LAGUERRE ESTIMATION UNDER CONSTRAINT AT A SINGLE POINT

F. COMTE[(1)] AND C. DION[(2)]

ABSTRACT. This paper presents a general methodology for nonparametric estimation of a function s related to a nonnegative real random variable X, under a constraint of type $s(0) = c$. Three different examples are investigated: the direct observations model (X is observed), the multiplicative noise model ($Y = XU$ is observed, with U following a uniform distribution) and the additive noise model ($Y = X + V$ is observed where V is a nonnegative nuisance variable with known density). When a projection estimator of the target function is available, we explain how to modify it in order to obtain an estimator which satisfies the constraint. We extend risk bounds from the initial to the new estimator. Moreover if the previous estimator is adaptive in the sense that a model selection procedure is available to perform the squared bias/variance trade-off, we propose a new penalty also leading to an oracle type inequality for the new constrained estimator. The procedure is illustrated on simulated data, for density and survival function estimation.

Keywords. Adaptive estimation, constrained estimator, Laguerre basis, nonparametric projection estimator

AMS Classification. 62G05-62G07

1. INTRODUCTION

In the statistical literature, different types of global constraints have been studied from nonparametric functional estimation point of view, such as convexity or monotonicity constraints. Specific procedures have been proposed to obtain for instance decreasing density estimators, see Huang and Wellner (1995), Balabdaoui and Wellner (2007). We may also mention the proposal of Chernozhukov *et al.* (2009) where an estimator of a weakly increasing function is modified to get a weakly increasing estimator, with no influence on the risk value. These authors propose an associated **R**-package `Rearrangement` which may be used in our setting when considering decreasing survival functions.

We are interested in a different question, namely: given an estimator built in the Laguerre basis, can we coherently modify it in order to fix its value in one specific point?

More precisely, consider a square integrable function s with support $\mathbb{R}^+$, as this currently occurs for lifetimes densities or survival functions in survival analysis, reliability and actuarial sciences. When s is square integrable, a natural idea is to consider its development in the Laguerre basis, defined by

$$\varphi_0(x) = \sqrt{2}e^{-x}, \quad \varphi_k(x) = \sqrt{2}L_k(2x)e^{-x} \text{ for } k \geq 1, \quad x \geq 0, \qquad (1.1)$$

with L_k the Laguerre polynomials

$$L_k(x) = \sum_{j=0}^{k}(-1)^j \binom{k}{j}\frac{x^j}{j!}. \qquad (1.2)$$

Indeed, the Laguerre basis is orthonormal for the integral scalar product on $\mathbb{R}^+$, $\langle s, t \rangle = \int_0^{+\infty} s(x)t(x)dx$. In other words, we write that $s = \sum_{j \geq 0} a_j(s)\varphi_j$ with $a_j(s) = \langle s, \varphi_j \rangle$. Then we consider that observations $Y_1, \ldots, Y_n$ related to s are available and allow us to build a projection estimator $\widehat{s}_m$ of s: $\widehat{s}_m = \sum_{j=0}^{m-1} \widehat{a}_j \varphi_j$ where $\widehat{a}_j$ for $j = 0, \ldots, m-1$ are known functions of the observations. Moreover we assume that $\mathbb{E}[\widehat{a}_j] = a_j(s)$ and call $\widehat{s}_m$ a projection estimator of s as it is an unbiased estimator of $s_m = \sum_{j=0}^{m-1} a_j(s)\varphi_j$, the orthogonal projection of s on the m-dimensional space $\mathcal{S}_m = \text{span}(\varphi_0, \ldots, \varphi_{m-1})$.

Now, our question in this paper is the following. If s is subject to a constraint $s(0) = c$, can we coherently modify $\widehat{s}_m$ and propose a new estimate $\widetilde{s}_m$ such that $\widetilde{s}_m(0) = c$. Indeed, obviously, if s is a survival function supported on $\mathbb{R}^+$, it must satisfy the constraint with $c = 1$; and there are examples where densities should be constrained by $c = 0$. As noted in Mabon (2015) and Comte and Dion (2016), this starting value is rather unstable in the Laguerre basis, so including the constraint in the procedure is likely to improve the estimator.

This is the problem solved in this work. We propose a general definition of the constrained estimator and prove that its mean-square integrated risk on $\mathbb{R}^+$ is comparable to the risk of $\widehat{s}_m$. The idea is to write the projection estimator $\widehat{s}_m$ as a minimum contrast estimator and then to modify this contrast by standard Lagrange multiplier strategy. If a model selection procedure allowing for a relevant choice of m is available for $\widehat{s}_m$, a modification is proposed, ensuring also a relevant choice of m for $\widetilde{s}_m$, under simple conditions.

The procedure is illustrated through several examples. First, s can simply be the density or the survival function of the random observations at hand, and we discuss this setting in relation with general procedures described in Efromovich (1999) or Massart (2007).

Our second example is the so-called multiplicative censoring model, a terminology introduced by Vardi (1989), with applications in survival analysis in van Es *et al.* (2000). In this model, the observations are the $Y_i = X_i U_i$ with X_i and U_i independent and U_i following a uniform distribution on $[0, 1]$. The random variables $(X_i)_{1 \leq i \leq n}$ are independent and identically distributed (*i.i.d.*) and so are the $(U_i)_{1 \leq i \leq n}$. We observe an *i.i.d.* sample of Y_i's while we are interested in the density or the survival function of X. Consider a setting where X is the true positive survival time of a patient, who has been sampled, while Y represents his observed survival time, then van Es *et al.* (2000) explain that it is natural to assume that the time point of sampling is uniformly distributed over the whole survival period of length X. Besides, Andersen and Hansen (2001) study this problem as an inverse problem, projection wavelet estimators are studied by Abbaszadeh *et al.* (2013), Brunel *et al.* (2016) propose kernel estimators of the density and the survival function, and Belomestny *et al.* (2016) consider Laguerre projection estimators when U_i follows a general $\beta(1, k)$ distribution (the case $k = 1$ corresponds to the uniform). Here, we study a projection Laguerre estimator of both the density and the survival function: our bound for density estimator is improved compared to all these works and in particular Belomestny *et al.* (2016), under a mild assumption; moreover, the bounds for the survival function are new in the Laguerre setting. They can however be related to the model studied in Comte and Dion (2016), where U follows a uniform density on $[1 - \mathsf{a}, 1 + \mathsf{a}]$ for $0 < \mathsf{a} < 1$: the application setting is then corresponding to amplification/attenuation of a signal. We show that we can deduce from these projection estimators, constrained estimators with the same theoretical properties as the first step estimators.

The third example is the convolution model, $Y_i = X_i + V_i$, where all variables are nonnegative, *i.i.d.*, and V is a nuisance process with known density; the function of interest is the density or the survival function of X_i while only the Y_i's are observed. Jirak and Reiß (2014) study a one-sided error regression model and explain an application to financial data: to infer bidders' private values from observed bids. The convolution model has been widely investigated, mainly with a Fourier approach (see Carroll and Hall, 1988; Fan, 1991; Comte *et al.*, 2006; Pensky and Vidakovic, 1999, for example). Recently Mabon (2015) proposed a projection estimator of the density and the survival function supported by $\mathbb{R}^+$. This approach relies on a Laguerre projection estimator and can be rather straightforwardly used in the present work to deduce constrained estimators.

The paper is organised as follows. In Section 2 we present the general method. The projection estimator in the Laguerre basis is constructed, and its constrained version is deduced, their risk are compared, and in particular the bias thanks to results from Bongioanni and Torrea (2009); Comte and Genon-Catalot (2015). The conditions to obtain a model selection result for $\widehat{s}_m$ and deduce a similar result for estimator $\widetilde{s}_m$ are given. In section 3, the results are applied in the case of direct observation of the variable of interest. Section 4, is dedicated to the multiplicative noise model case. The procedure is applied to the additive model in Section 5. In all cases, both density and survival

function estimation are considered. Finally some numerical results of the method in several examples are presented, to highlight the good performances of our estimators.

2. A general strategy

2.1. **Notations.** The space $\mathbb{L}^2(\mathbb{R}^+)$ is the space of square integrable functions on the positive real line. The associated $\mathbb{L}^2$-norm is denoted $\|t\|^2 = \int_{\mathbb{R}^+} |t(x)|^2 dx$. Finally, the supremum norm of a bounded function t is denoted by $\|t\|_\infty = \sup_{x \in \mathbb{R}^+} |t(x)|$. The Laguerre basis is defined by (1.1) and (1.2). It satisfies the orthonormality property $\langle \varphi_j, \varphi_k \rangle = \delta_{j,k}$ where $\delta_{j,k}$ is the Kronecker symbol, equal to 1 if $j = k$ and to zero otherwise. The following properties are used in the sequel (see Abramowitz and Stegun (1966)):

$$\forall j \geq 0, \|\varphi_j\|_\infty \leq \sqrt{2}, \text{ and } \varphi_j(0) = \sqrt{2}. \tag{2.1}$$

Any function of $\mathbb{L}^2(\mathbb{R}^+)$ can be decomposed on this basis.

2.2. **Estimation method and assumptions.** Let us denote the sample of observations: $(Y_i)_{1 \leq i \leq n}$ related to the variables of interest $(X_i)_{1 \leq i \leq n}$.

All along the paper our strategy is a projection strategy requiring that the following condition holds:

(A1)(s) $\quad s \in \mathbb{L}^2(\mathbb{R}^+)$.

Under **(A1)**(s), the development $s = \sum_{j \geq 0} a_j(s)\varphi_j$ with $a_j(s) = \langle s, \varphi_j \rangle$, holds in $\mathbb{L}^2$. As $\varphi_j(0) = \sqrt{2}$ for all j, the following condition ensures that $s(0)$ exists and defines the constraint:

(A2)(s) $\sum_{\ell \geq 0} |a_\ell(s)| < +\infty$ and $s(0) = c$.

Note that, by (2.1), this condition also implies that s is continuous and bounded, with $\|s\|_\infty \leq \sqrt{2} \sum_{\ell \geq 0} |a_\ell(s)| < +\infty$.

We consider the following estimator $\widehat{s}_m$ of s on the subspace $\mathcal{S}_m = \text{span}\{\varphi_0, \varphi_1, \ldots, \varphi_{m-1}\}$:

$$\widehat{s}_m = \sum_{j=0}^{m-1} \widehat{a}_j \varphi_j,$$

with $\widehat{a}_j$ computed from a known transformation of the observations $Y_1, \ldots, Y_n$. We assume that

$$\forall j \in \mathbb{N}, \quad \mathbb{E}[\widehat{a}_j] = a_j(s) \text{ with } a_j(s) := \langle s, \varphi_j \rangle.$$

Clearly this implies that

$$\mathbb{E}[\widehat{s}_m] = s_m := \sum_{j=0}^{m-1} a_j(s)\varphi_j,$$

where s_m is the orthogonal projection of s on $\mathcal{S}_m$. Thus $\widehat{s}_m$ is an unbiased estimator of s_m and is called a projection estimator of s. As a consequence, the following decomposition of the MISE (Mean Integrated Squared Error) holds:

$$\mathbb{E}\left[\|\widehat{s}_m - s\|^2\right] = \|s_m - s\|^2 + \mathbb{E}[\|\widehat{s}_m - s_m\|^2].$$

A useful description of the estimator is to note that $\widehat{s}_m$ is a minimum contrast estimator with respect to the contrast

$$\gamma_n(t) = \|t\|^2 - 2\langle t, \widehat{s}_m \rangle, \quad \text{for } t \in \mathcal{S}_m. \tag{2.2}$$

Indeed setting the gradient $(\partial \gamma_n(t)/\partial a_k)_{0 \leq k \leq m-1}$ to zero for $t = \sum_{j=0}^{m-1} a_j \varphi_j \in \mathcal{S}_m$ also leads to $\widehat{s}_m = \arg\min_{t \in \mathcal{S}_m} \gamma_n(t)$.

Now if we intend to propose an estimator $\widetilde{s}_m$ such that $\widetilde{s}_m(0) = c$, we consider the Lagrange multiplier method and the contrast

$$\widetilde{\gamma}_n(t, \lambda) = \|t\|^2 - 2\langle t, \widehat{s}_m \rangle - \lambda(t(0) - c), \quad \text{for } t \in \mathcal{S}_m. \tag{2.3}$$

Considering for $t = \sum_{j=0}^{m-1} a_j \varphi_j$ in $\mathcal{S}_m$,

$$\begin{cases} \dfrac{\partial}{\partial a_k} \widetilde{\gamma}_n(t, \lambda) = 2a_k - 2\widehat{a}_k - \lambda \varphi_k(0) \\[2mm] \dfrac{\partial}{\partial \lambda} \widetilde{\gamma}_n(t, \lambda) = - \left(\displaystyle\sum_{k=0}^{m-1} a_k \varphi_k(0) - c \right) \end{cases}$$

we get, using that $\varphi_j(0) = \sqrt{2}$,

$$\widetilde{s}_m = \sum_{j=0}^{m-1} \widetilde{a}_{j,m} \varphi_j, \quad \widetilde{a}_{j,m} = \widehat{a}_j - K_m, \text{ with } K_m = \frac{1}{m} \left(\sum_{\ell=0}^{m-1} \widehat{a}_\ell - \frac{c}{\sqrt{2}} \right). \tag{2.4}$$

Note that when $m = 1$, $\widetilde{a}_{0,1} = c/\sqrt{2}$ and $\widehat{s}_1(x) = \widehat{a}_0 \varphi_0(x)$ and $\widetilde{s}_1(x) = ce^{-x}$.

Remark 2.1. *The general formula for a constraint $\widetilde{s}_m(\mathsf{b}) = c$, $\mathsf{b} > 0$, would yield*

$$\widetilde{\widetilde{a}}_{j,m} = \widehat{a}_j - \frac{\widehat{s}_m(\mathsf{b}) - c}{\sum_{\ell=0}^{m-1} \varphi_\ell^2(\mathsf{b})} \varphi_j(\mathsf{b}). \tag{2.5}$$

Taking advantage of $\varphi_j(0) = \sqrt{2}$ is useful in the following computations, but not necessary. However, we do not have any example where it would be useful to set a constraint at another point than 0.

Finally, we have the following estimator

$$\widetilde{s}_m := \widehat{s}_m - K_m \sum_{j=0}^{m-1} \varphi_j, \tag{2.6}$$

with K_m given by (2.4). Our aim is to compare the risk bound on $\widetilde{s}_m$ to the one on $\widehat{s}_m$.

2.3. **Risk bound on the constrained estimator.** As, under **(A1)-(A2)**(s), $s(0) = c = \sqrt{2} \sum_{\ell \geq 0} a_\ell(s)$, we get

$$\mathbb{E}[K_m] = \frac{1}{m} \left(\sum_{\ell=0}^{m-1} a_\ell - \frac{c}{\sqrt{2}} \right) = \frac{1}{m} \sum_{\ell \geq m} a_\ell(s).$$

Therefore, the new estimator is a biased estimator of s_m since $\mathbb{E}[\widetilde{s}_m] = s_m - (m^{-1} \sum_{\ell \geq m} a_\ell(s)) \sum_{j=0}^{m-1} \varphi_j$. To evaluate the quality of this new estimator we prove the following result.

Proposition 2.2. *Under **(A1)-(A2)**(s), the MISE of the estimator $\widetilde{s}_m$ of s, given by Equation (2.4) satisfies,*

$$\mathbb{E}\left[\|\widetilde{s}_m - s\|^2 \right] = \mathbb{E}[\|\widehat{s}_m - s\|^2] + B_m - V_{n,m}, \tag{2.7}$$

where

$$B_m := \frac{1}{m} \left(\sum_{\ell \geq m} a_\ell(s) \right)^2, \quad V_{n,m} := \frac{1}{m} \mathrm{Var} \left(\sum_{j=0}^{m-1} \widehat{a}_j \right). \tag{2.8}$$

Therefore, the following bound holds

$$\mathbb{E}\left[\|\widetilde{s}_m - s\|^2 \right] \leq \|s - s_m\|^2 + \mathbb{E}[\|\widehat{s}_m - s_m\|^2] + \frac{1}{m} \left(\sum_{\ell \geq m} a_\ell(s) \right)^2. \tag{2.9}$$

The proofs of Proposition 2.2, as well as most other proofs, is relegated to Section 8.

Equation (2.7) implies that the MISE of $\widetilde{s}_m$ has the same order as the risk of $\widehat{s}_m$ up to two terms:

- B_m which depends only on m and may increase the bias,
- $V_{n,m}$ which is of variance type and may decrease the variance.

We provide hereafter a general study showing that, under mild assumption, the additional bias term B_m has the same order as $\|s - s_m\|^2$, and thus does not change the order of the bias.

Then, in each example, we shall study (or recall) the order the main variance term $\mathbb{E}[\|\widehat{s}_m - s_m\|^2]$, and compare it with the new variance term $V_{n,m}$. In most cases, we can prove that its order is less than the main variance bound, and thus should not compensate it. Note that it always holds that

$$V_{n,m} \leq \sum_{j=0}^{m-1} \mathrm{Var}(\widehat{a}_j) = \mathbb{E}[\|\widehat{s_m} - s_m\|^2]. \tag{2.10}$$

Remark 2.3. *We complete Remark 2.1 by noting that equality (2.7) holds in the case of a constraint at b and coefficients $\widetilde{\widetilde{a}}_{j,m}$ with B_m and $V_{n,m}$ replaced by*

$$B_m(\mathsf{b}) = \frac{(s_m(\mathsf{b}) - s(\mathsf{b}))^2}{\sum_{\ell=0}^{m-1} \varphi_\ell^2(\mathsf{b})} \quad and \quad V_{n,m}(\mathsf{b}) = \frac{\mathrm{Var}(\widehat{s}_m(\mathsf{b}))}{\sum_{\ell=0}^{m-1} \varphi_\ell^2(\mathsf{b})}.$$

The bound (2.10) is also true for $V_{n,m}(\mathsf{b})$.

2.4. **Bias order on Sobolev spaces.** Regularity spaces considered for Laguerre functions are Sobolev-Laguerre spaces defined by

$$W^\alpha(\mathbb{R}^+, L) = \left\{ p : \mathbb{R}^+ \to \mathbb{R}, \, p \in \mathbb{L}^2(\mathbb{R}^+), \, \sum_{k \geq 0} k^\alpha a_k^2(p) \leq L < +\infty \right\} \quad \text{with} \quad \alpha \geq 0 \tag{2.11}$$

where $a_k(p) = \langle p, \varphi_k \rangle$. These spaces have been introduced by Bongioanni and Torrea (2009) and the link with the coefficients of a function on a Laguerre basis was done by Comte and Genon-Catalot (2015). Now for $s \in W^\alpha(\mathbb{R}^+, L)$ defined by (2.11), we have

$$\|s - s_m\|^2 = \sum_{k=m}^\infty a_k^2(s) = \sum_{k=m}^\infty a_k^2(s) k^\alpha k^{-\alpha} \leq Lm^{-\alpha}.$$

Now, if $\alpha > 1$, by Cauchy-Schwarz Inequality, under **(A2)**(s) it comes

$$B_m = \frac{1}{m} \left(\sum_{\ell \geq m} a_\ell(s) \right)^2 \leq \frac{1}{m} \sum_{\ell \geq m} \ell^{-\alpha} \sum_{\ell \geq m} \ell^\alpha a_\ell^2(s) \leq \frac{L}{\alpha - 1} m^{-\alpha}$$

using that $\sum_{\ell \geq m} \ell^{-\alpha} \leq m^{1-\alpha}/(\alpha - 1)$. Therefore, the additional bias term B_m has the same order in m as the standard bias term. The following Corollary summarizes this finding.

Corollary 2.4. *Assume that **(A1)-(A2)**(s) hold, and moreover that $s \in W^\alpha(\mathbb{R}^+, L)$ for $\alpha > 1$, then the MISE of the estimator $\widetilde{s}_m$ of s, given by Equation (2.4) satisfies,*

$$\mathbb{E}\left[\|\widetilde{s}_m - s\|^2\right] \leq \frac{2\alpha}{\alpha - 1} Lm^{-\alpha} + \mathbb{E}[\|\widehat{s}_m - s_m\|^2]. \tag{2.12}$$

The consequence is that the order of the upper risk bounds of $\widetilde{s}_m$ and $\widehat{s}_m$ are the same, and also their rates of convergence.

2.5. **Model selection.** Now we discuss the selection of m in order to get an automatic squared bias-variance tradeoff. In our projection approach, the spaces $\mathcal{S}_m$ are nested and, to select an adequate dimension, we look for m in a finite set

$$\mathcal{M}_n = \{1, \ldots, m_{\max}\},$$

where, in any case, $m_{\max} \leq n$, and $m_{\max}$ is specifically defined in each context and can depend on n. The contrast function defined by (2.2) can thus be written, for any $t \in \mathcal{S}_m$,

$$\gamma_n(t) = \|t\|^2 - 2\langle t, \widehat{s}_{m_{\max}}\rangle. \tag{2.13}$$

This definition has the advantage that it no longer depends on m.

Assume that a selection procedure was settled up for the original estimator $\widehat{s}_m$. The bias term $\|s - s_m\|^2$ is estimated by $-\|\widehat{s}_m\|^2$: indeed, it is equal to $\|s\|^2 - \|s_m\|^2$, and $\|s\|^2$ is an unknown constant that can be dropped out from the minimization procedure. Thus, the selected dimension $\widehat{m}$ is defined by :

$$\widehat{m} = \operatorname*{argmin}_{m \in \mathcal{M}_n} \left\{ -\|\widehat{s}_m\|^2 + \widehat{\operatorname{pen}}_1(m) \right\},$$

where $\widehat{\operatorname{pen}}_1$ is a data driven increasing function of m, estimating the deterministic variance or an upper bound on it, denoted by $\operatorname{pen}_1(m)$. We require that this penalty satisfies three conditions. First, $\operatorname{pen}_1(m)$ is the smallest quantity such that

$$\mathbb{E}\left(\sup_{t \in B_{m,\widehat{m}}} \nu_n^2(t) - \frac{1}{4}\operatorname{pen}_1(m \vee \widehat{m})\right)_+ \leq \frac{C}{n}, \tag{2.14}$$

where C is a constant, $B_{m,\widehat{m}} := \{t \in \mathcal{S}_{m \vee \widehat{m}}, \|t\| = 1\}$, and ν_n is the centred empirical process, defined by

$$\nu_n(t) := \langle t, \widehat{s}_{m_{\max}} - s_{m_{\max}}\rangle. \tag{2.15}$$

Then, the two following conditions are required to link $\operatorname{pen}_1(m)$ and its estimator $\widehat{\operatorname{pen}}_1(m)$. The second condition is

$$\mathbb{E}[\widehat{\operatorname{pen}}_1(m)] \leq 2\operatorname{pen}_1(m). \tag{2.16}$$

Lastly, we assume that, for $a \in \{0, 1, 2\}$ and C' a constant,

$$\mathbb{E}\left[(\operatorname{pen}_1(\widehat{m}) - \widehat{\operatorname{pen}}_1(\widehat{m}))_+\right] \leq C'\frac{\log^a(n)}{n}. \tag{2.17}$$

Note that, in some examples, $\widehat{\operatorname{pen}}_1(m) = \operatorname{pen}_1(m)$ and then conditions (2.16)-(2.17) are straightforwardly satisfied. The underlying idea is that $\operatorname{pen}_1(m)$ has approximately the order of the variance bound, and $\widehat{\operatorname{pen}}_1(\cdot)$ can be computed from the observations.

It can be proved from (2.14)-(2.16)-(2.17) (see Massart, 2007, or the proof of Theorem 2.5) that

$$\mathbb{E}[\|\widehat{s}_{\widehat{m}} - s\|^2] \leq 3 \inf_{m \in \mathcal{M}_n} \left\{\|s - s_m\|^2 + 2\operatorname{pen}_1(m)\right\} + \frac{C''\log^a(n)}{n}, \tag{2.18}$$

where C'' is a constant depending on s but not on m or n.

For the new estimator, we introduce a new term of penalization pen_2: it can also be computed from the observations, given the constraint, and heuristically contains both B_m and $V_{n,m}$:

$$\widehat{\operatorname{pen}}(m) = \widehat{\operatorname{pen}}_1(m) + \widehat{\operatorname{pen}}_2(m) \quad \text{with} \quad \widehat{\operatorname{pen}}_2(m) = \frac{m}{2}K_m^2 = \frac{1}{2m}\left(\sum_{\ell=0}^{m-1} \widehat{a}_\ell - \frac{c}{\sqrt{2}}\right)^2. \tag{2.19}$$

Indeed, this second penalty term satisfies

$$\mathbb{E}[2\widehat{\operatorname{pen}}_2(m)] = \frac{1}{m}\mathbb{E}\left[\left(\sum_{\ell=0}^{m-1}(\widehat{a}_\ell - a_\ell) - \sum_{\ell \geq m} a_\ell\right)^2\right] = B_m + V_{n,m}. \tag{2.20}$$

Then we set

$$\widetilde{m} = \operatorname*{argmin}_{m \in \mathcal{M}_n} \{-\|\widehat{s}_m\|^2 + \widetilde{\operatorname{pen}}(m)\}. \tag{2.21}$$

Therefore, the estimate of the bias of $\widehat{s}_m$ given by $-\|\widehat{s}_m\|^2$ is increased by the B_m term contained in $\operatorname{pen}_2(m)$, and thus we get an estimate of the bias of $\widetilde{s}_m$. The $V_{n,m}$ contribution increases the variance part, but in a negligible way in most specific cases studied hereafter. A general bound following from (2.10) and (2.20) is

$$\mathbb{E}[2\widehat{\operatorname{pen}}_2(m)] \le \mathbb{E}[\|\widehat{s}_m - s_m\|^2] + \frac{1}{m}\left(\sum_{\ell \ge m} a_\ell(s)\right)^2. \tag{2.22}$$

Theorem 2.5. *Under **(A1)-(A2)**(s), if the empirical process ν_n satisfies (2.14)-(2.16)-(2.17), then the final estimator $\widetilde{s}_{\widetilde{m}}$ defined by (2.4) and (2.21) satisfies the oracle-type inequality:*

$$\mathbb{E}\left[\|\widetilde{s}_{\widetilde{m}} - s\|^2\right] \le 2 \inf_{m \in \mathcal{M}_n}\left\{3\mathbb{E}[\|\widehat{s}_m - s\|^2] + \frac{1}{m}\left(\sum_{\ell \ge m} a_\ell(s)\right)^2 + 6\operatorname{pen}_1(m)\right\} + \frac{C''' \log^a(n)}{n},$$

where C''' is a constant which does not depend on n, $a \in \{0,1,2\}$.

Clearly, assumptions (2.14)-(2.16)-(2.17) imply both (2.18) and Theorem 2.5. Inequality (2.18) means that $\widehat{s}_{\widehat{m}}$ is adaptive realizing the compromise between the bias term $\|s - s_m\|^2$ and the variance $\operatorname{pen}_1(m)$. The inequality in Theorem 2.5 states the same result for $\widetilde{s}_{\widetilde{m}}$, with the additional bias B_m. Both results are up to a negligible residual term $\log^a(n)/n$.

In each section hereafter, the above procedure is applied for different models. After a quick presentation of the context, the definition of the projection estimator is given, together with its constrained version, with common notation $(\widehat{a}_j, \widetilde{a}_j)$ for the coefficients on the Laguerre basis. Then a specific risk bound is provided associated to each example.

3. Direct observation model

We assume here that we have n direct observations of the variable of interest X: $X_1, \ldots, X_n$, which are nonnegative and *i.i.d.* with common density denoted by f and survival function denoted by S, $S(x) = \mathbb{P}(X > x)$.

3.1. Density estimation.
For density estimation from direct observations, we may wish to fix $f(0) = 0$. This occurs for instance if f is the convolution of two $\mathbb{R}^+$-supported densities g and h, that is $f(x) = \int_0^x g(x-u)h(u)du$; then under mild regularity assumptions on g and h, $f(0) = 0$.

Assume that f satisfies **(A1)**(f). Then the projection estimator of f (see *e.g* Efromovich, 1999; Massart, 2007) is

$$\widehat{f}_m = \sum_{j=0}^{m-1} \widehat{a}_j \varphi_j \text{ with } \widehat{a}_j = \frac{1}{n}\sum_{i=1}^{n}\varphi_j(X_i) \tag{3.1}$$

and

$$\widehat{f}_m = \arg\min_{t \in \mathcal{S}_m} \gamma_n(t), \quad \gamma_n(t) = \|t\|^2 - \frac{2}{n}\sum_{i=1}^{n} t(X_i).$$

Then it is easy to see that

$$\mathbb{E}[\|\widehat{f}_m - f\|^2] \le \|f - f_m\|^2 + \frac{2m}{n}$$

with $f_m = \sum_{j=0}^{m-1} a_j(f)\varphi_j$ the orthogonal projection of f on $\mathcal{S}_m$. The new estimator of f defined by

$$\widetilde{f}_m = \sum_{j=0}^{m-1} \widetilde{a}_{j,m}\varphi_j \quad \text{with } \widetilde{a}_{j,m} = \widehat{a}_j - K_m, \quad K_m = \frac{1}{m}\sum_{\ell=0}^{m-1} \widehat{a}_\ell \tag{3.2}$$

satisfies Proposition 2.2. More precisely, under assumption $(\mathbf{A2})(f)$, we can use the bound (2.9) and obtain the following result.

Proposition 3.1. *Consider the estimator $\widetilde{f}_m$ defined by (3.1)-(3.2), based on the i.i.d. sample $X_1, \ldots, X_n$ of random variables with common density f satisfying $(\mathbf{A1})$-$(\mathbf{A2})(f)$. Then, for any $m \geq 1$, we have*

$$\mathbb{E}[\|\widetilde{f}_m - f\|^2] \leq \|f - f_m\|^2 + \frac{2m}{n} + \frac{1}{m}\left(\sum_{\ell \geq m} a_\ell(f)\right)^2 \quad \text{and} \quad V_{n,m} \leq \frac{2\|f\|_\infty}{n}, \qquad (3.3)$$

where $V_{n,m}$ is defined by (2.8).

The bound on $V_{n,m}$ shows that this term has negligible order $O(1/n)$, and in particular negligible order w.r.t. the variance order (which is $O(m/n)$). It also follows from Corollary 2.4 that, if $f \in W^\alpha(\mathbb{R}^+, L)$ for $\alpha > 1$, then for $m_{\mathrm{opt}} = n^{1/(\alpha+1)}$,

$$\mathbb{E}[\|\widetilde{f}_{m_{\mathrm{opt}}} - f\|^2] \leq C(L, \alpha) n^{-\alpha/(1+\alpha)}.$$

This rate is the optimal rate on the Sobolev Laguerre space $W^\alpha(\mathbb{R}^+, L)$, as proved in Belomestny et al. (2016).

For the model selection step, as

$$\nu_n(t) = \frac{1}{n}\sum_{i=1}^n (t(X_i) - \langle t, f \rangle)$$

satisfies (2.14) with $\mathrm{pen}_1(m) = \kappa_0 m/n$, for κ_0 a constant (see Chapter 7 in Massart (2007)), and (2.16)-(2.17) are straightforwardly fulfilled for $\widehat{\mathrm{pen}}_1(m) = \mathrm{pen}_1(m)$, we can deduce that $\widetilde{m}$ defined by (2.19)-(2.21) provides an estimate $\widetilde{f}_{\widetilde{m}}$ satisfying the inequality of Theorem 2.5.

3.2. **Survival function estimation.** If $\mathbb{E}[X_1] < +\infty$ then the survival function S is squared integrable on $\mathbb{R}^+$: indeed $\int_0^{+\infty} S^2(x)dx \leq \int_0^{+\infty} S(x)dx = \mathbb{E}[X_1]$. Therefore, we can adopt a projection strategy on $\mathbb{L}^2(\mathbb{R}^+)$ for its estimation. An estimator of S is given by

$$\widehat{S}_m = \sum_{j=0}^{m-1} \widehat{a}_j \varphi_j, \quad \widehat{a}_j = \frac{1}{n}\sum_{i=1}^n \int_{\mathbb{R}^+} \varphi_j(x)\mathbb{1}_{X_i \geq x}dx, \quad \mathbb{E}[\widehat{a}_j] = a_j(S) = \langle \varphi_j, S \rangle. \qquad (3.4)$$

This estimator satisfies the following bound.

Proposition 3.2. *Assume that $\mathbb{E}[X_1] < +\infty$. Then the estimator $\widehat{S}_m$ (3.4) is an unbiased estimator of $S_m = \sum_{j=0}^{m-1} a_j(S)\varphi_j$ and it satisfies*

$$\mathbb{E}\left[\left\|\widehat{S}_m - S\right\|^2\right] \leq \|S_m - S\|^2 + \frac{\mathbb{E}[X_1]}{n}. \qquad (3.5)$$

Note that the estimator $\widehat{S}_m$ is also the minimizer of the following contrast:

$$\gamma_n(t) = \|t\|^2 - \frac{2}{n}\sum_{i=1}^n \int_{\mathbb{R}^+} t(x)\mathbb{1}_{X_i \geq x}dx, \quad \widehat{S}_m = \operatorname*{argmin}_{t \in \mathcal{S}_m} \gamma_n(t)$$

The new estimator of S noted $\widetilde{S}_m$ is defined in (2.4) with $c = 1$ and $s = S$. It satisfies $\widetilde{S}_m(0) = 1$ along with Proposition 2.2.

As the variance does not depend on m, we recover that the estimator can reach the parametric rate, by taking m as large as possible: for instance if $S \in W^\alpha(\mathbb{R}^+, L)$ for $\alpha > 1$, then choosing $m = n$ implies that $\widetilde{S}_n$ converges with parametric rate to S.

This is only a toy example since clearly, the simple empirical survival function $\widehat{S}_n(x) = n^{-1}\sum_{i=1}^n \mathbb{1}_{X_i \geq x}$ satisfies $\widehat{S}_n(0) = 1$ and $\mathbb{E}[\|\widehat{S}_n - S\|^2] \leq \mathbb{E}[X]/n$.

4. Multiplicative noise model

We consider in this section the multiplicative model. The common density and survival function of *i.i.d.* X_i's are still denoted by f and S but we observe

$$Y_i = X_i U_i, \ i = 1, \ldots, n,$$

where the U_i's follow a continuous uniform distribution on $[0,1]$: $U_i \sim \mathcal{U}([0,1])$.

4.1. **Useful Laguerre formulae.**

We state two useful properties of the Laguerre basis, which are specifically used for the multiplicative model, proved in Comte and Dion (2016). First, we have

$$\varphi_0'(x) = -\varphi_0(x), \quad \varphi_j'(x) = -\varphi_j(x) - 2\sum_{k=0}^{j-1} \varphi_k(x), \quad j \geq 1. \tag{4.1}$$

Moreover, the following relation (see Abramowitz and Stegun (1966))

$$\forall y \in \mathbb{R}^+, \ (y\varphi_j(y))' = -\frac{j}{2}\varphi_{j-1}(y) + \frac{1}{2}\varphi_j(y) + \frac{j+1}{2}\varphi_{j+1}(y), \quad j \geq 0, \tag{4.2}$$

with convention $\varphi_{-1} = 0$, implies in particular that $(y\varphi_j(y))'$ is bounded, with bound of order $O(j)$ (namely $\sqrt{2}(j+1)$).

Let us now give a useful property, implied by the model, see Brunel *et al.* (2016). If $t : \mathbb{R} \to \mathbb{R}$ is bounded, derivable, then

$$\mathbb{E}[t(Y_1) + Y_1 t'(Y_1)] = \mathbb{E}[t(X_1)]. \tag{4.3}$$

This relation follows from an integration by part and the fact that $\lim_{y \to 0} y f_Y(y) = \lim_{y \to +\infty} y f_Y(y) = 0$. Lastly, for any $t \in \mathbb{L}^2(\mathbb{R}^+)$, the following bound holds (see Comte and Genon-Catalot (2015)):

$$\mathbb{E}[(Y_1 t(Y_1))^2] \leq \|t\|^2 \mathbb{E}(X_1). \tag{4.4}$$

4.2. **Density estimation.**

The common density f_Y of the *i.i.d.* observations $(Y_i)_{1 \leq i \leq n}$ is given by

$$f_Y(y) = \int_y^{+\infty} \frac{f(x)}{x} dx, \quad y \in]0, +\infty[, \tag{4.5}$$

and the condition $f(0) = 0$ is required if we expect that f_Y is finite in 0 (even if it is not sufficient).

As f_Y is a non-increasing function, if it is bounded, then its supremum is $\int_0^{+\infty}(f(x)/x)dx$, and thus, the condition $\|f_Y\|_\infty < +\infty$ also requires $f(0) = 0$.

Equality (4.3) explains the projection estimator of the density f defined by Belomestny *et al.* (2016) in the Laguerre orthonormal basis of $\mathbb{L}^2(\mathbb{R}^+)$:

$$\widehat{f}_m = \sum_{j=0}^{m-1} \widehat{a}_j \varphi_j, \quad \widehat{a}_j = \frac{1}{n}\sum_{i=1}^n (Y_i \varphi_j'(Y_i) + \varphi_j(Y_i)). \tag{4.6}$$

Indeed by applying (4.3) to the function $t = \varphi_j$, we get $\mathbb{E}[\widehat{a}_j] = a_j(f) = \langle \varphi_j, f \rangle$. Thus $\widehat{f}_m$ is an unbiased estimator of the projection f_m on subspace $\mathcal{S}_m$. Moreover, we can prove the following result.

Proposition 4.1. *Assume that f satisfies **(A1)**(f). Then, the estimator $\widehat{f}_m$ given by (4.6) satisfies:*

$$\mathbb{E}[\|\widehat{f}_m - f\|^2] \leq \|f - f_m\|^2 + \frac{2m^3}{n} + \frac{3m}{2n}. \tag{4.7}$$

If in addition $\mathbb{E}[X_1] < \infty$, then we have

$$\mathbb{E}[\|\widehat{f}_m - f\|^2] \leq \|f - f_m\|^2 + 4\mathbb{E}[Y_1]\frac{m^2}{n} + 2\frac{m}{n}. \tag{4.8}$$

Note here that the constants appearing in the first bound (4.7) are numerical constants. This is due to the use of equality (4.2). Note also that $\mathbb{E}[X_1] = 2\mathbb{E}[Y_1]$. The bound (4.8) improves (4.7), which is stated in Belomestny *et al.* (2016), under the mild additional condition $\mathbb{E}[X_1]$. We keep this second bound in the sequel.

This estimator is also the minimizer of the following contrast:

$$\gamma_n(t) = \|t\|^2 - \frac{2}{n} \sum_{i=1}^{n} (Y_i t'(Y_i) + t(Y_i)), \quad \widehat{f}_m = \underset{t \in \mathcal{S}_m}{\operatorname{argmin}} \, \gamma_n(t)$$

Then, under **(A2)**(f), the corrected estimator defined by $\widetilde{f}_m = \sum_{j=0}^{m-1} \widetilde{a}_{j,m} \varphi_j$ with $\widetilde{a}_{j,m} = \widehat{a}_j - K_m$ and $K_m = (1/m) \sum_{\ell=0}^{m-1} \widehat{a}_\ell$ satisfies Proposition 2.2. Therefore we obtain, by using Equation (2.7), the following result.

Proposition 4.2. *Assume that f satisfies **(A1)-(A2)**(f), that $\mathbb{E}[X_1] < \infty$, and that f_Y is bounded. Then the estimator $\widetilde{f}_m$ defined by (3.2) and (4.6) satisfies*

$$\mathbb{E}[\|\widetilde{f}_m - f\|^2] \leq \|f - f_m\|^2 + 4\mathbb{E}[Y_1]\frac{m^2}{n} + 2\frac{m}{n} + \frac{1}{m}\left(\sum_{\ell \geq m} a_l(f)\right)^2 \quad and \quad V_{n,m} \leq \frac{m\|f_Y\|_\infty}{2n}.$$

Again, we obtain that the order of $V_{n,m}$ is negligible with respect to the main variance term, $4\mathbb{E}[Y_1]m^2/n$. From Corollary 2.4, the upper rate bound on Sobolev Laguerre space, i.e. for $f \in W^\alpha(\mathbb{R}^+, L)$ with $\alpha > 1$ and for $m_{\mathrm{opt}} = n^{1/(\alpha+2)}$, is of order $n^{-\alpha/(\alpha+2)}$ for both $\widehat{f}_{m_{\mathrm{opt}}}$ and $\widetilde{f}_{m_{\mathrm{opt}}}$.

We can propose a model selection method to select m automatically

$$\widehat{m} = \arg\min_{m \in \mathcal{M}_n} \left(-\|\widehat{f}_m\|^2 + \widehat{\mathrm{pen}}_1(m)\right) \quad \text{with} \quad \widehat{\mathrm{pen}}_1(m) = \kappa_1 \frac{m\log(2+m)}{n}(1 + 4\bar{Y}_n m),$$

where $\bar{Y}_n = n^{-1}\sum_{i=1}^n Y_i$, and

$$\mathcal{M}_n = \{m \in \{1, \ldots, n\}, \, m \leq \sqrt{n}\}.$$

The bound of order $\sqrt{n}$ for the dimensions considered in $\mathcal{M}_n$ simply ensures that the variance term remains bounded. We also denote by

$$\mathrm{pen}_1(m) = \kappa_1 \frac{m\log(2+m)}{n}(1 + 2\mathbb{E}[Y_1]m)$$

so $\mathbb{E}[\widehat{\mathrm{pen}}_1(m)] \leq 2\mathrm{pen}_1(m)$. We can prove the following result for the selected estimator.

Theorem 4.3. *Assume that **(A1)-(A2)**(f) hold and that $\mathbb{E}[X_1^2] < +\infty$. Then there exists a constant κ_1' such that for any $\kappa_1 \geq \kappa_1'$, we have*

$$\mathbb{E}[\|\widehat{f}_{\widehat{m}} - f\|^2] \leq 6 \inf_{m \in \mathcal{M}_n} \left\{\|f - f_m\|^2 + \mathrm{pen}_1(m)\right\} + C_1 \frac{\log(n)}{n},$$

where C_1 is a positive constant depending on $\mathbb{E}[Y_1]$ and $\mathbb{E}[Y_1^2]$.

The proof of Theorem 4.3 contains the proof that the empirical process

$$\nu_n(t) = \frac{1}{n}\sum_{i=1}^n (Y_i t'(Y_i) + t(Y_i) - \langle t, f\rangle)$$

associated with this problem satisfies Assumption (2.14) and that $\mathrm{pen}_1(m)$ and $\widehat{\mathrm{pen}}_1(m)$ satisfy Assumptions (2.16)-(2.17) with $a = 1$. Therefore, the procedure of Section 2.5 can be applied to select $\widetilde{m}$ for $\widetilde{f}$.

4.3. **Survival function estimation.** We estimate now the survival function S associated with f. The key formula here, obtained from (4.5), is

$$S_Y(y) = S(y) - y f_Y(y),$$

where S_Y is the survival function of Y. Thus, the coefficients of this function on the Laguerre basis are such that

$$a_j(S) = \langle S, \varphi_j \rangle = \mathbb{E}[Y \varphi_j(Y)] + \langle S_Y, \varphi_j \rangle.$$

As a consequence, we estimate the projection $S_m = \sum_{j=0}^{m-1} a_j(S)\varphi_j$ of S on $\mathcal{S}_m$ by

$$\widehat{S}_m = \sum_{j=0}^{m-1} \widehat{a}_j \varphi_j, \quad \widehat{a}_j = \frac{1}{n} \sum_{i=1}^{n} \left[\int_{\mathbb{R}^+} \varphi_j(x) \mathbb{1}_{Y_i \geq x} dx + Y_i \varphi_j(Y_i) \right]. \tag{4.9}$$

Then, similarly to Brunel *et al.* (2016) who study the kernel case, we can prove the following result.

Proposition 4.4. *If $\mathbb{E}[X_1] < +\infty$, the estimator $\widehat{S}_m$ (4.9) is an unbiased estimator of S_m and it satisfies*

$$\mathbb{E}\left[\left\| \widehat{S}_m - S \right\|^2 \right] \leq \| S_m - S \|^2 + \mathbb{E}\left[X_1 \right] \frac{m+1}{n}.$$

Then, using the same steps as before, the new estimator $\widetilde{S}_m$ of S defined in (2.4) with $c = 1$, is such that $\widetilde{S}_m(0) = 1$, and satisfies Proposition 2.2. Thus we get:

Proposition 4.5. *If $\mathbb{E}[X_1] < +\infty$, the estimator $\widetilde{S}_m$ of S satisfies*

$$\mathbb{E}\left[\left\| \widetilde{S}_m - S \right\|^2 \right] \leq \| S_m - S \|^2 + \mathbb{E}\left[X_1 \right] \frac{m+1}{n} + \frac{1}{m} \left(\sum_{\ell \geq m} a_\ell(S) \right)^2 \quad \text{and} \quad V_{n,m} \leq \frac{4\mathbb{E}[Y_1]}{n} \tag{4.10}$$

Remark 4.6. *The strategy for this multiplicative noise model can be extended to the model proposed in Comte and Dion (2016) defined by $Y_i = X_i U_i^{(\mathsf{a})}$ where $U_i^{(\mathsf{a})} \sim \mathcal{U}([1-\mathsf{a}, 1+\mathsf{a}])$ has a uniform density on a symmetric interval $[1-\mathsf{a}, 1+\mathsf{a}]$, $0 < \mathsf{a} < 1$. In this case, the estimation strategy of the survival function is set up in two steps. First the function*

$$\overline{G}(x) = \frac{1}{2a} \left[(1+\mathsf{a}) S \left(\frac{x}{1+\mathsf{a}} \right) - (1-\mathsf{a}) S \left(\frac{x}{1-\mathsf{a}} \right) \right] \tag{4.11}$$

is estimated by a projection estimator. As S is a survival function (not $\overline{G}$), $S(0) = 1$ and thus $\overline{G}(0) = 1$. Consequently, we can apply our constrained procedure to the unbiased projection estimator of $\overline{G}_m$ proposed in the paper. Then, the survival function is estimated by plugging in this estimate in the formula corresponding to the inversion of (4.11):

$$S_N(x) := \frac{2\mathsf{a}}{1+\mathsf{a}} \sum_{k=0}^{N-1} \left(\frac{1-\mathsf{a}}{1+\mathsf{a}} \right)^k \overline{G} \left(\left(\frac{1+\mathsf{a}}{1-\mathsf{a}} \right)^k (1+\mathsf{a}) x \right)$$

for $N \geq \log(n)/[3 \log((1+\mathsf{a})/(1-\mathsf{a}))]$. We would get $\widehat{S}_{N,m}$ and $\widetilde{S}_{N,m}$ in this context as well. For simplicity we omit the details.

For the selection procedure we can prove that

$$\widehat{m} = \arg \min_{m \in \mathcal{M}_n} \{ -\| \widehat{S}_m \|^2 + \widehat{\mathrm{pen}}_1(m) \}, \quad \widehat{\mathrm{pen}}_1(m) = 2\kappa_2 \bar{Y}_n m/n, \quad \mathcal{M}_n = \{1, \ldots, n\},$$

defines an estimate $\widehat{S}_{\widehat{m}}$ which makes a data driven bias variance trade-off.

Theorem 4.7. *Assume that (A1)-(A2)(S) hold and that* $\mathbb{E}[X_1^4] < +\infty$ *Then there exists a constant* κ_2' *such that for any* $\kappa_2 \geq \kappa_2'$, *we have*

$$\mathbb{E}[\|\widehat{S}_{\widehat{m}} - S\|^2] \leq 3 \inf_{m \in \mathcal{M}_n} \left\{ \|S - S_m\|^2 + 2\mathrm{pen}_1(m) \right\} + C \frac{\log^2(n)}{n},$$

where $\mathrm{pen}_1(m) = \kappa_2 \mathbb{E}[Y_1]/n$ *and* C *is a constant depending on* $\mathbb{E}[X_1^4]$ *and* $\mathbb{E}[X_1]$.

The proof of Theorem 4.7 relies on the study of the empirical process

$$\nu_n(t) = \frac{1}{n} \sum_{i=1}^{n} \int t(x) \mathbb{1}_{Y_i \geq x} dx + Y_i t(Y_i) - \mathbb{E}\left[\int t(x) \mathbb{1}_{Y_i \geq x} dx + Y_i t(Y_i) \right]$$

which satisfies Assumption (2.14) with $\mathrm{pen}_1(m) = 2\kappa_2 \mathbb{E}[Y_1]m/n$. We also have that pen_1 and $\widehat{\mathrm{pen}}_1$ satisfy Assumptions (2.16)-(2.17) with $a = 2$. As a consequence, the procedure of Section 2.5 can be applied to select $\widetilde{m}$ and build the final estimator $\widetilde{S}_{\widetilde{m}}$.

5. CONVOLUTION MODEL

Now we show that the constrained strategy also applies to the convolution model

$$Y_i = X_i + V_i, \quad i = 1, \ldots, n \tag{5.1}$$

where the $(X_i)_{1 \leq i \leq n}$ and the $(V_i)_{1 \leq i \leq n}$ are two independent sequences of *i.i.d.* nonnegative random variables. The X_i's have unknown density denoted by f and unknown survival function denoted by S, while the V_i's have known density g.

In the additive framework, the key property of the Laguerre basis (see Abramowitz and Stegun (1966)) is the following:

$$\varphi_k \star \varphi_j(x) = \int_0^x \varphi_k(u)\varphi_j(x-u)du = 2^{-1/2}\left(\varphi_{k+j}(x) - \varphi_{k+j+1}(x)\right), \tag{5.2}$$

showing that the convolution of two basis functions has a linear simple expression in function of two other basis functions.

5.1. **Density estimation.** Indeed, it follows from model (5.1) and the independence assumptions that the density f_Y of the observations Y_i is equal to

$$f_Y(x) = f \star g(x) = \int_0^x f(u)g(x-u)du.$$

Comte *et al.* (2017) noticed that, by using (5.2), this convolution equation can be rewritten:

$$\sum_{k=0}^{\infty} a_k(f_Y)\varphi_k(x) = \sum_{j=0}^{+\infty} \sum_{k=0}^{+\infty} a_j(f)a_k(g)\varphi_j \star \varphi_k(x)$$

$$= \sum_{k=0}^{\infty} \varphi_k(x) \sum_{\ell=0}^{k} 2^{-1/2}\left(a_{k-\ell}(g) - a_{k-\ell-1}(g)\right)a_\ell(f).$$

Therefore, for any integer m,

$$\vec{f}_{Ym} = \mathbf{G}_m \vec{f}_m,$$

with $\vec{f}_{Ym} = {}^t(a_0(f_Y), \ldots, a_{m-1}(f_Y))$, $\vec{f}_m = {}^t(a_0(f), \ldots, a_{m-1}(f))$, and $\mathbf{G}_m = ([\mathbf{G}_m]_{i,j})_{1 \leq i,j \leq m}$,

$$[\mathbf{G}_m]_{i,j} = \begin{cases} 2^{-1/2}a_0(g) & \text{if } i = j, \\ 2^{-1/2}\left(a_{i-j}(g) - a_{i-j-1}(g)\right) & \text{if } j < i, \\ 0 & \text{otherwise.} \end{cases} \tag{5.3}$$

The matrix $\mathbf{G}_m$ is known as g is known. An important feature of $\mathbf{G}_m$ is to be lower triangular and Toeplitz. As the diagonal elements $a_0(g) = \sqrt{2}\mathbb{E}[e^{-Y}] > 0$, the matrix $\mathbf{G}_m$ has nonzero determinant

and can be inverted. Starting from equality $\mathbf{G}_m^{-1}\vec{f}_{Ym} = \vec{f}_m$, Mabon (2015) proposes an estimator of f defined by

$$\widehat{f}_m(x) = \sum_{k=0}^{m-1} \widehat{a}_k \varphi_k(x) \quad \text{with} \quad \widehat{\vec{f}}_m = \mathbf{G}_m^{-1} \widehat{\vec{f}_{Ym}}, \tag{5.4}$$

where $\widehat{\vec{f}}_m = {}^t(\widehat{a}_0, \ldots, \widehat{a}_{m-1})$ and $\widehat{\vec{f}_{Ym}} = {}^t(\widehat{a}_0(Y), \ldots, \widehat{a}_{m-1}(Y))$, with $\widehat{a}_j(Y) = \dfrac{1}{n} \sum_{i=1}^n \varphi_j(Y_i)$. Here also $\widehat{f}_m$ is an unbiased estimator of f_m. The following risk bound on $\widehat{f}_m$ is proved in Mabon (2015):

$$\mathbb{E}[\|\widehat{f}_m - f\|^2] \leq \|f - f_m\|^2 + (2 \vee \|f_Y\|_\infty) \frac{\|\mathbf{G}_m^{-1}\|_F^2}{n},$$

where $\|\mathbf{A}\|_F^2 = \sum_{i,j} a_{i,j}^2$ the Frobenius or trace norm of a matrix $\mathbf{A}$.

If we know that f satisfies $f(0) = 0$, we can define $\widetilde{f}_m$ according to (2.4), this new estimator satisfies Proposition 2.2. Note that here we can only prove

$$V_{n,m} \leq \|f_Y\|_\infty \frac{\|\mathbf{G}_m^{-1}\|_{\text{op}}^2}{n}. \tag{5.5}$$

This bound is smaller than the variance term $\|\mathbf{G}_m^{-1}\|_F^2/n$, and numerically, this may improve the constants. However, it does not improve the order w.r.t. m. For instance, in the case where g corresponds to a $\gamma(p, \theta)$ density, both $\|\mathbf{G}_m^{-1}\|_F^2$ and $\|\mathbf{G}_m^{-1}\|_{\text{op}}^2$ have the same order $O(m^{2p})$ (see Comte et al. (2017)).

Model selection in the additive convolution model is studied in Mabon (2015), and relies on the study of

$$\nu_n(t) = \frac{1}{n} \sum_{i=1}^n \left\langle t, \sum_{j=0}^{m_{\max}-1} \left[\mathbf{G}_{m_{\max}}^{-1} (\vec{\varphi}_{m_{\max}}(Y_i) - \mathbb{E}(\vec{\varphi}_{m_{\max}}(Y_i))) \right]_j \varphi_j \right\rangle$$

where $\vec{\varphi}_m(x) = {}^t(\varphi_0(x), \ldots, \varphi_{m-1}(x))$. Note that, if $t \in \mathcal{S}_m$, $m_{\max}$ can be replaced by m thanks to the triangular feature of $\mathbf{G}_m$. Mabon (2015) proves that condition (2.14) is fulfilled for ν_n as above and the penalty defined by

$$\text{pen}_1(m) = \frac{\kappa_3(\|g\|_\infty \vee 1)}{n} (m\|\mathbf{G}_m^{-1}\|_{\text{op}}^2 \wedge \log(n)\|\mathbf{G}_m^{-1}\|_F^2).$$

Indeed the study of $\widehat{f}_{\widehat{m}}$ with $\widehat{m} = \arg\min_{m \in \mathcal{M}_n}(-\|\widehat{f}_m\|^2 + \text{pen}_1(m))$, and $\mathcal{M}_n = \{m \in \mathbb{N}, m\|\mathbf{G}_m^{-1}\|_{\text{op}}^2 \leq n\}$ implies that the not random penalty satisfies the conditions (2.16)-(2.17) with $a = 0$. Therefore, the procedure of Section 2.5 can be applied to select $\widetilde{m}$ and build the final estimator $\widetilde{S}_{\widetilde{m}}$.

5.2. **Survival function estimation.** For the survival function estimation, Mabon (2015) noticed that $S_Y(y) = S \star g(y) + S_V(y)$, where $S_Y(y) = \mathbb{P}(Y > y)$, $S_V(y) = \mathbb{P}(V > y)$. Then

$$a_j(S_Y) = \langle S_Y, \varphi_j \rangle = \mathbb{E}[\Phi_j(Y_1)] \text{ for } \Phi_j(y) := \int_0^y \varphi_j(x) dx,$$

and the proposed estimator is

$$\widehat{S}_m = \sum_{j=0}^{m-1} \widehat{a}_j(S) \varphi_j, \text{ with } \widehat{\vec{S}}_m = \mathbf{G}_m^{-1} \left(\widehat{\vec{S}_{Ym}} - \vec{S}_{Vm} \right),$$

where $\vec{S}_{Vm} = {}^t(a_0(S_V), \ldots, a_{m-1}(S_V))$, $\widehat{\vec{S}}_m = {}^t(\widehat{a}_0, \ldots, \widehat{a}_{m-1})$ and $\widehat{\vec{S}_{Ym}} = {}^t(\widehat{a}_0(Y), \ldots, \widehat{a}_{m-1}(Y))$, with $\widehat{a}_j(Y) = \dfrac{1}{n} \sum_{i=1}^n \Phi_j(Y_i)$.

The estimator $\widehat{S}_m$ is an unbiased estimator of S_m satisfying the following MISE bound: if S satisfies $(\mathbf{A1})(S)$ and $\mathbb{E}[Y_1] < +\infty$, then

$$\mathbb{E}[\|\widehat{S}_m - S\|^2] \leq \|S - S_m\|^2 + \frac{\mathbb{E}[Y_1]}{n}\|\mathbf{G}_m^{-1}\|_{\mathrm{op}}^2$$

(see Proposition 2.3.3 in Mabon, 2015). Clearly, $\widetilde{S}_m = \widehat{S}_m - K_m \sum_{j=0}^{m-1} \varphi_j$ with $K_m = m^{-1}(\sum_{\ell=0}^{m-1} \widehat{a}_\ell - 2^{-1/2})$ satisfies the bound in Proposition 2.2. Here we can prove

$$V_{n,m} = \frac{1}{m}\mathrm{Var}\left(\sum_{\ell=0}^{m-1} \widehat{a}_\ell\right) \leq \frac{\|\mathbf{G}_m^{-1}\|_F^2 \mathbb{E}[Y_1]}{nm}. \tag{5.6}$$

In the Gamma case mentioned above, i.e. if $g \sim \gamma(p,\theta)$, then $\|\mathbf{G}_m^{-1}\|_F^2/m = O(m^{2p-1})$ while $\|\mathbf{G}_m^{-1}\|_{\mathrm{op}}^2 = O(m^{2p})$ so that the bound (5.6) is such that $V_{n,m}$ can be negligible with respect to the variance.

Survival function estimation in the additive convolution model relies on the study of the empirical process $\nu_n(t) = \langle t, \widehat{S}_{m_{\max}} - S_{m_{\max}}\rangle$ with $\widehat{S}_{\widehat{m}}$ defined by

$$\widehat{\mathrm{pen}}_1(m) = \frac{2\kappa_4 \bar{Z}_n}{n}\|\mathbf{G}_m^{-1}\|_{\mathrm{op}}^2 \log(n)$$

and $\widehat{m} = \arg\min_{m \in \mathcal{M}_n}(-\|\widehat{S}_m\|^2 + \widehat{\mathrm{pen}}_1(m))$ for $\mathcal{M}_n = \{m, \|G_m^{-1}\|_{\mathrm{op}}^2 \log(n)/n \leq 1\}$. The empirical process fulfills (2.14) and the penalties pen_1 and $\widehat{\mathrm{pen}}_1(m)$ satisfy conditions (2.16)-(2.17) with $a = 0$. Thus, here again the procedure of section 2.5 can be applied to the constrained estimator.

6. Numerical study

6.1. Description of the practical procedure.
In the following Section we compare the new constrained estimator to the simple adaptive projection estimator in different cases. We illustrate the direct observation case presented in Section 3, the multiplicative noise model described in Section 4 and the additive model described in Section 5.

The size n of the samples is chosen $n = 100$ or $n = 1000$. Survival function estimation gives good results even for rather small samples. The variable of interest X is simulated from two different distributions:

- $X \sim \chi^2(10)/\sqrt{20}$ (denoted $X \sim \chi^2$),
- $X \sim 0.5\Gamma(2, 0.4) + 0.5\Gamma(11, 0.5)$ (denoted $X \sim M\Gamma$)

and in the additive model we choose to illustrate

- $V \sim \Gamma(2, 1/\sqrt{8})$.

Note that we checked that the choice $V \sim \exp(2)$ gives similar results.

Before the simulation phase, a calibration step is conducted. The universal constants κ_i appearing in all the procedures are calibrated with a large choice of setups, different from the ones of the simulation study. Empirical MISE are computed via Monte-Carlo experiments. We obtain $\kappa_0 = 0.2$ in the direct observation case, for the density estimation. In the multiplicative noise model, the constant for the density estimation is $\kappa_1 = 0.05$, and for the survival function estimation $\kappa_2 = 0.5$. For the additive model, the constants are $\kappa_3 = 0.03$ and $\kappa_4 = 0.001$ respectively.

6.2. Results and comments.
On Figure 1, we illustrate the estimation of the survival function from a sample of X distributed according to a mixture of Gamma distribution (the true function S is in plain black line). The figure is separated in two graphs, the left one represents the projection estimator $\widehat{S}$ and the right one the constrained estimator $\widetilde{S}$. We show the collection of estimators in each case for $D_{\max} = 10$ in orange (light grey). The final estimators $\widehat{S}_{m_{\max}}$ and $\widetilde{S}_{m_{\max}}$ are in bold red (black) and very close to the empirical estimator represented in dotted black line. This figure first shows that the estimator, chosen with highest dimension, performs at best, as suggested by the theoretical part.

	$X \sim \chi^2$		$X \sim M\Gamma$	
n	100	1000	100	1000
$\widehat{S}^{\mathrm{dir}}_{m_{\max}}$	0.433	0.045	1.449	0.118
$\widetilde{S}^{\mathrm{dir}}_{m_{\max}}$	0.431	0.044	1.447	0.117
$\widehat{S}^{\mathrm{mult}}_{\widehat{m}}$	2.421	0.178	3.603	0.438
$\widetilde{S}^{\mathrm{mult}}_{\widehat{m}}$	2.409	0.170	3.459	0.429
$\widehat{S}^{\mathrm{add}}_{\widehat{m}}$	0.915	0.075	1.539	0.247
$\widetilde{S}^{\mathrm{add}}_{\widetilde{m}}$	0.955	0.064	1.494	0.244

TABLE 1. Survival function estimation. MISE $\times 100$ for the estimators $\widehat{S}^{\mathrm{dir}}_{m_{\max}}$, $\widetilde{S}^{\mathrm{dir}}_{m_{\max}}$, $\widehat{S}^{\mathrm{mult}}_{\widehat{m}}$, $\widetilde{S}^{\mathrm{mult}}_{\widetilde{m}}$, $\widehat{S}^{\mathrm{add}}_{\widetilde{m}}$ with 200 repetitions.

	$X \sim \chi^2$		$X \sim M\Gamma$	
n	100	1000	100	1000
$\widehat{f}^{\mathrm{dir}}_{\widehat{m}}$	0.521	0.066	0.859	0.109
$\widetilde{f}^{\mathrm{dir}}_{\widetilde{m}}$	0.507	0.057	0.716	0.086
$\widehat{f}^{\mathrm{mult}}_{\widehat{m}}$	3.717	0.498	4.540	0.441
$\widetilde{f}^{\mathrm{mult}}_{\widehat{m}}$	2.898	0.433	3.035	0.391
$\widehat{f}^{\mathrm{add}}_{\widehat{m}}$	2.290	0.211	2.107	0.261
$\widetilde{f}^{\mathrm{add}}_{\widetilde{m}}$	2.472	0.207	1.363	0.236

TABLE 2. Density estimation. MISE $\times 100$ for the estimators $\widehat{f}^{\mathrm{dir}}_{\widehat{m}}$, $\widetilde{f}^{\mathrm{dir}}_{\widetilde{m}}$, $\widehat{f}^{\mathrm{mult}}_{\widehat{m}}$, $\widetilde{f}^{\mathrm{mult}}_{\widetilde{m}}$, $\widehat{f}^{\mathrm{add}}_{\widetilde{m}}$ with 200 repetitions.
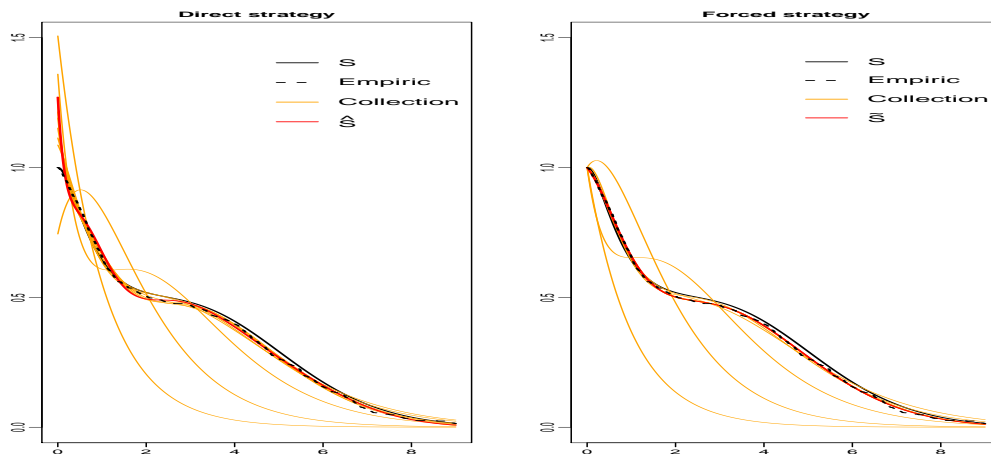

FIGURE 1. Direct observations. Survival function estimation. $X \sim MG$, $n = 100$. In dotted bold S, in plain orange (light grey) $\widehat{S}_m$ on the left and $\widetilde{S}_m$ on the right, in plain bold red (black) $\widehat{S}_{m_{\max}}$ on the left and $\widetilde{S}_{m_{\max}}$ on the right.

Then we clearly see (comparing left and right figures) that $\widetilde{S}_{m_{\max}}$ is better than $\widehat{S}_{m_{\max}}$. Besides, we also draw the empirical survival function which is the best estimation available, and our estimator $\widetilde{S}_{m_{\max}}$ is very close from it (right graph).

Figure 2 illustrates the estimation procedure detailed in Section 5 for the additive model. We draw 6 final survival function estimators $\widehat{S}_{\widehat{m}}$ in green (light grey) and $\widetilde{S}_{\widetilde{m}}$ in red (black), when X follows a chi-square distribution. They are all close to the real function S in dotted black. Nevertheless the constrained new estimator fits better the curve than the classical projection one.

Figure 3 shows the performance of the constrained density estimator in the multiplicative noise model, for X an exponential random variable with parameter 2 thus $f(x) = 2\exp(-2x)$ and $f(0) = 2$. The procedure can be applied, as explained in Section 2. The improvement of the new strategy (in red) is visually clear. Let us confirm this point on empirical risk results.

Tables 1 and 2 give the empirical MISE multiplied by 100, respectively for the survival function estimators and for the density estimators. We see the expected effect of the sample size: the larger
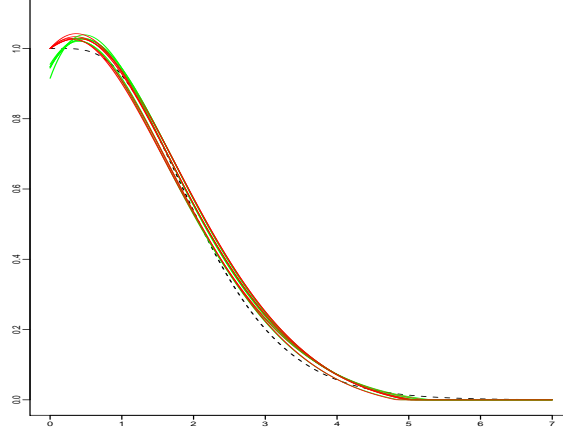
FIGURE 2. Additive model. Survival function estimation. $X \sim \chi^2$, $n = 200$. In dotted bold f, in plain green (light grey) 6 final estimators $\widehat{S}_{\widehat{m}}$, in plain red (black) 6 final estimators $\widetilde{S}_{\widetilde{m}}$.
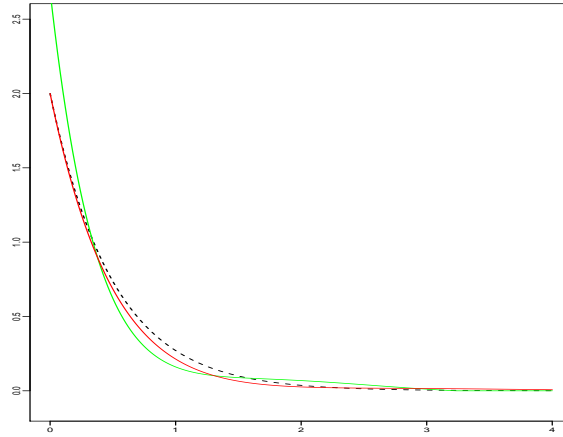

FIGURE 3. Multiplicative model. Density estimation. $X \sim \mathcal{E}(2)$, $n = 100$. In dotted bold f, in plain green (light grey) $\widehat{f}_{\widehat{m}}$, in plain red (black) $\widetilde{f}_{\widetilde{m}}$.

n, the smaller the MISE. Moreover, density estimation seems more difficult than survival function estimation. Nevertheless, we can see that the results of the new procedure, namely the three estimators $\widetilde{S}$ and $\widetilde{f}$ have smaller risks than the three simple projection estimators $\widehat{S}, \widehat{f}$ (except for the density estimation in the additive model for $n = 100$ and $X \sim \chi^2$).

This confirms the theoretical part witch shows comparative rates of convergence for the constrained and the non-constrained strategy. But this numerical study goes further and claims that the new estimators performs better and give a better fit of the estimated function than the previous ones.

7. CONCLUDING REMARKS

In this work, we consider a projection estimator $\widehat{s}_m$ built as an unbiased estimator of s_m, the projection of a function s on the space spanned by the m first functions of the Laguerre basis. We show that a general modification of $\widehat{s}_m$ can provide a new estimator $\widetilde{s}_m$ with fixed value in 0. The risk of the new estimator has slightly increased bias compared to $\widehat{s}_m$ but smaller variance. In any case, the order of the risk of the two estimators are the same. Moreover, if a model selection procedure is available for $\widehat{s}_m$, we can deduce a selection procedure for $\widetilde{s}_m$, leading to a data-driven trade-off

between the bias and the variance. Imposing a constraint at another point is possible, but it is not clear if it may be useful. In the case of survival function estimation with constraint set at 1 in 0, numerical experiments illustrate the improvement brought by our strategy, in three different contexts of observation: direct observation, multiplicative model and convolution model. These examples fitting in our setting are also studied from theoretical point of view.

The same kind of study may be conducted in other bases. For instance in the trigonometric basis on an interval, say $[0, 1]$, a constraint at 0 may also be similarly studied, but would imply the same constraint at 1. This is clearly not relevant for survival function estimation, but may be applied to density estimation, in the case of direct observation of the X_i's. Note that the specific properties of the Laguerre basis is crucial in the two other examples (multiplicative model and convolution model).

In the specific setting of Laguerre basis, other examples may be found by combining the models, as explained in Comte and Genon-Catalot (2017). Other examples fitting in our setting are given by considering observations of the form $X_i U_i + V_i$ or $(X_i + V_i)U_i$ with U_i following a uniform density on $[0, 1]$ and V_i a nonnegative random variable with known density. Also, if the V_i's have unknown density, but have been preliminarily observed, then the scalar products $\langle f_V, \varphi_j \rangle$ can be estimated: this context has been considered in Comte and Mabon (2016) and may be studied from our constraint point of view.

8. Proofs

8.1. **Proof of Proposition 2.2.** We have the general equality

$$\mathbb{E}\left[\|\widetilde{s}_m - s\|^2\right] = \|\mathbb{E}[\widetilde{s}_m] - s\|^2 + \mathbb{E}\left[\|\widetilde{s}_m - \mathbb{E}[\widetilde{s}_m]\|^2\right]. \tag{8.1}$$

We compute successively the bias and the variance terms. For the bias term, we have

$$\mathbb{E}[\widetilde{s}_m] = s_m - \frac{1}{m}\left(\sum_{\ell=0}^{m} a_\ell(s) - \frac{c}{\sqrt{2}}\right)\sum_{j=0}^{m-1} \varphi_j$$

and as $c/\sqrt{2} = \sum_{\ell \geq 0} a_\ell(s)$, we get $\mathbb{E}[\widetilde{s}_m] - s = \sum_{j \geq m} a_j(s)\varphi_j - m^{-1}\left(\sum_{\ell \geq m} a_\ell(s)\right)\sum_{j=0}^{m-1} \varphi_j$. As a consequence

$$\|\mathbb{E}[\widetilde{s}_m] - s\|^2 = \sum_{j \geq m} a_j^2(s) + \frac{1}{m}\left(\sum_{\ell \geq m} a_\ell(s)\right)^2. \tag{8.2}$$

For the variance term, we have $\widetilde{s}_m - \mathbb{E}[\widetilde{s}_m] = \sum_{j=0}^{m-1}\left[\widehat{a}_j - a_j(s) - \frac{1}{m}\sum_{\ell=0}^{m-1}(\widehat{a}_\ell - a_\ell(s))\right]\varphi_j$ and thus

$$\begin{aligned}
\|\widetilde{s}_m - \mathbb{E}[\widetilde{s}_m]\|^2 &= \sum_{j=0}^{m-1}\left[\widehat{a}_j - a_j(s) - \frac{1}{m}\sum_{\ell=0}^{m-1}(\widehat{a}_\ell - a_\ell(s))\right]^2 \\
&= \sum_{j=0}^{m-1}(\widehat{a}_j - a_j(s))^2 - \frac{1}{m}\left(\sum_{\ell=0}^{m-1}\widehat{a}_\ell - a_\ell(s)\right)^2.
\end{aligned}$$

This yields

$$\mathbb{E}\left[\|\widetilde{s}_m - \mathbb{E}[\widetilde{s}_m]\|^2\right] = \sum_{j=0}^{m-1}\mathrm{Var}(\widehat{a}_j) - \frac{1}{m}\mathrm{Var}\left(\sum_{\ell=0}^{m-1}\widehat{a}_\ell\right). \tag{8.3}$$

Plugging (8.3) and (8.2) into (8.1) gives equality (2.7). $\square$

8.2. **Proof of Theorem 2.5.** We consider the estimator $\widehat{s}_m$ as a minimum contrast estimator associated with γ_n defined by (2.13). It satisfies $\gamma_n(\widehat{s}_m) = -\|\widehat{s}_m\|^2$, $\widehat{s}_m = \underset{t \in \mathcal{S}_m}{\operatorname{argmin}}\ \gamma_n(t)$. It yields the relation for any $m, m' \in \mathcal{M}_n$ and $t \in \mathcal{S}_m$, $u \in \mathcal{S}_{m'}$,

$$
\begin{aligned}
\gamma_n(t) - \gamma_n(u) &= \|t - s\|^2 - \|u - s\|^2 + 2\langle t - u, s\rangle - 2\langle t - u, \widehat{s}_{m_{\max}}\rangle \\
&= \|t - s\|^2 - \|u - s\|^2 + 2\langle t - u, s_{m_{\max}}\rangle - 2\langle t - u, \widehat{s}_{m_{\max}}\rangle \\
&= \|t - s\|^2 - \|u - s\|^2 - 2\nu_n(t - u)
\end{aligned}
\tag{8.4}
$$

with $\nu_n(\cdot)$ defined by (2.15).

According to the relation: $\widetilde{s}_m = \widehat{s}_m - K_m \sum_{j=0}^{m-1} \varphi_j$ our strategy is to show a result on the estimator $\widehat{s}_{\widetilde{m}}$ and to deduce one for $\widetilde{s}_{\widetilde{m}}$. Notice that the definitions of $\widetilde{m}$ and $\widehat{s}_m$ give that, for all $m \in \mathcal{M}_n$,

$$
\gamma_n(\widehat{s}_{\widetilde{m}}) + \widehat{\mathrm{pen}}(\widetilde{m}) \leq \gamma_n(\widehat{s}_m) + \widehat{\mathrm{pen}}(m) \leq \gamma_n(s_m) + \widehat{\mathrm{pen}}(m).
$$

Therefore, $\gamma_n(\widehat{s}_{\widetilde{m}}) - \gamma_n(s_m) \leq \mathrm{pen}(m) - \mathrm{pen}(\widetilde{m})$ and with (8.4) we get

$$
\|\widehat{s}_{\widetilde{m}} - s\|^2 - \|s - s_m\|^2 - 2\nu_n(\widehat{s}_{\widetilde{m}} - s_m) \leq \widehat{\mathrm{pen}}(m) - \widehat{\mathrm{pen}}(\widetilde{m}).
$$

Therefore, denoting:

$$
\mathcal{B}_{m,m'} = \{t \in \mathcal{S}_{m \vee m'}, \|t\| = 1\},
\tag{8.5}
$$

we have

$$
\begin{aligned}
\|\widehat{s}_{\widetilde{m}} - s\|^2 &\leq \|s - s_m\|^2 + \widehat{\mathrm{pen}}(m) + 2\nu_n(\widehat{s}_{\widetilde{m}} - s_m) - \widehat{\mathrm{pen}}_1(\widetilde{m}) - \widehat{\mathrm{pen}}_2(\widetilde{m}) \\
&\leq \|s - s_m\|^2 + \widehat{\mathrm{pen}}(m) + \frac{1}{4}\|\widehat{s}_{\widetilde{m}} - s_m\|^2 + 4 \sup_{t \in \mathcal{B}_{m,\widetilde{m}}} \nu_n^2(t) - \widehat{\mathrm{pen}}_1(\widetilde{m}) - \widehat{\mathrm{pen}}_2(\widetilde{m})
\end{aligned}
$$

Writing that $\|\widehat{s}_{\widetilde{m}} - s_m\|^2 \leq 2\|\widehat{s}_{\widetilde{m}} - s\|^2 + 2\|s - s_m\|^2$ and gathering the terms implies

$$
\begin{aligned}
\frac{1}{2}\|\widehat{s}_{\widetilde{m}} - s\|^2 &\leq \frac{3}{2}\|s - s_m\|^2 + \widehat{\mathrm{pen}}(m) + 4 \sup_{t \in \mathcal{B}_{m,\widetilde{m}}} \left(\nu_n^2(t) - \frac{1}{4}\mathrm{pen}_1(\widetilde{m} \vee m)\right)_+ \\
&\quad + \mathrm{pen}_1(\widetilde{m} \vee m) - \widehat{\mathrm{pen}}_1(\widetilde{m}) - \widehat{\mathrm{pen}}_2(\widetilde{m}) \\
&\leq \frac{3}{2}\|s - s_m\|^2 + \widehat{\mathrm{pen}}(m) + 4 \sup_{t \in \mathcal{B}_{m,\widetilde{m}}} \left(\nu_n^2(t) - \frac{1}{4}\mathrm{pen}_1(\widetilde{m} \vee m)\right)_+ \\
&\quad + \mathrm{pen}_1(\widetilde{m}) + \mathrm{pen}_1(m) - \widehat{\mathrm{pen}}_1(\widetilde{m}) - \widehat{\mathrm{pen}}_2(\widetilde{m})
\end{aligned}
$$

as $\mathrm{pen}_1(\widetilde{m} \vee m) \leq \mathrm{pen}_1(\widetilde{m}) + \mathrm{pen}_1(m)$. Now taking expectation and using Assumption (2.14) and $\mathbb{E}[\widehat{\mathrm{pen}}_1(m)] \leq 2\mathrm{pen}_1(m)$ we get

$$
\begin{aligned}
\mathbb{E}[\|\widehat{s}_{\widetilde{m}} - s\|^2] &\leq 3\|s - s_m\|^2 + 6\mathrm{pen}_1(m) + 2\mathbb{E}[\widehat{\mathrm{pen}}_2(m)] + \frac{8C}{n} + 2\mathbb{E}[\mathrm{pen}_1(\widetilde{m}) - \widehat{\mathrm{pen}}_1(\widetilde{m})] \\
&\quad - 2\mathbb{E}[\widehat{\mathrm{pen}}_2(\widetilde{m})].
\end{aligned}
$$

As for all $m \in \mathcal{M}_n$, $\|\widetilde{s}_m - s\|^2 \leq 2\|\widehat{s}_m - s\|^2 + 2mK_m^2$, taking $m = \widetilde{m}$ with $mK_m^2/2 = \widehat{\mathrm{pen}}_2(m)$ leads to

$$
\begin{aligned}
\mathbb{E}[\|\widetilde{s}_{\widetilde{m}} - s\|^2] &\leq 2\mathbb{E}[\|\widehat{s}_{\widetilde{m}} - s\|^2] + 4\mathbb{E}[\widehat{\mathrm{pen}}_2(\widetilde{m})] \\
&\leq 6\|s - s_m\|^2 + 12\mathrm{pen}_1(m) + 4\mathbb{E}[\widehat{\mathrm{pen}}_2(m)] + \frac{16C}{n} + 4\mathbb{E}[(\mathrm{pen}_1(\widetilde{m}) - \widehat{\mathrm{pen}}_1(\widetilde{m}))_+].
\end{aligned}
$$

According to Assumption (2.17) the difference: $\mathbb{E}[(\mathrm{pen}_1(\widetilde{m}) - \widehat{\mathrm{pen}}_1(\widetilde{m}))_+]$ is under control. As, from (2.22), the penalty term $\widehat{\mathrm{pen}}_2$ satisfies:

$$
\mathbb{E}[2\widehat{\mathrm{pen}}_2(m)] \leq \mathbb{E}[\|\widehat{s}_m - s_m\|^2] + \frac{1}{m}\left(\sum_{\ell \geq m} a_\ell(s)\right)^2,
$$

we get for any $m \in \mathcal{M}_n$,

$$\mathbb{E}[\|\widetilde{s}_{\widetilde{m}} - s\|^2] \leq 6\|s - s_m\|^2 + 2\mathbb{E}[\|\widehat{s}_m - s_m\|^2] + 12\text{pen}_1(m) + \frac{2}{m}\left(\sum_{\ell \geq m} a_\ell(s)\right)^2 + \frac{C''''\log^a(n)}{n}.$$

Finally we have

$$\mathbb{E}[\|\widetilde{s}_{\widetilde{m}} - s\|^2] \leq 2\inf_{m \in \mathcal{M}_n}\left\{3\mathbb{E}[\|\widehat{s}_m - s\|^2] + \frac{1}{m}\left(\sum_{\ell \geq m} a_\ell(s)\right)^2 + 6\text{pen}_1(m)\right\} + \frac{C''''\log^a(n)}{n}$$

which the result of Theorem 2.5. $\square$

8.3. **Proof of Proposition 3.1.** We study here the MISE of the constrained estimator of the density of direct observations.

$$\begin{aligned}
V_{n,m} = \text{Var}\left(\sum_{\ell=0}^{m-1}\widehat{a}_\ell\right) &= \text{Var}\left(\sum_{\ell=0}^{m-1}\frac{1}{n}\sum_{i=1}^{n}\varphi_\ell(X_i)\right) = \sum_{0 \leq k,\ell \leq m-1}\text{cov}\left(\frac{1}{n}\sum_{i=1}^{n}\varphi_\ell(X_i), \frac{1}{n}\sum_{i=1}^{n}\varphi_k(X_i)\right) \\
&= \frac{1}{n}\sum_{0 \leq k,\ell \leq m-1}\text{cov}\left(\varphi_\ell(X_1), \varphi_k(X_1)\right) = \frac{1}{n}\text{Var}\left(\sum_{\ell=0}^{m-1}\varphi_\ell(X_1)\right) \\
&\leq \frac{1}{n}\mathbb{E}\left[\left(\sum_{\ell=0}^{m-1}\varphi_\ell(X_1)\right)^2\right] \leq \frac{\|f\|_\infty}{n}\int\left(\sum_{\ell=0}^{m-1}\varphi_\ell(x)\right)^2 dx.
\end{aligned}$$

As the φ_ℓ are orthonormal, we get $V_{n,m} \leq \|f\|_\infty m/n$, which ends the proof of Proposition 3.1. $\square$

8.4. **Proof of Proposition 3.2.** We study the projection estimator of the survival function for direct observations. We only have to look at the term of variance. We have

$$\begin{aligned}
\mathbb{E}\left[\left\|\widehat{S}_m - S_m\right\|^2\right] &= \sum_{j=0}^{m-1}\text{Var}(\widehat{a}_j) \leq \frac{1}{n}\sum_{j=0}^{m-1}\mathbb{E}\left[\left(\int_{\mathbb{R}^+}\varphi_j(x)\mathbb{1}_{X_1 \geq x}(x)dx\right)^2\right] \\
&= \frac{1}{n}\mathbb{E}\left[\sum_{j=0}^{m-1}\langle\varphi_j\mathbb{1}_{X_1 \geq .}\rangle^2\right] \leq \frac{1}{n}\mathbb{E}\left[\|\mathbb{1}_{X_1 \geq .}\|^2\right] = \frac{\mathbb{E}[X_1]}{n},
\end{aligned}$$

which gives Proposition 3.2. $\square$

8.5. **Proof of Proposition 4.1.** We study here the MISE of the projection estimator of the density in the multiplicative noise model. Let us look at the variance term: $\mathbb{E}[\|\widehat{f}_m - f_m\|^2]$. The Cauchy-Schwarz inequality and relation (4.2) imply:

$$\begin{aligned}
\mathbb{E}[\|\widehat{f}_m - f_m\|^2] &= \sum_{j=0}^{m-1}\text{Var}(\widehat{a}_j) \leq \frac{1}{n}\sum_{j=0}^{m-1}\mathbb{E}[(Y_1\varphi_j'(Y_1) + \varphi_j(Y_1))^2] = \frac{1}{n}\sum_{j=0}^{m-1}\mathbb{E}[((Y_1\varphi_j(Y_1))')^2] \\
&\leq \frac{1}{n}\sum_{j=0}^{m-1}3\mathbb{E}\left[\left(\frac{j}{2}\varphi_{j-1}(Y_1)\right)^2 + \left(\frac{1}{2}\varphi_j(Y_1)\right)^2 + \left(\frac{j+1}{2}\varphi_{j+1}(Y_1)\right)^2\right].
\end{aligned}$$

Here we use that

$$\mathbb{E}\left[\left(\frac{j}{2}\varphi_{j-1}(Y)\right)^2\right] = \int\left(\frac{j}{2}\varphi_{j-1}(y)\right)^2 f_Y(y)dy \leq \|\varphi_1\|_\infty^2\left(\frac{j}{2}\right)^2\int f_Y(y)dy \leq \frac{1}{2}j^2,$$

and it yields:

$$\mathbb{E}[\|\widehat{f}_m - f_m\|^2] \leq \frac{3}{n} \sum_{j=0}^{m-1} \left(\frac{j^2}{2} + \frac{1}{2} + \frac{(j+1)^2}{2} \right) \leq \frac{2m^3}{n} + \frac{3m}{2n},$$

and thus we obtain Equation (4.7).

Let us now prove Equation (4.8). We have $\|\widehat{f}_m - f\|^2 = \|f - f_m\|^2 + \|\widehat{f}_m - f_m\|^2$ by Pythagoras Theorem and as $\|\widehat{f}_m - f_m\|^2 = \sum_{j=0}^{m-1}(\widehat{a}_j - a_j)^2$ where $a_j = \mathbb{E}[\widehat{a}_j] = \langle f, \varphi_j \rangle$, we get

$$
\begin{aligned}
\mathbb{E}[\|\widehat{f}_m - f_m\|^2] &= \sum_{j=0}^{m-1} \mathrm{Var}(\widehat{a}_j) = \frac{1}{n} \sum_{j=0}^{m-1} \mathrm{Var}[Y_1 \varphi_j'(Y_1) + \varphi_j(Y_1)] \\
&\leq \frac{1}{n} \sum_{j=0}^{m-1} \mathbb{E}\left[(Y_1 \varphi_j'(Y_1) + \varphi_j(Y_1))^2 \right] \\
&= \frac{1}{n} \sum_{j=0}^{m-1} \left\{ \mathbb{E}\left[(Y_1 \varphi_j'(Y_1))^2 \right] + \mathbb{E}\left[2Y_1 \varphi_j'(Y_1)\varphi_j(Y_1) + \varphi_j(Y_1)^2 \right] \right\}
\end{aligned}
$$

Now we note that by Formula (4.3) applied to $t = \varphi_j^2$, we have

$$\mathbb{E}\left[2Y_1 \varphi_j'(Y_1)\varphi_j(Y_1) + \varphi_j(Y_1))^2 \right] = \mathbb{E}[\varphi_j^2(X_1)] \leq 2$$

and, as by Formula (4.2), we have $\|\varphi_j'\|^2 = 1 + 4j$, it follows from Equation (4.4), that

$$\mathbb{E}\left[(Y_1 \varphi_j'(Y_1))^2 \right] \leq \mathbb{E}[X_1]\|\varphi_j'\|^2 = (1 + 4j)\mathbb{E}[X_1].$$

Consequently

$$\mathbb{E}[\|\widehat{f}_m - f_m\|^2] \leq \frac{1}{n} \sum_{j=0}^{m-1} ((4j+1)\mathbb{E}[X] + 2) \leq 4\frac{m^2}{n}\mathbb{E}[Y_1] + 2\frac{m}{n},$$

using that $\mathbb{E}[X_1] = 2\mathbb{E}[Y_1]$. Adding this to the squared bias term $\|f - f_m\|^2$ gives Equation (4.8) and thus Proposition 4.1 is proved. $\square$

8.6. Proof of Proposition 4.2. We study here the MISE of the constrained estimator of the density in the multiplicative noise model . We have

$$\mathrm{Var}\left(\sum_{j=0}^{m-1} \widehat{a}_j \right) = \frac{1}{n}\mathrm{Var}\left(\sum_{j=0}^{m-1} (y\varphi_j(y))'(Y_1) \right).$$

Now using relation (4.2), we have

$$
\begin{aligned}
\sum_{j=0}^{m-1} (y\varphi_j(y))' &= \sum_{j=0}^{m-1} \left(-\frac{j}{2}\varphi_{j-1}(y) + \frac{1}{2}\varphi_j(y) + \frac{j+1}{2}\varphi_{j+1}(y) \right) \\
&= -\frac{1}{2} \sum_{j=0}^{m-2} (j+1)\varphi_j(y) + \frac{1}{2} \sum_{j=0}^{m-1} \varphi_j(y) + \frac{1}{2} \sum_{j=1}^{m} j\varphi_j(y) \\
&= -\frac{1}{2}\varphi_0(y) + \frac{1}{2}\varphi_0(y) + \frac{1}{2} \sum_{j=1}^{m-2} [-(j+1)\varphi_j(y) + \varphi_j(y) + j\varphi_j(y)] \\
&\quad + \frac{1}{2}\varphi_{m-1}(y) + \frac{1}{2}(m-1)\varphi_{m-1}(y) + \frac{1}{2}m\varphi_m(y) \\
&= \frac{m}{2}(\varphi_{m-1}(y) + \varphi_m(y))
\end{aligned}
$$

This implies that

$$\text{Var}\left(\sum_{j=0}^{m-1}\widehat{a}_j\right) \leq \frac{1}{n}\mathbb{E}\left[\left(\sum_{j=0}^{m-1}(y\varphi_j(y))'(Y_1)\right)^2\right] = \frac{m^2}{4n}\mathbb{E}\left[(\varphi_{m-1}(Y_1)+\varphi_m(Y_1))^2\right] \leq \frac{m^2\|f_Y\|_\infty}{2n}.$$

This, together with Equation (2.7) gives the result of Proposition 4.2. $\square$

8.7. **Proof of Theorem 4.3.** We define the contrast

$$\gamma_n(t) = \|t\|^2 - \frac{2}{n}\sum_{i=1}^{n}(t(Y_i)+Y_it'(Y_i)). \tag{8.6}$$

It is easy to check that $\widehat{f}_m = \underset{t\in\mathcal{S}_m}{\text{argmin}}\,\gamma_n(t)$ and to compute that $\gamma_n(\widehat{f}_m) = -\|\widehat{f}_m\|^2$. We notice that

$$\gamma_n(t) - \gamma_n(s) = \|t-f\|^2 - \|s-f\|^2 - 2\nu_n(t-s) \tag{8.7}$$

with

$$\nu_n(t) = \frac{1}{n}\sum_{i=1}^{n}t(Y_i) + Y_it'(Y_i) - \langle t,f\rangle = \frac{1}{n}\sum_{i=1}^{n}t(Y_i) + Y_it'(Y_i) - \mathbb{E}[t(Y_i)+Y_it'(Y_i)].$$

By definition of $\widehat{f}_{\widehat{m}}$, for all $m\in\mathcal{M}_n$, we have $\gamma_n(\widehat{f}_{\widehat{m}}) + \widehat{\text{pen}}_1(\widehat{m}) \leq \gamma_n(f_m) + \widehat{\text{pen}}_1(m)$. Denoting $m\vee m' = m^*$, and $\mathcal{B}_{m,m'}$ defined in (8.5), using (8.7) we get

$$\begin{aligned}
\|\widehat{f}_{\widehat{m}} - f\|^2 &\leq \|f-f_m\|^2 + \widehat{\text{pen}}_1(m) + 2\nu_n(\widehat{f}_{\widehat{m}}-f_m) - \widehat{\text{pen}}_1(\widehat{m}) \\
&\leq \|f-f_m\|^2 + \frac{1}{4}\|\widehat{f}_{\widehat{m}}-f_m\|^2 + 4\sup_{t\in\mathcal{B}_{m,\widehat{m}}}\nu_n^2(t) + \widehat{\text{pen}}_1(m) - \widehat{\text{pen}}_1(\widehat{m}) \\
&\leq \|f-f_m\|^2 + \frac{1}{2}\|\widehat{f}_{\widehat{m}}-f\|^2 + \frac{1}{2}\|f_m-f\|^2 + 4\sup_{t\in\mathcal{B}_{m,\widehat{m}}}\nu_n^2(t) + \widehat{\text{pen}}_1(m) - \widehat{\text{pen}}_1(\widehat{m})
\end{aligned}$$

Therefore we get

$$\begin{aligned}
\|\widehat{f}_{\widehat{m}} - f\|^2 &\leq 3\|f-f_m\|^2 + 8\sup_{t\in\mathcal{B}_{m,\widehat{m}}}\nu_n^2(t) + 2\widehat{\text{pen}}_1(m) - 2\widehat{\text{pen}}_1(\widehat{m}) \\
&\leq 3\|f-f_m\|^2 + 2\widehat{\text{pen}}_1(m) + 8\left(\sup_{t\in\mathcal{B}_{m,\widehat{m}}}\nu_n^2(t) - p(m,\widehat{m})\right) \\
&\quad + 8p(m,\widehat{m}) - 2\widehat{\text{pen}}_1(\widehat{m})
\end{aligned} \tag{8.8}$$

with

$$p(m,m') := 4\frac{(1+48\log(2+m^*))m^*(1+2\mathbb{E}[Y_1]m^*)}{n}$$

satisfying $4p(m,m') \leq \text{pen}_1(m) + \text{pen}_1(m')$ for $\kappa \geq \kappa_0 = 400$ with

$$\text{pen}_1(m) = \kappa\frac{\log(2+m)m(1+2\mathbb{E}[Y_1]m)}{n}, \quad \widehat{\text{pen}}_1(m) = \kappa\frac{\log(2+m)m(1+4\bar{Y}_nm)}{n}.$$

Let us state an intermediate result.

Lemma 8.1. *Under the assumption of Theorem 4.3,*

$$\mathbb{E}\left[\left(\sup_{t\in\mathcal{B}_{m,\hat{m}}}\nu_n^2(t) - p(m,\widehat{m})\right)_+\right] \leq \frac{K_1}{n}.$$

Taking expectation of (8.8), and plugging the results of Lemmas 8.1 implies

$$\mathbb{E}[\|\widehat{f}_{\widehat{m}} - f\|^2] \leq 3\|f - f_m\|^2 + 6\mathrm{pen}_1(m) + \frac{8K_1}{n} + 2\mathbb{E}\left(\mathrm{pen}_1(\widehat{m}) - \widehat{\mathrm{pen}}_1(\widehat{m})\right)_+.$$

Let $\Omega_n := \{\mathbb{E}[Y_1] - \bar{Y}_n| \leq \mathbb{E}[Y_1]/2\}$. Then according to this definition we have

$$\mathbb{E}\left[(\mathrm{pen}_1(\widehat{m}) - \widehat{\mathrm{pen}}_1(\widehat{m}))_+\right] \leq \mathbb{E}\left[\frac{2\kappa}{n}\log(2 + \widehat{m})\widehat{m}^2\left(\frac{\mathbb{E}[Y_1]}{2} - \bar{Y}_n\right)_+ \mathbb{1}_{\Omega_n^c}\right].$$

Moreover, Bienaymé-Tchebychev's inequality leads to

$$\mathbb{P}(\Omega_n^c) \leq \frac{4}{(\mathbb{E}[Y_1])^2}\frac{\mathrm{Var}(Y_1)}{n},$$

and Cauchy-Swharz's inequality implies,

$$\mathbb{E}\left[|\mathbb{E}[Y_1] - \bar{Y}_n|\mathbb{1}_{\Omega_n^c}\right] \leq \mathbb{E}\left[\left(\mathbb{E}[Y_1] - \bar{Y}_n\right)^2\right]^{1/2}\mathbb{P}(\Omega_n^c)^{1/2}$$

$$\leq \frac{\sqrt{\mathrm{Var}(Y_1)}}{\sqrt{n}}\frac{2\sqrt{\mathrm{Var}(Y_1)}}{\mathbb{E}[Y_1]\sqrt{n}} \leq \frac{2\mathrm{Var}(Y_1)}{n\mathbb{E}[Y_1]}.$$

Finally, using that $m \leq \sqrt{n}$ in $\mathcal{M}_n$, we get for $n > 1$,

$$\mathbb{E}\left[(\mathrm{pen}_1(\widehat{m}) - \widehat{\mathrm{pen}}_1(\widehat{m}))_+\right] \leq 4\kappa\log(2 + \sqrt{n})\frac{\mathrm{Var}(Y_1)}{\mathbb{E}[Y_1]}\frac{1}{n}$$

$$\leq 4\kappa\frac{\mathrm{Var}(Y_1)}{\mathbb{E}[Y_1]}\frac{\log(n)}{n} = C\frac{\log(n)}{n}.$$

This completes the proof of Theorem 4.3. $\square$

8.8. **Proof of Lemma 8.1.** First notice that,

$$\mathbb{E}\left[\left(\sup_{t \in \mathcal{B}_{m,\widehat{m}}} \nu_n^2(t) - p(m, \widehat{m})\right)_+\right] \leq \sum_{m' \in \mathcal{M}_n}\left(\sup_{t \in \mathcal{B}_{m,m'}} \nu_n^2(t) - p(m, m')\right)_+.$$

In the following we apply Talagrand's inequality to the above term. For that purpose, we compute the terms denoted by H^2, v and M in Theorem A.1.

We bound $\mathbb{E}\left[\sup_{t \in \mathcal{B}_{m,m'}} \nu_n^2(t)\right]$. For $t \in \mathcal{B}_{m,m'}$, using that $t \mapsto \nu_n(t)$ is linear and $t = \sum_{j=0}^{m^*-1}\langle t, \varphi_j\rangle\varphi_j$ with $\|t\|^2 = \sum_{j=0}^{m^*-1}\langle t, \varphi_j\rangle^2 = 1$, we get

$$\nu_n^2(t) = \left(\nu_n\left(\sum_{j=0}^{m^*-1}\langle t, \varphi_j\rangle\varphi_j\right)\right)^2 = \left(\sum_{j=0}^{m^*-1}\langle t, \varphi_j\rangle\nu_n(\varphi_j)\right)^2 \leq \sum_{j=0}^{m^*-1}\nu_n^2(\varphi_j).$$

Thus it follows from Proposition 4.1 (see the bound (4.8)),

$$\mathbb{E}\left[\sup_{t \in \mathcal{B}_{m,m'}} \nu_n^2(t)\right] \leq \sum_{j=0}^{m^*-1}\mathbb{E}[\nu_n(\varphi_j)^2] = \sum_{j=0}^{m^*-1}\frac{1}{n}\mathrm{Var}((Y_1\varphi_j'(Y_1) + \varphi_j(Y_1)) \leq \frac{2m^*}{n}(1 + 2m^*\mathbb{E}[Y_1]) =: H^2.$$

Then,

$$\mathrm{Var}(Y_1t'(Y_1) + t(Y_1)) \leq \mathbb{E}[(Y_1t'(Y_1) + t(Y_1))^2] \leq nH^2 =: v.$$

Finally, using Formula (4.2) and the fact that the basis function φ_j's are bounded by $\sqrt{2}$, we get

$$\sup_{t \in \mathcal{B}_{m,m'}}\sup_y |(yt(y))'| \leq \left(\sum_{j=0}^{m^*-1}(\sup_y(y\varphi_j(y))')^2\right)^{1/2} \leq \left(\sum_{j=0}^{m^*-1}(\sqrt{2}(j+1))^2\right)^{1/2}\sqrt{2/3}(m^*)^{3/2} =: M.$$

We obtain with $\alpha = \alpha(m^*) = 24\log(m^* + 2)$ in Theorem A.1:

$$\mathbb{E}\left[\left(\sup_{t \in \mathcal{B}_{m,m'}} \nu_n^2(t) - 2(1 + 2\alpha(m^*))\frac{2m^*}{n}(1 + 2m^*\mathbb{E}[Y_1])\right)_+\right]$$

$$\leq \frac{C}{n}\left(\frac{2m^*(1 + 2m^*\mathbb{E}[Y_1])}{(m^* + 2)^4} + \frac{(m^*)^3}{n}e^{-C_2\sqrt{\mathbb{E}[Y_1]}n^{1/4}}\right),$$

using that any $m \in \mathcal{M}_n$ satisfies $m \leq \sqrt{n}$.

Consequently,

$$\sum_{m' \in \mathcal{M}_n} \mathbb{E}\left[\left(\sup_{t \in \mathcal{B}_{m,m'}} \nu_n^2(t) - 2(1 + 2\alpha(m^*))\frac{2m^*}{n}(1 + 2m^*\mathbb{E}[Y_1])\right)_+\right] \leq \frac{K_1}{n}$$

where K_1 is a constant, which is the result announced in Lemma 8.1. $\square$

8.9. **Proof of Proposition 4.4.** We study here the risk of the projection estimator of the survival function in the multiplicative model. We have $\mathbb{E}[\|\widehat{S}_m - S\|^2] = \|S_m - S\|^2 + \mathbb{E}[\|\widehat{S}_m - S_m]\|^2]$, and we want to upper bound the term: $\mathbb{E}[\|\widehat{S}_m - S_m]\|^2]$.

$$\mathbb{E}\left[\left\|\widehat{S}_m - S_m\right\|^2\right] = \sum_{j=0}^{m-1} \text{Var}(\widehat{a}_j) = \frac{1}{n}\sum_{j=0}^{m-1}\text{Var}\left(\int_{\mathbb{R}^+} \varphi_j(x)\mathbb{1}_{Y_1 \geq x}dx + Y_1\varphi_j(Y_1)\right)$$

$$\leq \frac{2}{n}\sum_{j=0}^{m-1}\mathbb{E}\left[(\Phi_j(Y_1) + Y_1\varphi_j(Y_1))^2\right]$$

where $\Phi_j(x) = \int_0^x \varphi_j(u)du$. Now write that

$$\mathbb{E}\left[(\Phi_j(Y_1) + Y_1\varphi_j(Y_1))^2\right] = \mathbb{E}[\Phi_j^2(Y_1) + 2Y_1\Phi_j(Y_1)\varphi_j(Y_1)] + \mathbb{E}[Y_1^2\varphi_j^2(Y_1)].$$

Now we note that by Formula (4.3) applied to $t = \Phi_j^2$, we have

$$\mathbb{E}[\Phi_j^2(Y_1) + 2Y_1\Phi_j(Y_1)\varphi_j(Y_1)] = \mathbb{E}[\Phi_j^2(X_1)].$$

Moreover, it follows from (4.4) that

$$\mathbb{E}\left[(Y_1\varphi_j(Y_1))^2\right] \leq \mathbb{E}[X_1]\|\varphi_j\|^2 = \mathbb{E}[X_1].$$

Therefore,

$$\mathbb{E}\left[\left\|\widehat{S}_m - S_m\right\|^2\right] \leq \frac{1}{n}\sum_{j=0}^{m-1}\mathbb{E}\left[\left(\int_{\mathbb{R}^+} \varphi_j(x)\mathbb{1}_{X_1 \geq x}(x)dx\right)^2\right] + \frac{1}{n}\sum_{j=0}^{m-1}\mathbb{E}(Y_1^2\varphi_j^2(Y_1))$$

$$\leq \frac{1}{n}\mathbb{E}\left[\sum_{j=0}^{m-1}\langle\varphi_j, \mathbb{1}_{X_1 \geq \cdot}\rangle^2\right] + \frac{1}{n}\sum_{j=0}^{m-1}\mathbb{E}[X_1]$$

$$\leq \frac{1}{n}\mathbb{E}\left[\|\mathbb{1}_{X_1 \geq \cdot}\|^2\right] + \mathbb{E}[X_1]\frac{m}{n} = \mathbb{E}[X_1]\frac{m+1}{n}.$$

This is the result of Proposition 4.4. $\square$

8.10. **Proof of Proposition 4.5.** We study here the additional variance term of the constrained estimator of the survival function in the multiplicative model. As previously, we apply Formula (4.3) applied to $t = (\sum_{j=0}^{m-1} \Phi_j)^2$ (recall that $\Phi_j(x) = \int_0^x \varphi_j(u)du$), and we find

$$
\begin{aligned}
\operatorname{Var}\left(\sum_{j=0}^{m-1} \widehat{a}_j\right) &= \frac{1}{n}\operatorname{Var}\left(\sum_{\ell=0}^{m-1} \int_{\mathbb{R}^+} \varphi_\ell(u)\mathbb{1}_{Y_1 \geq u}du + Y_1\varphi_\ell(Y_1)\right) \\
&\leq \frac{1}{n}\mathbb{E}\left[\left(\sum_{\ell=0}^{m-1} \int_{\mathbb{R}^+} \varphi_\ell(u)\mathbb{1}_{X_1 \geq u}du\right)^2\right] + \frac{1}{n}\mathbb{E}\left[\left(\sum_{\ell=0}^{m-1} Y_1\varphi_\ell(Y_1)\right)^2\right].
\end{aligned}
$$

Now, we write by using Cauchy Schwarz and $\|\sum_{\ell=0}^{m-1} \varphi_\ell\|^2 = m$,

$$
\left(\sum_{\ell=0}^{m-1} \int_{\mathbb{R}^+} \varphi_\ell(u)\mathbb{1}_{x \geq u}du\right)^2 \leq \int_{\mathbb{R}^+} \left(\sum_{\ell=0}^{m-1} \varphi_\ell(u)\right)^2 du \int_{\mathbb{R}^+} \mathbb{1}_{x \geq u}du = mx,
$$

and we obtain

$$
\frac{1}{n}\mathbb{E}\left[\left(\sum_{\ell=0}^{m-1} \int_{\mathbb{R}^+} \varphi_\ell(u)\mathbb{1}_{X_1 \geq u}du\right)^2\right] \leq \mathbb{E}[X_1]\frac{m}{n}
$$

Besides, we use the property (4.4). It comes,

$$
\frac{1}{n}\mathbb{E}\left[\left(Y_1\sum_{\ell=0}^{m-1} \varphi_\ell(Y_1)\right)^2\right] \leq \frac{\mathbb{E}[X_1]}{n}\left\|\sum_{\ell=0}^{m-1} \varphi_\ell\right\|^2 = \mathbb{E}[X_1]\frac{m}{n}.
$$

We obtain finally:

$$
\operatorname{Var}\left(\sum_{j=0}^{m-1} \widehat{a}_j\right) \leq 2\mathbb{E}[X_1]\frac{m}{n} = 4\mathbb{E}[Y_1]\frac{m}{n}.
$$

This concludes the proof of Proposition 4.5. $\square$

8.11. **Proof of Theorem 4.7.** Let us now prove the oracle-type inequality for the survival function estimator in the multiplicative noise context. We prove that Assumptions (2.14)-(2.16)-(2.17) are fulfilled. Assumptions (2.16)-(2.17) follow from the density case, so we check (2.14). We set $p(m, m') = \operatorname{pen}(m \vee m')$ and we write

$$
\begin{aligned}
\mathbb{E}\left[\sup_{t \in B_{m,\widehat{m}}} \nu_n^2(t) - p(m, \widehat{m})\right]_+ &\leq 3\mathbb{E}\left[\sup_{t \in B_{m,\widehat{m}}} \nu_{n,1}^2(t) - p(m, \widehat{m})\right]_+ + 3\mathbb{E}\left[\sup_{t \in B_{m,\widehat{m}}} \nu_{n,2}^2(t)\right] \\
&\quad + 3\mathbb{E}\left[\sum_{t \in B_{m,\widehat{m}}} R_n^2(t)\right]_+
\end{aligned}
$$

where

$$
\nu_{n,1}(t) = \frac{1}{n}\sum_{i=1}^n \left(Y_i t(Y_i)\mathbb{1}_{Y_i \leq c_n} - \mathbb{E}[Y_i t(Y_i)\mathbb{1}_{Y_i \leq c_n}]\right), \quad \nu_{n,2}(t) = \frac{1}{n}\sum_{i=1}^n \left(Y_i t(Y_i)\mathbb{1}_{Y_i > c_n} - \mathbb{E}[Y_i t(Y_i)\mathbb{1}_{Y_i > c_n}]\right)
$$

and

$$
R_n(t) = \frac{1}{n}\sum_{i=1}^n \left(\int t(x)\mathbb{1}_{Y_i \geq x}dx - \mathbb{E}[\int t(x)\mathbb{1}_{Y_i \geq x}dx]\right).
$$

First for $m^* = m \vee \widehat{m}$,

$$
\begin{aligned}
\mathbb{E}\left[\left(\sup_{t\in B_{m,\widehat{m}}} R_n^2(t)\right)_+\right] & \leq \mathbb{E}\left[\sum_{j=1}^{m^*-1}\left(\int \varphi_j(x)\frac{1}{n}\sum_{i=1}^n(\mathbb{1}_{Y_i\geq x}-S_Y(x))dx\right)^2\right] \\
& \leq \mathbb{E}\left[\sum_{j=1}^{m_{\max}-1}\left(\int \varphi_j(x)\frac{1}{n}\sum_{i=1}^n(\mathbb{1}_{Y_i\geq x}-S_Y(x))dx\right)^2\right] \\
& \leq \frac{1}{n}\sum_{j=1}^{m_{\max}-1}\mathbb{E}\left[\left(\int \varphi_j(x)\left(\frac{1}{n}\sum_{i=1}^n\mathbb{1}_{Y_i\geq x}\right)dx\right)^2\right] \\
& \leq \frac{1}{n}\mathbb{E}[\|\mathbb{1}_{Y_1\geq .}\|^2] = \frac{\mathbb{E}[Y_1]}{n}.
\end{aligned}
\tag{8.9}
$$

Next we apply Talagrand's inequality (Theorem A.1) to $\nu_{n,1}$ and to that aim, we compute H^2, v and M. Clearly, denoting $m^* = m \vee m'$,

$$
\mathbb{E}\left[\sup_{t\in B_{m,m'}}\nu_{n,1}^2(t)\right] \leq \frac{\mathbb{E}[X_1]m^*}{n} = 2\frac{\mathbb{E}[Y_1]m^*}{n} := H^2
$$

Next using (4.4), we get

$$
\sup_{\|t\|=1}\mathrm{Var}(Y_1 t(Y_1)\mathbb{1}_{Y_1\leq c_n}) \leq \sup_{\|t\|=1}\mathbb{E}[Y_1^2 t^2(Y_1)] \leq \sup_{\|t\|=1}\mathbb{E}[X_1]\|t\|^2 = 2\mathbb{E}[Y_1] := v.
$$

Lastly, $\sup_{\|t\|=1}\sup_{y\in\mathbb{R}^+}|yt(y)\mathbb{1}_{y\leq c_n}| \leq c_n\sqrt{2m^*} = M$. Then we obtain, by taking $\alpha = 1/4$ in Theorem A.1, that, for $p(m,m') = 3\mathbb{E}[Y_1]m^*/n$

$$
\begin{aligned}
\mathbb{E}\left[\left(\sup_{t\in B_{m,\widehat{m}}}\nu_{n,1}^2(t)-p(m,\widehat{m})\right)_+\right] & \leq \sum_{m'\in\mathcal{M}_n}\mathbb{E}\left[\left(\sup_{t\in B_{m,m'}}\nu_{n,1}^2(t)-p(m,m)\right)_+\right] \\
& \leq \frac{C}{n}\sum_{m'\in\mathcal{M}_n}\left(\mathbb{E}[Y_1]e^{-bm^*/4}+c_n^2\frac{m^*}{n}e^{-c'\sqrt{\mathbb{E}[Y_1]}\sqrt{n}/c_n}\right)
\end{aligned}
$$

where C and c' are two numerical constants. Therefore, choosing

$$
c_n = \frac{c'}{\sqrt{2}}\sqrt{\mathbb{E}[Y_1]n}/\log(n)
$$

and using that $\mathrm{card}(\mathcal{M}_n) \leq n$ and $m^* \leq n$, yields

$$
\mathbb{E}\left[\left(\sup_{t\in B_{m,\widehat{m}}}\nu_{n,1}^2(t)-p(m,\widehat{m})\right)_+\right] \leq \frac{K_2}{n}.
\tag{8.10}
$$

Now,

$$
\begin{aligned}
\mathbb{E}\left[\sup_{t\in B_{m,\widehat{m}}}\nu_{n,2}^2(t)\right] & \leq \mathbb{E}\left[\sum_{j=1}^{m\vee\widehat{m}}\nu_{n,2}^2(\varphi_j)\right] \leq \mathbb{E}\left[\sum_{j=1}^{m_{\max}}\nu_{n,2}^2(\varphi_j)\right] \leq \frac{1}{n}\sum_{j=1}^{m_{\max}}\mathrm{Var}\left(Y_1\varphi_j(Y_1)\mathbb{1}_{Y_1>c_n}\right) \\
& \leq 2\mathbb{E}[Y_1^2\mathbb{1}_{Y_1>c_n}] \leq 2\frac{\mathbb{E}[Y_1^4]}{c_n^2} = \frac{\mathbb{E}[Y_1^4]}{\mathbb{E}[Y_1]}\frac{\log^2(n)}{n}.
\end{aligned}
\tag{8.11}
$$

Finally gathering (8.9), (8.10), (8.11) leads to the result of Theorem 4.7. $\square$

8.12. Proof of Inequality (5.5). Let $\vec{\varphi}_m(Y_1) = {}^t(\varphi_0(Y_1), \ldots, \varphi_{m-1}(Y_1))$. We have

$$
\begin{aligned}
n\mathrm{Var}\left(\sum_{\ell=0}^{m-1} \widehat{a}_\ell\right) &= \mathrm{Var}\left(\sum_{\ell=0}^{m-1}\left[\mathbf{G}_m^{-1}\vec{\varphi}_m(Y_1)\right]_\ell\right) \leq \mathbb{E}\left[\left(\sum_{\ell=0}^{m-1}\left[\mathbf{G}_m^{-1}\vec{\varphi}_m(Y_1)\right]_\ell\right)^2\right] \\
&\leq \mathbb{E}\left[\left(\sum_{\ell=0}^{m-1}\sum_{j=0}^{m-1}[\mathbf{G}_m^{-1}]_{\ell,j}\varphi_j(Y_1)\right)^2\right] = \mathbb{E}\left[\left(\sum_{j=0}^{m-1}\left(\sum_{\ell=0}^{m-1}[\mathbf{G}_m^{-1}]_{\ell,j}\right)\varphi_j(Y_1)\right)^2\right] \\
&\leq \|f_Y\|_\infty \sum_{j=0}^{m-1}\left(\sum_{\ell=0}^{m-1}[\mathbf{G}_m^{-1}]_{\ell,j}\right)^2 = \|f_Y\|_\infty\|\mathbf{G}_m^{-1}\vec{\mathbf{1}}_m\|^2 \\
&\leq \|f_Y\|_\infty\|\mathbf{G}_m^{-1}\|_{\mathrm{op}}^2\|\vec{\mathbf{1}}_m\|^2 = \|f_Y\|_\infty m\|\mathbf{G}_m^{-1}\|_{\mathrm{op}}^2.
\end{aligned}
$$

Consequently we get $m^{-1}\mathrm{Var}\left(\sum_{\ell=0}^{m-1}\widehat{a}_\ell\right) \leq \|f_Y\|_\infty\|\mathbf{G}_m^{-1}\|_{\mathrm{op}}^2/n$, which is (5.5). □

8.13. Proof of Inequality (5.6).

$$
\begin{aligned}
n\mathrm{Var}\left(\sum_{\ell=0}^{m-1}\widehat{a}_\ell\right) &= \mathrm{Var}\left(\sum_{\ell=0}^{m-1}\left[\mathbf{G}_m^{-1}\vec{\Phi}_m(Y_1)\right]_\ell\right) \leq \mathbb{E}\left[\left(\sum_{\ell=0}^{m-1}\left[\mathbf{G}_m^{-1}\vec{\Phi}_m(Y_1)\right]_\ell\right)^2\right] \\
&\leq \mathbb{E}\left[\left(\sum_{\ell=0}^{m-1}\sum_{j=0}^{m-1}[\mathbf{G}_m^{-1}]_{\ell,j}^2\sum_{\ell=0}^{m-1}\Phi_\ell^2(Y_1)\right)\right] \leq \|\mathbf{G}_m^{-1}\|_F^2\mathbb{E}\left[\sum_{\ell=0}^{m-1}\Phi_\ell^2(Y_1)\right] \\
&\leq \|\mathbf{G}_m^{-1}\|_F^2\mathbb{E}[Y_1].
\end{aligned}
$$

Therefore, we get (5.6). □

APPENDIX A. TALAGRAND'S INEQUALITY

The following result follows from the Talagrand concentration inequality.

Theorem A.1. *Consider* $n \in \mathbb{N}^*$, $\mathcal{F}$ *a class at most countable of measurable functions, and* $(X_i)_{i\in\{1,\ldots,n\}}$ *a family of real independent random variables. Define, for* $f \in \mathcal{F}$, $\nu_n(f) = (1/n)\sum_{i=1}^n(f(X_i) - \mathbb{E}[f(X_i)])$, *and assume that there are three positive constants* M, H *and* v *such that* $\sup_{f\in\mathcal{F}}\|f\|_\infty \leq M$, $\mathbb{E}[\sup_{f\in\mathcal{F}}|\nu_n(f)|] \leq H$, *and* $\sup_{f\in\mathcal{F}}(1/n)\sum_{i=1}^n\mathrm{Var}(f(X_i)) \leq v$. *Then for all* $\alpha > 0$,

$$
\begin{aligned}
\mathbb{E}\left[\left(\sup_{f\in\mathcal{F}}|\nu_n(f)|^2 - 2(1+2\alpha)H^2\right)_+\right] &\leq \frac{4}{b}\left(\frac{v}{n}\exp\left(-b\alpha\frac{nH^2}{v}\right)\right. \\
&\left. + \frac{49M^2}{bC^2(\alpha)n^2}\exp\left(-\frac{\sqrt{2}bC(\alpha)\sqrt{\alpha}}{7}\frac{nH}{M}\right)\right)
\end{aligned}
$$

with $C(\alpha) = (\sqrt{1+\alpha} - 1) \wedge 1$, *and* $b = \frac{1}{6}$.

REFERENCES

Abbaszadeh, M., Chesneau, C. and Doosti, H. (2013). Multiplicative censoring: estimation of a density and its derivatives under the l-p risk. *Revstat Statistical Journal* **11**, 255–276.

Abramowitz, M. and Stegun, I. A. (1966). Handbook of mathematical functions. *Applied Mathematics Series* **55**.

Andersen, K. and Hansen, M. (2001). Multiplicative censoring: density estimation by a series expansion approach. *Journal of Statistical Planning and Inference* **98**, 137 – 155.

Balabdaoui, F. and Wellner, J. A. (2007). Estimation of a k-monotone density: limit distribution theory and the spline connection. *Ann. Statist.* **35**, 2536–2564. ISSN 0090-5364.

Belomestny, D., Comte, F. and Genon-Catalot, V. (2016). Nonparametric laguerre estimation in the multiplicative censoring model. *Electron. J. Statist.* **10**, 3114–3152.

Bongioanni, B. and Torrea, J. L. (2009). What is a Sobolev space for the Laguerre function systems? *Studia Math.* **192**, 147–172.

Brunel, E., Comte, F. and Genon-Catalot, V. (2016). Nonparametric density and survival function estimation in the multiplicative censoring model. *TEST* **25**, 570–590. ISSN 1133-0686.

Carroll, R. and Hall, P. (1988). Optimal rates of convergence for deconvolving a density. *Journal of the American Statistical Association* **83**, 1184–1186.

Chernozhukov, V., Fernández-Val, I. and Galichon, A. (2009). Improving point and interval estimators of monotone functions by rearrangement. *Biometrika* **96**, 559–575.

Comte, F., Cuenod, C.-A., Pensky, M. and Rozenholc, Y. (2017). Laplace deconvolution and its application to dynamic contrast enhanced imaging. *J.R.Stat.Soc.Ser.B.Stat.Methodol.* **79**, 69–94.

Comte, F. and Dion, C. (2016). Nonparametric estimation in a multiplicative censoring model with symmetric noise. *Journal of Nonparametric Statistics* **28**, 1–34.

Comte, F. and Genon-Catalot, V. (2015). Adaptive laguerre density estimation for mixed poisson models. *Electronic Journal of Statistics* **9**, 1113–1149.

Comte, F. and Genon-Catalot, V. (2017). Laguerre basis for inverse problems. *Preprint Hal and Preprint MAP5 2017-05* .

Comte, F. and Mabon, G. (2016). Laguerre deconvolution with unknown matrix operator. *Preprint Hal and Preprint MAP5 2016-33* .

Comte, F., Rozenholc, Y. and Taupin, M.-L. (2006). Penalized contrast estimator for adaptive density deconvolution. *Can. J. Stat.* **34**, 431–452.

Efromovich, S. (1999). *Nonparametric curve estimation.* Springer Series in Statistics. Springer-Verlag, New York. ISBN 0-387-98740-1. Methods, theory, and applications.

Fan, J. (1991). On the optimal rates of convergence for nonparametric deconvolution problems. *Ann. Statist.* **19**, 1257–1272.

Huang, J. and Wellner, J. A. (1995). Estimation of a monotone density or monotone hazard under random censoring. *Scand. J. Statist.* **22**, 3–33. ISSN 0303-6898.

Jirak, A., M.and Meister and Reiß, M. (2014). Adaptive function estimation in nonparametric regression with one-sided errors. *The Annals of Statistics* **42**, 1970–2002.

Mabon, G. (2015). Adaptive deconvolution on the nonnegative real line. *To appear in Scandinavian Journal of Statistics* URL https://hal.archives-ouvertes.fr/hal-01076927,version2. Preprint HAL 01076927.

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

Pensky, M. and Vidakovic, B. (1999). Adaptative wavelet estimator for nonparametric density deconvolution. *Ann. Statist.* **27**, 2033–2053.

van Es, B., Klaassen, C. A. J. and Oudshoorn, K. (2000). Survival analysis under cross-sectional sampling: length bias and multiplicative censoring. *J. Statist. Plann. Inference* **91**, 295–312.

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