Estimation of the extreme survival probabilities from censored data

Ion Grama, Jean-Marie Tricot, Jean-François Petiot

BULETINUL ACADEMIEI DE ŞTIINŢE
A REPUBLICII MOLDOVA. MATEMATICA
Number 1(74), 2014, Pages 33–62
ISSN 1024–7696

Estimation of the extreme survival probabilities from censored data

Ion Grama, Jean-Marie Tricot and Jean-François Petiot

Abstract. The Kaplan-Meier nonparametric estimator has become a standard tool for estimating a survival time distribution in a right censoring schema. However, if the censoring rate is high, this estimator does not provide a reliable estimation of the extreme survival probabilities. In this paper we propose to combine the nonparametric Kaplan-Meier estimator and a parametric-based model into one construction. The idea is to fit the tail of the survival function with a parametric model while for the remaining to use the Kaplan-Meier estimator. A procedure for the automatic choice of the location of the tail based on a goodness-of-fit test is proposed. This technique allows us to improve the estimation of the survival probabilities in the mid and long term. We perform numerical simulations which confirm the advantage of the proposed method.

Mathematics subject classification: 62N01, 62N02, 62G32.
Keywords and phrases: Adaptive estimation, censored data, model selection, prediction, survival analysis, survival probabilities.

1 Introduction

Let $(X_i, C_i, Z_i)'$, $i = 1, ..., n$ be i.i.d. replicates of the vector $(X, C, Z)'$, where X and C are the survival and right censoring times and Z is a categorical covariate. It is supposed that X_i and C_i are conditionally independent given Z_i, $i = 1, ..., n$. We observe the sample $(T_i, \Delta_i, Z_i)'$, $i = 1, ..., n$, where $T_i = \min\{X_i, C_i\}$ is the observation time and $\Delta_i = 1_{\{X_i \leq C_i\}}$ is the failure indicator. Let $F(x|z)$, $x \geq x_0 \geq 0$ and $F_C(x|z)$, $x \geq x_0$ be the conditional distributions of X and C, given $Z = z$, respectively. In this paper we address the problem of estimation of the survival function $S_F(x|z) = 1 - F(x|z)$ when $x \geq x_0$ is large. The function S_F is traditionally estimated using the Kaplan-Meier nonparametric estimator (Kaplan and Meier [14]). Its properties have been extensively studied by numerous authors, including Fleming and Harrington [7], Andersen, Borgan, Gill and Keiding [2], Kalbfleisch and Prentice [13], Klein and Moeschberger [16]. However, in various practical applications, when the time x is close or exceeds the largest observed data, the predictions based on the Kaplan-Meier and related estimators are rather uninformative.

For illustration purposes we consider the well known PBC (primary biliary cirrhosis) data from a clinical trial analyzed in Fleming and Harrington [7]. In this trial one observes the censored survival times of two groups of patients: the first one ($Z = 1$) was given the DPCA (D-penicillamine drug) treatment and the second

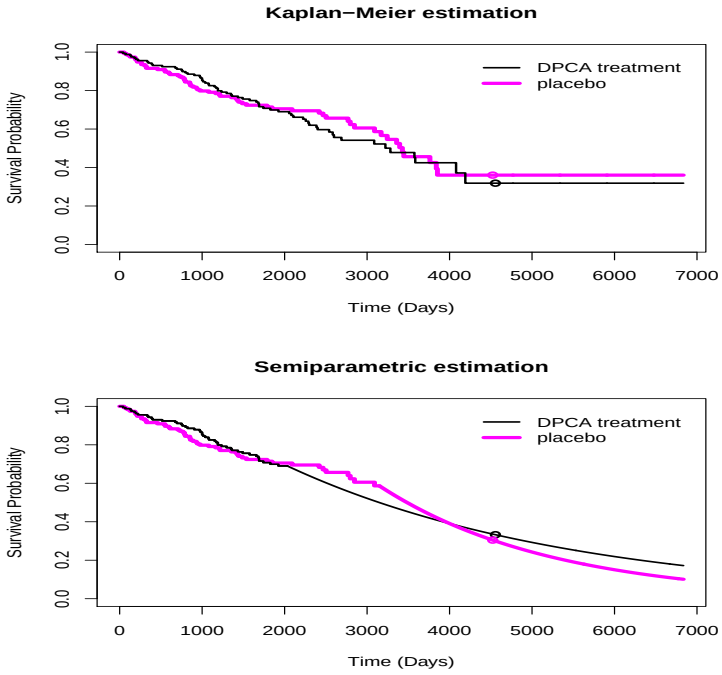

Figure 1. We compare two types of prediction of the survival probabilities in DPCA and placebo groups: on the top picture the prediction is based on the Kaplan-Meier estimation and on the bottom picture the prediction uses a semiparametric approach. The points on the curves correspond to the largest observation time in each group.

one is the control group ($Z = 0$). The overall censoring rate is about 60%. Here we consider only the group covariate and we are interested to compare the extreme survival probabilities of the patients under study in the two groups. In Figure 1 (top picture) we display the Kaplan-Meier nonparametric curves of the treatment and the control (placebo) groups. From these curves it seems difficult to infer whether the DPCA treatment has an effect on the survival probability. For instance at time $x = 4745$ (13 years) using the Kaplan-Meier nonparametric estimator (KM), one gets an estimated survival probability $\widehat{S}_{KM}(x|z = 0) = 0.3604$ for the control group and $\widehat{S}_{KM}(x|z = 1) = 0.3186$ for the DPCA treatment group. In this example and in many other applications one has to face the following two drawbacks. First, the estimated survival probabilities $\widehat{S}_{KM}(x|z)$ are constant for x beyond the largest (non-censored) survival time, which is not quite helpful for prediction purposes. Second, for this particular data set, the Kaplan-Meier estimation suggests that the DPCA treatment group has an estimated long term survival probability slightly lower than that of the control group, which can be explained by the high variability of $\widehat{S}_{KM}(x|z)$ for large x. These two points clearly rise the problem of correcting the behavior of the tail of the Kaplan-Meier estimator.

A largely accepted way to estimate the survival probabilities $S_F(x|z)$ for large x, is the parametric-based model fitting the hole data starting from the origin. Its

advantages are pointed out in Miller [18], however, it is well known that the bias model can be high if it is misspecified. The more flexible nonparametric Kaplan-Meier estimator would generally be preferred for estimating certain functionals of the survival curve, as argued in Meier, Karrison, Chappell and Xie [17]. In this paper we propose to combine the nonparametric Kaplan-Meier estimator and the parametric-based model into one construction which we call semiparametric Kaplan-Meier estimator (SKM). Our new estimator incorporates a threshold t in such a way that $S_F(x|z)$ is estimated by the Kaplan-Meier estimator for $x \leq t$ and by a parametric-based estimate for $x > t$. The main theoretical contribution of the paper is to show that with an appropriate choice of the threshold t such an estimate is consistent if the tail is correctly specified. In the case when the tail is misspecified we show by simulations that the method is robust. Denote by $\widehat{S}_t$ the resulting estimator of S_F, where the parametric-based model is the exponential distribution with mean θ. By simulations we have found that $\widehat{S}_{\widehat{t}}$, endowed with a data driven threshold $\widehat{t}$, outperforms the Kaplan-Meier estimator. As it is seen from Figure 1 (bottom picture), we obtain at $x = 4745$ the estimated survival probability $\widehat{S}_{\widehat{t}_0}(x|z=0) = 0.2739$ for the control group and $\widehat{S}_{\widehat{t}_1}(x|z=1) = 0.3150$ for the DPCA treatment group, where $\widehat{t}_0$ and $\widehat{t}_1$ are the corresponding data driven thresholds. Our predictions are recorded in Table 2 and seem to be more adequate than those based on the Kaplan-Meier estimation. We refer to Section 7, where this example is described in more details.

In Figures 2 we display the root of the mean squared error of the predictions of $S_F(x|z)$ based on the Kaplan-Meier and the proposed semiparametric Kaplan-Meier estimators as functions of the observation time x. This is an example where the exponential model for survival and censoring tails are misspecified. The errors are computed within a Monte-Carlo simulation study of size $M = 2000$ with a gamma distribution modeling the survival and censoring times which do not exhibit exponential behavior in the tail (see Section 6 and Example 2 of Section 2 for details). The advantage of the proposed semiparametric estimator over the Kaplan-Meier estimator can be clearly seen by comparing the two MSE curves. The MSE of the semiparametric estimator is much smaller than that of the Kaplan-Meier estimator for large observation times $x > q_{0.99}$ but also for mid range observation time values, for example $x \in [8, q_{0.99}]$, where $q_{0.99}$ is the 0.99-quantile of the distribution F. The proposed extensions of the nonparametric curves are particularly suited for predicting the survival probabilities in the case when the proportion of the censored times is large. This is the case of the mentioned simulated data where the mean censoring rate is about 77%. Note also that we get an improvement over the Kaplan-Meier estimator even for very low sample sizes like $n = 20$.

The proposed estimator $\widehat{S}_t$ is sensible to the choice of the threshold t. The main difficulty is to choose t small enough, so that the parametric-based part contains enough observation times to ensure a reliable prediction in the tail. At the same time one should choose t large enough in order to prevent from a large bias due to an inadequate tail fitting. The very important problem of the automatic choice of the

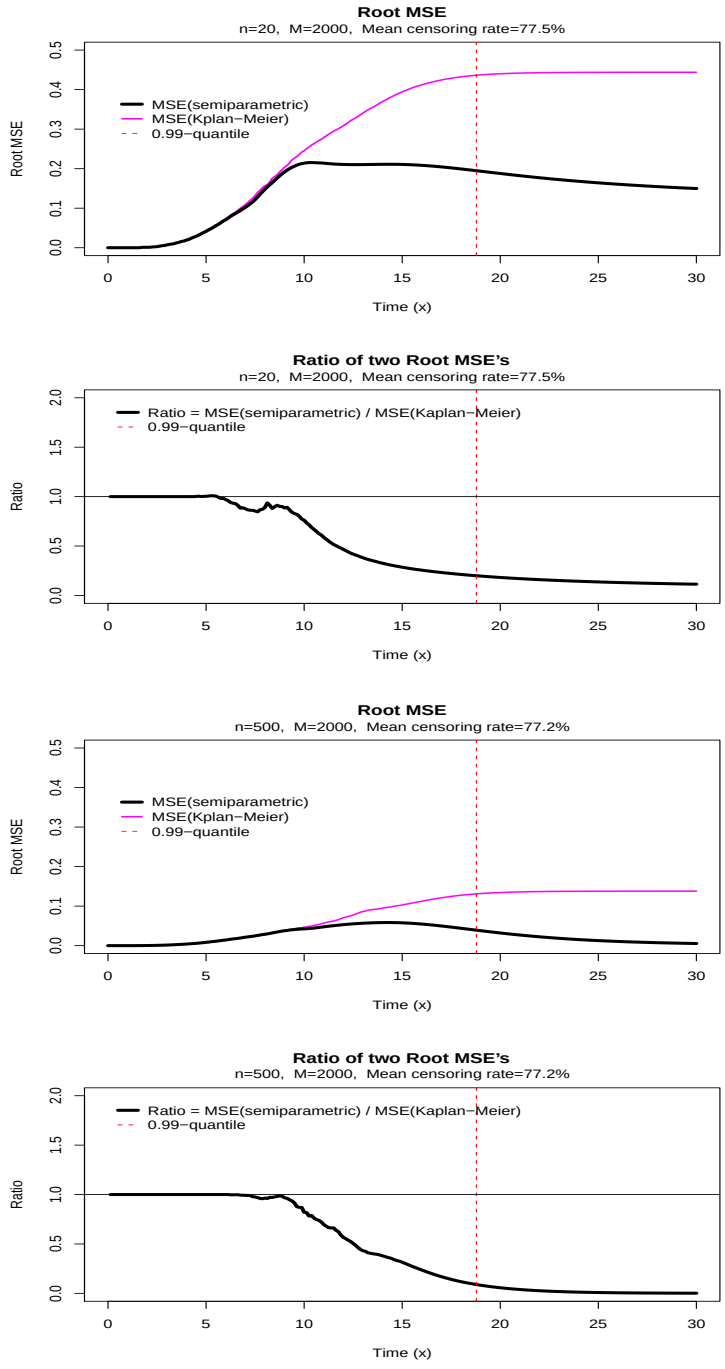

Figure 2. The lines 1, 3 (from the top) display simulated root MSE's of the Kaplan-Meier and semiparametric Kaplan-Meier estimators as functions of the time x. On the lines 2, 4 we show the ratio of the two root MSE's displayed on the lines 1, 3. The vertical dashed line is the 0.99 quantile of the true distribution of the survival time. The sample sizes are either $n = 20$ or $n = 500$.

threshold $\widehat{t}$ is treated in Section 5, where a procedure which we call *testing-pursuit-selection* is performed in two stages: First we test sequentially the null hypothesis that the proposed parametric-based model fits the data until we detect a chosen alternative. Secondly we select the best model among the accepted ones by penalized model selection. Therefore our testing-pursuit-selection procedure is actually also a goodness-of-fit test for the proposed parametric-based model. The resulting data driven estimator of the tail depends heavily on the testing procedure.

The approach developed here can be applied in conjunction with other techniques of prediction such as accelerated life testing, see Wei [22], Tseng, Hsieh and Wang [21], Escobar and Meeker [6] and extreme values estimation, see Hall [10], Hall and Welsh [11, 12], Dress [5], Grama and Spokoiny [8]. We refer also to Grama, Tricot and Petiot [9] for a related result concerning the approximation of the tail by the Cox model [4].

The case of continuous multivariate covariate Z in the context of a Cox model and the use of fitted tails other than the exponential can be treated by similar methods. The models which take into account the cure effects can be reduced to ours after removing the cure fraction. However, these problems are beyond the scope of the paper.

The paper is organized as follows. In Section 2 we introduce the main notations and give the necessary background. The main results of the paper about the consistency of the proposed estimators are stated in Sections 3 and 4. The automatic threshold selection procedure is described in Section 5. In Section 6 we give some simulation results and analyze the performance of the studied estimators. An application to real data is done in Section 7 and a conclusion in Section 8.

2 The model and background definitions

Assume that the survival and right censoring times arise from variables X and C which take their values in $[x_0, \infty)$, where $x_0 \geq 0$. Consider that X and C may depend on the categorical covariate Z with values in the set $\mathcal{Z} = \{0, ..., m\}$. The related conditional distributions $F(x|z)$ and $F_C(x|z)$, $x \geq x_0$, given $Z = z$, are supposed to belong to the set $\mathcal{F}$ of distributions with strictly positive density on $[x_0, \infty)$. Let $f_F(\cdot|z)$ and $S_F(\cdot|z) = 1 - F(\cdot|z)$ be the conditional density and survival functions of X, given $Z = z$. The corresponding conditional hazard function is $h_F(\cdot|z) = f_F(\cdot|z) / S_F(\cdot|z)$, given $Z = z$. Similarly, C has the conditional density $f_C(\cdot|z)$, survival function $S_C(\cdot|z)$ and hazard function $h_C(\cdot|z) = f_C(\cdot|z) / S_C(\cdot|z)$, given $Z = z$. We also assume the independence between X and C, conditionally with respect to Z. Let the observation time and the failure indicator be

$$T = \min\{X, C\} \quad \text{and} \quad \Delta = 1_{\{X \leq C\}},$$

where 1_B is the indicator function taking the value 1 on the event B and 0 otherwise. Let $P_{F,F_C}(dx, d\delta|z)$, $x \in [x_0, \infty)$, $\delta \in \{0, 1\}$ be the conditional distribution of the vector $\mathbf{Y} = (T, \Delta)'$, given $Z = z$. The density of P_{F,F_C} is

$$p_{F,F_C}(x, \delta|z) = f_F(x|z)^{\delta} S_F(x|z)^{1-\delta} f_C(x|z)^{1-\delta} S_C(x|z)^{\delta}, \tag{2.1}$$

where $x \in [x_0, \infty)$, $\delta \in \{0, 1\}$.

Let $z_i \in \mathcal{Z}$ be the observed value of the covariate Z_i, where Z_i, $i = 1, ..., n$ are i.i.d. copies of Z, and let $\mathbf{Y}_i = (T_i, \Delta_i)'$, $i = 1, ..., n$ be a sample of n vectors, where each vector $\mathbf{Y}_i$ has the conditional distribution $P_{F,F_C}(\cdot|z_i)$, given $Z_i = z_i$, for $i = 1, ..., n$. It is clear that, given $Z = z \in \mathcal{Z}$, the vectors $\mathbf{Y}_i$, $i \in \{j : z_j = z\}$ are i.i.d. .

In this paper the problem is to improve the nonparametric Kaplan-Meier estimators of the $m + 1$ survival probabilities $S_F(x|z) = 1 - F(x|z)$, $z \in \mathcal{Z}$, for large values of x. To this end, we fit the tail of $F(\cdot|z)$ by the exponential distribution with mean $\theta > 0$. Consider the following conditional semiparametric quasi-model

$$F_{\theta,t}(x|z) = \begin{cases} F(x|z), & x \in [x_0, t], \\ 1 - (1 - F(t|z)) \exp\left(-\frac{x-t}{\theta}\right), & x > t, \end{cases} \tag{2.2}$$

where $t \geq x_0$ is a nuisance parameter and $F(\cdot|z) \in \mathcal{F}$, $z \in \mathcal{Z}$ are functional parameters. The conditional density, survival and hazard functions of $F_{\theta,t}$ are denoted by $f_{F_{\theta,t}}$, $S_{F_{\theta,t}}$ and $h_{F_{\theta,t}}$, respectively. Note that $h_{F_{\theta,t}}(x|z) = 1/\theta$, for $x > t$.

The χ^2 entropy between two equivalent probability measures P and P_0 is defined by $\chi^2(P, P_0) = \int dP/dP_0 dP - 1$. By Jensen's inequality $\chi^2(P, P_0) \geq 0$.

Definition 2.1. *Let $F, F_C \in \mathcal{F}$ and $z \in \mathcal{Z}$. The tail of the distribution $F(\cdot|z)$ belongs to the domain of attraction of the exponential model under the right censoring schema if there exists a constant $\theta_z > 0$ such that*

$$\lim_{t \to \infty} \chi^2\left(P_{F,F_C}(\cdot|z), P_{F_{\theta_z,t},F_C}(\cdot|z)\right) = 0. \tag{2.3}$$

Below we give two examples when (2.3) is verified.

Example 1 (asymptotically constant hazards). Consider asymptotically constant survival and censoring hazard functions. This model can be related to the families of distributions in Hall [10], Hall and Welsh [11], Dress [5] and Grama and Spokoiny [8] for the extreme value models. Let $A > 0$, $\theta_{\max} > \theta_{\min} > 0$ be some constants. Consider that the survival time X has a hazard function $h_F(\cdot|z)$ such that for some $\theta_z \in (\theta_{\min}, \theta_{\max})$ and $\alpha_z > 0$,

$$|\theta_z h_F(\theta_z x|z) - 1| \leq A \exp(-\alpha_z x), \quad x \geq x_0. \tag{2.4}$$

Condition (2.4) means that $h_F(x|z)$ converges to θ_z^{-1} exponentially fast as $x \to \infty$. Substituting $\alpha_z = \alpha_z' \theta_z$, (2.4) gives $\left|h_F(x|z) - \theta_z^{-1}\right| \leq A' \exp(-\alpha_z' x)$, where $A' = A/\theta_{\min}$.

Similarly, let $M > 0$, $\gamma_{\max} > \gamma_{\min} > 0$, $\mu > 1$ be some constants. Assume that the hazard function $h_C(\cdot|z)$ of the censoring time C satisfies for some $\gamma_z \in (\gamma_{\min}, \gamma_{\max})$,

$$|\theta_z h_C(\theta_z x|z) - \gamma_z| \leq M(1 + x)^{-\mu}, \quad x \geq x_0. \tag{2.5}$$

Condition (2.5) is equivalent to saying that $h_C(x|z)$ approaches γ_z/θ_z polynomially fast as $x \to \infty$. Substituting $\gamma_z = \gamma_z' \theta_z$, (2.5) gives $|h_C(x|z) - \gamma_z'| \leq M' x^{-\mu}$, where $M' = M\theta_{\max}^{\mu}/\theta_{\min}$.

For example, conditions (2.4) and (2.5) are satisfied if F and F_C coincide with the re-scaled Cauchy distribution $K_{\mu,\theta}$ defined below. Let ξ be a variable with the positive Cauchy distribution $K(x) = 2\pi^{-1}\arctan(x)$, $x \geq 0$. We define the re-scaled Cauchy distribution by $K_{\mu,\theta}(x) = 1 - \frac{1-K(\exp((x-\mu)/\theta))}{1-K(\exp(-\mu/\theta))}$, where μ and θ are the location and scale parameters. The distribution $K_{\mu,\theta}$ can be seen as the excess distribution of the variable $\theta \log \xi + \mu$ over the threshold 0. The plots of the density $f_{K_{\mu,\theta}}$ related to $K_{\mu,\theta}$ for various values of parameters are given in Figure 4 (lines 1, 3). We leave to the reader the verification that $K_{\mu,\theta}$ fulfills (2.4) with $\theta_z = \theta$, $\alpha_z = 2$ and (2.5) with $\gamma_z = 1$. The distribution $K_{\mu,\theta}$ will be used in Section 6 to simulate survival and censoring times.

Example 2 (non-constant hazards). Now we consider the case when the hazard functions are not asymptotically constant. For instance, this is the case when the survival and censoring times have both gamma distributions. The numerical results presented in Figure 2 and Table 1 and discussed in Section 6 show that the approach of the paper works when conditions (2.4) and (2.5) are not satisfied.

The heuristic argument behind these experimental findings is as follows. Denote by $Q^{(t)}(x) = P(\xi \leq t + x | \xi \geq t)$, $x \geq 0$, the excess distribution of ξ over the threshold t, where ξ is a random variable with distribution Q. Let G_θ be the exponential distribution with mean θ. Obviously $G_\theta^{(t)} = G_\theta$. By simple re-normalization, the χ^2 entropy in (2.3) can be rewritten as follows:

$$\begin{aligned}\chi^2\left(P_{F,F_C}(\cdot|z), P_{F_{\theta_z,t},F_C}(\cdot|z)\right) &= S_F(t|z)\, S_C(t|z) \times \\ &\quad \chi^2(P_{F^{(t)},F_C^{(t)}}(\cdot|z), P_{G_{\theta_z},F_C^{(t)}}(\cdot|z)).\end{aligned} \tag{2.6}$$

Clearly from (2.6), Definition 2.1 is fulfilled if, as $t \to \infty$,

$$\chi^2(P_{F^{(t)},F_C^{(t)}}(\cdot|z), P_{G_{\theta_z},F_C^{(t)}}(\cdot|z)) \to 0, \tag{2.7}$$

which means that beyond the threshold t, the excess distribution $F^{(t)}(\cdot|z)$ is "well" approximated by an exponential distribution with parameter θ_z, for some $t > 0$. However (2.3) can be satisfied even if (2.7) may not hold, more precisely when

$$\chi^2(P_{F^{(t)},F_C^{(t)}}(\cdot|z), P_{G_{\theta_z},F_C^{(t)}}(\cdot|z)) = o\left(\frac{1}{S_F(t|z)}\right), \tag{2.8}$$

where $S_F(t|z) \to 0$ as $t \to \infty$. This means that the tail probabilities can be estimated by our approach even if the exponential model is misspecified for the tail.

3 Consistency of the estimator with fixed threshold

Define the quasi-log-likelihood by $\mathcal{L}_t(\theta|z) = \sum_{i=1}^n \log p_{F_{\theta,t},F_C}(T_i, \Delta_i|z_i)\, 1_{\{z_i=z\}}$, where $F_{\theta,t}$ is defined by (2.2) with parameters $\theta > 0$, $t \geq x_0$ and $F(\cdot|z) \in \mathcal{F}$, $z \in \mathcal{Z}$. Taking into account (2.1) and dropping the terms related to the censoring, the partial quasi-log-likelihood is

$$\mathcal{L}_t^{\text{part}}(\theta|z) = \sum_{T_i \leq t,\ z_i=z} \Delta_i \log h_{F_{\theta,t}}(T_i|z) - \sum_{T_i > t,\ z_i=z} \Delta_i \log \theta \tag{3.1}$$

$$-\sum_{T_i\leq t,\ z_i=z}\int_{x_0}^{T_i} h_{F_{\theta,t}}(v|z)\,dv-\sum_{T_i>t,\ z_i=z}\left(\int_{x_0}^{t} h_{F_{\theta,t}}(v)\,dv+\theta^{-1}(T_i-t)\right),$$

for fixed $z\in\mathcal{Z}$ and $t\geq x_0$. Maximizing $\mathcal{L}_t^{\text{part}}(\theta|z)$ in θ, obviously yields the estimator

$$\widehat{\theta}_{z,t}=\frac{\sum_{T_i>t,\ z_i=z}(T_i-t)}{\widehat{n}_{z,t}}, \tag{3.2}$$

where by convention $0/0=\infty$ and $\widehat{n}_{z,t}=\sum_{T_i>t,\ z_i=z}\Delta_i$ is the number of observed survival times beyond the threshold t.

The estimator of $S_F(x)$, for $x_0\leq x\leq t$, is easily obtained by standard nonparametric maximum likelihood approach due to Kiefer and Wolfowiz [15] (see also Bickel, Klaassen, Ritov and Wellner [3], Section 7.5). We use the product Kaplan-Meier (KM) estimator (with ties) defined by

$$\widehat{S}_{KM}(x|z)=\prod_{T_i\leq x}\left(1-d_i(z)/r_i(z)\right),\ x\geq x_0,$$

where $r_i(z)=\sum_{j=1}^{n}1_{\{T_j\geq T_i,\ z_j=z\}}$ is the number of individuals at risk at T_i and $d_i(z)=\sum_{j=1}^{n}1_{\{T_j=T_i,\Delta_j=1,\ z_j=z\}}$ is the number of individuals died at T_i (see Klein and Moeschberger [16], Section 4.2 and Kalbfleisch and Prentice [13]). The *semi-parametric fixed-threshold Kaplan-Meier estimator* (SFKM) of the survival function takes the form

$$\widehat{S}_t(x|z)=\begin{cases}\widehat{S}_{KM}(x|z), & x\in[x_0,t],\\ \widehat{S}_{KM}(t|z)\exp\left(-\frac{x-t}{\widehat{\theta}_{z,t}}\right), & x>t,\end{cases} \tag{3.3}$$

where $\exp\left(-(x-t)/\widehat{\theta}_{z,t}\right)=1$ if $\widehat{\theta}_{z,t}=\infty$. Similarly, it is possible to use the Nelson-Aalen nonparametric estimator (Nelson [19, 20], Aalen [1]) instead of the Kaplan-Meier one.

Consider the Kullback-Leibler divergence $\mathcal{K}(\theta',\theta)=\int\log(dG_{\theta'}/dG_\theta)\,dG_{\theta'}$ between two exponential distributions with means θ' and θ. By convention, $\mathcal{K}(\infty,\theta)=\infty$. It is easy to see that $\mathcal{K}(\theta',\theta)=\psi(\theta'/\theta-1)$, with $\psi(x)=x-\log(x+1)$, $x>-1$ and that there are two constants c_1 and c_2 such that $(\theta'/\theta-1)^2\leq c_1\mathcal{K}(\theta',\theta)\leq c_2(\theta'/\theta-1)^2$, when $|\theta'/\theta-1|$ is small enough.

The following theorem provides a rate of convergence of the estimator $\widehat{\theta}_{z,t}$ as function of the χ^2-entropy between P_{F,F_C} and $P_{F_{\theta,t},F_C}$. Let $\mathbb{P}$ be the joint distribution of the sample $\mathbf{Y}_i$, $i=1,...,n$ and $\mathbb{E}$ be the expectation with respect to $\mathbb{P}$. In the sequel, the notation $\alpha_n=O_{\mathbb{P}}(\beta_n)$ means that there is a positive constant c such that $\mathbb{P}(\alpha_n>c\beta_n,\beta_n<\infty)\to 0$ as $n\to\infty$, for any two sequences of positive possibly infinite variables α_n and β_n.

Theorem 3.1. *Let $z\in\mathcal{Z}$. For any $\theta_z>0$ (possibly depending on z) and $t\geq x_0$, it holds*

$$\mathcal{K}\left(\widehat{\theta}_{z,t},\theta_z\right)=O_{\mathbb{P}}\left(\frac{n}{\widehat{n}_{z,t}}\chi^2\left(P_{F,F_C}(\cdot|z),P_{F_{\theta_z,t},F_C}(\cdot|z)\right)+\frac{4\log n}{\widehat{n}_{z,t}}\right). \tag{3.4}$$

For any $z \in \mathcal{Z}$ and $\theta_z > 0$ the optimal rate of convergence is obtained when the terms in the right hand side of (3.4) are balanced, i.e. when $t = t_{z,n}$ is chosen such that

$$\chi^2\left(P_{F,F_C}(\cdot|z), P_{F_{\theta_z,t_{z,n}},F_C}(\cdot|z)\right) = O\left(\frac{\log n}{n}\right) \quad \text{as} \quad n \to \infty, \tag{3.5}$$

where $t_{z,n}$ may depend on z. It is easy to verify that, if the tail of the distribution $F(\cdot|z)$ belongs to the domain of attraction of the exponential model under the right censoring schema, a sequence $t_{z,n} \geq x_0$ satisfying (3.5) always exists.

From Theorem 3.1 we deduce the following:

Theorem 3.2. *Let $z \in \mathcal{Z}$. Assume that the distribution $F(\cdot|z)$ belongs to the domain of attraction of the exponential model under the right censoring schema and $t_{z,n}$ is a sequence satisfying (3.5). Then*

$$\mathcal{K}\left(\widehat{\theta}_{z,t_{z,n}}, \theta_z\right) = O_{\mathbb{P}}\left(\frac{\log n}{\widehat{n}_{z,t_{z,n}}}\right). \tag{3.6}$$

Using the two sided bound for the Kullback-leibler entropy between exponential laws stated before, from Theorem 3.2 we conclude that $\widehat{\theta}_{z,t_{z,n}}$ converges to θ_z at the usual $\left(\widehat{n}_{z,t_{z,n}}\right)^{-1/2}$ rate up to a $\log n$ factor: $\left(\widehat{\theta}_{z,t_{z,n}} - \theta_z\right)^2 = O_{\mathbb{P}}\left(\frac{\log n}{\widehat{n}_{z,t_{z,n}}}\right)$, provided that there are two constants $\theta_{\min}$ and $\theta_{\max}$ such that $0 < \theta_{\min} \leq \theta_z \leq \theta_{\max} < \infty$.

Furthermore, the rate of convergence of the estimator $\widehat{\theta}_{z,t_{z,n}}$ can be expressed in terms of $S_F(\cdot|z)$, $S_C(\cdot|z)$ and the sample size n, by giving a lower bound for $\widehat{n}_{z,t_{z,n}}$. To ensure such a bound we have to introduce two additional assumptions.

The first assumption involves the *conditional censoring rate function*

$$q_{F,F_C}(t|z) = \int_t^\infty S_{F,t}(x|z)\, f_{C,t}(x|z)\, dx \leq 1,\ t \geq x_0,\ z \in \mathcal{Z}, \tag{3.7}$$

where $S_{F,t}(x|z) = S_F(x|z)/S_F(t|z)$, $x \geq t$ is the conditional survival function related to the survival time X, given $X > t$, and $f_{C,t}(x|z) = f_C(x|z)/S_C(t|z)$, $x \geq t$ is the conditional density function related to the censoring time C, given $C > t$. The quantity $q_{F,F_C}(t|z)$ controls the proportion of the censored times among the observation times exceeding t. In particular if $t = x_0$, then $q_{F,F_C}(x_0|z) = \text{Prob}(X > C|z)$ is simply the mean censoring rate (given $Z = z$).

We assume that the conditional censoring rate function $q_{F,F_C}(\cdot|z)$ is separated from 1, i.e. that there are constants $r_0 \geq x_0$ and $q_0 < 1$, such that, for any $z \in \mathcal{Z}$ and any $t \geq r_0$,

$$q_{F,F_C}(t|z) \leq q_0. \tag{3.8}$$

Assumption (3.8) is verified, for instance, if $F(\cdot|z)$ and $F_C(\cdot|z)$ are exponential with intensities λ_X and λ_C respectively: in this case $q_{F,F_C}(t|z) = \lambda_C/(\lambda_C + \lambda_X)$, $t \geq 0$. It is also verified if distributions F and F_C meet (2.4) and (2.5). The trajectory of $q_{F,F_C}(\cdot|z)$ with F and F_C satisfying the two last conditions is plotted in Figure 4 (lines 2, 4).

The second assumption involves the number of individuals with profile $z \in \mathcal{Z}$: $n_z = \sum_{i=1}^{n} 1(z_i = z)$. We assume that there is a constant $\kappa \in (0,1]$ such that, for any $z \in \mathcal{Z}$,

$$n_z \geq \kappa\, n. \tag{3.9}$$

Lemma 3.3. *Assume that conditions (3.8) and (3.9) are satisfied. Then for every $t \geq r_0$, it holds $\mathbb{E}\widehat{n}_{z,t} \geq \kappa\ n(1-q_0) S_C(t|z) S_F(t|z)$ and $\mathbb{P}(\widehat{n}_{z,t} < \mathbb{E}\widehat{n}_{z,t}/2) \leq \exp(-\mathbb{E}\widehat{n}_{z,t}/8)$. Moreover, if the sequence $t_{z,n}$ is such that $\mathbb{E}\widehat{n}_{z,t_{z,n}} \to \infty$ as $n \to \infty$, then it holds $\mathbb{P}\left(\widehat{n}_{z,t_{z,n}} \geq \mathbb{E}\widehat{n}_{z,t_{z,n}}/2\right) \to 1$ as $n \to \infty$.*

As a simple consequence of Theorem 3.2 and Lemma 3.3 we have:

Theorem 3.4. *Assume conditions (3.8) and (3.9). Assume that the distribution $F(\cdot|z)$ belongs to the domain of attraction of the exponential model under the right censoring schema, $t_{z,n}$ is a sequence satisfying (3.5) and*

$$nS_C(t_{z,n}|z)\, S_F(t_{z,n}|z) \to \infty \text{ as } n \to \infty. \tag{3.10}$$

Then

$$\mathcal{K}\left(\widehat{\theta}_{z,t_{z,n}}, \theta_z\right) = O_{\mathbb{P}}\left(\frac{\log n}{nS_C(t_{z,n}|z)\, S_F(t_{z,n}|z)}\right).$$

4 Explicit computation of the rate of convergence

The results of the previous section show that the rate of convergence of the estimator $\widehat{\theta}_{z,t_{z,n}}$ depends on the survival functions $S_F(\cdot|z)$ and $S_C(\cdot|z)$ and on the sequences $t_{z,n}$. In order to derive a rate of convergence expressed only in terms of the sample size n we have to make additional assumptions on F and F_C. Moreover, we find minimal (up to one term expansion) threshold $t_{z,n}$ for which (3.5) holds true.

Our first result concerns the case when $h_C(\cdot|z)$ is separated from 0.

Theorem 4.1. *Assume conditions (3.8) and (3.9). Assume that $h_F(\cdot|z)$ satisfies (2.4), that there are positive constants $t_{\min}$ and $c_{\min}$ such that $h_C(x|z) \geq c_{\min}$ for any $x \geq t_{\min}$ and that*

$$S_C(t_{z,n}|z)\, n^{\frac{2\alpha_z}{1+2\alpha_z}} \log^{\frac{1}{1+2\alpha_z}} n \to \infty \text{ as } n \to \infty. \tag{4.1}$$

Then,

$$\mathcal{K}\left(\widehat{\theta}_{z,t_{z,n}}, \theta_z\right) = O_{\mathbb{P}}\left(\frac{\left(n^{-1}\log n\right)^{\frac{2\alpha_z}{1+2\alpha_z}}}{S_C(t_{z,n}|z)}\right), \tag{4.2}$$

where

$$t_{z,n} = \frac{\theta_z}{1+2\alpha_z}\log n + o(\log n).$$

Assume additionally that $S_C(t_{z,n}|z) \geq c_0 > 0$, which means that with positive probability there are large censoring times. Then the rate of convergence in (4.2) becomes $\left(n^{-1}\log n\right)^{\frac{2\alpha_z}{1+2\alpha_z}}$ for any $z \in \mathcal{Z}$.

Under the additional condition that $h_C(\cdot|z)$ satisfies (2.5) we have the following result:

Theorem 4.2. *Assume condition (3.9). Assume that $h_F(\cdot|z)$ satisfies (2.4) and $h_C(\cdot|z)$ satisfies (2.5). Then,*

$$\mathcal{K}\left(\widehat{\theta}_{z,t_{z,n}}, \theta_z\right) = O_{\mathbb{P}}\left(\left(\frac{\log n}{n}\right)^{\frac{2\alpha_z}{1+\gamma_z+2\alpha_z}}\right), \tag{4.3}$$

where

$$t_{z,n} = \frac{\theta_z}{1+\gamma_z+2\alpha_z}\log n + o(\log n).$$

We give some hints about the optimality of the rate in (4.3). Assume that the survival time X is exponential, i.e. $h_F(x|z) = \theta_z^{-1}$ for all $x \geq x_0$ and $z \in \mathcal{Z}$. This ensures that condition (2.4) is satisfied with any $\alpha > 0$. Assume conditions (2.5) and (3.9). If there are two constants $\theta_{\min}$ and $\theta_{\max}$ such that $0 < \theta_{\min} \leq \theta_z \leq \theta_{\max} < \infty$, (4.3) implies $\left|\widehat{\theta}_{z,t_{z,n}} - \theta_z\right| = O_{\mathbb{P}}\left(\left(n^{-1}\log n\right)^{\frac{\alpha}{1+\gamma_z+2\alpha}}\right)$, for any $\alpha > 0$. This rate becomes arbitrarily close to the $n^{-1/2}$ rate as $\alpha \to \infty$, since $\lim_{\alpha\to\infty} \alpha/(1+\gamma_z+2\alpha) \to 1/2$. Thus the estimator $\widehat{\theta}_{z,t_{z,n}}$ almost recovers the usual parametric rate of convergence as α becomes large whatever is $\gamma_z > 0$.

In the case when there are no censoring ($\gamma_z = 0$), after an exponential rescaling our problem can be reduced to that of the estimation of extreme index. If $\gamma_z \to 0$ our rate becomes close to $n^{-\frac{2\alpha_z}{1+2\alpha_z}}$, which is known to be optimal in the context of the extreme value estimation, see Dress [5] and Grama and Spokoiny [8]. So our result nearly recovers the best possible rate of convergence in this setting.

5 Testing and automatic selection of the threshold

In this section a procedure of selecting the adaptive estimator $\widehat{\theta}_z = \widehat{\theta}_{z,\widehat{t}_{z,n}}$ from the family of fixed threshold estimators $\widehat{\theta}_{z,t}$, $t \geq x_0$ is proposed. Here the adaptive threshold $\widehat{t}_{z,n}$ is obtained by a sequential testing procedure followed by a selection using a penalized maximum likelihood. This motivates our condensed terminology *testing-pursuit-selection* used in the sequel. The testing part is actually a multiple goodness-of-fit testing for the proposed parametric-based models, while the threshold $\widehat{t}_{z,n}$ can be seen as a data driven substitute for the theoretical threshold $t_{z,n}$ defined in Theorems 4.1 and 4.2 and in more general Theorems 3.2 and 3.4. For a discussion on the proposed approach we refer the reader to Section 3 of Grama and Spokoiny [8]. In the sequel, for simplicity of notations, we abbreviate $\widehat{t}_z = \widehat{t}_{z,n}$.

Define a semiparametric change-point distribution by

$$F_{\mu,s,\theta,t}(x|z) = \begin{cases} F(x|z), & x \in [x_0, s], \\ 1 - (1 - F(s|z)) \exp\left(-\frac{x-s}{\mu}\right), & x \in (s, t], \\ 1 - (1 - F(s|z)) \exp\left(-\frac{t-s}{\mu}\right) \exp\left(-\frac{x-t}{\theta}\right), & x > t, \end{cases}$$

for $\mu, \theta > 0$, $x_0 \leq s < t$ and $F(\cdot|z) \in \mathcal{F}$. As in Section 3 we find the maximum quasi-likelihood estimators $\widehat{\theta}_{z,t}$ of θ and $\widehat{\mu}_{z,s,t}$ of μ for fixed $z \in \mathcal{Z}$ and $x_0 \leq s < t$, which are given by (3.2) and

$$\widehat{\mu}_{z,s,t} = \frac{\widehat{n}_{z,s}\widehat{\theta}_{z,s} - \widehat{n}_{z,t}\widehat{\theta}_{z,t}}{\widehat{n}_{z,s,t}},$$

where $\widehat{n}_{z,s,t} = \sum_{s<T_i\leq t,\ z_i=z} \Delta_i$ and by convention $0 \cdot \infty = 0$ and $0/0 = \infty$.

Consider a constant $D > 0$, which will be the critical value in the testing procedure below. Let $k_0 \geq 3$ be a starting index and k_{step} be an increment for k. Let δ', δ'' be two positive constants such that $0 < \delta', \delta'' < 0.5$. The values k_0, k_{step}, δ', δ'' and D are the parameters of the procedure to be calibrated empirically. Without loss of generality, we consider that the T_i's are arranged in the decreasing order: $T_1 \geq ... \geq T_n$. The threshold t will be chosen in the set $\{T_1, ..., T_n\}$.

The *testing-pursuit-selection* procedure which we propose is performed in two stages. First we test the null hypothesis $\mathcal{H}_{T_k}(z) : F = F_{\theta,T_k}(\cdot|z)$ against the alternative $\widetilde{\mathcal{H}}_{T_k}(z) : F = F_{\mu,T_k,\theta,T_l}(\cdot|z)$ for some $\delta' k \leq l \leq (1-\delta'')k$, sequentially in $k = k_0 + ik_{step}$, $i = 0, ..., [n/k_{step}]$, until $\mathcal{H}_{T_k}(z)$ is rejected. Denote by $\widehat{k}_z$ the obtained break index and define the break time $\widehat{s}_z = T_{\widehat{k}_z}$. Second, using $\widehat{k}_z$ and $\widehat{s}_z$ define the adaptive threshold by $\widehat{t}_z = T_{\widehat{l}_z}$ with the adaptive index

$$\widehat{l}_z = \underset{\delta'\widehat{k}_z \leq l \leq (1-\delta'')\widehat{k}_z}{\operatorname{argmax}} \left\{ \mathcal{L}_{T_l}\left(\widehat{\theta}_{z,T_l}|z\right) - \mathcal{L}_{T_l}\left(\widehat{\theta}_{z,\widehat{s}_z}|z\right) \right\}, \tag{5.1}$$

where the term $\mathcal{L}_{T_l}\left(\widehat{\theta}_{z,\widehat{s}_z}|z\right)$ is a penalty for getting close to the break time $\widehat{s}_z$. The resulting adaptive estimator of θ_z is defined by $\widehat{\theta}_z = \widehat{\theta}_{z,\widehat{t}_z}$ and the *semiparametric adaptive-threshold Kaplan-Meier estimator* (SAKM) of the survival function is defined by $\widehat{S}_{\widehat{t}_z}(\cdot|z)$.

For testing $\mathcal{H}_{T_k}(z)$ against $\widetilde{\mathcal{H}}_{T_k}(z)$ we use the statistic

$$LR_{\max}(T_k|z) = \max_{\delta' k \leq l \leq (1-\delta'')k} LR(T_k, T_l|z), \tag{5.2}$$

where $LR(s, t|z)$ is the quasi-likelihood ratio test statistic for testing $\mathcal{H}_s(z) : F = F_{\theta,s}(\cdot|z)$ against the alternative $\widetilde{\mathcal{H}}_{s,t}(z) : F = F_{\mu,s,\theta,t}(\cdot|z)$. To compute (5.2), note that by simple calculations, using (3.1) and (3.2),

$$\mathcal{L}_t\left(\widehat{\theta}_{z,t}|z\right) - \mathcal{L}_t(\theta|z) = \widehat{n}_{z,t}\mathcal{K}\left(\widehat{\theta}_{z,t}, \theta\right), \tag{5.3}$$

where by convention $0 \cdot \infty = 0$. Similarly to (5.3), the quasi-likelihood ratio test statistic $LR(s,t|z)$ is given by

$$LR(s,t|z) = \widehat{n}_{z,s,t}\mathcal{K}\left(\widehat{\mu}_{z,s,t}, \widehat{\theta}_{z,s}\right) + \widehat{n}_{z,t}\mathcal{K}\left(\widehat{\theta}_{z,t}, \widehat{\theta}_{z,s}\right) \tag{5.4}$$

with the same convention. Note that, by (5.3), the second term in (5.4) can be viewed as the penalized quasi-log-likelihood

$$\begin{aligned} LR_{\text{pen}}(s,t|z) &= \mathcal{L}_t\left(\widehat{\theta}_{z,t}|z\right) - \mathcal{L}_t\left(\widehat{\theta}_{z,s}|z\right) \\ &= \widehat{n}_{z,t}\mathcal{K}\left(\widehat{\theta}_{z,t}, \widehat{\theta}_{z,s}\right). \end{aligned}$$

Our testing-pursuit-selection procedure reads as follows:

Step 1. Set the starting index $k = k_0$.

Step 2. Compute the test statistic for testing $\mathcal{H}_{T_k}(z)$ against $\widetilde{\mathcal{H}}_{T_k}(z)$:

$$LR_{\max}(T_k|z) = \max_{\delta' k \le l \le (1-\delta'')k} LR(T_k, T_l|z)$$

Step 3. If $k \le n - k_{step}$ and $LR_{\max}(T_k|z) \le D$, increase k by k_{step} and go to Step 2. If $k > n - k_{step}$ or $LR_{\max}(T_k|z) > D$, let $\widehat{k}_z = k$,

$$\widehat{l}_z = \underset{\delta'\widehat{k}_z \le l \le (1-\delta'')\widehat{k}_z}{\operatorname{argmax}} LR_{\text{pen}}\left(T_{\widehat{k}_z}, T_l|z\right),$$

take the adaptive threshold as $\widehat{t}_z = T_{\widehat{l}_z}$ and exit.

It may happen that with $k = k_0$ it holds $LR_{\max}(T_{k_0}|z) > D$, which means that the hypothesis that the tail is fitted by the exponential model, starting from T_{k_0}, is rejected. In this case we resume the procedure with a new augmented k_0, say with k_0 replaced by $[\nu_0 k_0]$, where $\nu_0 > 1$. Finally, if for each such k_0 it holds $LR_{\max}(T_{k_0}|z) > D$, we conclude that the tail of the model cannot be fitted with the proposed parametric tail and we estimate the tail by the Kaplan-Meier estimator. Therefore our testing-pursuit-procedure can be seen as well as a goodness-of-fit test for the tail.

Note that the Kullback-Leibler entropy $\mathcal{K}(\theta', \theta)$ is scale invariant, i.e. satisfies the identity $\mathcal{K}(\theta', \theta) = \mathcal{K}(\alpha\theta', \alpha\theta)$, for any $\alpha > 0$ and $\theta', \theta > 0$. Therefore the critical value D can be determined by Monte Carlo simulations from standard exponential observations. The choice of parameters of the proposed selection procedure is discussed in Section 6.

6 Simulation results

We illustrate the performance of the semiparametric estimator (3.3) with fixed and adaptive thresholds in a simulation study. The survival probabilities $S_F(x|z)$, for large values of x, are of interest.

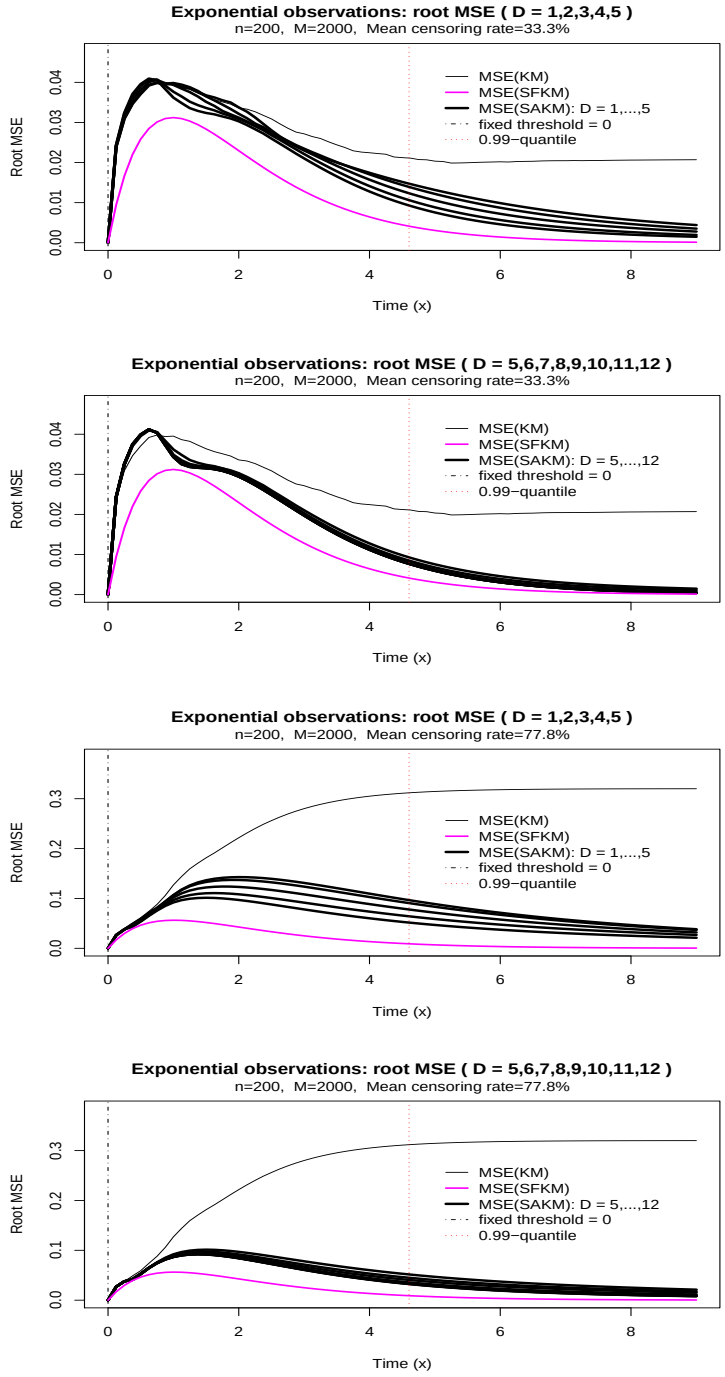

Figure 3. The lines 1, 3 (from the top) display the root type I MSE's of the Kaplan-Meier estimator (KM), of the semiparametric fixed-threshold Kaplan-Meier estimator (SFKM) with threshold fixed at 0 (which coincides with the exponential model) and of the semiparametric adaptive-threshold Kaplan-Meier estimator (SAKM) with $D = 1, 2, 3, 4, 5$. The lines 2, 4 display the same but with $D = 5, 6, 7, 8, 9, 10, 11, 12$. The mean censoring rate is either 33.3% or 77.8%.

The mean squared error (MSE) of an estimator $\widehat{S}(\cdot|z)$ of the true survival function $S_F(\cdot|z)$ is defined by $MSE_{\widehat{S}}(x|z) = \mathbb{E}\left(\widehat{S}(x|z) - S_F(x|z)\right)^2$. The quality of the estimator $\widehat{S}(\cdot|z)$ with respect to the Kaplan-Meier estimator $\widehat{S}_{KM}(\cdot|z)$ is measured by the ratio $R_{\widehat{S}}(x|z) = MSE_{\widehat{S}}(x|z) / MSE_{\widehat{S}_{KM}}(x|z)$.

Without loss of generality, we can assume that the covariate Z takes a fixed value z. In each study developed below, we perform $M = 2000$ Monte-Carlo simulations.

We start by giving some hints on the choice of the parameters $k_0, k_{step}, \delta', \delta''$ of the testing-pursuit-selection procedure in Section 5. The initial value k_0 controls the variability of the test statistic $LR_{\max}(T_k|z)$, $k \geq k_0$. We have fixed k_0 as a proportion of the initial sample size: $k_0 = n/10$. The choice $k_{step} = 5$ is made to speed up the computations. The parameters δ' and δ'' restrict the high variability of the test statistic $LR(T_k, T_l|z)$ when the change point $T_l \in [T_k, T_{k_0}]$ is close to the ends of the interval. The values $\delta' = 0.3$ and $\delta'' = 0.1$ are retained experimentally. Our simulations show that the adaptive procedure does not depend much on the choice of the parameters $k_0, k_{step}, \delta', \delta''$.

To choose the critical value D we analyze the type I MSE of the SAKM estimator, i.e. the MSE under the null hypothesis that the survival times $X_1, ..., X_n$ are i.i.d. standard exponential. We perform two simulations using i.i.d. exponential censoring times $C_1, ..., C_n$ with rates 0.5 and 3.5. The size is fixed at $n = 200$, but the results are quite similar for other sizes. The root MSE's as functions of the time x are given in Figure 3. For comparison, in Figure 3 we also included the MSE's corresponding to the parametric-based exponential modeling which coincides with the SFKM estimator having the threshold fixed at 0. Note that the MSE's calculated when the critical values are $D = 1, 2, 3, 4, 5$, decrease as D increases (see the lines 1, 3), while for $D = 5, 6, 7, 8, 9, 10, 11, 12$ the MSE's almost do not depend on D (see the lines 2, 4). The simulations show that the type I MSE decreases as D increases and stabilizes for $D \geq 5$. From these plots we conclude that the limits for the critical value D can be set between $D_0 = 5$ and $D_1 = 7$ without important loss in the type I MSE.

It is interesting to note that the adaptive threshold $\widehat{t}_z$ is relatively stable to changes of D. A typical trajectory of the test statistic $LR_{\max}(T_k|z)$ as function of T_k is drawn in Figure 7 (top). Despite the fact that the break time $\widehat{s}_z = T_{\widehat{k}_z}$ strongly depends on the critical value D (in this picture $D = 5.8$), we found that the adaptive threshold $\widehat{t}_z = T_{\widehat{l}_z}$, which maximizes the penalized quasi-log-likelihood $LR_{\text{pen}}\left(T_{\widehat{k}_z}, T_l|z\right)$ in Figure 7 (bottom), is stable to the local changes of the break time $\widehat{s}_z = T_{\widehat{k}_z}$ and thus is also quasi-stable to relatively small changes of D.

For our simulations we fix the value $D = 6$. Below we give some evidence that the SAKM estimator with this critical value has a reasonable type II MSE, under the hypothesis that the X_i's have a distribution F alternative to the standard exponential. Our simulations show that the type II MSE's are quite similar for several families we have tested. We have chosen the following two typical cases which are

representative for all these families.

Study case 1 (low tail censoring rate). We generate a sequence of $n = 200$ i.i.d. survival times $X_i,\ i = 1, ..., n$ from the re-scaled Cauchy distribution K_{μ_X,θ_X} with location parameter $\mu_X = 40$ and scale parameter $\theta_X = 5$ (see Section 2). The censoring times C_i, $i = 1, ..., n$ are i.i.d. from the re-scaled Cauchy distribution K_{μ_C,θ_C} with location parameter $\mu_C = \mu_X - 20 = 20$ and scale parameter $\theta_C = 2\theta_X = 10$. To give an overview of the variation of the censoring rate along the magnitude of X_i, we display the density functions of the survival and censoring times X_i and C_i in Figure 4. We also display the conditional censoring rate curve $q_{F,F_C}(t|z)$ as function of t. The (overall) mean censoring rate in this example corresponds to the starting point of the curve and is about 88% (horizontal dashed line in Figure 4, line 2). As $t \to \infty$ this curve decreases to the limit $\lim_{t\to\infty} q_{F,F_C}(t|z) = \theta_X/(\theta_C + \theta_X) = 1/3$, which means that the censoring rate for high observation times is about 33% (the right limit of the curve in Figure 4, line 2).

Study case 2 (high tail censoring rate). We take the same sample size $n = 200$. The X_i's, $i = 1, ..., n$ are i.i.d. from K_{μ_X,θ_X} with $\mu_X = 30$ and $\theta_X = 20$. The C_i's, $i = 1, ..., n$ are i.i.d. from K_{μ_C,θ_C} with $\mu_C = \mu_X + 10 = 40$ and $\theta_C = \theta_X/10 = 2$. In this case the (overall) mean censoring rate is about 40% (horizontal dashed line), however the conditional censoring rate in the tail is nearly equal to the limit $\lim_{t\to\infty} q_{F,F_C}(t|z) = \theta_X/(\theta_C + \theta_X) = 10/11$, i.e. is about 91% (see Figure 4, line 4).

We evaluate the performance of the SFKM and SAKM estimators $\widehat{S}_t(x|z)$ and $\widehat{S}_{\hat{t}_z}(x|z)$ with respect to the KM estimator $\widehat{S}_{KM}(x|z)$. In Figure 5 we display the root $MSE_{\widehat{S}}(x|z)$ (lines 1, 3) and the ratio $R_{\widehat{S}}(x|z)$ (lines 2, 4) for the three estimators as functions of the time x. From these plots we can see that both root $MSE_{\widehat{S}_t}(x|z)$ and root $MSE_{\widehat{S}_{\hat{t}_z}}(x|z)$ are equal to the root $MSE_{\widehat{S}_{KM}}(x|z)$ for small values of x and become smaller for large values of x, which shows that the SFKM and SAKM estimators improve the KM estimator.

In Figure 6 (lines 1, 3), for each fixed x, we show the confidence bands containing 90% of the values of $\widehat{S}_{KM}(x|z)$ and $\widehat{S}_{\hat{t}_z}(x|z)$. From these plots we see the ability of the model to fit the data and at the same time to give satisfactory predictions. Compared to those provided by the KM estimator which predicts a constant survival probability for large x, our predictions are more realistic.

In Figure 6 (lines 2, 4) we show the bias square and the variance of $\widehat{S}_{KM}(\cdot|z)$ and $\widehat{S}_{\hat{t}_z}(\cdot|z)$. From these plots we see that the variance of $\widehat{S}_{\hat{t}_z}(\cdot|z)$ is smaller than that of $\widehat{S}_{KM}(\cdot|z)$ in the two study cases. We conclude the same for their biases. However, the bias of $\widehat{S}_{KM}(\cdot|z)$ is large in the study case 2 (Figure 6, Line 4) because of a high conditional censoring rate in the tail (see Figure 4, line 4).

The case of non-constant hazards (see Example 2 of Section 2). The previous study is performed for models satisfying conditions (2.4) and (2.5). Now we consider the case when these conditions are not satisfied. Let X and C be generated from gamma distributions whose hazard rate function can be easily verified not to be asymptotically constant (in fact it is slowly varying at infinity). The survival time

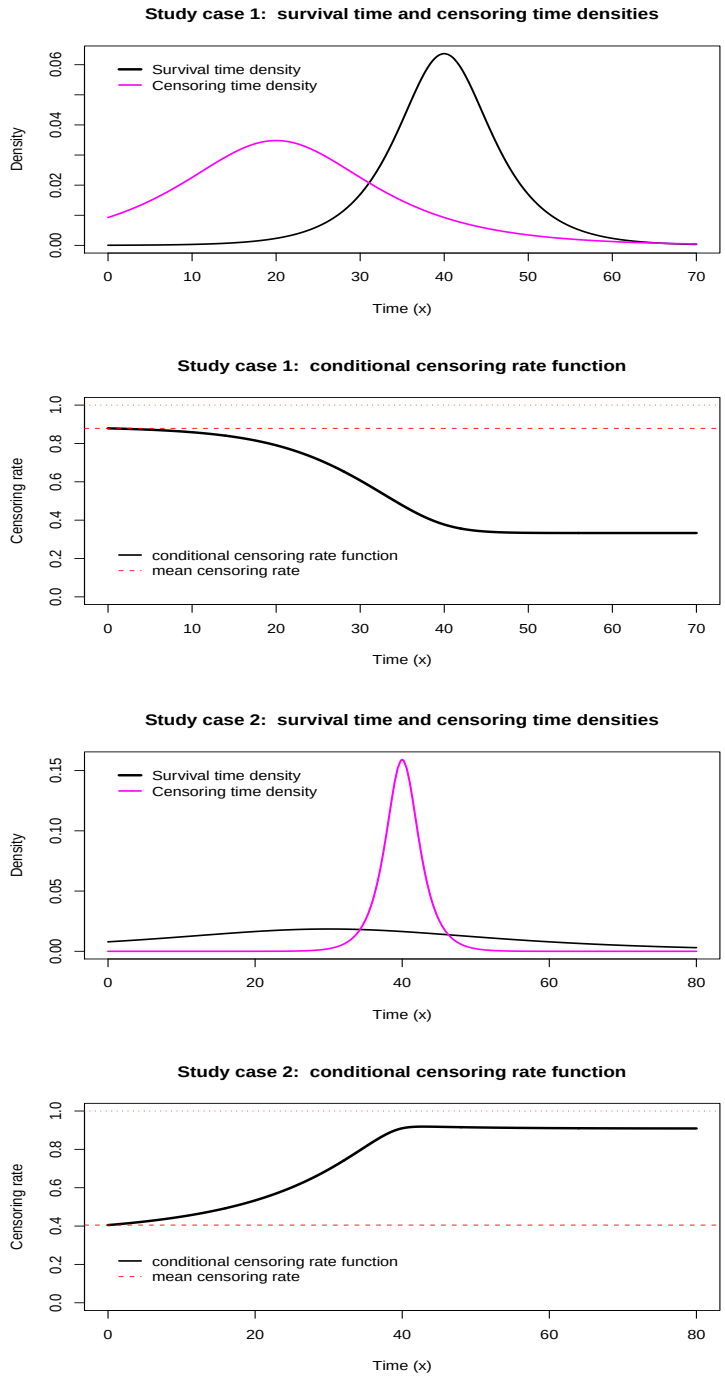

Figure 4. The lines 1, 3 (from the top) display the density functions of the survival and censoring times for study cases 1 and 2 (low and high tail censoring rates respectively). The lines 2, 4 display the conditional censoring rate $q_{F,F_C}(t|z)$ as function of the threshold t, for the two cases.

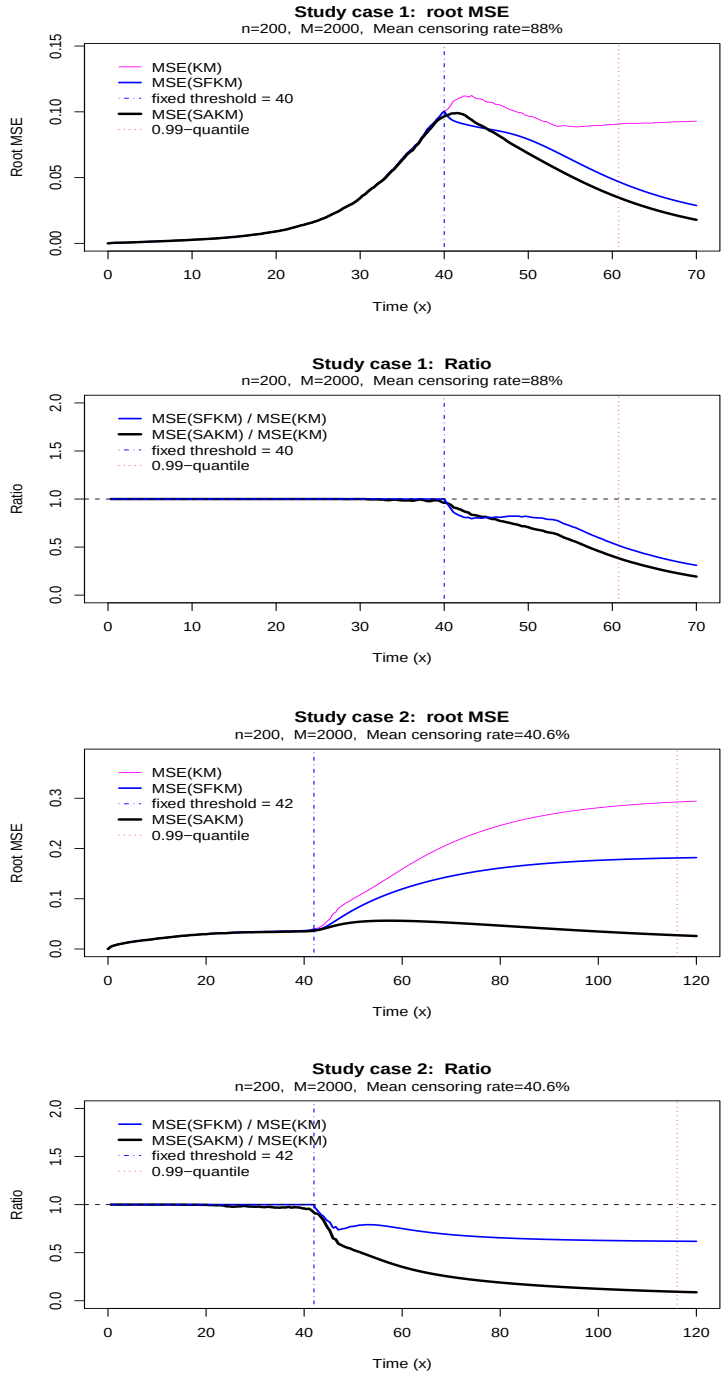

Figure 5. The lines 1, 3 (from the top) display the root type II MSE's of three estimators: $\widehat{S}_{KM}$ (KM), $\widehat{S}_t$ (SFKM) and $\widehat{S}_{\hat{t}_z}$ (SAKM). The lines 2, 4 display the corresponding ratios of the root type II MSE's on the lines 1, 3. The critical value D in the SAKM is set to 6.

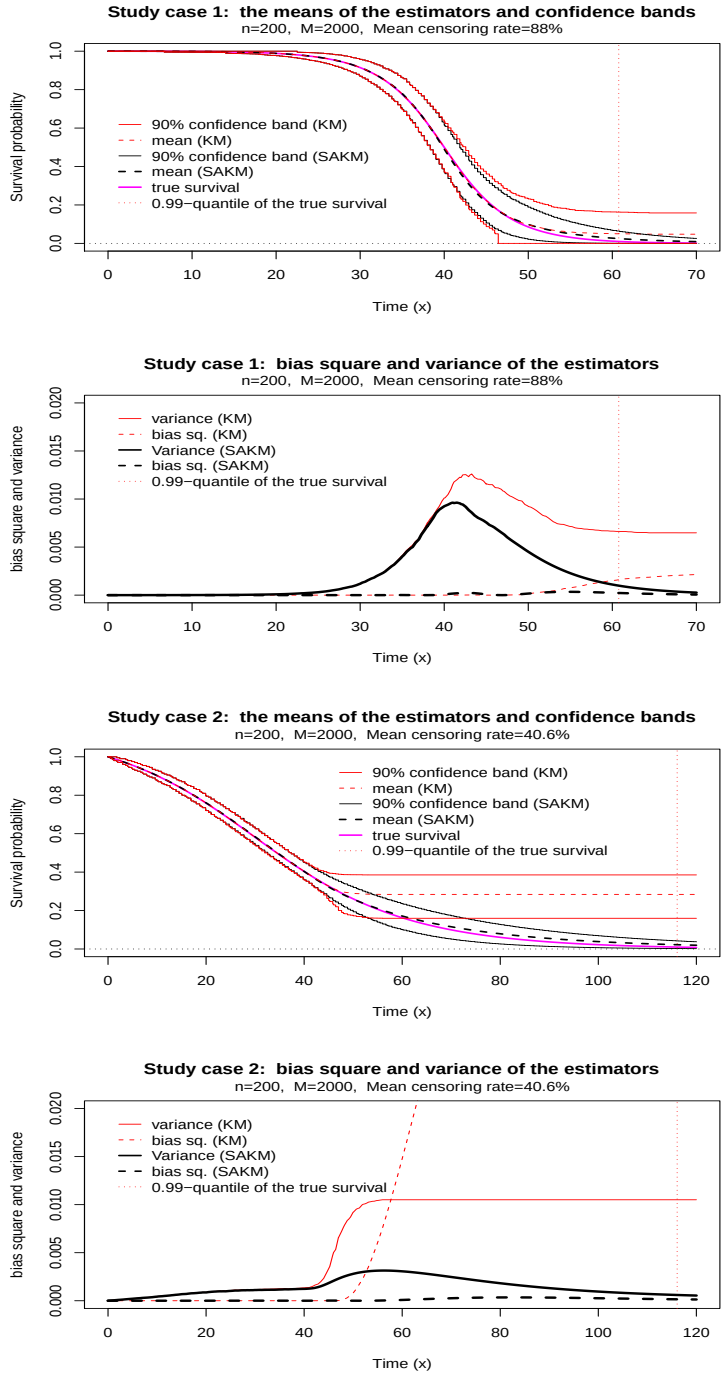

Figure 6. The lines 1, 3 (from the top) display the true survival S_F and the estimated means of $\widehat{S}_{KM}$ (KM) and $\widehat{S}_{\hat{t}_z}$ (SAKM). We give confidence bands containing 90% of the trajectories for each fixed time x. The lines 2, 4 display the corresponding biases square and variances.

Table 1. Simulations with gamma distributions for survival and censoring times

x	5	6	7	8	9	10	11	12	13
$S_F(x\|z)$	0.9682	0.9161	0.8305	0.7166	0.5874	0.4579	0.3405	0.2424	0.1658
Mean of $\widehat{S}_{\widehat{t}_z}(x\|z)$	0.9679	0.9159	0.8318	0.7107	0.5686	0.4504	0.3575	0.2853	0.2287
Mean of $\widehat{S}_{KM}(x\|z)$	0.9679	0.9159	0.8306	0.7160	0.5875	0.4581	0.3399	0.2472	0.1888
Root $MSE_{\widehat{S}_{\widehat{t}_z}}(x\|z)$	0.0135	0.0225	0.0336	0.0461	0.0552	0.0606	0.0702	0.0831	0.0940
Root $MSE_{\widehat{S}_{KM}}(x\|z)$	0.0135	0.0225	0.0345	0.0466	0.0604	0.0758	0.0933	0.1144	0.1284
x	14	15	16	17	18	19	20	21	22
$S_F(x\|z)$	0.1094	0.0699	0.0433	0.0261	0.0154	0.0089	0.0050	0.0028	0.0015
Mean of $\widehat{S}_{\widehat{t}_z}(x\|z)$	0.1841	0.1487	0.1205	0.0979	0.0798	0.0652	0.0534	0.0439	0.0361
Mean of $\widehat{S}_{KM}(x\|z)$	0.1586	0.1453	0.1411	0.1403	0.1402	0.1402	0.1402	0.1402	0.1402
Root $MSE_{\widehat{S}_{\widehat{t}_z}}(x\|z)$	0.0997	0.0998	0.0952	0.0876	0.0785	0.0690	0.0599	0.0515	0.0441
Root $MSE_{\widehat{S}_{KM}}(x\|z)$	0.1384	0.1503	0.1627	0.1731	0.1804	0.1850	0.1877	0.1893	0.1902

X is gamma with shape parameter 10 and rate parameter 1 and the censoring time C is gamma with shape parameter 8.5 and rate parameter 1.2. The mean censoring rate in this example is about 77%. The results of the simulations are given in Figure 2 ($n = 20$ and $n = 500$) and Table 1 ($n = 500$) for $\widehat{S}_{KM}(\cdot|z)$ and $\widehat{S}_{\widehat{t}_z}(\cdot|z)$. They show that for these distributions the SAKM estimator gives a smaller root MSE than the KM estimator even when the sample size is low ($n = 20$) and x is in the range of the data.

7 Application to real data

As an illustration we deal with the well known randomized trial in primary biliary cirrhosis (PBC) from Fleming and Harrington [7] (see Appendix D.1). PBC is a rare but fatal chronic liver disease and the analyzed event is the patient's death. The trial was open for patient registration between January 1974 and May 1984. The observations lasted until July 1986, when the disease and survival status of the patients where recorded. There where $n = 312$ patients registered for the clinical trial, including 125 patients who died. The censored times where recorded either for patients which had been lost to follow up or had undergone liver transplantation or was still alive at the study analysis time (July 1986). The number of censored times is 187 and the censoring rate is about 59.9%. The last observed time is 4556 which is a censored time. Ties occur for the following three times: 264, 1191 and 1690. So there are 122 separate times for which we can observe at least one event. Two treatment groups of patients where compared: the first one ($Z = 1$) of size $n_1 = 158$ was given the DPCA (D-penicillamine drug). The second group ($Z = 0$) of size $n_0 = 154$ was the control (placebo) group. In this example we consider only the group covariate. We are interested to predict the survival probabilities of the patients under study in both groups.

The survival curves based on the KM and SAKM estimators for each group are displayed in Figure 1 (top and bottom pictures respectively). The numerical results on the predictions appear in Table 2. In this table, the time is running from 3 years ($x = 1095$ days) up to 20 years ($x = 7300$) with the step 1 year equivalent to 365 days for convenience.

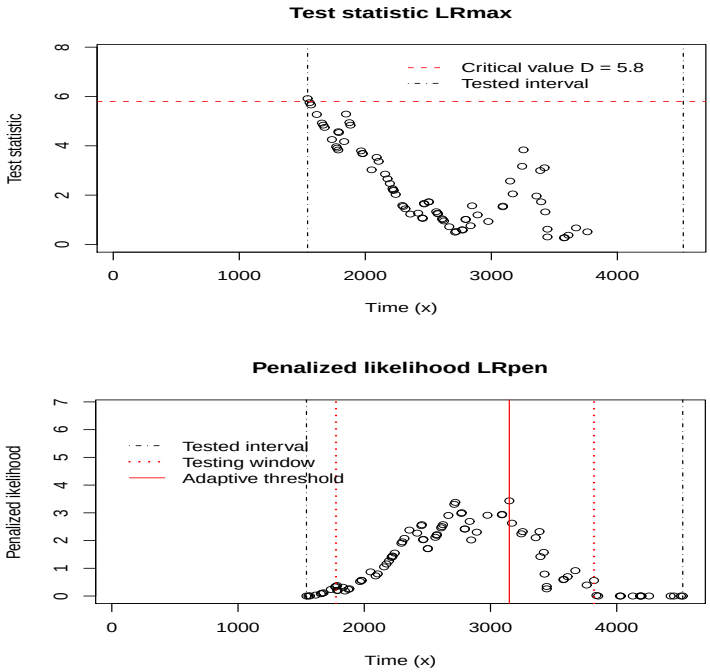

Figure 7. For the placebo group of PBC data we display the test statistics $LR_{\max}(T_k|z)$ as function of T_k (top) and $LR_{\text{pen}}(T_{\widehat{k}}, T_l|z)$ as function of T_l (bottom). The tested interval and the testing window are given by $[T_{\widehat{k}}, T_{k_0}]$ and $[T_{(1-\delta'')\widehat{k}}, T_{\delta'\widehat{k}}]$ respectively. The critical value D is fixed to 5.8.

Based on the usual KM estimator, the following two conclusions can be made: A1) The constant predictions for extreme survival probabilities in both groups appear to be too optimistic after the largest (non-censored) survival time. B1) The DPCA treatment appears to be less efficient than placebo in the long term. The statistical analysis with the SAKM estimator leads to more realistic conclusions: A2) The survival probabilities of each group extrapolate the tendency of the KM estimator as the time is increasing, and B2) the DPCA treatment is more efficient than placebo. For example, from the results in Table 2 we obtain that the survival probability in 20 years is about 2 times higher for the DPCA group than for the

Table 2. Predicted survival probabilities for PBC data

x : years	3	4	5	6	7	8	9	10	11
x : days	1095	1460	1825	2190	2555	2920	3285	3650	4015
DPCA: KM	0.8256	0.7635	0.7077	0.6613	0.5842	0.5417	0.4778	0.4247	0.4247
DPCA: SAKM	0.8256	0.7635	0.7077	0.6595	0.5934	0.5340	0.4805	0.4323	0.3890
Placebo: KM	0.7911	0.7398	0.7146	0.6950	0.6566	0.6055	0.5461	0.4563	0.3604
Placebo: SAKM	0.7911	0.7398	0.7146	0.6950	0.6566	0.6055	0.5497	0.4619	0.3881
x : years	12	13	14	15	16	17	18	19	20
x : days	4380	4745	5110	5475	5840	6205	6570	6935	7300
DPCA: KM	0.3186	0.3186	0.3186	0.3186	0.3186	0.3186	0.3186	0.3186	0.3186
DPCA: SAKM	0.3501	0.3150	0.2834	0.2550	0.2295	0.2065	0.1858	0.1672	0.1505
Placebo: KM	0.3604	0.3604	0.3604	0.3604	0.3604	0.3604	0.3604	0.3604	0.3604
Placebo: SAKM	0.3260	0.2739	0.2302	0.1934	0.1625	0.1365	0.1147	0.0964	0.0810

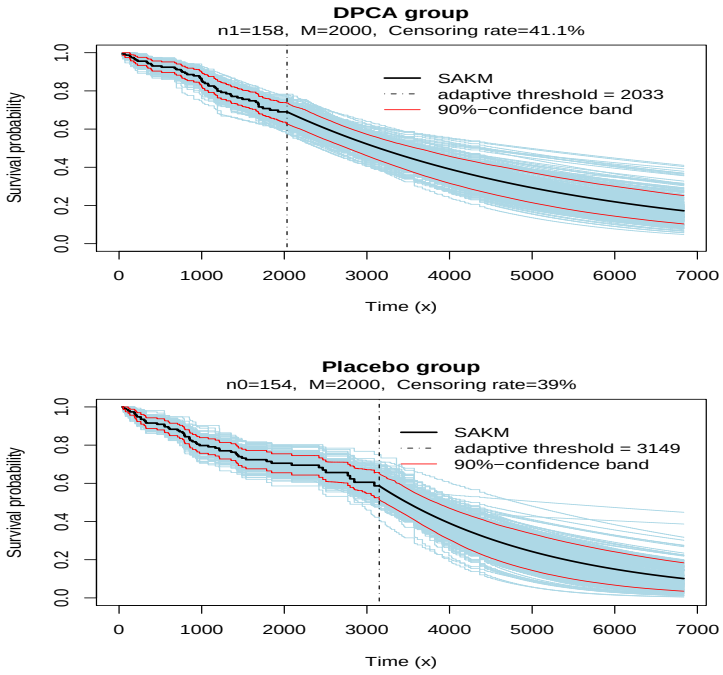

Figure 8. For PBC data, we display the pointwise bootstrap 90% confidence intervals for predicted probabilities: DPCA group (on top) and placebo (control) group (on bottom). We also display 100 bootstrap trajectories of the predicted probabilities for each group.

control group.

From the top picture of Figure 7 we see that the test statistic $LR_{\max}\left(T_k|0\right)$ for the control group ($Z=0$) reaches the critical value $D=5.8\in[D_0,D_1]$ for $k=\widehat{k}_0=90$. Thus the hypotheses $\mathcal{H}_{s_0}\left(0\right)$ was rejected for the break time $\widehat{s}_0=T_{\widehat{k}_0}=1542$. The adaptive threshold $\widehat{t}_0$ is chosen via the maximization of the penalized quasi-log-likelihood (5.1). In the bottom picture of Figure 7 we see that the maximum is attained for the adaptive index $\widehat{l}_0=30$ and threshold $\widehat{t}_0=T_{\widehat{l}_0}=3149$. Thus, our testing-pursuit-selection procedure has captured the "convex bump" on the control Kaplan-Meier curve (for $Z=0$) between the times 2000 and 3500, which is easily seen in the bottom picture of Figure 1.

The pointwise (in x) 0.9-confidence bootstrap intervals for the predicted probabilities $\widehat{S}_{\widehat{t}_1}\left(x|1\right)$ and $\widehat{S}_{\widehat{t}_0}\left(x|0\right)$ are displayed in Figure 8 (top for DPCA treatment group $Z=1$ and bottom for control group $Z=0$). Here $\widehat{t}_1=2033$ and $\widehat{t}_0=3149$ are the adaptive thresholds computed from the original sample. The adaptive estimators of the mean parameters θ_1 and θ_0 are respectively $\widehat{\theta}_{1,\widehat{t}_1}=3457.85$ and $\widehat{\theta}_{0,\widehat{t}_0}=2096.22$. We generated $M=2000$ bootstrap samples of size $n=312$ taken at random from the general sample gathering the data coming from the two groups. For the m-th bootstrap sample the SAKM estimators $\widehat{S}^{(m)}_{\widehat{t}_1^{(m)}}\left(x|1\right)$ and $\widehat{S}^{(m)}_{\widehat{t}_0^{(m)}}\left(x|0\right)$ are

computed as functions of x with their own adaptive thresholds $\widehat{t}_1^{(m)}$ and $\widehat{t}_0^{(m)}$.

8 Conclusion

This article deals with estimation of the survival probability in the framework of censored survival data. While the Kaplan-Meier estimator provides a flexible estimate of the survival function in the range of the data it can be improved for prediction of the extreme values, especially when the censoring rate is high. We propose a new approach based on the Kaplan-Meier estimator by adjusting a parametric correction to the tail beyond a given threshold t.

First we determine the rate of convergence of the corresponding estimators of the parameters in the adjusted model for a sequence of deterministic thresholds $t = t_{z,n}$ for each category z of the model covariate. This is done under the assumption that the hazard function is fitted by a constant in the sense that conditions (2.4) and (2.5) are satisfied. It is interesting to note that the rate of convergence depends not only on the class of survival time distributions but also on the class of censoring time distributions. By simulations we show that our approach is robust if the (survival and censoring) fitted tails are misspecified.

In applications the threshold t usually is not known. To overcome this we propose a testing-pursuit-selection procedure which yields an adaptive threshold $t = \widehat{t}_{z,n}$ in two stages: a sequential hypothesis testing and an adaptive choice of the threshold based on the maximization of a penalized quasi-log-likelihood. This testing-pursuit-selection procedure provides also a goodness-of-fit test for the parametric-based part of the model.

We perform numerical simulations with both the fixed and adaptive threshold estimators. Our simulations show that both estimators improve the Kaplan-Meier estimator not only in the long term, but also in a mid range inside the data. Comparing the fixed threshold and adaptive threshold estimators, we found that the adaptive choice of the threshold significantly improves on the quality of the predictions of the survival function.

We have seen that the quality of estimation of the extreme survival probabilities depends on the conditional censoring rate function, which describes the variations of the censoring rate as the time increases. The improvement over the Kaplan-Meier estimator is especially effective when the conditional censoring rate is high in the tail.

A Appendix: Proofs of the results

A.1 Auxiliary assertions

The following lemma plays the crucial role in the proof of our main results. Assume that $\mathbf{Y}_1, ..., \mathbf{Y}_n$ are i.i.d. with common distribution Q. Let $\mathbb{Q}$ be the joint distribution of $\mathbf{Y} = (\mathbf{Y}_1, ..., \mathbf{Y}_n)$. Let Q_1, Q_0 be two probability measures on $\mathbb{R}$ such

that Q, Q_0 and Q_1 are equivalent. Define the quasi-log-likelihood ratio by

$$\mathcal{L}(Q_1, Q_0) = \sum_{i=1}^{n} \log \frac{dQ_1}{dQ_0}(\mathbf{Y}_i).$$

Lemma A.1. *For any $x \geq 0$, $n \geq 1$, we have*

$$\mathbb{Q}\left(\mathcal{L}(Q_1, Q_0) > x + n\chi^2(Q, Q_0)\right) \leq \exp\left[-\frac{x}{2}\right].$$

Proof. By exponential Chebyshev's inequality, for any $y > 0$,

$$\mathbb{Q}(\mathcal{L}(Q_1, Q_0) > y) \leq \exp\left[-y/2 + \log \mathbb{Q} \exp\left(\frac{1}{2}\mathcal{L}(Q_1, Q_0)\right)\right]. \tag{A.1}$$

Since $\mathbf{Y}_1, ..., \mathbf{Y}_n$ are i.i.d. with common distribution Q, we get

$$\log \mathbb{Q} \exp\left(\frac{1}{2}\mathcal{L}(Q_1, Q_0)\right) = n \log Q\sqrt{\frac{dQ_1}{dQ_0}}. \tag{A.2}$$

By Holder's inequality $Q\left(\sqrt{dQ_1/dQ_0}\right) \leq \sqrt{Q(dQ/dQ_0)} = \sqrt{1 + \chi^2(Q, Q_0)}$. Using the last bound and (A.1), (A.2), it follows

$$\begin{aligned}\mathbb{Q}(\mathcal{L}(Q_1, Q_0) > y) &\leq \exp\left\{-\frac{y}{2} + \frac{n}{2}\log\left(1 + \chi^2(Q, Q_0)\right)\right\} \\ &\leq \exp\left\{-\frac{y}{2} + \frac{n}{2}\chi^2(Q, Q_0)\right\}.\end{aligned}$$

Letting $y = x + n\chi^2(Q, Q_0)$ completes the proof. □

Now we produce an exponential bound for the quasi-log-likelihood ratio

$$\mathcal{L}_t(\theta'|z) - \mathcal{L}_t(\theta|z) = \sum_{z_i = z} \log \frac{p_{F_{\theta',t},F_C}}{p_{F_{\theta,t},F_C}}(\mathbf{Y}_i|z).$$

Lemma A.2. *For any $\theta, \theta' \in \mathbb{R}$, $z \in \mathcal{Z}$ and any $x \geq 0$ it holds*

$$\mathbb{P}\left(\mathcal{L}_t(\theta'|z) - \mathcal{L}_t(\theta|z) > x + n\chi^2\left(P_{F,F_C}(\cdot|z), P_{F_{\theta,t},F_C}(\cdot|z)\right)\right) \leq \exp\left(-\frac{x}{2}\right).$$

Proof. Let $I_z = \{i : z_i = z\}$. Note that $\mathbf{Y}_i = (T_i, \Delta_i)'$, $i \in I_z$, are i.i.d. variables. We apply Lemma A.1 with $\mathbf{Y} = \{\mathbf{Y}_i : i \in I_z\}$ and $Q = P_{F,F_C}(\cdot|z)$, $Q_0 = P_{F_{\theta,t},F_C}(\cdot|z)$, $Q_1 = P_{F_{\theta',t},F_C}(\cdot|z)$, which ends the proof. □

Next, we give an exponential bound for the maximum quasi-log-likelihood ratio which permits to obtain a rate of convergence of $\widehat{\theta}_{z,t}$.

Lemma A.3. *For any $\theta > 0$, $t \geq x_0$ and any $x \geq 0$ it holds*

$$\mathbb{P}\left(\widehat{n}_{z,t}\mathcal{K}\left(\widehat{\theta}_{z,t}, \theta\right) > x + n\chi^2\left(P_{F,F_C}(\cdot|z), P_{F_{\theta,t},F_C}(\cdot|z)\right) + 2\log n\right) \leq 2\exp\left(-\frac{x}{2}\right),$$

where $z \in \mathcal{Z}$ and by convention $0 \cdot \infty = 0$.

Proof. We prove that

$$\mathbb{P}\left(\widehat{n}_{z,t}\mathcal{K}\left(\widehat{\theta}_{z,t},\theta\right)>y\right)\leq 2n\exp\left(-x/2\right)=2\exp\left(-x/2+\log n\right), \tag{A.3}$$

where $y=x+n\chi^2\left(P_{F,F_C}\left(\cdot|z\right),P_{F_{\theta,t},F_C}\left(\cdot|z\right)\right)\geq 0$.

Since $\widehat{n}_{z,t}\widehat{\theta}_{z,t}=\sum_{T_i>t,\ z_i=z}\left(T_i-t\right)$, by direct calculations, we have $\mathcal{L}_t\left(\theta'|z\right)-\mathcal{L}_t\left(\theta|z\right)=\widehat{n}_{z,t}\Lambda_z\left(\theta'\right)$, where $\Lambda_z\left(u\right)=\log\left(\theta/u\right)-\left(u^{-1}-\theta^{-1}\right)\widehat{\theta}_{z,t}$. Using that $\mathcal{K}\left(\theta',\theta\right)=\theta'/\theta-1-\log\left(\theta'/\theta\right)$, we deduce $\mathcal{K}\left(\widehat{\theta}_{z,t},\theta\right)=\Lambda_z\left(\widehat{\theta}_{z,t}\right)$. Denote for brevity $g\left(u,k\right)=\left(\log\left(\theta/u\right)-y/k\right)/\left(u^{-1}-\theta^{-1}\right),\ u\neq\theta$. Note that, for $0<u<\theta$ the inequality $k\Lambda_z\left(u\right)>y$ is equivalent to $g\left(u,k\right)>\widehat{\theta}_{z,t}$ and for $u>\theta$ the inequality $k\Lambda_z\left(u\right)>y$ is equivalent to $g\left(u,k\right)<\widehat{\theta}_{z,t}$. Moreover the function $g\left(u,k\right)$ has a maximum for $0<u<\theta$ and a minimum for $u>\theta$.

Let $\theta^+\left(k\right)=\arg\max_{0\leq u<\theta}\ g\left(u,k\right)$ and $\theta^-\left(k\right)=\arg\min_{u>\theta}\ g\left(u,k\right)$. Then

$$\begin{aligned}\left\{\widehat{n}_{z,t}\Lambda_z\left(\widehat{\theta}_{z,t}\right)>y,\ \widehat{\theta}_{z,t}<\theta\right\} &= \left\{g\left(\widehat{\theta}_{z,t},\widehat{n}_{z,t}\right)>\widehat{\theta}_{z,t},\ \widehat{\theta}_{z,t}<\theta\right\}\\ &\subset \left\{g\left(\theta^+\left(\widehat{n}_{z,t}\right),\widehat{n}_{z,t}\right)>\widehat{\theta}_{z,t},\ \widehat{\theta}_{z,t}<\theta\right\}\\ &= \left\{\widehat{n}_{z,t}\Lambda_z\left(\theta^+\left(\widehat{n}_{z,t}\right)\right)>y,\ \widehat{\theta}_{z,t}<\theta\right\}\\ &\subset \left\{\widehat{n}_{z,t}\Lambda_z\left(\theta^+\left(\widehat{n}_{z,t}\right)\right)>y\right\}.\end{aligned}$$

In the same way, we get $\left\{\widehat{n}_{z,t}\Lambda_z\left(\widehat{\theta}_{z,t}\right)>y,\ \widehat{\theta}_{z,t}>\theta\right\}\subset\left\{\widehat{n}_{z,t}\Lambda_z\left(\theta^-\left(\widehat{n}_{z,t}\right)\right)>y\right\}$. Since $\Lambda_z\left(\widehat{\theta}_{z,t}\right)=\mathcal{K}\left(\widehat{\theta}_{z,t},\theta\right)$ and $\mathcal{K}\left(\widehat{\theta}_{z,t},\theta\right)=0$ if $\widehat{\theta}_{z,t}=\theta$, these inclusions imply

$$\begin{aligned}\left\{\widehat{n}_{z,t}\mathcal{K}\left(\widehat{\theta}_{z,t},\theta\right)>y\right\} &\subset \left\{\widehat{n}_{z,t}\Lambda_z\left(\theta^+\left(\widehat{n}_{z,t}\right)\right)>y\right\}\\ &\quad\cup\left\{\widehat{n}_{z,t}\Lambda_z\left(\theta^-\left(\widehat{n}_{z,t}\right)\right)>y\right\}.\end{aligned} \tag{A.4}$$

From (A.4), we get

$$\begin{aligned}&\mathbb{P}\left(\widehat{n}_{z,t}\mathcal{K}\left(\widehat{\theta}_{z,t},\theta\right)>y\right)\\ \leq\ &\mathbb{P}\left(\widehat{n}_{z,t}\Lambda_z\left(\theta^+\left(\widehat{n}_{z,t}\right)\right)>y\right)+\mathbb{P}\left(\widehat{n}_{z,t}\Lambda_z\left(\theta^-\left(\widehat{n}_{z,t}\right)\right)>y\right)\\ \leq\ &\sum_{k=1}^{n}\mathbb{P}\left(\widehat{n}_{z,t}\Lambda_z\left(\theta^+\left(k\right)\right)>y\right)+\sum_{k=1}^{n}\mathbb{P}\left(\widehat{n}_{z,t}\Lambda_z\left(\theta^-\left(k\right)\right)>y\right).\end{aligned} \tag{A.5}$$

By Lemma A.2, it follows, for $k=1,...,n$, $\mathbb{P}\left(\widehat{n}_{z,t}\Lambda_z\left(\theta^{\pm}\left(k\right)\right)>y\right)\leq\exp\left(-x/2\right)$. Then, by (A.5), we get (A.3), which ends the proof. □

A.2 Proof of Theorems 3.1 and 3.2

Theorem 3.1 follows immediately from Lemma A.3 if we set $x=2\log n$. Theorem 3.2 is a consequence of Theorem 3.1 and (3.5).

A.3 Proof of Lemma 3.3

By (2.1) it follows that $\mathbb{E}\widehat{n}_{z,t} = \sum_{z_i=z} \int_t^\infty f_F(x|z)\, S_C(x|z)\, dx$. Therefore, integrating by parts, we have $\mathbb{E}\widehat{n}_{z,t} = n_z S_F(t|z)\, S_C(t|z)\, (1 - q_{F,F_C}(t|z))$. Using (3.8) proves the first assertion.

Denote, for brevity, $\xi_i = 1_{\{T_i>t,\Delta_i=1\}}$ and $p = \mathbb{P}(T_i > t, \Delta_i = 1)\, 1_{\{z_i=z\}}$. Then $\widehat{n}_{z,t} = \sum_{z_i=z} \xi_i$ and $\mathbb{E}\widehat{n}_{z,t} = n_z p$. Using exponential Chebyshev's inequality, for any $x > 0$ and any $u > 0$, we obtain

$$\mathbb{P}(\widehat{n}_{z,t} \leq \mathbb{E}\widehat{n}_{z,t} - x) \leq \exp\left(-ux + n_z p \frac{u^2}{2}\right).$$

Choosing $u = 1/2$ and $x = \mathbb{E}\widehat{n}_{z,t}/2$, we get $\mathbb{P}(\widehat{n}_{z,t} \leq \mathbb{E}\widehat{n}_{z,t}/2) \leq \exp(-n_z p/8)$, which proves the second assertion.

A.4 Proof of Theorem 4.1

Lemma A.4. *Assume that Q and Q_0 are two equivalent probability measures on a measurable space. Then*

$$\chi^2(Q, Q_0) \leq \int \left(\log \frac{dQ_0}{dQ}\right)^2 \exp\left(\left|\log \frac{dQ_0}{dQ}\right|\right) dQ.$$

Proof. Consider the convex function $g(x) = (x-1)^2/x$. Then $\chi^2(Q, Q_0) = \int g(dQ_0/dQ)\, dQ$. Since $(x-1)^2 \leq x^2 \log^2 x = \exp(2\log x) \log^2 x$ for $x \geq 1$, and $(x-1)^2 \leq \log^2 x$ for $x \in (0,1)$, we get $g(x) \leq \log^2 x \exp(|\log x|)$ for $x > 0$. □

We deduce Theorem 4.1 from Theorem 3.4. Let $z \in \mathcal{Z}$ and $t \geq x_0$. Consider the distance $\rho_t(h_1, h_2) = \sup_{x>t} |h_1(x) - h_2(x)|$, where h_1, h_2 are two non-negative functions. First we prove the following bound:

$$\chi^2\left(P_{F,F_C}(\cdot|z), P_{F_{\theta_z,t},F_C}(\cdot|z)\right) = O\left(S_C(t|z)\, S_F(t|z)\, \rho_t^2\right) \text{ as } t \to \infty. \tag{A.6}$$

By Lemma A.4,

$$\chi^2\left(P_{F,F_C}(\cdot|z), P_{F_{\theta_z,t},F_C}(\cdot|z)\right) \leq \int_{x_0}^\infty \left(\log \frac{dP_{F,F_C}}{dP_{F_{\theta_z,t},F_C}}(x, \delta|z)\right)^2 \tag{A.7}$$
$$\times \exp\left(\left|\log \frac{dP_{F,F_C}}{dP_{F_{\theta_z,t},F_C}}(x, \delta|z)\right|\right) P_{F,F_C}(dx, d\delta|z).$$

According to (2.1), for any $x > t$,

$$\begin{aligned}
\log \frac{dP_{F,F_C}}{dP_{F_{\theta_z,t},F_C}}(x, \delta|z) &= \log \frac{h_F(x|z)^\delta\, S_F(x|z)}{h_{F_{\theta_z,t}}(x|z)^\delta\, S_{F_{\theta_z,t}}(x|z)} \\
&= \delta \log \frac{h_F(x|z)}{\theta_z^{-1}} - \int_t^x \left(h_F(v|z) - \theta_z^{-1}\right) dv.
\end{aligned}$$

For brevity, we denote $\rho_t = \rho_t\left(h_F(\cdot|z), \theta_z^{-1}\right)$. Since $\log(1+u) \leq 2|u|$, for $u > -1/2$, it follows that

$$\left|\log \frac{dP_{F,F_C}}{dP_{F_{\theta_z,t},F_C}}(x,\delta|z)\right| \leq c\rho_t(1+(x-t)), \tag{A.8}$$

whenever $\rho_t \leq 1/(2\theta_{\min})$, where $c = \max\{2\theta_{\max}, 1\}$.

Denoting $g_{\rho_t}(x) = (1+x)^2 \exp(c\rho_t(1+x))$, from (A.7) and (A.8), we get

$$\begin{aligned}
& \chi^2\left(P_{F,F_C}(\cdot|z), P_{F_{\theta_z,t},F_C}(\cdot|z)\right) \\
\leq\ & c^2\rho_t^2 \int_{(t,\infty)\times\{0,1\}} g_{\rho_t}(x-t)\, p_{F,F_C}(x,\delta|z)\, \nu(dx,d\delta) \\
=\ & c^2\rho_t^2 \int_t^\infty \sum_{\delta\in\{0,1\}} g_{\rho_t}(x-t)\, f_F(x|z)^\delta\, S_F(x|z)^{1-\delta}\, f_C(x|z)^{1-\delta}\, S_C(x|z)^\delta\, dx.
\end{aligned}$$

Since $S_C(x) \leq S_C(t)$ and $S_F(x) \leq S_F(t)$, for $x \geq t$, we obtain

$$\begin{aligned}
& \chi^2\left(P_{F,F_C}(\cdot|z), P_{F_{\theta_z,t},F_C}(\cdot|z)\right) \\
\leq\ & c^2\rho_t^2 S_F(t|z)\, S_C(t|z) \int_t^\infty g_{\rho_t}(x-t)\left(\frac{f_F(x|z)}{S_F(t|z)} + \frac{f_C(x|z)}{S_C(t|z)}\right) dx.
\end{aligned}$$

From (2.4), $h_F(x|z)$ is bounded from below for x large enough:

$$\begin{aligned}
h_F(x|z) &\geq \theta_z^{-1}(1 - |\theta_z h_F(x|z) - 1|) \\
&\geq \theta_{\max}^{-1}\left(1 - A\exp\left(-\alpha_{\min}\frac{x}{\theta_z}\right)\right) \\
&\geq 1/(2\theta_{\max}),
\end{aligned}$$

whenever $x \geq t_{\min} = \theta_{\max}\log(2A)/\alpha_{\min}$, where $\alpha_{\min} = \min_{z\in\mathcal{Z}}\alpha_z$. This implies

$$\frac{S_F(x|z)}{S_F(t|z)} = \exp\left(-\int_t^x h_F(v|z)\, dv\right) \leq \exp(-c_0(x-t)),$$

where $c_0 = 1/(2\theta_{\max})$. Integrating by parts, for any $t \geq t_{\min}$,

$$\begin{aligned}
& \int_t^\infty g_{\rho_t}(x-t)\frac{f_F(x|z)}{S_F(t|z)}dx \\
=\ & \left[-g_{\rho_t}(x-t)\frac{S_F(x|z)}{S_F(t|z)}\right]_t^\infty + \int_t^\infty \frac{S_F(x|z)}{S_F(t|z)} g'_{\rho_t}(x-t)\, dx.
\end{aligned}$$

If $\rho_t \leq c_0/(2c)$, we have

$$\begin{aligned}
& \int_t^\infty g_{\rho_t}(x-t)\frac{f_F(x|z)}{S_F(t|z)}dx \\
\leq\ & \exp(c\rho_t) + \int_0^\infty (1+x)(2+c\rho_t(1+x))\exp(c\rho_t(1+x) - c_0 x)\, dx
\end{aligned}$$

$$\leq \quad \exp\left(\frac{c_2}{2}\right)\left(2+\frac{8}{c_0}+\frac{16}{c_0^2}\right)=O\left(1\right).$$

In the same way, conditions $h_C\left(x|z\right)\geq c_{\min}$, for $x\geq t_{\min}$ and $\rho_t\leq c_{\min}/\left(2c\right)$ imply, for $t\geq t_{\min}$,

$$\int_{t_{z,n}}^{\infty} g_{\rho_{t_{z,n}}}\left(x-t\right)\frac{f_C\left(x|z\right)}{S_C\left(t|z\right)}dx=O\left(1\right).$$

Putting together these bounds, yields (A.6).

Next, we find a sequence $t_{z,n}$ which verifies (3.5) and (3.10).

Since $S_C\left(t|z\right)\leq 1$, for verifying (3.5), it remains to find $t=t_{z,n}$ such that

$$S_F\left(t_{z,n}|z\right)\rho_{t_{z,n}}^2=O\left(\frac{\log n}{n}\right). \tag{A.9}$$

Recall that $\alpha_z'=\alpha_z/\theta_z$ and $\gamma_z'=\gamma_z/\theta_z$ (see Example in Section 2). To prove (A.9), we note that, by (2.4),

$$\begin{aligned} S_F\left(t_{z,n}\right) &= \exp\left(-\int_{x_0}^{t_{z,n}} h_F\left(v|z\right)dv\right) \\ &\leq \exp\left(-\int_{x_0}^{t_{z,n}}\left(\theta_z^{-1}-\theta_z^{-1}Ae^{-\alpha_z' v}\right)dv\right) \\ &= O\left(\exp\left(-\theta_z^{-1}\left(t_{z,n}-x_0\right)\right)\right) \end{aligned} \tag{A.10}$$

and, again by condition (2.4),

$$\rho_{t_{z,n}}^2=O\left(\exp\left(-2\alpha_z' t_{z,n}\right)\right). \tag{A.11}$$

Using (A.9), (A.10) and (A.11) we find $t_{z,n}$ from the following equation

$$\exp\left(-\left(\theta_z^{-1}+2\alpha_z'\right)t_{z,n}\right)=O\left(\frac{\log n}{n}\right).$$

The solution has the following expansion:

$$t_{z,n}=\frac{1}{\theta_z^{-1}+2\alpha_z'}\log n+o\left(\log n\right). \tag{A.12}$$

Thus (A.9) and consequently (3.5) are verified.

Now we prove (3.10). In the same way as in (A.10), we get

$$S_F\left(t_{z,n}\right)\geq\exp\left(-\theta_z^{-1}\left(t_{z,n}-x_0\right)-\frac{A}{\alpha_z}\exp\left(-\alpha_z' x_0\right)\right). \tag{A.13}$$

From (A.13) and (A.12), we get the following lower bound

$$nS_F\left(t_{z,n}|z\right) \quad \geq \quad n\exp\left(-\theta_z^{-1}t_{z,n}-c_1\right)$$

$$\begin{aligned} &\geq n \exp\left(-\frac{\log n - \log\log n}{1+2\alpha_z} - c_1\right) \\ &\geq c_2 n^{1-\frac{1}{1+2\alpha_z}} \log^{\frac{1}{1+2\alpha_z}} n \\ &= c_2 n^{\frac{2\alpha_z}{1+2\alpha_z}} \log^{\frac{1}{1+2\alpha_z}} n, \end{aligned} \tag{A.14}$$

where c_1, c_2 are some positive constants and n is large enough. Now condition (3.10) follows from (A.14) and from (4.1).

Assertion (4.2) follows from Theorem 3.4 using (A.14).

A.5 Proof of Theorem 4.2

As in the proof of Theorem 4.1 we verify (3.5) and (3.10). From (2.5) it follows

$$S_C\left(t_{z,n}|z\right) \leq \exp\left(-\gamma_z'\left(t_{z,n}-x_0\right) + \frac{M}{\mu-1}\left(1+x_0\right)^{-\mu+1}\right). \tag{A.15}$$

From (A.6), (A.10), (A.11) and (A.15), we have

$$\begin{aligned} \chi^2\left(P_{F,F_C}\left(\cdot|z\right), P_{F_{\theta_z,t_{z,n}},F_C}\left(\cdot|z\right)\right) &= O\left(S_C\left(t_{z,n}|z\right) S_F\left(t_{z,n}|z\right) \rho^2_{t_{z,n}}\right) \\ &= O\left(\exp\left(-\left(\gamma_z' + \theta_z^{-1} + 2\alpha_z'\right) t_{z,n}\right)\right). \end{aligned}$$

We find $t_{z,n}$ as the solution of the equation

$$\exp\left(-\left(\gamma_z' + \theta_z^{-1} + 2\alpha_z'\right) t_{z,n}\right) = O\left(\frac{\log n}{n}\right),$$

which gives $t_{z,n} = \left(\theta_z^{-1} + \gamma_z' + 2\alpha_z'\right)^{-1} \log n + o\left(\log n\right)$. Thus (3.5) is verified. Condition (3.10) follows from

$$\begin{aligned} nS_C\left(t_{z,n}|z\right) S_F\left(t_{z,n}|z\right) &\geq n \exp\left(-\gamma_z' - \theta_z^{-1} t_{z,n} - c_1\right) \\ &\geq n \exp\left(-\left(\theta_z^{-1} + \gamma_z'\right) \frac{\log n - \log\log n}{\theta_z^{-1} + \gamma_z' + 2\alpha_z'} - c_1\right) \\ &\geq c_2 n^{1-\frac{1+\gamma_z}{1+\gamma_z+2\alpha_z}} \log^{\frac{1+\gamma_z}{1+\gamma_z+2\alpha_z}} n \\ &= c_2 n^{\frac{2\alpha_z}{1+\gamma_z+2\alpha_z}} \log^{\frac{1+\gamma_z}{1+\gamma_z+2\alpha_z}} n, \end{aligned} \tag{A.16}$$

where c_1, c_2 are some positive constants and n is large enough.

The proof of (3.8) is based on similar arguments as in Section A.4.

References

[1] Aalen O. O. *Nonparametric inference in connection with multiple decrements models.* Scand. J. Statist., 1976, **3**, 15–27.

[2] Andersen P. K., Borgan Ø., Gill R. D. and Keiding N. *Statistical models based on counting processes.* Springer series in statistics. New York: Springer-Verlag, 1993.

[3] Bickel P. J., Klaassen C. A., Ritov Y. and Wellner J. A. *Efficient and adaptive estimation for semiparametric models*. The Johns Hopkins University Press, 1993.

[4] Cox D. R. *Regression models and life tables. J. Roy. Statist. Soc.* Ser. B, 1972, **34**, 187–220.

[5] Dress H. *Optimal rates of convergence for estimates of the extreme value index*. Ann. Statist., 1998, **26,** 434–448.

[6] Escobar A. and Meeker W. *A rewiew of accelerated test models*. Statistical Science, 2006, **21**, 552–577.

[7] Fleming T. and Harrington D. *Counting processes and survival analysis*. Wiley, 1991.

[8] Grama I. and Spokoiny V. *Statistics of extremes by oracle estimation*. Ann. Statist., 2008, **36**, 1619–1648.

[9] Grama I., Tricot J.-M. and Petiot J.-F. *Estimation of survival probabilities by adjusting a Cox model to the tail*. C.R. Acad. Sci. Paris, Ser. I., 2011, **349,** 807–811.

[10] Hall P. *On some simple estimates of an exponent of regular variation*. J. Roy. Statist. Soc. Ser. B, 1982, **44**, 37–42.

[11] Hall P. and Welsh A. H. *Best attainable rates of convergence for estimates of parameters of regular variation*. Ann. Statist., 1984, **12**, 1079–1084.

[12] Hall P. and Welsh A. H. *Adaptive estimates of regular variation*. Ann. Statist., 1985, **13,** 331–341.

[13] Kalbfleisch J. D. and Prentice R. L. *The Statistical analysis of failure time data*. Wiley, 2002.

[14] Kaplan E. I. and Meier P. *Nonparametric estimation from incomplete observation*. J. Amer. Statist. Assoc., 1958, **53**, 457–481.

[15] Kiefer J. and Wolfowitz J. *Consitency of the maximum likelihood estimator in the presence of infinitely many nuisance parameters*. Ann. Math. Statist., 1956, **27**, 887–906.

[16] Klein J. P. and Moeschberger M. L. *Survival analysis: techniques for censored and truncated data*. Springer, 2003.

[17] Meier P., Karrison T., Chappell R. and Xie, H. *The price of Kaplan-Meier*. Journal of the American Statistical Association, 2004, **99**, 890–896.

[18] Miller R. *What price Kaplan-Meier ?* Biometrics, 1983, **39**, 1077–1081.

[19] Nelson W. B. *Hazard plotting for incomplete failure data*. J. Qual. Technol., 1969, **1**, 27–52.

[20] Nelson W. B. *Theory and applications of hazard plotting for censored failure data*. Technometrics, 1972, **14**, 945–965.

[21] Tseng Y.-K., Hsieh F. and Wang J.-L. *Joint modeling of accelerated failure time and longitudinal data*. Biometrica, 2005, **92**, 587–603.

[22] Wei L. J. *The accelerated failure time model: a useful alternative to the Cox regression model in survival analysis*. Statistics in Medicine, 1992, **11**, 1871–1879.

Grama I.; Tricot J.M.; Petiot, J.F. *Received July 10, 2013*

Université de Bretagne Sud, LMBA, UMR CNRS 6205,
Vannes, France

E-mail: *ion.grama@univ-ubs.fr*;
jean-marie.tricot@univ-ubs.fr;
jean-francois@univ-ubs.fr