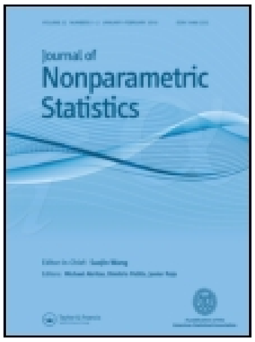

Journal of Nonparametric Statistics

ISSN: 1048-5252 (Print) 1029-0311 (Online) Journal homepage: https://www.tandfonline.com/loi/gnst20

Conditional mean residual life estimation

Elodie Brunel & Fabienne Comte

To cite this article: Elodie Brunel & Fabienne Comte (2011) Conditional mean residual life estimation, Journal of Nonparametric Statistics, 23:2, 471-495, DOI: 10.1080/10485252.2010.544731

To link to this article: https://doi.org/10.1080/10485252.2010.544731

Published online: 10 Feb 2011.

Journal of Nonparametric Statistics
Vol. 23, No. 2, June 2011, 471–495

Conditional mean residual life estimation

Elodie Brunel[a] and Fabienne Comte[b]*

[a] *I3M, UMR 5149 CNRS, Montpellier 2 University, France;* [b] *MAP5, UMR 8145 CNRS, Paris Descartes University, France*

(*Received 4 March 2010; final version received 29 November 2010*)

In this paper, we consider the problem of nonparametric mean residual life (MRL) function estimation in presence of covariates. We propose a contrast that provides estimators of the bivariate conditional MRL function, when minimised over different collections of linear finite-dimensional function spaces. Then we describe a model selection device to select the best estimator among the collection, in the mean integrated squared error sense. A non-asymptotic oracle inequality is proved for the estimator, which both ensures the good finite sample performances of the estimator and allows us to compute asymptotic rates of convergence. Lastly, examples and simulation experiments illustrate the method, together with a short real data study.

Keywords: adaptive methods; covariate; mean residual life; nonparametric estimation; survival analysis

MSC 2010: 63G05; 62N02

1. Introduction

In randomised clinical trials, survival times are often measured from randomisation or treatment implementations. But studying survival functions or hazard rates may be inadequate to answer a patient asking, during the trial, how much more time he/she still has or whether the new treatment improves his/her life expectancy. To correctly address these questions, life expectancy must be studied as a function of time, via the so-called mean residual life (MRL) function:

$$e(y) = \mathbb{E}(Y - y|Y > y), \quad y > 0, \tag{1}$$

where Y is a lifetime (i.e. a nonnegative random variable) with $\mathbb{E}(Y) < +\infty$. This function – average remaining life of a surviving subject – is of interest in several other application fields, such as reliability or actuarial studies. For a discussion concerning statistical applications of the MRL, we refer to Embrechts, Klüppelberg and Mikosch (1997). If we denote by F the cumulative distribution function (cdf) and by $\bar{F} = 1 - F$ the survival function, we have the following formula

*Corresponding author. Email: fabienne.comte@parisdescartes.fr

ISSN 1048-5252 print/ISSN 1029-0311 online

DOI: 10.1080/10485252.2010.544731
http://www.informaworld.com

for the MRL:

$$e(y) = \begin{cases} \int_y^{+\infty} \bar{F}(u)\,\mathrm{d}u/\bar{F}(y) & \text{if } \bar{F}(y) > 0 \\ 0 & \text{otherwise.} \end{cases}$$

This equality leads to several proposals of nonparametric estimators, built by plug-in of Kaplan-Meier survival estimators, see Hall and Wellner (1981) or Csörgö and Zitikis (1996) and the references therein. Under adequate assumptions, these estimators inherit the parametric rates of the Kaplan–Meier estimator, but unfortunately they are not smooth. To circumvent this drawback, regularised estimators based on kernel or spline smoothing have been proposed by Chaubey and Sen (1999), Na and Kim (1999) or Abdous and Berred (2005).

To measure the combined effect of a covariate X on the MRL, we shall rather define and study the conditional MRL:

$$e(y|x) = \mathbb{E}(Y - y|Y > y, X = x) = \begin{cases} \int_y^{+\infty} \bar{F}(u|x)\,\mathrm{d}u/\bar{F}(y|x) & \text{if } \bar{F}(y|x) > 0 \\ 0 & \text{otherwise,} \end{cases} \tag{2}$$

where $\bar{F}(y|x)$ is the conditional survival function of Y given $X = x$:

$$\bar{F}(y|x) = \mathbb{P}(Y > y|X = x) = \frac{\int_y^{+\infty} f_{(X,Y)}(x, u)\,\mathrm{d}u}{f_X(x)} \quad \text{if } f_X(x) > 0.$$

Here $f_{(X,Y)}$ denotes the joint probability density of (X, Y) and f_X denotes the marginal density of X. In semi-parametric regression analysis, Oakes and Dasu (1990) proposed a proportional MRL model to study the association with related covariates, when the response is completely observed. This model is studied in Maguluri and Zhang (1994). Then, Chen and Cheng (2005) and Chen, Jewell, Lei, and Cheng (2005) developed strategies in this model for censored response. Few nonparametric strategies have been developed yet. One is based on formula (2), replacing the conditional survival function by original estimators in a recent working paper by McLain and Ghosh (2009).

In this paper, we propose a minimum contrast estimator of the conditional MRL. It is also a purely nonparametric approach, which amounts to apply model selection methods to the specific MRL estimation problem. To be more precise, we propose a regression-type contrast which is original and interesting in this context. We minimise it over collections of finite-dimensional functional spaces spanned by orthonormal bases, called models. This produces a collection of estimators among which the best one, in a sense to be defined, is chosen by using a penalisation device. The resulting estimator is proved to satisfy an oracle-type inequality.

We describe in Section 2 our estimation strategy. First, we present the contrast which is minimised in the following. Then we give the conditions on the spaces over which the contrast is minimised: this corresponds to model collections for which examples are provided. The procedure is completed by a model selection performed via a penalisation of the minimal contrast. Then mean integrated squared error (MISE) bound is given in Section 3 and illustrated by asymptotic rate over Besov spaces. Illustrations are provided in Section 4: examples are given, simulations results are provided and the method is applied to a classical real data set. All proofs are relegated in Section 5.

2. Estimation strategy

2.1. *Definition of the contrast*

Let Y be a nonnegative random variable and X a one-dimensional covariate. We assume that the joint density $f_{(X,Y)}$ of (X, Y) is such that

$$\bar{F}_1(x, y) = \int_y^{+\infty} f_{(X,Y)}(x, u)\,\mathrm{d}u \quad \text{and} \quad \bar{F}_2(x, y) = \int_y^{+\infty} \bar{F}_1(x, u)\,\mathrm{d}u$$

are measurable and finite nonnegative functions. Then, it is interesting to remark that

$$e(y|x) = \frac{\bar{F}_2(x, y)}{\bar{F}_1(x, y)} \quad \text{if } \bar{F}_1(x, y) > 0.$$

This holds by simplification by $f_X(x)$ in formula (2). Next, we consider two functions S and T such that $\iint S^2(x, y)\bar{F}_1(x, y)\,\mathrm{d}x\,\mathrm{d}y < +\infty$ and $\iint T^2(x, y)\bar{F}_1(x, y)\,\mathrm{d}x\,\mathrm{d}y < +\infty$. We define the μ-scalar product of S and T by

$$\langle S, T\rangle_\mu = \iint S(x, y)T(x, y)\,\mathrm{d}\mu(x, y) \quad \text{with } \mathrm{d}\mu(x, y) = \bar{F}_1(x, y)\,\mathrm{d}x\,\mathrm{d}y \tag{3}$$

and by $\|.\|_\mu$ the associated norm: $\|T\|_\mu^2 = \langle T, T\rangle_\mu$. This is meaningful as $\bar{F}_1(x, y) \geq 0$ and we will work on fields where $\bar{F}_1(x, y) > 0$.

Let $T : (x, y) \mapsto T(x, y)$ be a bivariate measurable compactly supported function, with support denoted by $A = A_1 \times A_2$. We propose to study the following contrast for estimating the conditional MRL $e(y|x)$:

$$\Gamma_n(T) = \frac{1}{n}\sum_{i=1}^{n}\left(\int T^2(X_i, y)\mathbf{1}_{\{Y_i \geq y\}}\,\mathrm{d}y - 2\Psi_T(X_i, Y_i)\right), \tag{4}$$

where

$$\Psi_T(x, y) = \int_0^y (y - u)T(x, u)\,\mathrm{d}u.$$

This contrast is justified by the following result.

Proposition 2.1 *Assume that T and e are μ-square integrable. Then, under the assumption:*

(A0) *For all* $x \in A_1$, $\lim_{y\to+\infty} y\bar{F}_1(x, y) = 0$,

we have: $\mathbb{E}(\Gamma_n(T)) = \|T - e\|_\mu^2 - \|e\|_\mu^2$.

Note that, since $\bar{F}_1(x, y) = \bar{F}(y|x)f_X(x)$, Assumption (A0) is easily satisfied when the cdf $y \mapsto F(y|x)$, for fixed x, belongs to exponential family laws (and has exponential rate of decay w.r.t. y).

Proposition 2.1 implies that minimising the contrast $\Gamma_n(T)$ over a large set of functions should mean minimising the empirical counterpart of $\|T - e\|_\mu^2$ and lead to find the function T which is the ‘nearest’ of e among a given class of functions.

2.2. *Extension to the censored case*

If the variable of interest Y is censored, we can generalise the contrast function. Let the observations be X_i, $Z_i = Y_i \wedge C_i$, $\delta_i = \mathbf{1}_{\{Y_i \le C_i\}}$ where C is the censoring random variable. Assume that the independence assumption holds:

$$C \text{ is independent of } (X, Y).$$

Then, let $\bar{G}_n$ be the modified Kaplan and Meier (1958) estimator for $\bar{G}$, the survival function of the censoring sequence (C_i), $\bar{G}(x) = \mathbb{P}(C \ge x)$, as given in Lo, Mack and Wang (1989). It is defined by

$$\bar{G}_n(x) = \begin{cases} \displaystyle\prod_{i=1, Z_{(i)} \le x}^{n} \left(\frac{n-i+1}{n-i+2}\right)^{1-\delta_{(i)}} & \text{if } x \le Z_{(n)} \\ \bar{G}_n(Z_{(n)}) & \text{if } x > Z_{(n)}. \end{cases} \tag{5}$$

This modification of the Kaplan–Meier estimator is proposed because the estimate of $\bar{G}$ appears in a denominator.

Then the contrast of interest has to be corrected by classical inverse probability of censoring weighting as follows:

$$\Gamma_n^C(T) = \frac{1}{n}\sum_{i=1}^{n}\left(\int T^2(X_i, y)\mathbf{1}_{\{Z_i \ge y\}}\,\mathrm{d}y - 2\frac{\delta_i}{\bar{G}_n(Z_i)}\Psi_T^G(X_i, Z_i)\right), \tag{6}$$

where

$$\Psi_T^G(x, y) = \int_0^y (y-u)T(x,u)\bar{G}(u)\,\mathrm{d}u.$$

To justify the proposed contrast, it is easy to check that, under the independence assumption and for $\bar{G}_n$ replaced by $\bar{G}$, the expectation of Γ_n^C is the same as the one of Γ_n, with μ replaced by μ^G given by

$$\mathrm{d}\mu^G(x, y) = \bar{G}(y)\bar{F}_1(x, y)\,\mathrm{d}x\,\mathrm{d}y.$$

We would get for an estimator defined as a minimiser of Γ_n^C as in Section 2.5, results of the same flavour as the one obtained for the minimiser of Γ_n, but we do not provide the associated theoretical study to avoid additional technicalities.

Remark 2.1 To avoid the loss of efficiency due to the weighting of uncensored survival times only, other censoring corrections can be used. In the spirit of the transformation proposed first by Leurgans (1987) in a linear regression context, we may also consider the following contrast:

$$\tilde{\Gamma}_n^C(T) = \frac{1}{n}\sum_{i=1}^{n}\left[\int T^2(X_i, y)\mathbf{1}_{\{Z_i \ge y\}}\,\mathrm{d}y - 2\int_0^{Z_i}\frac{1}{\bar{G}_n(w)}\left(\int_0^w \bar{G}_n(v)T(X_i, v)\,\mathrm{d}v\right)\mathrm{d}w\right]. \tag{7}$$

It also satisfies:

$$\mathbb{E}(\tilde{\Gamma}_n^C(T)) = \|T - e\|_{\mu^G}^2 - \|e\|_{\mu^G}^2. \tag{8}$$

This result is proved in Section 5.2.

In the censored case, the compact set of estimation in the y-direction, A_2 should be such that $A_2 \subsetneq [0, \inf(\tau_F, \tau_G)]$, where $\tau_F = \inf\{t, 1 - F(t) > 0\}$, $\tau_G = \inf\{t, 1 - G(t) > 0\}$.

2.3. *Assumptions and collections of linear spaces*

Let us mention first that we provide an estimator of e on a compact set only. This does not mean that we assume the function to be compactly supported (see the examples in Section 4) but that we estimate the function e restricted to a compact set only.

We denote this compact by $A = A_1 \times A_2$ and the collection of spaces are defined with respect to the compact set. We will use norms referring to this compact set, for $T \in (\mathbb{L}^2 \cap \mathbb{L}^\infty)(A)$:

$$\|T\|_A^2 = \iint_A T^2(x, y)\, \mathrm{d}x\, \mathrm{d}y, \quad \|T\|_{\infty,A} = \sup_{(x,y)\in A} |T(x, y)|.$$

Moreover, our assumptions are also related to the compact set:

(A1) There exist $\bar{F}_0$, $f_1 > 0$ such that $\forall (x, y) \in A_1 \times A_2$, $\bar{F}_1(x, y) \geq \bar{F}_0$ and $f_X(x) \leq f_1$.
(A2) $\forall (x, y) \in A_1 \times A_2, e(y|x) \leq \|e\|_{\infty,A} < +\infty$.

Assumptions (A1) and (A2) are weak because the bounds are required on a compact set only. As $\bar{F}_1(x, y)/f_X(x) = \bar{F}(y|x)$ is a conditional survival function, it is bounded by 1, thus $f_X(x) \leq f_1$ in Assumption (A1) implies $\bar{F}_1(x, y) \leq f_1$ for $(x, y) \in A$. Therefore Assumption (A1) implies that $\forall (x, y) \in A$, $\bar{F}_0 \leq \bar{F}_1(x, y) \leq f_1$, i.e. the reference measure of the problem here $\mathrm{d}\mu(x, y) = \bar{F}_1(x, y)\mathrm{d}x\, \mathrm{d}y$ is equivalent to the Lebesgue measure on A.

Now, we introduce a collection $\{S_m : m \in \mathcal{M}_n\}$ of projection spaces: S_m is called a model and $\mathcal{M}_n$ is a set of multi-indexes. For each m, the space S_m of functions with support in $A = A_1 \times A_2$ is defined by:

$$S_m = F_m \otimes \mathcal{H}_n = \left\{ h, \quad h(x, z) = \sum_{j\in J_m} \sum_{k\in\mathcal{K}_n} a_{j,k}\varphi_j^m(x)\psi_k(z),\ a_{j,k} \in \mathbb{R} \right\},$$

where F_m and $\mathcal{H}_n$ are subspaces of $(\mathbb{L}^2 \cap \mathbb{L}^\infty)(\mathbb{R})$, respectively, spanned by two orthonormal bases: $(\varphi_j^m)_{j\in J_m}$ with $|J_m| = D_m$, where D_m is varying and $(\psi_k)_{k\in\mathcal{K}_n}$ with $|\mathcal{K}_n| = \mathcal{D}_n^{(2)}$ is fixed. For all j and all k, the supports of φ_j^m and ψ_k are, respectively, included in A_1 and A_2.

For the sake of simplicity, we take $A_1 = [0, 1]$ and consider for F_m the space spanned by the following regular piecewise polynomial basis: φ_j^m for $j = (\ell, d)$ is a polynomial of degree $d \in \{0, \ldots, r\}$ (where r is fixed) on the interval $[(\ell - 1)/2^D, \ell/2^D[$ with $\ell \in \{1, \ldots, 2^D\}$. Here, indexes j are not integers, but pairs of integers. Thus, we have $m = (D, r)$, $J_m = \{j = (\ell, d) \in \mathbb{N} \times \mathbb{N},\ 1 \leq \ell \leq 2^D, 0 \leq d \leq r\}$, $D_m = (r + 1)2^D$. The collection $(F_m)_m$ is nested in the sense that $D_m \leq D_{m'} \Rightarrow F_m \subset F_{m'}$ and fulfils

$$\forall x \in A_1, \sum_{j\in J_m} (\varphi_j^m(x))^2 \leq \phi_1 D_m, \tag{9}$$

with $\phi_1 = \sqrt{r + 1}$. We denote by $\mathcal{F}_n$ the largest space of the collection with maximal dimension denoted by $\mathcal{D}_n^{(1)} \leq \sqrt{n}/\log(n)$.

We take $\mathcal{H}_n = \mathcal{F}_n$ so that $\dim(\mathcal{H}_n) = \mathcal{D}_n^{(2)} \leq \sqrt{n}/\log(n)$. But, as it is possible to take a space generated by a different basis, we keep distinct notations.

The collection $S_m = F_m \otimes \mathcal{H}_n$ is nested and included in $\mathcal{S}_n = \mathcal{F}_n \otimes \mathcal{H}_n$ with dimension $N_n = \mathcal{D}_n^{(1)}\mathcal{D}_n^{(2)} \leq n/\log^2(n)$.

Other bases like trigonometric polynomials (see Barron, Birgé and Massart 1999) or wavelet bases (see Cohen, Daubechies and Vial 1993) can be used. Note that the histogram basis is a particular case of the polynomial basis with $r = 0$. Practical computations with histograms are easier than with any other basis, which explains that it is used in Section 4.

Remark 2.2 From a theoretical point of view, we may consider that the covariate X belongs to $\mathbb{R}^d$ and consider models of the form $S_m = F_{m_1} \otimes \cdots F_{m_d} \otimes \mathcal{H}_n$. The convergence rate of the estimator would be slower because of the curse of dimensionality.

2.4. *Minimising the contrast*

The first step is to study the minimisation of $\Gamma_n(T)$ over S_m. To that end, let $T(x, y) = \sum_{j\in J_m} \sum_{k\in\mathcal{K}_n} a_{j,k}\varphi_j^m(x)\psi_k(y)$ be a function in S_m. To compute $\hat{e}_m$, we have to solve:

$$\forall j_0 \in J_m, \quad \forall k_0 \in \mathcal{K}_n, \quad \frac{\partial\Gamma_n(T)}{\partial a_{j_0,k_0}} = 0$$

or equivalently for all $j_0 \in J_m$, $k_0 \in \mathcal{K}_n$,

$$\sum_{j\in J_m}\sum_{k\in\mathcal{K}_n} a_{j,k}\frac{1}{n}\sum_{i=1}^n \varphi_j^m(X_i)\varphi_{j_0}^m(X_i)\int \psi_k(z)\psi_{k_0}(z)\mathbf{1}_{\{Y_i\geq z\}}\,\mathrm{d}z$$
$$= \frac{1}{n}\sum_{i=1}^n \varphi_{j_0}^m(X_i)\int_0^{Y_i}(Y_i - u)\psi_{k_0}(u)\,\mathrm{d}u.$$

Let $\mathrm{vec}(\cdot)$ denote the operator that stacks the columns of a matrix into a vector. The above equation can be summarised by

$$G_m\hat{A}_m = \Upsilon_m,$$

where $\hat{A}_m$ denotes the vector $\mathrm{vec}((\hat{a}_{j,k})_{j\in J_m,k\in\mathcal{K}_n})$ of the coefficients of the development of the estimator in the basis,

$$G_m := \left(\frac{1}{n}\sum_{i=1}^n \varphi_j^m(X_i)\varphi_\ell^m(X_i)\int \psi_k(z)\psi_p(z)\mathbf{1}_{\{Y_i\geq z\}}\,\mathrm{d}z\right)_{(j,k),(\ell,p)\in(J_m\times\mathcal{K}_n)^2}$$

and

$$\Upsilon_m := \mathrm{vec}\left(\left(\frac{1}{n}\sum_{i=1}^n \varphi_j^m(X_i)\int_0^{Y_i}(Y_i - u)\psi_k(u)\,\mathrm{d}u\right)_{j\in J_m,k\in\mathcal{K}_n}\right).$$

The matrix G_m is a square matrix with size $|J_m||\mathcal{K}_n| \times |J_m||\mathcal{K}_n|$. It has nonnegative eigenvalues. Indeed, if $u = \mathrm{vec}((u_{j,k})_{j\in J_m,k\in\mathcal{K}_n})$ is a vector and $u^\top$ denotes its transpose, then

$$u^\top G_m u = \frac{1}{n}\sum_{i=1}^n \int\left(\sum_{j,k} u_{j,k}\varphi_j^m(X_i)\psi_k(z)\right)^2 \mathbf{1}_{\{Y_i\geq z\}}\,\mathrm{d}z \geq 0.$$

We denote by $\mathrm{Sp}(G_m)$ the spectrum of G_m, i.e. the set of the nonnegative eigenvalues of the matrix G_m. Lastly, the matrix G_m can also be written

$$G_m = \frac{1}{n}\sum_{i=1}^n \Phi_m^{(i)} \otimes \Psi_m^{(i)},$$

where $\Phi_m^{(i)} \otimes \Psi_m^{(i)}$ is the tensorial product of two square matrices $\Phi_m^{(i)} := (\varphi_j^m(X_i)\varphi_{j_0}^m(X_i))_{(j,j_0)\in J_m^2}$ and $\Psi_m^{(i)} := (\int \psi_k(z)\psi_{k_0}(z)\mathbf{1}_{\{Y_i\geq z\}}\,\mathrm{d}z)_{(k,k_0)\in\mathcal{K}_n^2}$. For the practical implementation of the estimator, we need to compute the inverse of the matrix G_m.

2.5. *Definition of the estimator*

As G_m may be non invertible, we modify the definition of $\hat{e}_m$ in the following way:

$$\hat{e}_m := \begin{cases} \arg\min_{T\in S_m} \Gamma_n(T) & \text{on } \hat{H}_m \\ 0 & \text{on } \hat{H}_m^c \end{cases}, \tag{10}$$

where

$$\hat{H}_m := \left\{ \min \mathrm{Sp}(G_m) \geq \max\left(\frac{\hat{\bar{F}}_0}{3}, n^{-1/2}\right) \right\}. \tag{11}$$

The quantity $\hat{\bar{F}}_0$ is an estimator of the bound $\bar{F}_0$ (the minimum of $\bar{F}_1$ on A; see Assumption (A1)). We require that it fulfils the following assumption.

(A3) There exists a constant C such that $\mathbb{P}(|\hat{\bar{F}}_0 - \bar{F}_0| > \bar{F}_0/2) \leq C/n^5$.

An estimator $\hat{\bar{F}}_0$ satisfying Assumption (A3) is defined in Comte, Gaïffas and Guilloux (in press). For the sake of completeness, we recall here its definition. Let

$$\hat{\bar{F}}_m(x, y) = \sum_{j\in J_m, k\in\mathcal{K}_n} \hat{b}_{j,k}\varphi_j^m(x)\psi_k(y), \quad \text{with } \hat{b}_{j,k} = \frac{1}{n}\sum_{i=1}^{n} \varphi_j^m(X_i)\int \psi_k(y)\mathbf{1}_{\{Y_i\geq y\}}\,\mathrm{d}y.$$

It is easy to see that $\mathbb{E}(\hat{b}_{j,k}) = \iint \varphi_j^m(x)\psi_k(y)\bar{F}_1(x, y)\,\mathrm{d}x\,\mathrm{d}y = \langle\varphi_j^m \otimes \psi_k, \bar{F}_1\rangle$, so that $\hat{\bar{F}}_m(x, y)$ is the natural projection estimator of $\bar{F}_1$. Then take

$$\hat{\bar{F}}_0 = \inf_{(x,y)\in A} \hat{\bar{F}}_{m^*}(x, y) \tag{12}$$

where m^* is chosen such that $\log(n) \leq D_{m^*} \leq n^{1/4}/\sqrt{\log(n)}$, and $\mathcal{D}_n^{(2)} = \mathrm{Int}(n^{1/4}/\sqrt{\log(n)})$, where $\mathrm{Int}(x)$ denotes the integer part of x.

Then Proposition 1 in Comte et al. (in press) gives the conditions on $\bar{F}_1$ under which Assumption (A3) is fulfilled.

The final step is to select the relevant space via the penalised criterion. Here, only one direction requires model selection, namely the x-direction. Indeed, the y-direction keeps the good properties of empirical estimators provided that the y-space is simply chosen as large as possible, as it would be done if no covariate was involved. Therefore, we select a model $\hat{m}$ defined by

$$\hat{m} = \arg\min_{m\in\mathcal{M}_n} (\Gamma_n(\hat{e}_m) + \mathrm{pen}(m)), \tag{13}$$

where $\mathrm{pen}(m)$ is defined in Theorem 3.1. Our estimator of e on A is then $\tilde{e} = \hat{e}_{\hat{m}}$.

3. Oracle inequality and rate of convergence

3.1. *MISE bound*

We can prove an oracle-type inequality under the following assumption:

(A4) $Y_1, \ldots, Y_n$ are $\mathbb{R}^+$-supported and $\mathbb{E}(Y_1^6) < +\infty$.

Moreover, we denote by e_m the $\mathbb{L}^2$-orthogonal projection on S_m of e restricted to A and by $\ell(A_2)$ the length of the interval A_2. Our main theorem is the following.

THEOREM 3.1 *Assume that* Assumptions (A0)–(A4) *hold. Then there exists a numerical constant* κ *such that, for*

$$\text{pen}(m) = \kappa\phi_1 \frac{\mathbb{E}(Y_1^3) + \ell(A_2)\mathbb{E}(Y_1^2)}{\bar{F}_0}\frac{D_m}{n}, \tag{14}$$

the estimator $\tilde{e} = \hat{e}_{\hat{m}}$ *with* $\hat{m}$ *defined by Equation* (13) *satisfies:*

$$\mathbb{E}(\|\tilde{e} - e\|_A^2) \leq C \inf_m (\|e - e_m\|_A^2 + \text{pen}(m)) + \frac{C'}{n}, \tag{15}$$

where C *is a constant depending on* $\bar{F}_0$ *and* C' *is a constant depending on* $\mathbb{E}(Y_1^6)$, $\bar{F}_0$, $\|e\|_{\infty,A}$.

Inequality (15) shows that the estimator automatically makes the compromise between the square bias $\|e - e_m\|_A^2$ and the variance term which is proportional to the order of the penalty D_m/n. It is worth mentioning that this term does not depend on $\mathcal{D}_n^{(2)}$. This is why only the x-dimension D_m has to be selected whereas the y-dimension $\mathcal{D}_n^{(2)}$ is just taken as large as possible, in order to get smaller bias.

We emphasise that the constant terms in the penalty do not have the same status. The constant κ is numerical and does not depend on any unknown quantity. It is universal in the sense that it is not affected by the sampling changes and it can be calibrated over a wide range of models by simulation experiments. The constant ϕ_1 is known when the basis is chosen.

On the other hand, the unknown quantities $\mathbb{E}(Y_1^2)$ and $\mathbb{E}(Y_1^3)$ must be estimated by empirical moments, and $\bar{F}_0$ must be replaced by $\hat{\bar{F}}_0$. The effective penalty is thus random. For an example of theoretical study of such random penalty, see Comte et al. (in press). The results of Theorem 3.1 can be generalised with an estimated penalty, but in an asymptotic setting.

Remark 3.1 If we work with the assumption that there exists a positive constant $B < +\infty$ such that $0 < Y_1 \leq B$ a.s., which is an assumption stronger than Assumption (A4), the proof of the result (15) is simpler, and we can take the penalty

$$\text{pen}_B(m) = \kappa'\phi_1 \frac{B^3}{\bar{F}_0}\frac{D_m}{n}. \tag{16}$$

Here, the bound B is unknown. Nevertheless, it is possible to estimate it by $\hat{B} = \max_{1\leq i\leq n} Y_i$. In this case, the numerical choice associated to κ' will not be the same as the choice for κ.

3.2. *Rates of convergence on Besov spaces*

We can deduce from Theorem 3.1 the order of the risk and the rate of convergence of the estimator. For that purpose, assume that e restricted to A belongs to the anisotropic Besov space $\mathcal{B}_{2,\infty}^{\boldsymbol{\alpha}}(A)$ on A with regularity $\boldsymbol{\alpha} = (\alpha_1, \alpha_2)$ (for details on Besov spaces, see DeVore and Lorentz (1993)). Roughly speaking, this means that e has regularity of order α_1 in the x-direction and of order α_2 in the y-direction.

COROLLARY 3.1 *Assume that* e *restricted to* A *belongs to the anisotropic Besov space* $B_{2,\infty}^{\boldsymbol{\alpha}}(A)$ *with regularity* $\boldsymbol{\alpha} = (\alpha_1, \alpha_2)$ *such that* $\alpha_1 > 1/2$ *and* $\alpha_2 > 1$. *We consider the spaces of piecewise* *polynomials described in Subsection* 2.3 *with* $r > \alpha_i - 1$, $i = 1, 2$. *Then*, *under the assumptions of Theorem* 3.1,

$$\mathbb{E}(\|e\mathbf{1}_A - \tilde{e}\|^2) = O(n^{-2\alpha_1/(2\alpha_1+1)}).$$

The proof of Corollary 3.1 is standard and thus omitted (see Brunel, Comte and Lacour 2010).

It is noteworthy that we obtain a rate of convergence which would be standard for the estimation of a function of *one variable* with regularity α_1.

4. Examples and illustrations

We give numerical illustrations for some classical regression models used in lifetime analysis. The description and the parametric inference of these models are detailed in Chapter 6 of Lawless (2003). It is easy to check that all our examples satisfy Assumption (A0).

4.1. *Accelerated failure time model*

Let $\sigma > 0$ and $\mu : \mathbb{R} \mapsto \mathbb{R}$ and consider the model:

$$\ln(Y) = \mu(X) + \sigma\varepsilon \quad \text{and} \quad \varepsilon \text{ independent of } X.$$

Then we have

$$\begin{aligned}\mathbb{P}(Y > y|X = x) = \mathbb{P}(\ln(Y) > \ln(y)|X = x) &= \mathbb{P}\left(\varepsilon > \frac{\ln(y) - \mu(x)}{\sigma}|X = x\right)\\ &= \bar{F}_\varepsilon\left(\frac{\ln(y) - \mu(x)}{\sigma}\right),\end{aligned}$$

where $\bar{F}_\varepsilon$ stands for the survival function of the noise ε. Therefore, we can write

$$e(y|x) = \int_y^{+\infty} \frac{\bar{F}_\varepsilon((\ln(u) - \mu(x))/\sigma)\,\mathrm{d}u}{\bar{F}_\varepsilon((\ln(y) - \mu(x))/\sigma)}.$$

- *Example 1*: Take $\bar{F}_\varepsilon(x) = \exp(-\exp(x))$ and $\sigma = 1$. Then $e_1(y|x) = \mathrm{e}^{\mu(x)}$, the conditional expectation does not depend on y.
- *Example 2*: Take $\bar{F}_\varepsilon(x) = \exp(-\exp(x))$ and $\sigma = 2$. Then

$$\begin{aligned}&\int_y^{+\infty} \exp(-\sqrt{u}\mathrm{e}^{-\mu(x)/2})\,\mathrm{d}u\\ &\quad = 2\mathrm{e}^{\mu(x)/2}\int_{\sqrt{y}\mathrm{e}^{-\mu(x)/2}}^{+\infty} \mathrm{e}^{-v}v\,\mathrm{d}v = 2\mathrm{e}^{\mu(x)}(1 + \sqrt{y}\mathrm{e}^{-\mu(x)/2})\exp(-\sqrt{y}\mathrm{e}^{-\mu(x)/2}).\end{aligned}$$

Thus,

$$e_2(y|x) = 2\mathrm{e}^{\mu(x)}(1 + \sqrt{y}\mathrm{e}^{-\mu(x)/2}). \tag{17}$$

- *Example 3*: Take $\bar{F}_\varepsilon(x) = (1 + \exp(x))^{-1}$ and $\sigma = 1/2$. Then $\bar{F}(y|x) = 1/(1 + y^2\mathrm{e}^{-2\mu(x)})$ and

$$\int_y^{+\infty} \frac{\mathrm{d}u}{1 + u^2\mathrm{e}^{-2\mu(x)}} = \mathrm{e}^{\mu(x)}\int_{y\mathrm{e}^{-\mu(x)}}^{+\infty} \frac{\mathrm{d}v}{1 + v^2} = \mathrm{e}^{\mu(x)}\arctan(y^{-1}\mathrm{e}^{\mu(x)}).$$

This yields

$$e_3(y|x) = (\mathrm{e}^{\mu(x)} + y^2\mathrm{e}^{-\mu(x)}) \arctan(y^{-1}\mathrm{e}^{\mu(x)}). \tag{18}$$

We shall take affine functions $\mu(x)$.

4.2. *Generalised Cox model*

The standard Cox model assumes that the conditional hazard rate α can be decomposed in the following multiplicative way: $\alpha(y|x) = \exp(\beta x)\alpha_0(y)$. It was generalised by Castellan and Letué (2000) for nonparametric estimation purpose by the general equation

$$\alpha(y|x) = \exp(\mu(x))\alpha_0(y).$$

As conditional cumulative hazard denoted by $A(y|x)$ is related to conditional survival function $\bar{F}(y|x)$ by: $A(y|x) = \int_0^y \alpha(u|x)\,\mathrm{d}u = -\ln(\bar{F}(y|x))$, we have

$$\bar{F}(y|x) = \exp\left(-\int_0^y \alpha(u|x)\,\mathrm{d}u\right).$$

Let us denote $A_0(y) = \int_0^y \alpha_0(u)\,\mathrm{d}u$. We find

$$e(y|x) = \int_y^{+\infty} \exp(-\mathrm{e}^{\mu(x)}(A_0(v) - A_0(y)))\,\mathrm{d}v.$$

It is worth noting that for $A_0(y) = \lambda y$, that is, constant hazard $\alpha_0(y) = \lambda$, this model gives $e(y|x) = e^{-\mu(x)}/\lambda$ which is the same model as the first accelerated failure time model above.

- *Example 4*: We can consider the case $\alpha_0(y) = \lambda y$ and $A_0(y) = \lambda y^2/2$. Let $\Phi(u) = \int_{-\infty}^u \exp(-v^2/2)\,\mathrm{d}v$. Then we find

$$e_4(y|x) = \frac{1}{\sqrt{\lambda}} \exp\left[\frac{1}{2}(\lambda y^2\mathrm{e}^{\mu(x)} - \mu(x))\right](1 - \Phi(\sqrt{\lambda}y\mathrm{e}^{\mu(x)/2})). \tag{19}$$

4.3. *Additive hazards models*

Additive hazards models are sometimes useful and are defined, with the same notations as in Section 4.2 by: $\alpha(y|x) = \alpha_0(y) + \exp(f(x))$, $f \geq 0$. Simple calculations give:

$$\bar{F}(y|x) = \exp\left(-\int_0^y \alpha(u|x)\,\mathrm{d}u\right) = \exp[-A_0(y) - yf(x)].$$

Then, we find $e(y|x) = \int_y^{+\infty} \exp[A_0(y) - A_0(v) + f(x)(y - v)]\,\mathrm{d}v$.

- *Example 5*: If we take an exponential baseline hazard with parameter λ, $A_0(y) = \lambda y$, we get:

$$e_5(y|x) = \int_y^{+\infty} \exp[(\lambda + f(x))(y - v)]\,\mathrm{d}v = \frac{1}{\lambda + f(x)}. \tag{20}$$

4.4. *Monte-Carlo study*

We study the numerical performances of our penalised estimator by generating samples $(X_i, Y_i)_{i=1}^n$ following the models described in the previous sections.

- *Example 1*: $e_1(y|x) = \mathrm{e}^{\mu(x)}$ with $\mu(x) = ax + b$ with $a = 2$, $b = -2$ and $X \sim \mathcal{U}([0, 1])$. We take $A_1 \times A_2 = [0.05, 0.9] \times [0.015, 1.16]$.
- *Example 2*: $e_2(y|x)$ given by Equation (17) with $\mu(x) = ax + b$ with $a = 1$, $b = -2$ and $X \sim \mathcal{U}([0, 1])$. We take $A_1 \times A_2 = [0.05, 0.9] \times A_2 = [1.10^{-3}, 1.4]$.
- *Example 3*: $e_3(y|x)$ given by Equation (18) with $\mu(x) = ax + b$ with $a = 0.5$, $b = -2$ and $X \sim \chi^2(8)/16$. We take $A_1 \times A_2 = [0.16, 0.8] \times [0.04, 0.5]$.
- *Example 4*: $e_4(y|x)$ given by Equation (19) with $\lambda = 2$ $\mu(x) = ax$ with $a = 5$ and $X \sim \chi^2(8)/16$. We take $A_1 \times A_2 = [0.16, 0.8] \times A_2 = [0.04, 0.6]$.
- *Example 5*: $e_5(y|x)$ given by Equation (20) with $f(x) = x^5$, $\lambda = 0.8$ and $X \sim \mathcal{U}([0, 1])$. We take $A_1 \times A_2 = [0.05, 0.9] \times [0.05, 2.4]$.

The sets $A = A_1 \times A_2$ are fixed intervals, roughly calibrated with respect to each distribution. In practice, we would have chosen the compact sets of estimation with respect to the data (and their extreme values). Here, we fixed them for reproducibility of the experiments and in order to have the same set of estimation for all paths. Clearly, this is possible in a simulation setting only. We illustrate the practical implementation of our estimator for dyadic histogram bases.

We compute the empirical MISE over $N = 500$ replications of samples of size 100 and 500 and $N = 200$ replications for samples of size 1000, by averaging over the paths $j = 1, \ldots, N$, the quantities

$$\frac{\ell(A_1)\ell(A_2)}{K^2} \sum_{k,\ell=1}^{K} (\tilde{e}^{(j)}(y_\ell|x_k) - e(y_\ell|x_k))^2, \tag{21}$$

where $\ell(A_i)$ is the length of the interval A_i, $i = 1, 2$, $(x_k)_{1\le k\le K}$ and $(y_k)_{1\le k\le K}$ are uniform subdivisions of A_1 and A_2, respectively, and $\tilde{e}^{(j)}$ is the estimator associated to the jth sample path. Note that the computed error given by Equation (21) is the empirical version of the $\mathbb{L}^2$-risk

$$\mathbb{E}\left(\iint_{A_1\times A_2} (\tilde{e}(y_\ell|x_k) - e(y_\ell|x_k))^2 \, \mathrm{d}x_k \, \mathrm{d}y_\ell\right),$$

which corresponds to integrated errors in both x- and y-directions.

We compute the values of penalised contrasts for the different models, and in all cases, the algorithm selects the x-dimension $D_m^{(1)}$ less than $\sqrt{n}/\log(n)$ while the y-dimension $\mathcal{D}_n^{(2)}$ is fixed to the maximal value $\sqrt{n}/\log(n)$.

Here, we need to give some explanations on the choice of the constants involved in the penalty. The penalty term is chosen as follows:

$$C\left(\max_{1\le i\le n}(Y_i)\right)^3 \frac{D_{m_1}}{n}. \tag{22}$$

It corresponds to the empirical version of Equation (16) where $C = \kappa\phi_1/\bar{F}_0$ and the constant κ has to be calibrated. Here, $\phi_1 = 1$ (histogram basis) so that the constant reduces to $\kappa/\bar{F}_0$. We estimate $\bar{F}_0$ by $\hat{\bar{F}}_0$ as defined by Equation (12) in Section 2.5 and fix the dimensions $D_{m^*} = 2^{m^*}$,

Table 1. Empirical MISE in Examples 1–5 for estimators with penalty (16), standard deviations of the MISE given in parenthesis and mean value of the constant $C_1 = \kappa/\hat{\hat{F}}_0$.

	$n = 100$	$n = 500$	$n = 1000$
Example 1	0.0246 (0.0244)	0.0175 (0.0099)	0.0146 (0.0037)
C_1	9.98	48.49	49.19
Example 2	0.1637 (0.344)	0.1091 (0.0645)	0.0809 (0.0542)
C_1	9.45	16.55	17.85
Example 3	0.0229 (0.0854)	0.0105 (0.1184)	0.0039 (0.0533)
C_1	9.57	8.92	8.43
Example 4	0.0023 (0.0005)	0.0016 (0.0004)	0.0009 (0.0004)
C_1	9.97	48.32	49.7
Example 5	0.2640 (0.2048)	0.1190 (0.0444)	0.0890 (0.0449)
C_1	9.43	16.40	13.19

$\mathcal{D}_n^{(2)} = 2^{m_2}$ with $m^* = m_2 = \mathrm{Int}[0.25\ln(n)/\ln(2)] = 2$ (for $n = 500, 1000$), in accordance with the conditions given on the dimensions in Section 2.5. We take $\kappa = 1$ and we bound the value of $\hat{\hat{F}}_0$ in case it achieves too small values by taking for constant C, $C_1 = \min(1/\hat{\hat{F}}_0; n/10)$ for $n = 100$ or $n = 500$ and $C_1 = \min(1/\hat{\hat{F}}_0; n/50)$ for $n = 1000$. Indeed, the value of C_1 has to be upper bounded in practice to avoid over-penalisation. The results are reported in Table 1: they indicate that for all models considered here, the value of C_1 has the adequate penalisation effect. The line where C_1 is given in Table 1 provides the mean value obtained for this quantity over the replications: we can see that its behaviour can be quite different depending on the model. For Examples 1 and 4, the value of $1/\hat{\hat{F}}_0$ is often over the maximal bound $n/10$ or $n/50$: we can see in Table 1 that the mean value of C_1 in these cases is near to the maximal value. Note that the values of the average MISEs given in Table 1 are not completely satisfactory in regard to their associated standard deviations, in particular for a small sample with size $n = 100$. This is indeed small in the context of (bivariate) estimation. We can see in Table 1 that for larger sample sizes $n = 500$ and $n = 1000$, we obtain better values.

Another calibration question we address is the choice between penalties given by Equations (16) and (14). Since our real data study involves a small sample with size of order 100, we compare the two formulae for sample size $n = 100$. The results given in the first two columns of Table 2 report the MISE values for penalty (16) with constant C taken as fixed at $C_0 = 10$ first, and then with constant C_1 computed as described before. In our study, we also consider the choice (14). The empirical version used in practice is:

$$C'(\bar{Y^3} + \ell(A_2)\bar{Y^2})\frac{D_{m_1}}{n}.$$

Table 2. Empirical MISE in Examples 1–5 for $n = 100$ and different penalisations (500 replications) with standard deviations given in parenthesis.

	Fixed C_0 penalty (16)	Estimated C_1 penalty (16)	Estimated C_2 penalty (14)
Example 1	0.0372 (0.0249)	0.0246 (0.0244)	0.0167 (0.0126)
	$C_0 = 10$	$C_1 = 9.98$	$C_2 = 99.74$
Example 2	0.198 (0.344)	0.1637 (0.344)	0.1322 (0.2560)
	$C_0 = 10$	$C_1 = 9.45$	$C_2 = 90.67$
Example 3	0.0274 (0.0854)	0.0229 (0.0854)	0.0104 (0.0295)
	$C_0 = 10$	$C_1 = 9.57$	$C_2 = 79.84$
Example 4	0.0045 (0.0018)	0.0023 (0.0005)	0.0012 (0.0003)
	$C_0 = 10$	$C_1 = 9.97$	$C_2 = 99.27$
Example 5	0.3203 (0.2691)	0.2640 (0.2048)	0.2958 (0.2304)
	$C_0 = 10$	$C_1 = 9.43$	$C_2 = 96.7$

with $\bar{Y}^p = (1/n)\sum_{i=1}^{n} Y_i^p$ for $p = 2, 3$ and $C' = \kappa'\phi_1/\bar{F}_0$. Here, we fix $\kappa' = 10$, we still have $\phi_1 = 1$ and we take for the constant C' the choice $C_2 = 10\min(1/\hat{\bar{F}}_0; n/10)$. The results of this experiment are reported in Table 2. We can conclude from the table that the MISE values with their standard deviations are improved with the last proposal. But this may not be true for larger sample size. This explains nevertheless why we choose the penalisation given by Equation (14) with constant C_2 for the next section.

On the whole, the automatic selection works well. We provide in Figure 1 a view of typical estimates and typical improvement between sizes $n = 500$ and $n = 2000$. It has been noticed by Abdous and Berred (2005) that the estimation of MRL is anyway difficult, in particular for

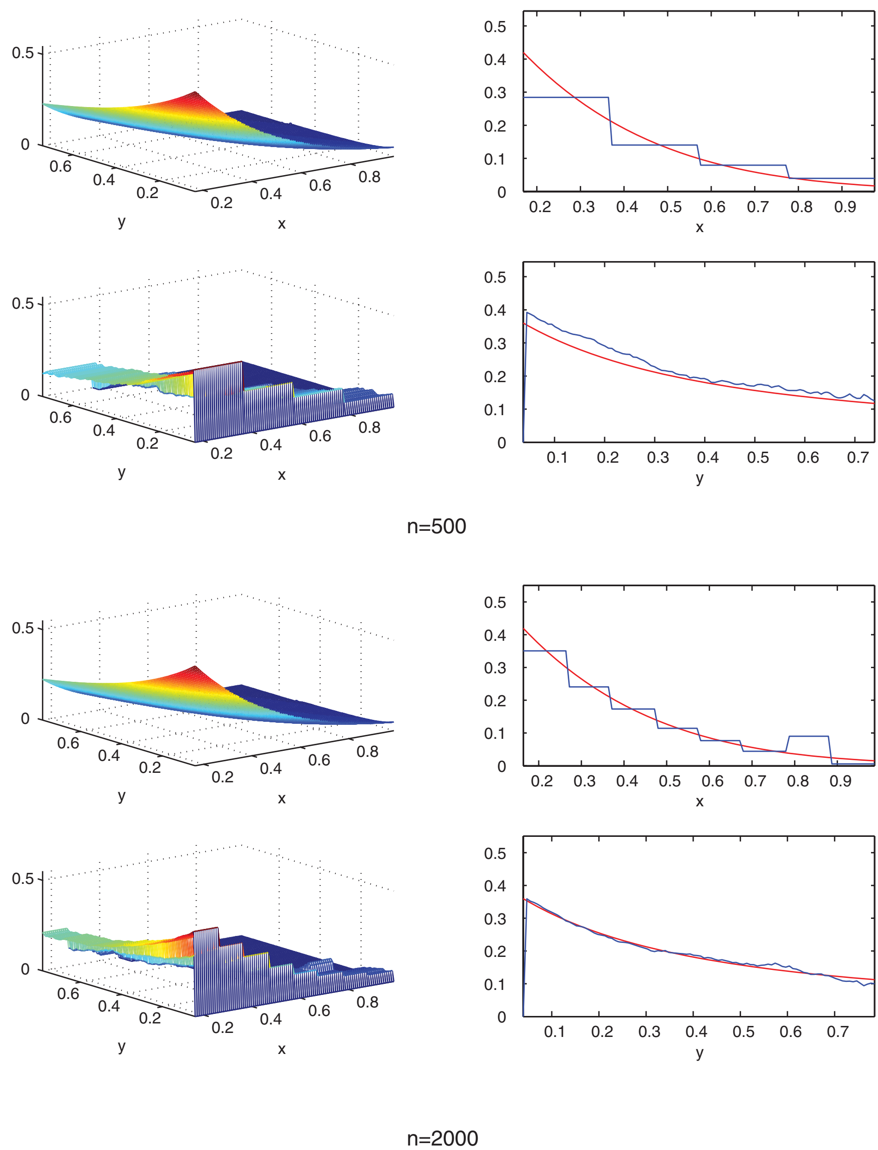

Figure 1. Penalised estimator for Example 4. Top left: true conditional MRL e; bottom left: estimator $\tilde{e}$; top right: $x \mapsto e(y|x)$ and $x \mapsto \tilde{e}(y|x)$ for a fixed value of y; bottom right: $y \mapsto e(y|x)$ and $y \mapsto \tilde{e}(y|x)$ for a fixed value of x.

small samples, even in the absence of a covariate. This is the reason why we do not start our representations to $n = 100$ in a multivariate setting.

4.5. *Application to a real data set*

As an illustration, we consider data from the Veterans Administration Lung Cancer Trial presented by Prentice (1973), in which males with advanced inoperable lung cancer received chemotherapy. Several covariates were observed for each patient. Hereafter, we study the survival times of 128 uncensored patients which range from 1 to 587 and we focus on the patients' age as a covariate. These data were also studied in Chen and Cheng (2005) but only for a subgroup of 97 patients with no prior therapy, and for two other covariates: a categorical covariate which gives the tumor type, which we cannot handle with our method and the performance status which is a therapeutic score ranging from 0 to 100 and evaluating the autonomy of the patient. Figure 2 (top left) plots the MRL estimate as a bivariate function of the survival time and the patient's age. In view of the simulation study, for such small sample sizes, respectively, $n = 128$ and $n = 97$, we recommend to use the penalty choice (14) with $C_2 = 10\min(\hat{\hat{F}}_0; n/10)$. The curve shows three parts with respect to the age, with change points corresponding somehow to early, middle and old ages. This behaviour also appears on Figure 2 (bottom left) where the estimated MRL is shown for three fixed values of ages 41, 54 and 68. While the MRL is quite the same for ages 41 and 68, it is of interest that the MRL is twice better for middle age 54. We could think that there is an optimal age (the middle age)

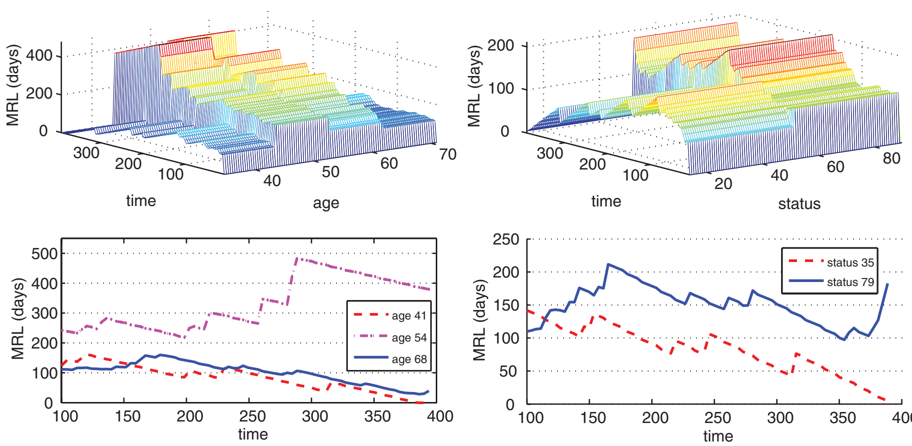

Figure 2. Estimated MRL as a function of the survival time and the age (top left) and as a function of the survival time and the status (top right). At bottom left, estimated MRL for given ages 41, 54 and 68 and at bottom right for fixed status 35 and 79.

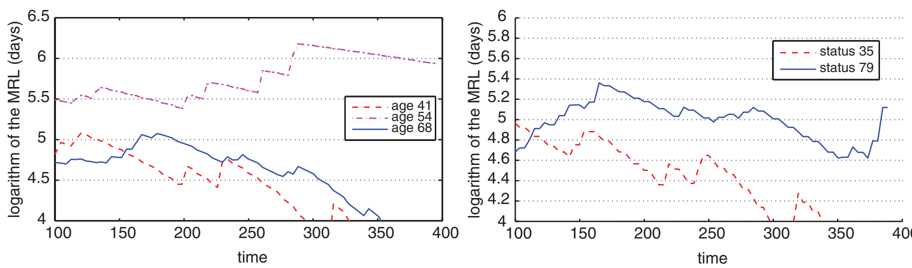

Figure 3. Log-scale for estimated MRL with age (left) and status (right).

to receive the treatment. We also estimate the MRL conditionally to the performance status as a continuous covariate with the subgroup of 97 patients with no prior therapy. The estimated MRL is shown in Figure 2 (right). For the performance status, we detect as for the age two parts in the curve corresponding to low and high status. Moreover, our purely nonparametric estimator gives a new graphical method to check whether the proportionality assumption is true or not. In fact, the Oakes–Dasu model states that $e(y|x) = e_0(x)\exp(bx)$, where b is a real parameter and $e_0(\cdot)$ is the baseline MRL. Then, in log-scale, the curves in Figure 3 should be parallel to each other for any fixed value of the covariate. But, it seems that for the covariate age, the Oakes–Dasu model cannot be reasonably considered. Besides, for the covariate performance status, as mentioned in the work by Chen and Cheng (2005), after an initial period from 0 to 150 days, the curves corresponding, respectively, to statuses 35 and 79 appear quite parallel, which may suggest the adequacy of the proportionality assumption.

5. Proofs

5.1. *Proof of Proposition* 2.1

Let us compute

$$\mathbb{E}(\Gamma_n(T)) = \mathbb{E}\left(\int T^2(X_1, y)\mathbf{1}_{\{Y_1\geq y\}}\,\mathrm{d}y - 2\Psi_T(X_1, Y_1)\right). \tag{23}$$

First, the Fubini–Tonelli theorem implies that

$$\begin{aligned}\mathbb{E}\left(\int T^2(X_1, y)\mathbf{1}_{\{Y_1\geq y\}}\,\mathrm{d}y\right) &= \iint\left(\int T^2(x, y)\mathbf{1}_{\{u\geq y\}}\,\mathrm{d}y\right) f_{(X,Y)}(x, u)\,\mathrm{d}x\,\mathrm{d}u\\ &= \iint T^2(x, y)\bar{F}_1(x, y)\,\mathrm{d}x\,\mathrm{d}y,\end{aligned} \tag{24}$$

and the last term is finite as T is μ-square integrable. Secondly, an integration by part yields

$$\int_u^{+\infty}(y-u)f_{(X,Y)}(x, y)\,\mathrm{d}y = [-(y-u)\bar{F}_1(x, y)]_{y=u}^{y=+\infty} + \int_u^{+\infty}\bar{F}_1(x, y)\,\mathrm{d}y = \bar{F}_2(x, u)$$

under the condition (A0). Therefore, with the Fubini–Tonelli theorem first and next the Cauchy–Schwarz inequality, we have

$$\begin{aligned}&\iint\int_0^y (y-u)|T(x, u)|f_{(X,Y)}(x, y)\,\mathrm{d}u\,\mathrm{d}x\,\mathrm{d}y\\ &\quad= \iint |T(x, u)|\bar{F}_2(x, u)\,\mathrm{d}u\,\mathrm{d}x\\ &\quad= \iint |T(x, u)|e(u|x)\bar{F}_1(x, u)\,\mathrm{d}u\,\mathrm{d}x\\ &\quad\leq \left(\iint T^2(x, u)\bar{F}_1(x, u)\,\mathrm{d}u\,\mathrm{d}x\iint e^2(u|x)\bar{F}_1(x, u)\,\mathrm{d}u\,\mathrm{d}x\right)^{1/2}.\end{aligned}$$

The last bound is finite since T and e are μ-square integrable. Thus, the Fubini theorem can be applied to get:

$$\begin{aligned}
\mathbb{E}(\Psi_T(X_1, Y_1)) &= \iint \int_0^y (y-u)T(x,u)\,\mathrm{d}u f_{(X,Y)}(x,y)\,\mathrm{d}x\,\mathrm{d}y \\
&= \iint \left(\int \mathbf{1}_{\{u\le y\}}(y-u) f_{(X,Y)}(x,y)\,\mathrm{d}y \right) T(x,u)\,\mathrm{d}x\,\mathrm{d}u \\
&= \iint T(x,u)\bar{F}_2(x,u)\,\mathrm{d}x\,\mathrm{d}u. \qquad (25)
\end{aligned}$$

Now, combining Equations (23)–(25) yields

$$\mathbb{E}(\Gamma_n(T)) = \iint [T^2(x,y) - 2T(x,y)e(y|x)]\bar{F}_1(x,y)\,\mathrm{d}x\,\mathrm{d}y = \|T\|_\mu^2 - 2\langle T, e\rangle_\mu$$

To end the proof, we can see that $\mathbb{E}(\Gamma_n(T)) = \|T-e\|_\mu^2 - \|e\|_\mu^2$.

5.2. *Proof of equality* (8)

Consider $\tilde{\Gamma}_n^C(T)$ given by Equation (7). Clearly,

$$\mathbb{E}\left(\int T^2(X_1, y)\mathbf{1}_{\{Z_1\ge y\}} dy \right) = \iint T^2(x,y)\bar{F}_1(x,y)\bar{G}(y)\mathrm{d}x\,\mathrm{d}y.$$

On the other hand, we get

$$\begin{aligned}
&\mathbb{E}\left[\int_0^Z \frac{1}{\bar{G}(w)} \left(\int_0^w T(X,v)\bar{G}(v)\,\mathrm{d}v \right) \mathrm{d}w \right] \\
&\quad = \mathbb{E}\left[\mathbf{1}_{Y\le C} \int_0^Y \frac{1}{\bar{G}(w)} \left(\int_0^w T(X,v)\bar{G}(v)\,\mathrm{d}v \right) \mathrm{d}w \right] \\
&\qquad + \mathbb{E}\left[\mathbf{1}_{C<Y} \int_0^C \frac{1}{\bar{G}(w)} \left(\int_0^w T(X,v)\bar{G}(v)\,\mathrm{d}v \right) \mathrm{d}w \right] \\
&\quad = \mathbb{E}\left[\bar{G}(Y) \int_0^Y \frac{1}{\bar{G}(w)} \left(\int_0^w T(X,v)\bar{G}(v)\,\mathrm{d}v \right) \mathrm{d}w \right] \\
&\qquad + \mathbb{E}\left[\bar{F}_1(X,C) \int_0^C \frac{1}{\bar{G}(w)} \left(\int_0^w T(X,v)\bar{G}(v)\,\mathrm{d}v \right) \mathrm{d}w \right] \\
&\quad = \iint T(x,v)\bar{G}(v) \left[\int_v^{+\infty} \left(\int_v^y \frac{1}{\bar{G}(w)}\,\mathrm{d}w \right) \bar{G}(y) f_{(X,Y)}(x,y)\,\mathrm{d}y \right] \mathrm{d}v\,\mathrm{d}x \\
&\qquad + \iint T(x,v)\bar{G}(v) \left[\int_v^{+\infty} \left(\int_v^y \frac{1}{\bar{G}(w)}\,\mathrm{d}w \right) \bar{F}_1(x,y) g(y)\,\mathrm{d}y \right] \mathrm{d}v\,\mathrm{d}x.
\end{aligned}$$

Denoting by $\bar{L}(x, y) = \int_y^{+\infty} f_{(X,Z)}(x, u)\,\mathrm{d}u = \bar{F}_1(x, y)\bar{G}(y)$, an integration by part yields:

$$\begin{aligned}
&\int_v^{+\infty} \left(\int_v^y \frac{1}{\bar{G}(w)}\,\mathrm{d}w\right)(\bar{G}(y) f_{(X,Y)}(x, y) + \bar{F}_1(x, y))g(y)\,\mathrm{d}y \\
&\quad = \left[-\left(\int_v^y \frac{1}{\bar{G}(w)}\,\mathrm{d}w\right)\bar{L}(x, y)\right]_v^{+\infty} + \int_v^{+\infty} \frac{\bar{L}(x, y)}{\bar{G}(y)}\,\mathrm{d}y \\
&\quad = \int_v^{+\infty} \bar{F}_1(x, y)\,\mathrm{d}y = \bar{F}_2(x, v).
\end{aligned}$$

We obtain

$$\mathbb{E}\left[\int_0^Z \frac{1}{\bar{G}(w)}\left(\int_0^w T(X, v)\bar{G}(v)\,\mathrm{d}v\right)\mathrm{d}w\right] = \iint T(x, v)e(x, v)\bar{L}(x, v)\,\mathrm{d}x\,\mathrm{d}v.$$

This implies Equation (8).

5.3. *Proof of Theorem 3.1*

Many auxiliary results used hereafter are borrowed from the work by Comte et al. (in press).

The following 'empirical' norm is involved by the definition of the contrast. For $T \in S_m$, let

$$\|T\|_n^2 := \frac{1}{n}\sum_{i=1}^n \int T^2(X_i, y)\mathbf{1}_{\{Y_i \geq y\}}\,\mathrm{d}y.$$

It is related with the scalar product defined by (3) by $\mathbb{E}(\|T\|_n^2) = \|T\|_\mu^2$. Next, we have the following relation between the norm $\|\cdot\|_n$ and the contrast Γ_n:

$$\Gamma_n(T) - \Gamma_n(S) = \|T - e\|_n^2 - \|S - e\|_n^2 - 2\nu_n(T - S), \tag{26}$$

where

$$\nu_n(T) = \frac{1}{n}\sum_{i=1}^n \left(\Psi_T(X_i, Y_i) - \int T(X_i, y)\mathbf{1}_{\{Y_i \geq y\}} e(y|X_i)\,\mathrm{d}y\right). \tag{27}$$

We shall use in the proof the following sets (recall that $\hat{H}_m$ is defined by Equation (11):

$$\hat{H} := \bigcap_{m \in \mathcal{M}_n} \hat{H}_m, \quad \Delta := \left\{\forall T \in \mathcal{S}_n : \left|\frac{\|T\|_n^2}{\|T\|_\mu^2} - 1\right| \leq \frac{1}{2}\right\} \quad \text{and} \quad \Omega := \left\{\left|\frac{\hat{\bar{F}}_0}{\bar{F}_0} - 1\right| \leq \frac{1}{2}\right\}. \tag{28}$$

For $m \in \mathcal{M}_n$, we recall that e_m is the orthogonal projection on $F_m \otimes \mathcal{H}_n$ of e restricted to A. The following bounds hold:

$$\begin{aligned}
\mathbb{E}(\|\hat{e}_{\hat{m}} - e\|_A^2) &\leq 2\|e - e_m\|_A^2 + 2\mathbb{E}(\|\hat{e}_{\hat{m}} - e_m\|_A^2\mathbf{1}(\Delta \cap \Omega)) \\
&\quad + 2\mathbb{E}(\|\hat{e}_{\hat{m}} - e_m\|_A^2\mathbf{1}(\Delta^c \cap \Omega)) + 2\mathbb{E}(\|\hat{e}_{\hat{m}} - e_m\|_A^2\mathbf{1}(\Omega^c)) \\
&\leq 2\|e - e_m\|_A^2 + 2\mathbb{E}(\|\hat{e}_{\hat{m}} - e_m\|_A^2\mathbf{1}(\Delta \cap \Omega)) \\
&\quad + 4\mathbb{E}((\|\hat{e}_{\hat{m}}\|^2 + \|e\|_A^2)\mathbf{1}(\Delta^c \cap \Omega)) + 4\mathbb{E}((\|\hat{e}_{\hat{m}}\|^2 + \|e\|_A^2)\mathbf{1}(\Omega^c)).
\end{aligned} \tag{29}$$

We use that $\mathbb{P}(\Omega^c) \leq C/n^5$ under (A3) and the following results:

PROPOSITION 5.1 *Provided that* $\mathbb{E}[Y_1^6] < +\infty$, *we have* $\mathbb{E}(\|\hat{e}_{\hat{m}}\|^4) \leq C' n^3$, *where* $C' = \phi_1^2 \mathbb{E}(Y_1^6)/9$.

PROPOSITION 5.2 *If* (A1) *is fulfilled*, *we have* $\mathbb{P}(\Delta^c) \leq C_k/n^k$ *for any* $k \geq 1$, *when* n *is large enough*, *where* C_k *is a constant depending on* $\bar{F}_0$ *and the basis.*

Proposition 5.1 is proved in Section 5.4 hereafter and Proposition 5.2 is Proposition 4 of Comte et al. (in press).

Thus, using Propositions 5.1 and 5.2 and Assumption (A3), we get

$$\begin{aligned}
&\mathbb{E}((\|\hat{e}_{\hat{m}}\|^2 + \|e\|_A^2)\mathbf{1}(\Delta^c \cap \Omega)) + \mathbb{E}((\|\hat{e}_{\hat{m}}\|^2 + \|e\|_A^2)\mathbf{1}(\Omega^c)) \\
&\quad \leq \|e\|_A^2(\mathbb{P}(\Omega^c) + \mathbb{P}(\Delta^c)) + \mathbb{E}^{1/2}(\|\hat{e}_{\hat{m}}\|^4)(\mathbb{P}^{1/2}(\Omega^c) + \mathbb{P}^{1/2}(\Delta^c)) \leq \frac{C_2}{n}.
\end{aligned} \tag{30}$$

Thus, it remains to study $\mathbb{E}(\|\hat{e}_{\hat{m}} - e_m\|_A^2 \mathbf{1}(\Delta \cap \Omega))$. The following lemma can be proved as in Comte et al. (in press).

LEMMA 5.1 *The following embedding holds*: $\Delta \cap \Omega \subset \hat{H} \cap \Omega$. *As a consequence*, *for all* $m \in \mathcal{M}_n$, *the matrices* G_m *are invertible on* $\Delta \cap \Omega$.

Now, on $\Delta \cap \Omega$ we have $\Gamma_n(\hat{e}_{\hat{m}}) + \mathrm{pen}(\hat{m}) \leq \Gamma_n(e_m) + \mathrm{pen}(m)$, where $\hat{e}_{\hat{m}} \in F_{\hat{m}} \otimes \mathcal{H}_n$ and $e_m \in F_m \otimes \mathcal{H}_n$. It follows from Equations (26) and (27) and from the inequality $2xy \leq x^2/\theta^2 + \theta^2 y^2$, with $x, y, \theta \in \mathbb{R}^+$ (here $\theta = 2$), that, on $\Delta \cap \Omega$,

$$\begin{aligned}
\|\hat{e}_{\hat{m}} - e_m\|_n^2 &\leq 2\langle \hat{e}_{\hat{m}} - e_m, e - e_m\rangle_n + \mathrm{pen}(m) + 2\nu_n(\hat{e}_{\hat{m}} - e_m) - \mathrm{pen}(\hat{m}) \\
&\leq \frac{1}{4}\|\hat{e}_{\hat{m}} - e_m\|_n^2 + 4\|e - e_m\|_n^2 + \mathrm{pen}(m) + \frac{1}{4}\|\hat{e}_{\hat{m}} - e_m\|_\mu^2 \\
&\quad + 4 \sup_{T \in B^\mu_{m,\hat{m}}(0,1)} \nu_n^2(T) - \mathrm{pen}(\hat{m}),
\end{aligned}$$

where $B^\mu_{m,m'}(0,1) := \{T \in (F_m + F_{m'}) \otimes \mathcal{H}_n : \|T\|_\mu \leq 1\}$. This yields

$$\begin{aligned}
\frac{3}{4}\|\hat{e}_{\hat{m}} - e_m\|_n^2 &\leq 4\|e - e_m\|_n^2 + \mathrm{pen}(m) + \frac{1}{4}\|\hat{e}_{\hat{m}} - e_m\|_\mu^2 \\
&\quad + 4\left(\sup_{T \in B^\mu_{m,\hat{m}}(0,1)} \nu_n^2(T) - p(m,\hat{m})\right)_+ + 4p(m,\hat{m}) - \mathrm{pen}(\hat{m}),
\end{aligned}$$

where $p(m, m') \geq 0$ is defined in the following proposition.

PROPOSITION 5.3 *Let* $p(m, m') = \kappa((\mathbb{E}(Y_1^3) + \ell(A_2)\mathbb{E}(Y_1^2))/4\bar{F}_0)((D_m + D_{m'})/n)$, *where* κ *is a numerical constant. Under the assumptions of Theorem* 3.1, *we have*

$$\mathbb{E}\left(\sup_{T \in B^\mu_{m,\hat{m}}(0,1)} (\nu_n^2(T) - p(m,\hat{m}))_+ \mathbf{1}(\Delta)\right) \leq \frac{C_1'}{n}.$$

Now, we can see that the penalty is such that

$$\forall m, m', \quad 4p(m, m') \leq \mathrm{pen}(m) + \mathrm{pen}(m'), \tag{31}$$

and use the definition of Δ. We obtain on $\Delta \cap \Omega$:

$$\frac{1}{2}\|\hat{e}_{\hat{m}} - e_m\|_\mu^2 \leq 4\|e - e_m\|_n^2 + 2\mathrm{pen}(m) + \frac{1}{4}\|\hat{e}_{\hat{m}} - e_m\|_\mu^2 + 4\left(\sup_{T \in B^\mu_{m,\hat{m}}(0,1)} \nu_n^2(T) - p(m,\hat{m})\right)_+$$

and thus on $\Delta \cap \Omega$:

$$\frac{1}{4}\|\hat{e}_{\hat{m}} - e_m\|_\mu^2 \leq 4\|e - e_m\|_n^2 + 2\mathrm{pen}(m) + 4\left(\sup_{T \in B^\mu_{m,\hat{m}}(0,1)} \nu_n^2(T) - p(m,\hat{m})\right)_+ .$$

Taking the expectation of the last Inequality and using proposition 5.3, we get:

$$\frac{1}{4}\mathbb{E}(\|\hat{e}_{\hat{m}} - e_m\|_\mu^2 \mathbf{1}(\Delta \cap \Omega)) \leq 4\|e - e_m\|_\mu^2 + 2\mathrm{pen}(m) + \frac{C_1}{n}. \tag{32}$$

Combining (29), (30) and (32) leads to

$$\begin{aligned}\mathbb{E}(\|\hat{e}_{\hat{m}} - e\|_A^2) &\leq 2\|e_m - e\|_A^2 + 8\bar{F}_0^{-1}\left(4\|e - e_m\|_\mu^2 + 2\mathrm{pen}(m) + \frac{C_1}{n}\right) + \frac{C_2}{n} \\ &\leq 2(1 + 16\bar{F}_0^{-1})\|e_m - e\|_A^2 + 16\bar{F}_0^{-1}\mathrm{pen}(m) + \frac{C_3}{n}\end{aligned} \tag{33}$$

for any $m \in \mathcal{M}_n$. This concludes the proof of Theorem 3.1.

5.4. *Proof of Proposition* 5.1

Let us note that $\hat{e}_{\hat{m}}$ is either 0 or $\arg\min_{T \in S_{\hat{m}}} \Gamma_n(T)$. In the second case, $\min \mathrm{Sp}(G_{\hat{m}}) \geq \max(\hat{\bar{F}}_0, n^{-1/2})$ and thus

$$\begin{aligned}\|\hat{e}_{\hat{m}}\|^2 &= \sum_{j,k} (\hat{a}^{\hat{m}}_{j,k})^2 = \|A_{\hat{m}}\|^2 = \|G_{\hat{m}}^{-1}\Upsilon_{\hat{m}}\|^2 \leq \left(\frac{1}{\min \mathrm{Sp}(G_{\hat{m}})}\right)^2 \|\Upsilon_{\hat{m}}\|^2 \\ &\leq \min\left(\frac{1}{\hat{\bar{F}}_0^2}, n\right) \sum_{j,k}\left(\frac{1}{n}\sum_{i=1}^n \varphi_j^{\hat{m}}(X_i) \int_{A_2} \mathbf{1}_{(0 \leq u \leq Y_i)}(Y_i - u)\psi_k(u)\,\mathrm{d}u\right)^2 \\ &\leq n\frac{1}{n}\sum_{i=1}^n \sum_j (\varphi_j^{\hat{m}}(X_i))^2 \sum_k \left(\int_{A_2} \mathbf{1}_{(0 \leq u \leq Y_i)}(Y_i - u)\psi_k(u)\,\mathrm{d}u\right)^2.\end{aligned}$$

Thus, $\|\hat{e}_{\hat{m}}\|^2 \leq \phi_1 \mathcal{D}_n^{(1)} \sum_{i=1}^n \int_{A_2} \mathbf{1}_{(0 \leq u \leq Y_i)}(Y_i - u)^2\,\mathrm{d}u \leq \phi_1 \mathcal{D}_n^{(1)} \sum_{i=1}^n Y_i^3/3$, and it follows that

$$\mathbb{E}(\|\hat{e}_{\hat{m}}\|^4) \leq \phi_1^2 (\mathcal{D}_n^{(1)})^2 \mathbb{E}\left[\left(\sum_{i=1}^n \frac{Y_i^3}{3}\right)^2\right] \leq \frac{\phi_1^2 (\mathcal{D}_n^{(1)})^2 n^2 \mathbb{E}(Y_1^6)}{9} \leq \frac{\phi_1^2 \mathbb{E}(Y_1^6)}{9} n^3.$$

5.5. *Proof of Proposition* 5.3

We use several times the same very useful inequality based on the property that the squared norm of the orthogonal projection of a function is less than the squared norm of the function itself. We use this property as follows. For any function $h \in \mathbb{L}^2(A_2)$,

$$\sum_k \left(\int_{A_2} h(v)\psi_k(v)\,\mathrm{d}v \right)^2 \leq \|h\|^2. \tag{34}$$

Let $W = (X, Y)$ and $\xi_T(W) = \Psi_T(X, Y) - \int T(X, v)\mathbf{1}_{\{Y \geq v\}} e(v|X)\,\mathrm{d}v$. To study the empirical process, we split $\xi_T(W)$ in three parts $\xi_T(W) = \xi_{T,1}(W) + \xi_{T,2}(W) - \xi_{T,3}(W)$, with

$$\begin{aligned}
\xi_{T,1}(W) &= \Psi_T(X, Y)\mathbf{1}_{\{Y \leq k_n\}} - \mathbb{E}(\Psi_T(X, Y)\mathbf{1}_{\{Y \leq k_n\}}),\\
\xi_{T,2}(W) &= \Psi_T(X, Y)\mathbf{1}_{\{Y > k_n\}} - \mathbb{E}(\Psi_T(X, Y)\mathbf{1}_{\{Y > k_n\}}),\\
\xi_{T,3}(W) &= \int T(X, v)\mathbf{1}_{\{Y \geq v\}} e(v|X)\,\mathrm{d}v - \mathbb{E}\left(\int T(X, v)\mathbf{1}_{\{Y \geq v\}} e(v|X)\,\mathrm{d}v \right).
\end{aligned}$$

Then $\nu_n(T) = (1/n)\sum_{i=1}^n \xi_T(W_i)$ can be split in the same way: $\nu_n(T) = \nu_{n,1}(T) + \nu_{n,2}(T) - \nu_{n,3}(T)$ with $\nu_{n,k}(T) = (1/n)\sum_{i=1}^n \xi_{T,k}(W_i)$ for $k = 1, 2, 3$. We choose

$$k_n = \left(\frac{3n}{\log^4(n)} \right)^{1/3}, \tag{35}$$

Now, the main tool of the proof is the checkout of the Talagrand inequality Talagrand (1996).

LEMMA 5.2 *Let $W_1, \ldots, W_n$ be i.i.d. random variables and $(\xi_T)_{T \in \mathcal{B}}$ a set of bounded functions and $\mathcal{B}$ a unit ball of a finite-dimensional subspace of $\mathbb{L}^2(A)$. Let $\nu_n(T) = (1/n)\sum_{i=1}^n \xi_T(W_i)$ where $\mathbb{E}[\xi_T(W_1)] = 0$, and suppose that*:

$$\text{(i)} \sup_{T \in \mathcal{B}} \|\xi_T\|_\infty \leq M_1, \quad \text{(ii)} \sup_{T \in \mathcal{B}} \operatorname{Var}[\xi_T(W_1)] \leq v \quad \text{and} \quad \text{(iii)}\ \mathbb{E}\left(\sup_{T \in \mathcal{B}} |\nu_n(T)| \right)^2 \leq H^2.$$

Then, there exists constants $K > 0$, $K_1 > 0$ and $K_2 > 0$ such that:

$$\mathbb{E}\left[\sup_{T \in \mathcal{B}} |\nu_n(T)|^2 - 2H^2 \right] \leq K \left[\frac{v}{n} e^{-K_1(nH^2/v)} + \frac{M_1^2}{n^2} e^{-K_2(nH/M_1)} \right]$$

We apply Talagrand's inequality given in Lemma 5.2 to the terms involving $\nu_{n,1}$ and $\nu_{n,3}$ in the following inequality:

$$\begin{aligned}
\mathbb{E}\left(\sup_{T \in B^\mu_{m,\hat{m}}(0,1)} \left(\nu_n^2(T) - p(m, \hat{m})\right)_+ \mathbf{1}(\Delta) \right) &\leq 3\mathbb{E}\left(\sup_{T \in B^\mu_{m,\hat{m}}(0,1)} \left(\nu_{n,1}^2(T) - \frac{p_1(m, \hat{m})}{6} \right)_+ \right)\\
&\quad + 3\mathbb{E}\left(\sup_{T \in B^\mu_n(0,1)} (\nu_{n,2}^2(T)) \right)\\
&\quad + 3\mathbb{E}\left(\sup_{T \in B^\mu_{m,\hat{m}}(0,1)} \left(\nu_{n,3}^2(T) - \frac{p_3(m, \hat{m})}{6} \right)_+ \right),
\end{aligned}$$

where $p(m, m') = p_1(m, m') + p_3(m, m')$ and $B^\mu_n(0, 1) = \{T \in \mathcal{S}_n : \|T\|_\mu \leq 1\}$.

Study of $\nu_{n,2}$. Recall that $\mathbb{E}(Y_1^6) < +\infty$. We write:

$$\mathbb{E}\left(\sup_{T\in B_n^\mu(0,1)}(\nu_{n,2}^2(T))\right) \leq \frac{1}{\bar{F}_0}\sum_{j,k}\mathbb{E}(\nu_{n,2}^2(\varphi_j^n\otimes\psi_k)),$$

where $(\varphi_j^n\otimes\psi_k)_{j,k}$ denotes here an orthonormal basis of $\mathcal{S}_n$ w.r.t the norm $\|\cdot\|_A$. This implies, as $\mathbb{E}(\nu_{n,2}^2(T)) = (1/n)\mathrm{Var}(\xi_{T,2}(W))$ and using Equation (34), that

$$\mathbb{E}\left(\sup_{T\in B_n^\mu(0,1)}(\nu_{n,2}^2(T))\right) \leq \frac{2}{n\bar{F}_0}\sum_{j,k}\left[\mathbb{E}\left(\int(Y_1-v)\mathbf{1}_{\{v\leq Y_1\}}\varphi_j^n(X_1)\psi_k(v)\,\mathrm{d}v\mathbf{1}_{\{Y_1>k_n\}}\right)^2\right],$$

and with Equation (9)

$$\begin{aligned}\mathbb{E}\left(\sup_{T\in B_n^\mu(0,1)}(\nu_{n,2}^2(T))\right) &\leq \frac{2\phi_1\mathcal{D}_n^{(1)}}{n\bar{F}_0}\mathbb{E}\left(\int_{A_2}(Y_1-v)^2\mathbf{1}_{\{v\leq Y_1\}}\,\mathrm{d}v\mathbf{1}_{\{Y_1>k_n\}}\right)^2\\ &\leq \frac{2\phi_1}{\sqrt{n}\bar{F}_0}\mathbb{E}(Y_1^3\mathbf{1}_{\{Y_1>k_n\}}) \leq \frac{2\phi_1}{\sqrt{n}\bar{F}_0}\frac{\mathbb{E}(Y_1^{3+p})}{k_n^p}\\ &\leq \frac{2\phi_1\mathbb{E}(Y_1^{3+p})}{3^{p/3}\bar{F}_0}\frac{(\log(n))^{4p/3}}{n^{1/2+p/3}} \leq \frac{C}{n}\end{aligned}$$

as soon as we take $p > 3/2$, e.g. $p = 2$.

Study of $\nu_{n,1}$.

We apply Talagrand's Inequality, and for this purpose, we will have to check (i), (ii) and (iii) and to compute M_1, v and H^2 defined in Lemma 5.2.

(i) Search for bound M_1. Under (A1), we have,

$$\sup_{T\in B_{m,m'}^\mu(0,1)}\ \sup_{x\in\mathbb{R},y\in\mathbb{R}}|\xi_{T,1}(x,y)| \leq \frac{1}{\sqrt{\bar{F}_0}}\sup_{T\in B_{m,m'}(0,1)}\ \sup_{x\in\mathbb{R},y\in\mathbb{R}}|\xi_{T,1}(x,y)|$$

Here $B_{m,m'}(0,1) = \{T\in F_{m\vee m'}\otimes\mathcal{H}_n : \|T\|_A\leq 1\}$ and $T = \sum_{j,k}a_{j,k}\varphi_j\psi_k$, where $(\varphi_j\otimes\psi_k)_{(j,k)}$ is an orthonormal basis of $F_{m\vee m'}\otimes\mathcal{H}_n$ w.r.t. the norm $\|\cdot\|_A$, where $F_{m\vee m'} = F_m + F_{m'}$ and $\dim(F_{m\vee m'}) = \max(D_m, D_{m'})$ (nested collection).

$$\begin{aligned}|\Psi_T(x,y)| &= \left|\sum_{j,k}a_{j,k}\varphi_j(x)\int(y-v)\mathbf{1}_{\{v\leq y\}}\psi_k(v)\,\mathrm{d}v\mathbf{1}_{\{y\leq k_n\}}\right|\\ &\leq \left(\sum_{j,k}a_{j,k}^2\sum_j(\varphi_j(x))^2\sum_k\left(\int(y-v)\mathbf{1}_{\{v\leq y\}}\psi_k(v)\,\mathrm{d}v\mathbf{1}_{\{y\leq k_n\}}\right)^2\right)^{1/2}\\ &\leq \|T\|_A\left(\sum_j(\varphi_j(x))^2\int(y-v)^2\mathbf{1}_{\{v\leq y\}}\,\mathrm{d}v\mathbf{1}_{\{y\leq k_n\}}\right)^{1/2} \quad \text{with inequality (5.12)}\\ &\leq \sqrt{\phi_1\left(\frac{k_n^3}{3}\right)\max(D_m,D_{m'})} \leq \frac{\sqrt{\phi_1(D_m+D_{m'})n}}{\log^2(n)}\end{aligned}$$

with (9) and recalling that $k_n = (3n/\log^4(n))^{1/3}$ (see Equation (35)). Therefore,

$$\sup_{T\in B^{\mu}_{m,m'}(0,1)} \|\xi_T\|_\infty \le 2\frac{\sqrt{\phi_1(D_m+D_{m'})n}}{\log^2(n)\sqrt{\bar F_0}} := M_1.$$

(ii) Search for bound v. First, let $B=\max(D_m, D_{m'})^{1/5}$ and write

$$\begin{aligned}\mathrm{Var}[\xi_{T,1}(W_1)] &\le \mathbb{E}[\Psi_T^2(X_1,Y_1)\mathbf{1}_{\{Y_1<k_n\}}] \le \mathbb{E}[\Psi_T^2(X_1,Y_1)]\\ &\le \mathbb{E}[\Psi_T^2(X_1,Y_1)\mathbf{1}_{\{Y_1\le B\}}] + \mathbb{E}[\Psi_T^2(X_1,Y_1)\mathbf{1}_{\{Y_1>B\}}].\end{aligned}$$

Now, we study each term. First, we have for $T\in B^{\mu}_{m,m'}(0,1)$

$$\begin{aligned}&\mathbb{E}[\Psi_T^2(X_1,Y_1)\mathbf{1}_{\{Y_1\le B\}}] \le \mathbb{E}\left[\left(\int T(X_1,v)(Y_1-v)\mathbf{1}_{\{v\le Y_1\}}\,\mathrm{d}v\right)^2\mathbf{1}_{\{Y_1\le B\}}\right]\\ &\le \ell(A_2)B^2\mathbb{E}\left[\int_{A_2} T^2(X_1,v)\mathbf{1}_{\{v\le Y_1\}}\,\mathrm{d}v\right]\\ &= \ell(A_2)B^2\|T\|_\mu^2 = O(B^2) = O((\max(D_m,D_{m'})^{2/5}).\end{aligned}$$

On the other hand, for $T(x,y)=\sum_{j,k} a_{j,k}\varphi_j(x)\psi_k(y)$, and $\|T\|_\mu^2\le 1$, we have,

$$\begin{aligned}\mathbb{E}[\Psi_T^2(X_1,Y_1)\mathbf{1}_{\{Y_1>B\}}] &\le \sum_{j,k} a_{j,k}^2 \sum_{j,k}\mathbb{E}[\varphi_j^2(X_1)((Y_1-v)\psi_k(v)\mathbf{1}_{\{v\le Y_1\}}\mathbf{1}_{\{Y_1>B\}}\,\mathrm{d}v)^2]\\ &\le \frac{1}{\bar F_0}\mathbb{E}\left[\sum_j \varphi_j^2(X_1)\int_{A_2}(Y_1-v)^2\mathbf{1}_{\{v\le Y_1\}}\mathbf{1}_{\{Y_1>B\}}\,\mathrm{d}v\right]\\ &\le \frac{\phi_1\max(D_m,D_{m'})}{\bar F_0}\,\frac{\mathbb{E}(Y_1^3\mathbf{1}_{\{Y_1>B\}})}{3}\\ &\le \frac{\phi_1\mathbb{E}(Y_1^6)\max(D_m,D_{m'})}{3\bar F_0 B^3} = O((\max(D_m,D_{m'})^{2/5}).\end{aligned}$$

Therefore, $\sup_{T\in B^{\mu}_{m,m'}(0,1)}\mathrm{Var}[\xi_{T,1}(W_1)]\le C(D_m+D_{m'})^{2/5} := v$, where C is a constant depending on $\ell(A_2)$, ϕ_1, $\bar F_0$ and $\mathbb{E}(Y_1^6)$.

(iii) Search for bound H^2 : Let us write here $T=\sum_{j,k} a_{j,k}\varphi_j\psi_k$, where $(\varphi_j\otimes\psi_k)_{(j,k)}$ is an orthonormal basis of $(F_m+F_{m'})\otimes\mathcal{H}_n$ w.r.t. the norm $\|.\|_A$.

$$\begin{aligned}&\mathbb{E}\left(\sup_{T\in B^{\mu}_{m,m'}(0,1)}|\nu_{n,1}^2(T)|\right) \le \frac{1}{\bar F_0}\mathbb{E}\left(\sup_{T\in B_{m,m'}(0,1)}|\nu_{n,1}^2(T)|\right) \le \frac{1}{\bar F_0}\sum_{j,k}\mathbb{E}(\nu_{n,1}^2(\varphi_j\otimes\psi_k))\\ &\le \frac{1}{n\bar F_0}\sum_{j,k}\mathbb{E}\left(\int (Y_1-v)\mathbf{1}_{\{v\le Y_1\}}\varphi_j(X_1)\psi_k(v)\,\mathrm{d}v\right)^2\end{aligned}$$

$$\leq \frac{1}{n\bar{F}_0}\sum_j \mathbb{E}\left[\varphi_j^2(X_1)\int (Y_1 - v)^2 \mathbf{1}_{\{v\leq Y_1\}}\,\mathrm{d}v\right]$$

$$\leq \frac{1}{n\bar{F}_0}\sum_j \iint \varphi_j^2(x)\int (y - v)^2 \mathbf{1}_{\{v\leq y\}}\,\mathrm{d}v f_{(X,Y)}(x, y)\,\mathrm{d}x\,\mathrm{d}y$$

$$\leq \frac{1}{n\bar{F}_0}\iint \left(\sum_j \varphi_j^2(x)\right)\frac{y^3}{3} f_{(X,Y)}(x, y)\,\mathrm{d}x\,\mathrm{d}y \leq \frac{\phi_1 \max(D_m, D_{m'})\mathbb{E}(Y_1^3)}{3n\bar{F}_0}.$$

Therefore, $\mathbb{E}(\sup_{T\in B^{\mu}_{m,m'}(0,1)} |\nu^2_{n,1}(T)|) \leq (\phi_1(D_m + D_{m'})\mathbb{E}(Y_1^3))/(3n\bar{F}_0) := H^2$.

Applying Lemma 5.2 yields that

$$\mathbb{E}\left(\sup_{T\in B^{\mu}_{m,m'}(0,1)} \nu^2_{n,1}(T) - 2H^2\right)$$
$$\leq K'\left(\frac{(D_m + D_{m'})^{2/5}}{n}\mathrm{e}^{-K_1'(D_m+D_{m'})^{3/5}} + \frac{D_m + D_{m'}}{n\log^4(n)}\mathrm{e}^{-K_2'\log^2(n)}\right)$$
$$\leq \frac{K''}{n}(D_m^{2/5}\mathrm{e}^{-K_3' D_m^{3/5}}\mathrm{e}^{-K_3' D_{m'}^{3/5}} + D_{m'}^{2/5}\mathrm{e}^{-K_3' D_{m'}^{3/5}}) + \frac{2K''}{n^{1/2}\log^4(n)}\mathrm{e}^{-K_2'\log^2(n)}$$

using that $(x + y)^a \geq (x^a + y^a)/2$ for $a = 2/5$ or $a = 3/5$. As $k^{2/5}\exp(-Ck^{3/5})$ is bounded and with finite sum for $k \in \mathbb{N}$ and $K''|\mathcal{M}_n|n^{-1/2}\log^{-4}(n)\mathrm{e}^{-K_2'\log^2(n)}$ is $O(1/n)$, it follows that

$$\mathbb{E}\left(\sup_{T\in B^{\mu}_{m,\hat{m}}(0,1)} \nu^2_{n,1}(T) - 2H^2\right) \leq \sum_{m'\in\mathcal{M}_n}\mathbb{E}\left(\sup_{T\in B^{\mu}_{m,m'}(0,1)} \nu^2_{n,1}(T) - 2H^2\right) \leq \frac{C}{n}.$$

Study of $\nu_{n,3}$.

(i) Search for M_1. First, let $T(x, y) = \sum_{j,k} a_{j,k}\varphi_j(x)\psi_k(y) \in B^{\mu}_{m,m'}(0, 1)$, we note that

$$\int_{A_2} T^2(x, v)\,\mathrm{d}v = \sum_{j,j'}\left(\sum_k a_{j,k}a_{j',k}\right)\varphi_j(x)\varphi_{j'}(x)$$

For $b_j = (\sum_k a_{j,k}^2)^{1/2}$, we have $\sum_j b_j^2 \leq \sum_{j,k} a_{j,k}^2 \leq 1/\bar{F}_0$, and

$$\int_{A_2} T^2(x, v)\,\mathrm{d}v \leq \sum_{j,j'} b_j b_{j'}|\varphi_j(x)\varphi_{j'}(x)| = \left(\sum_j b_j|\varphi_j(x)|\right)^2$$
$$\leq \sum_j b_j^2 \sum_j \varphi_j^2(x) \leq \left(\frac{1}{\bar{F}_0}\right)\|\sum_j \varphi_j^2\|_\infty \leq \left(\frac{1}{\bar{F}_0}\right)\phi_1\max(D_m, D_{m'}).$$

This yields

$$\left|\int T(x,v)\mathbf{1}_{\{y\geq v\}}e(v|x)\,\mathrm{d}v\right| \leq \left(\int_{A_2} T^2(x,v)\,\mathrm{d}v\int_{A_2} e^2(v|x)\mathbf{1}_{A_1}(x)\,\mathrm{d}v\right)^{1/2}$$
$$\leq \left[\left(\frac{1}{\bar{F}_0}\right)\phi_1\max(D_m,D_{m'})\right]^{1/2}\sup_{(x,v)\in A}|e(v|x)|\sqrt{\ell(A_2)}$$
$$= \left(\frac{\ell(A_2)\|e\|_{\infty,A}\phi_1}{\bar{F}_0}\right)^{1/2}\sqrt{D_m+D_{m'}} := M_1$$

(ii) Search for v.

$$\mathbb{E}\left(\int T(X_1,v)\mathbf{1}_{\{Y_1\geq v\}}e(v|X_1)\,\mathrm{d}v\right)^2$$
$$\leq \mathbb{E}\left[\int_{A_2} T^2(X_1,v)\mathbf{1}_{\{v\leq Y_1\}}e^2(v|X_1)\,\mathrm{d}v\right]$$
$$\leq \iint_A T^2(x,v)e^2(v|x)\int \mathbf{1}_{\{v\leq y\}}f_{(X,Y)}(x,y)\,\mathrm{d}y\,\mathrm{d}v\,\mathrm{d}x$$
$$\leq \iint_A T^2(x,v)e^2(v|x)\bar{F}_1(x,v)\,\mathrm{d}v\,\mathrm{d}x \leq \|e\|^2_{\infty,A}\|T\|^2_\mu = \|e\|^2_{\infty,A} := v.$$

(iii) Search for H^2. We also have with the same argument,

$$\sum_{j,k}\mathbb{E}\left(\int \varphi_j(X_1)\psi_k(v)\mathbf{1}_{\{Y_1\geq v\}}e(v|X_1)\,\mathrm{d}v\right)^2 = \sum_j \mathbb{E}\left[\varphi_j^2(X_1)\int_{A_2}\mathbf{1}_{\{Y_1\geq v\}}e^2(v|X_1)\,\mathrm{d}v\right]$$
$$\leq \phi_1\max(D_m,D_{m'})\int_{A_2}\mathbb{E}(e^2(v|X_1))\,\mathrm{d}v$$
$$\leq \phi_1(D_m+D_{m'})\int_{A_2}\mathbb{E}[\mathbb{E}((Y_1-v)^2|Y_1>v,X_1)]\,\mathrm{d}v$$
$$\leq \phi_1(D_m+D_{m'})\ell(A_2)\mathbb{E}(Y_1^2).$$

Therefore, $\mathbb{E}(\sup_{T\in B^\mu_{m,m'}(0,1)}|\nu^2_{n,3}(T)|) \leq (\phi_1(D_m+D_{m'})\mathbb{E}(Y_1^2)\ell(A_2))/(n\bar{F}_0) := H^2$. Applying Lemma 5.2 yields that

$$\mathbb{E}\left(\sup_{T\in B^\mu_{m,m'}(0,1)}\nu^2_{n,3}(T)-2H^2\right) \leq K'\left(\frac{1}{n}\mathrm{e}^{-K_1'(D_m+D_{m'})}+\frac{1}{n}\mathrm{e}^{-K_2'\sqrt{n}}\right),$$

and this gives the result in the same manner as in the previous case.

References

Abdous, B., and Berred, A. (2005), 'Mean Residual Life Estimation', *Journal of Statistical Planning and Inference*, 132, 3–19.

Barron, A.R., Birgé, L., and Massart, P. (1999), 'Risk Bounds for Model Selection Via Penalization', *Probability Theory and Related Fields*, 113, 301–413.

Brunel, E., Comte, F., and Lacour, C. (2010), 'Minimax Estimation of the Conditional Cumulative Distribution Function Under Random Censorship', *Sankhya A*, 72, Part 2, 293–330.

Castellan, G., and Letué, F. (2000), 'Estimation of the Cox Regression Function Via Model Selection', Ph.D. thesis. http://ljk.imag.fr/membres/Frederique.Letue/.

Chaubey, Y.P., and Sen, P.K. (1999), 'On Smooth Estimation of Mean Residual Life', *Journal of Statistical Planning and Inference*, 75, 223–236.

Chen, Y.Q., and Cheng, S. (2005), 'Semiparametric Regression Analysis of Mean Residual Life With Censored Survival Data', *Biometrika*, 92, 19–29.

Chen, Y.Q., Jewell, N.P., Lei, X., Cheng, S.C. (2005), 'Semiparametric Estimation of Proportional Mean Residual Life Model in Presence of Censoring', *Biometrics*, 61, 170–178.

Cohen, A., Daubechies, I., and Vial, P. (1993), 'Wavelets on the Interval and Fast Wavelet Transforms', *Applied and Computational Harmonic Analysis*, 1, 54–81.

Comte, F., Gaïffas, S., and Guilloux, A. (in press), 'Adaptive Estimation of the Conditional Intensity of Marker-Dependent Counting Processes', *Annales de l'Inst. Henri Poincaré Probability and Statistics*. http://www.imstat.org/aihp/accepted.html.

Csörgö, M., and Zitikis, R. (1996), 'Mean Residual Life Processes' *Annals of Statistics*, 24, 1717–1739.

DeVore, R.A., and Lorentz, G.G. (1993), *Constructive Approximation*, Berlin: Springer-Verlag.

Embrechts, P., Klüppelberg, C., and Mikosch, T. (1997), *Modelling Extremal Events*, eds. I. Karatzas and M. Yor, Berlin: Springer.

Hall, W.J., and Wellner, J. (1981), 'Mean Residual Life', in *Statistics and Related Topics (Ottawa, Ontario, 1980)*, eds. M. Csorgo, D.A. Dawson, J.N.K. Rao and A.K.Md.E. Saleh, Amsterdam, New York: North-Holland, pp. 169–184.

Kaplan, E.L., and Meier, P. (1958), 'Nonparametric Estimation from Incomplete Observations', *Journal of the American Statistical Association*, 53, 457–481.

Lawless, J.F. (2003), *Statistical Models and Methods for Lifetime Data*, 2nd ed., Wiley Series in Probability and Statistics. Wiley-Interscience [John Wiley & Sons], Hoboken, NJ.

Leurgans, S. (1987), 'Linear Models, Random Censoring and Synthetic Data', *Biometrika*, 74 , 301–309.

Lo, S.H., Mack, Y.P., and Wang, J.L. (1989), 'Density and Hazard Rate Estimation for Censored Data Via Strong Representation of the Kaplan-Meier Estimator', *Probability and Theory Related Fields*, 80, 461–473.

Maguluri, G., and Zhang, C.-H. (1994), 'Estimation in the Mean Residual Life Regression Model', *Journal of the Royal Statistical Society Series B*, 56, 477–489.

McLain, A.C., and Ghosh S.K. (2009), 'Nonparametric Estimation of the Conditional Mean Residual Life Function Based on Censored Data', NC State Department of Statistics, Technical report # 2627. http://www.lib.ncsu.edu/resolver/1840.4/4103.

Na, M.H., and Kim, J.J. (1999), 'On Inference for Mean Residual Life', *Communications in Statististics – Theory and Methods*, 28, 2917–2933.

Oakes, D., and Dasu, T. (1996), 'A Note on Residual Life', *Biometrika*, 77, 409–410.

Prentice, R.I. (1973), 'Exponential Survivals with Censoring and Explanatory Variables', *Biometrika*, 60, 279–288.

Talagrand, M. (1996), 'New Concentration Inequalities in Product Spaces', *Inventiones Mathematicae*, 126, 505–563.