Bernoulli **26**(1), 2020, 557–586
https://doi.org/10.3150/19-BEJ1140

Consistent semiparametric estimators for recurrent event times models with application to virtual age models

ERIC BEUTNER[1], LAURENT BORDES[2] and LAURENT DOYEN[3]

[1]*Vrije Universiteit Amsterdam, Amsterdam, The Netherlands. E-mail: e.a.beutner@vu.nl*
[2]*Laboratoire de Mathématiques et de leurs Applications – IPRA, Univ. Pau and Pays de l'Adour, CNRS, F-64000 Pau, France. E-mail: laurent.bordes@univ-pau.fr*
[3]*Laboratoire Jean Kuntzmann, Univ. Grenoble Alpes, CNRS, F-38000 Grenoble, France. E-mail: laurent.doyen@univ-grenoble-alpes.fr*

Virtual age models are very useful to analyse recurrent events. Among the strengths of these models is their ability to account for treatment (or intervention) effects after an event occurrence. Despite their flexibility for modeling recurrent events, the number of applications is limited. This seems to be a result of the fact that in the semiparametric setting all the existing results assume the virtual age function that describes the treatment (or intervention) effects to be known. This shortcoming can be overcome by considering semiparametric virtual age models with parametrically specified virtual age functions. Yet, fitting such a model is a difficult task. Indeed, it has recently been shown that for these models the standard profile likelihood method fails to lead to consistent estimators. Here we show that consistent estimators can be constructed by smoothing the profile log-likelihood function appropriately. We show that our general result can be applied to most of the relevant virtual age models of the literature. Our approach shows that empirical process techniques may be a worthwhile alternative to martingale methods for studying asymptotic properties of these inference methods. A simulation study is provided to illustrate our consistency results together with an application to real data.

Keywords: effective age process; recurrent event data; semiparametric inference; smoothed profile likelihood; virtual age process

1. Introduction

Virtual age models also called effective age models are useful in understanding the dynamical aspects of recurrent events arising in epidemiology (see, e.g., relapse times of a disease), economics (see, e.g., cycles between economic crises), industry (see, e.g., repair times of a system), climate (occurrence times of extremal climatic events), etc. The strength of these models is that they allow to model explicitly the effect of an intervention or treatment performed just after an event occurrence. Virtual age models have been introduced by Kijima [12,13] and basically they assume that the intensity at time t of a counting process $N(t)=\sum_{j\geq 1}\mathbb{1}_{\{X_j\leq t\}}$, where $X_0=0<X_1<X_2<\cdots$ are the event times, can be written as $(\lambda\circ\varepsilon)(t)$ where λ is a deterministic hazard rate function and ε is (possibly) a random function that may depend on the history of the process or some covariates. When $\varepsilon(t)\equiv t$, the process N is a nonhomogeneous Poisson process with intensity λ, while if $\varepsilon(t) = t - X_{N(t-)}$, the process N is a renewal process. As a consequence we see that the function ε specifies the virtual age or effective age just after an event which has to reflect the effect of the treatment in case of a disease, or the effect of an economic policy, or the efficiency of a maintenance in industry etc. Clearly, in many applications this effect is a priori unknown, and must be estimated based on data. To be able to do so, Doyen and Gaudoin [7] introduced some parametrized versions ε^θ of the effective age function ε where the Euclidean parameter θ measures the efficiency of a medical treatment, an industrial maintenance policy, etc.

In the semiparametric or nonparametric setting, all the existing results assume that the virtual age function is known. For instance, Dorado et al. [6] proposed an estimator of Λ, the cumulative hazard rate function of λ, for a general (yet known) effective age function and studied its asymptotic properties. During the last two decades the model considered by Dorado et al. [6] has been enriched by adding covariates effects and frailties. One of the most complex versions of these models has been proposed by Peña [17] and recent semiparametric estimation and asymptotic results based on the profiled likelihood estimation method (in the sense of Murphy and van der Vaart [16]) have been obtained by Adekpedjou and Stocker [1] and Peña [18]. We also refer to Lindqvist [15] for a review of basic modeling approaches for maintenance models for repairable systems in reliability as well as to Lawless [14] and Cook and Lawless [5] for extensive discussions on the various existing models and inference methods available to deal with recurrent events data. However, as mentioned above, assuming that the effective age function is known limits the applicability of these models. One possibility to overcome this shortcoming is to consider virtual age models for which both the virtual age function and λ are parametrically specified and to apply the usual maximum likelihood approach. Alternatively, one can consider parametrically specified virtual age functions with a nonparametrically specified λ. Constructing consistent estimators for such a model is not straightforward. Beutner et al. [3] recently showed that the standard profile likelihood method fails to lead to consistent estimators. This phenomenon has also been observed for the semiparametric accelerated failure time model where the standard profile likelihood function does not lead to consistent estimators for the unknown Euclidean parameter. To overcome this failure of the standard profile likelihood method, Zeng and Lin [26] showed that profiling out with a smoothed version of the pseudo-estimator of the unknown baseline hazard rate function is enough to obtain consistency and efficiency of the profile likelihood estimator at the price of adding a new parameter required to define the level of regularization. These authors generalized their approach to the case of recurrent events in Zeng and Lin [27].

In this paper, we present a general result that shows that consistent estimators for virtual age models with parametrically specified virtual age function and nonparametrically specified λ can be obtained by smoothing the profiled log-likelihood function appropriately. We apply the general result to a large class of models including most of the relevant virtual age models of the literature like, for instance, the Doyen and Gaudoin [7] Arithmetic Reduction of Age (ARA) models that include Kijima [12] Type-I and Type-II models. To the best of our knowledge, all the asymptotic results obtained so far for these models are based on appropriately adapted martingale methods. Adapting traditional martingale methods is necessary because virtual age models require to switch from the calendar time scale to the effective age scale where martingale properties are not satisfied (see, e.g., Peña et al. [20]); the effective age scale follows an idea introduced by Sellke and Siegmund [22] (see also Sellke [21]). In addition to overcoming the limitation of an a priori known effective age function our approach also shows that empirical process techniques may be a worthwhile alternative to martingale methods for studying the asymptotic properties of these inference methods. Basically our empirical processes are based on independent and identical copies of $\boldsymbol{Z} = (\mathcal{T}, \boldsymbol{X})$ with probability distribution P where $\mathcal{T}$ is a right censoring time and $\boldsymbol{X} = (X_1, X_2, \dots)$ is the nondecreasing sequence of event times. The right censored counting process N is defined by $N(t) = \sum_{j\ge 1} \mathbb{1}_{\{X_j \le t\wedge \mathcal{T}\}}$ for $t \in [0, s]$ where $[0, s]$ is the period of study. Then defining the class of functions $\mathcal{H}_s = \{z = (\tau, x_1, x_2, \dots) \mapsto h_t(z) = \sum_{j\ge 1} \mathbb{1}_{\{x_j \le t\wedge \tau\}}; t \in [0, s]\}$ and using the usual empirical processes notations (see van der Vaart and Wellner [25] or van der Vaart [24]) we can identify $(N(t))_{t\in[0,s]}$ and $(\delta_{\boldsymbol{Z}} h)_{h\in\mathcal{H}_s}$ where $\delta_{\boldsymbol{Z}}$ is the Dirac measure at the random (infinite dimensional) point $\boldsymbol{Z}$.

The paper is organized as follows. In the next section, we define the class of virtual age models and propose a general method of estimation based on a regularized version of the profiled likelihood function. Then general conditions are given under which consistency results are obtained for both the Euclidean and the functional parameter. In Section 3, we show that the general conditions are fulfilled for a large class of semiparametric models with very few assumptions on the model parameters. Identifiability of the model parameters is also discussed under various censoring schemes. A short simulation study is provided in Section 4 to illustrate empirically our consistency results. Section 5 is devoted to estimation of parametric and semiparametric ARA models using reliability data. The final section is devoted to concluding remarks and perspectives.

2. Consistent estimation by profiling and smoothing

2.1. General model and notations

As outlined above the aim is to do inference for a simple counting process $N = \{N(t), t \in [0, s]\}$ where $s \in \mathbb{R}^+$. As usual, N is defined on a measurable space $(\Omega, \mathcal{F})$ endowed with the natural filtration. The jump times or event times of N will be denoted by $\boldsymbol{X} = (X_1, X_2, \dots)$. They are assumed to take their values in $\mathcal{X}$ which denotes the set of increasing sequences of non negative real numbers without accumulation point, that is, $\boldsymbol{x} = (x_n)_{n\in\mathbb{N}} \in \mathcal{X}$ if $0 \le x_1 < x_2 < \cdots$ and $x_n \to +\infty$ as $n \to \infty$. We specify the statistical model for N via its compensator which is assumed to be of the form

$$A(t) = \int_0^t Y(u)\lambda\bigl(\varepsilon^\theta(u)\bigr)\,du$$

with respect to the natural filtration. Here the process Y equals either 0 or 1. The predictable stochastic process ε^θ describes the virtual age. The unknown model parameters are $\lambda \in \mathcal{A}$ where $\mathcal{A}$ is an infinite set of hazard rate functions on $\mathbb{R}^+$ and the Euclidean parameter $\theta \in \Theta$. Regarding the sample paths of ε^θ which describe the possible developments of the virtual age over time we assume that they take values in $\mathcal{S}^\theta = \{e^\theta : \mathbb{R}^+ \to \mathbb{R}^+ | e^\theta(t) = e_0^\theta(t)\mathbb{1}_{[x_0,x_1]}(t) + \sum_{j\ge 2} e_{j-1}^\theta(t)\mathbb{1}_{(x_{j-1},x_j]}(t),$ for some $\boldsymbol{x} \in \mathcal{X}\}$, where $x_0 = 0$. We allow e_{j-1}^θ, $j \ge 1$, to depend on $x_0, \dots, x_{j-1}$ so that e^θ may depend on $\boldsymbol{x}$. Note, however, that the dependency of e^θ and e_{j-1}^θ on $\boldsymbol{x}$ and $x_0, \dots, x_{j-1}$, respectively, is not made explicit. The interpretation is that $e_{j-1}^\theta(t)$ gives us the effective age at calendar time $t \in (x_{j-1}, x_j]$. Here are two examples. Further examples can, for instance, be found in Lindqvist [15] and Peña [19].

Example 2.1 (ARA$_1$ **or Kijima I with non-random repair).** For an ARA$_1$ model, see [7, 13] and [12] for this model, we have $\theta \in [0, 1]$, $e_0^\theta(t) = t$ which means that regardless of θ calendar time and virtual age coincide until the first treatment/intervention, and for $j \geq 2$ we have $e_{j-1}(t) = t - \theta x_{j-1}$. Let us first look at the two extreme cases $\theta = 0$ and $\theta = 1$. If $\theta = 0$ then for all $j \geq 2$ we have $e_{j-1}(t) = t$ so that $e^\theta(t) = t$, $t \geq 0$. Hence, calendar time and virtual age always coincide. Thus, N will be a non-homogeneous Poisson process with intensity λ; see also first paragraph of the Introduction. If $\theta = 1$, then for all $j \geq 2$ we have $e_{j-1}^\theta(t) = t - x_{j-1}$ and consequently $e^\theta(t) = t - x_{j-1}$ for $t \in (x_{j-1}, x_j]$. Because e_{j-1}^θ describes the virtual age for the calendar time interval $t \in (x_{j-1}, x_j]$ we see that after an event occurrence at x_{j-1} the virtual age is set back to 0. This, of course, is a renewal process. More generally, the events' effects in this model correspond to the rejuvenation induced by the jth event time (for $j \geq 1$) which satisfies $e_j^\theta(x_j-) - e_{j-1}^\theta(x_j) = -\theta(x_j - x_{j-1})$. Consequently, the reduction of the virtual age is proportional to the difference between the last and the second to last jump (with proportionality factor θ).

Example 2.2 (ARA$_\infty$ **or Kijima II with non-random repair).** For an ARA$_\infty$ model, see [7,13] and [12] for this model, we have $\Theta = [0, 1]$ and $e_{j-1}^\theta(t) = t - \theta \sum_{l=1}^{j-1} (1-\theta)^{j-1-l} x_l$, $j \geq 1$, where we use the convention $\sum_{i=a}^b \cdot = 0$ for $a > b$. We see that for this specification the functions e_{j-1}^θ depend on the entire jump history. Clearly, as before, we have for $\theta = 1$ a renewal processes and taking $\theta = 0$ results in non homogeneous Poisson processes. Here again, the events' effects in this model correspond to the rejuvenation $e_j^\theta(x_j-) - e_{j-1}^\theta(x_j) = -\theta e_{j-1}^\theta(x_j)$. For an ARA$_\infty$ case, the event effect reduces the virtual age proportional to its value at the considered event time.

Because we assumed the predictable process ε^θ to take values only in $\mathcal{S}^\theta$ it will be of the form $\varepsilon_0^\theta(t)\mathbb{1}_{[X_0, X_1]}(t) + \sum_{j \geq 2} \varepsilon_{j-1}^\theta(t)\mathbb{1}_{(X_{j-1}, X_j]}(t)$ where we again suppress the dependency of ε^θ and ε_{j-1}^θ on $\boldsymbol{X}$ and $X_0, \ldots, X_{j-1}$, respectively. In Example 2.1, it then reads as

$$\varepsilon^\theta(t) = t \quad \text{for } t \leq X_1, \qquad \varepsilon^\theta(t) = t - \theta X_{j-1} \quad \text{for } X_{j-1} < t \leq X_j \text{ and } j \geq 2,$$

and in Example 2.2 as

$$\varepsilon^\theta(t) = t - \theta \sum_{l=1}^{j-1} (1-\theta)^{j-1-l} X_l \quad \text{for } X_{j-1} < t \leq X_j \text{ and } j \geq 1.$$

Now let us turn to inference on λ and θ. For this we allow N to be censored, that is, we consider $N(t) = \sum_{j \geq 1} \mathbb{1}_{\{X_j \leq t \wedge \mathcal{T}\}}$ for $t \geq 0$, so that $Y(u) = \mathbb{1}_{\{\mathcal{T} \geq u\}}$ for a censoring time $\mathcal{T}$. Under Assumption B1 on the virtual age function stated in Section 2.3, we know that $M(t) = N(t) - \int_0^t Y(u)\lambda(\varepsilon^\theta(u))\,du$ is a square integrable martingale with respect to the natural filtration. For inference on λ and θ, it is beneficial to disentangle them. To this end, Peña et al. [20]

introduced doubly indexed processes that, by a change of variables (see, for instance, Peña [17]), are given by

$$M^{\theta}(s,t) = N^{\theta}(s,t) - \int_0^t Y^{\theta}(s,u)\lambda(u)\,du,$$

where $N^{\theta}(s,t) = f_{\theta,t}(\boldsymbol{Z})$ and $Y^{\theta}(s,t) = g_{\theta,t}(\boldsymbol{Z})$ are defined by

$$f_{\theta,t}(z) = \sum_{j\geq 1} \mathbb{1}_{\{e^{\theta}(x_j)\leq t; x_j \leq s\wedge\tau\}} = \sum_{j\geq 1} \mathbb{1}_{\{e^{\theta}_{j-1}(x_j)\leq t; x_j \leq s\wedge\tau\}}$$

and

$$\begin{aligned} g_{\theta,t}(z) &= \mathbb{1}_{\{t\leq x_1\wedge s\wedge\tau\}} + \sum_{j\geq 2} \mathbb{1}_{\{e^{\theta}(x_{j-1}+)<t\leq e^{\theta}(x_j\wedge s\wedge\tau); x_{j-1}<s\wedge\tau\}} \\ &= \mathbb{1}_{\{t\leq x_1\wedge s\wedge\tau\}} + \sum_{j\geq 2} \mathbb{1}_{\{e^{\theta}_{j-1}(x_{j-1})<t\leq e^{\theta}_{j-1}(x_j\wedge s\wedge\tau); x_{j-1}<s\wedge\tau\}}, \end{aligned}$$

respectively, where we suppress the dependence of f and g on s. Note that this particular form of $g_{\theta,t}$ is valid under Assumption B1 below. Although it has moment properties comparable to those of martingales (in particular $PM^{\theta}(s,t) = 0$), the process $t \mapsto M^{\theta}(s,t)$ is no longer a martingale, making the study of estimators based on $t \mapsto M^{\theta}(s,t)$ rather complicated (see, e.g., Dorado et al. [6], Peña et al. [20], Peña [18]).

2.2. Estimation method

Now we consider $\mathbb{Z}_n = \{\boldsymbol{Z}_1, \dots, \boldsymbol{Z}_n\}$ where the $\boldsymbol{Z}_i = (\mathcal{T}_i, \boldsymbol{X}_i)$ are $n > 1$ i.i.d. copies of $\boldsymbol{Z} = (\mathcal{T}, \boldsymbol{X})$. We write $\boldsymbol{X}_i = (X_{i,j})_{j\geq 1}$ and $X_{i,0} = 0$. Since the model is known up to parameters $(\theta, \lambda) \in \Theta \times \mathcal{A}$, following Jacod [11] (see also Andersen et al. [2], Section II.7, Peña [19], Beutner et al. [3]) the log-likelihood function can be written as

$$\begin{aligned} \ell_{n,s}(\theta,\lambda|\mathbb{Z}_n) &= \frac{1}{n}\sum_{i=1}^{n}\int_0^{\infty} \log\lambda(u) N_i^{\theta}(s,du) - Y_i^{\theta}(s,u)\Lambda(du), \\ &= \frac{1}{n}\sum_{i=1}^{n}\int_0^{M} \log\lambda(u) N_i^{\theta}(s,du) - Y_i^{\theta}(s,u)\Lambda(du), \end{aligned} \tag{2.1}$$

where, as in the Introduction, Λ is the cumulative hazard rate function corresponding to λ. As above, the existence of M is given by Assumption B1 below, and $N_i^{\theta}(s,t) = f_{\theta,t}(\boldsymbol{Z}_i)$ and $Y_i^{\theta}(s,t) = g_{\theta,t}(\boldsymbol{Z}_i)$, $1 \leq i \leq n$. Considering the empirical measure $\mathbb{P}_n = \frac{1}{n}\sum_{i=1}^{n}\delta_{\boldsymbol{Z}_i}$ we write

$$\bar{N}_n^{\theta}(s,t) = \mathbb{P}_n f_{\theta,t} = \frac{1}{n}\sum_{i=1}^{n} N_i^{\theta}(s,t) \quad \text{and} \quad \bar{Y}_n^{\theta}(s,t) = \mathbb{P}_n g_{\theta,t} = \frac{1}{n}\sum_{i=1}^{n} Y_i^{\theta}(s,t).$$

Since the process $t \mapsto \bar{N}_n^\theta(s,t) - \int_0^t \bar{Y}_n^\theta(s,u)\lambda(u)\,du$ is centered, a "method-of-moment" type estimator for $\Lambda(t) = \int_0^t \lambda(u)\,du$ is defined by

$$\Lambda_n^\theta(s,t) = \int_0^t \frac{\bar{N}_n^\theta(s,du)}{\bar{Y}_n^\theta(s,u)}.$$

We call $\Lambda_n^\theta(s,t)$ a pseudo-NPMLE of Λ since for θ known it is a NPMLE (Nonparametric Maximum Likelihood Estimator) of Λ as proved in Beutner et al. [3]. The first general version of this estimator was derived by Dorado et al. [6] and it has been extended to more general models using the doubly indexed counting processes of Peña et al. [20]. The common approach to estimate θ based on (2.1) profiles out Λ and λ using Λ_n^θ and $\breve{\lambda}_n^\theta$, respectively, where $\breve{\lambda}_n^\theta$ equals the jump heights of Λ_n^θ at the jump points of $\bar{N}_n^\theta$; for full details see [3], Section 2.2. The resulting profile likelihood function equals (see again [3], Section 2.2)

$$\int_0^M \log\bigl(\breve{\lambda}_n^\theta(s,u)\bigr)\,d\bar{N}_n^\theta(s,du). \tag{2.2}$$

Yet, in contrast to other semiparametric models this approach fails for the model considered here; see Sections 2 and 4 in [3]. On the other hand, one would expect that (2.2) converges to

$$\ell_s(\theta) = \int_0^M \log\bigl(\lambda^\theta(s,t)\bigr)\nu^\theta(s,dt), \tag{2.3}$$

where

$$\lambda^\theta(s,t) = \frac{d\Lambda^\theta(s,t)}{dt} \quad \text{with } \Lambda^\theta(s,t) = \int_0^t \frac{\nu^\theta(s,du)}{y^\theta(s,u)}dt \text{ so that } \lambda^\theta(s,t) = \frac{\frac{d\nu^\theta}{dt}(s,t)}{y^\theta(s,t)},$$

with

$$\nu^\theta(s,t) = Pf_{\theta,t} = \mathbb{E}N^\theta(s,t) \quad \text{and} \quad y^\theta(s,t) = Pg_{\theta,t} = \mathbb{E}Y^\theta(s,t).$$

Furthermore, it can be showed (cf. Proposition 2.3 below) that $\ell_s(\theta)$ has a unique maximum at the true θ_0. Because $\bar{N}_n^\theta$ only depends on the model and the data this suggests the inconsistency of the classical profile likelihood method proved by Beutner et al. [3] comes from the fact that $\breve{\lambda}_n^\theta$ is an inappropriate estimator to have convergence of (2.2) to (2.3) and consequently that (2.2) is inappropriate for estimating the Euclidean parameter. It then further indicates that a consistent estimator for θ might be obtained from (2.2) if we use another technique for the estimation of λ. To this end, define for $t \in [0, M]$

$$\lambda_n^\theta(s,t) = \frac{1}{b_n}\int_{\mathbb{R}} \kappa\left(\frac{t-u}{b_n}\right)\Lambda_n^\theta(s,du), \tag{2.4}$$

where $\Lambda_n^\theta(s,du) \equiv 0$ on (M,∞), κ is a probability density function (kernel function) and b_n is a bandwidth tending to 0 as n tends to infinity. Then introducing

$$\ell_{n,s}(\theta) = \int_0^M \log\bigl(\lambda_n^\theta(s,t)\bigr)\bar{N}_n^\theta(s,dt),$$

we estimate θ by $\theta_n = \arg\max_{\theta\in\Theta} \ell_{n,s}(\theta)$, $\Lambda(t)$ by $\Lambda_n(s,t) = \Lambda_n^{\theta_n}(s,t)$ and $\lambda_n(s,t) = \lambda_n^{\theta_n}(s,t)$.

2.3. Consistency results

In the sequel, we denote the total variation $\int_a^b |f(s,dt)|$ of $t \mapsto f(s,t)$ on $[a,b]$ by $\mathrm{TV}_f[a,b]$, and if it is finite, we say that $t \mapsto f(s,t) \in \mathrm{BV}[a,b]$. Hereafter any true model parameter is indexed by 0. Our main consistency results are obtained under assumptions below. First, Assumptions A contains technical properties of the kernel function and the bandwidth.

Assumptions A (On the kernel and the bandwidth).

A1. The bandwidth $(b_n)_{n\in\mathbb{N}}$ satisfies $b_n = cn^{-d}$ for $d \in (0, 1/2)$ and a fixed real number $c > 0$.

A2. The kernel function κ is a pdf with support in $[-1, 1]$ and $\kappa \in \mathrm{BV}[-1, 1]$.

Assumptions B below impose conditions on the statistical model. These conditions must be checked once the effective age functions e^{θ}_{j-1}, Θ and $\mathcal{A}$ are specified. In Section 3, we show that these conditions hold for a large class of models.

Assumptions B (On the model and its parameters).

B1. e^{θ}_0 is the identity function. For $t \in (x_{j-1}, x_j]$ the functions e^{θ}_{j-1}, $j \geq 2$, are non negative. For a fixed constant $s > 0$ there exists an M such that for $j \geq 2$ we have $e^{\theta}_{j-1}(t) \leq M$ for $t \leq s$. In addition, for $j \geq 2$ the functions e^{θ}_{j-1} are nondecreasing, continuous and differentiable with derivative equal to 1.

B2. (i) There exists $\alpha > 0$ and $A < \infty$ such that $\alpha \leq y^{\theta}(s,t) \leq A$ for all $(\theta, t) \in \Theta \times [0, M]$. In addition, $(\theta, t) \mapsto y^{\theta}(s,t)$ is continuous on $\Theta \times (0, M)$, $\sup_{\theta\in\Theta} \int_0^M |y^{\theta}(s, dt)| < \infty$ and $\sup_{\theta\in\Theta} \int_0^M |\frac{1}{y^{\theta}(s,dt)}| < \infty$.

(ii) There exists $\tilde{\alpha} > 0$ and $\tilde{A} < \infty$ such that $\tilde{\alpha} \leq \lambda^{\theta}(s,t) \leq \tilde{A}$ for all $(\theta, t) \in \Theta \times [0, M]$. In addition, $(\theta, t) \mapsto \lambda^{\theta}(s,t)$ is continuous on $\Theta \times (0, M)$ and $\sup_{\theta\in\Theta} \int_0^M |\lambda^{\theta}(s, dt)| < \infty$.

B3. If the following equation holds with probability one

$$\int_0^M \log \lambda^{\theta}(s,u) N^{\theta}(s,du) - Y^{\theta}(s,u)\Lambda^{\theta}(s,du) = \int_0^M \log \lambda_0(u) N^{\theta_0}(s,du) - Y^{\theta_0}(s,u)\Lambda_0(du) \tag{2.5}$$

we have $\theta = \theta_0$.

Assumption C (Convergence rate of basic processes). With b_n as in Assumption A.1. we have with probability 1

$$b_n^{-1} \sup_{(\theta,t)\in\Theta\times[0,M]} \left|\bar{N}^{\theta}_n(s,t) - \nu^{\theta}(s,t)\right| \to 0 \quad \text{and} \quad b_n^{-1} \sup_{(\theta,t)\in\Theta\times[0,M]} \left|\bar{Y}^{\theta}_n(s,t) - y^{\theta}(s,t)\right| \to 0.$$

We show below that under Assumptions A–C we have that θ_n is strongly consistent. Moreover, it will be seen that under these assumptions $\Lambda_n(s,\cdot)$ and $\lambda_n(s,\cdot)$ are both uniformly strongly consistent. If one only wants to prove convergence in probability and uniform convergence in probability, respectively, one can replace the almost sure convergence in Assumption C by convergence in probability, and the results below continue to hold if one replaces there almost surely by in probability. The "in probability" version of Assumption C would obviously hold if one assumes that $\mathcal{F}=\{z\mapsto f_{\theta,t}(z);(\theta,t)\in\Theta\times[0,M]\}$ and $\mathcal{G}=\{z\mapsto g_{\theta,t}(z);(\theta,t)\in\Theta\times[0,M]\}$ are P-Donsker classes of functions. This in turn would be implied by assumption D1 below. Of course, there are other methods to ensure that $\mathcal{F}$ and $\mathcal{G}$ are P-Donsker; see, for instance, [8,9,24] and [25].

Lemma 2.1 below gives sufficient conditions for Assumption C to be satisfied. Because these conditions are also extremely useful later on to show that ARA_m models satisfy Assumptions B we state these conditions as

Assumptions D.

D1. For all $r\in\mathbb{N}$, there exists $C_r<\infty$ such that $\sum_{k\geq 1}k^r P(X_k\leq s)\leq C_r$;

D2. $\mathcal{F}=\{z\mapsto f_{\theta,t}(z);(\theta,t)\in\Theta\times[0,M]\}$ and $\mathcal{G}=\{z\mapsto g_{\theta,t}(z);(\theta,t)\in\Theta\times[0,M]\}$ have ϵ-bracketing numbers of polynomial order.

Lemma 2.1. *Assume that Assumptions D hold. Then Assumption C holds.*

Proof. Let us start with $\bar{N}_n^\theta$. Let $\epsilon>0$, proving the result is by the Borel–Cantelli lemma equivalent to prove that $\sum_{n\geq 1}P\{b_n^{-1}\sup_{f\in\mathcal{F}}|(\mathbb{P}_n-P)f|>\epsilon\}<\infty$. The proof we give for this relies on Theorem 2.14.9 in van der Vaart and Wellner [25]. This theorem requires the functions of the function class to have sup-norm less than or equal to one. To achieve this for an integer k, we set $\mathcal{X}_k=\{(x_j)_{j\in\mathbb{N}}\in\mathcal{X};x_j>s\text{ for }j>k\}$ and $\mathcal{Z}_k=\mathbb{R}^+\times\mathcal{X}_k$. Then for $k\geq 0$ $\mathcal{Z}_k\subset\mathcal{Z}_{k+1}$ and because $\mathcal{X}$ has no accumulation point $\mathcal{Z}=\bigcup_{k\geq 0}\mathcal{Z}_k$. Define $\mathcal{F}_k=\{k^{-1}f_{|\mathcal{Z}_k};f\in\mathcal{F}\}$ for any positive integer k. Note that if f belongs to $\mathcal{F}_k$ then $\|f\|_\infty\leq 1$. We have for a nondecreasing sequence of integers $(k_n)_{n\geq 1}$ and $\mathbb{G}_n=\sqrt{n}(\mathbb{P}_n-P)$

$$
\begin{aligned}
&\Big\{b_n^{-1}\sup_{f\in\mathcal{F}}\big|(\mathbb{P}_n-P)f\big|>\epsilon\Big\}\\
&\quad\subset\Big\{\sup_{f\in\mathcal{F}}|\mathbb{G}_nf|>b_nn^{1/2}\epsilon\Big\}\\
&\quad\subset\bigg(\bigg\{\sup_{f\in\mathcal{F}_{k_n}}|\mathbb{G}_nf|>\frac{b_nn^{1/2}\epsilon}{k_n}\bigg\}\cap\Big\{\max_{1\leq i\leq n}N_i(s)\leq k_n\Big\}\bigg)\cup\Big\{\max_{1\leq i\leq n}N_i(s)>k_n\Big\}\\
&\quad\subset\bigg\{\sup_{f\in\mathcal{F}_{k_n}}|\mathbb{G}_nf|>\frac{b_nn^{1/2}\epsilon}{k_n}\bigg\}\cup\Big\{\max_{1\leq i\leq n}N_i(s)>k_n\Big\}.
\end{aligned}
$$

By Assumption D2 there exist positive constants c' and V such that $N_{[]}(\epsilon,\mathcal{F},L^2(P))\leq c'/\epsilon^V$, thus from Theorem 2.14.9 in van der Vaart and Wellner [25] we have for $k_n=\lfloor n^\beta\rfloor$ with $\beta\in$

$(0, 1/2 - d)$

$$P\left\{\sup_{f\in\mathcal{F}_{k_n}} |\mathbb{G}_n f| > \frac{b_n n^{1/2}\epsilon}{k_n}\right\} \le P\left\{\sup_{f\in\mathcal{F}_{k_n}} |\mathbb{G}_n f| > \frac{b_n n^{1/2}\epsilon}{n^{\beta}}\right\}$$
$$\le c''\epsilon^V n^{V(1-2d-2\beta)/2} \exp\bigl(-2c^2\epsilon^2 n^{1-2d-2\beta}\bigr),$$

where c'' is another positive constant, thus the right-hand side of the last inequality is the general term of a converging series since $1 - 2d - 2\beta > 0$. It remains to prove that $\sum_{n\ge1} P\{\max_{1\le i\le n} N_i(s) > k_n\} < \infty$. We have

$$\begin{aligned}
&\sum_{n\ge1} P\Bigl\{\max_{1\le i\le n} N_i(s) > k_n\Bigr\}\\
&\quad\le \sum_{n\ge1}\bigl(1 - \bigl(P\bigl\{N(s)\le k_n\bigr\}\bigr)^n\bigr)\\
&\quad\le \sum_{n\ge1} n P\bigl\{N(s) > k_n\bigr\} \le \sum_{n\ge1} n P\{X_{\lfloor n^{\beta}\rfloor} \le s\}\\
&\quad\le \sum_{k\ge1} (k+1)^{\lceil 1/\beta\rceil}\,\mathrm{Card}\bigl\{n\in\mathbb{N};\ \bigl\lfloor n^{\beta}\bigr\rfloor = k\bigr\} P\{X_k\le s\}\\
&\quad\le 2^{\lceil 1/\beta\rceil}\sum_{k\ge1} k^{\lceil 1/\beta\rceil}\,\mathrm{Card}\bigl\{n\in\mathbb{N};\ \bigl\lfloor n^{\beta}\bigr\rfloor = k\bigr\} P\{X_k\le s\}\\
&\quad\le \frac{2^{2\lceil 1/\beta\rceil-1}}{\beta}\sum_{k\ge1} k^{2\lceil 1/\beta\rceil-1} P\{X_k\le s\} < +\infty \quad \text{(by Assumption D1)},
\end{aligned}$$

since $\mathrm{Card}\{n\in\mathbb{N};\ \lfloor n^{\beta}\rfloor = k\} \le (2^{\lceil 1/\beta\rceil-1}/\beta)k^{\lceil 1/\beta\rceil-1}$. Indeed for $k\ge1$ we have $\lfloor n^{\beta}\rfloor = k \Leftrightarrow k \le n^{\beta} < k+1 \Leftrightarrow k^{1/\beta} \le n < (k+1)^{1/\beta}$. Put $w(k) = k^{1/\beta}$. Then by the mean value theorem and since $\beta\in(0,1)$ we have $w(k+1) - w(k) \le w'(k+1) = (1/\beta)(k+1)^{1/\beta-1} \le (2^{\lceil 1/\beta\rceil-1}/\beta)k^{\lceil 1/\beta\rceil-1}$ and $\mathrm{Card}\{n\in\mathbb{N};\ \lfloor n^{\beta}\rfloor = k\} \le w(k+1) - w(k)$.

The second convergence result concerning $\bar{Y}_n^{\theta}$ holds by using the same arguments. □

We now give preliminary results to show our consistency results.

Proposition 2.1. *Under Assumptions A–C we have*

$$b_n^{-1}\sup_{(\theta,t)\in\Theta\times[0,M]}\bigl|\Lambda_n^{\theta}(s,t) - \Lambda^{\theta}(s,t)\bigr| \to 0.$$

with probability one.

Proof. We have by using the integration by parts formula

$$\left|b_n^{-1}\int_0^t\left\{\frac{\bar{N}_n^{\theta}(s,du)}{\bar{Y}_n^{\theta}(s,u)} - \frac{\nu^{\theta}(s,du)}{y^{\theta}(s,u)}\right\}\right|$$

$$
\begin{aligned}
&\leq \left| b_n^{-1} \int_0^t \frac{y^\theta(s,u) - \bar{Y}_n^\theta(s,u)}{y^\theta(s,u)\bar{Y}_n^\theta(s,u)} \bar{N}_n^\theta(s,du) \right| + \frac{b_n^{-1}}{y^\theta(s,t)} \left| \bar{N}_n^\theta(s,t) - \nu^\theta(s,t) \right| \\
&\quad + b_n^{-1} \left| \int_0^t \left(\bar{N}_n^\theta(s,u) - \nu^\theta(s,u) \right) d\left(\frac{1}{y^\theta(s,u)} \right) \right| \\
&\leq b_n^{-1} \sup_{(\theta,t)\in\Theta\times[0,M]} \left| \bar{Y}_n^\theta(s,t) - y^\theta(s,t) \right| \times \int_0^M \left| \frac{\bar{N}_n^\theta(s,du)}{\bar{Y}_n^\theta(s,u) y^\theta(s,u)} \right| \\
&\quad + b_n^{-1} \sup_{(\theta,t)\in\Theta\times[0,M]} \left| \bar{N}_n^\theta(s,t) - \nu^\theta(s,t) \right| \times \left\{ \frac{1}{\alpha} + \int_0^M \left| d\left(\frac{1}{y^\theta(s,u)} \right) \right| \right\}.
\end{aligned}
$$

Due to Assumption C it remains to show that $\sup_{\theta\in\Theta} \int_0^M |\frac{\bar{N}_n^\theta(s,du)}{\bar{Y}_n^\theta(s,u) y^\theta(s,u)}| = O_{a.s.}(1)$, since by B2(i) we have $\sup_{\theta\in\Theta} \int_0^M |\frac{1}{y^\theta(s,du)}| < \infty$.

Let $0 = t_0 < t_1 < \cdots < t_k = M$ be a partition of $[0, M]$ we have

$$
\begin{aligned}
&\sum_{i=1}^k \left| \frac{\bar{N}_n^\theta(s,t_i)}{\bar{Y}_n^\theta(s,t_i) y^\theta(s,t_i)} - \frac{\bar{N}_n^\theta(s,t_{i-1})}{\bar{Y}_n^\theta(s,t_{i-1}) y^\theta(s,t_{i-1})} \right| \\
&\quad \leq \sum_{i=1}^k \left| \frac{\bar{N}_n^\theta(s,t_i)\bar{Y}_n^\theta(s,t_{i-1}) y^\theta(s,t_{i-1}) - \bar{N}_n^\theta(s,t_{i-1})\bar{Y}_n^\theta(s,t_i) y^\theta(s,t_i)}{\bar{Y}_n^\theta(s,t_i) y^\theta(s,t_i)\bar{Y}_n^\theta(s,t_{i-1}) y^\theta(s,t_{i-1})} \right| \\
&\quad \lesssim \sum_{i=1}^k \left\{ \left| \bar{N}_n^\theta(s,t_i) - \bar{N}_n^\theta(s,t_{i-1}) \right| \bar{Y}_n^\theta(s,t_{i-1}) y^\theta(s,t_{i-1}) \right. \\
&\qquad + \left| \bar{Y}_n^\theta(s,t_i) - \bar{Y}_n^\theta(s,t_{i-1}) \right| \bar{N}_n^\theta(s,t_{i-1}) y^\theta(s,t_{i-1}) \\
&\qquad \left. + \left| y^\theta(s,t_i) - y^\theta(s,t_{i-1}) \right| \bar{N}_n^\theta(s,t_{i-1}) \bar{Y}_n^\theta(s,t_i) \right\} \\
&\quad \lesssim \bar{N}_n^\theta(s,M) \left\{ 1 + \mathrm{TV}_{\bar{Y}_n^\theta(s,\cdot)}[0,M] + \mathrm{TV}_{y^\theta(s,\cdot)}[0,M] \right\}
\end{aligned}
$$

using Assumption C and Assumption B2(i) to delete the denominator, and where $\lesssim$ means less or equal up to a multiplicative constant. Since the functions $t \mapsto e_j^\theta(t)$ are assumed to be nondecreasing, we have $g_{\theta,t} = g_{\theta,t}^+ - g_{\theta,t}^-$ where

$$
g_{\theta,t}^+(z) = 1 + \sum_{j\geq 2} \mathbb{1}_{\{e_{j-1}^\theta(x_{j-1})<t;\, x_{j-1}<s\wedge\tau\}}, \tag{2.6}
$$

and

$$
g_{\theta,t}^-(z) = \sum_{j\geq 1} \mathbb{1}_{\{e_{j-1}^\theta(x_j\wedge s\wedge\tau)<t;\, x_{j-1}<s\wedge\tau\}}. \tag{2.7}
$$

For all $\theta \in \Theta$, $t \mapsto \mathbb{P}_n g^+_{\theta,t}$ and $t \mapsto \mathbb{P}_n g^-_{\theta,t}$ are two nonnegative, nondecreasing functions. In addition, both functions are bounded by $\bar{N}^\theta_n(s, M)$ which converges almost surely (from Assumption C) to $\nu^\theta(s, M) = \int_0^M y^\theta(s,t)\lambda^\theta(s,t)\,dt \le MA\tilde{A}$ (by Assumption B2). Thus, for n sufficiently large, we have $\sup_{\theta\in\Theta} \mathrm{TV}_{\bar{Y}^\theta_n(s,\cdot)}[0, M] \le 4MA\tilde{A}$. This finishes the proof. $\square$

Proposition 2.2. *Under Assumptions A–C we have*:

(i) *For all closed intervals* $[a, b] \subset (0, M)$, *and as n tends to infinity*

$$\sup_{(\theta,t)\in\Theta\times[a,b]} \left|\lambda^\theta_n(s,t)/\lambda^\theta(s,t) - 1\right| \to 0, \quad a.s.$$

(ii) *There exists* $(\alpha', A') \in (0, 1] \times [1, \infty)$ *such that for almost all* ω *we have for all* n *sufficiently large* (*with sufficiently large depending on* ω) *and all* $(\theta, t) \in \Theta \times [0, M]$, $\alpha' \le \lambda^\theta_n(s,t)/\lambda^\theta(s,t) \le A'$.

Proof. First, we note that for all $(\theta, t) \in \Theta \times [0, M]$

$$\lambda^\theta_n(s,t) - \lambda^\theta(s,t) = \frac{1}{b_n}\int_{\mathbb{R}} \kappa\left(\frac{t-u}{b_n}\right)\big(\Lambda^\theta_n(s,du) - \Lambda^\theta(s,du)\big) \tag{2.8}$$

$$+ \frac{1}{b_n}\int_{\mathbb{R}} \kappa\left(\frac{t-u}{b_n}\right)\Lambda^\theta(s,du) - \lambda^\theta(s,t). \tag{2.9}$$

Write $A^\theta_n(s,t)$ for the right-hand side of (2.8) and $B^\theta_n(s,t)$ for the term in (2.9). By the integration by parts formula, the right continuity of $t \mapsto \Lambda^\theta_n(s,t)$, and Assumption A2 on the kernel function κ we have

$$\begin{aligned}\left|A^\theta_n(s,t)\right| &\le \frac{1}{b_n}\int_{\mathbb{R}} \left|\Lambda^\theta_n(s,u) - \Lambda^\theta(s,u)\right|\left|d\kappa\big((t-u)/b_n\big)\right| \\ &\le \frac{C}{b_n}\sup_{(\theta,t)\in\Theta\times[0,M]} \left|\Lambda^\theta_n(s,t) - \Lambda^\theta(s,t)\right|,\end{aligned}$$

which tends to 0 with probability one by Proposition 2.1. For the remaining term, we have

$$\begin{aligned}\left|B^\theta_n(s,t)\right| &\le \left|\int_{\mathbb{R}} \kappa(u)\big(\lambda^\theta(s,t-b_n u) - \lambda^\theta(s,t)\big)\,du\right| \\ &\le \sup_{|u|\le b_n} \left|\lambda^\theta(s,t+u) - \lambda^\theta(s,t)\right| \quad \text{(by Assumption A2).}\end{aligned}$$

By Assumption B2(ii), $(\theta, t) \mapsto \lambda^\theta(s,t)$ is uniformly continuous on $\Theta \times [a, b]$, then $\sup_{(\theta,t)\in\Theta\times[a,b]} |B^\theta_n(s,t)| \to 0$ tends to 0. From B2(ii), we have that $\lambda^\theta(s,t) \ge \tilde{\alpha}$ which finishes the proof of (i).

Using Assumptions B2(ii) and A2, we have for n sufficiently large

$$\tilde{\alpha}/2 \le \lambda_n^\theta(s,t) = A_n^\theta(s,t) + \frac{1}{b_n}\int_{\mathbb{R}} \kappa\left(\frac{t-u}{b_n}\right)\Lambda^\theta(s,du) \le 2\tilde{A}$$

with probability one. Consequently, we can take $\alpha' = \tilde{\alpha}/(2\tilde{A})$ and $A' = 2\tilde{A}/\tilde{\alpha}$. □

Proposition 2.3. *Under Assumptions A–C we have*:

(i) *As n tends to infinity*

$$\sup_{\theta\in\Theta}\left|\ell_{n,s}(\theta) - \ell_s(\theta)\right| \to 0, \quad a.s. \tag{2.10}$$

(ii) *$\theta \mapsto \ell_s(\theta)$ is continuous on Θ and $\ell_s(\theta) < \ell_s(\theta_0)$ for all $\theta \in \Theta\backslash\{\theta_0\}$.*

Proof. Let us prove (i). First, note that for $x \in [-1/2, 1/2]$ we have $|\log(1+x)-x| \le |x|$ (apply the mean value theorem to $x \mapsto \log(1+x) - x$). We have

$$\ell_{n,s}(\theta) - \ell_s(\theta) = \int_0^M \log\left(\frac{\lambda_n^\theta(s,t)}{\lambda^\theta(s,t)}\right)\bar{N}_n^\theta(s,dt) \tag{2.11}$$

$$+ \int_0^M \log\big(\lambda^\theta(s,t)\big)\big[\bar{N}_n^\theta(s,dt) - \nu^\theta(s,dt)\big]. \tag{2.12}$$

Write $A_{n,s}(\theta)$ for the right-hand side of (2.11) and $B_{n,s}(\theta)$ for the term in (2.12). We have for all $\epsilon > 0$

$$\begin{aligned}
&\left|A_{n,s}(\theta)\right| \\
&\quad\le \int_\epsilon^{M-\epsilon}\left|\log\left(1+\frac{\lambda_n^\theta(s,t)-\lambda^\theta(s,t)}{\lambda^\theta(s,t)}\right) - \frac{\lambda_n^\theta(s,t)-\lambda^\theta(s,t)}{\lambda^\theta(s,t)}\right|\bar{N}_n^\theta(s,dt) \\
&\qquad + \int_\epsilon^{M-\epsilon}\left|\frac{\lambda_n^\theta(s,t)-\lambda^\theta(s,t)}{\lambda^\theta(s,t)}\right|\bar{N}_n^\theta(s,dt) + \big(\log(A') - \log(\alpha')\big)\int_{[0,\epsilon]\cup[M-\epsilon,M]}\bar{N}_n^\theta(s,dt) \\
&\quad\le 2\sup_{(\theta,t)\in\Theta\times[\epsilon,M-\epsilon]}\left|\frac{\lambda_n^\theta(s,t)-\lambda^\theta(s,t)}{\lambda^\theta(s,t)}\right|\bar{N}_n^\theta(s,M) + \left|A_{n,s}(\theta)\right|\mathbb{1}_{\{\sup_{(\theta,t)\in\Theta\times[\epsilon,M-\epsilon]}|\frac{\lambda_n^\theta(s,t)}{\lambda^\theta(s,t)}-1|\ge 1/2\}} \\
&\qquad + \big(\log(A') - \log(\alpha')\big) \\
&\qquad \times\left(2\epsilon\sup_{(\theta,t)\in\Theta\times[0,M]}\left|\frac{d\nu^\theta}{dt}(s,t)\right| + 4\sup_{(\theta,t)\in\Theta\times[0,M]}\left|\bar{N}_n^\theta(s,t) - \nu^\theta(s,t)\right|\right),
\end{aligned}$$

where the number 4 in the last equation comes from the upper bound of $\sum_{t\in\{0,\epsilon,M-\epsilon,M\}}|\bar{N}_n^\theta(s,t) - \nu^\theta(s,t)|$. Notice that $0 \le \frac{d\nu^\theta}{dt}(s,t) = y^\theta(s,t)\times\lambda^\theta(s,t) \le A\tilde{A}$. Thus, $\sup_{\theta\in\Theta}|A_{n,s}(\theta)| \to 0$ a.s. from Proposition 2.2 and Assumption C. In addition by the integration by parts formula, we can write

$$\begin{aligned}
&\big|B_{n,s}(\theta)\big| \\
&\quad\le \big|\log\big(\lambda^\theta(s,M)\big)\big(\bar N_n^\theta(s,M)-\nu^\theta(s,M)\big)\big| \\
&\qquad + \int_0^M \big|\bar N_n^\theta(s,t)-\nu^\theta(s,t)\big|\left|\frac{\lambda^\theta(s,dt)}{\lambda^\theta(s,t)}\right| \\
&\quad\le \sup_{(\theta,t)\in\Theta\times[0,M]}\big|\bar N_n^\theta(s,t)-\nu^\theta(s,t)\big| \\
&\qquad\times\left\{\big|\log(\tilde\alpha)\big|+\big|\log(\tilde A)\big|+\frac{1}{\tilde\alpha}\times\sup_{\theta\in\Theta}\int_0^M\big|\lambda^\theta(s,dt)\big|\right\}.
\end{aligned}$$

We conclude that $\sup_{\theta\in\Theta}|B_{n,s}(\theta)|\to 0$ a.s. by Assumption C and Assumptions B2(ii). This finishes the proof of (i).

Let us prove (ii). First, note that $\ell_s(\theta)=\int_0^M\log(\lambda^\theta(s,t))\lambda^\theta(s,t)y^\theta(s,t)\,dt$, thus the continuity follows from the dominated convergence theorem. In fact, we have that $t\mapsto\log(\lambda^\theta(s,t))\lambda^\theta(s,t)y^\theta(s,t)$ is continuous on $(0,M)$ for any fixed $\theta\in\Theta$, continuous on Θ for any fixed $t\in(0,M)$ and its absolute value is bounded by $(\log(A')-\log(\alpha'))\tilde AA$. Peña [18] defined the random function $A(s,t;\theta_0)=\int_0^t Y^{\theta_0}(s,u)\Lambda_0(du)$ and showed (Proposition 1) that $\nu^{\theta_0}(s,t)=\mathbb{E}N^{\theta_0}(s,t)=\mathbb{E}A(s,t;\theta_0)$ for $t\in[0,M]$. Thus, for all $t\in[0,M]$, we have

$$\Lambda^{\theta_0}(s,t)=\int_0^t\frac{\mathbb{E}A(s,du;\theta_0)}{y^{\theta_0}(s,u)}=\int_0^t\frac{y^{\theta_0}(s,u)\Lambda_0(du)}{y^{\theta_0}(s,u)}=\Lambda_0(t) \tag{2.13}$$

since by Assumption B2(i) $t\mapsto y^{\theta_0}(s,t)$ is bounded away from zero on $[0,M]$. In addition we have

$$\begin{aligned}
&\mathbb{E}\int_0^M Y^\theta(s,u)\Lambda^\theta(s,du) \\
&\quad=\mathbb{E}\int_0^M Y^\theta(s,u)\frac{\mathbb{E}N^\theta(s,du)}{\mathbb{E}Y^\theta(s,u)} \\
&\quad=\int_0^M\mathbb{E}N^\theta(s,du)=\mathbb{E}\int_0^M\bar N^\theta(s,du)=\mathbb{E}N(s)\quad\text{(from Assumption B1)},
\end{aligned}$$

which means that the right-hand side in the first equality does not depend on θ. As a consequence, we have $\tilde\theta=\arg\max_{\theta\in\Theta}\ell_s(\theta)=\arg\max_{\theta\in\Theta}\tilde\ell_s(\theta)$, where

$$\tilde\ell_s(\theta)=\mathbb{E}\left[\int_0^M\big\{\log\big(\lambda^\theta(s,u)\big)N^\theta(s,du)-Y^\theta(s,u)\Lambda^\theta(s,du)\big\}\right].$$

It is easy to check that the likelihood function $L_s(\theta, \lambda)$ of one observation $(N(t), Y(t))_{0\le t\le s}$ is defined by

$$L_s(\theta,\lambda) = \prod_{u\in[0,s]} \big[Y(u)\lambda\big(\varepsilon^\theta(u)\big)\big]^{N(\Delta u)} \times \exp\Big(-\int_0^s Y(u)\lambda\big(\varepsilon^\theta(u)\big)\,du\Big)$$
$$= \exp\Big(\int_0^M \log\lambda(u) N^\theta(s,du) - Y^\theta(s,u)\Lambda(du)\Big).$$

Now we define

$$\tilde{\ell}_s(\theta,\lambda) = \mathbb{E}\Big[\int_0^M \big\{\log\big(\lambda(u)\big)N^\theta(s,du) - Y^\theta(s,u)\Lambda(du)\big\}\Big]$$
$$= \int_0^M \big\{\log\big(\lambda(u)\big)\nu^\theta(s,du) - y^\theta(s,u)\Lambda(du)\big\} = \mathbb{E}\big[\log L_s(\theta,\lambda)\big].$$

Then $\tilde{\ell}_s(\theta) = \tilde{\ell}_s(\theta, \lambda^\theta(s,\cdot))$. Let us note $\tilde{\lambda}(\cdot) = \lambda^{\tilde{\theta}}(s,\cdot)$ and $(\widetilde{\theta,\lambda})$ the maximum argument of $\tilde{\ell}_s(\theta,\lambda)$ for $(\theta,\lambda) \in \Theta \times \mathcal{A}^*$, where $\mathcal{A}^*$ is the set of hazard rate functions on $[0, M]$ bounded away from zero. Note that $\tilde{\ell}_s(\widetilde{\theta,\lambda}) \ge \tilde{\ell}_s(\tilde{\theta}) = \tilde{\ell}_s(\tilde{\theta},\tilde{\lambda})$. In addition, for all $\lambda \in \mathcal{A}^*$ and $\theta \in \Theta$ we have

$$\tilde{\ell}_s\big(\theta,\lambda^\theta(s,\cdot)\big) - \tilde{\ell}_s(\theta,\lambda) = \int_0^M y^\theta(s,t)\lambda(t)\varphi\Big(\frac{\lambda^\theta(s,t)}{\lambda(t)}\Big)\,dt,$$

where $\varphi(x) = x\log x - x + 1$ is nonnegative on $(0, +\infty)$, thus $\tilde{\ell}_s(\theta) = \tilde{\ell}_s(\theta, \lambda^\theta(s,\cdot)) \ge \tilde{\ell}_s(\theta,\lambda)$. Then, because $\tilde{\theta}$ is a maximizer of $\tilde{\ell}_s$ we have $\tilde{\ell}_s(\tilde{\theta}) \ge \tilde{\ell}_s(\widetilde{\theta,\lambda})$. It follows that $\tilde{\ell}_s(\tilde{\theta}) = \tilde{\ell}_s(\widetilde{\theta,\lambda}) \ge \tilde{\ell}_s(\theta_0,\lambda_0)$, or equivalently $\mathbb{E}[\log L_s(\tilde{\theta},\tilde{\lambda})] \ge \mathbb{E}[\log L_s(\theta_0,\lambda_0)]$, which by the Kullback–Leibler inequality yields that

$$\int_0^M \log\lambda^{\tilde{\theta}}(s,u)N^{\tilde{\theta}}(s,du) - Y^{\tilde{\theta}}(s,u)\Lambda^{\tilde{\theta}}(s,du)$$
$$= \int_0^M \log\lambda_0(u)N^{\theta_0}(s,du) - Y^{\theta_0}(s,u)\Lambda_0(du) \tag{2.14}$$

with probability one, thus Assumption B3 finishes the proof. □

Theorem 2.1. *Under Assumptions A–C we have for* $\Lambda_n(\cdot) = \Lambda_n^{\theta_n}(s,\cdot)$

$$\theta_n \to \theta_0 \quad \textit{and} \quad \sup_{t\in[0,s]} \big|\Lambda_n(t) - \Lambda_0(t)\big| \to 0,$$

with probability one. In addition, for any closed set $[a, b]$ *included in* $(0, M)$, *and for* $\lambda_n(\cdot) = \lambda_n^{\theta_n}(s,\cdot)$ *we also have*

$$\sup_{t\in[a,b]} \big|\lambda_n(t) - \lambda_0(t)\big| \to 0,$$

with probability one.

Proof. Because ℓ_s is continuous at θ_0 for a $\epsilon > 0$ there exists an $\eta > 0$ such that

$$\begin{aligned}
\{\|\theta_n - \theta_0\| > \epsilon\} &\subset \{\ell_s(\theta_n) < \ell_s(\theta_0) - \eta\} \\
&\subset \{\ell_s(\theta_n) - \ell_{n,s}(\theta_n) + \ell_{n,s}(\theta_n) < \ell_s(\theta_0) - \eta\} \\
&\subset \Big\{- \sup_{\theta \in \Theta} \big|\ell_{n,s}(\theta) - \ell_s(\theta)\big| + \ell_{n,s}(\theta_0) < \ell_s(\theta_0) - \eta\Big\} \\
&\subset \Big\{\eta < 2 \sup_{\theta \in \Theta} \big|\ell_{n,s}(\theta) - \ell_s(\theta)\big|\Big\}
\end{aligned}$$

thus $\theta_n \to \theta_0$ with probability one from (2.10) in Proposition 2.3. In addition, we have from (2.13)

$$\begin{aligned}
&\sup_{t \in [0,M]} \big|\Lambda_n(t) - \Lambda_0(t)\big| \\
&\quad \le \sup_{(\theta,t) \in \Theta \times [0,M]} \big|\Lambda_n^{\theta}(s,t) - \Lambda^{\theta}(s,t)\big| + \sup_{t \in [0,M]} \big|\Lambda^{\theta_n}(s,t) - \Lambda^{\theta_0}(s,t)\big|.
\end{aligned}$$

We obtained the desired result by Proposition 2.1 and since $(\theta, t) \mapsto \Lambda^{\theta}(s,t)$ is continuous (and then uniformly continuous) on $\Theta \times [0, M]$. This last point is a consequence of the dominated convergence theorem and Assumption B2(ii). A similar proof can be performed for λ_n. □

3. Application to ARA$_m$ models

In Examples 2.1 and 2.2, we have seen two particular ARA models, and as mentioned in the Introduction the general class of Arithmetic Reduction of Age (ARA) models has been introduced by Doyen and Gaudoin [7]; a description of how interventions after an event affect the virtual age in these models as well as their relation to Kijima [12] can also be found in Doyen and Gaudoin [7]. What is important here is that for ARA$_m$ models with $m \in \overline{\mathbb{N}} \equiv \mathbb{N} \cup \{+\infty\}$ (we will use for $m = \infty$ the conventions $-m = -\infty$ as well as $-m + a = -m$ and $m + a = m$ with $a \in \mathbb{R}$) we have for $j \ge 0$

$$e_j^{\theta}(t) = t - \theta \sum_{i=0}^{(j-1)\wedge(m-1)} (1-\theta)^i x_{j-i}, \tag{3.1}$$

where $\theta \in [0, 1]$ and we use, as in Example 2.2, the convention $\sum_{i=a}^{b} \cdot = 0$ for $a > b$. After some algebra this can be rewritten as

$$\begin{aligned}
e_j^{\theta}(t) &= t - x_j + \sum_{i=0}^{(j-2)\wedge(m-2)} (1-\theta)^{i+1} (x_{j-i} - x_{j-i-1}) + (1-\theta)^{j \wedge m} x_{1 \vee (j-m+1)} \\
&= t - x_j + e_j^{\theta}(x_j).
\end{aligned} \tag{3.2}$$

Thus, for any $j \geq 1$ and $t \in (x_j, x_{j+1}]$ we have that e_j^θ is nonnegative and the function $\theta \mapsto e_j^\theta(t)$ is nonincreasing, it means the greater θ is, the less aged the system is. As a consequence the parameter $\theta \in [0, 1]$ represents the efficiency of interventions. Indeed notice that for $\theta = 0$ we have $e^0(x_j+) = e_j^0(x_j) = x_j$ which corresponds to the well known "as bad as old" situation in reliability applications, where after an event occurrence the system is only restored to its condition just prior to failure. On the other hand, we have for $\theta = 1$ that $e^1(x_j+) = e_j^1(x_j) = 0$ which corresponds to the "as good as new" situation, for which after an event occurrence the system is renewed. The corresponding random processes are respectively a non homogeneous Poisson process and a renewal process.

Below are assumptions concerning the properties of the baseline hazard rate function and the distribution of the random censoring time.

Assumptions E.

E1. The baseline hazard rate λ_0 is non constant, continuous, and uniformly bounded away from zero on $[0, s]$ with lower bound different from 0, and it belongs to $\mathrm{BV}[0, s]$.

E2. The censoring time $\mathcal{T}$ is a positive random variable, independent of $\boldsymbol{X}$, with pdf $f_\mathcal{T}$ bounded on $[0, s]$, survival function $S_\mathcal{T}$ such that $S_\mathcal{T}(s) > 0$, and for all $c \in [0, s]$ and $\epsilon > 0$, $P(\mathcal{T} \in [c - \epsilon, c + \epsilon] \cap [0, s]) > 0$.

Proposition 3.1. *Suppose that Assumptions E are satisfied and that the effective age function follows an* ARA_m *model for* $m \in \overline{\mathbb{N}}$. *Then Assumptions B and D are satisfied.*

The consequence of Proposition 3.1 is that Theorem 2.1 holds for ARA_m models ($m \in \overline{\mathbb{N}}$) under very few Assumptions (A and E). Note that if $f_\mathcal{T}$ is uniformly bounded away from zero on $[0, s+\delta]$, $\delta > 0$, then $S_\mathcal{T}(s) > 0$ and for all $c \in [0, s]$ and $\epsilon > 0$, $P(\mathcal{T} \in [c-\epsilon, c+\epsilon] \cap [0, s]) > 0$.

The next subsection is devoted to the verification of B1–B3 and D1 and D2. As indicated above for ARA_m models it seems beneficial to first show that D1 and D2 hold, because they are useful to check B2 and B3. Hence, below we start with B1 and then go to D1, D2 and from there to B2 and B3. Note that because Assumption B3 depends on the censoring scheme we will discuss the verification of this assumption under various censoring schemes in the last subsection.

3.1. Proof of Proposition 3.1

Verification of B1

By construction of ARA_m models e_0^θ is the identity function. From (3.2), we have that $e_j^\theta(t) \geq 0$ for $t \in (x_j, x_{j+1}]$. Moreover, for any $j \geq 0$ we have from (3.1) that $e_j^\theta(s) \leq s$ thus we can set $M = s$, and $t \mapsto e_j^\theta(t)$ is nondecreasing, continuous and differentiable with derivative equal to 1. In addition, by (3.2) the functions $\theta \mapsto e_j^\theta(t)$ are nonincreasing, then we have $e_j^\theta(t) \geq e_j^1(t) = t - x_j \geq 0$ for $t \in (x_j, x_{j+1}]$.

Verification of D1

Please see [4], Section 7.

Verification of $D2$

The aim of this section is to prove that the classes of functions $\mathcal{F}$ and $\mathcal{G}$ are P-Donsker with polynomial bracketing numbers if the functions e_j^θ are specified as in (3.1). Since these functions fulfill condition B1 we know from the proof of Proposition 2.1 that we have $g_{\theta,t} = g_{\theta,t}^+ - g_{\theta,t}^-$. Define $\mathcal{G}^+$ and $\mathcal{G}^-$ to be the sets generated by the functions $g_{\theta,t}^+$ (see (2.6)) and $g_{\theta,t}^-$ (see (2.7)), respectively. In turn $g_{\theta,t}^-$ can be rewritten as $g_{\theta,t}^- = f_{\theta,t-} + h_{\theta,t}^+ + h_{\theta,t}^-$ where $f_{\theta,t-} = \lim_{u\uparrow t} f_{\theta,u}$, with $\mathcal{H}^+$ and $\mathcal{H}^-$ the sets of functions mapping from $\mathcal{Z}$ to $\mathbb{R}$ defined by

$$h_{\theta,t}^+(z) = \mathbb{1}_{\{e^\theta(s)<t;\tau>s\}} = \sum_{j\geq 2} \mathbb{1}_{\{e_{j-1}^\theta(s)<t;x_{j-1}<s\leq x_j;\tau>s\}},$$

and

$$h_{\theta,t}^-(z) = \mathbb{1}_{\{e^\theta(\tau)<t;\tau\leq s\}} = \sum_{j\geq 1} \mathbb{1}_{\{e_{j-1}^\theta(\tau)<t;x_{j-1}<\tau\leq x_j;\tau\leq s\}},$$

respectively. Notice that for $h_{\theta,t}^+$ the indicator function corresponding to e_0^θ is always equal to zero since $s \geq t$. The next lemma whose proof can be found in [4], Section 10, gives a general expression for $e_{j-1}^\theta(x_j)$, $e_j^\theta(x_j)$, $e_j^\theta(s)$, and $e_{j-1}^\theta(\tau)$ that allows to show the Donsker property of $\mathcal{F}$, $\mathcal{G}^+$, $\mathcal{H}^+$ and $\mathcal{H}^-$ at once.

Lemma 3.1. *For any* $\theta \in [0, 1]$ *and* $j \geq 1$ *let* $v_j(w)$ *be either* (i) $e_{j-1}^\theta(x_j)$, (ii) $e_j^\theta(x_j)$, (iii) $e_j^\theta(s)$ *or* (iv) $e_{j-1}^\theta(\tau)$. *Then* $v_j^\theta(w)$ *can be written in the generic form*

$$v_j^\theta(w) = sd + (-1)^d \phi(\theta)\big(u_j + \tilde{v}_{j-1}^\theta(\boldsymbol{x}_{j-1})\big), \tag{3.3}$$

where $d \in \{0, 1\}$, $\phi(\theta) > 0$ *except for* (*ii*) *when* $\theta = 1$ *and* (*iii*) *when* $\theta = 0$, *and* $\boldsymbol{x}_{j-1} = (x_1, \dots, x_{j-1})$. *In addition, the map* $\theta \to sd + (-1)^d\phi(\theta)(u_j + \tilde{v}_{j-1}^\theta(\boldsymbol{x}_{j-1}))$ *is monotone and there exists a constant* c *such that for all pairs* $(\theta, \tilde{\theta}) \in \Theta^2$ *we have*

$$\big|\tilde{v}_{j-1}^\theta(\boldsymbol{x}_{j-1}) - \tilde{v}_{j-1}^{\tilde{\theta}}(\boldsymbol{x}_{j-1})\big| \leq c(j-1)x_{j-1}|\theta - \tilde{\theta}|,$$

and

$$\big|\tilde{v}_{j-1}^\theta(\boldsymbol{x}_{j-1})\big| \leq c(j-1)x_{j-1}.$$

Furthermore, for (i), (ii), *and* (iii) *we have* $u_j \in [0, s]$ *if* $x_j \leq s$ *and the same holds for* (iv) *if* $\tau \leq s$.

On the other hand, if $\upsilon_j^\theta(W)$ *denotes either* (i) $\varepsilon_{j-1}^\theta(X_j)$, (ii) $\varepsilon_j^\theta(X_j)$, (iii) $\varepsilon_j^\theta(s)$ *or* (iv) $\epsilon_{j-1}^\theta(\mathcal{T})$ *for any* $\theta \in [0, 1]$ *and* $j \geq 1$, *then* $\upsilon_j^\theta(W)$ *still satisfies* (3.3) *where* W, U_j *and* $\boldsymbol{X}_{j-1}$ *are the stochastic counterparts of the quantities* w, u_j *and* $\boldsymbol{x}_{j-1}$. *And in all four cases, the conditional pdf* $f_{U_j|\boldsymbol{X}_{j-1}}$ *is almost surely bounded from above by a constant* c_0 *on* $[0, s]$.

Showing the Donsker property of $\mathcal{G}$ can be further simplified because we have $g_{\theta,t} = g^+_{\theta,t} - f_{\theta,t} - h^+_{\theta,t} - h^-_{\theta,t}$ so that it follows from Lemma 3.2 below that $\mathcal{G}$ is P-Donsker with polynomial ϵ-bracketing number if this holds for the classes of functions $\mathcal{G}^+$, $\mathcal{F}$, $\mathcal{H}^+$ and $\mathcal{H}^-$.

Lemma 3.2. *If Ψ and $\widetilde{\Psi}$ are P-Donsker classes of functions with ϵ-bracketing numbers of polynomial order, then it is also true for both $\Psi^+ = \Psi + \widetilde{\Psi}$ and $\Psi^- = \Psi - \widetilde{\Psi}$.*

Proof. The result follows from the inequality

$$N_{[]}\big(\epsilon, \Psi^\pm, L^2(P)\big) \le N_{[]}\big(\epsilon/2, \Psi, L^2(P)\big) \times N_{[]}\big(\epsilon/2, \widetilde{\Psi}, L^2(P)\big)$$

and the fact that Ψ and $\widetilde{\Psi}$ have polynomial bracketing numbers. □

The next lemma whose proof can also be found in [4], Section 10, gives sufficient conditions for a class of functions of the type of $\mathcal{F}$, $\mathcal{G}^+$, $\mathcal{H}^+$ and $\mathcal{H}^-$ to be Donsker with polynomial bracketing numbers.

Lemma 3.3. *A class of functions $\Psi = \{z \in \mathcal{Z} \mapsto \psi_{\theta,\xi}(z) \in \mathbb{R}; (\theta, \xi) \in \Theta \times [0, +\infty)\}$ is P-Donsker with ϵ-bracketing numbers of polynomial order if:*

a. *for all $z \in \mathcal{Z}$, the function $\psi_{\theta,\xi}(z)$ is non decreasing (or nonincreasing) in θ for any fixed $\xi \in [0, +\infty)$ and also non decreasing (or nonincreasing) in ξ for any fixed $\theta \in \Theta$,*

b. *$\forall \epsilon \in (0, 1]$, $\exists M_\epsilon \le \tilde{c}/\epsilon$ such that $\forall (\theta, \tilde{\theta}) \in \Theta^2$ we have*

$$\forall (\xi, \tilde{\xi}) \in [0, M_\epsilon]^2, \quad P(\psi_{\theta,\xi} - \psi_{\tilde{\theta},\tilde{\xi}})^2 \le c\big\|(\theta - \tilde{\theta}, \xi - \tilde{\xi})\big\|_2,$$
$$\forall (\xi, \tilde{\xi}) \in [M_\epsilon, \infty)^2, \quad P(\psi_{\theta,\xi} - \psi_{\tilde{\theta},\tilde{\xi}})^2 \le \epsilon$$

for some universal constants $(c, \tilde{c}) \in (0, \infty)^2$.

For any function $\Phi_\theta : t \in [0, M] \mapsto \Phi_\theta(t) \in \Xi \subset [0, \infty)$ for $\theta \in \Theta$, the class of functions $\widetilde{\Psi} = \{z \in \mathcal{Z} \mapsto \widetilde{\psi}_{\theta,t}(z) = \psi_{\theta,\Phi_\theta(t)}(z); (\theta, t) \in \Theta \times [0, M]\}$ is also P-Donsker with ϵ-bracketing numbers of polynomial order.

Now by combining the previous lemma with the generic form of the effective age functions found in Lemma 3.1 we arrive at the following proposition.

Proposition 3.2. *$\mathcal{F}$, $\mathcal{G}^+$, $\mathcal{H}^-$ and $\mathcal{H}^+$ are P-Donsker classes of functions with ϵ-bracketing numbers of polynomial order.*

Proof. The proof is based on Lemma 3.3. Using the notations of this lemma let us consider that $\widetilde{\psi}_{\theta,t}(z)$ is one of the functions $f_{\theta,t}(z)$, $g^+_{\theta,t}(z)$, $h^-_{\theta,t}(z)$, or $h^+_{\theta,t}(z)$ where by Lemma 3.1 $\Phi_\theta(t) = (-1)^d(t - sd)/\phi(\theta)$ is well defined if $\phi(\theta) \neq 0$. In Lemma 3.1, there are only two cases (in part (ii) for $\theta = 1$ and in part (iii) for $\theta = 0$) where $\phi(\theta) = 0$ which correspond to $g^+_{1,t}(z) = 1 + \sum_{j\ge 2} \mathbb{1}_{\{x_{j-1} < s\wedge\tau\}}$ and $h^+_{0,t}(z) = 0$, respectively, since $s \ge t \in [0, M]$. In both cases, the resulting function does not depend on t so that the Donsker property is not affected as we just have to add one function. For $(t_1, t_2) \in [0, s]^2$ and $\theta_1, \theta_2 \in \{\theta \in \Theta; \phi(\theta) \neq 0\}$, we set $\xi_1 = \Phi_{\theta_1}(t_1)$ and $\xi_2 = \Phi_{\theta_2}(t_2)$. For $z \in \mathcal{Z}$ we have

$$
\begin{aligned}
&\left|\widetilde{\psi}_{\theta_2,t_2}(z) - \widetilde{\psi}_{\theta_1,t_1}(z)\right| \\
&\quad\le \sum_{j\ge 1} |\mathbb{1}_{\{u_j \leqq \xi_2 - \tilde{v}^{\theta_2}_{j-1}(\boldsymbol{x}_{j-1})\}} - \mathbb{1}_{\{u_j \leqq \xi_1 - \tilde{v}^{\theta_1}_{j-1}(\boldsymbol{x}_{j-1})\}}|\mathbb{1}_{\{u_j\in[0,s]; x_{j-1}\le s\}} \\
&\quad\le \sum_{j\ge 1} \mathbb{1}_{\{\xi_1\wedge\xi_2 - \tilde{v}^{\theta_1}_{j-1}(\boldsymbol{x}_{j-1})\vee\tilde{v}^{\theta_2}_{j-1}(\boldsymbol{x}_{j-1}) \le u_j \le \xi_1\vee\xi_2 - \tilde{v}^{\theta_1}_{j-1}(\boldsymbol{x}_{j-1})\wedge\tilde{v}^{\theta_2}_{j-1}(\boldsymbol{x}_{j-1}); u_j\in[0,s]; x_{j-1}\le s\}},
\end{aligned}
$$

where $\leqq$ represents either $\le$ or $<$. In addition, notice that for any $(a_j)_{j\ge1} \in [0, 1]^{\mathbb{N}}$ we have $(\sum_{j\ge1} a_j)^2 \le 2\sum_{j=1}^{\infty} j a_j$. Let c_0 be the upper bound of the conditional pdf of U_j on $[0, s]$ (for all $j \ge 1$) given in Lemma 3.1. With the notation $\widetilde{\psi}_{\theta,t}(z) \equiv \psi_{\theta,\xi}(z)$, we have

$$
\begin{aligned}
&P(\psi_{\theta_2,\xi_2} - \psi_{\theta_1,\xi_1})^2 \\
&\quad\le 2c_0 \sum_{j=1}^{\infty} j \int_{\mathbb{R}^{j-1}} \left(|\xi_2 - \xi_1| + \left|\tilde{v}^{\theta_2}_{j-1}(\boldsymbol{x}_{j-1}) - \tilde{v}^{\theta_1}_{j-1}(\boldsymbol{x}_{j-1})\right|\right)\mathbb{1}_{\{x_{j-1}\le s\}} F_{\boldsymbol{X}_{j-1}}(d\boldsymbol{x}_{j-1}) \\
&\quad\le 2c_0\left(|\xi_2 - \xi_1| \sum_{j=1}^{\infty} j P(X_{j-1} \le s) + |\theta_2 - \theta_1| cs \sum_{j=1}^{\infty} j(j-1) P(X_{j-1} \le s)\right) \\
&\quad\le c_0' \left\|(\theta_2 - \theta_1, \xi_2 - \xi_1)\right\|_2,
\end{aligned}
$$

where the second to last inequality follows from Lemma 3.1 and the last from Assumption D1 and the norm equivalence.

In addition, Lemma 3.1 implies $U_j + \tilde{v}^{\theta}_{j-1}(\boldsymbol{X}_{j-1}) \le s(1 + cj)$. Consequently, for (ξ_1, ξ_2) in $[s(1 + ck), +\infty)^2$ and $k \in \mathbb{N}$, we have

$$
\begin{aligned}
&P(\psi_{\theta_2,\xi_2} - \psi_{\theta_1,\xi_1})^2 \\
&\quad\le P\left(\sum_{j>k} |\mathbb{1}_{\{U_j \leqq \xi_2 - \tilde{v}^{\theta_2}_{j-1}(\boldsymbol{X}_{j-1})\}} - \mathbb{1}_{\{U_j \leqq \xi_1 - \tilde{v}^{\theta_1}_{j-1}(\boldsymbol{X}_{j-1})\}}|\mathbb{1}_{\{U_j\in[0,s]; X_{j-1}\le s\}}\right)^2 \\
&\quad\le 2\sum_{j>k} (j-1) P(X_{j-1} \le s) \equiv \tilde{\epsilon}_{k+1}.
\end{aligned}
$$

By Assumption D1 $(\tilde{\epsilon}_k)_{k\ge1}$ is the remaining term of a convergent series with positive terms, thus it tends to 0 as k tends to infinity and for $\epsilon \in (0, 1]$, there exists $k_\epsilon \in \mathbb{N}$ such that $\tilde{\epsilon}_{k_\epsilon+1} < \epsilon \le \tilde{\epsilon}_{k_\epsilon}$. To complete the proof using Lemma 3.3, fixing $M_\epsilon = s(1 + c(k_\epsilon + 1))$ we have to show that ϵM_ϵ is bounded. Indeed

$$
\epsilon M_\epsilon \le \epsilon s\big(1 + c(k_\epsilon + 1)\big) \le \tilde{\epsilon}_{k_\epsilon} s\big(1 + c(k_\epsilon + 1)\big)
$$

$$\leq 2s\big(1+c(k_\epsilon+1)\big)\sum_{j>k_\epsilon}(j-1)P(X_{j-1}\leq s)$$
$$\leq 4s(1+c)\sum_{j\geq 1} j^2 P(X_j\leq s)<+\infty,$$

where the last inequality is again a consequence of Assumption D1. □

Verification of B2

To prove that Assumption B2 holds, which is done in [4], Section 8, we can follow the steps we used to show that Assumption D2 holds once we have a closed form expression for $\lambda^\theta(s,t)$ which is needed for B2(ii). To do so, recall that for $t\in[0,s]$ and $\theta\in\Theta$, we have $\lambda^\theta(s,t)=\frac{\partial \nu^\theta(s,t)}{\partial t}\times(y^\theta(s,t))^{-1}$ with

$$\nu^\theta(s,t)=P(X_1\leq s\wedge\mathcal{T}\wedge t)+\sum_{j\geq 2}P\big(\varepsilon^\theta_{j-1}(X_j)\leq t;X_j\leq s\wedge\mathcal{T}\big).$$

Notice that $\mathcal{G}^\theta_1(s,t)=f_0(t)S_{\mathcal{T}}(t)$ corresponds to the derivative of $t\mapsto P(X_1\leq s\wedge\mathcal{T}\wedge t)$. For $j\geq 2$, let $\mathcal{G}^\theta_j(s,t)$ denote the derivative of $t\mapsto P(\varepsilon^\theta_{j-1}(X_j)\leq t;X_j\leq s\wedge\mathcal{T})$. For $j\geq 1$ with $U_{j+1}=X_{j+1}-X_j$ and $\mathbf{Z}_j=(\mathcal{T},\boldsymbol{X}_j)$ we have

$$\begin{aligned}\mathcal{G}^\theta_{j+1}(s,t)&=\mathbb{E}\Big[\frac{d}{dt}P\,\big\{U_{j+1}\leq\big(t-\varepsilon^\theta_j(X_j)\big)\wedge(s\wedge\mathcal{T}-X_j)|\mathbf{Z}_j\big\}\mathbb{1}_{\{X_j\leq s\wedge\mathcal{T}\}}\Big]\\&=\mathbb{E}\big[f_{U_{j+1}|\boldsymbol{X}_j}\big(t-\varepsilon^\theta_j(X_j)\big)\mathbb{1}_{\{0<t-\varepsilon^\theta_j(X_j)\leq s\wedge\mathcal{T}-X_j;X_j\leq s\wedge\mathcal{T}\}}\big]\\&=\mathbb{E}\Big[\frac{f_0(t+\varepsilon^{\theta_0}_j(X_j)-\varepsilon^\theta_j(X_j))}{S_0(\varepsilon^{\theta_0}_j(X_j))}\mathbb{1}_{\{\varepsilon^\theta_j(X_j)<t\leq\varepsilon^\theta_j(s\wedge\mathcal{T});X_j\leq s\wedge\mathcal{T}\}}\Big],\end{aligned}\tag{3.4}$$

where the last equality follows from (10.1) in [4].

Verification of B3

The aim is to verify that ARA_m models are identifiable if equation (2.5) holds with probability one. That equation can be rewritten as

$$\begin{aligned}&\sum_{j\geq1}\log\lambda^\theta\big(s,e^\theta_{j-1}(x_j)\big)\mathbb{1}_{\{x_j\leq s\wedge\tau\}}\\&\qquad-\sum_{j\geq1}\big[\Lambda^\theta\big(s,e^\theta_{j-1}(x_j\wedge s\wedge\tau)\big)-\Lambda^\theta\big(s,e^\theta_{j-1}(x_{j-1})\big)\big]\mathbb{1}_{\{x_{j-1}\leq s\wedge\tau\}}\\&=\sum_{j\geq1}\log\lambda_0\big(e^{\theta_0}_{j-1}(x_j)\big)\mathbb{1}_{\{x_j\leq s\wedge\tau\}}\end{aligned}$$

$$-\sum_{j\geq 1}\big[\Lambda_0\big(e^{\theta_0}_{j-1}(x_j\wedge s\wedge\tau)\big)-\Lambda_0\big(e^{\theta_0}_{j-1}(x_{j-1})\big)\big]\mathbb{1}_{\{x_{j-1}\leq s\wedge\tau\}}, \tag{3.5}$$

and identifiability of the model corresponds to the fact that $\theta=\theta_0$ and $\lambda^\theta(s,t)=\lambda_0(t)$ for all $t\in[0,s]$. For every $c\in(0,s)$ and every $\epsilon>0$, let $B_c(\epsilon)$ be the interval $\{y\in[0,s]; |y-c|\leq\epsilon\}$. Identifiability in the random right censoring case will follow from the next Lemma. It is based on the fact that for ARA_m models with $m\in\overline{\mathbb{N}}$, the effective age functions restricted to the first and second inter arrival times do not depend on m and are defined by $e^\theta_0(t)=t$ for all $t\in[0,x_1]$, and $e^\theta_1(t)=t-\theta x_1$ for all $t\in(x_1,x_2]$.

Lemma 3.4. *Let equation (3.5) hold with probability one. Suppose that for all* $(c,c_1,\epsilon)\in(0,s)^3$ *satisfying* $c_1<c$ *and* $B_{c_1}(\epsilon)\cap B_c(\epsilon)=\varnothing$, *we have*

(a) $P(\{\mathbf{Z}\in\mathcal{Z}; \mathcal{T}<s\leq X_1, \mathcal{T}\in B_c(\epsilon)\})>0$,
(b) $P(\{\mathbf{Z}\in\mathcal{Z}; s<X_2, X_1\in B_{c_1}(\epsilon), \mathcal{T}\in B_c(\epsilon)\})>0$,

then the model parameters are identifiable.

Proof. Let us prove that $t\mapsto\lambda^\theta(s,t)$ is continuous on $[0,s]$ for any $\theta\in\Theta$. From the proof of Assumption B2, we can directly deduce the continuity for $\theta\in(0,1)$. Moreover, since the right-hand side of Inequality (8.1) in [4] with $\theta=\tilde{\theta}$ and $\phi(\theta)=0$ vanishes, the continuity also holds for $\theta\in\{0,1\}$.

Second, for $0\leq\mathcal{T}<s\leq X_1$, we obtain $\Lambda^\theta(s,\mathcal{T})=\Lambda_0(\mathcal{T})$ for all $\mathcal{T}\in[0,s]$ by (3.5). Assume that $\Lambda^\theta(s,t)\neq\Lambda_0(t)$ for some $t\in(0,s)$. Then by continuity of the functions $\Lambda^\theta(s,\cdot)$ (Assumption B2(ii)) and $\Lambda_0(\cdot)$ (Assumption E1) there exists an interval $B_t(\epsilon)$ around t such that $\Lambda^\theta(s,u)\neq\Lambda_0(u)$, $\forall u\in B_t(\epsilon)$. But then under Assumption (a), we see that (3.5) cannot hold with probability one. Hence, we must have $\Lambda^\theta(s,t)=\Lambda_0(t)$ for all $t\in[0,s]$. Since $t\mapsto\lambda^\theta(s,t)$ is continuous on $[0,s]$ and because by Assumption E1 λ_0 is continuous, we have $\lambda^\theta(s,t)=\lambda_0(t)$.

Finally, λ_0 is assumed to be nonconstant over $[0,s]$, then there exist $z\in(0,s]$ and $d>0$ such that $|\lambda_0(z)-\lambda_0(0)|>d$. Only the case where $\lambda_0(z)>\lambda_0(0)$ is considered, as the proof works similarly in the opposite case. In addition, $\lambda_0(\cdot)$ is continuous, then there exists two non empty, disjoint intervals $B_0(\epsilon)$ and $B_z(\epsilon)$, around respectively, 0 and z, such that $\lambda_0(x)\geq\lambda_0(y)+d/2$, $\forall x\in B_z(\epsilon)$, $\forall y\in B_0(\epsilon)$. Let us denote $c_1=2\epsilon/3$, $\tilde{\epsilon}=\epsilon/3$. Now on for $s<X_2$, $\mathcal{T}\in B_z(\tilde{\epsilon})$, $X_1\in B_{c_1}(\tilde{\epsilon})$, equation (3.5) reduces to

$$\begin{aligned}
0&=\big|\Lambda_0(\mathcal{T}-\theta X_1)-\Lambda_0\big((1-\theta)X_1\big)-\Lambda_0(\mathcal{T}-\theta_0X_1)+\Lambda_0\big((1-\theta_0)X_1\big)\big|\\
&=\left|\int_{\mathcal{T}-\theta_0X_1}^{\mathcal{T}-\theta X_1}\lambda_0(u)\,du-\int_{(1-\theta_0)X_1}^{(1-\theta)X_1}\lambda_0(u)\,du\right|\\
&=\left|\int_0^{(\theta_0-\theta)X_1}\lambda_0(\mathcal{T}+u-\theta_0X_1)-\lambda_0\big(u+(1-\theta_0)X_1\big)\,du\right|\\
&\geq|\theta_0-\theta|X_1\,d/2\geq|\theta_0-\theta|\epsilon d/6.
\end{aligned}$$

The inequality holds since $(\mathcal{T}+u-\theta_0X_1)\in B_z(\epsilon)$ and $(u+(1-\theta_0)X_1)\in B_0(\epsilon)$. Then, under Assumption (b), the last inequality necessarily implies $\theta=\theta_0$. $\square$

Finally, let us show that for $m \in \overline{\mathbb{N}}$ conditions (a) and (b) of Lemma 3.4 hold under the assumptions of Proposition 3.1. First for $(c, \epsilon) \in (0, s)^2$ we have

$$P\big(\mathcal{T} \in B_c(\epsilon); X_1 > s\big) = P\big(\mathcal{T} \in B_c(\epsilon)\big) \times S_0(s) > 0,$$

by independence of $\mathcal{T}$ and $\boldsymbol{X}$, Assumption E2, and assumptions on the distribution $\mathcal{T}$ in E2. Thus (a) holds.

In addition (b) holds since for $(c, c_1, \epsilon) \in (0, s)^3$ such that $c_1 < c$ and $B_{c_1}(\epsilon) \cap B_c(\epsilon) = \varnothing$ we have

$$\begin{aligned}
&P\big(X_1 \in B_{c_1}(\epsilon); \mathcal{T} \in B_c(\epsilon); X_2 > s\big) \\
&\quad = P\big(X_1 \in B_{c_1}(\epsilon); X_2 - X_1 > s - X_1\big) \times P\big(\mathcal{T} \in B_c(\epsilon)\big) \\
&\quad = \int_{(c_1-\epsilon)\vee 0}^{c_1+\epsilon} f_0(x_1) \int_{s-x_1}^{+\infty} \frac{f_0(v + (1-\theta_0)x_1)}{S_0((1-\theta_0)x_1)}\, dv\, dx_1 \times P\big(\mathcal{T} \in B_c(\epsilon)\big) \\
&\quad = \int_{(c_1-\epsilon)\vee 0}^{c_1+\epsilon} \frac{f_0(x_1) S_0(s - \theta_0 x_1)}{S_0((1-\theta_0)x_1)}\, dx_1 \times P\big(\mathcal{T} \in B_c(\epsilon)\big) \\
&\quad \geq S_0(s) \times \int_{(c_1-\epsilon)\vee 0}^{c_1+\epsilon} \lambda_0(x_1)\, dx_1 \times P\big(\mathcal{T} \in B_c(\epsilon)\big) \\
&\quad \geq S_0(s) \times \epsilon \times \inf_{t \in [0,s]} \lambda_0(t) \times P\big(\mathcal{T} \in B_c(\epsilon)\big) > 0.
\end{aligned}$$

3.2. Alternative censoring schemes

We also prove that Assumptions B1–B3 and D1 and D2 are satisfied under alternative censoring schemes [4], Section 9:

- Type-II censoring: $\mathcal{T} = X_k$ for some fixed $k \geq 2$. Since by construction of the estimation method the period of study ends at some fixed s, this censoring scheme, and the following one, can be viewed as hybrid censoring.
- Random Type-II censoring: $\mathcal{T} = X_N$ for a positive integer random variable N, independent of $\boldsymbol{X}$ and such that $P(N \geq 2) > 0$.
- Non random Type-I censoring: we can assume, without loss of generality, that $s = \mathcal{T}$. In this case, we need the additional assumption that λ_0 is monotone.

4. Numerical illustration

The objective of this section is to illustrate the consistency results for ARA$_m$ models. We shall focus on the estimation of θ_0 whose semiparametric estimation is the main contribution of the paper. In our simulations, the baseline hazard rate function corresponds to a shifted Weibull distribution defined by $\lambda_0(t) = 0.1 \times (t + 0.5)^2$ which satisfies Assumption E1. All simulation results are based on $N = 1000$ simulated samples. We consider two different censoring schemes:

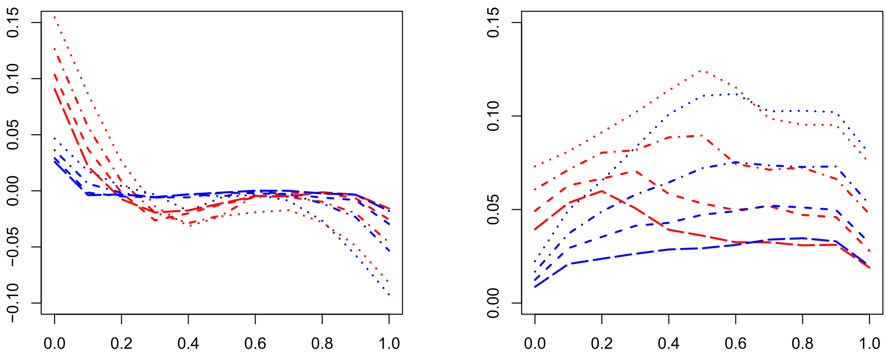

Figure 1. Empirical bias (left plot) and standard deviation (right plot) based on $N = 1000$ estimates of θ_0 varying in $[0, 1]$ for an ARA_1 model (red) and an ARA_∞ model (blue) with sample size $n = 50$ (dotted lines), $n = 100$ (dashed dotted lines), $n = 200$ (dashed lines), $n = 400$ (long dashed lines) for Type-I censoring with $\tau = s = 5$ and bandwidth $b = 0.3$.

Type-I censoring for which $\mathcal{T}_i = \tau$ for all $1 \le i \le n$ where τ is a constant, and Type-II censoring for which $\mathcal{T}_i = X_{i,k}$ for all $1 \le i \le n$ where k is an integer independent of i.

Consistency results are illustrated for both ARA_1 and ARA_∞ models, for four sample sizes $n \in \{50, 100, 200, 400\}$ and the two censoring schemes are Type-I with $s = \tau = 5$ and Type-II with $k = 4$. The information contained in both censoring schemes is of the same order since the expected number of events per trajectory is $E[N_5] \approx 6.2$ for $\theta = 0$, and $E[N_5] \approx 3.4$ for $\theta = 1$ (ARA_1 and ARA_∞ models are equivalent in both cases).

Figures 1 and 2 show the bias and the standard deviation of θ_n for θ_0 varying in $[0, 1]$, for several sample sizes, ARA_1 and ARA_∞ models and Type-I and Type-II censoring. All the sim-

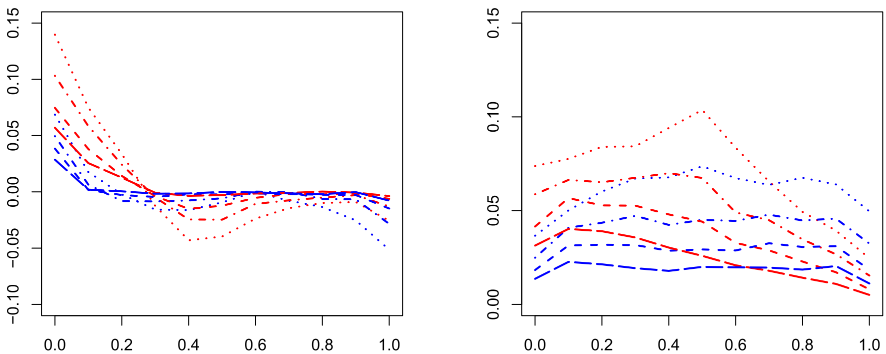

Figure 2. Empirical bias (left plot) and standard deviation (right plot) based on $N = 1000$ estimates of θ_0 varying in $[0, 1]$ for an ARA_1 model (red) and an ARA_∞ model (blue) with sample size $n = 50$ (dotted lines), $n = 100$ (dashed dotted lines), $n = 200$ (dashed lines), $n = 50$ (long dashed lines) for Type-II censoring with $k = 4$ (s sufficiently large) and bandwidth $b = 0.3$.

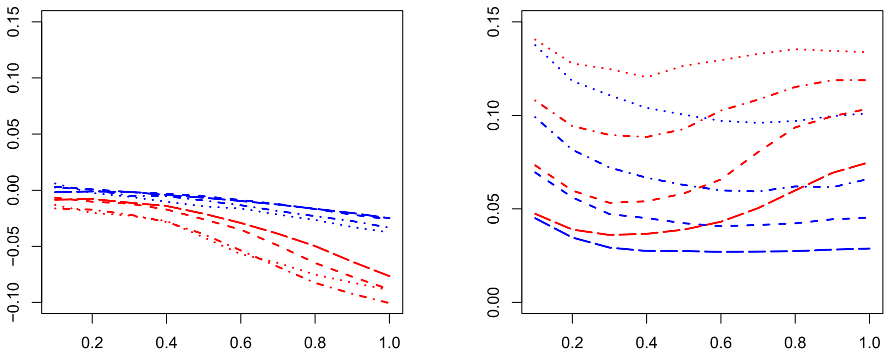

Figure 3. For the bandwidth b varying in [0.1, 1] calculation of the empirical bias (left plot) and standard deviation (right plot) of $N = 1000$ estimates of $\theta_0 = 0.5$ for an ARA_1 model (red) and an ARA_∞ model (blue) with sample size $n = 50$ (dotted lines), $n = 100$ (dashed dotted lines), $n = 200$ (dashed lines), $n = 50$ (long dashed lines) for Type-I censoring with $\tau = s = 5$.

ulation results are obtained for the same bandwidth fixed to 0.3. We observe that going from $n = 100$ to $n = 400$ reduces approximatively the standard deviation by a half whatever the model and the censoring scheme. It means that the usual root-of-n convergence rate probably holds for θ_n. We also observe that whatever the model the bias decreases as n increases. Fitting θ for the ARA_1 model seems to be more difficult than it is for an ARA_∞ model, especially in a neighborhood of $\theta_0 = 0$. Note that both the bias and the standard deviation vary with θ_0. For instance, the bias is considerably higher for θ close to the left-hand and right-end point of the parameter space $\Theta = [0, 1]$. However, one has to keep in mind that the bandwidth b is fixed to 0.3 whatever the sample size and the value of θ_0. Additional simulation results are available in Supplement B.1 for $b \in [0.1, 1]$.

Now by fixing the value of θ_0 to 0.5 and letting b vary in [0.1, 1], we see in Figure 3 (resp. Figure 4) that for both censoring schemes the bias is non increasing (resp. non decreasing) and the standard deviation is non monotonic (resp. nonincreasing). For instance, the bias decreases as b increases for Type-I censoring while the opposite is true for Type-II censoring. Note also that small biases do not necessarily occur jointly with small standard deviations. However, we can see in Supplement B.2 and Supplement B.3 (where we have the same graphs for $\theta_0 \in [0, 1]$) that the results in Figures 3 and 4 are rather specific to $\theta_0 = 0.5$. Globally these simulations show that the choice of the bandwidth has to be adapted to the unknown value of θ_0 and that it would ideally be depend on both the censoring scheme and on the underlying model.

The last simulation result is given in Figure 5 where the influence of s on the estimator of $\theta_0 = 0.5$ is studied under Type-II censoring. Additional simulation results are available in Supplement B.4. We can again see that it is more difficult to fit an ARA_1 model than an ARA_∞ model. If s is too small the quality of estimation gets poorer. Indeed in this case the information regarding θ, that is present in the $X_{i,j}$s only for $j \geq 2$, may be very poor. In fact, whatever the value of θ_0 (see Supplement B.4) the larger s, the smaller the bias and the standard deviations. It means, as expected, that we should not consider s to be a tuning parameter, the larger it is, the better.

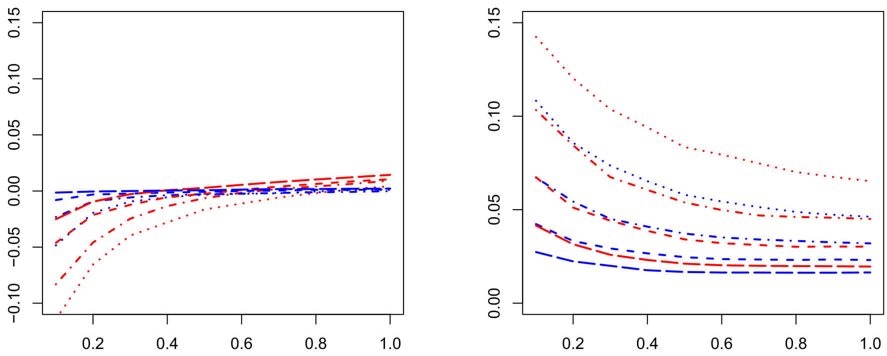

Figure 4. For the bandwidth b varying in $[0.1, 1]$ calculation of the empirical bias (left plot) and standard deviation (right plot) of $N = 1000$ estimates of $\theta_0 = 0.5$ for an ARA_1 model (red) and an ARA_∞ model (blue) with sample size $n = 50$ (dotted lines), $n = 100$ (dashed dotted lines), $n = 200$ (dashed lines), $n = 50$ (long dashed lines) for Type-II censoring with $k = 4$ (s sufficiently large).

5. Application to reliability data

In this section, we consider a real data set recently analyzed by Guerra de Toledo et al. [10] in the context of evaluating the maintenance policy for diesel engines of off-road mining trucks. The data are displayed in Supplement A.2. When an engine fails, it goes through corrective maintenance, which restores it to use conditions. During the repair the component that caused the failure is fixed. Moreover, the main components are reviewed and replaced if they are about to fail. Hence, just after the repair, the system is better than it was just before the failure, but worse than a brand-new one. In addition, some preventive maintenance actions are also implemented, they consist of a full overhaul of the engine, restoring it to the brand-new condition ("as good as

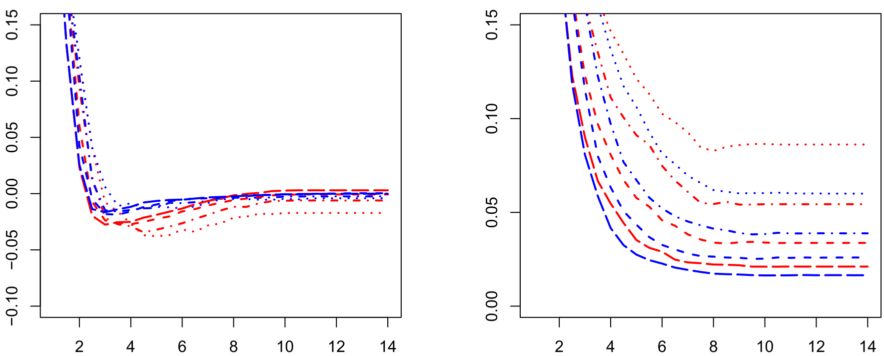

Figure 5. For s varying in $[1, 14]$ calculation of the empirical bias (left plot) and standard deviation (right plot) of $N = 1000$ estimates of $\theta_0 = 0.5$ for an ARA_1 model (red) and an ARA_∞ model (blue) with sample size $n = 50$ (dotted lines), $n = 100$ (dashed dotted lines), $n = 200$ (dashed lines), $n = 50$ (long dashed lines) for Type-II censoring with $k = 4$ and bandwidth $b = 0.5$.

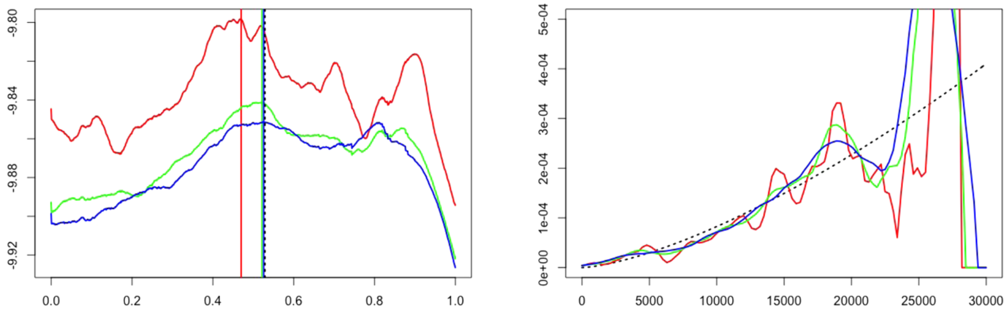

Figure 6. On the left-hand side are the profile likelihood functions for $b = 1000$ (red), $b = 2000$ (green) and $b = 3000$ (blue), the vertical lines correspond to the locations of the maxima whereas the black dashed vertical line corresponds to the parametric estimate in [10]. On the right-hand side are the corresponding estimates of the baseline hazard rate function, the black dashed curve corresponds to the parametric estimator in [10].

new”). Thus, after a preventive maintenance the system is considered as a new one. The analyzed data set consists of 193 sequences of event times. For each sequence, the terminal event is either a censoring time (51) due to a preventive maintenance, or the first (88), the second (43), the third (10) or the fourth (1) failure time. Guerra de Toledo et al. [10] fitted a parametric ARA_1 model with a power-law baseline hazard rate function. Here we fit the same model semiparametrically, that is without a parametric assumption on the baseline hazard rate function. The estimation results are summarized in Figure 6. First, we remark that for a bandwidth $b \in \{2000; 3000\}$ the parametric and semiparametric estimation results are rather close. However, we can see that for $b = 3000$ the profile likelihood function has two peaks. Moreover, the function values at the two peaks are almost equal. This indicates that our estimate for θ may dependent in an unsatisfactory way on the bandwidth. To analyze this further, we plotted in Figure 7 the function $b \mapsto \theta_n(b)$. In this figure, the estimates of θ based on the smoothed profile likelihood function (2.4) are displayed by stars.

The displayed circles correspond to the estimates of θ if we use in the profile likelihood function a double kernel based estimate for λ, that is,

$$\lambda_n^\theta(s,t) = \frac{b_n^{-1}\int_{\mathbb{R}} \kappa((t-u)/b_n)\bar{N}_n^\theta(s,du)}{\int_{\mathbb{R}} \mathcal{K}((t-u)/b'_n)\bar{Y}_n^\theta(s,du)}$$

for an additional bandwidth b'_n and $\mathcal{K}$ the cumulative distribution function associated to κ. In Figure 7, we used $b'_n = b_n$. Overall the results are rather close, however the estimate of θ based on the double kernel is less sensitive to the choice of the bandwidth. This can be seen from the fact that the estimate of θ jumps to 0.8 when the bandwidth exceeds 3500. Overall, this shows the need for research on robust data-driven methods to calibrate the bandwidth.

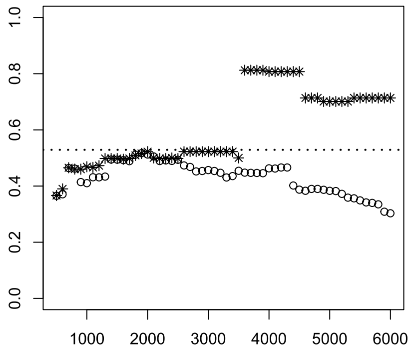

Figure 7. Maximizers of the smoothed profile likelihood function based on (2.4) (stars) or on (5) (circles).

6. Concluding remarks and perspectives

Virtual age models are powerful tools to describe recurrent events especially if one want to measure the evolution of successive inter-arrival time distributions. In this paper, we develop a consistent semiparametric estimation method that allows to overcome the deficiency of the standard profile likelihood to produce consistent estimators for a large class of virtual age models namely the ARA models. The application of empirical processes methods turns out to be a efficient way to study the asymptotic behavior of our estimators. This point deserves to be emphasized since it probably allows the study of large classes of estimation methods whenever it is possible to build estimators depending on the empirical measure based on i.i.d. random elements on sequence spaces. Additionally there are two main perspectives to this work which are complementary. The first one is theoretical, it concerns the study of the central limit behavior of estimators and the way to derive confidence intervals and bands for the unknown parameters of the model. A major difficulty in doing so is to find results analogously to Lemma 3.1 and Lemma 3.3 for the function classes involved in the derivative of $\ell_{n,s}$ w.r.t. θ. Without such results a unified treatment of ARA_m models with $m \in \overline{\mathbb{N}}$ and different censoring schemes seems not possible. The other one is both computational and theoretical since the simulation study has made clear that even if consistency can be illustrated numerically, the finite sample behavior of the estimators depends on the choice of the bandwidth b (and even two bandwidths if $t \mapsto \lambda_n^\theta(s,t)$ is estimated by (5)), for which a data-driven selection criterion should be provided.

Acknowledgements

The authors thank an associate editor and a referee for particularly helpful and insightful comments that led to an improved version of the manuscript.

Supplementary Material

Supplement A.1: Additional proofs (DOI: 10.3150/19-BEJ1140SUPPA; .pdf). This supplement material gives the proof of Lemmas 3.1 and 3.3, details the proof of Assumptions B2 and D1, and also demonstrates that Assumptions B1–B3 and D1–D2 are satisfied under alternative censoring schemes.

Supplement A.2: Data set (DOI: 10.3150/19-BEJ1140SUPPB; .pdf). This supplement material displays the data set considered in Section 5.

Supplement B.1: Numercial illustration (DOI: 10.3150/19-BEJ1140SUPPC; .zip). File "VS-theta_CensI.mp4" represents the empirical bias, standard deviation and Mean Square Error (MSE) of $N = 1000$ estimates of θ_0 versus the true parameter value, for an ARA_1 model (red) and an ARA_∞ model (blue) with sample size $n = 50$ (dotted lines), $n = 100$ (dashed dotted lines), $n = 200$ (dashed lines), $n = 400$ (long dashed lines) for Type-I censoring with $\tau = s = 5$ and bandwidth b varying in $[0.1, 1]$.

Supplement B.2: Numercial illustration (DOI: 10.3150/19-BEJ1140SUPPD; .zip). File "VS-theta_CensII.mp4" is similar to Supplement B.1, but for Type-II censoring with $k = 4$ (s sufficiently large).

Supplement B.3: Numercial illustration (DOI: 10.3150/19-BEJ1140SUPPE; .zip). File "VSb_CensI.mp4" is similar to Supplement B.1, but plots are versus bandwidth b, for Type-I censoring with $\tau = s = 5$ and a true parameter value θ_0 varying in $[0, 1]$.

Supplement B.4: Numercial illustration (DOI: 10.3150/19-BEJ1140SUPPF; .zip). File "VSb_CensII.mp4" is similar to Supplement B.3, but for Type-II censoring with $k = 4$ (s sufficiently large).

Supplement B.5: Numercial illustration (DOI: 10.3150/19-BEJ1140SUPPG; .zip). File "VSs_CensII.mp4" is similar to Supplement B.1, but plots are versus s, for Type-II censoring with $k = 4$ (s sufficiently large) and a true parameter value θ_0 varying in $[0, 1]$ and $b = 0.5$.

References

[1] Adekpedjou, A. and Stocker, R. (2015). A general class of semiparametric models for recurrent event data. *J. Statist. Plann. Inference* **156** 48–63. MR3267512 https://doi.org/10.1016/j.jspi.2014.07.012

[2] Andersen, P.K., Borgan, Ø., Gill, R.D. and Keiding, N. (1993). *Statistical Models Based on Counting Processes*. *Springer Series in Statistics*. New York: Springer. MR1198884 https://doi.org/10.1007/978-1-4612-4348-9

[3] Beutner, E., Bordes, L. and Doyen, L. (2017). The failure of the profile likelihood method for a large class of semi-parametric models. *Bernoulli* **23** 3650–3684. MR3654819 https://doi.org/10.3150/16-BEJ861

[4] Beutner, E., Bordes, L. and Doyen, L. (2020). Supplement to "Consistent semiparametric estimators for recurrent event times models with application to virtual age models." https://doi.org/10.3150/19-BEJ1140SUPPA, https://doi.org/10.3150/19-BEJ1140SUPPB, https://doi.org/10.3150/19-BEJ1140SUPPC, https://doi.org/10.3150/19-BEJ1140SUPPD, https://doi.org/

10.3150/19-BEJ1140SUPPE, https://doi.org/10.3150/19-BEJ1140SUPPF, https://doi.org/10.3150/19-BEJ1140SUPPG.

[5] Cook, R.J. and Lawless, J.F. (2007). *The Statistical Analysis of Recurrent Events*. *Statistics for Biology and Health*. New York: Springer. MR3822124

[6] Dorado, C., Hollander, M. and Sethuraman, J. (1997). Nonparametric estimation for a general repair model. *Ann*. *Statist*. **25** 1140–1160. MR1447744 https://doi.org/10.1214/aos/1069362741

[7] Doyen, L. and Gaudoin, O. (2004). Classes of imperfect repair models based on reduction of failure intensity or virtual age. *Reliab*. *Eng*. *Syst*. *Saf*. **84** 45–56.

[8] Dudley, R.M. (1999). *Uniform Central Limit Theorems*. *Cambridge Studies in Advanced Mathematics* **63**. Cambridge: Cambridge Univ. Press. MR1720712 https://doi.org/10.1017/CBO9780511665622

[9] Giné, E. and Nickl, R. (2016). *Mathematical Foundations of Infinite-Dimensional Statistical Models*. *Cambridge Series in Statistical and Probabilistic Mathematics*. New York: Cambridge Univ. Press. MR3588285 https://doi.org/10.1017/CBO9781107337862

[10] Guerra de Toledo, M.L., Freitas, M.A., Colosimo, E.A. and Gilardoni, G.L. (2016). Optimal periodic maintenance policy under imperfect repair: A case study of off-road engines. *IIE Trans*. **48** 747–758.

[11] Jacod, J. (1974/75). Multivariate point processes: Predictable projection, Radon–Nikodým derivatives, representation of martingales. *Z*. *Wahrsch*. *Verw*. *Gebiete* **31** 235–253. MR0380978 https://doi.org/10.1007/BF00536010

[12] Kijima, M. (1989). Some results for repairable systems with general repair. *J*. *Appl*. *Probab*. **26** 89–102. MR0981254 https://doi.org/10.1017/s0021900200041826

[13] Kijima, M., Morimura, H. and Suzuki, Y. (1988). Periodical replacement problem without assuming minimal repair. *European J*. *Oper*. *Res*. **37** 194–203. MR0963925 https://doi.org/10.1016/0377-2217(88)90329-3

[14] Lawless, J.F. (1995). The analysis of recurrent events for multiple subjects. *J*. *R*. *Stat*. *Soc*. *Ser*. *C*. *Appl*. *Stat*. **44** 487–498.

[15] Lindqvist, B.H. (2006). On the statistical modeling and analysis of repairable systems. *Statist*. *Sci*. **21** 532–551. MR2369984 https://doi.org/10.1214/088342306000000448

[16] Murphy, S.A. and van der Vaart, A.W. (2000). On profile likelihood. *J*. *Amer*. *Statist*. *Assoc*. **95** 449–485. With comments and a rejoinder by the authors. MR1803168 https://doi.org/10.2307/2669386

[17] Peña, E.A. (2006). Dynamic modeling and statistical analysis of event times. *Statist*. *Sci*. **21** 487–500. MR2369983 https://doi.org/10.1214/088342306000000349

[18] Peña, E.A. (2016). Asymptotics for a class of dynamic recurrent event models. *J*. *Nonparametr*. *Stat*. **28** 716–735. MR3555454 https://doi.org/10.1080/10485252.2016.1225733

[19] Peña, E.A., Slate, E.H. and González, J.R. (2007). Semiparametric inference for a general class of models for recurrent events. *J*. *Statist*. *Plann*. *Inference* **137** 1727–1747. MR2323859 https://doi.org/10.1016/j.jspi.2006.05.004

[20] Peña, E.A., Strawderman, R.L. and Hollander, M. (2001). Nonparametric estimation with recurrent event data. *J*. *Amer*. *Statist*. *Assoc*. **96** 1299–1315. MR1946578 https://doi.org/10.1198/016214501753381922

[21] Sellke, T. (1988). Weak convergence of the Aalen estimator for a censored renewal process. In *Statistical Decision Theory and Related Topics*, *IV*, *Vol*. 2 (*West Lafayette*, *Ind*., 1986) 183–194. New York: Springer. MR0927132

[22] Sellke, T. and Siegmund, D. (1983). Sequential analysis of the proportional hazards model. *Biometrika* **70** 315–326. MR0712020 https://doi.org/10.1093/biomet/70.2.315

[23] Shaked, M. and Shanthikumar, J.G. (2007). *Stochastic Orders*. *Springer Series in Statistics*. New York: Springer. MR2265633 https://doi.org/10.1007/978-0-387-34675-5

[24] van der Vaart, A.W. (1998). *Asymptotic Statistics*. *Cambridge Series in Statistical and Probabilistic Mathematics* **3**. Cambridge: Cambridge Univ. Press. MR1652247 https://doi.org/10.1017/CBO9780511802256

[25] van der Vaart, A.W. and Wellner, J.A. (1996). *Weak Convergence and Empirical Processes*: *With Applications to Statistics*. *Springer Series in Statistics*. New York: Springer. MR1385671 https://doi.org/10.1007/978-1-4757-2545-2
[26] Zeng, D. and Lin, D.Y. (2007). Efficient estimation for the accelerated failure time model. *J*. *Amer*. *Statist*. *Assoc*. **102** 1387–1396. MR2412556 https://doi.org/10.1198/016214507000001085
[27] Zeng, D. and Lin, D.Y. (2010). A general asymptotic theory for maximum likelihood estimation in semiparametric regression models with censored data. *Statist*. *Sinica* **20** 871–910. MR2682647

Received March 2018 and revised July 2019