\{Y_i\in A\}}\right\} - \int_A\varphi_j(x)\varphi_k(x)\bar{F}_X(x)dx.$$

Hence,

$$\Omega_{\rho_0}^c \subset \left\{\inf_{\vec{\alpha}\in\mathbb{R}^{D_{N_n}},\left\|\sum_{j\in\mathbb{J}_{N_n}}\alpha_j\varphi_j\right\|=1}\left|{}^t\vec{\alpha}\Phi\vec{\alpha}\right| > \rho_0\bar{F}_0\right\} \tag{40}$$

Now, the Cauchy-Schwarz inequality applied twice and Assumption $(\mathcal{M}_2)$ imply that, for all $\vec{\alpha} \in \mathbb{R}^{D_{N_n}}$, such that $\| \sum_{j\in\mathbb{J}_{N_n}} \alpha_j \varphi_j \| = 1$,

$$\begin{aligned}
{}^t\vec{\alpha}\Phi\vec{\alpha} &= \sum_{j,k\in\mathbb{J}_{N_n}} \alpha_j\alpha_k\Phi_{j,k} = \sum_{j\in\mathbb{J}_{N_n}} \alpha_j \sum_{k\in\mathbb{J}_{N_n}} \alpha_k\Phi_{j,k} \leq \sum_{j\in\mathbb{J}_{N_n}} \alpha_j \left(\sum_{k\in\mathbb{J}_n} \alpha_k^2\right)^{1/2} \left(\sum_{k\in\mathbb{J}_{N_n}} \Phi_{j,k}^2\right)^{1/2} \\
&\leq \left(\sum_{j\in\mathbb{J}_{N_n}} \alpha_j^2\right)\left(\sum_{j,k\in\mathbb{J}_{N_n}} \Phi_{j,k}^2\right)^{1/2} \leq \frac{1}{\mathfrak{d}_1}\left\|\sum_{j\in\mathbb{J}_{N_n}} \alpha_j\varphi_j\right\|^2 \left(\sum_{j,k\in\mathbb{J}_{N_n}} \Phi_{j,k}^2\right)^{1/2}, \\
&= \frac{1}{\mathfrak{d}_1}\left(\sum_{j,k\in\mathbb{J}_{N_n}} \Phi_{j,k}^2\right)^{1/2}
\end{aligned}$$

which amounts to upper-bound the spectral radius of Φ by its Frobenius norm. From this and Equation (40) we obtain

$$\Omega_{\rho_0}^c \subset \left\{ \sum_{j,k\in\mathbb{J}_{N_n}} \Phi_{j,k}^2 > \mathfrak{d}_1^2\rho_0^2\bar{F}_0^2 \right\}$$

and

$$\mathbb{P}\left(\Omega_{\rho_0}^c\right) \leq \sum_{j,k\in\mathbb{J}_{N_n}} \mathbb{P}\left(\Phi_{j,k}^2 > \frac{\mathfrak{d}_1^2\rho_0^2\bar{F}_0^2}{D_{N_n}^2}\right) = \sum_{j,k\in\mathbb{J}_{N_n}} \mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^n Z_{j,k}^{(i)} - \mathbb{E}\left[Z_{j,k}^{(i)}\right]\right| > \frac{\mathfrak{d}_1\rho_0\bar{F}_0}{D_{N_n}}\right) \tag{41}$$

where $Z_{j,k}^{(i)} = \int_A \varphi_j(x)\varphi_k(x)\mathbf{1}_{Y_i\geq x}dx + Y_i\varphi_j(Y_i)\varphi_k(Y_i)\mathbf{1}_{\{Y_i\in A\}}$. We now control independently the two terms of $Z_{j,k}^{(i)}$: $Z_{j,k}^{(i)} = Z_{j,k}^{(i,1)} + Z_{j,k}^{(i,2)}$ with $Z_{j,k}^{(i,1)} = \int_A \varphi_j(x)\varphi_k(x)\mathbf{1}_{Y_i\geq x}dx$ and $Z_{j,k}^{(i,2)} = Y_i\varphi_j(Y_i)\varphi_k(Y_i)\mathbf{1}_{\{Y_i\in A\}}$. If the basis $(\varphi_j)_{j\in\mathbb{J}_{N_n}}$ is localised (Assumption $(\mathcal{M}_{4,2})$), we have a slighly more accurate inequality:

$$\mathbb{P}\left(\Omega_{\rho_0}^c\right) \leq \sum_{j\mathbb{J}_{N_n}} \sum_{k\in\mathbb{J}_{N_n},\varphi_j\varphi_k\neq 0} \mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^n Z_{j,k}^{(i)} - \mathbb{E}\left[Z_{j,k}^{(i)}\right]\right| > \frac{\mathfrak{d}_1\rho_0\bar{F}_0}{\sqrt{c_1 D_{N_n}}}\right), \tag{42}$$

since there are only $c_1 D_{N_n}$ non-zero terms in the sum over j and k (and not $D_{N_n}^2$).

Control of $Z_{j,k}^{(i,1)}$. First remark that $|Z_{j,k}^{(i,1)}|$ is almost surely bounded by $\mathfrak{d}_2$ (thanks to assumption $(\mathcal{M}_2)$). Hence, the Hoeffding Inequality (Hoeffding, 1963), gives us the following bound: for all $x > 0$,

$$\mathbb{P}\left(\frac{1}{n}\sum_{i=1}^n Z_{j,k}^{(i,1)} - \mathbb{E}[Z_{j,k}^{(i,1)}] \geq x\right) \leq 2\exp\left(-\frac{2x^2}{4\mathfrak{d}_2^2 n}\right) = 2\exp\left(-\frac{x^2}{2\mathfrak{d}_2^2 n}\right).$$

This implies that

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^n Z_{j,k}^{(i,1)} - \mathbb{E}\left[Z_{j,k}^{(i,2)}\right]\right| > \frac{\mathfrak{d}_1\rho_0\bar{F}_0}{2D_{N_n}}\right) \leq 2\exp\left(-n\frac{\mathfrak{d}_1^2\rho_0^2\bar{F}_0^2}{4\mathfrak{d}_2^2 D_{N_n}^2}\right), \tag{43}$$

and for a localised basis

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^n Z_{j,k}^{(i,1)} - \mathbb{E}\left[Z_{j,k}^{(i,2)}\right]\right| > \frac{\mathfrak{d}_1\rho_0\bar{F}_0}{2\sqrt{c_1 D_{N_n}}}\right) \leq 2\exp\left(-n\frac{\mathfrak{d}_1^2\rho_0^2\bar{F}_0^2}{4\mathfrak{d}_2^2 c_1 D_{N_n}}\right). \tag{44}$$

Control of $Z_{j,k}^{(i,2)}$. We apply the following Bernstein Inequality (which immediatly follows from Birgé and Massart 1998, p.366).

Lemma 6.7. *(Bernstein Inequality) Let $T_1, T_2, \dots, T_n$ be independent random variables and $S_n(T) = \sum_{i=1}^n (T_i - \mathbb{E}[T_i])$. Assume that*

$$Var(T_1) \leq v^2 \text{ and } \forall \ell \geq 2, \ \frac{1}{n}\sum_{i=1}^n \mathbb{E}\left[|T_i|^\ell\right] \leq \frac{\ell!}{2} v^2 b_0^{l-2}.$$

Then, for $\eta > 0$,

$$\mathbb{P}\left(\frac{1}{n}|S_n(T)| \geq \eta\right) \leq 2\exp\left(-\frac{n\eta^2/2}{v^2 + b_0\eta}\right).$$

A distinction must be made depending on the property of the basis. If the basis is bounded (Assumption $(\mathcal{M}_{4,1})$)

$$\mathbb{E}\left[\left(Z_{j,k}^{(i,2)}\right)^2\right] = \mathbb{E}\left[Y_1^2 \varphi_j^2(Y_1)\varphi_k^2(Y_1)\right] \leq c_0^4 \mathbb{E}[Y_1^2] =: v^2,$$

and, for all integer $\ell \geq 2$,

$$\mathbb{E}\left[\left(Z_{j,k}^{(i,2)}\right)^\ell\right] = \mathbb{E}\left[Y_1^\ell \varphi_j^\ell(Y_1)\varphi_k^\ell(Y_1)\right] \leq c_0^{2\ell}\mathbb{E}[Y_1^\ell \mathbf{1}_{\{Y_1 \in A\}}] \leq c_0^{2\ell}\mathbf{a}^{\ell-2}\mathbb{E}[Y_1^2] \leq \frac{\ell!}{2} v^2 b_0^{\ell-2}.$$

with $b_0 = c_0^2 \mathbf{a}$. Then applying Bernstein Inequality (Lemma 6.7), we have

$$\begin{aligned}\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^n Z_{j,k}^{(i,2)} - \mathbb{E}\left[Z_{j,k}^{(i,2)}\right]\right| > \frac{\mathfrak{d}_1 \rho_0 \bar{F}_0}{2D_{N_n}}\right) &\leq 2\exp\left(-\frac{n}{D_{N_n}^2}\frac{\mathfrak{d}_1^2\rho_0^2\bar{F}_0^2}{8c_0^4\mathbb{E}[Y_1^2] + 4c_0^2\mathbf{a}\mathfrak{d}_1\rho_0\bar{F}_0/D_{N_n}}\right) \\ &\leq 2\exp\left(-c^{**}\frac{n}{D_{N_n}^2}\right), \end{aligned} \tag{45}$$

with $c^{**} = \rho_0^2\bar{F}_0^2/(8c_0^4\mathbb{E}[Y_1^2] + 4c_0^2\mathbf{a}\mathfrak{d}_1\rho_0\bar{F}_0)$ (using the fact that $D_{N_n} \geq 1$). Now, if the basis is not bounded but satisfies $(\mathcal{M}_{4,2})$,

$$\begin{aligned}\mathbb{E}\left[\left(Z_{j,k}^{(i,2)}\right)^2\right] &= \int_A y^2\varphi_j^2(y)\varphi_k^2(y)f_Y(y)dy \leq \int_A y\varphi_j^2(y)\varphi_k^2(y)f_Y(y)dy \leq \tilde{c}_2 D_{N_n}, \\ &\leq \mathbf{a}\|\varphi_j\|_{\infty,A}^2 \int_A \varphi_k^2(y)dy \leq \mathbf{a} \times c_2 D_{N_n} \times \mathfrak{d}_2 := v^2,\end{aligned}$$

with $yf_Y(y) \leq 1$, $y \leq \mathbf{a}$, (i) of $(\mathcal{M}_{4,2})$ and $(\mathcal{M}_2)$. For all integer $\ell \geq 2$, with the same reasoning,

$$\begin{aligned}\mathbb{E}\left[\left(Z_{j,k}^{(i,2)}\right)^\ell\right] &\leq \int_A y^\ell\varphi_j^\ell(y)\varphi_k^\ell(y)f_Y(y)dy \leq \mathbf{a}^{\ell-1}\int_A \varphi_j^\ell(y)\varphi_k^\ell(y)dy, \\ &\leq \mathbf{a}^{\ell-1}\|\varphi_j\|_{\infty,A}^\ell\|\varphi_k\|_{\infty,A}^{\ell-2}\int_A \varphi_k^2(y)dy \leq (c_2D_{N_n})^{\ell/2}(c_2D_{N_n})^{(\ell-2)/2} \times \mathfrak{d}_2, \\ &= c_2 D_{N_n}^{\ell-1}\mathfrak{d}_2 \leq \frac{\ell!}{2}v^2 b_0^{\ell-2},\end{aligned}$$

with $b_0 = D_{N_n}$. This gives, from Lemma 6.7,

$$\begin{aligned}\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n} Z_{j,k}^{(i,2)} - \mathbb{E}\left[Z_{j,k}^{(i,2)}\right]\right| > \frac{\mathfrak{d}_1\rho_0\bar{F}_0}{2\sqrt{c_1 D_{N_n}}}\right) &\leq 2\exp\left(-\frac{n}{D_{N_n}}\frac{\mathfrak{d}_1^2\rho_0^2\bar{F}_0^2}{c_1(8c_2\mathfrak{d}_2 D_{N_n} + 4\sqrt{D_{N_n}}\mathfrak{d}_1\rho_0\bar{F}_0/\sqrt{c_1})}\right)\\ &\leq 2\exp\left(-c^{**}\frac{n}{D_{N_n}^2}\right), \qquad (46)\end{aligned}$$

with $c^{**} = \mathfrak{d}_1^2\rho_0^2\bar{F}_0^2/(c_1(8c_2\mathfrak{d}_2 + 4\mathfrak{d}_1\rho_0\bar{F}_0/\sqrt{c_1}))$.

Now, gathering inequalities (41), (43) and (45) or (42), (44) and (46) (depending on the assumption $(\mathcal{M}_{4,1})$ or $(\mathcal{M}_{4,2})$ on the basis) ends the proof of Lemma 6.5.

6.6.3. *Proof of Lemma 6.6.* We start from the decomposition given in the proof of Lemma 6.4. We thus have to bound the terms of the right-hand-side of the following inequality:

$$\begin{aligned}\mathbb{E}\left[\left(\sup_{t\in B_{m\vee\widehat{m}}} \nu_n^2(t) - p(m,\widehat{m})\right)_+\right] &\leq 4\sum_{l\in\{1,2,(3,2)\}}\sum_{m'=1}^{N_n}\mathbb{E}\left[\left(\sup_{t\in B_{m\vee m'}}\left(\nu_n^{(l)}(t)\right)^2 - \frac{p_l(m,m')}{3}\right)_+\right]\\ &\quad + 4\mathbb{E}\left[\sup_{t\in B_{N_n}}\left(\nu_n^{(3,1)}(t)\right)^2\right]\end{aligned}$$

with

$$\begin{aligned}p_1(m,m') &= \bar{\kappa}_0\Phi_0^2(\bar{F}_0\mathfrak{d}_1)^{-1}\frac{D_{m\vee m'}}{n},\\ p_2(m,m') &= \bar{\kappa}_0\mathbf{a}^2\Phi_1^2(\bar{F}_0\mathfrak{d}_1)^{-1}\frac{D_{m\vee m'}^3}{n},\\ p_3(m,m') &= \bar{\kappa}_0\mathbf{a}\Phi_0^2\|h\|^2(\mathfrak{d}_1\bar{F}_0)^{-1}\frac{D_{m\vee m'}}{n}.\end{aligned}$$

For the last term, we obtain from (38) that $\mathbb{E}\left[\sup_{t\in B_{N_n}}\left(\nu_n^{(3,1)}(t)\right)^2\right] \leq c_3\|h\|^2/(\mathfrak{d}_1\bar{F}_0 n)$.

For the three other terms, the guiding idea to bound each of them is to apply the following version of the Talagrand Inequality which can be found, for instance, in Lacour (2008).

Theorem 6.1 (Talagrand Inequality)**.** *Let $\mathcal{F}$ be a set of uniformly bounded functions, which have a countable dense sub-family for the infinite norm. Let $(V_1,\ldots,V_n)$ be independent random variables and*

$$Z = \sup_{f\in\mathcal{F}}\left|\frac{1}{n}\sum_{i=1}^{n}\left(f(V_i) - \mathbb{E}\left[f(V_i)\right]\right)\right|.$$

Consider b, v, and H, such that

$$b \geq \sup_{f\in\mathcal{F}}\|f\|_\infty, v \geq \sup_{f\in\mathcal{F}}\frac{1}{n}\sum_{i=1}^{n} Var(f(V_i)) \text{ and } H \geq \mathbb{E}[Z].$$

Then, for every $\varepsilon > 1$, there exist numerical positive constants C_1, C_2, c_1 and c_2 such that

$$\mathbb{E}\left[\left(Z^2 - \varepsilon H^2\right)_+\right] \leq C_1\frac{v}{n}\exp\left(-c_1\frac{nH^2}{v}\right) + C_2\frac{b^2}{n^2}\exp\left(-c_2\frac{nH}{b}\right).$$

Control of $\nu_n^{(1)}$. The aim is to apply the Talagrand Inequality (Theorem 6.1). We first compute the key quantities involved in the concentration results.

The quantity $(H^{(1)})^2 = \Phi_0^2(\bar{F}_0\mathfrak{d}_1)^{-1}(D_{m\vee m'}/n)$ follows from the proof of Lemma 6.4.

Let us compute $b^{(1)}$, a uniform upper-bound for the sup-norm of the functions to which the inequality is applied. For $t \in B_{m\vee m'}$, we have by Inequality (15) in Lemma 6.2, $\sup_{t\in B_{m\vee m'}} \|t\|^2_{\infty,A} \leq \Phi_0^2(\bar{F}_0\mathbf{a})^{-1}D_{m\vee m'} =: \left(b^{(1)}\right)^2$. We turn now on the upper-bound $v^{(1)}$ on $\sup_{t\in B_{m\vee m'}} n^{-1}\sum_{i=1}^n \mathrm{Var}(t(Y_i))$:

$$\begin{aligned}
\sup_{t\in B_{m\vee m'}} \frac{1}{n}\sum_{i=1}^n \mathrm{Var}(t(Y_i)) &\leq \sup_{t\in B_{m\vee m'}} \mathbb{E}\left[t^2(Y)\right] = \sup_{t\in B_{m\vee m'}} \int_A t^2(y)f_Y(y)dy \\
&\leq \sup_{t\in B_{m\vee m'}} \|t\|_{\infty,A}\int_A |t(y)|f_Y(y)dy \leq (\sqrt{\mathfrak{d}_1\bar{F}_0})^{-1}\Phi_0\|f_Y\|\|t\|\sqrt{D_{m\vee m'}}, \\
&\leq \sqrt{\mathfrak{d}_1}^{-1}\Phi_0\|f_Y\|\sqrt{D_{m\vee m'}} =: v^{(1)}.
\end{aligned}$$

We apply Theorem 6.1 and get, for all $\varepsilon^{(1)} > 1$,

$$\mathbb{E}\left[\left(\sup_{t\in B_{m\vee m'}}\left(\nu_n^{(1)}(t)\right)^2 - \varepsilon^{(1)}\Phi_0^2(\mathfrak{d}_1\bar{F}_0)^{-1}\frac{D_{m\vee m'}}{n}\right)_+\right] \leq C\left\{\frac{1}{n}e^{-c_1^{(1)}\sqrt{D_{m\vee m'}}} + \frac{D_{m\vee m'}}{n^2}e^{-c_2^{(2)}\sqrt{n}}\right\},$$

where $C, c_1^{(1)}, c_2^{(1)} > 0$ are some constants. We sum over m, and choose $\bar{\kappa}_0$ in the penalty $p(m,m')$ such that $p_1(m,m') \geq 3\varepsilon^{(1)}$. This leads to

$$\sum_{m'=1}^{N_n}\mathbb{E}\left[\left(\sup_{t\in B_{m\vee m'}}\left(\nu_n^{(1)}(t)\right)^2 - \frac{p_1(m,m')}{3}\right)_+\right] \leq \frac{C}{n}. \tag{47}$$

Control of $\nu_n^{(2)}$. The proof is very similar to the one of the previous paragraph. We recall that $\nu_n^{(2)}(t) := \frac{1}{n}\sum_{i=1}^n Y_it'(Y_i) - \mathbb{E}[Y_it'(Y_i)]$ and compute the quantity $b^{(2)}$, $v^{(2)}$ and $H^{(2)}$ required to apply the Talagrand Inequality. Again, from the proof of Lemma 6.4, we get $(H^{(2)})^2 = \mathbf{a}^2\Phi_1^2(2\bar{F}_0\mathfrak{d}_1)^{-1}(D^3_{m\vee m'}/n)$.

Let us now prove the uniform boundedness. Let $t \in B_{m\vee m'}$, from the Cauchy-Schwarz Inequality, assumptions $(\mathcal{M}_2)$ and $(\mathcal{M}_3)$, we get,

$$\begin{aligned}
\sup_{t\in B_{m\vee m'}}\sup_{x\in A}\left\{x^2t'(x)^2\right\} &\leq \mathbf{a}^2 \sup_{\vec{\alpha}\in\mathbb{R}^{D_{m\vee m'}},\ {}^t\vec{\alpha}\vec{\alpha}\leq\bar{F}_0^{-1}}\ \sup_{x\in A}\left(\sum_{j\in\mathbb{J}_{m\vee m'}}\alpha_j\varphi_j'(x)\right)^2 \\
&\leq \mathbf{a}^2(\mathfrak{d}_1\bar{F}_0)^{-1}\sup_{x\in A}\sum_{j\in\mathbb{J}_{m\vee m'}}\varphi_j'(x)^2 \leq \mathbf{a}^2\Phi_1^2(\bar{F}_0\mathfrak{d}_1)^{-1}D^3_{m\vee m'} =: \left(b^{(2)}\right)^2.
\end{aligned}$$

Now, $\sup_{t\in B_{m\vee m'}} n^{-1}\sum_{i=1}^n \mathrm{Var}(Y_it'(Y_i)) \leq \sup_{t\in B_{m\vee m'}}\mathbb{E}[Y^2t'(Y)^2]$. But, using $yf_Y(y) \leq 1$ for any $y \geq 0$, we get for any $t \in B_{m\vee m'}$,

$$\mathbb{E}\left[Y^2t'(Y)^2\right] = \int_0^{\mathbf{a}} y^2(t'(y))^2f_Y(y)dy \leq \mathbf{a}\|t'\|^2 \leq \mathbf{a}\Phi_2^2D^2_{m\vee m'}/(\mathfrak{d}_1\bar{F}_0),$$

with the last inequality of $(\mathcal{M}_3)$. Therefore,

$$\sup_{t\in B_{m\vee m'}}\frac{1}{n}\sum_{i=1}^n \mathrm{Var}(Y_it'(Y_i)) \leq \mathbf{a}\Phi_1^2(\bar{F}_0\mathfrak{d}_1)^{-1}D^2_{m\vee m'} := \left(v^{(2)}\right)^2.$$

Then, by Theorem 6.1, for all $\varepsilon^{(2)} > 1$, there exist $C, c_1^{(2)}, c_2^{(2)} > 0$ such that

$$\mathbb{E}\left[\left(\sup_{t\in B_{m\vee m'}}\left(\nu_n^{(2)}(t)\right)^2 - \varepsilon^{(2)}(H^{(2)})^2\right)_+\right] \leq C\left\{\frac{D_{m\vee m'}^2}{n}e^{-c_1^{(2)}D_{m\vee m'}} + \frac{D_{m\vee m'}^3}{n^2}e^{-c_2^{(2)}\sqrt{n}}\right\}.$$

We sum over m, and choose $\bar{\kappa}_0$ in $p(m,m')$ such that $p_2(m,m') \geq 3\varepsilon^{(2)}$. This leads to

$$\sum_{m'=1}^{N_n}\mathbb{E}\left[\left(\sup_{t\in B_{m\vee m'}}\left(\nu_n^{(2)}(t)\right)^2 - \frac{p_2(m,m')}{3}\right)_+\right] \leq \frac{C}{n}. \tag{48}$$

Control of $\nu_n^{(3,2)}$. We begin as usual by computing the quantities involved in the Talagrand Inequality with $(H^{(3,2)})^2 = \bar{F}_0^{-1}\mathfrak{d}_1 D_{m\vee m'}\|h\|^2/n$, from the proof of Lemma 6.4. Next, thanks to (15),

$$\begin{aligned}\sup_{t\in B_{m\vee m'}}\sup_{y\in A}|yt(y)h(y)| &\leq \mathbf{a}\sup_{\vec{\alpha}\in\mathbb{R}^{D_{m\vee m'}},\vec{\alpha}\vec{\alpha}^t\leq\bar{F}_0^{-1}}\sup_{y\in A}\left|\sum_{j\in\mathbb{J}_{m\vee m'}}\alpha_j\varphi_j(y)h(y)\right| \\ &\leq \Phi_0(\bar{F}_0\mathfrak{d}_1\mathbf{a})^{-1/2}\sqrt{D_m}\|h\|_{\infty,A} =: b^{(3,2)}.\end{aligned}$$

Then, as $yf_Y(y) \leq 1$ (see (5)),

$$\begin{aligned}\sup_{t\in B_{m\vee m'}}\frac{1}{n}\sum_{i=1}^n \mathrm{Var}\left(Y_i t(Y_i)h(Y_i)\right) &\leq \sup_{t\in B_{m\vee m'}}\mathbb{E}\left[Y^2t^2(Y)h^2(Y)\right], \\ &= \sup_{t\in B_{m\vee m'}}\int_A y^2t^2(y)h^2(y)f_Y(y)dy, \\ &\leq \sup_{t\in B_{m\vee m'}}\mathbf{a}\|h\|_{\infty,A}^2\|t\|^2, \\ &\leq \mathbf{a}(\mathfrak{d}_1\bar{F}_0)^{-1}\|h\|_{\infty,A}^2 := \left(v^{(3,2)}\right)^2.\end{aligned}$$

Applying Talagrand's Inequality of Theorem 6.1 is exactly similar to what has been done for the previous processes $\nu_n^{(1)}$ and $\nu_n^{(2)}$. Since $D_m^2 \geq \ln^2(n)$, and as soon as $\ln^2(n) \geq \|h\|^2$, we obtain

$$\begin{aligned}&\sum_{m'=1}^{N_n}\mathbb{E}\left[\left(\sup_{t\in B_{m\vee m'}}\left(\nu_n^{(3,2)}(t)\right)^2 - \varepsilon^{(3,2)}\bar{F}_0^{-1}\frac{D_{m\vee m'}^3}{n}\right)_+\right] \\ &\quad\leq \sum_{m'=1}^{N_n}\mathbb{E}\left[\left(\sup_{t\in B_{m\vee m'}}\left(\nu_n^{(3,2)}(t)\right)^2 - \varepsilon^{(3,2)}\bar{F}_0^{-1}\|h_A\|^2\frac{D_{m\vee m'}}{n}\right)_+\right] \leq \frac{C}{n}.\end{aligned}$$

If $\bar{\kappa}_0$ in the penalty $p(m,m')$ is such that $p_3(m,m') \geq 6\varepsilon^{(3,2)}$, this leads to

$$\sum_{m'=1}^{N_n}\mathbb{E}\left[\left(\sup_{t\in B_{m\vee m'}}\left(\nu_n^{(3,2)}(t)\right)^2 - \frac{p_3(m,m')}{6}\right)_+\right] \leq \frac{C}{n}. \tag{49}$$

We end the proof of Lemma 6.6 by gathering (47), (48), and (49).

6.7. **Proof of Proposition 6.1.** The guideline for the proof is the following.

Part 1: We prove that there exist $L'' > 0$ and $0 < \varepsilon' < \min(\mathbf{a}, \varepsilon)$ such that $f_X \in \mathcal{H}^\beta([0, \mathbf{a}+\varepsilon'], L'')$.

Part 2: We prove that the result of Part 1 implies that there exists $L' > 0$ such that $f_Y \in \mathcal{H}^{\beta+1}([\mathbf{a}-\varepsilon', \mathbf{a}+\varepsilon'], L')$.

Proof of Part 1. We note $H(x) = -\ln(\bar{F}_X(x))$, under Assumption (H_2), and by continuity of $\bar{F}_X$, there exists $0 < \varepsilon' < \min(\mathbf{a}, \varepsilon)$ such that $\bar{F}_X(x) > 0$ for all $x \in [0, \mathbf{a}+\varepsilon']$. Hence, on the set $[\mathbf{a}-\varepsilon', \mathbf{a}+\varepsilon']$, the function H is well defined and differentiable and we have $H'(x) = h(x)$, which implies that $H \in \mathcal{H}^{\beta+1}([0, \mathbf{a}+\varepsilon], L)$. Moreover, since $\bar{F}_X(x) = \exp(-H(x))$, the function $\bar{F}_X$ is $(\ell+1)$-times differentiable on $[0, \mathbf{a}+\varepsilon']$ (with $\ell = \lfloor \beta \rfloor$) and consequently f_X is ℓ-times differentiable.

We now prove that there exists $L'' > 0$ such that, for all $x, y \in [0, \mathbf{a}+\varepsilon']$,

$$\left| f_X^{(\ell)}(x) - f_X^{(\ell)}(y) \right| \leq L''|x-y|^{\beta-\ell} \tag{50}$$

which ends the proof of Part 1.

By induction, we easily show that

$$\bar{F}_X^{(\ell)}(x) = -h^{(\ell-1)}(x)e^{-H(x)} + R_\ell(x)$$

with $R_1 \equiv 0$ and, for all $k = 1, ..., \ell-1$, $R_{k+1} = h^{(k-1)}(x)h(x)e^{-H(x)} + R_k'(x)$.

Hence, for all $x, y \in [0, \mathbf{a}+\varepsilon']$,

$$\begin{aligned}
|f_X^{(\ell)}(x) - f_X^{(\ell)}(y)| &= |\bar{F}_X^{(\ell+1)}(x) - \bar{F}_X^{(\ell+1)}(y)| \\
&\leq |h^{(\ell)}(x)e^{H(x)} - h^{(\ell)}(y)e^{H(y)}| + |R_\ell(x) - R_\ell(y)| \\
&\leq |h^{(\ell)}(x)e^{H(x)} - h^{(\ell)}(y)e^{H(x)}| + |h^{(\ell)}(y)e^{H(x)} - h^{(\ell)}(y)e^{H(y)}| \\
&\quad + |R_\ell(x) - R_\ell(y)| \\
&\leq e^{\|H\|_{\infty,[0,\mathbf{a}+\varepsilon]}} |h^{(\ell)}(x) - h^{(l)}(y)| + \|h^\ell\|_{\infty,[0,\mathbf{a}+\varepsilon]} \left| e^{H(x)} - e^{H(y)} \right| \\
&\quad + |R_\ell(x) - R_\ell(y)|
\end{aligned} \tag{51}$$

Since $h \in \mathcal{H}^\beta([0, \mathbf{a}+\varepsilon'], L)$, both H and $h^{(\ell)}$ are continuous on the compact set $[0, \mathbf{a}+\varepsilon']$ (we recall that $\bar{F}_X > 0$ on this set) , hence $\|H\|_{\infty,[0,\mathbf{a}+\varepsilon]} < \infty$ and $\|h^\ell\|_{\infty,[0,\mathbf{a}+\varepsilon]} < \infty$. Moreover, e^H is differentiable, hence, since $[0, \mathbf{a}+\varepsilon']$ is compact, it is Lipschitz continuous. By induction, we can also prove that R_ℓ is differentiable (and then Lipschitz continuous). Hence, from Equation (51), using again the fact that $h \in \mathcal{H}^\beta([0, \mathbf{a}+\varepsilon'], L)$ and the compactness of $[0, \mathbf{a}+\varepsilon']$, we know that there exists $L_1 > 0$ such that

$$|f_X^{(\ell)}(x) - f_X^{(\ell)}(y)| \leq e^{\|H\|_{\infty,[0,\mathbf{a}+\varepsilon]}} L|x-y|^{\beta-\ell} + L_1|x-y| \leq L''|x-y|^{\beta-\ell},$$

with $L'' = e^{\|H\|_{\infty,[0,\mathbf{a}+\varepsilon]}} L + L_1(a+\varepsilon')^{1-\beta+\ell}$ which implies Equation (50) and ends the proof of Part 1.

Proof of Part 2. From the fact that $yf_Y'(y) = -f_X(x)$ (see Equation (4)), it can be shown by induction that, if f_X is ℓ-times differentiable on $[0, \mathbf{a}+\varepsilon']$, then f_Y is $(\ell+1)$-times differentiable on $[\mathbf{a}-\varepsilon', \mathbf{a}+\varepsilon']$ (we recall that $\mathbf{a}-\varepsilon' > 0$) and

$$f_Y^{(\ell+1)}(y) = -\frac{f_X^{(\ell)}(y)}{y} + \tilde{R}_\ell(y)$$

with $\widetilde{R}_0 \equiv 0$ and $\widetilde{R}_{\ell+1}(y) = \frac{f^{(\ell)}(y)}{y^2} + \widetilde{R}'_\ell(y)$.

Hence, for all $x, y \in [\mathbf{a}-\varepsilon', \mathbf{a}+\varepsilon']$, using the fact that $f_X \in \mathcal{H}^\beta([0, \mathbf{a}+\varepsilon'], L'')$, and that $\widetilde{R}_\ell$ is differentiable – which can be proved by induction –, hence Lipschitz continuous on the compact set $[\mathbf{a}-\varepsilon', \mathbf{a}+\varepsilon']$, there exists a constant $L_2 > 0$ such that

$$\begin{aligned}
|f_Y^{(\ell+1)}(y) - f_Y^{(\ell+1)}(x)| &\leq \left|\frac{f_X^{(\ell)}(x)}{x} - \frac{f_X^{(\ell)}(y)}{y}\right| + |\widetilde{R}_\ell(x) - \widetilde{R}_\ell(y)| \\
&\leq \left|\frac{f_X^{(\ell)}(x)}{x} - \frac{f_X^{(\ell)}(x)}{y}\right| + \left|\frac{f_X^{(\ell)}(x)}{y} - \frac{f_X^{(\ell)}(y)}{y}\right| + |\widetilde{R}_\ell(x) - \widetilde{R}_\ell(y)| \\
&\leq \|f_X^{(\ell)}\|_{\infty,[\mathbf{a}-\varepsilon',\mathbf{a}+\varepsilon']} \left|\frac{1}{x} - \frac{1}{y}\right| + \frac{1}{\mathbf{a}-\varepsilon'}|f_X^{(\ell)}(x) - f_X^{(\ell)}(y)| \\
&\quad + |\widetilde{R}_\ell(x) - \widetilde{R}_\ell(y)| \\
&\leq \frac{\|f_X^{(\ell)}\|_{\infty,[\mathbf{a}-\varepsilon',\mathbf{a}+\varepsilon']}}{\mathbf{a}-\varepsilon'}|x-y| + \frac{L''}{\mathbf{a}-\varepsilon'}|x-y|^{\beta-\ell} + L_2|x-y|.
\end{aligned}$$

This implies that $f_Y \in \mathcal{H}^{\beta+1}([\mathbf{a}-\varepsilon', \mathbf{a}+\varepsilon'], L')$ with $L' = \frac{\|f_X^{(\ell)}\|_{\infty,[\mathbf{a}-\varepsilon',\mathbf{a}+\varepsilon']}}{\mathbf{a}-\varepsilon'} + \frac{L''}{\mathbf{a}-\varepsilon'} + L_2(\mathbf{a}-\varepsilon')^{1-\beta-\ell}$.

References

M. Abbaszadeh, C. Chesneau, and H. Doosti. Nonparametric estimation of density under bias and multiplicative censoring via wavelet methods. *Statist. Probab. Lett.*, 82(5):932–941, 2012.

M. Abbaszadeh, C. Chesneau, and H. Doosti. Multiplicative censoring: estimation of a density and its derivatives under the L_p-risk. *REVSTAT*, 11(3):255–276, 2013.

K. E. Andersen and M. B. Hansen. Multiplicative censoring: density estimation by a series expansion approach. *J. Statist. Plann. Inference*, 98(1-2):137–155, 2001.

A. Antoniadis, G. Grégoire, and G. Nason. Density and hazard rate estimation for right-censored data by using wavelet methods. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 61(1):63–84, 1999.

M. Asgharian, M. Carone, and V. Fakoor. Large-sample study of the kernel density estimators under multiplicative censoring. *Ann. Statist.*, 40(1):159–187, 2012.

A. Barron, L. Birgé, and P. Massart. Risk bounds for model selection via penalization. *Probab. Theory Related Fields*, 113(3):301–413, 1999.

J.-P. Baudry, C. Maugis, and B. Michel. Slope heuristics: overview and implementation. *Stat. Comput.*, 22(2):455–470, 2012.

K. Bertin, C. Lacour, and V. Rivoirard. Adaptive pointwise estimation of conditional density function. *Ann. Inst. Henri Poincaré Probab. Stat.*, 52(2):939–980, 2016.

L. Birgé and P. Massart. Minimum contrast estimators on sieves: exponential bounds and rates of convergence. *Bernoulli*, 4(3):329–375, 1998.

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

E. Brunel, F. Comte, and V. Genon-Catalot. Nonparametric density and survival function estimation in the multiplicative censoring model. *TEST*, 2015. to appear.

Y. P. Chaubey, C. Chesneau, and H. Doosti. Adaptive wavelet estimation of a density from mixtures under multiplicative censoring. *Statistics*, 49(3):638–659, 2015.

C. Chesneau. Wavelet estimation of a density in a GARCH-type model. *Comm. Statist. Theory Methods*, 42(1):98–117, 2013.

F. Comte. *Estimation non-paramétrique*. Spartacus IDH, 2015.

F. Comte and C. Dion. Nonparametric Estimation in a Multiplicative Censoring Model with Symmetric Noise. *J. Nonparametr. Stat.*, 2016. to appear.

F. Comte, S. Gaïffas, and A. Guilloux. Adaptive estimation of the conditional intensity of marker-dependent counting processes. *Ann. Inst. Henri Poincaré Probab. Stat.*, 47(4):1171–1196, 2011.

C. de Boor. *A practical guide to splines*, volume 27 of *Applied Mathematical Sciences*. Springer-Verlag, New York-Berlin, 1978.

R. A. DeVore and G. G. Lorentz. *Constructive approximation*, volume 303 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 1993.

S. Efromovich. Minimax theory of nonparametric hazard rate estimation: efficiency and adaptation. *Ann. Inst. Statist. Math.*, 68(1):25–75, 2016.

R. M. Gray. Toeplitz and circulant matrices: A review. *Foundations and Trends® in Communications and Information Theory*, 2(3):155–239, 2006.

W. Härdle, G. Kerkyacharian, D. Picard, and A. Tsybakov. *Wavelets, approximation, and statistical applications*, volume 129 of *Lecture Notes in Statistics*. Springer-Verlag, New York, 1998.

W. Hoeffding. Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.*, 58:13–30, 1963.

F. Killmann and E. von Collani. A note on the convolution of the uniform and related distributions and their use in quality control. *Economic Quality Control*, 16(1):17–41, 2001.

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

P. Massart. *Concentration inequalities and model selection*, volume 1896 of *Lecture Notes in Mathematics*. Springer, Berlin, 2007. Lectures from the 33rd Summer School on Probability Theory held in Saint-Flour, July 6–23, 2003, With a foreword by Jean Picard.

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

P. N. Patil. On the least squares cross-validation bandwidth in hazard rate estimation. *Ann. Statist.*, 21(4):1792–1810, 1993b.

S. Plancade. Model selection for hazard rate estimation in presence of censoring. *Metrika*, 74 (3):313–347, 2011.

G. Rebelles. Pointwise adaptive estimation of a multivariate density under independence hypothesis. *Bernoulli*, 21(4):1984–2023, 2015.

B. Van Es, P. Spreij, and H. van Zanten. Nonparametric volatility density estimation for discrete time models. *J. Nonparametr. Stat.*, 17(2):237–251, 2005.

Y. Vardi. Multiplicative censoring, renewal processes, deconvolution and decreasing density: nonparametric estimation. *Biometrika*, 76(4):751–761, 1989.

Y. Vardi and C.-H. Zhang. Large sample study of empirical distributions in a random-multiplicative censoring model. *Ann. Statist.*, 20(2):1022–1039, 1992.