To appear in *Statistics: A Journal of Theoretical and Applied Statistics*
Vol. 00, No. 00, Month 201X, 1–32

Semiparametric inference for the recurrent events process by means of a single-index model

Olivier Bouaziz[a*], Ségolen Geffray[b] and Olivier Lopez[c,d]

[a]*MAP5, Université Paris Descartes, 45 rue des Saints Pères, 75270 Paris Cedex 06*; [b]*IRMA, Université de Strasbourg, 7 rue René-Descartes 67084 Strasbourg Cedex*; [c]*ENSAE Paris-Tech & CREST, Laboratoire de Finance et d'Assurance, 3 avenue Pierre Larousse, 92245 Malakoff Cedex France*; [d] *Sorbonne Universités, UPMC Université Paris VI, EA 3124, LSTA, 4 place Jussieu 75005 Paris*

(*Received 00 Month 201X; final version received 00 Month 201X*)

In this paper, we introduce new parametric and semiparametric regression techniques for a recurrent event process subject to random right censoring. We develop models for the cumulative mean function and provide asymptotically normal estimators. Our semiparametric model which relies on a single-index assumption can be seen as a dimension reduction technique that, contrary to a fully nonparametric approach, is not stroke by the curse of dimensionality when the number of covariates is high. We discuss data-driven techniques to choose the parameters involved in the estimation procedures and provide a simulation study to support our theoretical results.

Key words and phrases: Asymptotic normality, dimension reduction, empirical processes, recurrent events, right-censoring, single-index model.

This work is supported by French Agence Nationale de la Recherche (ANR) ANR Grant "Prognostic" ANR-09-JCJC-0101-01.

1. Introduction

The modeling of recurrent events has become a crucial issue in various application fields of statistical inference such as clinical and epidemiological studies, insurance or actuarial science, in particular in the presence of a terminal event. Among many examples, one can mention the modeling of asthma, of epileptic seizures in the presence of death or of repeated warranty claims with possibility of contract breaking. In these settings, we consider the problem of predicting or identifying the causes which influence the number of such events occurring during a given time period. A natural way to measure the impact of covariates on the recurrence of these events consists of estimating the cumulative function of some recurrent event process of interest conditionally on covariates. In this paper, our

Corresponding author*. Email: olivier.bouaziz@parisdescartes.fr

aim consists of developing both parametric and semiparametric inference for regression models suited to the conditional cumulative mean function in the presence of independent right-censoring and terminal event.

Let $\widetilde{N}(t)$ represent the number of recurrent events occurring in the time interval $[0, t]$ for $t \geq 0$. In the literature, various regression models tailored for the recurrent event process $\widetilde{N}(\cdot)$ have been considered. These regression models can be roughly divided into two categories which are Cox-type regression models and accelerated-failure-time (AFT)-type regression models. Cox-type regression models (see eg Andersen and Gill [1982] or Sun and Wei [2000]) typically assume that a vector of covariates Z acts on the conditional cumulative mean function through a multiplicative effect on an unspecified baseline function $\mu_0(\cdot)$. For instance, Sun and Wei [2000] assume that:

$$E[\widetilde{N}(t)|Z] = \exp(\beta' Z)\mu_0(t) \tag{1.1}$$

where β is an unknown vector of parameters. AFT-type regression models (see eg Lin et al. [1998]) typically assume that a vector of covariates Z acts on the conditional cumulative mean function through a time-scaling effect on an unspecified baseline function $\mu_0(\cdot)$:

$$E[\widetilde{N}(t)|Z] = \mu_0(t \exp(\beta' Z)) \tag{1.2}$$

even though variations of AFT-like regression models exist, for example, in Ghosh [2004] under the form

$$E[\widetilde{N}(t)|Z] = \exp(-\beta' Z)\, \mu_0(t \exp(\beta' Z)). \tag{1.3}$$

However, none of the aforementioned papers takes into account terminal event occurrence or dependent censoring. One of the most developed inferential approaches in the presence of both dependent and independent censoring can be found in Ghosh and Lin [2003] where the recurrent event process $\widetilde{N}(\cdot)$ is modeled by means of AFT-type regression models. Frailty and time-dependent covariates extensions of Cox-type regression models have also been considered in Huang and Wang [2004] and Huang et al. [2010].

Technically speaking, the main advantage of these kinds of models stands in the simplicity of the regression function. But they unfortunately face the disadvantage (with respect to a purely nonparametric approach) of relying on strong modeling assumptions that may not hold in practice. However, it turns out that, while allowing full flexibility, the nonparametric approach is known to fail when the number of covariates is high (greater than 3 in practice) which is the so-called "curse-of-dimensionality". On the other hand, single-index models (see e.g. Ichimura [1993], Härdle et al. [1993], Xia et al. [2002] and Bouaziz and Lopez [2010]) aim to achieve a compromise between a parametric approach and a nonparametric one. The basic idea behind this class of models is to assume that the regression function depends on an unknown linear combination of the covariates. In our

framework, this corresponds to assuming that $E[\widetilde{N}(t)|Z] = E[\widetilde{N}(t)|\theta_0'Z]$, for some unknown vector θ_0. Hence if this parameter was known, estimating $E[\widetilde{N}(t)|Z]$ would reduce to a nonparametric problem in dimension one. This dimension reduction assumption allows to achieve better convergence rates and still ensures enough flexibility to be adapted to a large number of practical cases. Moreover, these models can also generalize standard models such as the Cox or AFT regression models.

In this work, we first study a general parametric regression model for a recurrent event process in the presence of independent right-censoring and terminal event. We then study a semiparametric generalization which relies on a single-index assumption. We propose new procedures to estimate both the index and the conditional cumulative mean regression function and provide a detailed asymptotic study of the proposed estimators. Compared to uncensored single-index models adapted to mean-regression, see e.g. Ichimura [1993], the presence of censoring usually deteriorates the quality of estimation in the tail of the distribution. Therefore, in our specific setting of regression for recurrent events with censoring and terminal event, we introduce a weight function designed to compensate for the lack of information induced by independent censoring. The main novelty of our procedure stands in the fact that this weight function may be chosen using data-driven techniques. We then discuss a data-driven way of calibrating the parameters involved in the estimation procedures.

The paper is organized as follows. In Section 2, we define the parametric and semiparametric models, introduce the corresponding estimators and explain the general methodology. Asymptotic results for our new estimators are presented in Section 3. Simulation studies are carried out in Section 4 to investigate on the performance of our methods for moderate sample size. Technical results are postponed to the Appendix in Section 6.

2. Model assumptions and methodology

In this section, we present the regression framework. Specifically, Section 2.1 introduces the different regression models considered in this paper. Section 2.2 presents the estimation procedures. They are based on a least-squares type criterion involving a rescaled process and a weighting measure introduced in Section 2.2.1 which enable to correct the impact of censoring. The parametric case is dealt with in Subsection 2.2.2 and the semiparametric case is treated in Subsection 2.2.3.

2.1 Regression models for the recurrent event process

Consider the process $\widetilde{N}(\cdot)$ where we recall that $\widetilde{N}(t)$ stands for the number of recurrent events occurring in the time interval $[0, t]$ for any $t \geq 0$. Let D denote the random variable representing the time until occurrence of a terminal event. In clinical applications, this variable D may stand for the death time of a patient. For insurance applications, D can represent the warranty length (which can be random if the client has the possibility of

breaking the contract) or the lifetime of the insured good. Then introduce the recurrent event process $N^*(\cdot) = \widetilde{N}(\cdot \wedge D)$ (where $a \wedge b$ denotes the infimum between a and b) which can be seen as a piecewise constant function with jumps only on $[0, D]$. In this paper, we aim to infer on the cumulative conditional mean function given for $t \geq 0$ by

$$\mu(t|z) = E\left[N^*(t)|Z = z\right]$$

where Z is a $d-$dimensional vector of covariates. Note that our interest is focused on the process $N^*(\cdot)$ and not on $\widetilde{N}(\cdot)$. The role of the variable D is to stop the process $\tilde{N}$ but, on the opposite of Ghosh and Lin [2003] for instance, we are not interested in modeling its distribution. Note also that no assumptions regarding the dependence between N^* and D are made. Our setting is similar to the one of Ghosh and Lin [2000], Dauxois and Sencey [2009] or Bouaziz et al. [2013].

We now present the two different models for μ that are studied throughout this paper.

Model 1 : parametric case.

$$\mu(t|z) = \mu_0(t, z; \theta_0), \tag{2.1}$$

where θ_0 is unknown in some parameter space $\Theta \subset \mathbb{R}^{d'}$ (where d' may be different from d) and μ_0 is a known function.

Model 2 : semiparametric case.

$$\mu(t|z) = \mu_{\theta_0}(t, \theta_0' z), \tag{2.2}$$

where θ_0 is unknown in some parameter space $\Theta \subset \mathbb{R}^d$, where
$\mu_\theta(t, u) = E[N^*(t)|\theta' Z = u]$ and where the family of functions $\mathcal{F} = \{\mu_\theta : \theta \in \Theta\}$ is unknown. We impose that the first component of θ_0 is 1 to identify this parameter. Indeed, if we do not make any assumption on θ_0, this parameter would only be defined up to a multiplicative constant. Note that another equivalent condition would consist of imposing that θ_0 is of norm 1 for any given norm on $\mathbb{R}^d$. Model 2 is a single-index model (see e.g. Härdle et al. [1993]), since it consists of assuming that $\mu(t|z)$ depends on the covariates only through a linear combination $\theta_0' z$. In the notation that we introduce in (2.2) and that we will use throughout this paper, we emphasize the fact that the function μ depends on θ in two different ways. The notation μ_θ indicates that, for two different values of θ, the conditional distribution that one is considering is not the same. Indeed, the distribution of $N^*(\cdot)$ conditionally on $\theta_1' Z$ is, in general, not the same as the distribution of $N^*(\cdot)$ conditionally to $\theta_2' Z$ for two different vectors θ_1 and θ_2. Moreover, this function $\mu_{\theta_0}(t, u)$ is evaluated at the point $\theta_0' Z$. This distinction is essential to obtain some crucial properties of single-index models, and explains why the vector of partial derivatives with respect to θ has a relatively complex form (see Lemma 5 in Supplementary material). To ensure

consistency of single-index approaches, a continuity property of the map $\theta \to \mu_\theta(\cdot,\cdot)$ (which is a map from Θ towards some space of functions) will be required. To illustrate the notations, consider the particular case where

$$E[N^*(t)|Z] = (\theta_0' Z + 5)\mu_0(t)$$

for some function $\mu_0(\cdot)$. A quick computation shows that

$$\begin{aligned} E\left[N^*(t)|\theta' Z\right] &= E\left[E[N^*(t)|Z]|\theta' Z\right] \\ &= (\theta_0' E[Z|\theta' Z] + 5)\mu_0(t), \end{aligned}$$

which gives

$$\mu_\theta(t,u) = (\theta_0' E[Z|\theta' Z = u] + 5)\mu_0(t).$$

The appealing feature of Model 1 stands in the simplicity of the regression function. However, like every parametric procedure, it relies on strong assumptions which have few chances to hold in practice. On the opposite, a fully nonparametric procedure requires fewer assumptions but suffers from the so-called "curse of dimensionality" when the number of covariates is high. Therefore, Model 2 appears as a good compromise between the parametric approach and the nonparametric one. Indeed it is more flexible than a fully parametric one but is not stroke by the curse of dimensionality since it relies on a dimension reduction assumption. Moreover, Model 2 can be seen as a generalization of the models exposed in equations (1.1) to (1.3).

One does not generally observe $N^*(\cdot)$ on the whole time interval $[0, D]$ because the random variable D is subject to right-censoring. Let C be a positive random variable standing for the censoring time. The observation time T is then given by $T = D \wedge C$. Hence, instead of observing $N^*(t)$ for $t \in [0, D]$, one only observes $N(t) = N^*(t \wedge C)$ for $t \in [0, D]$. Letting $\delta = I(D \leq C)$, the observations consist of n i.i.d. replications $(T_i, \delta_i, Z_i, N_i(\cdot))_{1 \leq i \leq n}$ of $(T, \delta, Z, N(\cdot))$. Let us introduce the cumulative distribution functions of the observed variables in the censored data model:

$$\begin{cases} H(t) = P(T \leq t), \\ F(t) = P(D \leq t), \\ G(t) = P(C \leq t). \end{cases}$$

We also define $\tau_H = \inf\{t : H(t) = 1\}$ the right endpoint of the support of the random variable T. In the sequel, we need the two following assumptions to identify these

distribution functions.

Assumption 1. *For a counting process L let us denote $dL(t) = L(t) - L(t-)$ (where $L(t-) = \lim_{u \to t, u<t} L(u)$) the jump of process L at time t. Assume that*

$$\begin{cases} P\big(dN^*(C) \neq 0\big) = 0, \\ P(D = C) = 0. \end{cases}$$

This is a common assumption in the context of recurrent events which prevents us from ties between the occurrence times of death, censoring and recurrent events.

Assumption 2. *We write down $A \perp\!\!\!\perp B$ when two random variables A and B are independent. Assume that*

$$\begin{cases} C \perp\!\!\!\perp (N^*(\cdot), D), \\ P(C \leq t | N^*(\cdot), Z, D) = P(C \leq t | N^*(\cdot), D) \text{ for } t \in [0, \tau_H]. \end{cases}$$

Assumption 2 holds in the particular case where C is independent of $(N^*(\cdot), D, Z)$ but is slightly more general since it does not require the independence between C and Z. Similar assumptions are often considered in the literature on the Kaplan-Meier estimator for the survival distribution function with covariates. To study the Kaplan Meier estimator, Stute [1993] assumed that $P(C \leq t|Z, D) = P(C \leq t|D)$ and $C \perp\!\!\!\perp D$ (see the discussion in Stute [1993] for this assumption). Our Assumption 2 is the natural extension of the assumption proposed by Stute [1993], but now in the presence of a recurrent event process N^*.

Alternatively, one could assume that $C \perp\!\!\!\perp (N^*(\cdot), D)$ conditionally on Z. Using such an assumption instead of Assumption 2 would require to modify the approach described below, by replacing the Kaplan-Meier estimator of the distribution of C by a conditional Kaplan-Meier estimator as the one proposed by Beran [1981] and studied by Dabrowska [1987]. In the following, we do not focus on the theoretical behaviour of this modification, which is left to future research.

2.2 Estimation procedure

2.2.1 Heuristics for the rescaled process and the weighting measure

Our objective is to estimate $E[N^*(t)|Z]$ successively under Model 1 and Model 2 from the i.i.d. sample $(T_i, \delta_i, Z_i, N_i(\cdot))_{1 \leq i \leq n}$. In our regression framework, going back to the definition of the conditional expectation, it is quite natural to perform estimation of $E[N^*(t)|Z]$ using minimization of a least-squares-type criterion both in Model 1 and Model 2. With this method in mind, consider a least-squares criterion which is integrated over $[0, \tau_H)$ to control the trajectory of the process of interest over this time interval.

This gives the following criterion, say under Model 1,

$$\int_0^{\tau_H} E\left[(\mu_0(t,Z;\theta) - N^*(t))^2\right] dt$$

which is to be minimized with respect to θ. One of the difficulties we face when estimating the conditional expectation of $N^*(t)$ is that the process $N^*(\cdot)$ is not directly observed because of censoring. Hence, empirical versions of criteria like the above one can not be computed since they rely on $N^*(\cdot)$. To circumvent this difficulty we introduce a rescaled process $Y(\cdot)$ which is designed to compensate the censoring effects. We define for any t in $[0, \tau_H)$

$$Y(t) = \int_0^t \frac{dN(s)}{1 - G(s-)}. \tag{2.3}$$

The logic behind this rescaled process is similar to the approach used by Leurgans [1987] in a censored regression framework. In the definition (2.3), the denominator is decreasing when s grows to infinity. This means that we give more weight to the events we observe when s is large and compensate for the lack of observations due to censoring for large s. To go further in the definition of our least-squares criterion, notice that under Assumptions 1 and 2, we have for any s in $[0, \tau_H)$

$$\begin{aligned} E[dN(s)|Z] &= E[dN^*(s \wedge C)|Z] \\ &= E[dN^*(s) I(s \leq C)|Z] \\ &= E[dN^*(s)|Z](1 - G(s-)) \end{aligned} \tag{2.4}$$

so that

$$E[Y(t)|Z] = E[N^*(t)|Z].$$

The consequence is that we can now consider a modified least-squares criterion based on the estimated rescaled process $Y(\cdot)$, say again under Model 1,

$$\int_0^{\tau_H} E\left[(\mu_0(t,Z;\theta) - Y(t))^2\right] dt,$$

that is the integrated squared error (see for example Bowman [1984] in the context of cross validation for kernel estimators). The other difficulty we have to face is that we have to ensure the finiteness of our least-squares criterion which is not guaranteed with the above definition. To circumvent this other difficulty, we will use a weighting measure w specifically designed to ensure the finiteness of our criterion so that our criterion to be minimized will be of the form

$$\int_0^{\tau_H} E\left[(\mu_0(t,Z;\theta) - Y(t))^2\right] dw(t).$$

We now consider in details separately the parametric case and the semiparametric case in the two subsections to come.

2.2.2 The parametric case

Suppose that Model 1 is satisfied. Let w denote a measure such that $w\big([0,\infty)\big) < \infty$ and such that the quantity

$$M_w(\theta,\mu_0) = \int_0^{\tau_H} E\big[\mu_0(t,Z;\theta)^2\big]dw(t) - 2\int_0^{\tau_H} E\big[Y(t)\mu_0(t,Z;\theta)\big]dw(t)$$

is finite. Let us notice that $\mu_0(\cdot,\cdot;\theta)$ and $Y(\cdot)$ may tend to infinity when $t \to \tau_H$. This remark leads us to introduce a supplementary condition on w distinct from the fact that this measure has a finite total mass. The true parameter value θ_0 satisfies

$$\theta_0 = \arg\min_{\theta\in\Theta} M_w(\theta,\mu_0) \tag{2.5}$$

regardless of the choice of w. To estimate θ_0, it is natural to replace the function M_w by an empirical version. However, the rescaled process $Y(\cdot)$ can not be computed in practice since it relies on the distribution function G which is usually unknown. To circumvent this other difficulty, we introduce an empirical counterpart of $Y(\cdot)$. Let $T_{(n)}$ denote the last order statistic of the sample $(T_i)_{i=1,\ldots,n}$. The distribution function G can be consistently estimated on $[0,T_{(n)}]$ by the Kaplan-Meier estimator of G denoted by $\hat{G}$ and given for t in $[0,T_{(n)}]$ by

$$\hat{G}(t) = 1 - \prod_{i:T_i\le t}\left(1 - \frac{1}{\sum_{j=1}^n I(T_j \ge T_i)}\right)^{1-\delta_i}.$$

Consequently, the process $Y(\cdot)$ can itself be estimated for t in $[0,T_{(n)}]$ by

$$\hat{Y}(t) = \int_0^t \frac{dN(s)}{1-\hat{G}(s-)}. \tag{2.6}$$

The empirical version of M_w considered here is then

$$M_{n,w}(\theta,\mu_0) = \frac{1}{n}\sum_{i=1}^n \int_0^{T_{(n)}} \mu_0(t,Z_i;\theta)^2 dw(t) - \frac{2}{n}\sum_{i=1}^n \int_0^{T_{(n)}} \hat{Y}_i(t)\mu_0(t,Z;\theta)dw(t).$$

This allows us to define an estimator of θ_0 as

$$\hat{\theta}(w) = \arg\min_{\theta\in\Theta} M_{n,w}(\theta,\mu_0). \tag{2.7}$$

In the above definition, we emphasize the fact that this estimator depends on the choice of the measure w. This measure w is an important feature of our procedure. First, in some situations, the statistician may wish to give more weight to some time intervals

which are of higher importance. Moreover, the measure w is also useful to control the rescaled process. Indeed, in Equation (2.6), the denominator goes to zero when s grows large and w can be precisely designed to avoid the practical problems caused by these too small denominators. Therefore, the finite sample behaviour of our estimation procedure strongly relies on a wise choice of the measure w.

The asymptotic results derived in Section 3 allow us to obtain asymptotic representations of $\hat{\theta}(w)$ as a process indexed by w which hold uniformly in $w \in \mathcal{W}$ where $\mathcal{W}$ is a set of measures in which the statistician plans to choose w. We discuss in Section 3.4 the adaptive choice of w.

2.2.3 The semiparametric case

Suppose that Model 2 is satisfied. In the semiparametric case, the family of functions μ_θ is unknown. However, the criterion used for the parametric case can be slightly modified to estimate θ_0. We can write

$$\theta_0 = \arg\min_{\theta\in\Theta} M_w(\theta, \mu_\theta),$$

where

$$M_w(\theta, \mu_\theta) = \int_0^{\tau_H} E\big[\mu_\theta(t, \theta' Z)^2\big] dw(t) - 2\int_0^{\tau_H} E\big[Y(t)\mu_\theta(t, \theta' Z)\big] dw(t)$$

and where w is now chosen such that $M_w(\theta, \mu_\theta) < \infty$ in addition to having finite total mass.

Using a family of nonparametric estimators $\hat{\mu}_\theta$ of μ_θ, we define the estimator of θ_0 as

$$\hat{\theta}(w) = \arg\min_{\theta\in\Theta} M_{n,w}(\theta, \hat{\mu}_\theta), \tag{2.8}$$

where

$$M_{n,w}(\theta, \hat{\mu}_\theta) = n^{-1}\sum_{i=1}^{n}\int_0^{T_{(n)}} \hat{\mu}_\theta(t, \theta' Z_i)^2 dw(t) - 2n^{-1}\sum_{i=1}^{n}\int_0^{T_{(n)}} \hat{Y}_i(t)\hat{\mu}_\theta(t, \theta' Z_i) dw(t).$$

We give indications in Section 2.2.4 on how to perform the minimization of such a contrast in practice. In Section 3.3, we derive an asymptotic representation of $\hat{\theta}(w)$ (see Theorem 3) regardless of the type of nonparametric estimators $\hat{\mu}_\theta$ used in the computation and provided these nonparametric estimators satisfy a list of uniform convergence conditions. Nevertheless, let us give a precise example of $\hat{\mu}_\theta$ using kernel estimators. The convergence properties of this type of estimator are derived in Section 6.3.

Using the same arguments as in (2.4), we have from the identifiability Assumptions 1 and 2,

$$\mu_\theta(t, u) = \int_0^t \frac{E[dN(s)|\theta' Z = u]}{1 - G(s-)}. \tag{2.9}$$

We estimate the numerator in (2.9) using a kernel estimator and the denominator by the Kaplan-Meier estimator $\hat{G}$, leading to

$$\hat{\mu}_{\theta,h}(t,u) = \int_0^t \frac{\sum_{i=1}^n K\left(\frac{\theta' Z_i - u}{h}\right) dN_i(s)}{\sum_{j=1}^n K\left(\frac{\theta' Z_j - u}{h}\right)\left(1-\hat{G}(s-)\right)}, \tag{2.10}$$

where K is a kernel function and h a bandwidth sequence going to zero. In Section 6.3, we list some conditions on K and h. How to choose the bandwidth from the data in practice is considered at the end of Section 3.6.

It is important to mention that, in this semiparametric approach, knowledge of the family of functions $\{\mu_\theta : \theta \in \Theta\}$ is never required for computing the estimator, since these functions are replaced by nonparametric estimators. This family of functions will only appear in the theoretical validation of the procedure, as the limit of the estimators $\hat{\mu}_\theta$.

2.2.4 Minimization of the contrast (2.8)

The contrast $M_{n,w}(\theta, \hat{\mu}_\theta)$ can be tricky to minimize in practice, since it depends on nonparametric estimators. A first possibility consists of using iterative algorithms (see e.g. Xia et al. [2002]). Another possibility is to use a direct maximization as the one described by Härdle et al. [1993] or Delecroix et al. [2006] in the case of single-index mean regression. This technique is particularly suited to the use of kernel estimators, which depend on a bandwidth parameter h (as the ones described in Section 3.6 and used in Section 4). To emphasize this dependence, we will use the notation $\hat{\mu}_{\theta,h}$.

In this case, the function $M_{n,w}(\theta, \hat{\mu}_{\theta,h})$ can be seen as a function of both θ and h. Härdle et al. [1993] proposed to choose jointly $\hat{\theta}$ and an adaptive bandwidth $\hat{h}$ by taking $(\hat{\theta}, \hat{h}) = \arg\min_{\theta\in\Theta, h\in\mathcal{H}} M_{n,w}(\theta, \hat{\mu}_{\theta,h})$, where $\mathcal{H}$ denotes a set of bandwidths among which one wishes to select the most appropriate. In Section 4, we use a finite grid of bandwidths $\mathcal{H}$, so that minimizing this contrast with respect to h does not raise any additional technical issue. In the case of mean-regression, Härdle et al. [1993] have shown that $\hat{h}$ obtained using this technique is asymptotically equivalent to the (uncomputable) h^* obtained using cross-validation if we had an exact knowledge of the nonparametric part. The same result is proved in our context, in Section 3.6.

Nevertheless, the question of initializing the minimization algorithm with the proper starting point is more delicate. In practice, one may use the average derivative technique, see Powell et al. [1989] in the case of mean-regression. The main advantage of this technique is that it produces closed formulas to compute an estimator of the index θ_0. Therefore, the average derivative technique is often used in order to provide starting points (see e.g. Delecroix et al. [2003]). Denoting $\nabla_z \mu(t|z)$ the vector of partial derivatives of μ with respect to z, and $\mu'_{\theta_0}(t,u) = \partial_u \mu_{\theta_0}(t,u)$, it follows from the single-index assumption that $\nabla_z \mu(t|z) = \theta_0 \mu'_{\theta_0}(t, \theta_0' z)$. Hence, $\int_0^{\tau_H} \nabla_z \mu(t|z) dw(t)$ is colinear to θ_0. This quantity can be estimated by $\int_0^{T_{(n)}} \nabla_z \tilde{\mu}(t|z) dw(t)$, where $\tilde{\mu}$ is a nonparametric estimator

of $\mu(t|z)$ (which does not take the single-index assumption into account). Computing

$$\beta_n = \frac{1}{n}\sum_{i=1}^{n}\int_0^{T_{(n)}} \nabla_z \tilde{\mu}(t|Z_i)dw(t),$$

this quantity should be close to be colinear to θ_0 provided that $\tilde{\mu}$ is consistent. Then, one can compute $\hat{\theta}_{prel} = \beta_n/\beta_{n,1}$, where $\beta_{n,1}$ denotes the first component of β_n. This preliminary estimator $\hat{\theta}_{prel}$, has a first component equal to one, as required.

As an estimator $\tilde{\mu}$, one can use

$$\tilde{\mu}(t|z) = \int_0^t \frac{\sum_{i=1}^n \tilde{K}\left(\frac{Z_i - z}{h_0}\right) dN_i(s)}{\sum_{j=1}^n \tilde{K}\left(\frac{Z_j - z}{h_0}\right)\left(1 - \hat{G}(s-)\right)},$$

where $\tilde{K}$ is a (multivariate) kernel function, and h_0 a preliminary bandwidth. As every nonparametric estimator in high dimension, the rate of convergence of this estimator may be slow, but it will still be consistent. Moreover, the fact that β_n is computed from a mean of nonparametric estimator improves the quality of this preliminary estimation. To simplify the notations we use the same bandwidth h_0 for each component of Z_j but different bandwidths may be used in practice.

3. Asymptotic results

In this part, we provide asymptotic properties for our estimators. In Section 3.1, we expose our main lemma, which is the keystone of our theoretical results. In the next two sections we give asymptotic representations of $\hat{\theta}(w)$ for the parametric and semiparametric models. We then discuss the adaptive choice of the measure w in order to improve the performance of our procedure in Section 3.4. The variance of the limiting process is estimated in Section 3.5 and the choice of the bandwidth h in (2.10) is highlighted in Section 3.6.

All these results are presented for both models 1 and 2 and a large class of measures $\mathcal{W}$. The technical assumptions needed for the estimation procedures are listed in Section 6.1. In particular, our results hold true if we consider a class $\{\mu_\theta(\cdot,\cdot), \theta \in \Theta\}$ (or $\{\mu_0(\cdot,\cdot;\theta), \theta \in \Theta\}$ in the parametric case) of polynomial functions and if we take $\mathcal{W}$ as a set of piecewise constant bounded measures with a finite number of jumps. However, notice that the assumptions presented in Section 6.1 are more general and correspond to a larger area of practical situations.

3.1 The main lemma

From a theoretical viewpoint, the main issue stands in studying the difference between Y and its estimated version. The following lemma provides an asymptotic representation for a class of empirical sums in which the process $\hat{Y}$ is involved.

Such kind of asymptotic representations have become very valuable tools for inference

in survival analysis, since they allow to transform a non i.i.d. quantity into an other one that can be easily studied using the central limit theorem. See e.g. Stute [1995], Van Keilegom and Akritas [1999], Sánchez Sellero et al. [2005] or Lopez [2009] for some similar results in other frameworks.

Lemma 1. *Let $\mathcal{F}$ be a class of functions with bounded envelope $\bar{F}$ satisfying Property 1 and assume that Assumptions 3 and 4 hold. Define, for any function $f \in \mathcal{F}$,*

$$S_n(f,w) = \frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau_H} Y_i(t) f(t,Z_i) dw(t)$$

and

$$\hat{S}_n(f,w) = \frac{1}{n}\sum_{i=1}^{n}\int_0^{T_{(n)}} \hat{Y}_i(t) f(t,Z_i) dw(t).$$

(1) Assume that $\sup_{w\in\mathcal{W}} E[S_n(\bar{F},w)] < \infty$. Let

$$\mathcal{I}_w(T_i,\delta_i,f) = \int_0^{\tau_H}\int_0^t \eta_{s-}(T_i,\delta_i) E[f(t,Z) d\mu(s|Z)] dw(t)$$

where

$$d\mu(s|Z) = \frac{\partial \mu(s|Z)}{\partial s} ds$$
$$\eta_t(T,\delta) = \frac{(1-\delta) I(T\le t)}{1-H(T-)} - \int_0^t \frac{I(T\ge s) dG(s)}{[1-H(s-)][1-G(s-)]}$$

Then, for all $f \in \mathcal{F}$,

$$\hat{S}_n(f,w) - S_n(f,w) = \frac{1}{n}\sum_{i=1}^{n} \mathcal{I}_w(T_i,\delta_i,f) + R_n(f,w),$$

where

$$\sup_{w\in\mathcal{W}, f\in\mathcal{F}} |R_n(f,w)| = o_P(n^{-1/2}).$$

Moreover, if the measures w are all supported in $[0,\tau]$ for some $\tau < \tau_H$, then

$$\sup_{w\in\mathcal{W}, f\in\mathcal{F}} |R_n(f,w)| = O_P(n^{-1}\log n).$$

(2) If $\hat{f}$ denotes a family of nonparametric estimators of functions $f \in \mathcal{F}$ satisfying $\sup_{f\in\mathcal{F}} \|\hat{f} - f\|_\infty = o_P(1)$, then

$$\sup_{w\in\mathcal{W}} |\hat{S}_n(\hat{f},w) - \hat{S}_n(f,w)| = o_P(n^{-1/2}).$$

Moreover, if the measures w are all supported in $[0,\tau]$ for some $\tau < \tau_H$, then

$$\sup_{w\in\mathcal{W}} |\hat{S}_n(\hat{f},w) - \hat{S}_n(f,w)| = O_P(n^{-1}\log n).$$

The proof is postponed to Section 6.2. With the estimated rescaled process $\hat{Y}$ at hand, we can now propose $M-$estimation procedures to estimate the regression function in both the parametric and semiparametric cases.

3.2 Asymptotic normality of $\hat{\theta}$ in the parametric case

Let $\Longrightarrow$ denote the weak convergence.

Theorem 2. *Assume that (2.1) holds. Under Assumptions 1 to 5, the estimator in (2.7) admits the following asymptotic representation*

$$\hat{\theta}(w) - \theta_0 = \Sigma_{w,p}^{-1}\left\{\frac{1}{n}\sum_{i=1}^{n}\left(\int_0^{\tau_H}[Y_i(t) - \mu_0(t,Z_i;\theta_0)]\nabla_\theta\mu_0(t,Z_i;\theta_0)dw(t) + \int_0^{\tau_H}\int_0^t \eta_{s-}(T_i,\delta_i)E[\nabla_\theta\mu_0(t,Z;\theta_0)d\mu_0(s,Z;\theta_0)]dw(t)\right)\right\} + R_n(w),$$

where $\sup_{w\in\mathcal{W}}\|R_n(w)\| = o_P(n^{-1/2})$. *As a consequence, for any* $w\in\mathcal{W}$,

$$\sqrt{n}\big(\hat{\theta}(w) - \theta_0\big) \Longrightarrow \mathcal{N}(0, V_{w,p}),$$

where $V_{w,p} = \Sigma_{w,p}^{-1}\Delta_{w,p}\Sigma_{w,p}^{-1}$ *with*

$$\Sigma_{w,p} = \nabla_\theta^2 M_w(\theta_0,\mu_0),$$
$$\Delta_{w,p} = E\left[\mathcal{I}_w(T,\delta,\nabla_\theta\mu_0(\cdot,\cdot;\theta_0))\mathcal{I}_w(T,\delta,\nabla_\theta\mu_0(\cdot,\cdot;\theta_0))'\right].$$

Proof. Write

$$M_{n,w}(\theta,\mu_0) = -2\hat{S}_n(\mu_0(\cdot,\cdot;\theta),w) + n^{-1}\sum_{i=1}^{n}\int_0^{T_{(n)}}\mu_0(t,Z_i;\theta)^2dw(t). \tag{3.1}$$

Then, using the asymptotic representation of Lemma 1, one gets

$$\begin{aligned} M_{n,w}(\theta,\mu_0) &= -\frac{2}{n}\sum_{i=1}^{n}\int_0^{T_{(n)}} Y_i(t)\mu_0(t,Z_i;\theta)dw(t) \\ &\quad + \frac{1}{n}\sum_{i=1}^{n}\int_0^{T_{(n)}}\mu_0(t,Z_i;\theta)^2dw(t) \\ &\quad - \frac{2}{n}\sum_{i=1}^{n}\int_0^{\tau_H}\int_0^t \eta_{s-}(T_i,\delta_i)E[\mu_0(t,Z;\theta)d\mu(s|Z)]dw(t) + R_n(\mu_0(\cdot,\cdot;\theta),w) \\ &= \mathcal{T}_1(\theta,w) + \mathcal{T}_2(\theta,w) + \mathcal{T}_3(\theta,w) + R_n(\mu_0(\cdot,\cdot;\theta),w), \end{aligned}$$

where $\sup_{\theta,w} |R_n(\mu_0(\cdot,\cdot,\theta),w)| = o_P(1)$, since Assumption 5 ensures that $\{\mu_0(\cdot,\cdot;\theta), \theta \in \Theta\}$ satisfies the conditions of Lemma 1. The sum of the first two terms $\mathcal{T}_1(\theta,w)+\mathcal{T}_2(\theta,w)$ converges towards $M_w(\theta,\mu_0)$ from the law of large numbers. Moreover, Assumption 5 ensures that this convergence is uniform with respect to θ and w. On the other hand, the expectation of $\mathcal{T}_3(\theta,w)$ is 0, and the convergence of $\mathcal{T}_3(\theta,w)$ is uniform with respect to θ and w since it is of the form $n^{-1}\sum_{i=1}^n \int_0^{\tau_H} \psi(\theta,t,T_i,\delta_i,Z_i)dw(t)$, with ψ satisfying a Lipschitz property with respect to θ (this is a consequence of Properties 2 and 3 for $\{\mu_0(\cdot,\cdot;\theta), \theta \in \Theta\}$). Finally, this ensures $\sup_{\theta,w} |M_{n,w}(\theta,\mu_0) - M_w(\theta,\mu_0)| = o_P(1)$ and consequently $\sup_w \|\hat{\theta}(w) - \theta_0\| = o_P(1)$.

To obtain the uniform CLT property for $\hat{\theta}(w)$, use a Taylor expansion of $\nabla_\theta M_{n,w}(\theta,\mu_0)$ around θ_0:

$$\nabla_\theta M_{n,w}(\hat{\theta},\mu_0) = \nabla_\theta M_{n,w}(\theta_0,\mu_0) + \nabla^2_\theta M_{n,w}(\tilde{\theta},\mu_0)(\hat{\theta}-\theta_0), \tag{3.2}$$

for some $\tilde{\theta}$ between $\hat{\theta}$ and θ_0. The left-hand side of (3.2) is zero by definition of $\hat{\theta}$. Moreover, the matrix $\nabla^2_\theta M_{n,w}(\tilde{\theta},\mu_0)$ is almost surely invertible for n large enough under Assumption 5 since $\hat{\theta}$ (and consequently $\tilde{\theta}$) tends to θ_0 almost surely. This leads to

$$\hat{\theta} - \theta_0 = -\nabla^2_\theta M^{-1}_{n,w}(\tilde{\theta},\mu_0)\nabla_\theta M_{n,w}(\theta_0,\mu_0).$$

Write

$$\begin{aligned}\nabla^2_\theta M_{n,w}(\tilde{\theta},\mu_0) = -2\Bigg[\hat{S}_n(\nabla^2_\theta\mu_0(\cdot,\cdot;\tilde{\theta}),w) - \frac{1}{n}\sum_{i=1}^n \int_0^{\tau_H} \Big(\nabla_\theta\mu_0(t,Z_i;\tilde{\theta})\nabla_\theta\mu_0(t,Z_i;\tilde{\theta})' \\ +\mu_0(t,Z_i;\tilde{\theta})\nabla^2_\theta\mu_0(t,Z_i;\tilde{\theta})\Big)\, dw(t)\Bigg] + R_n(\theta,w),\end{aligned}$$

where $R_n(\theta,w)$ comes from the change in the integration bounds of $[0,T_{(n)}]$ by $[0,\tau_H]$ and can be seen to tend uniformly to zero from Lebesgue's dominated convergence since the term inside the integral is bounded. From Lemma 1, the almost sure convergence of $\tilde{\theta}$ and the fact that $\{\nabla^2_\theta\mu_0(\cdot,\cdot,\theta), \theta \in \Theta\}$ satisfies Property 3 (see Assumption 5), we get that $\hat{S}_n(\nabla^2_\theta\mu_0(\cdot,\cdot;\tilde{\theta}),w)$ converges to $\int_0^{\tau_H} E[Y(t)\nabla^2_\theta\mu_0(t,Z;\theta_0)]dw(t)$ uniformly in w. The second part converges uniformly to its expectation over Θ as a consequence of the Glivenko-Cantelli property of classes of functions satisfying Property 3. This shows that

$$\sup_w \|\nabla^2_\theta M^{-1}_{n,w}(\tilde{\theta},\mu_0) - \nabla^2_\theta M^{-1}_w(\theta_0,\mu_0)\| = o_P(1).$$

On the other hand, we write

$$\nabla_\theta M_{n,w}(\theta_0,\mu_0) = -2\left[\hat{S}_n(\nabla_\theta\mu_0(\cdot,\cdot;\theta_0),w) - \frac{1}{n}\sum_{i=1}^n\int_0^{\tau_H}\mu_0(t,Z;\theta_0)\nabla_\theta\mu_0(t,Z;\theta_0)dw(t)\right] + \frac{2}{n}\sum_{i=1}^n\int_{T_{(n)}}^{\tau_H}\mu_0(t,Z;\theta_0)\nabla_\theta\mu_0(t,Z;\theta_0)dw(t).$$

Using Lebesgue's dominated convergence theorem, the last term tends uniformly to zero at a $n^{-1/2}$ rate. Finally, the asymptotic representation follows from Lemma 1. □

3.3 Asymptotic normality of $\hat{\theta}$ in the semiparametric case

Theorem 3. *Assume that (2.2) holds. Under Assumptions 1 to 4 and 6 to 9, the estimator in (2.8) admits the following asymptotic representation*

$$\hat{\theta}(w)-\theta_0 = \Sigma_{w,sp}^{-1}\left\{\frac{1}{n}\sum_{i=1}^n\left(\int_0^{\tau_H}[Y_i(t)-\mu_{\theta_0}(t,\theta_0'Z_i)]\nabla_\theta\mu_{\theta_0}(t,Z_i)dw(t) + \int_0^{\tau_H}\int_0^t\eta_{s-}(T_i,\delta_i)E[\nabla_\theta\mu_{\theta_0}(t,Z)d\mu_{\theta_0}(s,\theta_0'Z)]dw(t)\right)\right\} + R_n(w),$$

where $\sup_{w\in\mathcal{W}}\|R_n(w)\| = o_P(n^{-1/2})$. *As a consequence, for any* $w\in\mathcal{W}$,

$$\sqrt{n}(\hat{\theta}(w)-\theta_0) \Longrightarrow \mathcal{N}(0,V_{w,sp}),$$

where $V_{w,sp} = \Sigma_{w,sp}^{-1}\Delta_{w,sp}\Sigma_{w,sp}^{-1}$ *with*

$$\Sigma_{w,p} = \nabla_\theta^2 M_w(\theta_0,\mu_{\theta_0}),$$
$$\Delta_{w,p} = E\left[\mathcal{I}_w(T,\delta,\nabla_\theta\mu_{\theta_0}(\cdot,\cdot;\theta_0))\mathcal{I}_w(T,\delta,\nabla_\theta\mu_{\theta_0}(\cdot,\cdot;\theta_0))'\right].$$

Proof. The consistency of the preliminary estimator can be proved in the same way as in the proof of Theorem 2. Following the decomposition (3.1), one can show that $M_{n,w}(\theta,\hat{\mu}_\theta) = M_{n,w}(\mu_\theta) + R_n(\theta,w)$, where $\sup_{\theta,w}|R_n(\theta,w)| = o_P(1)$. Indeed, the second part of Lemma 1 allows us to replace $\hat{S}_n(\hat{\mu}_\theta,w)$ by $S_n(\hat{\mu}_\theta,w)$ up to some uniformly negligible remainder term. Next, observe that

$$\frac{1}{n}\sum_{i=1}^n\int_0^{T_{(n)}}\hat{\mu}_\theta(t,\theta'Z_i)^2dw(t) = \frac{1}{n}\sum_{i=1}^n\int_0^{T_{(n)}}\mu_\theta(t,\theta'Z_i)^2dw(t) + \frac{1}{n}\sum_{i=1}^n\int_0^{T_{(n)}}[\hat{\mu}_\theta(t,\theta'Z_i)^2-\mu_\theta(t,\theta'Z_i)^2]dw(t),$$

where the second term goes to zero uniformly in θ and w thanks to the uniform convergence assumptions on $\hat{\mu}_\theta$. This shows that $\sup_{\theta,w}|M_{n,w}(\theta,\hat{\mu}_\theta) - M_{n,w}(\theta,\mu_\theta)| = o_P(1)$

and the uniform convergence of $M_{n,w}(\theta, \mu_\theta)$ is obtained following the path of the proof of Theorem 2.

Asymptotic normality comes from the fact that

$$\hat{\theta} - \theta_0 = -\nabla^2_\theta M^{-1}_{n,w}(\tilde{\theta}, \hat{\mu}_{\tilde{\theta}}) \nabla_\theta M_{n,w}(\theta_0, \hat{\mu}_{\theta_0}).$$

The fact that

$$\sup_w \|\nabla^2_\theta M^{-1}_{n,w}(\tilde{\theta}, \hat{\mu}_{\tilde{\theta}}) - \nabla^2_\theta M^{-1}_w(\theta_0, \mu_{\theta_0})\| = o_P(1)$$

can be shown in the same way as in the proof of Theorem 2 using now the second part of Lemma 1. The big issue consists of proving the asymptotic representation of $\nabla_\theta M_{n,w}(\theta_0, \hat{\mu}_{\theta_0})$. Write

$$\nabla_\theta M_{n,w}(\theta_0, \hat{\mu}_{\theta_0}) = -2 \left[\hat{S}_n(\nabla_\theta \hat{\mu}_{\theta_0}(\cdot, \theta_0' \cdot), w) - \frac{1}{n} \sum_{i=1}^n \int_0^{T_{(n)}} \hat{\mu}_{\theta_0}(t, \theta_0' Z_i) \nabla_\theta \hat{\mu}_{\theta_0}(t, Z_i) dw(t) \right].$$

Using the second part of Lemma 1, this can be rewritten as

$$\begin{aligned}
&\nabla_\theta M_{n,w}(\theta_0, \hat{\mu}_{\theta_0}) \\
&\quad = \nabla_\theta M_{n,w}(\theta_0, \mu_{\theta_0}) \\
&\qquad - \frac{2}{n} \sum_{i=1}^n \int_0^{\tau_H} \bar{\mu}_{\theta_0}(t, \theta_0' Z_i)^{\lambda_1+\lambda_2} \big(\mu_{\theta_0}(t, \theta_0' Z_i) - Y_i(t)\big) \frac{\nabla_\theta \mu_{\theta_0}(t, Z_i) - \nabla_\theta \hat{\mu}_{\theta_0}(t, Z_i)}{\bar{\mu}_{\theta_0}(t, \theta_0' Z_i)^{\lambda_1+\lambda_2}} dw(t) \\
&\qquad + \frac{2}{n} \sum_{i=1}^n \int_0^{\tau_H} \frac{\hat{\mu}_{\theta_0}(t, \theta_0' Z_i) - \mu_{\theta_0}(t, \theta_0' Z_i)}{\bar{\mu}_{\theta_0}(t, \theta_0' Z_i)^{\lambda_1+\lambda_2}} \bar{\mu}_{\theta_0}(t, \theta_0' Z_i)^{\lambda_1+\lambda_2} \nabla_\theta \mu_{\theta_0}(t, Z_i) dw(t) \\
&\qquad + \frac{2}{n} \sum_{i=1}^n \int_0^{\tau_H} \frac{\big(\hat{\mu}_{\theta_0}(t, \theta_0' Z_i) - \mu_{\theta_0}(t, \theta_0' Z_i)\big)\big(\nabla_\theta \hat{\mu}_{\theta_0}(t, Z_i) - \nabla_\theta \mu_{\theta_0}(t, Z_i)\big)}{\bar{\mu}_{\theta_0}(t, \theta_0' Z_i)^{2(\lambda_1+\lambda_2)} \bar{\mu}_{\theta_0}(t, \theta_0' Z_i)^{-2(\lambda_1+\lambda_2)}} dw(t) \\
&\qquad + R_{4n}(w) \\
&\quad = \nabla_\theta M_{n,w}(\theta_0, \mu_{\theta_0}) + R_{1n}(w) + R_{2n}(w) + R_{3n}(w) + R_{4n}(w),
\end{aligned}$$

where $R_{4n}(w)$ comes from Lemma 1 and the change in the integration bound of $[0, T_{(n)}]$ by $[0, \tau_H]$. Using the same arguments as in the proof of Theorem 2, we deduce that $\sup_w \|R_{4n}(w)\| = o_P(n^{-1/2})$. Using the uniform convergence rates of $\hat{\mu}_{\theta_0}$ and of $\nabla_\theta \hat{\mu}_{\theta_0}$, we get straightforwardly that $\sup_w \|R_{3n}(w)\| = o_P(n^{-1/2})$. Using the uniform convergence of $\nabla_\theta \hat{\mu}_{\theta_0}$, we see that the term R_{1n} can be decomposed into

$$R_{1n}(w) = n^{-1} \sum_{i=1}^n \big(f_w(Z_i, Y_i) - f_{n,w}(Z_i, Y_i)\big),$$

where f_w and $f_{n,w}$ both belong (almost surely for n large enough) to the class $\mathcal{G}$ defined in Assumption 9 and with $\sup_w \|f_w - f_{n,w}\|_\infty \to 0$ a.s. Therefore, using the asymp-

totic equicontinuity of the Donsker class $\mathcal{G}$ (see e.g. Section 2.1.2 in Van der Vaart and Wellner [1996]), this shows that

$$\sup_w \|R_{1n}(w) - \int \big(f_w(z,y) - f_{n,w}(z,y)\big) dP_{Z,Y}(z,y)\| = o_P(n^{-1/2}).$$

Moreover, it is clear that $\int (f_w(z,y) - f_{n,w}(z,y)) dP_{Z,Y}(z,y) = 0$ using the fact that $\nabla_\theta \mu_{\theta_0}(t,z) - \nabla_\theta \hat{\mu}_{\theta_0}(t,z)$ is a function of z only and that $E[\mu_{\theta_0}(t,\theta_0' Z_i) - Y_i(t)|Z_i] = 0$.

The term $R_{2n}(w)$ can be handled in the same way using now the Donsker class $\mathcal{H}$ in Assumption 9, observing that $\hat{\mu}_{\theta_0}(t,\theta_0' z) - \mu_{\theta_0}(t,\theta_0' z)$ is a function of $\theta_0' z$ only and getting from Lemma 5 in Supplementary Material, that $E[\nabla_\theta \mu_{\theta_0}(t,Z)|\theta_0' Z] = 0$. □

3.4 Adaptive choice of w

The representations of Theorems 2 and 3 hold uniformly in $w \in \mathcal{W}$. Therefore, consider some data-driven measure $\hat{w}_n \in \mathcal{W}$, tending to some asymptotic measure w_0 in the sense that $\int \phi(t) d\{\hat{w}_n - w_0\}(t)$ tends to 0 in probability for all function ϕ in $L^1(d\hat{w}_n) \cap L^1(dw_0)$. Moreover, assume that

$$\sup_{w \in \mathcal{W}: w \to w_0} \left| n^{-1} \sum_{i=1}^n \int \phi(t; T_i, \delta_i, Z_i, Y_i) d\{w - w_0\}(t) \right| = o_P(n^{-1/2}), \tag{3.3}$$

for any function ϕ with $E[\phi(t; T_i, \delta_i, Z_i, Y_i)] = 0$. This assumption can easily be fulfilled by considering a simple class of measures, ensuring that $\{(T, Z, \delta, Y) \to n^{-1} \sum_{i=1}^n \int \phi(t; T, \delta, Z, Y) dw(t) : w \in \mathcal{W}\}$ is a Donsker class of functions. In this situation, one can deduce that (for example in the semiparametric case),

$$\hat{\theta}(\hat{w}) - \theta_0 = \Sigma_{w,sp}^{-1} \left\{ \frac{1}{n} \sum_{i=1}^n \left(\int_0^{\tau_H} [Y_i(t) - \mu_{\theta_0}(t, \theta_0' Z_i)] \nabla_\theta \mu_{\theta_0}(t, Z_i) dw_0(t) \right.\right.$$
$$\left.\left. + \int_0^{\tau_H} \int_0^t \eta_{s-}(T_i, \delta_i) E[\nabla_\theta \mu_{\theta_0}(t, Z) d\mu_{\theta_0}(s, \theta_0' Z)] dw_0(t) \right) \right\} + R_n'(\hat{w}), \tag{3.4}$$

where $R_n'(\hat{w}) = o_P(n^{-1/2})$. This can be done by combining the asymptotic representation of $\hat{\theta}(\hat{w}) - \hat{\theta}(w_0)$ and of $\hat{\theta}(w_0) - \theta_0$, and using (3.3) by studying the difference between the two main terms.

The idea is to consider an asymptotically optimal measure w_0. For example, suppose that one is searching for an estimator that minimizes $E[\|\hat{\theta} - \theta_0\|^2]$. Denote $V_{w,sp}(i)$ the $i-$th diagonal element of the asymptotic covariance matrix $V_{w,sp}$. In view of our objective, one wishes to consider w_0 as the minimizer among measures $\mathcal{W}$ of $n^{-1} \sum_{i=1}^d V_{w,sp}(i)$. In practice, $\hat{w}$ can be taken as the minimizer over $\mathcal{W}$ of $n^{-1} \sum_{i=1}^d \hat{V}_{w,sp}(i)$, where $\hat{V}_{w,sp}$ denotes an estimated version of the asymptotic covariance matrix. The quantity $n^{-1} \sum_{i=1}^d \hat{V}_{w,sp}(i)$ can be seen as an estimator of the mean-squared error of $\hat{\theta}(w)$. Such an adaptive weight function will converge towards w_0 under mild conditions, provided that there is some continuity over $\mathcal{W}$ of the map $w \to V_{w,sp}$. Then, represen-

tation (3.4) ensures that $\hat{\theta}(\hat{w})$ is asymptotically equivalent to the optimal estimator $\hat{\theta}(w_0)$.

3.5 Estimation of the variance

We show how to estimate the variance in the representation of Theorem 3 and we propose an estimator of the mean squared error of θ_0. Denote by $\xi_{n,w}$ the term between brackets in the representation of Theorem 3 so that

$$\hat{\theta}(w) - \theta_0 = \Sigma_{w,sp}^{-1} \xi_{n,w} + R_n(w),$$

where $\sup_{w \in \mathcal{W}} \|R_n(w)\| = o_P(n^{-1/2})$. The quantity $\xi_{n,w}$ can be estimated in the following way

$$\hat{\xi}_{n,w} = \frac{1}{n} \sum_{i=1}^{n} \hat{\psi}(\delta_i, Z_i, T_i, Y_i; w),$$

where

$$\hat{\psi}(\delta, Z, T, Y; w) = \int_0^{T_{(n)}} \big(Y(t) - \hat{\mu}_{\hat{\theta}}(t, \hat{\theta}' Z)\big) \nabla_\theta \hat{\mu}_{\hat{\theta}}(t, Z) dw(t) + \int_0^{T_{(n)}} \int_0^t \hat{\eta}_{s-}(T, \delta) n^{-1} \sum_{i=1}^{n} \big(\nabla_\theta \hat{\mu}_{\hat{\theta}}(t, Z_i) d\hat{\mu}_{\hat{\theta}}(s, \hat{\theta}' Z_i)\big) dw(t),$$

$$\hat{\eta}_t(T, \delta) = \frac{(1-\delta) I(T \leq t)}{1 - \hat{H}(T-)} - \int_0^t \frac{I(T \geq s) d\hat{G}(s)}{\big(1 - \hat{H}(s-)\big)\big(1 - \hat{G}(s-)\big)}$$

and $\hat{H}$ is the empirical estimator of H.

Therefore, the quantity $\Delta_{w,sp}$ can be estimated by

$$\hat{\Delta}_{w,sp} = \frac{1}{n} \sum_{i=1}^{n} \left(\hat{\psi}(\delta, Z, T, Y; w) - \frac{1}{n} \sum_{i=1}^{n} \hat{\psi}(\delta, Z, T, Y; w) \right)^{\otimes 2},$$

where $\otimes 2$ denotes the product of the matrix with its transpose. To consistently estimate $\Sigma_{w,sp}$, we use

$$\hat{\Sigma}_{w,sp} = \frac{1}{n} \sum_{i=1}^{n} \int_0^{T_{(n)}} \nabla_\theta \hat{\mu}_{\hat{\theta}}(t, Z_i) \nabla_\theta \hat{\mu}_{\hat{\theta}}(t, Z_i)' dw(t).$$

A consistent estimator of $V_{w,sp}$ can then be computed from $\hat{V}_{w,sp} = \hat{\Sigma}_{w,sp}^{-1} \hat{\Delta}_{w,sp} \hat{\Sigma}_{w,sp}^{-1}$. The consistency of $\hat{V}_{w,sp}$ comes from the uniform consistency of the estimator $\hat{H}$ and $\hat{G}$ (and of Kaplan-Meier integrals with respect to $\hat{G}$), from the consistency of $\hat{\theta}$, and from the uniform consistency of the nonparametric estimators of $\hat{\mu}_\theta$ and of its partial derivatives.

3.6 Estimation of the nonparametric part

In the semiparametric model, estimation of the finite dimensional parameter θ_0 is only the first step of the method. With our estimator $\hat{\theta}$ at hand, we wish to estimate the conditional mean function $\mu(t|z)$. Different strategies can be proposed to perform this estimation. For this final estimator, there is no theoretical need to use the same kind of nonparametric estimator as in the computation of $\hat{\theta}$. Proposition 4 below states that, under some convergence assumptions for the nonparametric estimator used in this second step, the asymptotic behavior of the final semiparametric estimator of μ is identical to the asymptotic behavior of a purely nonparametric estimator in the case where θ_0 is exactly known.

Proposition 4. *Let Θ^* be some neighborhood of θ_0, and let $\mathcal{T}$ be a set on which $\sup_{\theta\in\Theta^*,t\in\mathcal{T},z\in\mathcal{Z}} \|\nabla_\theta \mu_{\theta_0}(t,z)\| < \infty$. Let $\hat{\mu}_\theta$ be a family of nonparametric estimators of μ_θ satisfying the assumption*

$$\sup_{\theta\in\Theta^*,t\in\mathcal{T},z\in\mathcal{Z}} \|\nabla_\theta \hat{\mu}_\theta(t,z) - \nabla_\theta \mu_\theta(t,z)\| = o_P(1). \tag{3.5}$$

Then, we have

$$\sup_{t\in\mathcal{T},z\in\mathcal{Z}} |\hat{\mu}_{\hat{\theta}}(t,\hat{\theta}'z) - \hat{\mu}_{\theta_0}(t,\theta_0'z)| = O_P(n^{-1/2}).$$

Proof. This is a direct consequence of a Taylor expansion of $\hat{\mu}_{\hat{\theta}}$ around θ_0. From Theorem 3 we have $\hat{\theta} - \theta_0 = O_P(n^{-1/2})$. Then, the boundedness of $\nabla_\theta \mu_{\theta_0}(t,z)$ and the uniform convergence in assumption (3.5) give the result. □

As explained in Section 2.2.4, we propose to simultaneously select the adaptive bandwidth $\hat{h}$ and the index parameter $\hat{\theta}$ in the following way:

$$(\hat{\theta},\hat{h}) = \underset{\theta\in\Theta,h\in\mathcal{H}}{\arg\min}\, M_{n,w}(\theta,\hat{\mu}_{\theta,h}). \tag{3.6}$$

The uniform in bandwidth consistency of the kernel estimators we use (see Section 6.3) ensures us that $\hat{\theta}$ has the same asymptotic properties as in Theorem 3. On the other hand, Proposition 5 below shows that the adaptive bandwidth $\hat{h}$ is asymptotically equivalent to the bandwidth we could obtain using a classical cross-validation technique in the case where the parameter θ_0 is exactly known.

Proposition 5. *For some positive constants a, b and α, let $\mathcal{H} = [an^{-\alpha}, bn^{-\alpha}]$ be a set of bandwidths satisfying Assumption 10 and let*

$$h_0 = \underset{h\in\mathcal{H}}{\arg\min}\, M_{n,w}(\theta_0,\hat{\mu}_{\theta_0,h}).$$

Under the assumptions of Theorem 3 and provided that $\sup_{h\in\mathcal{H},t\in\mathbb{R}_+,z\in\mathcal{Z}} |\hat{\mu}_{\theta,h}(t,\theta'z) -$

$\mu_{\theta,h}(t,\theta' z)| = o_P(1)$, *we have*

$$\hat{h}/h_0 \to 1 \ a.s.$$

Proof. Define $\phi(h/h_0) = M_{n,w}(\theta_0, \hat{\mu}_{\theta_0,h})$ and $\phi_n(h/h_0) = \arg\min_{\theta\in\Theta} M_{n,w}(\theta, \hat{\mu}_{\theta,h})$. By definition of h_0 and $\hat{h}$ we have $\arg\min_{x\in[a,b]} \phi(x) = 1$ and $\hat{h}/h_0 = \arg\min_{x\in[a,b]} \phi_n(x)$. Now write

$$\phi_n(x) = \phi(x) - \frac{2}{n}\sum_{i=1}^{n}\int_0^{\tau_H} \hat{Y}_i(t)\big(\hat{\mu}_{\theta,xh_0}(t,\theta' Z_i) - \hat{\mu}_{\theta,h_0}(t,\theta' Z_i)\big)dw(t)$$
$$+ \frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau_H} \big(\hat{\mu}_{\theta,xh_0}(t,\theta' Z_i)^2 - \hat{\mu}_{\theta,h_0}(t,\theta' Z_i)^2\big)dw(t) + M_{n,w}(\theta, \hat{\mu}_{\theta,h_0}).$$

Using Lemma 1 and the uniform in bandwidth consistency of $\hat{\mu}_{\theta,h}$, the second and third terms in the decomposition tend to zero uniformly in x. On the other hand, the last term does not depend on x. This shows that $\hat{h}/h_0 \to 1$ a.s. □

4 Simulations

We present here some empirical evidence of the good behavior of our semiparametric estimation procedure for finite sample sizes.

The variables D_i are generated according to a Weibull distribution with parameter $(10, 1.1)$ and we consider 4-dimensional covariates $Z_i \sim \otimes^4\mathcal{U}[1,2]$ for $i = 1,\dots,n$. Conditionally on Z_i, the processes $N_i^*(\cdot)$ are generated in the following way: we first generate homogeneous Poisson processes $\widetilde{N}_i(\cdot)$ with intensity $\theta_0' Z_i + 5$ where $\theta_0 = (1, 1.6, 1.25, 0.7)'$ and then compute $N_i^*(t) = \widetilde{N}_i(t \wedge D_i)$ for $i = 1,\dots,n$. This ensures that

$$E[N_i^*(t)|Z_i] = (\theta_0' Z_i + 5)\int_0^t (1 - F(s-))ds, \quad i = 1,\dots,n,$$

such that condition (2.2) is verified. The censoring distribution is selected to be Weibull with parameters $(4, \lambda)$. Taking $\lambda = 1.38$ or $\lambda = 1$ leads respectively to 30% or 70% of censoring and an average of 11.5 or 10 recurrent events per sample. We decide to estimate μ_θ with the kernel estimator defined in (2.10) with the Epanechnikov kernel. In our simulation study, we emphasize the impact of the two parameters involved in our semiparametric procedure, namely the bandwidth of the nonparametric kernel estimators and the measure w.

First, we consider the case of a fixed bandwidth and show how the adaptive choice of $\hat{w}$ can improve the estimation performance of the parameter θ_0. The kernel estimator is computed using a bandwidth $h_0 = 1.1$. We consider a set of discrete measures $w(\cdot)$ supported on $\mathcal{I} = \{0.1, 0.2, \dots, 1.2\}$. The range of values of $\mathcal{I}$ is chosen accordingly to the range of values of the recurrent event times such that $\mathcal{I}$ is a representative subset of the whole trajectory of $N(\cdot)$. With this choice, for any function f, the integral with respect

Table 1. Biases, variances and MSE of $\tilde{\theta}, \hat{\theta}_{\hat{w},h_0}, \hat{\theta}_{cox}$ and $\hat{\theta}_{\mathrm{A}FT}$ for 30% of censored data

$p = 30\%$	Bias	Variance	MSE
$\tilde{\theta}$	$\begin{pmatrix} 0.0816 \\ 0.0897 \\ 0.0289 \end{pmatrix}$	$\begin{pmatrix} 0.0938 & -0.0122 & -0.0369 \\ -0.0122 & 0.0731 & -0.0211 \\ -0.0369 & -0.0211 & 0.0872 \end{pmatrix}$	0.2697
$\hat{\theta}_{\hat{w},h_0}$	$\begin{pmatrix} 0.0451 \\ 0.0439 \\ 0.0291 \end{pmatrix}$	$\begin{pmatrix} 0.0129 & -0.0034 & -0.0064 \\ -0.0034 & 0.0122 & -0.0054 \\ -0.0064 & -0.0054 & 0.0160 \end{pmatrix}$	0.0459
$\hat{\theta}_{cox}$	$\begin{pmatrix} -1.4615 \\ -1.1464 \\ -0.6454 \end{pmatrix}$	$\begin{pmatrix} 0.0190 & -0.0002 & -0.0011 \\ -0.0002 & 0.0183 & 0.0006 \\ -0.0011 & 0.0006 & 0.0191 \end{pmatrix}$	3.9232
$\hat{\theta}_{\mathrm{A}FT}$	$\begin{pmatrix} -5.7300 \\ -6.4569 \\ -8.2406 \end{pmatrix}$	$\begin{pmatrix} 1.7817 & -0.7809 & -1.3722 \\ -0.7809 & 1.7994 & -0.9062 \\ -1.3722 & -0.9062 & 3.0520 \end{pmatrix}$	149.0650

to w reduces to a finite sum. Indeed, we have

$$\int f(t)dw(t) = \sum_{k\in\mathcal{I}} f(k)w(\{k\}).$$

Moreover, we consider an adaptive choice of the weight w among the following family $\{w_a, a = (a_1, \ldots, a_4)$ with $a_i \in \{0.25, 0.5, 0.75, 1\}, i = 1, \ldots, 4\}$ with

$$w_a(\{k\}) = \begin{cases} 1 & \text{for } k = 0.1, \ldots, 0.8 \\ a_i & \text{for } k = 0.8 + i, i = 1, \ldots, 4 \end{cases}$$

which makes 256 possible choices. The intuition is that our procedure should allocate smaller weights to large values of T_i since the behaviour of the Kaplan-Meier estimator is known to be less effective in this part of the distribution (and contributes significantly to the variance). Our estimator $\hat{\theta}_{\hat{w},h_0}$ is then compared to the estimator $\tilde{\theta}$ obtained for the measure w_0 which puts mass 1 at every point of $\mathcal{I}$. We also compare our estimator $\hat{\theta}_{\hat{w},h_0}$ to the estimators $\hat{\theta}_{cox}$ and $\hat{\theta}_{\mathrm{A}FT}$ obtained respectively under Model (1.1) and Model (1.3) by Andersen and Gill [1982] and Ghosh [2004].

In Tables 1 and 2, we report our results over 1 000 simulations of samples of size 100 for two different rates of censoring ($p = 30\%$ and $p = 70\%$). Recalling that the first component of θ_0 is imposed to be one, we only have to estimate the three other components. For each estimator, the Mean Squared Error (MSE) $E(\|\hat{\theta} - \theta_0\|^2)$ is decomposed into bias and variance.

We also computed the average weights of $\hat{w}$ for the last four points of $\mathcal{I}$. For 30% of censoring, we have: $E[\hat{w}(\{0.9\})] = 0.685, E[\hat{w}(\{1\})] = 0.579, E[\hat{w}(\{1.1\})] = 0.586$ and $E[\hat{w}(\{1.2\})] = 0.569$ and for 70% of censoring, $E[\hat{w}(\{0.9\})] = 0.684, E[\hat{w}(\{1\})] =$

Table 2. Biases, variances and MSE of $\tilde{\theta}, \hat{\theta}_{\hat{w},h_0}, \hat{\theta}_{cox}$ and $\hat{\theta}_{\mathrm{A}FT}$ for 70% of censored data

$p = 70\%$	Bias	Variance	MSE
$\tilde{\theta}$	$\begin{pmatrix} 0.0957 \\ 0.0995 \\ 0.0155 \end{pmatrix}$	$\begin{pmatrix} 0.1507 & -0.0067 & -0.1060 \\ -0.0067 & 0.1083 & -0.0305 \\ -0.1060 & -0.0305 & 0.1782 \end{pmatrix}$	0.4566
$\hat{\theta}_{\hat{w},h_0}$	$\begin{pmatrix} 0.0460 \\ 0.0457 \\ 0.0262 \end{pmatrix}$	$\begin{pmatrix} 0.0179 & -0.0031 & -0.0093 \\ -0.0031 & 0.0141 & -0.0067 \\ -0.0093 & -0.0067 & 0.0210 \end{pmatrix}$	0.0579
$\hat{\theta}_{cox}$	$\begin{pmatrix} -1.4633 \\ -1.1407 \\ -0.6370 \end{pmatrix}$	$\begin{pmatrix} 0.0218 & 0.0000 & -0.0008 \\ 0.0000 & 0.0223 & 0.0004 \\ -0.0008 & 0.0004 & 0.0211 \end{pmatrix}$	3.9132
$\hat{\theta}_{\mathrm{A}FT}$	$\begin{pmatrix} -5.8857 \\ -6.5324 \\ -7.8799 \end{pmatrix}$	$\begin{pmatrix} 1.3032 & -0.6241 & -0.9075 \\ -0.6241 & 1.4428 & -0.7425 \\ -0.9075 & -0.7425 & 2.1692 \end{pmatrix}$	144.3226

$0.606, E[\hat{w}(\{1.1\})] = 0.593$ and $E[\hat{w}(\{1.2\})] = 0.538$. Clearly, choosing the measure from the data improves both the bias and the variance of our estimator. Moreover the weights of $\hat{w}(\{k\})$ get smaller for large values of k, especially when the proportion of censored data is high. Consequently, the adaptive measure seems to have a significant impact on the quality of the estimation of θ_0. Note that even if the Cox estimator has very small variance components, our estimators have significantly smaller biases and MSE than the Cox and AFT estimators. This suggests that the Cox and AFT models do not fit well the data and consequently give really poor estimates of θ_0. Since our model assumptions are less restrictive, this explains why our estimator outperforms the Cox and AFT estimators.

Next, we show how the choice of the parameter h influences the quality of estimation. We consider the fixed measure w_0 which puts the same weights 1 at each point. The bandwidth $\hat{h}$ is chosen adaptively in a regular grid of length 0.05 in the set $[0.2, 1.8]$. The average bandwidths over the 1 000 samples were computed and equal to 1.141 for 30% of censoring and 1.126 for 70% of censoring.

The performance of the resulting estimator $\hat{\theta}_{w_0,\hat{h}}$ presented in Table 3 is then compared with the estimators of the previous tables. We observe significant improvement of its MSE compared to $\tilde{\theta}$. As previously, our estimator outperforms $\hat{\theta}_{cox}$ and $\hat{\theta}_{\mathrm{A}FT}$ in term of bias and MSE. We also see that choosing an adaptive bandwidth with a fixed measure or choosing an adaptive measure with a fixed bandwidth leads to a similar quality of estimation of θ_0. However this is no longer true for a high censoring rate: for 70% of censored data, the MSE of $\hat{\theta}_{\hat{w},h_0}$ is almost 2 times lower than the MSE of $\hat{\theta}_{w_0,\hat{h}}$. This shows that the adaptive measure is well suited to the case of censored data: when a large proportion of recurrent events are censored the adaptive measure can compensate the lack of observations due to censoring and allows us to obtain a very accurate estimation of θ_0.

More simulations results are presented in the Supplementary Material paper with a different setup. We consider a recurrent event process where the number of events in a

time interval has a negative binomial distribution. This entails an increase in the variance estimates but other conclusions are similar.

Table 3. Biases, variances and MSE of $\hat{\theta}_{w_0,\hat{h}}$ for 30% and 70% of censored data

	Bias	Variance	MSE
$\hat{\theta}_{w_0,\hat{h}}$, $p = 30\%$	$\begin{pmatrix} 0.0405 \\ 0.0384 \\ 0.0393 \end{pmatrix}$	$\begin{pmatrix} 0.0143 & -0.009 & -0.0108 \\ -0.009 & 0.0143 & -0.0095 \\ -0.0108 & -0.0095 & 0.0150 \end{pmatrix}$	0.0483
$\hat{\theta}_{w_0,\hat{h}}$, $p = 70\%$	$\begin{pmatrix} 0.0395 \\ 0.0383 \\ 0.0362 \end{pmatrix}$	$\begin{pmatrix} 0.0391 & -0.0228 & -0.0113 \\ -0.0228 & 0.0208 & -0.0094 \\ -0.0113 & -0.0094 & 0.0357 \end{pmatrix}$	0.1099

5 Conclusion

We proposed a new procedure to estimate the conditional cumulative mean function of a recurrent event process. We considered both parametric and semiparametric models for the conditional cumulative mean function. Our semiparametric single-index model can be seen as a generalization of both the Cox model and the accelerated failure time model. Moreover, a new feature of our procedure stands in the measure w involved in our estimators which is designed to prevent us from problems in the tail of the distribution due to the presence of censoring. Then, we proposed a data-driven method to choose this measure adaptively. Our criterion is based on the minimization of the mean squared error for the estimation of θ_0 but our procedure is flexible enough to allow the use of any other criteria more adapted to the context. For example, we could consider a criterion directly based on the error of the estimation of μ.

In this work, we mainly focused on kernel estimators for estimating the nonparametric part of our model, providing methods to choose the smoothing parameter from the data. Nevertheless, all our results are valid for any nonparametric estimator of μ_θ provided it satisfies some convergence properties (see Assumption 7). Hence, other kinds of estimators may be used provided they satisfy these conditions.

6 Appendix

6.1 Exposition and discussion of assumptions

In this section we state the technical assumptions on which the proofs of the results of Section 3 are based. We first present some general properties on a class of functions in order to use empirical processes theory.

Let $\mathcal{F} = \{f : (t,z) \in [0,\tau_H] \times \mathcal{Z} \mapsto f(t,z)\}$ be a class of functions with envelope $\bar{F}$ i.e. such that $|f(t,z)| \leq |\bar{F}(t,z)|$ for every (t,z) and f. Define, for a probability measure Q, the norm $\|\cdot\|_{p,Q}$ as the norm of $L^p(Q)$. The covering number of the class $\mathcal{F}$ for the measure Q denoted by $N(\varepsilon, \mathcal{F}, \|\cdot\|_{p,Q})$ is the smallest number of $L^p(Q)-$balls of radius ε needed to cover the set $\mathcal{F}$. The uniform covering number is defined as $N(\varepsilon, \mathcal{F}, \|\cdot\|_p) =$

$\sup_Q N(\varepsilon\|\bar{F}\|_{p,Q}, \mathcal{F}, \|\cdot\|_{p,Q})$ where the supremum is taken over all probability measures. In what follows, we say that a class of functions $\mathcal{F}$ is a $\|\cdot\|_p - VC-$class of functions if there exists two positive constants γ and c such that $N(\varepsilon, \mathcal{F}, \|\cdot\|_p) \leq c\varepsilon^{-\gamma}$. Moreover, a class of functions for which the uniform law of large number holds true is said to be Glivenko Cantelli and a class of functions for which the uniform central limit theorem holds true is said to be Donsker. We refer the reader to Van der Vaart and Wellner [1996] for more details on these definitions.
A class of functions $\mathcal{F}$ is said to satisfy one of the following properties if the corresponding condition holds.

Property 1. *For a class of functions $\mathcal{F} = \{f : (t,z) \in [0,\tau_H] \times \mathcal{Z} \mapsto f(t,z)\}$ and for any $\tau < \tau_H$, define*

$$\mathcal{F}_\tau = \{f(t,\cdot), t \in [0,\tau]\},$$

which is a set of functions defined on $\mathcal{Z}$. Then, for any $\tau < \tau_H$, $\mathcal{F}_\tau$ is a VC-class of functions.

Property 2. *For a class of functions $\mathcal{F} = \{f : (t,z) \in [0,\tau_H] \times \mathcal{Z} \mapsto f(t,z)\}$, the family of functions defined by $\{(z,y) \mapsto \int_0^{\tau_H} y(t)f(t,z)dw(t), f \in \mathcal{F}, w \in \mathcal{W}\}$ is Glivenko Cantelli.*

In Section 4 in Supplementary material, we give a general type of sufficient conditions to fulfill these properties. It is easy to check that these technical assumptions are verified when the following conditions hold simultaneously:

- $\mathcal{F}$ is a class of polynomial functions $f(t,z)$ (with bounded coefficients),
- $dE[Y(t)] = g(t)dt$ for some polynomial function $g(t)$,
- the class of measures is of the form $\mathcal{W} = \{w : dw(t) = W(t)d\tilde{w}(t)\}$ where $W(t)$ is a decreasing function (of order t^{-k} for k sufficiently high or exponential) and where $\tilde{w}$ belongs to a class of monotone positive uniformly bounded functions sufficiently small (for example, piecewise constant bounded functions with a finite number of jumps).

Property 3. *Let $\mathcal{F} = \{f_\theta : (t,z) \in [0,\tau_H] \times \mathcal{Z} \mapsto f_\theta(t,z), \theta \in \Theta\}$ be a family of functions indexed by θ. For any $f_{\theta_1}, f_{\theta_2} \in \mathcal{F}$ and $z \in \mathcal{Z}$, we have*

$$\sup_{w \in \mathcal{W}} \int_0^{\tau_H} \|f_{\theta_1}(t,z) - f_{\theta_2}(t,z)\| dw(t) \leq c\|\theta_1 - \theta_2\|,$$

where c is a positive constant.

We now introduce the assumptions needed to derive the asymptotic normality of $\hat{\theta}$ in the parametric and semiparametric models.

Assumptions for the parametric model.
In the estimation procedures, we consider integrated versions of the rescaled process with

respect to a measure w belonging to a class of measures $\mathcal{W}$. Detailed comments on this family and its role in the statistical procedure are discussed in Section 3.4. We need the following assumption for this class of measures.

Assumption 3. *Assume there exists some probability measure w_0 and a positive constant c_0 such that, for any $w \in \mathcal{W}$,*

$$\int_t^{\tau_H} dw(s) \leq c_0 W_0(t),$$

where $W_0(t) = \int_t^{\tau_H} dw_0(s)$ can be written as

$$W_0(t) = W_1(t)W_2(t)$$

where W_1 and W_2 are two positive and non-increasing functions satisfying

(1) $\int_0^{\tau_H} W_1^2(t)(1-F(t-))^{-1}(1-G(t-))^{-2}dG(t) < \infty$,
(2) $\int_0^{\tau_H} W_2(t)E[dN^*(t)] < \infty$,
(3) $\lim_{t\to\tau_H} W_2(t) = 0$.

In particular, Assumption 3 holds when all the measures w have their support included in a common compact subspace strictly included in $[0, \tau_H]$. On the other hand, since the function W_1 controls $1-\hat{G}(s-)$ in $\hat{Y}(s)$ for s in the vicinity of the tail of the distribution, Assumption 3 also allows to consider measures w which are supported in the whole interval $[0, \tau_H]$. Taking $W_1(t) = (1-H(t-))^{1/2}(1-G(t-))^{\varepsilon}$ for some $\varepsilon > 0$ would be sufficient to obtain (1). Moreover, in the case where $\tau_H = \infty$, if we suppose that, for $\beta_1 > 0$, we have $E[N^*(t)] \sim \beta_1 t$ when $t \to \infty$, we could take for example $W_2(t) = t^{-\beta_2}$ for $\beta_2 > 1$ to fulfill (2) and (3).

We also need the following Hölder condition on the process N. This is a technical assumption used in the proof of our main lemma.

Assumption 4. *Suppose there exists $\gamma > 0$ such that*

$$E\left[\sup_{t\leq\tau, t'\leq\tau} \frac{|N(t)-N(t')|}{|t-t'|^{\gamma}}\right] < \infty.$$

Let $\nabla_\theta \mu_0(s, z; \theta_1)$ (resp. $\nabla^2_\theta \mu_0(s, z; \theta_1)$) denote the vector of partial derivatives (resp. the Hessian matrix) of $\mu_0(s, z; \theta)$ with respect to all the components of θ evaluated at θ_1. The following assumption can be understood as a regularity assumption on the regression model.

Assumption 5. *Assume that, for all $w \in \mathcal{W}$, the matrix*
$\Sigma_{w,p} = \int_0^{\tau_H} E[\nabla_\theta \mu_0(t, Z, \theta_0)\nabla_\theta \mu_0(t, Z, \theta_0)']dw(t)$ *is invertible. Moreover, assume that the classes of functions $\{\mu_0(\cdot,\cdot;\theta), \theta \in \Theta\}$, $\{\nabla_\theta \mu_0(\cdot,\cdot;\theta), \theta \in \Theta\}$ and $\{\nabla^2_\theta \mu_0(\cdot,\cdot;\theta), \theta \in \Theta\}$ satisfy Properties 1, 2 and 3.*

Additional assumptions for the semiparametric model.

The following assumption is similar to Assumption 5. Here, $\nabla_\theta \mu_{\theta_1}(s,z)$ (resp. $\nabla^2_\theta \mu_{\theta_1}(s,z)$) denotes the vector of partial derivatives (resp. the Hessian matrix) of $\mu_\theta(s,\theta' z)$ with respect to all the components of θ evaluated at θ_1. Note that the gradient vector $\nabla_\theta \mu_{\theta_1}(s,z)$ does not only depend on $\theta' z$ but also depends on the whole vector z. We give an explicit expression of this gradient in Lemma 5, in Supplementary material.

Assumption 6. *Assume that, for all $w \in \mathcal{W}$, the matrix* $\Sigma_{w,sp} = \int_0^{\tau_H} E[\nabla_\theta \mu_{\theta_0}(t,Z) \nabla_\theta \mu_{\theta_0}(t,Z)'] dw(t)$ *is invertible. Moreover, assume that the classes of functions $\{\mu_\theta(\cdot,\cdot), \theta \in \Theta\}$, $\{\nabla_\theta \mu_\theta(\cdot,\cdot), \theta \in \Theta\}$ and $\{\nabla^2_\theta \mu_\theta(\cdot,\cdot), \theta \in \Theta\}$ satisfy Properties 1, 2 and 3.*

This assumption is hard to check in practice, since the family of functions $\{\mu_\theta : \theta \in \Theta\}$ may have a complex form, which may be impossible to determine explicitly without additional assumptions on the model. Nevertheless, such kind of assumptions are commonplace in the single-index literature and cannot easily be removed. Indeed, they ensure some regularity of the map $\theta \to \mu_\theta(\cdot,\cdot)$ with respect to θ. Without such regularity assumptions, performing single-index estimation is hopeless, since any error of estimation of θ_0 will be amplified by the irregularity of the map $\theta \to \mu_\theta(\cdot,\cdot)$. In practice, looking at the stability of the estimated functions $\hat{\mu}_\theta$ for different values of θ may inform if this assumption is likely to hold.

In order to enable a data-driven procedure, we need uniform convergence properties for the nonparametric estimators $\hat{\mu}_\theta$.

Assumption 7. *Define* $\bar{\mu}_\theta(t,u) = \sup(\mu_\theta(t,u), 1)$.

(1) Assume that

$$\sup_{t \le T_{(n)}, \theta \in \Theta, z \in \mathcal{Z}} \left| \frac{\hat{\mu}_\theta(t,\theta' z) - \mu_\theta(t,\theta' z)}{\bar{\mu}_{\theta_0}(t,\theta_0' z)^{\lambda_1+\lambda_2}} \right| = o_P(1),$$

$$\sup_{t \le T_{(n)}, \theta \in \Theta, z \in \mathcal{Z}} \left\| \frac{\nabla_\theta \hat{\mu}_\theta(t,z) - \nabla_\theta \mu_\theta(t,z)}{\bar{\mu}_{\theta_0}(t,\theta_0' z)^{\lambda_1+\lambda_2}} \right\| = o_P(1),$$

$$\sup_{t \le T_{(n)}, \theta \in \Theta, z \in \mathcal{Z}} \left\| \frac{\nabla^2_\theta \hat{\mu}_\theta(t,z) - \nabla^2_\theta \mu_\theta(t,z)}{\bar{\mu}_{\theta_0}(t,\theta_0' z)^{\lambda_1+\lambda_2}} \right\| = o_P(1),$$

where λ_1, λ_2 are such that $\lambda_1 + \lambda_2 \ge 1$.

(2) Assume also that

$$\sup_{t \le T_{(n)}, z \in \mathcal{Z}} \left| \frac{\hat{\mu}_{\theta_0}(t,\theta_0' z) - \mu_{\theta_0}(t,\theta_0' z)}{\bar{\mu}_{\theta_0}(t,\theta_0' z)^{\lambda_1+\lambda_2}} \right| = O_P(\varepsilon_n),$$

$$\sup_{t \le T_{(n)}, z \in \mathcal{Z}} \left\| \frac{\nabla_{\theta_0} \hat{\mu}_{\theta_0}(t,z) - \nabla_\theta \mu_{\theta_0}(t,z)}{\bar{\mu}_{\theta_0}(t,\theta_0' z)^{\lambda_1+\lambda_2}} \right\| = O_P(\varepsilon_n'),$$

where $\varepsilon_n \varepsilon_n' = o_P(n^{-1/2})$.

Assumption 8. *Assume that*

$$\sup_{z\in\mathcal{Z}}\int_0^{\tau_H}\mu_{\theta_0}(t,\theta_0'z)^{2(\lambda_1+\lambda_2)}dw(t)<\infty,$$

where λ_1,λ_2 *were defined in Assumption 7.*

The following assumption is essential to the empirical processes theory used in our proofs. We assume that the nonparametric estimators and μ_{θ_0} belong to some Donsker classes of functions.

Assumption 9. *Assume that there exists some Donsker classes of functions* $\mathcal{G}$ *and* $\mathcal{H}$ *such that for all* $w\in\mathcal{W}$

$$(z,y)\longmapsto\int_0^{\tau_H}\big(\mu_{\theta_0}(t,\theta_0'z)-y(t)\big)\nabla_\theta\mu_{\theta_0}(t,z)dw(t)\in\mathcal{G},$$
$$z\longmapsto\int_0^{\tau_H}\mu_{\theta_0}(t,\theta_0'z)\nabla_\theta\mu_{\theta_0}(t,z)dw(t)\in\mathcal{H}.$$

Moreover, assume that, almost surely for n *large enough,*

$$(z,y)\longmapsto\int_0^{\tau_H}(\mu_{\theta_0}(t,\theta_0'z)-y(t))\nabla_\theta\hat{\mu}_{\theta_0}(t,z)dw(t)\in\mathcal{G},$$
$$z\longmapsto\int_0^{\tau_H}\hat{\mu}_{\theta_0}(t,\theta_0'z)\nabla_\theta\mu_{\theta_0}(t,z)dw(t)\in\mathcal{H}.$$

To give examples of such kind of classes, consider $\mathcal{F}$ and $\mathcal{W}$ as defined in the discussion following Property 2 and suppose, in addition, that the functions $(t,u)\to W_0(t)f(t,u)$ for $f\in\mathcal{F}$ (f is defined on $\mathbb{R}^2$ since $\theta_0'z\in\mathbb{R}$) are twice continuously differentiable with bounded derivatives up to order 2.

Defining $\mathcal{F}'=\{(u,y)\to\int_0^{\tau_H}(f_1(t,u)-y(t))f_2(t,u)dw(t), w\in\mathcal{W}, f_1,f_2\in\mathcal{F}\}$, it follows from the results of Section 4 in Supplementary material and from the decomposition of the gradient vector $\nabla_\theta\mu_{\theta_0}(t,z)$ obtained in Lemma 5 that we can decompose $\int_0^{\tau_H}(\mu_{\theta_0}(t,\theta_0'z)-y(t))\nabla_\theta\hat{\mu}_{\theta_0}(t,z)dw(t)=\phi_1(\theta_0'z,y)+z\phi_2(\theta_0'z,y)$, with ϕ_1 and ϕ_2 in $\mathcal{F}'$. A similar decomposition can be used on $\int_0^{\tau_H}\hat{\mu}_{\theta_0}(t,\theta_0'z)\nabla_\theta\mu_{\theta_0}(t,z)dw(t)$. Hence, we can consider the class of functions $\mathcal{H}=\mathcal{G}=\mathcal{F}'+z\mathcal{F}'$.

6.2 Proof of Lemma 1

Let

$$S_n^{T_{(n)}}(f,w)=\frac{1}{n}\sum_{i=1}^n\int_0^{T_{(n)}}Y_i(t)f(Z_i,t)dw(t).$$

Write

$$\hat{S}_n(f,w) = S_n^{T_{(n)}}(f,w) + \frac{1}{n}\sum_{i=1}^{n}\int_0^{T_{(n)}} f(Z_i,t)\int_0^t \frac{\big(\hat{G}(s-)-G(s-)\big)dN_i(s)}{\big(1-G(s-)\big)\big(1-\hat{G}(s-)\big)}dw(t)$$
$$= S_n^{T_{(n)}}(f,w) + R_n(f,w).$$

Decompose f into its positive and negative parts denoted respectively by f^+ and f^-. The expectations of the two resulting sums $S_n^{T_{(n)}}(f^+,w) - S_n(f^+,w)$ and $S_n^{T_{(n)}}(f^-,w) - S_n(f^-,w)$ go to zero faster than $n^{-1/2}$ using Lebesgue's dominated convergence. This entails that

$$\sup_{f\in\mathcal{F},w\in\mathcal{W}} |S_n^{T_{(n)}}(f,w) - S_n(f,w)| = o_P(n^{-1/2}).$$

Let $\tau < \tau_H$ and define $w_\tau(t) = w(t)I(t \le \tau)$. On $[0,\tau]$, we use the asymptotic i.i.d. expansion of the Kaplan-Meier estimator $\hat{G}$ proposed by Gijbels and Veraverbeke [1991] which can also be deduced from Stute [1995]:

$$\frac{\hat{G}(t)-G(t)}{1-G(t)} = \frac{1}{n}\sum_{j=1}^{n}\eta_t(T_j,\delta_j) + \tilde{R}_n(t),$$

where $\sup_{t\le\tau}|\tilde{R}_n(t)| = O_P(n^{-1}\log n)$ and

$$\eta_t(T,\delta) = \frac{(1-\delta)I(T\le t)}{1-H(T-)} - \int_0^t \frac{I(T\ge s)dG(s)}{\big(1-H(s-)\big)\big(1-G(s-)\big)}.$$

Moreover, recall that $\sup_{t\le\tau}|\hat{G}(t)-G(t)| = O_P(n^{-1/2})$ (see Gill [1983], Theorem 2.1) and that $\sup_{t\le\tau}(1-G(t))(1-\hat{G}(t))^{-1} = O_P(1)$ (see Gill [1983], Lemma 2.6). Then, write

$$R_n(f,w_\tau) = \frac{1}{n^2}\sum_{i,j}\int_0^{T_{(n)}} f(Z_i,t)\int_0^t \frac{\eta_{s-}(T_j,\delta_j)dN_i(s)}{1-G(s-)}dw_\tau(t) + R_n'(f,w_\tau).$$

Using the fact that $\mathcal{F}$ is an uniformly bounded class, that $\int dw_\tau \le c_0$ from Assumption 3 and that $E[N_i(\tau)] < \infty$ for all τ, we deduce that $\sup_{f,w}|R_n'(f,w_\tau)| = O_P(n^{-1})$. The first term in $R_n(f,w_\tau)$ can be rewritten as

$$\frac{1}{n}\sum_{j=1}^{n}\int_0^{T_{(n)}}\int_0^t \eta_{s-}(T_j,\delta_j)E\big[f(Z,t)d\mu(s|Z)\big]dw_\tau(t) + \int\left(\frac{1}{n^2}\sum_{i,j}\psi^{f,t}(Z_i,N_i,T_j,\delta_j)\right)dw_\tau(t),$$

where

$$\psi^{f,t}(Z_i,N_i,T_j,\delta_j) = \int_0^t \eta_{s-}(T_j,\delta_j)\left\{\frac{f(Z_i,t)dN_i(s)}{1-G(s-)} - E\big[f(Z,t)d\mu(s|Z)\big]\right\}.$$

Observe that, with probability tending to one, the upper bound $T_{(n)}$ in the integrals can

be replaced by $\tau < \tau_H$. Let $f, f' \in \mathcal{F}$ and $t, t' \in [0, \tau]$. We have

$$|\psi^{f,t}(Z_i, N_i, T_j, \delta_j) - \psi^{f',t'}(Z_i, N_i, T_j, \delta_j)| \leq c_\tau \Bigg(\|f - f'\|_\infty N_i(\tau) + |t - t'|^\gamma \sup_{t,t' \leq \tau} \frac{N_i(t) - N_i(t')}{|t - t'|^\gamma} \Bigg), \quad (6.1)$$

where $c_\tau < \infty$ and $\gamma > 0$. Let $\mathcal{H}_\tau$ denote the set of all functions $\psi^{f,t}$ when f ranges $\mathcal{F}$ and t ranges $[0, \tau]$. It follows from (6.1) and Assumption 4 that $\mathcal{H}_\tau$ is a $\|\cdot\|_2-$VC-class of functions. From this, using the Glivenko-Cantelli property of $\mathcal{H}_\tau$,

$$\sup_{f,t \leq \tau} \left| \frac{1}{n^2} \sum_{i=1}^n \psi^{f,t}(Z_i, N_i, T_i, \delta_i) \right| = O_P(n^{-1})$$

and

$$\sup_{f,t \leq \tau} \left| \frac{1}{n^2} \sum_{i \neq j} \psi^{f,t}(Z_i, N_i, T_j, \delta_j) \right| = O_P(n^{-1}),$$

since this can be seen as the supremum of a second order degenerate $U-$process indexed by $\mathcal{H}_\tau$ (see Sherman [1994]). This leads to the i.i.d. representation for $\hat{S}_n(f, w_\tau)$ for any $\tau < \tau_H$.
Similarly, write

$$\hat{S}_n(\hat{f}, w_\tau) = S_n^{T_{(n)}}(\hat{f}, w_\tau) + R_n(\hat{f} - f, w_\tau) + R_n(f, w_\tau)$$

and using the fact that $\sup_{f \in \mathcal{F}} \|f - \hat{f}\|_\infty = o_P(1)$ and that $\sup_{t \leq \tau} |\hat{G}(t) - G(t)| = O_P(n^{-1/2})$, we deduce that $\sup_{f,w} |R_n(\hat{f} - f, w_\tau)| = o_P(n^{-1/2})$. The representation for $\hat{S}_n(\hat{f}, w_\tau)$ follows.

Now, we make τ tend to τ_H. Let $\hat{P}_n(f, w) = \hat{S}_n(f, w) - S_n^{T_{(n)}}(f, w)$ and $P_n^\tau(f, w) = \hat{S}_n(f, w_\tau) - S_n^{T_{(n)}}(f, w_\tau)$. Since the class $\mathcal{F}$ is uniformly bounded, we get

$$\begin{aligned} |\hat{P}_n(f, w) - P_n^\tau(f, w)| &\leq \frac{M}{n} \sum_{i=1}^n \int_\tau^{T_{(n)}} \int_0^t \frac{|\hat{G}(s-) - G(s-)|}{\big(1 - G(s-)\big)\big(1 - \hat{G}(s-)\big)} dN_i(s) dw(t) \\ &\leq \frac{M'}{n} \sum_{i=1}^n \int_0^{T_{(n)}} \frac{W_0(s \vee \tau) |\hat{G}(s-) - G(s-)| dN_i(s)}{\big(1 - G(s-)\big)\big(1 - \hat{G}(s-)\big)}, \end{aligned}$$

(with $a \vee b$ denoting the maximum between a and b) where the last inequality is obtained from Fubini's theorem and Assumption 3. From Theorem 1.2 in Gill [1983], Assumption 3 and the fact that $\sup_{t \leq T_{(n)}} (1 - G(t-))(1 - \hat{G}(t-))^{-1} = O_P(1)$ (see again Gill [1983],

we get that

$$|\hat{P}_n(f,w) - P_n^\tau(f,w)| \leq \frac{A_n}{n}\sum_{i=1}^{n}\int_0^{T_{(n)}} \frac{W_2(s\vee\tau)dN_i(s)}{1-G(s-)},$$

where $A_n = O_P(n^{-1/2})$. The result follows from Lemma 6 in Supplementary material.

6.3 Assumptions for the uniform convergence of the nonparametric estimators

In the supplementary material, we show that the kernel estimator $\hat{\mu}_{\theta,h}$ defined by (2.10) satisfies the convergence rates required by Assumption 7. This requires the following assumption on the kernel and the bandwidth.

Assumption 10. *Assume that*

(1) K has a compact support, say $[-1,1]$, $\int_{\mathbb{R}} K(s)ds = 1$ and $\sup_x |K(x)| < \infty$,
(2) K is a twice differentiable and second-order kernel with derivatives of order 0, 1 and 2 of bounded variation,
(3) $\mathcal{K} := \{K\big((x-\cdot)/h\big) : h > 0, x \in \mathbb{R}^d\}$ is a pointwise measurable class of functions,
(4) $h \in \mathcal{H}_n \subset [an^{-\alpha}, bn^{-\alpha}]$ with $a, b > 0$ and $\alpha \in (1/8, 1/5)$.

From the definition of our estimator, problems arise when studying $\hat{\mu}_{\theta,h}$ for t in the tail of the distribution. This is a common problem when studying Kaplan-Meier estimators but it can be circumvented by some moment conditions on the response and censoring distribution. For instance, in the classical censored framework, Stute [1995] used the function C_G to compensate the bad behavior of the Kaplan Meier estimator in the tail of the distribution. Therefore we also require the following assumption, which gives a similar moment condition but adapted to our recurrent event context.

Assumption 11. *Assume that, for some $\varepsilon > 0$,*

$$\sup_{t,u} \frac{C_G(t)^{7/20+\varepsilon}}{\bar{\mu}_{\theta_0}(t,u)^{\lambda_1}} < \infty$$

and

$$\sup_{t,u} \frac{\int_0^t \big(1-G(s-)\big)E[N^*(s)dN^*(s)]}{\big(1-G(t-)\big)^2 \bar{\mu}_{\theta_0}(t,u)^{2\lambda_2}} < \infty,$$

where λ_1 and λ_2 are defined in Assumption 3.

These conditions allow us to consider a process N^* and variables D and C that are supported on the whole interval $[0, \tau_H]$. However they will hold true only if there is enough information on the recurrent event process in the tails of the distribution. For further illustration take $\mu_{\theta_0}^k(t,u) \sim c_k(1-G(t))^{-\beta_1}$, for $k = 1$ and 2, for t in a neighborhood of τ_H, $u \to \infty$ and where c_1, c_2, β_1 are three positive constants. Take also, for $c_3 > 0$ and

$\beta_2 > 0$, $1 - F(t) \sim c_3(1 - G(t))^{\beta_2}$ for t in a neighborhood of τ_H. Then it can be shown that these conditions are verified for example in the case $\beta_1 > 1$, $\beta_2 = 1$ and $\lambda_1 = \lambda_2 = 1$.

References

[1] Andersen, P. and R. Gill (1982). Cox's regression model for counting processes: a large sample study. *Ann. Statist. 10*, 1100–1120.

[2] Beran, R. (1981). Nonparametric regression with randomly censored survival data. *Technical report, Univ. California, Berkeley*.

[3] Bouaziz, O., F. Comte, and A. Guilloux (2013). Nonparametric estimation of the intensity function of a recurrent event process. *Statistica Sinica 23*(2), 635–665.

[4] Bouaziz, O. and O. Lopez (2010). Conditional density estimation in a censored single-index regression model. *Bernoulli 16*(2), 514–542.

[5] Bowman, A. W. (1984). An alternative method of cross-validation for the smoothing of density estimates. *Biometrika 71*(2), 353–360.

[6] Dabrowska, D. M. (1987). Nonparametric regression with censored survival time data. *Scand. J. Statist. 14*(3), 181–197.

[7] Dauxois, J.-Y. and S. Sencey (2009). Non-parametric tests for recurrent events under competing risks. *Scand. J. Stat. 36*(4), 649–670.

[8] Delecroix, M., W. Härdle, and M. Hristache (2003). Efficient estimation in conditional single-index regression. *J. Multivariate Anal. 86*(2), 213–226.

[9] Delecroix, M., M. Hristache, and V. Patilea (2006). On semiparametric M-estimation in single-index regression. *J. Statist. Plann. Inference 136*(3), 730–769.

[10] Ghosh, D. (2004). Accelerated rates regression models for recurrent failure time data. *Lifetime Data Anal. 10*(3), 247–261.

[11] Ghosh, D. and D. Lin (2000). Nonparametric analysis of recurrent events and death. *Biometrics 56*, 554–562.

[12] Ghosh, D. and D. Lin (2003). Semiparametric analysis of recurrent events in the presence of dependent censoring. *Biometrics 59*, 877–885.

[13] Gijbels, I. and N. Veraverbeke (1991). Almost sure asymptotic representation for a class of functionals of the Kaplan-Meier estimator. *Ann. Statist. 19*(3), 1457–1470.

[14] Gill, R. (1983). Large sample behavior of the product-limit estimator on the whole line. *Ann. Statist. 11*, 49–58.

[15] Härdle, W., P. Hall, and H. Ichimura (1993). Optimal smoothing in single-index models. *Ann. Statist. 21*(1), 157–178.

[16] Huang, C.-Y., J. Qin, and M.-C. Wang (2010). Semiparametric analysis for recurrent event data with time-dependent covariates and informative censoring. *Biometrics 66*(1), 39–49.

[17] Huang, C.-Y. and M.-C. Wang (2004). Joint modeling and estimation for recurrent event processes and failure time data. *J. Amer. Statist. Assoc. 99*(468), 1153–1165.

[18] Ichimura, H. (1993). Semiparametric least squares (SLS) and weighted SLS estimation of single-index models. *J. Econometrics 58*(1-2), 71–120.

[19] Leurgans, S. (1987). Linear models, random censoring and synthetic data. *Biometrika 74*(2), 301–309.

[20] Lin, D. Y., L. J. Wei, and Z. Ying (1998). Accelerated failure time models for counting processes. *Biometrika 85*(3), 605–618.

[21] Lopez, O. (2009). Single-index regression models with right-censored responses. *J. Statist.*

Plann. Inference 139(3), 1082–1097.

[22] Powell, J. L., J. H. Stock, and T. M. Stoker (1989). Semiparametric estimation of index coefficients. *Econometrica 57*(6), 1403–1430.

[23] Sánchez Sellero, C., W. González Manteiga, and I. Van Keilegom (2005). Uniform representation of product-limit integrals with applications. *Scand. J. Statist. 32*(4), 563–581.

[24] Sherman, R. P. (1994). Maximal inequalities for degenerate U-processes with applications to optimization estimators. *Ann. Statist. 22*(1), 439–459.

[25] Stute, W. (1993). Consistent estimation under random censorship when covariables are present. *J. Multivariate Anal. 45*(1), 89–103.

[26] Stute, W. (1995). The central limit theorem under random censorship. *Ann. Statist. 23*(2), 422–439.

[27] Sun, J. and L. J. Wei (2000). Regression analysis of panel count data with covariate-dependent observation and censoring times. *J. R. Stat. Soc. Ser. B Stat. Methodol. 62*(2), 293–302.

[28] Van der Vaart, A. W. and J. A. Wellner (1996). *Weak convergence and empirical processes.* Springer Series in Statistics. New York: Springer-Verlag. With applications to statistics.

[29] Van Keilegom, I. and M. G. Akritas (1999). Transfer of tail information in censored regression models. *Ann. Statist. 27*(5), 1745–1784.

[30] Xia, Y., H. Tong, W. K. Li, and L.-X. Zhu (2002). An adaptive estimation of dimension reduction space. *J. R. Stat. Soc. Ser. B Stat. Methodol. 64*(3), 363–410.