The failure of the profile likelihood method for semi-parametric effective age models

Eric Beutner, Laurent Bordes, Laurent Doyen

The failure of the profile likelihood method for semi-parametric effective age models

Eric Beutner[a] Laurent Bordes[b] Laurent Doyen[c,*]

* Corresponding author

[a] Department of Quantitative Economics, Maastricht University, P.O. Box 616, NL-6200 MD Maastricht, The Netherlands, e.beutner@maastrichtuniversity.nl

[b] Laboratoire de Mathématiques et de leurs Applications, UMR CNRS 5142 Université de Pau et des Pays de l'Adour, France, laurent.bordes@univ-pau.fr

[c] Univ. Grenoble Alpes, LJK, F-38000 Grenoble, France CNRS, LJK, F-38000 Grenoble, France, laurent.doyen@imag.fr

* This work was partially supported by the French National Agency of Research (ANR), project AMMSI, number ANR-2011-BS01-021.

Abstract

We consider a semi-parametric model for recurrent events. The model consists of an unknown hazard rate function, the infinite-dimensional parameter of the model, and a parametrically specified effective age function. We will present a condition on the family of effective age functions under which the profile likelihood function evaluated at the parameter vector $\boldsymbol{\theta}$, say, exceeds the profile likelihood function evaluated at the parameter vector $\tilde{\boldsymbol{\theta}}$, say, with probability p. From this we derive a condition under which profile likelihood inference for the finite-dimensional parameter of the model leads to inconsistent estimates. Examples will be presented. In particular, we will provide an example where the profile likelihood function is monotone with probability one regardless of the true data generating process.

Keywords: Recurrent event data; Semi-parametric statistical model; Effective age process; Profile likelihood inference; virtual age process.

1. Introduction

Recurrent event data arise from the study of processes that generate events repeatedly over time. Such processes occur in many settings such as biomedicine, clinical trials and engineering to mention a few. For a list of references and some examples of recurrent event data see, for instance, Nelson (2003), Cook and Lawless (2007) and Aalen et al. (2008). In this article, our starting point is a semi-parametric model for recurrent events that was introduced by Peña and Hollander (2004); see also Equation (1) below. Among other things, the model incorporates the effects of interventions after each event occurrence through an effective age process (or virtual age process). Probably the best known effective age process is the one arising from a renewal process where, after each event occurrence, the effective age is set back to zero. For further information on effective age processes see also Hollander and Sethuramam (2004), Last and Szekli (1998), Lindqvist (2006), and Peña (2006). Statistical results for the model introduced in Peña and Hollander (2004) can be found in Peña et al. (2007) as well as in Dorado et al. (1997) and Adekpedjou and Stocker (2015) who consider sub-models for which they prove consistency and derive weak convergence results. See also Gärtner (2003) who considers a slightly different data collection process. The most general results on consistency and weak convergence were obtained very recently by Peña (2014) who restricts the general model given in Equation (1) below only by considering the case without frailties. In these articles it is assumed that the effective age function is entirely known. This implies that the way the interventions influence the effective age must be known by the statistician. Here we question whether this assumption can be weakened in a semi-parametric model. More precisely, we analyse whether the profile likelihood function can be used to derive consistent estimators when the effective age process is not assumed to be known but parametrically specified.

Inference based on the likelihood function and its variants has a long history; for general accounts and a recent review see, for instance, Barndorff-Nielsen (1988), Barndorff-Nielsen and Cox (1994), Davison (2003), Severini (2000) and Reid (2013). When the parameter is of the form (ζ, η) with η being a nuisance parameter, inference for ζ is often based on the profile likelihood function or modifications and adjustments to it. This approach has been applied in both parametric and semi-parametric problems; see, for example, Barndorff-Nielsen (1988), Barndorff-Nielsen and Cox (1994), Davison (2003), Fraser (2003), McCullagh and Tibshirani (1990), Reid (2013), Scott and Wild (1997), Severini (2000), Severini and Wong (1992), and Slud and Vonta (2005). For some semi-parametric models like Cox's proportional hazards model, asymptotic normality of the profile (partial) likelihood estimator has been known for awhile; see, for instance, Andersen et al. (1993) and Huang et al. (2012) for some recent extension. A general result in the semi-parametric context was proved by Murphy and van der Vaart (2000) who showed that profile likelihood inference for the finite-dimensional parameter behaves like ordinary likelihood inference whenever some functional-analytic conditions are satisfied; see also Hirosi (2011) who gave a weaker set of conditions. The result by Murphy and van der Vaart (2000) implies, for example, asymptotic normality of the profile likelihood estimator, and it was successfully applied by many authors and in different settings; see, for example, Breslow et al. (2003), Braekers and Veraverbeke (2005), Claeskens and Carroll (2007), Xu et al. (2009), and Zeng and Lin (2010). However, it is worth recalling that the standard approach, i.e. profiling out the infinite-dimensional parameter through a right-continuous step function, may lead to an inconsistent estimator for the finite-dimensional parameter of the model. For a single event model with covariates (accelerated failure time model) this can be easily seen; see, for instance, Zeng and Lin (2007).

As mentioned above, we address the following question: Suppose we profile out the infinite-dimensional parameter by using a right-continuous step function. Can we use the resulting profile likelihood function of the above mentioned semi-parametric model for recurrent events if the effective age process is parametrically specified? Here infinite-dimensional parameter refers to the integrated λ that is used in Model 2.1 below, i.e. by the infinite-dimensional parameter we mean the integrated hazard rate function. Denote the set in which the finite dimensional parameter lies by Θ, and let $\boldsymbol{\theta} \in \Theta$ and $\tilde{\boldsymbol{\theta}} \in \Theta$. We shall give a condition on the family of effective age processes under which the profile likelihood function at $\boldsymbol{\theta}$ is not less than at $\tilde{\boldsymbol{\theta}}$ with probability p, say. Additionally, we shall present an extension of this condition under which the profile likelihood function at $\boldsymbol{\theta}$ exceeds the profile likelihood function at $\tilde{\boldsymbol{\theta}}$ with probability p', say. From this one can easily derive a corollary providing conditions that rule out the possibility to obtain a consistent estimator based on the profile likelihood function. Examples will be presented to which the conditions given in our main results can be easily applied. In particular, we provide an example where p equals one regardless of the true probability measure and of the sample size. In the same example, we find a lower bound for p' that does not depend on the true probability measure or on the sample size. Still for the same example, we will infer from our main results that the profile likelihood function is monotonically decreasing with probability one whatever the true probability measure and the sample size. Furthermore, we present statistical models containing the renewal process and the non-homogeneous Poisson process as special cases for which it will turn out that the profile likelihood function at the parameter corresponding to the non-homogeneous Poisson process is never less than at the parameter corresponding to the renewal process regardless of whether the data come from a renewal process or an non-homogeneous Poisson process. The rest of this article is organized as follows. In Section 2 we define the model for recurrent events that we consider, explain its relation to the model introduced by Peña and Hollander (2004), detail how the profile likelihood is derived, and present our main results as well as examples to which they apply. Following the standard procedure the derivation of the profile likelihood is based on a formula valid if the true model is continuous, whereas profiling out the infinite-dimensional parameter is done w.r.t. a jump function see Section 2.2 and in particular Remark 2.5. In the literature this technique to profile out the infinite-dimensional parameter is often compared to a technique that profiles out the infinite-dimensional parameter using a formula valid for a 'discrete model'. We study this technique for the model considered in this article through a simulation study in Section 3. The theoretical derivations of this technique for the model considered here are carried out in Appendix B. Additional simulation results illustrating the conditions imposed in our main result are also presented in Section 3. Moreover, the results shown there will illustrate the decrease in the above mentioned example. All proofs are given in Appendix A. Appendix C contains results on the identifiability of the examples presented below.

2. Main result

Throughout, we shall use the following conventions: $\mathbb{N} := \{1, 2, \ldots\}$, $\mathbb{N}_0 := \mathbb{N} \cup \{0\}$, $\mathbb{R}_+ := \{x \in \mathbb{R} | x \geq 0\}$, the subscript $_0$ indicates the true, but unknown parameter, $x \wedge y$ stands for the minimum of x and y, and for a function f we denote by $f(x-)$ and $f(x+)$ the left-hand and right-hand limit of f at x, respectively. Convergence in probability is denoted by $\xrightarrow{\mathbb{P}}$ and $\overset{\mathbb{P}}{\nrightarrow}$ means that convergence in probability does not take place. For a simple counting process N we denote by $0 = S_0 < S_1 < S_2 < \ldots$ the sequence of jump times. In the next section we define the model we analyse, then in Section 2.2 we explain the estimators and show that the estimator for the cumulative hazard rate can be considered to be a non-parametric maximum likelihood estimator (NPMLE), and in Section 2.3 we give our main results.

2.1. The model

We will consider the following statistical model:

Model 2.1 *Let $N = \{N(s), 0 \leq s \leq s^*\}$, $s^* \in \mathbb{R}_+$, be a (simple) counting process on some measurable space $(\Omega, \mathcal{F})$ endowed with a filtration $\mathbf{F} = \{\mathcal{F}_s, 0 \leq s \leq s^*\}$ satisfying the usual conditions. Let $\mathbb{P}^{(\lambda,\boldsymbol{\theta})}$, $\lambda \in \Gamma, \boldsymbol{\theta} \in \Theta$, where Γ denotes the set of all hazard rate functions and $\Theta \subset \mathbb{R}^d$, be a set of probability measures on $(\Omega, \mathcal{F})$ such that under $\mathbb{P}^{(\lambda,\boldsymbol{\theta})}$ the $\mathbf{F}$-compensator $A = \{A(s), 0 \leq s \leq s^*\}$ of N is given by*

$$A(s) = \int_0^s Y(u)\lambda(\varepsilon^{\boldsymbol{\theta}}(u))\, du,$$

where the process $Y = \{Y(s), 0 \leq s \leq s^*\}$ is predictable, non-increasing and fulfils $Y(s) \in \{0,1\}$, $\forall s \in [0, s^*]$. For every $\boldsymbol{\theta} \in \Theta$ we have that $\varepsilon^{\boldsymbol{\theta}} = \{\varepsilon^{\boldsymbol{\theta}}(s), 0 \leq s \leq s^*\}$ is a predictable process with the following additional properties:

(a) $\varepsilon^{\boldsymbol{\theta}}(0,\omega) = c_0$ $\mathbb{P}^{(\lambda,\boldsymbol{\theta})}$*-a.s. for some* $c_0 \in \mathbb{R}_+$*;*

(b) $s \to \varepsilon^{\boldsymbol{\theta}}(s,\omega)$ *is* $\mathbb{P}^{(\lambda,\boldsymbol{\theta})}$*-a.s. non-negative;*

(c) We have $\mathbb{P}^{(\lambda,\boldsymbol{\theta})}$*-a.s. that* $s \to \varepsilon^{\boldsymbol{\theta}}(s,\omega)$ *is continuous on* $(S_{k-1}(\omega), S_k(\omega)]$*,* $k \in \mathbb{N}$*, and differentiable on* $(S_{k-1}(\omega), S_k(\omega))$*,* $k \in \mathbb{N}$*, with positive derivatives. The restriction of* $\varepsilon^{\boldsymbol{\theta}}$ *to the random time interval* $(S_{k-1}(\omega), S_k(\omega)]$ *is denoted by* $\varepsilon^{\boldsymbol{\theta}}_{k-1}$*,* $k \in \mathbb{N}$*.*

Examples for the Model 2.1 will be given below; see Examples 2.2, 2.3, and 2.4. Notice that, as usual, $N(s)$ denotes the number of events over the period $(0, s]$ for an observable unit, and the time of the ith recurrent event is denoted by S_i. The interpretation of the predictable process Y is as follows: the unit is still under observation, i.e. at risk, if and only if $Y(s) = 1$. In the following we assume that Y is of the form $Y = \{Y(s), 0 \leq s \leq s^*\}$ with $Y(s) = \mathbb{1}_{\{\tau \geq s\}}$, where τ is some positive (random) variable so that $[0, \tau]$ is the (random) observation interval. Here and in the following $\mathbb{1}$ denotes the indicator function. Further, we refer to $\varepsilon^{\boldsymbol{\theta}}$, $\boldsymbol{\theta} \in \Theta$, as the effective age process and to $\boldsymbol{\theta}$ as the effective age parameter. It describes the effect of interventions applied to the observational unit after experiencing a recurrent event. Notice that we do not require that $\varepsilon^{\boldsymbol{\theta}}(s,\omega) \leq s$. This means that the effective age might be larger than the current time s and that we do not exclude harmful interventions, i.e. interventions that increase the effective age instead of reducing it. To clarify the meaning of the effective age process let us consider two well-known processes:

1. Renewal process: Replacing the observational unit by a new one results in a renewal process with effective age process that equals $s - S_{k-1}(\omega)$ on $(S_{k-1}(\omega), S_k(\omega)]$ at time s;
2. Non-homogeneous Poisson process: Somehow on the other side of the spectrum is the non-homogeneous Poisson process with intensity function λ, because its effective age process at time s equals just s.

Model 2.1 is in the same spirit as the model introduced by Peña and Hollander (2004); see also Peña (2006) and Peña et al. (2007). In their model the compensator is assumed to be of the form

$$A(s|Z, \mathbf{X}(u), 0 \leq u \leq s) = \int_0^s Z\, Y(u)\, \rho(N(u-), \boldsymbol{\alpha})\, \psi(\boldsymbol{\beta}^T \mathbf{X}(u))\, \lambda(\varepsilon(u))\, du, \tag{1}$$

where Z is a frailty variable, ρ is a mapping from $\mathbb{N}_0$ to $\mathbb{R}_+$ of known functional form depending on some unknown parameter vector $\boldsymbol{\alpha} \in \mathfrak{A} \subset \mathbb{R}^p$ with $\rho(0; \boldsymbol{\alpha}) = 1$, for all $\boldsymbol{\alpha} \in \mathfrak{A}$, ψ is a known mapping from $\mathbb{R}$ to $\mathbb{R}_+$ with $\boldsymbol{\beta} \in \mathfrak{B} \subset \mathbb{R}^q$ an unknown parameter

vector and $\mathbf{X} = \{\mathbf{X}(s), 0 \leq s \leq s^*\}$ is an $\mathbb{R}^q$-valued stochastic process interpreted as the possibly time-varying covariates. The superscript ^{T} denotes the transpose. The predictable process $\varepsilon = \{\varepsilon(s), 0 \leq s \leq s^*\}$ fulfils the properties (a)–(c) mentioned in Model 2.1. The parameter of interest for the model defined by Equation (1) is thus $(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0, \lambda_0)$. As mentioned in the introduction statistical results on this model or sub-models of it can be found in Dorado et al. (1997), Peña et al. (2007), Adekpedjou and Stocker (2015), and Peña (2014). Thus, on one hand the statistical model introduced in Peña and Hollander (2004) is more general than Model 2.1, because Model 2.1 takes the functions ρ and ψ to be identically equal to one. On the other hand Model 2.1 is more general than the one introduced by Peña and Hollander, because it allows for a class of predictable processes $\varepsilon^{\boldsymbol{\theta}}$, $\boldsymbol{\theta} \in \Theta$, whereas Peña and Hollander and the above references assume that the process $\varepsilon = \{\varepsilon(s), 0 \leq s \leq s^*\}$ is known. Clearly, Model 2.1 could be extended to contain the model defined by Equation (1). However, as our focus here is on the extent to which profile likelihood inference for effective age models is possible we restrict ourselves to Model 2.1. At this point, it is worth mentioning that the model given by Equation (1) contains, for instance, Cox's proportional hazards model (with $Z \equiv 1, \rho(N(s-), \boldsymbol{\alpha}) \equiv 1, X(s) = \text{constant}, \psi(\boldsymbol{\beta}^T X) = \exp(\boldsymbol{\beta}^T X)$, and $\varepsilon(s) = s$) for which profile likelihood inference for $\boldsymbol{\beta}$ leads to a consistent and asymptotically normally distributed estimator; see, for instance, Andersen and Gill (1982).
We now present some examples for Model 2.1.

Example 2.2 *(ARA_1 or Kijima I with non-random repair)* For an ARA_1 model we have $\varepsilon^{\theta}_{k-1}(s, \omega) = s - \theta \cdot S_{k-1}(\omega)$ with $\Theta = [0, 1]$. This is the same model as a Kijima I model with non-random repair; see, for instance, Kijima et al. (1988), Kijima (1989) and Dorado et al. (1997). Notice that this model contains renewal processes with $\theta = 1$ and non-homogeneous Poisson processes with $\theta = 0$. $\diamond$

Example 2.3 *(ARA_∞ or Kijima II with non-random repair)* For an ARA_∞ model we have: $\varepsilon^{\theta}_{k-1}(s, \omega) = s - \theta \sum_{l=1}^{k-1} (1-\theta)^{k-1-l} S_l(\omega)$ with $\Theta = [0, 1]$. This is the same model as a Kijima II model with non-random repair; see again Kijima et al. (1988), Kijima (1989) and Dorado et al. (1997). Clearly, as in Example 2.2 we see that for $\theta = 1$ we get renewal processes and taking $\theta = 0$ results in non-homogeneous Poisson processes. $\diamond$

Example 2.4 Gonzáles et al. (2005) fitted a sub-model of the one given in Example 2.3 taking into account covariates via the equation as given in (1) to the data of 63 patients having a subtype of indolent non-Hodgkins lymphomas. They restricted $\Theta = [0, 1]$ in Example 2.3 to the discrete set $\{0, 0.5, 1\}$. Then a 0 stands for no response to the therapy/intervention, 0.5 means a partial remission, and 1 indicates a perfect intervention. $\diamond$

The notion of ARA (Arithmetic Reduction of Age) has been introduced in Doyen and Gaudoin (2004). Further examples for effective age processes can be found in Doyen and Gaudoin (2004), Lindqvist (2006), and Peña et al. (2007).

2.2. The estimators and NPMLE

We now begin by introducing a profile likelihood method for estimating the unknown parameter vector $\boldsymbol{\theta}$. Let $N_1, \ldots, N_m$ be m independent copies of a counting process as described in Model 2.1. For each counting process denote by Y_i, $1 \leq i \leq m$, the process as introduced in Model 2.1 and let $0 = S_{i,0} < S_{i,1} < \ldots$ be the jump times of the process N_i. By $\varepsilon_i^{\boldsymbol{\theta}}$ we denote the predictable process arising in the definition of N_i, $1 \leq i \leq m$, and by $\varepsilon_{i,j-1}^{\boldsymbol{\theta}}$ its restriction to the time interval $(S_{i,j-1}, S_{i,j}]$. Following Jacod (1975) (see also Andersen et al. (1993, II.7) and Peña et al. (2007)) the full likelihood process equals

$$
\begin{aligned}
L_{m,F}(s|\lambda, \varepsilon^{\boldsymbol{\theta}}, \mathbf{D}_m(s)) = & \prod_{i=1}^{m} \prod_{u=0}^{s} \left[Y_i(u) \lambda(\varepsilon_i^{\boldsymbol{\theta}}(u)) \right]^{N_i(\Delta u)} \\
& \times \exp\left[- \sum_{i=1}^{m} \int_0^s Y_i(u) \lambda(\varepsilon_i^{\boldsymbol{\theta}}(u)) \, du \right], \qquad (2)
\end{aligned}
$$

where at time s the data $\mathbf{D}_m(s)$ equals $\mathbf{D}_m(s) := \{N_1(u), \ldots, N_m(u), Y_1(u), \ldots, Y_m(u), 0 \leq u \leq s\}$. To obtain the profile likelihood function from (2) we first introduce an estimator for Λ_0 the cumulative hazard rate of λ_0. In doing so, following a technique of Peña et al. (2001) who extended an idea of Gill (1981) and Selke (1988), we define double indexed processes; see also Selke and Siegmund (1983) who seem to be the first to consider double indexed processes in survival analysis. Below we demonstrate that the resulting estimator may be considered to be a NPMLE. Firstly, define the double indexed process $N_i^{\boldsymbol{\theta}}$, $1 \leq i \leq m$, by

$$
N_i^{\boldsymbol{\theta}}(s,t) := \int_0^s Z_i^{\boldsymbol{\theta}}(u,t) \, dN_i(u), \quad 0 \leq s \leq s^*, \, 0 \leq t < \infty,
$$

with $Z_i^{\boldsymbol{\theta}}(u,t) := \mathbb{1}_{\{\varepsilon_i^{\boldsymbol{\theta}}(u) \leq t\}}, 1 \leq i \leq m$. $N_i^{\boldsymbol{\theta}}(s,t)$ denotes the number of events over the period $(0, s]$ for the ith unit whose effective age at time of occurrence was at most t. Thus, the first time variable s of $N_i^{\boldsymbol{\theta}}$ stands for the observation time and the second time variable t for the effective age time. Notice that $N_i^{\boldsymbol{\theta}}$ depends on the effective age parameter $\boldsymbol{\theta}$ in contrast to N_i. Secondly, we define what has been called the adjusted at risk process (or generalized at risk process) $Y_i^{\boldsymbol{\theta}} = \{Y_i^{\boldsymbol{\theta}}(s,t), 0 \leq s \leq s^*, 0 \leq t < \infty\}$ by

$$
\begin{aligned}
Y_i^{\boldsymbol{\theta}}(s,t) := \sum_{j=1}^{N_i(s-)} & \gamma_{i,j-1}^{\boldsymbol{\theta}}(t) \cdot \mathbb{1}_{(\varepsilon_{i,j-1}^{\boldsymbol{\theta}}(S_{i,j-1}+), \varepsilon_{i,j-1}^{\boldsymbol{\theta}}(S_{i,j})]}(t) \\
& + \gamma_{i,N_i(s-)}^{\boldsymbol{\theta}}(t) \cdot \mathbb{1}_{(\varepsilon_{i,N_i(s-)}^{\boldsymbol{\theta}}(S_{i,N_i(s-)}+), \varepsilon_{i,N_i(s-)}^{\boldsymbol{\theta}}(s \wedge \tau_i)]}(t),
\end{aligned} \tag{3}
$$

where the functions $\gamma_{i,j-1}^{\boldsymbol{\theta}}$ are defined by

$$
\gamma_{i,j-1}^{\boldsymbol{\theta}}(t) := \frac{1}{\left(\varepsilon_{i,j-1}^{\boldsymbol{\theta}}\right)' \left(\left(\varepsilon_{i,j-1}^{\boldsymbol{\theta}}\right)^{-1}(t)\right)}
$$

with $(\varepsilon_{i,j-1}^{\boldsymbol{\theta}})'$ denoting the derivative of $\varepsilon_{i,j-1}^{\boldsymbol{\theta}}$ w.r.t. observational time s and $(\varepsilon_{i,j-1}^{\boldsymbol{\theta}})^{-1}$ denoting the inverse w.r.t. observational time. With the help of $Y_i^{\boldsymbol{\theta}}$ one can rewrite the integral arising in the full likelihood (cf. Equation (2)) in terms of λ evaluated at the observational time s instead of at the effective age time $\varepsilon_i(s)$; see Equation (6) below. Notice that the jth event of the ith unit contributes to the risk set at the time pair (s,t) if it fulfils three conditions: Firstly, it occurred during the observation period [0,s), secondly the effective age of the jth event of the ith unit is larger than or equal to t, and thirdly the effective age of the ith unit immediately after the intervention succeeding the $(j-1)$th event is less than t. For further information on the adjusted at-risk process see Peña (2014). We define

$$
\mathcal{S}_m^{\boldsymbol{\theta}}(s,t) := \sum_{i=1}^{m} Y_i^{\boldsymbol{\theta}}(s,t). \tag{4}
$$

For fixed $\boldsymbol{\theta}$ we define the following method-of-moments estimator $\widehat{\Lambda}_m$ for Λ_0:

$$
\widehat{\Lambda}_m(s,t|\boldsymbol{\theta}) := \int_0^t \frac{J_m^{\boldsymbol{\theta}}(s,u)}{S_m^{\boldsymbol{\theta}}(s,u)} \left[\sum_{i=1}^{m} N_i^{\boldsymbol{\theta}}(s,du)\right], \tag{5}
$$

where $J_m^{\boldsymbol{\theta}}(s,u) := \mathbb{1}_{\{S_m^{\boldsymbol{\theta}}(s,u)>0\}}$. A justification for calling $\widehat{\Lambda}_m$ a method-of-moment estimator can be found in Peña et al. (2007) after their Proposition 1. Moreover, for $\boldsymbol{\theta}$ known and s fixed this estimator is consistent and converges, after being suitably normalized, to a Gaussian process (cf. Peña (2014)).

We now demonstrate that for observational time s and effective age parameter $\boldsymbol{\theta}$ fixed the estimator $\widehat{\Lambda}_m(s,t|\boldsymbol{\theta})$ can be seen to be a NPMLE. For this, notice first of all that the substitution rule

$$
\int_a^b f(x)\,dx = \int_{\phi(a)}^{\phi(b)} f(\phi^{-1}(x))(\phi^{-1}(x))'\,dx
$$

implies (with $f = \lambda \circ \varepsilon^{\boldsymbol{\theta}}$ and $\phi = \varepsilon^{\boldsymbol{\theta}}$) that the full likelihood (2) can be written as

$$L_{m,F}(s|\lambda, \varepsilon^{\boldsymbol{\theta}}, \mathbf{D}_m(s)) = \prod_{i=1}^{m} \prod_{u=0}^{s} \left[Y_i(u) \lambda(\varepsilon_i^{\boldsymbol{\theta}}(u)) \right]^{N_i(\Delta u)} \times \exp\left[- \int_0^\infty S_m^{\boldsymbol{\theta}}(s,u)\, d\Lambda(u) \right]. \tag{6}$$

Now if we take Λ to be a jump function with jumps at $\varepsilon_{k,\ell-1}^{\boldsymbol{\theta}}(S_{k,\ell})$, $1 \le k \le m$, $1 \le \ell \le N_k(s)$, i.e. the effective age of the kth unit at the time of the ℓth event, and if we denote these jumps by $\lambda_{k,\ell}^{\boldsymbol{\theta}}$ the log of the full likelihood becomes

$$\begin{aligned} \log\left(L_{m,F}(s|\lambda^{\boldsymbol{\theta}}, \varepsilon^{\boldsymbol{\theta}}, \mathbf{D}_m(s)) \right) &= \sum_{k=1}^{m} \sum_{\ell=1}^{N_k(s)} \log\left(\lambda_{k,\ell}^{\boldsymbol{\theta}} \right) \\ &- \sum_{k=1}^{m} \sum_{\ell=1}^{N_k(s)} \lambda_{k,\ell}^{\boldsymbol{\theta}} \left[\sum_{(i,j) \in I_{k,\ell}^{\boldsymbol{\theta}}} \gamma_{i,j-1}^{\boldsymbol{\theta}}(\varepsilon_{k,\ell-1}^{\boldsymbol{\theta}}(S_{k,\ell})) \right] \\ &- \sum_{k=1}^{m} \sum_{\ell=1}^{N_k(s)} \lambda_{k,\ell}^{\boldsymbol{\theta}} \left[\sum_{i \in I_k^{\boldsymbol{\theta},\tau_k}} \gamma_{i,N_i(s-)}^{\boldsymbol{\theta}} \left(\varepsilon_{k,\ell-1}^{\boldsymbol{\theta}}(S_{k,\ell}) \right) \right], \end{aligned} \tag{7}$$

where for every pair (k,ℓ), $1 \le k \le m, 1 \le \ell \le N_k(s)$, the sets $I_{k,\ell}^{\boldsymbol{\theta}}$ are defined by

$$I_{k,\ell}^{\boldsymbol{\theta}} := \{(i,j), 1 \le i \le m, 1 \le j \le N_i(s-) | \varepsilon_{i,j-1}^{\boldsymbol{\theta}}(S_{i,j-1}+) < \varepsilon_{k,\ell-1}^{\boldsymbol{\theta}}(S_{k,\ell}) \le \varepsilon_{i,j-1}^{\boldsymbol{\theta}}(S_{i,j})\}$$

and for every k, $1 \le k \le m$, the sets $I_k^{\boldsymbol{\theta},\tau_k}$ are defined by

$$I_k^{\boldsymbol{\theta},\tau_k} := \{i, 1 \le i \le m | \varepsilon_{i,N_i(s-)}^{\boldsymbol{\theta}}(S_{i,N_i(s-)}+) < \varepsilon_{k,\ell-1}^{\boldsymbol{\theta}}(S_{k,\ell}) \le \varepsilon_{i,N_i(s-)}^{\boldsymbol{\theta}}(s \wedge \tau_i)\}.$$

Clearly, maximizing (7) with respect to $\lambda_{k,\ell}^{\boldsymbol{\theta}}$ has solution given by

$$\lambda_{k,\ell}^{\boldsymbol{\theta}} = \frac{1}{\sum_{(i,j) \in I_{k,\ell}^{\boldsymbol{\theta}}} \gamma_{i,j-1}^{\boldsymbol{\theta}}(\varepsilon_{k,\ell-1}^{\boldsymbol{\theta}}(S_{k,\ell})) + \sum_{i \in I_k^{\boldsymbol{\theta},\tau_k}} \gamma_{i,N_i(s-)}^{\boldsymbol{\theta}} \left(\varepsilon_{k,\ell-1}^{\boldsymbol{\theta}}(S_{k,\ell}) \right)}. \tag{8}$$

Hence, upon substituting the NPMLE $\widehat{\Lambda}_m$ for Λ in the full likelihood we obtain from Equation (6) that for every fixed $\boldsymbol{\theta}$ the resulting log profile likelihood function $\ell_{m,P}$, up to a constant, equals

$$\ell_{m,P}(s|\boldsymbol{\theta}, \widehat{\Lambda}_m, \mathbf{D}_m(s)) = -\int_0^s \sum_{i=1}^{m} \log\left(\mathcal{S}_m^{\boldsymbol{\theta}}(s, \varepsilon_i^{\boldsymbol{\theta}}(w)) \right) dN_i(w),$$

because replacing Λ by $\widehat{\Lambda}_m$ in the argument of the exponential function in the full likelihood (cf. Equation 2), we obtain

$$\exp\left[-\int_0^\infty S_m^{\boldsymbol{\theta}}(s,u)\,d\widehat{\Lambda}(u)\right] = \exp\left[-\sum_{i=1}^m N_i(s)\right].$$

It is worth mentioning that the argument of exp being free of the finite-dimensional parameter is not a peculiarity of the model we consider here. For instance, for Cox's proportional hazards model, when plugging in the NPMLE into the full likelihood the argument of the exponential function is free of the regression parameter; see, for instance, Johansen (1983).

Remark 2.5 As mentioned in the introduction we followed the standard procedure to derive a NPMLE to profile out the infinite-dimensional component that is to say our NPMLE was taken to be a jump function whereas formula (2) is valid only in the continuous case. It is, therefore, common to additionally study the profile likelihood function if instead of formula (2) the corresponding formula for the discrete case is used. We shall do the same here. In Appendix B we derive all relevant formulas as well as the resulting profile likelihood function and in Section 3 this profile likelihood function is studied through a simulation study. $\diamond$

Remark 2.6 Notice that (8) boils down to well-known NPMLEs in special cases. Recall that the model in Example 2.2 contains, for instance, the renewal process (corresponding to $\theta = 1$) with effective age process $s - S_{k-1}(\omega)$ and the non-homogeneous Poisson process (corresponding to $\theta = 0$) whose effective age process equals s during the observation period. In these cases noting that $\gamma_{i,j-1} \equiv 1$ we obtain from (8) with $T_{i,j} := S_{i,j} - S_{i,j-1}$ and $\tau_{(i)}$ the ith smallest value among the $\tau_1, \ldots, \tau_m$ the well-known NPMLEs

$$\lambda_{k,\ell}^{\theta=1} = \frac{1}{\sum_{i=1}^m \sum_{j=1}^{N_i(s-)} \mathbb{1}_{\{T_{k,\ell} \le T_{i,j}\}} + \sum_{i=1}^m \mathbb{1}_{\{T_{k,\ell} \le (s\wedge\tau_i) - S_{i,N_i(s-)}\}}},$$

and

$$\lambda_{k,\ell}^{\theta=0} = \frac{1}{m-i+1}, \quad \text{whener } \tau_{(i-1)} < s_{k,\ell} \le \tau_{(i)}$$

with $\tau_{(0)} := 0$; see, for instance, Peña et al. (2001) and Lawless (1995), respectively. $\diamond$

2.3. Main result

In this section we present our main result. We first state an assumption that is needed in the theorem and provide examples when the assumption is satisfied. We further analyse this assumption in the simulation study in Section 3. The other assumptions made in our main result are illustrated below Theorem 2.10.

Assumption 2.7 *For $Y_i^{\boldsymbol{\theta}}$, $1 \le i \le m$, as defined in (3) we have for every $\boldsymbol{\theta} \in \Theta$ with probability 1*

$$Y_i^{\boldsymbol{\theta}}(s^*, t) = \sum_{j=1}^{J_i(s^*)} \gamma_{i,j-1}^{\boldsymbol{\theta}}(t) \cdot \mathbb{1}_{(\varepsilon_{i,j-1}^{\boldsymbol{\theta}}(S_{i,j-1}+), \varepsilon_{i,j-1}^{\boldsymbol{\theta}}(S_{i,j})]}(t), \quad 0 \le t < \infty,$$

where $J_i(s^*)$, $1 \le i \le m$, are random variables taking values in $\mathbb{N}_0$.

Assumption 2.7 means that $\gamma_{i,j-1}^{\boldsymbol{\theta}}(t) \cdot \mathbb{1}_{(\varepsilon_{iN_i(s-)}^{\boldsymbol{\theta}}(S_{iN_i(s-)}+), \varepsilon_{iN_i(s-)}^{\boldsymbol{\theta}}(s^* \wedge \tau_i)]}(t)$ is either of no relevance for $Y_i^{\boldsymbol{\theta}}(s^*, t)$ or of the form $\gamma_{i,j-1}^{\boldsymbol{\theta}}(t) \cdot \mathbb{1}_{(\varepsilon_{i,j-1}^{\boldsymbol{\theta}}(S_{i,j-1}+), \varepsilon_{i,j-1}^{\boldsymbol{\theta}}(S_{i,j})]}(t)$. We now give two examples that fulfil Assumption 2.7.

Example 2.8 *(Type-II censoring)* Let $\tau_i = S_{i,n_i}$ with $n_i \in \mathbb{N}$, $1 \le i \le m$, and $s^* \ge \max_i \tau_i$. In this case we have $s^* \wedge \tau_i = s^* \wedge S_{i,n_i} = S_{i,n_i}$. Moreover, if $s^* > S_{i,n_i}$, then $(\varepsilon_{i,N_i(s^*-)}^{\boldsymbol{\theta}}(S_{i,N_i(s^*-)}+), \varepsilon_{i,N_i(s^*-)}^{\boldsymbol{\theta}}(s^* \wedge \tau_i)]$ equals the empty set and the representation in Assumption 2.7 holds with $J_i(s^*) = N_i(s^*-) = N_i(s^*)$. Finally, if $s^* = S_{i,n_i}$, then we have

$$(\varepsilon_{i,N_i(s^*-)}^{\boldsymbol{\theta}}(S_{i,N_i(s^*-)}+), \varepsilon_{i,N_i(s^*-)}^{\boldsymbol{\theta}}(s^* \wedge \tau_i)] = (\varepsilon_{i,N_i(s^*-)}^{\boldsymbol{\theta}}(S_{i,N_i(s^*-)}+), \varepsilon_{i,N_i(s^*-)}^{\boldsymbol{\theta}}(S_{i,N_i(s^*)})],$$

and the representation in Assumption 2.7 holds with $J_i(s^*) = N_i(s^*-) + 1 = N_i(s^*)$. $\diamond$

Example 2.9 *(Compact support and finite number of interventions)* Let λ_0 be such that $\int_0^v \lambda_0(u)\, du = \infty$ for some $v \in \mathbb{R}_+$. Additionally, suppose that we consider the model of Example 2.2 with $\tau_i = s^* \wedge S_{i2}$, where $s^* > 2v$. By definition of τ_i we observe at most two events for the ith unit. Moreover, in the model of Example 2.2, the largest value we can observe for S_{i2} equals $v + [v - v(1-\theta)] = v + \theta v$ which is maximal for $\theta = 1$. Hence, since $s^* > 2v$ and $\int_0^v \lambda_0(u)\, du = \infty$, we have $\tau_i = S_{i2}$, $1 \le i \le m$, so that

$$(\varepsilon_{i,N_i(s^*-)}^{\boldsymbol{\theta}}(S_{i,N_i(s^*-)}+), \varepsilon_{i,N_i(s^*-)}^{\boldsymbol{\theta}}(s^* \wedge \tau_i)] = (\varepsilon_{i,2}^{\boldsymbol{\theta}}(S_{i,2}+), \varepsilon_{i,2}^{\boldsymbol{\theta}}(S_{i,2})]$$

equals the empty set, and the representation in Assumption 2.7 holds with $J_i(s^*) = N_i(s^*-)$. $\diamond$

We now state our main result whose proof, as mentioned in the introduction, is given in Appendix A.

Theorem 2.10 *Let Assumption 2.7 be satisfied, N_i, $1 \le i \le m$, etc. be as above, and denote by $(\mathbb{P}^{\lambda_0, \boldsymbol{\theta}_0})^m$ the m-fold product measure of $\mathbb{P}^{\lambda_0, \boldsymbol{\theta}_0}$. Moreover, let $\boldsymbol{\theta}$ and $\tilde{\boldsymbol{\theta}}$ be such that there exists a $c > 0$ with the following property: For every $t \ge 0$ we have that $(\mathbb{P}^{\lambda_0, \boldsymbol{\theta}_0})^m(\gamma_{i,j-1}^{\boldsymbol{\theta}}(t) \le c, \gamma_{i,j-1}^{\tilde{\boldsymbol{\theta}}}(t) \ge c, 1 \le i \le m, 1 \le j \le J_i(s^*), 0 \le t < \infty) = 1$. Then*

(a) *Denote by* $A_{m,\boldsymbol{\theta},\tilde{\boldsymbol{\theta}}}$ *the set of all* ω*'s such that for all pairs* (i,j), $1 \leq i \leq m$, $1 \leq j \leq J_i(s^*)$, *and all pairs* (k,ℓ) $1 \leq k \leq m$, $1 \leq \ell \leq J_k(s^*)$, *we have that*

$$\varepsilon^{\boldsymbol{\theta}}_{i,j-1}(S_{i,j-1}(\omega)) < \varepsilon^{\boldsymbol{\theta}}_{k,\ell-1}(S_{k,\ell}(\omega))$$

implies that

$$\varepsilon^{\tilde{\boldsymbol{\theta}}}_{i,j-1}(S_{i,j-1}(\omega)) < \varepsilon^{\tilde{\boldsymbol{\theta}}}_{k,\ell-1}(S_{k,\ell}(\omega)).$$

Then we have

$$(\mathbb{P}^{\lambda_0,\boldsymbol{\theta}_0})^m \left(\ell_{P,m}\left(s^*|\boldsymbol{\theta},\widehat{\Lambda}_m,\mathbf{D}_m(s)\right) \geq \ell_{P,m}\left(s^*|\tilde{\boldsymbol{\theta}},\widehat{\Lambda}_m,\mathbf{D}_m(s)\right)\right) \geq (\mathbb{P}^{\lambda_0,\boldsymbol{\theta}_0})^m\left(A_{m,\boldsymbol{\theta},\tilde{\boldsymbol{\theta}}}\right).$$

(b) *Denote by* $B_{m,\boldsymbol{\theta},\tilde{\boldsymbol{\theta}}}$ *the set of all* $\omega \in A_{m,\boldsymbol{\theta},\tilde{\boldsymbol{\theta}}}$ *for which we additionally have that there are at least two pairs* $(\underline{i},\underline{j})$, $1 \leq \underline{i} \leq m$, $1 \leq \underline{j} \leq J_{\underline{i}}(s^*)$, *and* $(\underline{k},\underline{\ell})$, $1 \leq \underline{k} \leq m$, $1 \leq \underline{\ell} \leq J_{\underline{k}}(s^*)$, *such that*

$$\varepsilon^{\tilde{\boldsymbol{\theta}}}_{\underline{i},\underline{j}-1}(S_{\underline{i},\underline{j}-1}(\omega)) < \varepsilon^{\tilde{\boldsymbol{\theta}}}_{\underline{k},\underline{\ell}-1}(S_{\underline{k},\underline{\ell}}(\omega))$$

but

$$\varepsilon^{\boldsymbol{\theta}}_{\underline{i},\underline{j}-1}(S_{\underline{i},\underline{j}-1}(\omega)) \geq \varepsilon^{\boldsymbol{\theta}}_{\underline{k},\underline{\ell}-1}(S_{\underline{k},\underline{\ell}}(\omega)).$$

Then we have

$$(\mathbb{P}^{\lambda_0,\boldsymbol{\theta}_0})^m \left(\ell_{P,m}\left(s^*|\boldsymbol{\theta},\widehat{\Lambda}_m,\mathbf{D}_m(s)\right) > \ell_{P,m}\left(s^*|\tilde{\boldsymbol{\theta}},\widehat{\Lambda}_m,\mathbf{D}_m(s)\right)\right) \geq (\mathbb{P}^{\lambda_0,\boldsymbol{\theta}_0})^m\left(B_{m,\boldsymbol{\theta},\tilde{\boldsymbol{\theta}}}\right).$$

From Theorem 2.10 one can easily derive a criterion for inconsistency of the maximizer of the log-likelihood function denoted by $\hat{\boldsymbol{\theta}}_m$. We state the result as a corollary whose proof is omitted.

Corollary 2.11 *Denote by* $B(\boldsymbol{\theta}_0,\epsilon)$ *an* ϵ*-ball around* $\boldsymbol{\theta}_0$ *and assume that* $\boldsymbol{\theta}$ *is such that for some* $m' \in \mathbb{N}$ *we have for all* $m \geq m'$ *that* $(\mathbb{P}^{\lambda_0,\boldsymbol{\theta}_0})^m\left(B_{m,\boldsymbol{\theta},\tilde{\boldsymbol{\theta}}}\right) \geq c$, $c > 0$, $\forall \tilde{\boldsymbol{\theta}} \in B(\boldsymbol{\theta}_0,\epsilon)$. *Then*

$$\hat{\boldsymbol{\theta}}_m \overset{\mathbb{P}}{\nrightarrow} \boldsymbol{\theta}_0, \quad \textit{as } m \to \infty.$$

Before giving examples to which Theorem 2.10 and Corollary 2.11 apply we present the following example that makes Theorem 2.10 and Corollary 2.11 plausible in a simplified setting.

Example 2.12 Consider the model of Example 2.2 (the same reasoning applies to the model of Example 2.3), take $m = 2$, let both samples be Type-II censored and take *arbitrary* event times $s_{1,1}, \ldots, s_{1,n_1}$ and $s_{2,1}, \ldots, s_{2,n_2}$. Then with $s = \max\{s_{1,n_1}, s_{2,n_2}\}$ the function $\mathcal{S}_2^0(s, \cdot)$ (see Equation (4)) equals:

$$\mathbb{1}_{(0,s_{1,1}]}(\cdot) + \mathbb{1}_{(s_{1,1},s_{1,2}]}(\cdot) + \ldots + \mathbb{1}_{(s_{1,n_1-1},s_{1,n_1}]}(\cdot) + \ldots + \mathbb{1}_{(s_{2,n_2-1},s_{2,n_2}]}(\cdot),$$

whereas the function $\mathcal{S}_2^1(s, \cdot)$ equals

$$\mathbb{1}_{(0,s_{1,1}]}(\cdot) + \mathbb{1}_{(0,s_{1,2}]}(\cdot) + \ldots + \mathbb{1}_{(0,s_{1,n_1}]}(\cdot) + \ldots + \mathbb{1}_{(0,s_{2,n_2}]}(\cdot).$$

Clearly, plugging in an arbitrary $t \geq 0$ into $\mathcal{S}_2^0$ there are at most two intervals that contain t. This is because the intervals coming from the first and the second sample, respectively, do not overlap. On the other hand, looking at the function $\mathcal{S}_2^1$ we see that for small values of t there are $n_1 + n_2$ overlapping intervals. As t increases the number of overlapping intervals decreases from $n_1 + n_2$ to $n_1 + n_2 - 1$ to $n_1 + n_2 - 2$ and so on. Apparently, this behaviour does not depend on how the samples $s_{1,1}, \ldots, s_{1,n_1}$ and $s_{2,1}, \ldots, s_{2,n_2}$ were generated. Of course, such a behaviour rules out the possibility to get a consistent estimator based on the profile likelihood function. $\diamond$

We now provide examples to which Theorem 2.10 and Corollary 2.11 can be applied.

Example 2.13 *(ARA_1 or Kijima I with non-random repair)* Consider again the model of Example 2.2 and notice first of all that $\gamma_{i,j-1}^{\theta} \equiv 1$ for every $\theta \in [0, 1]$. Let $0 \leq \theta < \tilde{\theta} \leq 1$. We first discuss part (a) of Theorem 2.10 for this model. For *arbitrary* positive real numbers x, y and z with $y < z$ we have that

$$x - \theta x < z - \theta y \Rightarrow x - \tilde{\theta} x < z - \tilde{\theta} y.$$

Indeed, the functions $f_1, f_2 : [0, 1] \to \mathbb{R}$, defined by $f_1(\bar{\theta}) := x - \bar{\theta} x$ and $f_2(\bar{\theta}) := z - \bar{\theta} y$, respectively, are both monotonically decreasing. By assumption $f_1(\theta) < f_2(\theta)$. Moreover, $f_1(1) = 0$ and $f_2(1) > 0$. Hence, $f_1(\underline{\theta}) < f_2(\underline{\theta})$, $\forall \underline{\theta} > \theta$, because f_1 and f_2 are linear. Hence, for this model $(\mathbb{P}^{\lambda_0,\theta_0})^m(A_{m,\theta,\tilde{\theta}}) = 1$ for every $m \in \mathbb{N}$ whenever $\theta < \tilde{\theta}$ regardless of (λ_0, θ_0) (recall that we consider a simple counting process so that we have $S_{k,\ell} > S_{k,\ell-1}$ with probability one and notice that $S_{k,\ell}$ corresponds to z and $S_{k,\ell-1}$ to y).

We now turn to part (b) of Theorem 2.10. For x, y and z as before and $0 \leq \theta \leq \tilde{\theta} < 1$ the condition

$$x - \tilde{\theta} x < z - \tilde{\theta} y, \quad \text{but} \quad x - \theta x \geq z - \theta y$$

is equivalent to

$$\frac{z - y}{1 - \theta} + y \leq x < \frac{z - y}{1 - \tilde{\theta}} + y. \tag{9}$$

Now, if we consider the data generating process of Example 2.8 with $n_i \geq 2$ and $m \geq 2$, then we have for $\theta < \tilde{\theta}$ from (9)

$$
\begin{aligned}
(\mathbb{P}^{\lambda_0,\theta_0})^m \left(B_{m,\theta,\tilde{\theta}}\right) &\geq (\mathbb{P}^{\lambda_0,\theta_0})^m \left(\frac{S_{2,2}-S_{2,1}}{1-\theta} + S_{2,1} \leq S_{1,1} < \frac{S_{2,2}-S_{2,1}}{1-\tilde{\theta}} + S_{2,1}\right) \\
&= \int_{\mathbb{R}^2} (\mathbb{P}^{\lambda_0,\theta_0})^m \left(\frac{s_{2,2}-s_{2,1}}{1-\theta} + s_{2,1} \leq S_{1,1} < \frac{s_{2,2}-s_{2,1}}{1-\tilde{\theta}} + s_{2,1}\right) dF^{S_{2,2},S_{2,1}}_{\lambda_0,\theta_0}(s_{2,2},s_{2,1}) \\
&= \int_{\mathbb{R}^2} \left[F^{S_{1,1}}_{\lambda_0,\theta_0}\left(\frac{s_{2,2}-s_{2,1}}{1-\tilde{\theta}} + s_{2,1}\right) - F^{S_{1,1}}_{\lambda_0,\theta_0}\left(\frac{s_{2,2}-s_{2,1}}{1-\theta} + s_{2,1}\right)\right] dF^{S_{2,2},S_{2,1}}_{\lambda_0,\theta_0}(s_{2,2},s_{2,1}),
\end{aligned}
\tag{10}
$$

where the first equality follows from the independence of the units, and $F^{S_{2,2},S_{2,1}}_{\lambda_0,\theta_0}$ denotes the joint distribution function of $(S_{2,1}, S_{2,2})$ under $(\mathbb{P}^{\lambda_0,\theta_0})^m$ and similar $F^{S_{1,1}}_{\lambda_0,\theta_0}$ denotes the distribution function of $S_{1,1}$ under $(\mathbb{P}^{\lambda_0,\theta_0})^m$. Clearly, if λ_0 is such that the corresponding cumulative distribution function F_0, which equals here $F^{S_{1,1}}_{\lambda_0,\theta_0}$, is strictly increasing on $\mathbb{R}$, the integrand in (10) is positive whenever $\theta < \tilde{\theta}$. With slightly more effort other cases as, for instance, an F_0 which is constant on some intervals can be discussed. Let us finally consider the condition of Corollary 2.11. Assume that $\theta_0 \in [0,1]$ is not equal to 0. Consider $[\theta_0-\epsilon, \theta_0+\epsilon]$ with $[\theta_0-\epsilon, \theta_0+\epsilon] \subsetneq [0,1]$ and let $0 \leq \theta < \theta_0 - \epsilon$. Then, from Equation (10) we have a lower bound for $(\mathbb{P}^{\lambda_0,\theta_0})^m (B_{m,\theta,\theta_0-\epsilon})$ that does not depend on m, $m \geq 2$. Clearly, this lower bound also holds for $(\mathbb{P}^{\lambda_0,\theta_0})^m \left(B_{m,\theta,\tilde{\theta}}\right)$, $\tilde{\theta} \in [\theta_0-\epsilon, \theta_0+\epsilon]$, as the integrand in Equation (10) is non-decreasing in $\tilde{\theta}$ for θ fixed. $\diamond$

Example 2.14 *(ARA_∞ or Kijima II with non-random repair)* Consider again the model of Example 2.3 and notice that we again have that $\gamma^\theta_{i,j-1} \equiv 1$ for every $\theta \in [0,1]$. Let $0 \leq \theta < \tilde{\theta} \leq 1$. Then for *arbitrary* $s_1 < ... < s_{i-1}$ and $\bar{s}_1 < ... < \bar{s}_k$ (positive) real numbers the condition in part (a) of Theorem 2.10 reads as

$$
\begin{aligned}
&s_{i-1} - \theta \sum_{\ell=1}^{i-1} (1-\theta)^{i-1-\ell} s_\ell < \bar{s}_k - \theta \sum_{\ell=1}^{k-1} (1-\theta)^{k-1-\ell} \bar{s}_\ell \\
\Rightarrow &s_{i-1} - \tilde{\theta} \sum_{\ell=1}^{i-1} (1-\tilde{\theta})^{i-1-\ell} s_\ell < \bar{s}_k - \tilde{\theta} \sum_{\ell=1}^{k-1} (1-\tilde{\theta})^{k-1-\ell} \bar{s}_\ell.
\end{aligned}
\tag{11}
$$

This implication may not hold for every pair $(\theta, \tilde{\theta})$ with $\theta < \tilde{\theta}$ regardless of $s_1 < ... < s_{i-1}$ and $\bar{s}_1 < ... < \bar{s}_k$; see Section 3 for more details on that. However, we see that it holds for $0 \leq \theta < 1$ and $\tilde{\theta} = 1$ so that $(\mathbb{P}^{\lambda_0,\theta_0})^m (A_{m,\theta,1}) = 1$ and Theorem 2.10 now implies

$$
(\mathbb{P}^{\lambda_0,\theta_0})^m (l_{P,m}(s^*|\theta) \geq l_{P,m}(s^*|1)) = 1, \quad 0 \leq \theta < 1.
$$

Furthermore, for $i = 2$ and $k = 2$ we find that Equation (11) boils down to Equation (9) (with $x = s_1$, $y = \bar{s}_1$ and $z = \bar{s}_2$). It therefore follows from Example 2.13 that $(\mathbb{P}^{\lambda_0,\theta_0})^m (B_{m,\theta,1}) > 0$, $0 \leq \theta < 1$. ◇

3. Simulation results

In this section, as mentioned in the introduction, we provide some simulation results illustrating Assumption 2.7 imposed in Theorem 2.10. We have seen in Example 2.8 that under Type-II censoring Assumption 2.7 holds. It is also clear from Example 2.9 that, in general, under Type-I censoring Assumption 2.7 is not fulfilled. Moreover, in Example 2.14 we left some questions open. Here, we will analyse these questions further by Monte Carlo simulations. Furthermore, as mentioned in the introduction and in Remark 2.5 the behaviour of the "discrete log profile likelihood" is studied by Monte Carlo simulations as well. In all the simulations the hazard rate function used in the definition of Model 2.1 was taken to come from a right truncated Weibull distribution with reliability function S_d defined for $d > 0$ by

$$S_d(t) := \frac{e^{-t^2} - e^{-d^2}}{1 - e^{-d^2}} \mathbb{1}_{[0,d)}(t), \quad t > 0.$$

Hence the corresponding hazard rate function λ_d equals

$$\lambda_d(t) = \frac{2te^{-t^2}}{e^{-t^2} - e^{-d^2}} \mathbb{1}_{[0,d)}(t), \quad t > 0.$$

We simulated data following an ARA_1 model (see Example 2.2) and an ARA_∞ model (see Example 2.3), respectively, for various values of (θ, d). In addition we introduced two types of censoring: Type-I and Type-II censoring. The sample size m was always taken to be equal to 100. Under Type-I censoring the data observed were given by $S_{i,1} < \cdots < S_{i,r_i}$ for (random) r_i, $i = 1, \ldots, 100$, and S_{i,r_i+1}, $i = 1, \ldots, 100$, was right censored by $\tau_i = \tau$, $i = 1, \ldots, 100$, for some non-random $\tau > 0$. In case of Type-I censoring Assumption 2.7 is not fulfilled. If the censoring time is random and equal to $S_{i,r}$, $i = 1, \ldots, 100$, for some non-random r the data are Type-II censored (see Example 2.8) and Assumption 2.7 holds. Combining the two models, ARA_1 and ARA_∞, and the two types of censoring one obtains four possible combinations. The results, i.e. ten realizations of the function $\ell_{P,100}$, for Type-II censoring and the following values of $(\theta_0, d, r) \in \{0, 0.5, 1\} \times \{5\} \times \{2, 5\}$ are given in Fig. 1 and 3. Here θ_0 denotes as in Examples 2.13 and 2.14 the true parameter. Type-I censored data were simulated for $(\theta_0, d, \tau) \in \{0.1, 0.5, 1\} \times \{2\} \times \{2.2\}$ and $(\theta_0, d, \tau) \in \{0, 0.5, 1\} \times \{6\} \times \{5.9\}$. The reason for taking $(\theta_0, d, \tau) = (0.1, 2, 2.2)$ instead of $(\theta_0, d, \tau) = (0, 2, 2.2)$ is that θ_0 corresponds to a non-homogeneous Poisson process for which with $d = 2$ the observations are not censored. The results, i.e. again ten realizations of the function $\ell_{P,100}$, are given in Fig. 2 and 4.

Examples 2.8 and 2.13 together imply that conditions (a) and (b) of Theorem 2.10 are met for an ARA_1 model under Type-II censoring with probability one and a positive probability, respectively. It is also clear that Assumption 2.7 is not fulfilled for Type-I censoring. The results for Type-II censoring are given in Fig. 1. Clearly, the ten realizations of the function $\ell_{P,100}$ are decreasing in θ as it was proved in Example 2.13 by verifying condition (a) of Theorem 2.10 for all pairs $(\theta, \tilde{\theta})$ with $\theta < \tilde{\theta}$. We also see that all realizations of $\ell_{P,100}$ in Fig. 1 seem to be strictly decreasing as a function of θ so that the probabilities with which condition (b) of Theorem 2.10 hold might be quite large for the above λ_ds. The picture is slightly different for the combination ARA_1 and Type-I censoring. Of course assumptions (a) and (b) of Theorem 2.10 on the statistical model are not affected by considering Type-I censored data instead of Type-II censored data, but assumption (Assumption 2.7) on the sampling procedure is not met. Comparing the left- and right-hand side of Fig. 2 we see that on the left-hand side not all realizations of $\ell_{P,100}$ are monotonically decreasing whereas on the right-hand side all realizations seem to lead to a monotonically decreasing $\ell_{P,100}$. This is probably a result of the fact that due to the increased observation period ($\tau = 2.2$ on the left-hand side and $\tau = 5.9$ on the right-hand side) the part of $\ell_{P,100}$ stemming from Type-I censored observations (exactly these observations seem to prevent $\ell_{P,100}$ from being monotonically decreasing in case of Type-I censoring) gets outweighed by the number of observed failure times.

Now we briefly discuss the simulation results for the ARA_∞ model and the two types of censoring. In Example 2.14 we did not prove that conditions (a) and (b) of Theorem 2.10 hold for all pairs $(\theta, \tilde{\theta})$ with $\theta < \tilde{\theta}$. The simulation results for an ARA_∞ model under Type-II censoring (Assumption 2.7 is then fulfilled) suggest that condition (a) (and maybe even (b)) of Theorem 2.10 may also hold with probability one for all pairs $(\theta, \tilde{\theta})$ with $\theta < \tilde{\theta}$ as for the ARA_1 model. However, this is not the case as can be seen from Fig. 5 where we plotted the difference between the right- and left-hand side in the first displayed equation of Example 2.14 for $i = 9$, $s_1 = 5.0, s_2 = 7.1, s_3 = 12.2, s_4 = 16.3, s_5 = 17.0, s_6 = 20.5, s_7 = 22.5, s_8 = 27$, and $k = 9$, $\bar{s}_1 = 3.4, \bar{s}_2 = 7.9, \bar{s}_3 = 10, \bar{s}_4 = 14.0, \bar{s}_5 = 19.6, \bar{s}_6 = 22.6, \bar{s}_7 = 23.3, \bar{s}_8 = 26.0, \bar{s}_9 = 27.1$ as a function of θ. However, the realizations shown in Fig. 3 suggest that the probabilities of the events in condition (a) and (b) of Theorem 2.10, respectively, are relatively large. In case of an ARA_∞ model and Type-I censoring Assumption 2.7 is not met and condition (a) of Theorem 2.10 does not hold for all pairs $(\theta, \tilde{\theta})$ with $\theta < \tilde{\theta}$. Nevertheless, the simulation results shown in Fig. 4 suggest that the profile likelihood estimator remains inconsistent.

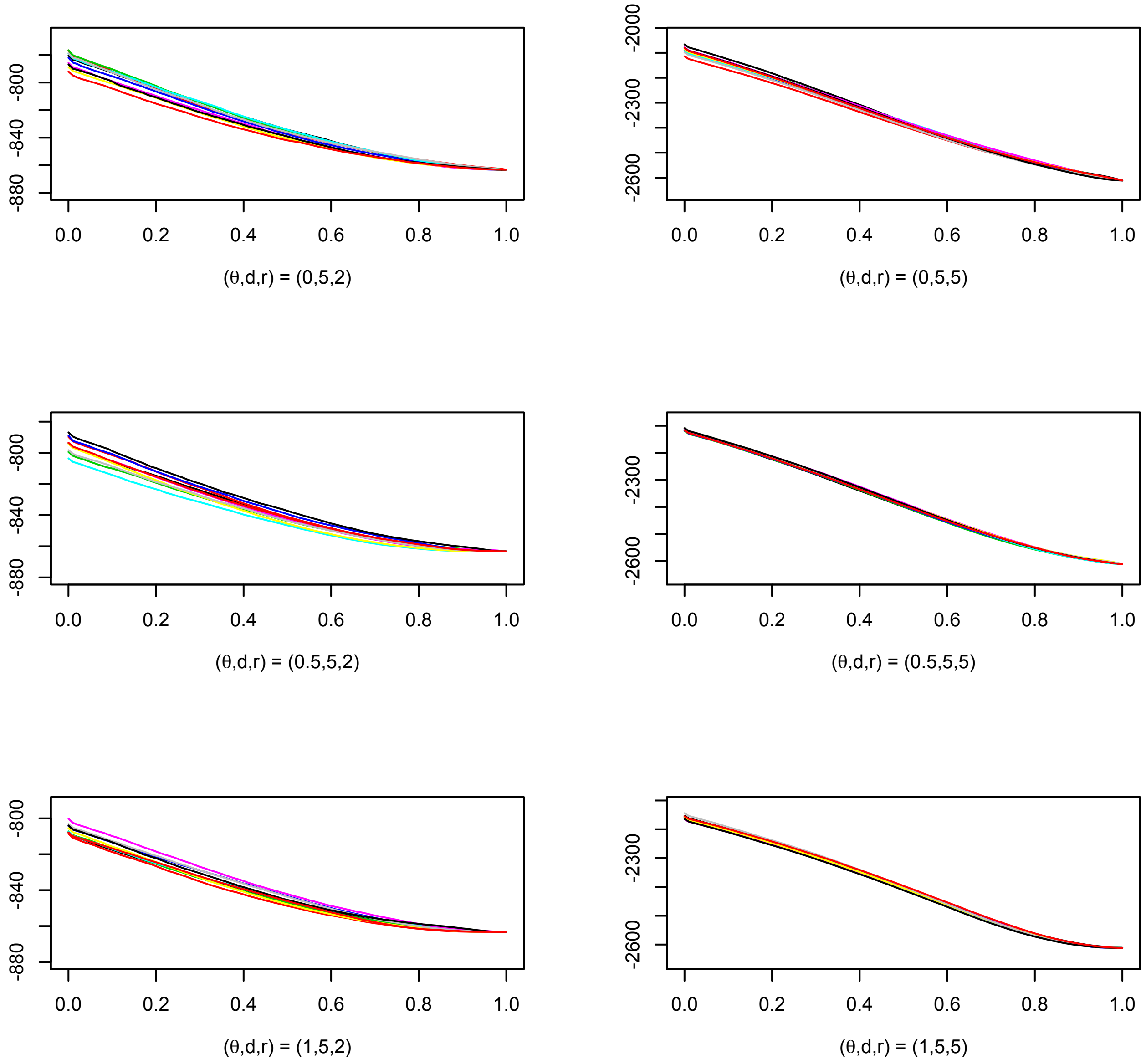

Figure 1: Each of the 6 figures contains ten graphs of $\theta \mapsto \ell_{m,P}(s|\theta, \widehat{\Lambda}_m, \mathbf{D}_m(s))$ for $\theta \in [0,1]$ and $m = 100$ obtained from simulated data for which $\lambda = \lambda_d$, the effective age follows the ARA_1 assumption of Example 2.2, and for $\tau_i = S_{i,r}$ for $1 \leq i \leq m$ (Type II censoring, see Example 2.8). Here θ denotes the true parameter.

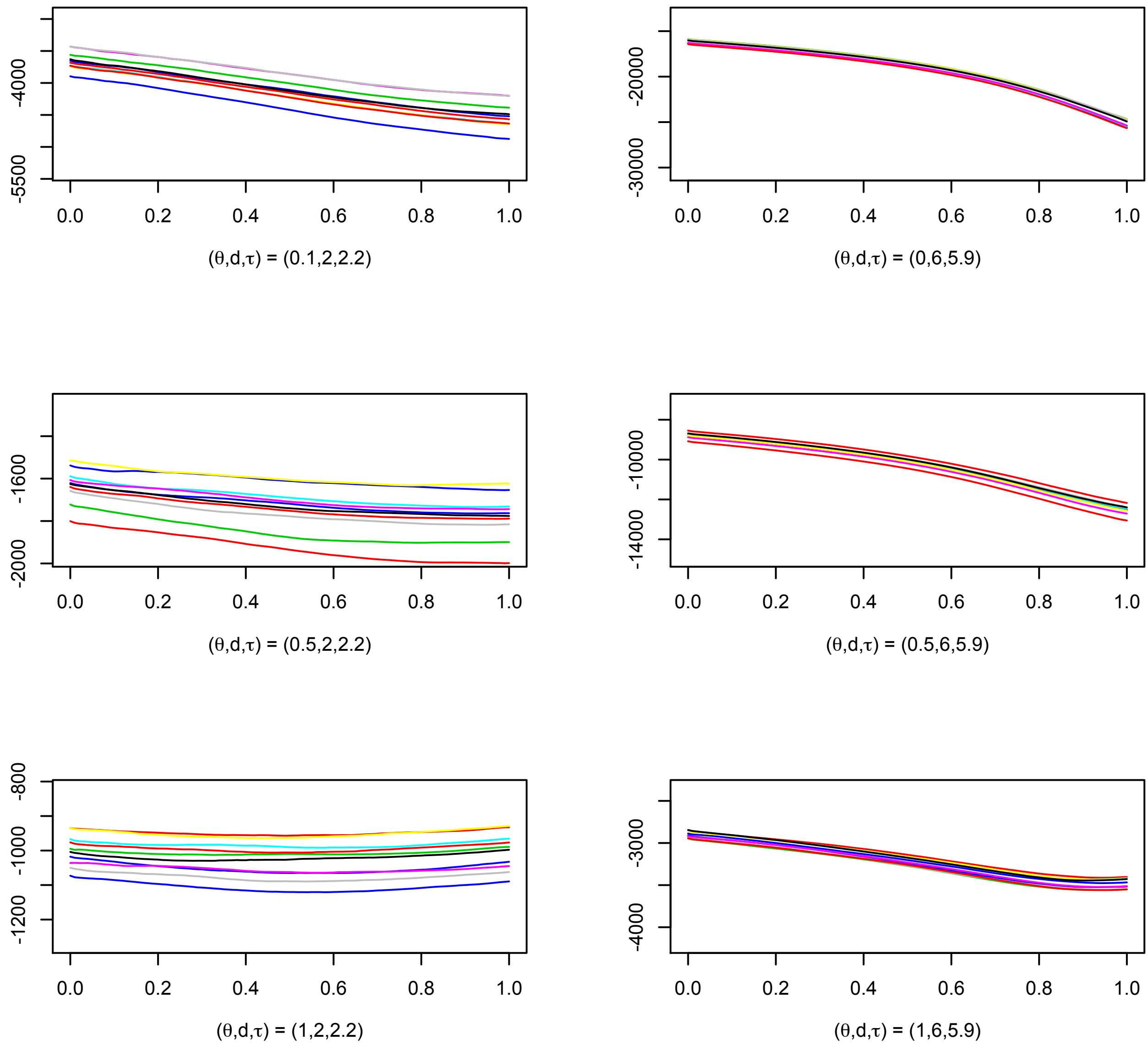

Figure 2: Each of the 6 figures contains ten graphs of $\theta \mapsto \ell_{m,P}(s|\theta, \widehat{\Lambda}_m, \mathbf{D}_m(s))$ for $\theta \in [0,1]$ and $m = 100$ obtained from simulated data for which $\lambda = \lambda_d$, the effective age follows the ARA_1 assumption of Example 2.2, and $s = \tau_i = \tau$ for all $1 \le i \le m$ (Type I censoring). Here θ denotes the true parameter.

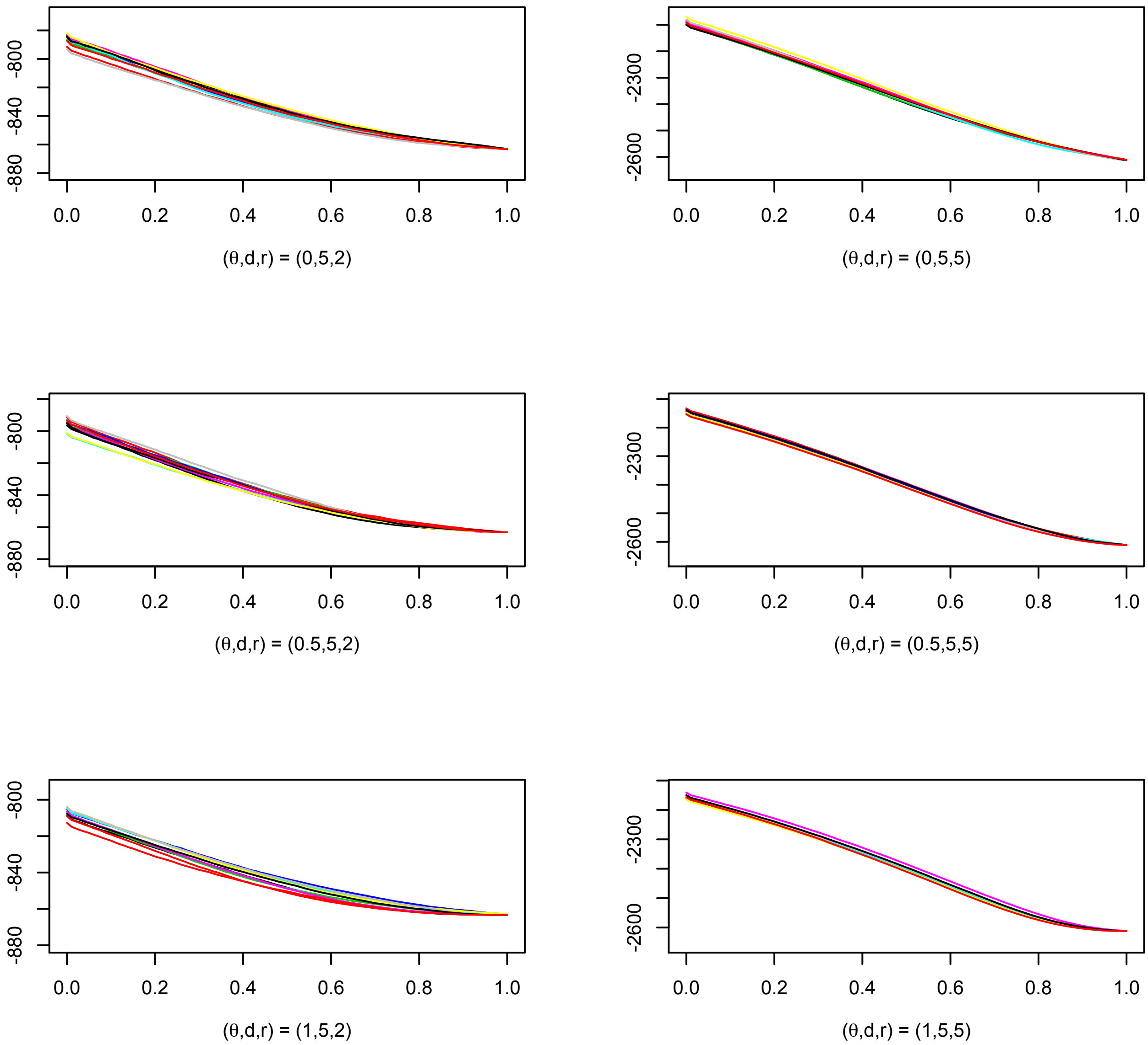

Figure 3: Each of the 6 figures contains ten graphs of $\theta \mapsto \ell_{m,P}(s|\theta, \widehat{\Lambda}_m, \mathbf{D}_m(s))$ for $\theta \in [0,1]$ and $m = 100$ obtained from simulated data for which $\lambda = \lambda_d$, the effective age follows the ARA_∞ assumption of Example 2.3, and for $\tau_i = S_{i,r}$ for $1 \leq i \leq m$ (Type II censoring, see Example 2.8). Here θ denotes the true parameter.

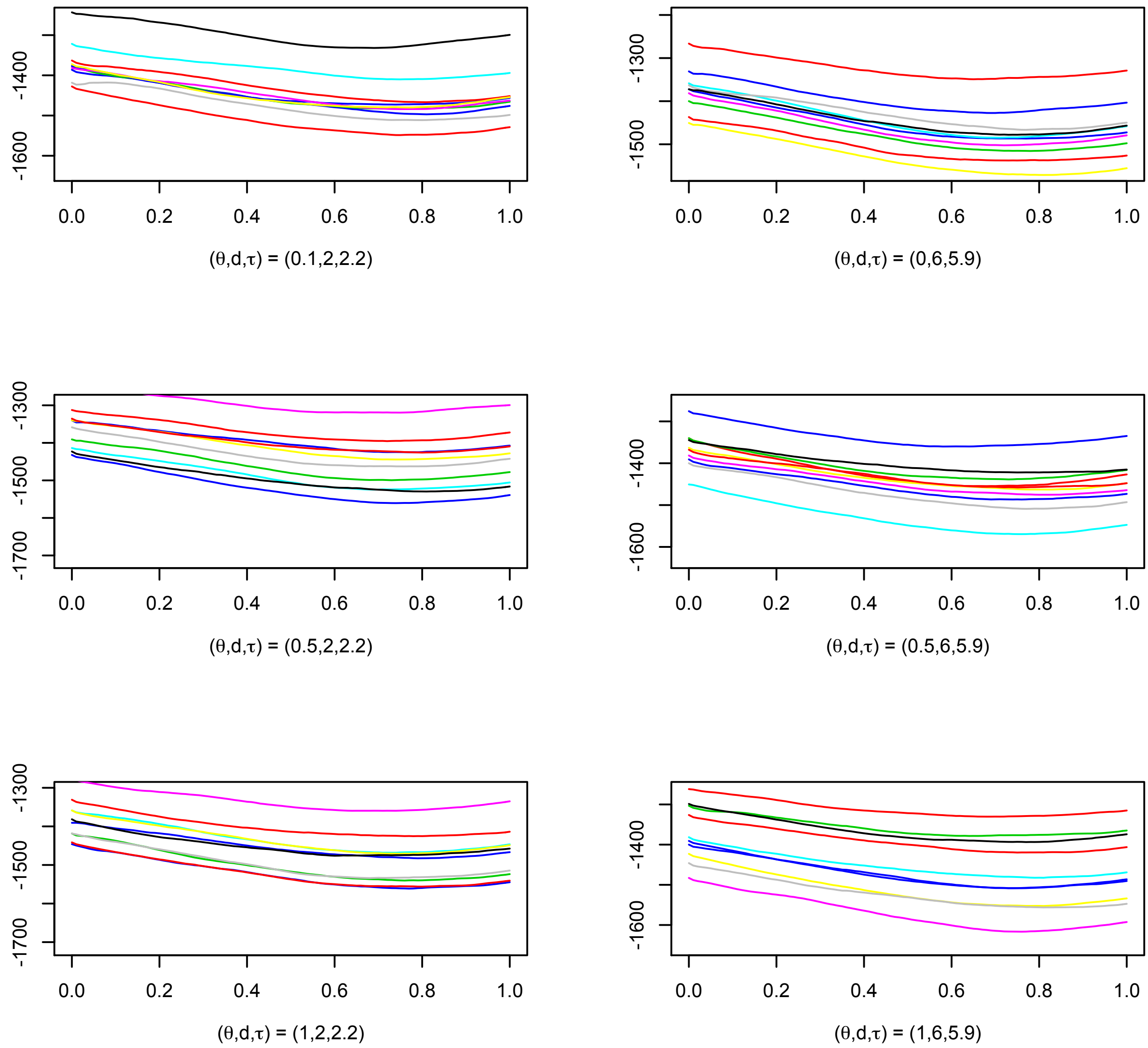

Figure 4: Each of the 6 figures contains ten graphs of $\theta \mapsto \ell_{m,P}(s|\theta, \widehat{\Lambda}_m, \mathbf{D}_m(s))$ for $\theta \in [0,1]$ and $m = 100$ obtained from simulated data for which $\lambda = \lambda_d$, the effective age follows the ARA_∞ assumption of Example 2.3, and $s = \tau_i = \tau$ for all $1 \leq i \leq m$ (Type I censoring). Here θ denotes the true parameter.

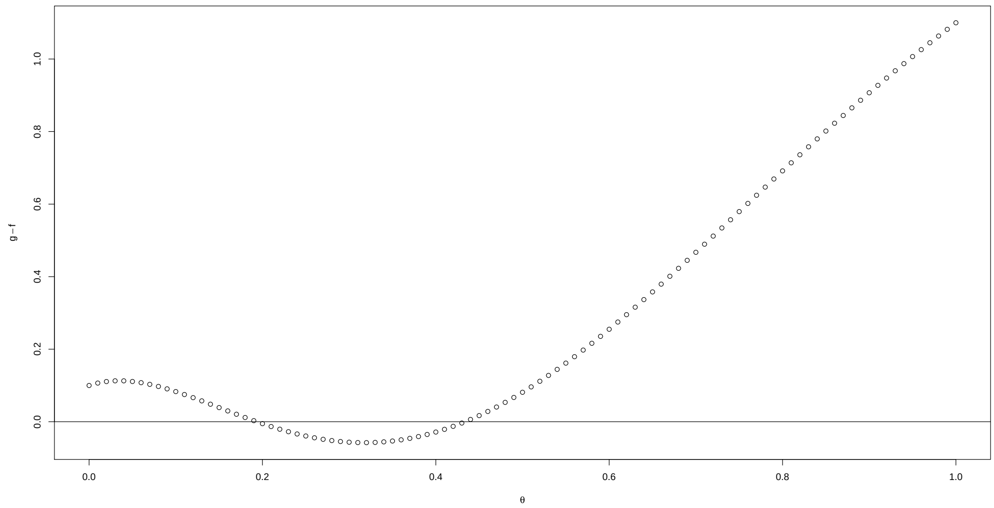

Figure 5: Difference between $g(\theta) := \bar{s}_k - \theta \sum_{\ell=1}^{k-1}(1-\theta)^{k-1-\ell}\bar{s}_\ell$ and $f(\theta) := s_{i-1} - \theta \sum_{\ell=1}^{i-1}(1-\theta)^{i-1-\ell}s_\ell$, where $i = 9$, $s_1 = 5.0, s_2 = 7.1, s_3 = 12.2, s_4 = 16.3, s_5 = 17.0, s_6 = 20.5, s_7 = 22.5, s_8 = 27$, $k = 9$ and $\bar{s}_1 = 3.4, \bar{s}_2 = 7.9, \bar{s}_3 = 10, \bar{s}_4 = 14.0, \bar{s}_5 = 19.6, \bar{s}_6 = 22.6, \bar{s}_7 = 23.3, \bar{s}_8 = 26.0, \bar{s}_9 = 27.1$, see Example 2.14.

We finish this section by the comparison of the "discrete" and the "continuous log profile likelihood" as explained in Remark 2.5. For the above given right-truncated Weibull distribution and various values of (θ_0, d, r) and a sample size of $m = 100$ we simulated five samples from an ARA_1 model. For each simulated sample the two versions (continuous and discrete) of the profile likelihood function were calculated for $\theta \in [0, 1]$. The continuous profile likelihood curves are plotted with solid black lines whereas the discrete profile likelihood curves are plotted with dotted red lines. The results are given in Fig. 6. The obtained results are consistent with Theorem 2.10 where we use the continuous version of the profile log–likelihood function (solid lines). In addition we observe that using a version of the profile log–likelihood accounting the fact that the NPMLE of the baseline cumulative hazard is discrete (Appendix B) still leads to monotone profile log–likelihood functions. Finally notice that we see in Fig. 6 that the differences of the realizations of the discrete and continuous version are for all $\theta \in [0, 1]$ close to $\log \exp\left(-\sum_{i=1}^m N_i(s)\right) = -m \cdot r$. This term is just a product integral (the second double product in (44)). That the differences of the realizations are almost constant as a function of θ is due to the fact that m is large compared to r. For m small and r large the differences are far from being constant as a function of θ.

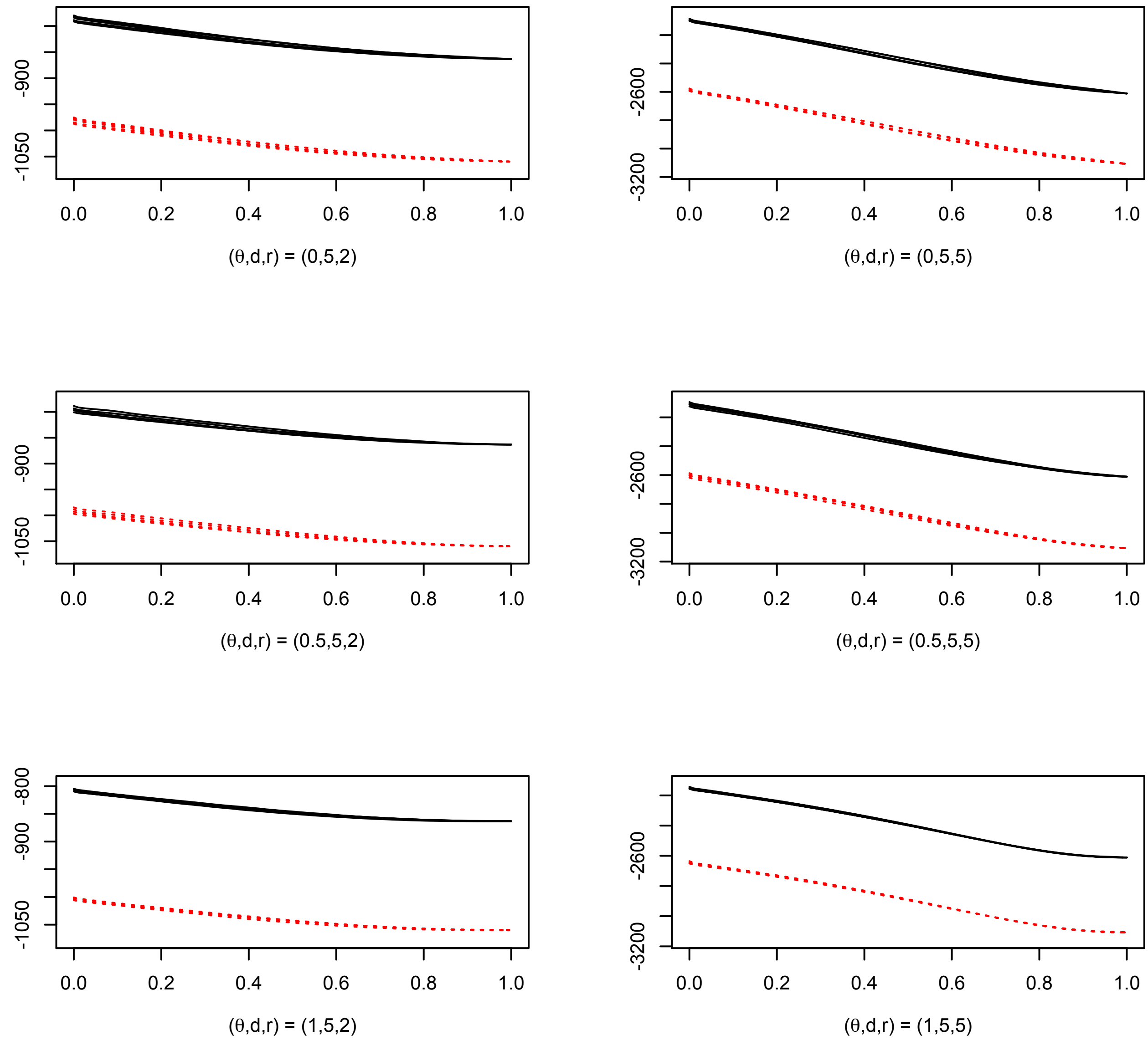

Figure 6: Each of the 6 figures contains five graphs of $\theta \mapsto \ell_{m,P}(s|\theta, \widehat{\Lambda}_m, \mathbf{D}_m(s))$ (solid lines) for $\theta \in [0,1]$ and $m = 100$ obtained from simulated data for which $\lambda = \lambda_d$, the effective age follows the ARA$_1$ assumption of Example 2.2, and for $\tau_i = S_{i,r}$ for $1 \leq i \leq m$ (Type II censoring, see Example 2.8). The dotted red lines are the realizations of the discrete profile likelihood curves. Here θ denotes the true parameter.

4. Conclusion

We have seen in Example 2.13 a model for recurrent events for which the profile likelihood approach with a right-continuous step function as an estimator for Λ fails in all respects as the profile likelihood function in that example is monotonically decreasing with probability one regardless of the sample size and the true underlying probability measure. As mentioned in the introduction a similar behaviour may occur in a single event model with covariates. It is also clear from the two sentences preceding Remark 2.5 that this behaviour cannot only be attributed to the fact that the argument of the exponential in the profile likelihood function is free of the virtual age parameter $\boldsymbol{\theta}$. This is simply because the same goes for the profile likelihood function in Cox's proportional hazards model for which the profile likelihood approach works. Moreover, Appendix C rules out the possibility that the failure of the profile likelihood approach results from an identifiability issue. To conclude: It seems to be a fine line that divides semi-parametric models for which the profile likelihood approach works from those semi-parametric models for which it fails. An exact description, i.e. an 'if and only if' statement, might be difficult or even impossible to obtain. Nevertheless, a few general features are worth summarizing. When the NPMLE $\hat{\Lambda}(\cdot;\boldsymbol{\theta})$ is a right-continuous step function the profile likelihood method fails for the semi-parametric accelerated failure time model (with a single event). This is because for this model $\boldsymbol{\theta}$ influences only the locations of the jumps of $\hat{\Lambda}(\cdot;\boldsymbol{\theta})$ but not the jump heights. This is in contrast to Cox's proportional hazards model where the jump heights of $\hat{\Lambda}(\cdot;\boldsymbol{\theta})$ do depend on $\boldsymbol{\theta}$ whereas the locations of the jumps are the same for all $\boldsymbol{\theta}$. Here we have proved a result that gives conditions under which profile likelihood inference does not work for the effective age parameter in a semi-parametric recurrent event model. In the examples presented the jump heights of the NPMLE $\hat{\Lambda}(\cdot;\theta)$ depend on θ. Nevertheless, the profile likelihood method fails. This might be a result of the fact that θ does not only affect the jump heights, but also the locations. One may wonder whether the locations being independent of $\boldsymbol{\theta}$ (together with some regularity conditions) is a sufficient condition for the profile likelihood method to work.

A. Proofs and auxiliary lemmas

Before we give the proof of Theorem 2.10 we state two lemmas. The first lemma will be used in the proof of part (a) and part (b) whereas the second lemma will only be used in the proof of part (b).

Lemma A.1 *Let I_1 and I_2 be two subsets of a finite set J with $I_1 \neq J$ and $|I_1| = |I_2|$, where for any set K we denote by $|K|$ its cardinality. Moreover, assume that for at least* *one element i_1 of I_1 we have $i_1 \notin I_2$. Then there is at least one element $i_2 \in J$ such that $i_2 \in I_2$, but $i_2 \notin I_1$.*

Proof The claim is immediate from the facts that $I_1 \subsetneq J$ and $I_2 \subsetneq J$ have the same cardinality and that $i_1 \in I_1$, but $i_1 \notin I_2$. $\square$

The following lemma is obvious and its proof is therefore omitted.

Lemma A.2 *Let x_i, y_i, $1 \leq i \leq I$, and $\tilde{x}_i$, $\tilde{y}_i$, $1 \leq i \leq I$, be (non-negative) real numbers.*

(a) Define the functions G and $\tilde{G}$ both with domain $\{1, \ldots, I\}$ by

$$G(j) := \sum_{i=1}^{I} \mathbb{1}_{(x_i,\infty)}(y_j)$$

and

$$\tilde{G}(j) := \sum_{i=1}^{I} \mathbb{1}_{(\tilde{x}_i,\infty)}(\tilde{y}_j).$$

Then:

(i) If for a given $j \in \{1, \ldots, I\}$ we have that for all $i \in \{1, \ldots, I\}$ the relation $x_i < y_j$ implies the following relation $\tilde{x}_i < \tilde{y}_j$, then

$$G(j) \leq \tilde{G}(j).$$

(ii) If additionally to the assumption in part (i) we have that there is an $\underline{i} \in \{1, \ldots, I\}$ such that $\tilde{x}_{\underline{i}} < \tilde{y}_j$ but $x_{\underline{i}} \geq y_j$ then

$$G(j) < \tilde{G}(j).$$

(b) Denote by $y_{(i)}$, $1 \leq i \leq I$, and $\tilde{y}_{(i)}$, $1 \leq i \leq I$, the increasingly ordered values of the y_i, $1 \leq i \leq I$, and of the $\tilde{y}_i$, $1 \leq i \leq I$, respectively. Define the functions G^{ord} and $\tilde{G}^{ord}$ both with domain $\{1, \ldots, I\}$ by

$$G^{ord}(j) := \sum_{i=1}^{I} \mathbb{1}_{(x_i,\infty)}(y_{(j)})$$

and

$$\tilde{G}^{ord}(j) := \sum_{i=1}^{I} \mathbb{1}_{(\tilde{x}_i,\infty)}(\tilde{y}_{(j)}).$$

Then we have for $j \leq k$

$$G^{ord}(j) \leq G^{ord}(k) \text{ and } \tilde{G}^{ord}(j) \leq \tilde{G}^{ord}(k).$$

Proof of Theorem 2.10 Throughout the proof whenever appropriate we suppress the dependence on ω, otherwise it is made explicit. We first prove part (a). Notice first that under Assumption 2.7 we have for every $\breve{\boldsymbol{\theta}}$ that the log profile likelihood $\ell_{P,m}(s^*|\breve{\boldsymbol{\theta}}, \widehat{\Lambda}_m, \mathbf{D}_m(s))$ equals

$$-\sum_{k=1}^{m}\sum_{\ell=1}^{J_k(s^*)} \log\left(\mathcal{S}_m^{\breve{\boldsymbol{\theta}}}(s^*, \varepsilon_k^{\breve{\boldsymbol{\theta}}}(S_{k\ell}))\right) = -\sum_{k=1}^{m}\sum_{\ell=1}^{J_k(s^*)} \log\left(\mathcal{S}_m^{\breve{\boldsymbol{\theta}}}(s^*, \varepsilon_{k,\ell-1}^{\breve{\boldsymbol{\theta}}}(S_{k,\ell}))\right). \quad (12)$$

Now, let ω be arbitrary. Put $U(\omega) := \sum_{k=1}^{m} J_k(s^*, \omega)$. Denoting the increasingly ordered values of $\varepsilon_{k,\ell-1}^{\breve{\boldsymbol{\theta}}}(S_{k,\ell}(\omega))$, $1 \le k \le m, 1 \le \ell \le J_k(s^*, \omega)$, by $\left(\varepsilon_{k,\ell-1}^{\breve{\boldsymbol{\theta}}}(S_{k,\ell}(\omega))\right)_{(p)}$, $p = 1, \ldots, U(\omega)$, the right-hand side of Equation (12) can be rewritten as

$$-\sum_{p=1}^{U(\omega)} \log\left[\mathcal{S}_m^{\breve{\boldsymbol{\theta}}}\left(s^*, (\varepsilon_{k,\ell-1}^{\breve{\boldsymbol{\theta}}}(S_{k,\ell}(\omega)))_{(p)}\right)\right]. \quad (13)$$

From Equation (4) we now obtain that under Assumption 2.7 the quantity in (13) equals

$$\begin{aligned}-\sum_{p=1}^{U(\omega)} \log\Big[\sum_{i=1}^{m}\sum_{j=1}^{J_i(s^*,\omega)} &\gamma_{i,j-1}^{\breve{\boldsymbol{\theta}}}((\varepsilon_{k,\ell-1}^{\breve{\boldsymbol{\theta}}}(S_{k,\ell}(\omega)))_{(p)}) \\ &\times \mathbb{1}_{(\varepsilon_{i,j-1}^{\breve{\boldsymbol{\theta}}}(S_{i,j-1}(\omega)),\varepsilon_{i,j-1}^{\breve{\boldsymbol{\theta}}}(S_{i,j}(\omega))]}((\varepsilon_{k,\ell-1}^{\breve{\boldsymbol{\theta}}}(S_{k,\ell}(\omega)))_{(p)})\Big]. \quad (14)\end{aligned}$$

We see that under the assumption $(\mathbb{P}^{\lambda_0,\boldsymbol{\theta}_0})^m(\gamma_{i,j-1}^{\boldsymbol{\theta}}(t) \le c, \gamma_{i,j-1}^{\tilde{\boldsymbol{\theta}}}(t) \ge c, 1 \le i \le m, 1 \le j \le J_i(s^*), 0 \le t < \infty) = 1$ our claim

$$(\mathbb{P}^{\lambda_0,\boldsymbol{\theta}_0})^m\left(\ell_{P,m}\left(s^*|\boldsymbol{\theta}, \widehat{\Lambda}_m, \mathbf{D}_m(s)\right) \ge \ell_{P,m}\left(s^*|\tilde{\boldsymbol{\theta}}, \widehat{\Lambda}_m, \mathbf{D}_m(s)\right)\right) \ge (\mathbb{P}^{\lambda_0,\boldsymbol{\theta}_0})^m\left(A_{m,\boldsymbol{\theta},\tilde{\boldsymbol{\theta}}}\right)$$

would follow if

$$\begin{aligned}&-\sum_{p=1}^{U(\omega)} \log\Big[\sum_{i=1}^{m}\sum_{j=1}^{J_i(s^*,\omega)} \mathbb{1}_{(\varepsilon_{i,j-1}^{\boldsymbol{\theta}}(S_{i,j-1}(\omega)),\varepsilon_{i,j-1}^{\boldsymbol{\theta}}(S_{i,j}(\omega))]}((\varepsilon_{k,\ell-1}^{\boldsymbol{\theta}}(S_{k,\ell}(\omega)))_{(p)})\Big] \\ &\ge -\sum_{p=1}^{U(\omega)} \log\Big[\sum_{i=1}^{m}\sum_{j=1}^{J_i(s^*,\omega)} \mathbb{1}_{(\varepsilon_{i,j-1}^{\tilde{\boldsymbol{\theta}}}(S_{i,j-1}(\omega)),\varepsilon_{i,j-1}^{\tilde{\boldsymbol{\theta}}}(S_{i,j}(\omega))]}((\varepsilon_{k,\ell-1}^{\tilde{\boldsymbol{\theta}}}(S_{k,\ell}(\omega)))_{(p)})\Big], \quad (15)\end{aligned}$$

were true for (almost) all $\omega \in A_{m,\boldsymbol{\theta},\tilde{\boldsymbol{\theta}}}$, because we have

$$\begin{aligned}&-\sum_{p=1}^{U(\omega)} \log\Big[\sum_{i=1}^{m}\sum_{j=1}^{J_i(s^*,\omega)} \gamma_{i,j-1}^{\boldsymbol{\theta}}((\varepsilon_{k,\ell-1}^{\boldsymbol{\theta}}(S_{k,\ell}(\omega)))_{(p)}) \\ &\qquad\qquad \times \mathbb{1}_{(\varepsilon_{i,j-1}^{\boldsymbol{\theta}}(S_{i,j-1}(\omega)),\varepsilon_{i,j-1}^{\boldsymbol{\theta}}(S_{i,j}(\omega))]}((\varepsilon_{k,\ell-1}^{\boldsymbol{\theta}}(S_{k,\ell}(\omega)))_{(p)})\Big] \\ &\ge -\sum_{p=1}^{U(\omega)} \log\Big[c\sum_{i=1}^{m}\sum_{j=1}^{J_i(s^*,\omega)} \mathbb{1}_{(\varepsilon_{i,j-1}^{\boldsymbol{\theta}}(S_{i,j-1}(\omega)),\varepsilon_{i,j-1}^{\boldsymbol{\theta}}(S_{i,j}(\omega))]}((\varepsilon_{k,\ell-1}^{\boldsymbol{\theta}}(S_{k,\ell}(\omega)))_{(p)})\Big] \quad (16)\end{aligned}$$

and

$$
\begin{aligned}
&-\sum_{p=1}^{U(\omega)} \log \Big[c \sum_{i=1}^{m} \sum_{j=1}^{J_i(s^*,\omega)} \mathbb{1}_{(\varepsilon^{\tilde{\boldsymbol{\theta}}}_{i,j-1}(S_{i,j-1}(\omega)),\varepsilon^{\tilde{\boldsymbol{\theta}}}_{i,j-1}(S_{i,j}(\omega))]}((\varepsilon^{\tilde{\boldsymbol{\theta}}}_{k,\ell-1}(S_{k,\ell}(\omega)))_{(p)}) \Big] \\
&\geq -\sum_{p=1}^{U(\omega)} \log \Big[\sum_{i=1}^{m} \sum_{j=1}^{J_i(s^*,\omega)} \gamma^{\tilde{\boldsymbol{\theta}}}_{i,j-1}((\varepsilon^{\tilde{\boldsymbol{\theta}}}_{k,\ell-1}(S_{k,\ell}(\omega)))_{(p)}) \\
&\qquad\qquad \times \mathbb{1}_{(\varepsilon^{\tilde{\boldsymbol{\theta}}}_{i,j-1}(S_{i,j-1}(\omega)),\varepsilon^{\tilde{\boldsymbol{\theta}}}_{i,j-1}(S_{i,j}(\omega))]}((\varepsilon^{\tilde{\boldsymbol{\theta}}}_{k,\ell-1}(S_{k,\ell}(\omega)))_{(p)}) \Big]. \qquad (17)
\end{aligned}
$$

Now notice that for arbitrary $\breve{\boldsymbol{\theta}}$ we can rewrite

$$
-\sum_{p=1}^{U(\omega)} \log \Big[\sum_{i=1}^{m} \sum_{j=1}^{J_i(s^*,\omega)} \mathbb{1}_{(\varepsilon^{\breve{\boldsymbol{\theta}}}_{i,j-1}(S_{i,j-1}(\omega)),\varepsilon^{\breve{\boldsymbol{\theta}}}_{i,j-1}(S_{i,j}(\omega))]}((\varepsilon^{\breve{\boldsymbol{\theta}}}_{k,\ell-1}(S_{k,\ell}(\omega)))_{(p)}) \Big]
$$

as

$$
-\sum_{p=1}^{U(\omega)} \left(\log \left[\sum_{i=1}^{m} \sum_{j=1}^{J_i(s^*,\omega)} \left((\varepsilon^{\breve{\boldsymbol{\theta}}}_{k,\ell-1}(S_{k,\ell}(\omega)))_{(p)} \right) - (p-1) \right] \right), \qquad (18)
$$

because the condition

$$
(\varepsilon^{\breve{\boldsymbol{\theta}}}_{k,\ell-1}(S_{k,\ell}(\omega)))_{(p)} \leq \varepsilon^{\breve{\boldsymbol{\theta}}}_{i,j-1}(S_{i,j}(\omega))
$$

fails for exactly $(p-1)$ pairs (i,j), $1 \leq i \leq m, 1 \leq j \leq J_i(s^*)$. As the representation (18) holds for every element of Θ, we see (15) would follow if for every p and (almost) every $\omega \in A_{m,\boldsymbol{\theta},\tilde{\boldsymbol{\theta}}}$ it were true that for every pair (i,j)

$$
\varepsilon^{\boldsymbol{\theta}}_{i,j-1}(S_{i,j-1}(\omega)) < \varepsilon^{\boldsymbol{\theta}}_{g^{\boldsymbol{\theta}}_{(p)},h^{\boldsymbol{\theta}}_{(p)}-1}(S_{g^{\boldsymbol{\theta}}_{(p)},h^{\boldsymbol{\theta}}_{(p)}}(\omega)) \Rightarrow \varepsilon^{\tilde{\boldsymbol{\theta}}}_{i,j-1}(S_{i,j-1}(\omega)) < \varepsilon^{\tilde{\boldsymbol{\theta}}}_{k^{\tilde{\boldsymbol{\theta}}}_{(p)},\ell^{\tilde{\boldsymbol{\theta}}}_{(p)}-1}(S_{k^{\tilde{\boldsymbol{\theta}}}_{(p)},\ell^{\tilde{\boldsymbol{\theta}}}_{(p)}}(\omega)). \qquad (19)
$$

Here $\varepsilon^{\boldsymbol{\theta}}_{g^{\boldsymbol{\theta}}_{(p)},h^{\boldsymbol{\theta}}_{(p)}-1}(S_{g^{\boldsymbol{\theta}}_{(p)},h^{\boldsymbol{\theta}}_{(p)}}(\omega))$ denotes the p smallest value among the $\varepsilon^{\boldsymbol{\theta}}_{g,h-1}(S_{g,h}(\omega))$, $1 \leq g \leq m$, $1 \leq h \leq J_g(s^*)$, and $\varepsilon^{\tilde{\boldsymbol{\theta}}}_{k^{\tilde{\boldsymbol{\theta}}}_{(p)},\ell^{\tilde{\boldsymbol{\theta}}}_{(p)}-1}(S_{k^{\tilde{\boldsymbol{\theta}}}_{(p)},\ell^{\tilde{\boldsymbol{\theta}}}_{(p)}}(\omega))$ denotes the p smallest value among the $\varepsilon^{\tilde{\boldsymbol{\theta}}}_{k,\ell-1}(S_{k,\ell}(\omega))$, $1 \leq k \leq m$, $1 \leq \ell \leq J_k(s^*)$. Consequently, $(g^{\boldsymbol{\theta}}_{(p)}, h^{\boldsymbol{\theta}}_{(p)})$ denotes the index pair of the p smallest observation among the $\varepsilon^{\boldsymbol{\theta}}_{g,h-1}(S_{g,h}(\omega))$, $1 \leq g \leq m$, $1 \leq h \leq J_g(s^*)$ and similar for the pair $(k^{\tilde{\boldsymbol{\theta}}}_{(p)}, \ell^{\tilde{\boldsymbol{\theta}}}_{(p)})$. Before continuing we need some more notation. For two pairs (i,j) and (k,ℓ) of natural numbers $(i,j) = (k,\ell)$ means $i = k$ and $j = \ell$. Moreover, for any pair (i,j), $1 \leq i \leq m, 1 \leq j \leq J_i(s^*)$, and $\breve{\boldsymbol{\theta}} \in \{\boldsymbol{\theta}, \tilde{\boldsymbol{\theta}}\}$ we denote by $\mathrm{rk}^{\breve{\boldsymbol{\theta}}}(i,j)$ the rank of $\varepsilon^{\breve{\boldsymbol{\theta}}}_{i,j-1}(S_{i,j}(\omega))$ among the $\varepsilon^{\breve{\boldsymbol{\theta}}}_{k,\ell-1}(S_{k,\ell}(\omega))$, $1 \leq k \leq m$, $1 \leq \ell \leq J_k(s^*)$, i.e. $\mathrm{rk}^{\breve{\boldsymbol{\theta}}}(i,j)$ is equal to 1 if $\varepsilon^{\breve{\boldsymbol{\theta}}}_{i,j-1}(S_{i,j}(\omega))$ is the smallest among the

$\varepsilon^{\breve{\boldsymbol\theta}}_{k,\ell-1}(S_{k,\ell}(\omega))$, $1 \le k \le m$, $1 \le \ell \le J_k(s^*)$. For the rest of the proof of part (a) let $\omega \in A_{m,,\boldsymbol\theta,\tilde{\boldsymbol\theta}}$.

Case 1: Suppose that $g^{\boldsymbol\theta}_{(p)} = k^{\tilde{\boldsymbol\theta}}_{(p)}$ and $h^{\boldsymbol\theta}_{(p)} = \ell^{\tilde{\boldsymbol\theta}}_{(p)}$. Then we immediately see that (19) is just the assumption stated in part (a).

Case 2: Suppose that $(g^{\boldsymbol\theta}_{(p)}, h^{\boldsymbol\theta}_{(p)}) \neq (k^{\tilde{\boldsymbol\theta}}_{(p)}, \ell^{\tilde{\boldsymbol\theta}}_{(p)})$ and that we additionally have $\mathrm{rk}^{\boldsymbol\theta}(k^{\tilde{\boldsymbol\theta}}_{(p)}, \ell^{\tilde{\boldsymbol\theta}}_{(p)}) > p$. Then (19) holds, because we have

$$\varepsilon^{\boldsymbol\theta}_{k^{\tilde{\boldsymbol\theta}}_{(p)},\ell^{\tilde{\boldsymbol\theta}}_{(p)}-1}(S_{k^{\tilde{\boldsymbol\theta}}_{(p)},\ell^{\tilde{\boldsymbol\theta}}_{(p)}}(\omega)) > \varepsilon^{\boldsymbol\theta}_{g^{\boldsymbol\theta}_{(p)},h^{\boldsymbol\theta}_{(p)}-1}(S_{g^{\boldsymbol\theta}_{(p)},h^{\boldsymbol\theta}_{(p)}}(\omega))$$

due to the fact that $\mathrm{rk}^{\boldsymbol\theta}(k^{\tilde{\boldsymbol\theta}}_{(p)}, \ell^{\tilde{\boldsymbol\theta}}_{(p)}) > p$ and the implication

$$\varepsilon^{\boldsymbol\theta}_{i,j-1}(S_{i,j-1}(\omega)) < \varepsilon^{\boldsymbol\theta}_{k^{\tilde{\boldsymbol\theta}}_{(p)},\ell^{\tilde{\boldsymbol\theta}}_{(p)}-1}(S_{k^{\tilde{\boldsymbol\theta}}_{(p)},\ell^{\tilde{\boldsymbol\theta}}_{(p)}}(\omega)) \Rightarrow \varepsilon^{\tilde{\boldsymbol\theta}}_{i,j-1}(S_{i,j-1}(\omega)) < \varepsilon^{\tilde{\boldsymbol\theta}}_{k^{\tilde{\boldsymbol\theta}}_{(p)},\ell^{\tilde{\boldsymbol\theta}}_{(p)}-1}(S_{k^{\tilde{\boldsymbol\theta}}_{(p)},\ell^{\tilde{\boldsymbol\theta}}_{(p)}}(\omega))$$

is just the assumption made in part (a).

Case 3: Let $(g^{\boldsymbol\theta}_{(p)}, h^{\boldsymbol\theta}_{(p)}) \neq (k^{\tilde{\boldsymbol\theta}}_{(p)}, \ell^{\tilde{\boldsymbol\theta}}_{(p)})$ and $\mathrm{rk}^{\boldsymbol\theta}(k^{\tilde{\boldsymbol\theta}}_{(p)}, \ell^{\tilde{\boldsymbol\theta}}_{(p)}) < p$ and assume additionally that $\mathrm{rk}^{\tilde{\boldsymbol\theta}}(g^{\boldsymbol\theta}_{(p)}, h^{\boldsymbol\theta}_{(p)}) < p$. Then (19) is true, because we have

$$\varepsilon^{\tilde{\boldsymbol\theta}}_{g^{\boldsymbol\theta}_{(p)},h^{\boldsymbol\theta}_{(p)}-1}(S_{g^{\boldsymbol\theta}_{(p)},h^{\boldsymbol\theta}_{(p)}}(\omega)) < \varepsilon^{\tilde{\boldsymbol\theta}}_{k^{\tilde{\boldsymbol\theta}}_{(p)},\ell^{\tilde{\boldsymbol\theta}}_{(p)}-1}(S_{k^{\tilde{\boldsymbol\theta}}_{(p)},\ell^{\tilde{\boldsymbol\theta}}_{(p)}}(\omega))$$

due to the fact that $\mathrm{rk}^{\tilde{\boldsymbol\theta}}(g^{\boldsymbol\theta}_{(p)}, h^{\boldsymbol\theta}_{(p)}) < p$ and the implication

$$\varepsilon^{\boldsymbol\theta}_{i,j-1}(S_{i,j-1}(\omega)) < \varepsilon^{\boldsymbol\theta}_{g^{\boldsymbol\theta}_{(p)},h^{\boldsymbol\theta}_{(p)}-1}(S_{g^{\boldsymbol\theta}_{(p)},h^{\boldsymbol\theta}_{(p)}}(\omega)) \Rightarrow \varepsilon^{\tilde{\boldsymbol\theta}}_{i,j-1}(S_{i,j-1}(\omega)) < \varepsilon^{\tilde{\boldsymbol\theta}}_{g^{\boldsymbol\theta}_{(p)},h^{\boldsymbol\theta}_{(p)}-1}(S_{g^{\boldsymbol\theta}_{(p)},h^{\boldsymbol\theta}_{(p)}}(\omega))$$

is again just the assumption stated in part (a).

Case 4: Let $(g^{\boldsymbol\theta}_{(p)}, h^{\boldsymbol\theta}_{(p)}) \neq (k^{\tilde{\boldsymbol\theta}}_{(p)}, \ell^{\tilde{\boldsymbol\theta}}_{(p)})$ and $\mathrm{rk}^{\boldsymbol\theta}(k^{\tilde{\boldsymbol\theta}}_{(p)}, \ell^{\tilde{\boldsymbol\theta}}_{(p)}) < p$ and assume additionally that $\mathrm{rk}^{\tilde{\boldsymbol\theta}}(g^{\boldsymbol\theta}_{(p)}, h^{\boldsymbol\theta}_{(p)}) > p$. Notice first of all that $\mathrm{rk}^{\boldsymbol\theta}(k^{\tilde{\boldsymbol\theta}}_{(p)}, \ell^{\tilde{\boldsymbol\theta}}_{(p)}) < p$ and $\mathrm{rk}^{\tilde{\boldsymbol\theta}}(g^{\boldsymbol\theta}_{(p)}, h^{\boldsymbol\theta}_{(p)}) > p$ together imply that there is at least one pair (e, f), $1 \le e \le m$, $1 \le f \le J_e(s^*)$ such that $\mathrm{rk}^{\boldsymbol\theta}(e, f) > p$ and $\mathrm{rk}^{\tilde{\boldsymbol\theta}}(e, f) < p$. Indeed, let

$$I^{\boldsymbol\theta} := \{(v, w), 1 \le v \le m, 1 \le w \le J_v(s^*) | \mathrm{rk}^{\boldsymbol\theta}(v, w) < \mathrm{rk}^{\boldsymbol\theta}(g, h) = p\}$$

and

$$I^{\tilde{\boldsymbol\theta}} := \{(v, w), 1 \le v \le m, 1 \le w \le J_v(s^*) | \mathrm{rk}^{\tilde{\boldsymbol\theta}}(v, w) < \mathrm{rk}^{\boldsymbol\theta}(g, h) = p\}.$$

Because by assumption we have $(k^{\tilde{\boldsymbol\theta}}_{(p)}, \ell^{\tilde{\boldsymbol\theta}}_{(p)}) \in I^{\boldsymbol\theta}$ and $(k^{\tilde{\boldsymbol\theta}}_{(p)}, \ell^{\tilde{\boldsymbol\theta}}_{(p)}) \notin I^{\tilde{\boldsymbol\theta}}$, Lemma A.1 implies that there is at least one pair $(e, f) \in I^{\tilde{\boldsymbol\theta}}$ such that $(e, f) \notin I^{\boldsymbol\theta}$. Hence, $\mathrm{rk}^{\tilde{\boldsymbol\theta}}(e, f) < p$ and $\mathrm{rk}^{\boldsymbol\theta}(e, f) \ge p$. Now, if $\mathrm{rk}^{\boldsymbol\theta}(e, f) = p$ we must have $(e, f) = (g^{\boldsymbol\theta}_{(p)}, h^{\boldsymbol\theta}_{(p)})$. However, this is impossible, because on one hand we have

$$(e, f) \in I^{\tilde{\boldsymbol\theta}}$$

and on the other hand

$$\mathrm{rk}^{\tilde{\boldsymbol{\theta}}}(g^{\boldsymbol{\theta}}_{(p)}, h^{\boldsymbol{\theta}}_{(p)}) > p.$$

This proves $\mathrm{rk}^{\boldsymbol{\theta}}(e,f) > p$ and therefore there is indeed at least one pair with the stated properties. Now we get from the assumption in part (a)

$$\varepsilon^{\boldsymbol{\theta}}_{ij-1}(S_{i,j-1}(\omega)) < \varepsilon^{\boldsymbol{\theta}}_{e,f-1}(S_{e,f}(\omega)) \Rightarrow \varepsilon^{\tilde{\boldsymbol{\theta}}}_{i,j-1}(S_{i,j-1}(\omega)) < \varepsilon^{\tilde{\boldsymbol{\theta}}}_{e,f-1}(S_{e,f}(\omega)).$$

This clearly implies (19), because

$$\varepsilon^{\boldsymbol{\theta}}_{e,f-1}(S_{e,f}(\omega)) > \varepsilon^{\boldsymbol{\theta}}_{g^{\boldsymbol{\theta}}_{(p)},h^{\boldsymbol{\theta}}_{(p)}-1}(S_{g^{\boldsymbol{\theta}}_{(p)},h^{\boldsymbol{\theta}}_{(p)}}(\omega))$$

and

$$\varepsilon^{\tilde{\boldsymbol{\theta}}}_{e,f-1}(S_{e,f}(\omega)) < \varepsilon^{\tilde{\boldsymbol{\theta}}}_{k^{\tilde{\boldsymbol{\theta}}}_{(p)},\ell^{\tilde{\boldsymbol{\theta}}}_{(p)}-1}(S_{k^{\tilde{\boldsymbol{\theta}}}_{(p)},\ell^{\tilde{\boldsymbol{\theta}}}_{(p)}}(\omega)).$$

This finishes the proof of part (a).

We now begin with the proof of the statement in part (b). Consider an arbitrary $\omega \in B_{m,\boldsymbol{\theta},\tilde{\boldsymbol{\theta}}}$. To shorten the notation we introduce the functions $F^{\breve{\boldsymbol{\theta}}}_{\omega}$, $\breve{\boldsymbol{\theta}} \in \{\boldsymbol{\theta}, \tilde{\boldsymbol{\theta}}\}$, defined on $\{(g,h) | 1 \le g \le m, 1 \le h \le J_g(s^*,\omega)\}$ by

$$F^{\breve{\boldsymbol{\theta}}}_{\omega}(g,h) := \sum_{i=1}^{m} \sum_{j=1}^{J_i(s^*,\omega)} \mathbb{1}_{(\varepsilon^{\breve{\boldsymbol{\theta}}}_{i,j-1}(S_{i,j-1}(\omega)),\infty)}(\varepsilon^{\breve{\boldsymbol{\theta}}}_{g,h-1}(S_{g,h}(\omega))),\ \breve{\boldsymbol{\theta}} \in \{\boldsymbol{\theta}, \tilde{\boldsymbol{\theta}}\}.$$

Before continuing with the actual proof, we state the following facts about the functions $F^{\boldsymbol{\theta}}_{\omega}$ and $F^{\tilde{\boldsymbol{\theta}}}_{\omega}$. Firstly, under the assumption made in part (a) of Theorem 2.10 we obtain from Lemma A.2 part (a), (i) that for every pair (g,h)

$$F^{\boldsymbol{\theta}}_{\omega}(g,h) \le F^{\tilde{\boldsymbol{\theta}}}_{\omega}(g,h). \tag{20}$$

Secondly, the assumptions in part (a) and part (b) of Theorem 2.10 together imply in view of Lemma A.2 (a), (ii) that

$$F^{\boldsymbol{\theta}}_{\omega}(\underline{k},\underline{\ell}) < F^{\tilde{\boldsymbol{\theta}}}_{\omega}(\underline{k},\underline{\ell}). \tag{21}$$

Thirdly, notice that part (b), (i) of Lemma A.2 implies for $\breve{\boldsymbol{\theta}} \in \{\boldsymbol{\theta}, \tilde{\boldsymbol{\theta}}\}$

$$F^{\breve{\boldsymbol{\theta}}}_{\omega}(g,h) \le F^{\breve{\boldsymbol{\theta}}}_{\omega}(k,\ell),\ \text{ if } \mathrm{rk}^{\breve{\boldsymbol{\theta}}}(g,h) \le \mathrm{rk}^{\breve{\boldsymbol{\theta}}}(k,\ell). \tag{22}$$

Fourthly, in the proof of part (a) of Theorem 2.10 we showed (19) so that we have

$$F^{\boldsymbol{\theta}}_{\omega}(g,h) \le F^{\tilde{\boldsymbol{\theta}}}_{\omega}(k,\ell),\ \text{ if } \mathrm{rk}^{\boldsymbol{\theta}}(g,h) = \mathrm{rk}^{\tilde{\boldsymbol{\theta}}}(k,\ell). \tag{23}$$

We now start considering all possible cases. In each case considered it is sufficient to show that the inequality in (23) is strict for two pairs (g,h) and (k,ℓ). This follows by

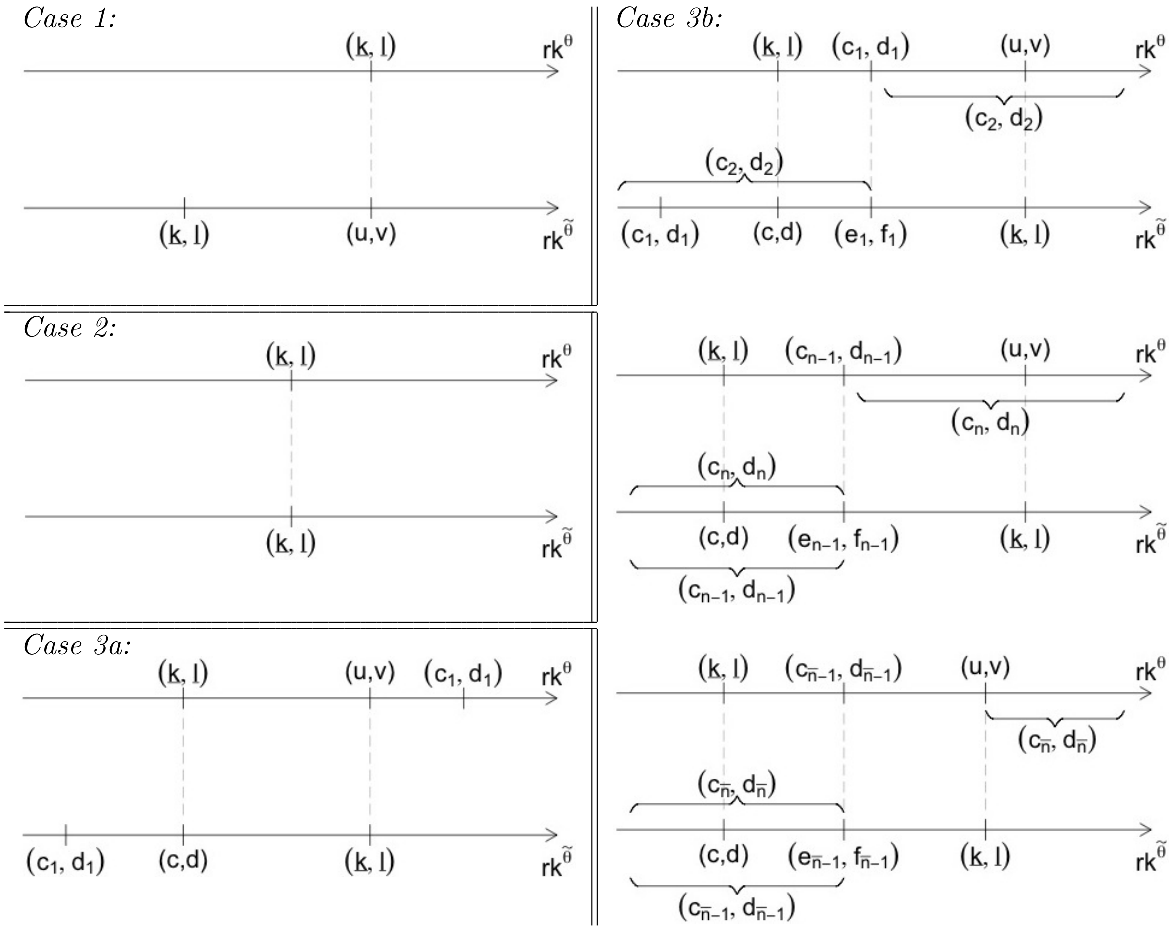

Figure 7: Illustration of the different cases considered in the proof of part (b) of Theorem 2.10. The ranks of the pairs under θ are given on the upper axis in ascending order and under $\tilde{\theta}$ on the lower axis. For Case 3b the three graphics on the right-hand side illustrate the process of extracting $\bar{n}$ with the property as given in (32).

combining (16) and (17). Of course, the two pairs might be equal. To ease the reading of the proof each case is illustrated by a graphic; see Fig. 7.

Case 1: $\mathrm{rk}^{\tilde{\boldsymbol{\theta}}}(\underline{k},\underline{\ell}) < \mathrm{rk}^{\boldsymbol{\theta}}(\underline{k},\underline{\ell})$. Then let the pair (u,v) be such that $\mathrm{rk}^{\tilde{\boldsymbol{\theta}}}(u,v) = \mathrm{rk}^{\boldsymbol{\theta}}(\underline{k},\underline{\ell})$ and notice that we consequently have $\mathrm{rk}^{\tilde{\boldsymbol{\theta}}}(u,v) > \mathrm{rk}^{\tilde{\boldsymbol{\theta}}}(\underline{k},\underline{\ell})$. We now obtain

$$F_\omega^{\tilde{\boldsymbol{\theta}}}(u,v) \overset{(22)}{\geq} F_\omega^{\tilde{\boldsymbol{\theta}}}(\underline{k},\underline{\ell}) \overset{(21)}{>} F_\omega^{\boldsymbol{\theta}}(\underline{k},\underline{\ell}).$$

Hence, the inequality in (23) is strict for the pairs (u,v) and $(\underline{k},\underline{\ell})$ that fulfil $p = \mathrm{rk}^{\tilde{\boldsymbol{\theta}}}(u,v) = \mathrm{rk}^{\boldsymbol{\theta}}(\underline{k},\underline{\ell})$.

Case 2: $\mathrm{rk}^{\tilde{\boldsymbol{\theta}}}(\underline{k},\underline{\ell}) = \mathrm{rk}^{\boldsymbol{\theta}}(\underline{k},\underline{\ell})$. Then we immediately see that the inequality in (23) is strict for $p = \mathrm{rk}^{\tilde{\boldsymbol{\theta}}}(\underline{k},\underline{\ell}) = \mathrm{rk}^{\boldsymbol{\theta}}(\underline{k},\underline{\ell})$ by (21).

Case 3: $\mathrm{rk}^{\tilde{\boldsymbol{\theta}}}(\underline{k},\underline{\ell}) > \mathrm{rk}^{\boldsymbol{\theta}}(\underline{k},\underline{\ell})$. Before continuing with Case 3 we remark that we will assume that

$$F_\omega^{\boldsymbol{\theta}}(g,h) = F_\omega^{\tilde{\boldsymbol{\theta}}}(k,\ell), \quad \text{if } \mathrm{rk}^{\boldsymbol{\theta}}(g,h) = \mathrm{rk}^{\tilde{\boldsymbol{\theta}}}(k,\ell). \tag{24}$$

The reason why we can assume Equation (24) is that if (24) did not hold, we would have found two pairs for which the inequality in (23) would be strict. However, this would finish the proof. To make the remainder of the proof easily accessible we state the following lemma.

Lemma A.3 *Under the assumption stated in (24) we have that*

$$rk^{\boldsymbol{\theta}}(i,j) \geq rk^{\boldsymbol{\theta}}(a,b) = rk^{\tilde{\boldsymbol{\theta}}}(\alpha,\beta) \geq rk^{\tilde{\boldsymbol{\theta}}}(i,j)$$

implies

$$F_\omega^{\boldsymbol{\theta}}(i,j) = F_\omega^{\boldsymbol{\theta}}(a,b) = F_\omega^{\tilde{\boldsymbol{\theta}}}(\alpha,\beta) = F_\omega^{\tilde{\boldsymbol{\theta}}}(i,j).$$

Proof of Lemma A.3

$$F_\omega^{\boldsymbol{\theta}}(i,j) \overset{(22)}{\geq} F_\omega^{\boldsymbol{\theta}}(a,b) \overset{(24)}{=} F_\omega^{\tilde{\boldsymbol{\theta}}}(\alpha,\beta) \overset{(22)}{\geq} F_\omega^{\tilde{\boldsymbol{\theta}}}(i,j) \overset{(20)}{\geq} F_\omega^{\boldsymbol{\theta}}(i,j).$$

Hence, we can replace all $\geq$ by $=$ which finishes the proof. □

Proof of Theorem 2.10 continued. Let the pairs (u,v) and (c,d) be such that

$$\mathrm{rk}^{\boldsymbol{\theta}}(u,v) = \mathrm{rk}^{\tilde{\boldsymbol{\theta}}}(\underline{k},\underline{\ell}) \text{ and } \mathrm{rk}^{\boldsymbol{\theta}}(\underline{k},\underline{\ell}) = \mathrm{rk}^{\tilde{\boldsymbol{\theta}}}(c,d). \tag{25}$$

From (24) we then get

$$F_\omega^{\tilde{\boldsymbol\theta}}(\underline{k},\underline{\ell}) = F_\omega^{\boldsymbol\theta}(u,v), \tag{26}$$

because $\mathrm{rk}^{\boldsymbol\theta}(u,v) = \mathrm{rk}^{\tilde{\boldsymbol\theta}}(\underline{k},\underline{\ell})$. Define now the sets

$$I_1^{\boldsymbol\theta} := \{(a,b), 1 \leq a \leq m, 1 \leq b \leq J_a(s^*) | \mathrm{rk}^{\boldsymbol\theta}(a,b) \leq \mathrm{rk}^{\boldsymbol\theta}(\underline{k},\underline{\ell})\}$$

and

$$I_1^{\tilde{\boldsymbol\theta}} := \{(a,b), 1 \leq a \leq m, 1 \leq b \leq J_a(s^*) | \mathrm{rk}^{\tilde{\boldsymbol\theta}}(a,b) \leq \mathrm{rk}^{\boldsymbol\theta}(\underline{k},\underline{\ell})\}.$$

Since $(\underline{k},\underline{\ell}) \in I_1^{\boldsymbol\theta}$ and $(\underline{k},\underline{\ell}) \notin I_1^{\tilde{\boldsymbol\theta}}$, it follows from Lemma A.1 that there must be a pair (c_1,d_1) such that $(c_1,d_1) \in I_1^{\tilde{\boldsymbol\theta}}$ and $(c_1,d_1) \notin I_1^{\boldsymbol\theta}$, i.e.

$$\mathrm{rk}^{\tilde{\boldsymbol\theta}}(c_1,d_1) \leq \mathrm{rk}^{\boldsymbol\theta}(\underline{k},\underline{\ell}) \overset{(25)}{=} \mathrm{rk}^{\tilde{\boldsymbol\theta}}(c,d) \text{ and } \mathrm{rk}^{\boldsymbol\theta}(c_1,d_1) > \mathrm{rk}^{\boldsymbol\theta}(\underline{k},\underline{\ell}). \tag{27}$$

Now make the following assumption that we call

Case 3a:
$$\mathrm{rk}^{\boldsymbol\theta}(c_1,d_1) \geq \mathrm{rk}^{\boldsymbol\theta}(u,v). \tag{28}$$

Under this assumption combining the left-hand side of (25) with the left-hand side of (27) we obtain from Lemma A.3

$$F_\omega^{\boldsymbol\theta}(c_1,d_1) = F_\omega^{\boldsymbol\theta}(u,v) = F_\omega^{\tilde{\boldsymbol\theta}}(\underline{k},\underline{\ell}) = F_\omega^{\tilde{\boldsymbol\theta}}(c_1,d_1)$$

which implies that

$$F_\omega^{\boldsymbol\theta}(c_1,d_1) = F_\omega^{\tilde{\boldsymbol\theta}}(\underline{k},\underline{\ell}). \tag{29}$$

Combining the left and right-hand side of (27) we have

$$\mathrm{rk}^{\tilde{\boldsymbol\theta}}(c_1,d_1) \leq \mathrm{rk}^{\boldsymbol\theta}(\underline{k},\underline{\ell}) = \mathrm{rk}^{\tilde{\boldsymbol\theta}}(c,d) < \mathrm{rk}^{\boldsymbol\theta}(c_1,d_1)$$

and we obtain from Lemma A.3

$$F_\omega^{\boldsymbol\theta}(c_1,d_1) = F_\omega^{\boldsymbol\theta}(\underline{k},\underline{\ell}) = F_\omega^{\tilde{\boldsymbol\theta}}(c,d) = F_\omega^{\tilde{\boldsymbol\theta}}(c_1,d_1). \tag{30}$$

However, in view of (29) the first equality in Equation (30) gives $F_\omega^{\boldsymbol\theta}(\underline{k},\underline{\ell}) = F_\omega^{\tilde{\boldsymbol\theta}}(\underline{k},\underline{\ell})$ which this is a contradiction to (21). Hence the assumption made in Equation (26) does not hold and we must have

$$F_\omega^{\tilde{\boldsymbol\theta}}(\underline{k},\underline{\ell}) > F_\omega^{\boldsymbol\theta}(u,v),$$

which finishes the proof if we are in *Case 3a*, i.e. the assumption stated in (28) holds. Now if (28) does not hold, we consider

Case 3b:
$$\mathrm{rk}^{\boldsymbol\theta}(c_1,d_1) < \mathrm{rk}^{\boldsymbol\theta}(u,v).$$

In this case continue extracting pairs (c_n, d_n), $n \geq 2$, with

$$(c_n, d_n) \in I_n^{\tilde{\boldsymbol{\theta}}} := \{(a,b), 1 \leq a \leq m, 1 \leq b \leq J_a(s^*) | \mathrm{rk}^{\tilde{\boldsymbol{\theta}}}(a,b) \leq \mathrm{rk}^{\boldsymbol{\theta}}(c_{n-1}, d_{n-1})\}, \text{ and}$$
$$(c_n, d_n) \notin I_n^{\boldsymbol{\theta}} := \{(a,b), 1 \leq a \leq m, 1 \leq b \leq J_a(s^*) | \mathrm{rk}^{\boldsymbol{\theta}}(a,b) \leq \mathrm{rk}^{\boldsymbol{\theta}}(c_{n-1}, d_{n-1})\}. \tag{31}$$

Clearly, the pairs (c_n, d_n) and the sets as given in (31) exist by Lemma A.1 at least as long as $\mathrm{rk}^{\boldsymbol{\theta}}(c_{n-1}, d_{n-1}) < \mathrm{rk}^{\boldsymbol{\theta}}(u,v) = \mathrm{rk}^{\tilde{\boldsymbol{\theta}}}(\underline{k}, \underline{\ell})$, because by construction $(c_{n-1}, d_{n-1}) \in I_n^{\tilde{\boldsymbol{\theta}}}$ and $(c_{n-1}, d_{n-1}) \in I_n^{\boldsymbol{\theta}}$, but $(\underline{k}, \underline{\ell}) \notin I_n^{\tilde{\boldsymbol{\theta}}}$ and $(\underline{k}, \underline{\ell}) \in I_n^{\boldsymbol{\theta}}$. Because $\mathrm{rk}^{\boldsymbol{\theta}}(c_n, d_n)$ is strictly increasing for every ω (with the property that we have to consider Case 3 and that we are not in *Case 3a*, i.e. the assumption stated in (28) does not hold) there is an $\bar{n} \geq 2$ such that

$$\mathrm{rk}^{\boldsymbol{\theta}}(c_{\bar{n}}, d_{\bar{n}}) \geq \mathrm{rk}^{\boldsymbol{\theta}}(u,v) \overset{(25)}{=} \mathrm{rk}^{\tilde{\boldsymbol{\theta}}}(\underline{k}, \underline{\ell}) \text{ and } \mathrm{rk}^{\boldsymbol{\theta}}(c_{\bar{n}-1}, d_{\bar{n}-1}) < \mathrm{rk}^{\boldsymbol{\theta}}(u,v). \tag{32}$$

Notice that Equation (32) together with (31) implies

$$\mathrm{rk}^{\tilde{\boldsymbol{\theta}}}(c_{\bar{n}}, d_{\bar{n}}) < \mathrm{rk}^{\boldsymbol{\theta}}(u,v) \overset{(32)}{=} \mathrm{rk}^{\tilde{\boldsymbol{\theta}}}(\underline{k}, \underline{\ell}). \tag{33}$$

Let the pair (e_{n-1}, f_{n-1}) be such that

$$\mathrm{rk}^{\tilde{\boldsymbol{\theta}}}(e_{n-1}, f_{n-1}) = \mathrm{rk}^{\boldsymbol{\theta}}(c_{n-1}, d_{n-1}),\ 2 \leq n \leq \bar{n}. \tag{34}$$

Then

$$\mathrm{rk}^{\tilde{\boldsymbol{\theta}}}(c_n, d_n) \leq \mathrm{rk}^{\tilde{\boldsymbol{\theta}}}(e_{n-1}, f_{n-1}) \overset{(34)}{=} \mathrm{rk}^{\boldsymbol{\theta}}(c_{n-1}, d_{n-1}) < \mathrm{rk}^{\boldsymbol{\theta}}(c_n, d_n),\ 2 \leq n \leq \bar{n}. \tag{35}$$

Applying Lemma A.3 to (35) we now obtain for every n, $2 \leq n \leq \bar{n}$

$$F_\omega^{\boldsymbol{\theta}}(c_n, d_n) = F_\omega^{\boldsymbol{\theta}}(c_{n-1}, d_{n-1}) = F_\omega^{\tilde{\boldsymbol{\theta}}}(e_{n-1}, f_{n-1}) = F_\omega^{\boldsymbol{\theta}}(c_n, d_n). \tag{36}$$

Furthermore, combining (32) and (33) we have

$$\mathrm{rk}^{\boldsymbol{\theta}}(c_{\bar{n}}, d_{\bar{n}}) \geq \mathrm{rk}^{\boldsymbol{\theta}}(u,v) = \mathrm{rk}^{\tilde{\boldsymbol{\theta}}}(\underline{k}, \underline{\ell}) > \mathrm{rk}^{\tilde{\boldsymbol{\theta}}}(c_{\bar{n}}, d_{\bar{n}}).$$

Applying once again Lemma A.3 we obtain

$$F_\omega^{\boldsymbol{\theta}}(c_{\bar{n}}, d_{\bar{n}}) = F_\omega^{\boldsymbol{\theta}}(u,v) = F_\omega^{\tilde{\boldsymbol{\theta}}}(\underline{k}, \underline{\ell}) = F_\omega^{\tilde{\boldsymbol{\theta}}}(c_{\bar{n}}, d_{\bar{n}}). \tag{37}$$

In particular,

$$F_\omega^{\tilde{\boldsymbol{\theta}}}(\underline{k}, \underline{\ell}) = F_\omega^{\boldsymbol{\theta}}(c_{\bar{n}}, d_{\bar{n}}), \tag{38}$$

and we know from Equation (30) which is valid here as well as it was derived from Equation (27)

$$F_\omega^{\boldsymbol{\theta}}(\underline{k}, \underline{\ell}) = F_\omega^{\boldsymbol{\theta}}(c_1, d_1). \tag{39}$$

Together with (36) Equations (38) and (39) now imply that

$$F_{\omega}^{\tilde{\boldsymbol{\theta}}}(\underline{k},\underline{\ell}) = F_{\omega}^{\boldsymbol{\theta}}(\underline{k},\underline{\ell}).$$

However, this is a contradiction to (21) showing that also in *Case 3b* the assumed equality in (24) cannot hold. This finishes the proof of part (b). □

B. Discrete hazard rates

Recall that for a probability measure $\mathbb{P}$ on $\mathbb{R}$ with a *finite* support on $\{z_1,\dots,z_K\}$, say, the (discrete) hazard rate λ and the survival function S are given by the formulas

$$\lambda(t) := \frac{\mathbb{P}(\{t\})}{\mathbb{P}([t,\infty))}, \quad \text{and } \mathbb{P}((t,\infty)) =: S(t) = \prod_{i:z_i\le t}(1-\lambda(z_i)). \tag{40}$$

Recall also that the relation $\mathbb{P}([t,\infty)) = S(t-) = \prod_{i:z_i<t}(1-\lambda(z_i))$ implies that

$$\mathbb{P}(\{t\}) = \lambda(t)\cdot S(t-).$$

Now we shall define an effective age model corresponding to a sequence of discrete probability measures $\mathbb{P}_d$ with supports $\{z_1,\dots,z_{K_1},z_{K_1+1},\dots,z_{K_d}\}$, $d\in\mathbb{N}$ (the method could also be used for a continuous probability measure). This model has exactly the same transition kernels as Model 2.1. To do so, assume that for every $\boldsymbol{\theta}\in\Theta$ we have given a sequence of functions $\varepsilon_0^{\boldsymbol{\theta}},\varepsilon_1^{\boldsymbol{\theta}},\dots$, $\boldsymbol{\theta}\in\Theta$, that are interpreted as the effective age functions for the time interval between the jth and the $(j-1)$th event, with the following properties:

(a) $\varepsilon_0^{\boldsymbol{\theta}}:\mathbb{R}_+\to\mathbb{R}_+$,

(b) $\varepsilon_{j-1}^{\boldsymbol{\theta}}:\mathbb{D}\subset\mathbb{R}_+^j\to\mathbb{R}_+$, $j=2,3,\dots$ and $\mathbb{D}$ denotes the domain;

(c) The functions $\varepsilon_{j-1;s_1,\dots,s_{j-1}}^{\boldsymbol{\theta}}(\cdot) := \varepsilon_{j-1}^{\boldsymbol{\theta}}(s_1,\dots,s_{j-1},\cdot)$ are strictly increasing for every fixed $s_1,\dots,s_{j-1}$.

Here $\mathbb{R}_+^d := \{x \in \mathbb{R}^d | x_i \geq 0, i = 1, \ldots, d\}$. For every $d \in \mathbb{N}$ we now define a probability measure $\mathbb{Q}_d$ on $\mathbb{R}^d$ by defining for $j = 1, \ldots, d$ transition kernels

$$\begin{aligned}
\mathbb{Q}_j\left(s_j | s_1, \ldots, s_{j-1}, \varepsilon^{\boldsymbol{\theta}}_{j-1}\right) &:= \frac{\mathbb{P}_d\left(\{\varepsilon^{\boldsymbol{\theta}}_{j-1;s_1,\ldots,s_{j-1}}(s_j)\}\right)}{\mathbb{P}_d\left((\varepsilon^{\boldsymbol{\theta}}_{j-1;s_1,\ldots,s_{j-1}}(s_{j-1}), \infty)\right)}, \\
&= \lambda_d\left(\varepsilon^{\boldsymbol{\theta}}_{j-1;s_1,\ldots,s_{j-1}}(s_j)\right) \frac{\mathbb{P}_d\left([\varepsilon^{\boldsymbol{\theta}}_{j-1;s_1,\ldots,s_{j-1}}(s_j), \infty)\right)}{\mathbb{P}_d\left((\varepsilon^{\boldsymbol{\theta}}_{j-1;s_1,\ldots,s_{j-1}}(s_{j-1}), \infty)\right)} \\
&= \lambda_d\left(\varepsilon^{\boldsymbol{\theta}}_{j-1;s_1,\ldots,s_{j-1}}(s_j)\right) \frac{S_d\left(\varepsilon^{\boldsymbol{\theta}}_{j-1;s_1,\ldots,s_{j-1}}(s_j)-\right)}{S_d\left(\varepsilon^{\boldsymbol{\theta}}_{j-1;s_1,\ldots,s_{j-1}}(s_{j-1})\right)} \\
&\quad \text{for } s_j > s_{j-1}, \qquad (41)
\end{aligned}$$

where we use the convention that $0/0 := 0$, $(\varepsilon^{\boldsymbol{\theta}}_{j-1;s_1,\ldots,s_{j-1}})^{-1}$ denotes the inverse of $\varepsilon^{\boldsymbol{\theta}}_{j-1;s_1,\ldots,s_{j-1}}$, $s_0 = 0$, and λ_d and S_d are the hazard rate and the survival function corresponding to $\mathbb{P}_d$, respectively. In Equation (41) we have implicitly assumed that $\mathbb{P}_d$ and the functions $\varepsilon^{\boldsymbol{\theta}}_{j-1;s_1,\ldots,s_{j-1}}$ match properly, i.e. $\mathbb{P}_d\left((\varepsilon^{\boldsymbol{\theta}}_{j-1;s_1,\ldots,s_{j-1}}(s_{j-1}), \infty)\right) > 0$, $j = 1, 2, \ldots$.

Remark B.1 Constructing effective age models through Equation (41) is similar to the approaches in Dorado et al. (1997) and Last and Szekli (1998).

Example B.2 For the model discussed in Example 2.2 we have for $\theta \in [0, 1]$ that $\varepsilon^\theta_0(s) := s$, $s \in \mathbb{R}_+$. Moreover, we have for $\theta \in [0, 1]$ that $\varepsilon^\theta_{d-1} : \mathbb{R}^j_+ \to \mathbb{R}_+$, $j \geq 2$, with $\varepsilon^\theta_{j-1}(\mathbf{s}) := s_j - \theta s_{j-1}$, $\mathbf{s} = (s_1, \ldots, s_{j-1}, s_j) \in \mathbb{R}^j_+$.
Furthermore, for the model discussed in Example 2.3 we have for $\theta \in [0, 1]$ that $\varepsilon^\theta_0(s) := s$, $s \in \mathbb{R}_+$. Moreover, we have for $\theta \in [0, 1]$ that $\varepsilon^\theta_{j-1} : \mathbb{R}^j_+ \to \mathbb{R}_+$, $j \geq 2$, with $\varepsilon^\theta_{j-1}(\mathbf{s}) := s_j - \theta \sum_{\ell=1}^{j-1} (1-\theta)^{j-1-\ell} s_\ell$, $\mathbf{s} = (s_1, \ldots, s_{j-1}, s_j) \in \mathbb{R}^d_+$.

Now, let $[0, s]$, $\mathbf{D}_m(s)$, $N_i(s)$ and τ_i be as in Section 2 and denote the event occurrence times by $\tilde{s}_{ij}$, $1 \leq i \leq m$, $1 \leq j \leq N_i(s)$. Then the full likelihood corresponding to the model defined by (41) is given by with $d = \max\{N_1(s), \dots, N_m(s)\}$

$$
\begin{aligned}
L_F(s|\mathbb{P}_d, (\varepsilon^{\boldsymbol{\theta}}_{i,j-1})_{1\le i\le m, j\in\mathbb{N}}, \mathbf{D}_m(s)) &= \prod_{i=1}^{m}\prod_{j=1}^{N_i(s)} \lambda_d\left(\varepsilon^{\boldsymbol{\theta}}_{i,j-1;\tilde{s}_{i,1},\dots,\tilde{s}_{i,j-1}}(\tilde{s}_{i,j})\right)\\
&\quad\times \prod_{i=1}^{m}\prod_{j=1}^{N_i(s)} \frac{S_d\left(\varepsilon^{\boldsymbol{\theta}}_{i,j-1;\tilde{s}_{i,1},\dots,\tilde{s}_{i,j-1}}(\tilde{s}_{i,j}-)\right)}{S_d\left(\varepsilon^{\boldsymbol{\theta}}_{i,j-1;\tilde{s}_{i,1},\dots,\tilde{s}_{i,j-1}}(\tilde{s}_{i,j-1})\right)}\\
&\quad\times \prod_{i=1}^{m} \frac{S_d\left(\varepsilon^{\boldsymbol{\theta}}_{i,N_i(s);\tilde{s}_{i,1},\dots,\tilde{s}_{i,N_i(s)}}(\tau_i\wedge s)\right)}{S_d\left(\varepsilon^{\boldsymbol{\theta}}_{i,N_i(s);\tilde{s}_{i,1},\dots,\tilde{s}_{i,N_i(s)}}(\tilde{s}_{i,N_i(s)})\right)}\\
&\overset{(40)}{=} \prod_{i=1}^{m}\prod_{j=1}^{N_i(s)} \lambda_d\left(\varepsilon^{\boldsymbol{\theta}}_{i,j-1;\tilde{s}_{i,1},\dots,\tilde{s}_{i,j-1}}(\tilde{s}_{i,j})\right)\\
&\quad\times\left(\prod_{i=1}^{m}\prod_{j=1}^{N_i(s)}\prod_{k\in I^{\boldsymbol{\theta}}_{i,j}}(1-\lambda_d(z_k))\right)\\
&\quad\times\left(\prod_{i=1}^{m}\prod_{k\in I^{\boldsymbol{\theta},\tau_i}_{i}}(1-\lambda_d(z_k))\right),
\end{aligned}
\tag{42}
$$

where for every pair (i,j), $1\le i\le m$, $1\le j\le N_i(s)$, the sets $I^{\boldsymbol{\theta}}_{i,j}$ are defined by

$$
I^{\boldsymbol{\theta}}_{i,j} := \{k\in\{1,\dots,K\} | \varepsilon^{\boldsymbol{\theta}}_{i,j-1;\tilde{s}_{i,1},\dots,\tilde{s}_{i,j-1}}(\tilde{s}_{i,j-1}) < z_k < \varepsilon^{\boldsymbol{\theta}}_{i,j-1;\tilde{s}_{i,1},\dots,\tilde{s}_{i,j-1}}(\tilde{s}_{i,j})\}
$$

and for every i, $1\le i\le m$, the sets $I^{\boldsymbol{\theta},\tau_i}_{i}$ are defined by

$$
I^{\boldsymbol{\theta},\tau_i}_{i} := \{k\in\{1,\dots,K\} | \varepsilon^{\boldsymbol{\theta}}_{i,N_i(s);\tilde{s}_{i,1},\dots,\tilde{s}_{i,N_i(s)}}(\tilde{s}_{i,N_i(s)}) < z_k \le \varepsilon^{\boldsymbol{\theta}}_{i,N_i(s);\tilde{s}_{i,1},\dots,\tilde{s}_{i,N_i(s)}}(\tau_i\wedge s)\}.
$$

To see that Equation (43) is indeed the discrete time analogue of Equation (2) notice that

$$
\begin{aligned}
&\left(\prod_{i=1}^{m}\prod_{j=1}^{N_i(s)}\prod_{k\in I^{\boldsymbol{\theta}}_{i,j}}(1-\lambda_d(z_k))\right)\cdot\left(\prod_{i=1}^{m}\prod_{k\in I^{\boldsymbol{\theta},\tau_i}_{i}}(1-\lambda_d(z_k))\right)\\
&=\left(\prod_{i=1}^{m}\prod_{j=1}^{N_i(s)}\prod_{u\in(I^{\boldsymbol{\theta}}_{i,j})^{-1}}\left(1-\lambda_d\left(\varepsilon^{\boldsymbol{\theta}}_{i,j-1;\tilde{s}_{i,1},\dots,\tilde{s}_{i,j-1}}(u)\right)\right)\right)\\
&\quad\times\left(\prod_{i=1}^{m}\prod_{u\in(I^{\boldsymbol{\theta},\tau_i}_{i})^{-1}}\left(1-\lambda_d\left(\varepsilon^{\boldsymbol{\theta}}_{i,N_i(s);\tilde{s}_{i,1},\dots,\tilde{s}_{i,N_i(s)}}(u)\right)\right)\right),
\end{aligned}
\tag{43}
$$

where for every pair (i,j), $1 \le i \le m$, $1 \le j \le N_i(s)$, the *intervals* $(I^{\boldsymbol{\theta}}_{i,j})^{-1}$ are defined by

$$(I^{\boldsymbol{\theta}}_{i,j})^{-1} := \{u \in \mathbb{R}^+ | \tilde{s}_{i,j-1} < u < \tilde{s}_{i,j}\}$$

and for every i, $1 \le i \le m$, the *intervals* $(I^{\boldsymbol{\theta},\tau_i}_{i})^{-1}$ are defined by

$$(I^{\boldsymbol{\theta},\tau_i}_{i})^{-1} := \{u \in \mathbb{R}^+ | \tilde{s}_{i,N_i(s)} < u \le \tau_i \wedge s\},$$

and the product w.r.t. u stands for the product integral.

Now we are going to maximise (42) for $\boldsymbol{\theta}$ fixed. As in Section 2.2, maximisation is w.r.t. all discrete probability measures that put (positive) mass at the points $\varepsilon^{\boldsymbol{\theta}}_{k,\ell-1;\tilde{s}_{k,1},\dots,\tilde{s}_{k,\ell-1}}(\tilde{s}_{k,\ell})$, $1 \le k \le m$, $1 \le \ell \le N_k(s)$, which we assume to be different. Also as in Section 2.2 we denote the corresponding hazard rates at these points by $\lambda^{\boldsymbol{\theta}}_{k,\ell}$. Then the full likelihood (42) after rearranging can be written as

$$\prod_{k=1}^{m} \prod_{\ell=1}^{N_k(s)} \lambda^{\boldsymbol{\theta}}_{k,\ell} \prod_{k=1}^{m} \prod_{\ell=1}^{N_k(s)} (1-\lambda^{\boldsymbol{\theta}}_{k,\ell})^{|I^{\boldsymbol{\theta},B}_{k,\ell}|+|I^{\boldsymbol{\theta},\tau_k,B}_{k}|}, \tag{44}$$

where for every pair (k,ℓ), $1 \le k \le m$, $1 \le \ell \le N_k(s)$, the sets $I^{\boldsymbol{\theta},B}_{k,\ell}$ are defined (cf. Section 2.2) as

$$I^{\boldsymbol{\theta},B}_{k,\ell} := \{(i,j),\, 1 \le i \le m, 1 \le j \le N_i(s) | \\ \varepsilon^{\boldsymbol{\theta}}_{i,j-1;\tilde{s}_{i,1},\dots,\tilde{s}_{i,j-1}}(\tilde{s}_{i,j-1}) < \varepsilon^{\boldsymbol{\theta}}_{k,\ell-1;\tilde{s}_{k,1},\dots,\tilde{s}_{k,\ell-1}}(\tilde{s}_{k,\ell}) < \varepsilon^{\boldsymbol{\theta}}_{i,j-1;\tilde{s}_{i,1},\dots,\tilde{s}_{i,j-1}}(\tilde{s}_{i,j})\}$$

and for every k, $1 \le k \le m$, the sets $I^{\boldsymbol{\theta},\tau_i,B}_{k}$ are defined (cf. Section 2.2) by

$$I^{\boldsymbol{\theta},\tau_i,B}_{k} := \{i,\, 1 \le i \le m | \\ \varepsilon^{\boldsymbol{\theta}}_{i,N_i(s);\tilde{s}_{i,1},\dots,\tilde{s}_{i,N_i(s)}}(\tilde{s}_{i,N_i(s)}) < \varepsilon^{\boldsymbol{\theta}}_{k,\ell-1;\tilde{s}_{k,1},\dots,\tilde{s}_{k,\ell-1}}(\tilde{s}_{k,\ell}) \le \varepsilon^{\boldsymbol{\theta}}_{i,N_i(s);\tilde{s}_{i,1},\dots,\tilde{s}_{i,N_i(s)}}(\tau_i \wedge s)\}.$$

Moreover, as in Appendix A for any set I we denote by $|I|$ its cardinality. As the function $x \to x(1-x)^k$, $k \in \mathbb{N}_0$ with $x \in [0,1]$, is maximised at $x = 1/(k+1)$ we see that Equation (44) is maximised at

$$\lambda^{\boldsymbol{\theta}}_{k,\ell} = \frac{1}{|I^{\boldsymbol{\theta},B}_{k,\ell}| + |I^{\boldsymbol{\theta},\tau_k,B}_{k}| + 1}.$$

Notice that this NPMLE coincides with the one given in Section 2.2 if the derivatives of the effective functions w.r.t. observational time are equal to one, because we have the relations $|I^{\boldsymbol{\theta},B}_{k,\ell}| + 1 = |I^{\boldsymbol{\theta}}_{k,\ell}|$ and $|I^{\boldsymbol{\theta},\tau_k,B}_{k}| = |I^{\boldsymbol{\theta},\tau_k}_{k}|$.

C. Identifiability in Examples 2.2 and 2.3

First note that the distribution of (S_1, S_2) (or (T_1, T_2)) is the same under Kijima I or Kijima II models; see Examples 2.2 and 2.3. Indeed, the joint distribution of (T_1, T_1) is defined by its density function

$$g^{(\theta,f)}_{T_1,T_2}(t_1,t_2) = f(t_1)f(t_2 + (1-\theta)t_1)/S((1-\theta)t_1)\mathbb{1}_{\{t_1\geq 0\}}\mathbb{1}_{\{t_2\geq 0\}},$$

for $(t_1, t_2) \in \mathbb{R}^2_+$. Here f is an unknown probability density function with support $[0, \infty)$ (S is the corresponding survival function) and θ is an unknown Euclidean parameter in $[0, 1]$. Let μ_k be the Lebesgue measure on $\mathbb{R}^k$ ($k \geq 1$). Proving the identifiability of (θ, f) requires to show the following one–to–one property: $g^{(\theta,f)}_{T_1,T_2} = g^{(\tilde\theta,\tilde f)}_{T_1,T_2}$ μ_2–a.e. implies $(\theta, f) = (\tilde\theta, \tilde f)$. The fact that $f = \tilde f$ is straightforward by integrating $g^{(\theta,f)}_{T_1,T_2}(t_1, t_2)$ with respect to t_2 on $\mathbb{R}$. Thus identifiability reduces to proving that $\theta = \tilde\theta$ results from

$$f(t_2 + (1-\theta)t_1)/S((1-\theta)t_1) = f(t_2 + (1-\tilde\theta)t_1)/S((1-\tilde\theta)t_1) \quad \mu_2 - \text{a.e.} \tag{45}$$

Integrating (45) with respect to t_2 on $[s, \infty)$ leads to

$$S(s + (1-\theta)t_1)/S((1-\theta)t_1) = S(s + (1-\tilde\theta)t_1)/S((1-\tilde\theta)t_1) \quad \mu_2 - \text{a.e.} \tag{46}$$

Using (45) (with $t_2 = s$) and (46) we obtain

$$\lambda(s + (1-\theta)t_1) = \lambda(s + (1-\tilde\theta)t_1) \quad \mu_2 - \text{a.e.} \tag{47}$$

Now suppose that there exists a non empty open interval (a, b) such that λ is one–to–one on (a, b). Let $\tilde\theta \neq \theta$. Then for any pair (s, t_1) quantities $s + (1-\theta)t_1$ and $s + (1-\tilde\theta)t_1$ belong to (a, b) simultaneously if and only if

$$\begin{cases} a - (1-\theta)t_1 < s < b - (1-\theta)t_1, \\ a - (1-\tilde\theta)t_1 < s < b - (1-\tilde\theta)t_1. \end{cases}$$

The latter holds if and only if

$$a - (1-\theta\vee\tilde\theta)t_1 < b - (1-\theta\wedge\tilde\theta)t_1 \Leftrightarrow t_1 < \frac{b-a}{|\tilde\theta-\theta|}.$$

Then for μ_2 almost all $(t_1, s) \in \{(x, y) \in \mathbb{R}^2_+ | x < (b-a)/|\tilde\theta-\theta|, a - (1-\theta\vee\tilde\theta)t_1 < y < b - (1-\theta\wedge\theta)t_1\}$ we must have $s + (1-\theta)t_1 = s + (1-\tilde\theta)t_1$. Hence $\theta = \tilde\theta$ which proves the semi-parametric identifiability of Kijima I and Kijima II models whenever the first two failures can be observed.

References

Aalen, O.O., Borgan, Ø. and Gjessing, H.K. (2008). *Survival and Event History Analysis: A Process Point of View*, Springer Science + Business Media, New York.

Adekpedjou, A. and Stocker, R. (2015). A general class of semiparametric models for recurrent event data. *Journal of Statistical Planning and Inference*, 156, 48–63.

Andersen, P.K. and Gill, R.D. (1982). Cox's Regression Model for Counting Processes: A Large Sample Study. *Annals of Statistics*, 10, 1100-1120.

Andersen, P.K., Borgan, Ø, Gill, R.D. and Keidling, N. (1993). *Statistical Models Based on Counting Processes*. Springer-Verlag, New York.

Barndorff-Nielsen, O.E. (1988). *Parametric Statistical Models and Inference*, Springer-Verlag, New York.

Barndorff-Nielsen, O.E. and Cox, D.R. (1994). *Inference and Asymptotics*, Chapman and Hall/CRC, Boca Raton.

Braekers, R. and Veraverbeke, N. (2005). Cox's regression model under partially informative censoring. *Communications in Statistics - Theory and Methods*, 34, 1793–1811.

Breslow, N., McNeney, B. and Wellner, J.A. (2003). Large sample theory for semiparametric regression models with two-phase, outcome dependent sampling. *Annals of Statistics*, 31, 1110–1139.

Claeskens, G. and Carroll, R.J. (2007). An asymptotic theory for model selection inference in general semiparametric problems. *Biometrika*, 94, 249–265.

Cook, R.J. and Lawless, J. (2007). *The Statistical Analysis of Recurrent Events*, Springer Science + Business Media, New York.

Davison, A.C. (2003). *Statistical models*, Cambridge University Press, Cambridge.

Dorado, C., Hollander, M. and Sethuramam, J. (1997). Nonparametric estimation for a general repair model. *Annals of Statistics*, 25, 1140–1160.

Doyen, L. and Gaudoin, O. (2004). Classes of imperfect repair models based on reduction of failure intensity or virtual age. Reliability Engineering and System Safety, 84, 45–56.

Fraser, D.A.S. (2003). Likelihood for Component Parameters. *Biometrika*, 90, 327–339.

Gärtner, M. (2003). Estimation in a general repair model based on left-truncated data. *Scandinavian Journal of Statistics*, 30, 61–73.

Gill, R.D., (1981). Testing with replacement and the product-limit estimator. *Annals of Statistics*, 9, 853–860.

González, J.R., Peña, E.A. and Slate, E.H. (2005). Modelling intervention effects after cancer relapses. *Statistics in Medicine*, 24, 3959–3975.

Hirosi, Y. (2011). Efficiency of profile likelihood in semi-parametric models. *Annals of the Institute of Statistical Mathematics*, 63, 1247–1275.

Hollander, M. and Sethuramam, J. (2004). Nonparametric Methods for Repair models. In: N. Balakrishnan and C.R. Rao (eds.), Handbook of Statistics, Vol. 23, Advances in Survival Analysis, Elsevier, Amsterdam, 747–763.

Huang, C.-J., Kin, J., and Follmann, D.A. (2012). A maximum pseudo-profile likelihood estimator for the Cox model under length-biased sampling. *Biometrika*, 99, 199–210.

Jacod, J. (1975). Multivariate point processes: Predictable projection, Radon-Nikodym derivatives, representation of martingales. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 3, 235–253.

Johansen, S. (1983). An Extension of Cox's Regression Model. *International Statistical Review*, 51, 165–174.

Kijima, M., Morimura, H. and Suzuki, Y. (1988). Periodical replacement problem without assuming minimal repair. *European Journal of Operational Research*, 37, 194–203.

Kijima, M. (1989). Some results for repairable systems with general repair. *Journal of Applied Probability*, 26, 89–102.

Last, G. and Szekli, R. (1998). Asymptotic and monotonicity properties of some repairable systems. *Advances in Applied Probability*, 30, 1089–1110.

Lawless, J.F. (1995). The analysis of recurrent events for multiple subjects. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 44, 487–498.

Lindqvist, B.H. (2006). On the statistical modeling and analysis of repairable systems. *Statistical Science*, 21, 532–551.

McCullagh, P. and Tibshirani, R. (1990). Adjustments to profile likelihood. *Journal of the Royal Statistical Society: Series B*, 52, 325–344.

Murphy, S.A. and van der Vaart, A.W. (2000). On profile likelihood (with discussion), *Journal of the American Statistical Association*, 95, 449–485.

Nelson, W.B. (2003). *Recurrent Events Data Analysis for Product Repairs, Disease Recurrences, and Other Applications.* ASA-SIAM Series on Statistics and Applied Probability, SIAM, Philadelphia.

Peña, E.A., Strawderman, R.L. and Hollander, M. (2001). Nonparametric estimation with recurrent event data. *Journal of the American Statistical Association*, 96, 1299–1315.

Peña, E.A. and Hollander, M. (2004). Models for recurrent events in reliability and survival analysis. In: R. Soyer, T. Mazzuchi and N. Singpurwalla (Eds.), Mathematical Reliability: An Expository Perspective, Kluwer, Boston, 105–123.

Peña, E.A. (2006). Dynamic Modeling and Statistical Analysis of Event Times. *Statistical Science*, 21, 487–500.

Peña, E.A., Slate, E.H. and González, J.R. (2007). Semiparametric inference for a general class of models for recurrent events. *Journal of Statistical Planning and Inference*, 137, 1727–1747.

Peña, E.A. (2014). Asymptotics for a class of dynamic recurrent event models. arXiv 1410.8729v2.

Reid, N. (2013). Aspects of likelihood inference. *Bernoulli*, 19, 1404–1418.

Scott, J.A. and Wild, C.J. (1997). Fitting regression models to case-control data by maximum likelihood. *Biometrika*, 84, 57–71.

Selke, T. and Siegmund, D. (1983). Sequential Analysis of the Proportional Hazards Model. *Biometrika*, 70, 315–326.

Sellke, T., 1988. Weak convergence of the Aalen estimator for a censored renewal process. In: Gupta, S. and Berger, J. (Eds.), Statistical Decision Theory and Related Topics IV, Vol. 2, Springer-Verlag, New York, 183–194.

Severini, T.A. and Wong, W.H. (1992). Profile likelihood and conditionally parametric models. *Annals of Statistics*, 20, 1768–1802.

Severini, T.A. (2000). *Likelihood Methods in Statistics.* Oxford University Press, Oxford.

Slud, E.V. and Vonta, F. (2005). Efficient semiparametric estimators via modified profile likelihood. *Journal of Statistical Planning and Inference*, 129, 339–367.

Xu, R., Vaida, F. and Harrington, D.P. (2009). Using profile likelihood for semiparametric model selection with application to proportional hazards mixed model. *Statistica Sinica*, 19, 819–842.

Zeng, D. and Lin, D.Y. (2007). Efficient estimation for the accelerated failure time model. *Journal of the American Statistical Association*, 102, 1387–1396.

Zeng, D. and Lin, D.Y. (2010). A general asymptotic theory for maximum likelihood estimation in semiparametric regression models with censored data. *Statistica Sinica*, 20, 871–910.