DE GRUYTER Int. J. Biostat. 2021; aop

Jun Ma*, Dominique-Laurent Couturier, Stephane Heritier and Ian C. Marschner

Penalized likelihood estimation of the proportional hazards model for survival data with interval censoring

https://doi.org/10.1515/ijb-2020-0104
Received July 15, 2020; accepted October 1, 2021; published online October 27, 2021

Abstract: This paper considers the problem of semi-parametric proportional hazards model fitting where observed survival times contain event times and also interval, left and right censoring times. Although this is not a new topic, many existing methods suffer from poor computational performance. In this paper, we adopt a more versatile penalized likelihood method to estimate the baseline hazard and the regression coefficients simultaneously. The baseline hazard is approximated using basis functions such as M-splines. A penalty is introduced to regularize the baseline hazard estimate and also to ease dependence of the estimates on the knots of the basis functions. We propose a Newton–MI (multiplicative iterative) algorithm to fit this model. We also present novel asymptotic properties of our estimates, allowing for the possibility that some parameters of the approximate baseline hazard may lie on the parameter space boundary. Comparisons of our method against other similar approaches are made through an intensive simulation study. Results demonstrate that our method is very stable and encounters virtually no numerical issues. A real data application involving melanoma recurrence is presented and an R package 'survivalMPL' implementing the method is available on R CRAN.

Keywords: asymptotic properties; automated smoothing; constrained optimization; interval censoring; semi-parametric proportional hazard model.

1 Introduction

In this paper, we consider the problem of fitting semi-parametric proportional hazards (PH) models, also known as Cox regression, where observed survival times contain event times and left, right and interval censoring times. This type of data is also called the partly interval-censored data in [1]. We will present an efficient algorithm to compute the constrained maximum penalized likelihood (MPL) estimate of PH models where a penalty function is employed to regularize the baseline hazard estimate. This penalty function is also used to relax dependence between the estimates and the knots of the basis functions employed to approximate the baseline hazard. We will also provide asymptotic results for the constrained MPL estimates, accommodating possible active constraints, i.e., baseline hazard parameters equal to zero.

Likelihood based semi-parametric PH model estimation for survival data with interval censoring has been considered by many researchers; see, for example, [1–9], and references therein. Our method in this paper is designed to estimate the regression coefficients and baseline hazard of the semi-parametric PH model.

***Corresponding author: Jun Ma**, Department of Mathematics and Statistics, Macquarie University, Macquarie Park, Australia, E-mail: jun.ma@mq.edu.au
Dominique-Laurent Couturier, Cancer Research UK – Cambridge Institute, University of Cambridge, Cambridge, Cambridgeshire, UK; and MRC Biostatistics Unit, University of Cambridge, Cambridge, Cambridgeshire, UK
Stephane Heritier, School of Public Health and Preventive Medicine, Monash University, Melbourne, Australia
Ian C. Marschner, NHMRC Clinical Trials Centre, University of Sydney, Camperdown, Australia

For these existing methods, there are two common issues: (i) the non-negative constraint on the baseline hazard is not addressed properly; and (ii) the asymptotic covariance matrix for the regression coefficients estimates can be difficult to compute.

Issue (i) is often addressed by transformations. For example, Joly et al. [4] adopt squared coefficients for their baseline hazard approximation, and Cai and Betens [9] express the baseline hazard by exponentiating a spline function. However, transformations do not provide ideal solutions. Indeed, the square coefficient method can cause multiple local maxima to the objective function, resulting in algorithms sensitive to their initial values, and exponentiating a spline function does not allow portions of the baseline hazard to equal 0, which can happen with long survival times, as well as lead to computational hardship when updating the cumulative baseline hazard at each iteration after the spline coefficients are updated. This computational hardship is mainly due to the integration component of the cumulative baseline hazard which has no closed-form expressions in general, possibly leading to much longer computational time. Common approaches to addressing issue (ii) include bootstrapping (e.g. [5]) and efficient score functions (e.g. [10]). However, bootstrapping is computationally intensive, while efficient scores are generally difficult to compute.

Our approach is designed to address the above two common difficulties for likelihood estimates. In this paper we first present a computationally efficient procedure for constrained MPL estimation of the PH model, then we develop asymptotic properties for the constrained MPL estimates. As shown in our simulation study, the asymptotic results produce accurate standard error estimates for both regression coefficients and baseline hazard.

In Section 2 we develop an iterative scheme to solve this problem. Optimal smoothing using a marginal likelihood is explained in Section 3. Asymptotic properties for these estimates are presented in Section 4, with proofs given in Supplementary Material. Section 5 contains the results from a simulation study and also the results from a melanoma data application. Concluding remarks are included in Section 6.

2 Penalized likelihood estimation

For individual i, where $i = 1, \ldots, n$, let Y_i be the random variable representing the time to onset of the event-of-interest. The value for Y_i may not be obtained exactly and in this case we observe interval censored times denoted by a bivariate random vector $\boldsymbol{C}_i = (C_i^{\mathrm{L}}, C_i^{\mathrm{R}})^{\mathrm{T}}$, where $C_i^{\mathrm{L}} \geq 0$, $C_i^{\mathrm{R}} > C_i^{\mathrm{L}}$ and the superscript T denotes matrix transpose. Note that it is possible for $C_i^{\mathrm{L}} = 0$ or $C_i^{\mathrm{R}} = +\infty$, corresponding to left or right censoring respectively. For any i, we can observe either Y_i or $\boldsymbol{C}_i$ but not both. If $\boldsymbol{C}_i$ is observed then we know $Y_i \in (C_i^{\mathrm{L}}, C_i^{\mathrm{R}})$. We assume independent interval censoring given the covariates. More specifically, this independent assumption means that $P(T_i \leq t | C_i^{\mathrm{L}} = c_i^{\mathrm{L}}, C_i^{\mathrm{R}} = c_i^{\mathrm{R}}, C_i^{\mathrm{L}} \leq T_i < C_i^{\mathrm{R}}) = P(T_i \leq t | c_i^{\mathrm{L}} \leq T_i < c_i^{\mathrm{R}})$ (refer to Chapter 1 of [6] and [11] for details).

Let δ_i be the indicator for event times and δ_i^{R}, δ_i^{L} and δ_i^{I} be the indicators for right, left and interval censoring times respectively. Clearly, $\delta_i = 1 - \delta_i^{\mathrm{R}} - \delta_i^{\mathrm{L}} - \delta_i^{\mathrm{I}}$. In this paper, we denote the observations for subject i by $(t_i^{\mathrm{L}}, t_i^{\mathrm{R}}, \mathbf{x}_i)$, where $(t_i^{\mathrm{L}}, t_i^{\mathrm{R}})$ can represent the observed event time (indicated by $t_i^{\mathrm{L}} = t_i^{\mathrm{R}}$) or censoring times and $\mathbf{x}_i$ is a p-row vector for values on covariates. Clearly, if $t_i^{\mathrm{L}} = 0$ the corresponding t_i^{R} is the left censoring time, while $t_i^{\mathrm{R}} = \infty$ gives t_i^{L} the right censoring time. Note that for the cases of left, right and event time, effectively only a single time point is involved, and so we can denote them by a single time t_i (associated with δ_i^{L}, δ_i^{R} or δ_i) when there is no confusion.

From the given observations, we wish to estimate the PH model:

$$h(t|\mathbf{x}_i) = h_0(t)\exp(\mathbf{x}_i\boldsymbol{\beta}), \tag{1}$$

where $h(t|\mathbf{x}_i)$ denotes the hazard function for individual i, $h_0(t)$ represents the baseline hazard and $\boldsymbol{\beta}$ is a p-vector of regression coefficients. Let $\mathbf{X}$ be the model matrix whose ith row is given by $\mathbf{x}_i$. Clearly, it requires $h_0(t) \geq 0$ so that both $h_0(t)$ and $h(t|\mathbf{x}_i)$ are valid hazard functions. Let $H_0(t)$ be the cumulative baseline hazard function. Then, $H(t|\mathbf{x}_i) = H_0(t)\exp(\mathbf{x}_i\boldsymbol{\beta})$ is the cumulative hazard function and $S(t|\mathbf{x}_i) = \exp(-H(t|\mathbf{x}_i))$ is the survival function. Note that, to simplify notations, we will remove $\mathbf{x}_i$ from $h(t|\mathbf{x}_i)$, $S(t|\mathbf{x}_i)$ and $H(t|\mathbf{x}_i)$ when there is no confusion, and respectively denote them by $h_i(t)$, $S_i(t)$ and $H_i(t)$.

In this paper we consider the MPL method to fit model (1), where $h_0(t)$ and $\boldsymbol{\beta}$ are estimated simultaneously and a penalty is used to regularize the $h_0(t)$ estimate. Since $h_0(t)$ is an infinite dimensional parameter, its estimation from a finite number of observations is ill-conditioned. This problem can be addressed by approximating $h_0(t)$ using a finite number of non-negative basis functions (e.g. [4]), that is

$$h_0(t) = \sum_{u=1}^{m} \theta_u \psi_u(t), \tag{2}$$

where $\psi_u(t) \geq 0$ are basis functions and m can vary with sample size n but with a slower rate in that $m \to \infty$ as $n \to \infty$ but $m/n \to 0$. Note that m is related to the number of knots defining the basis functions. Possible choices for basis functions include indicator functions, M-splines and Gaussian density functions.

The requirement that $h_0(t) \geq 0$ can now be imposed more simply through $\boldsymbol{\theta} \geq 0$, where $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_m)^{\mathrm{T}}$ and $\boldsymbol{\theta} \geq 0$ is interpreted element-wise. Approximation using basis functions in PH models has been adopted by many authors, including Zhang et al. [8] for spline based sieve maximum likelihood estimation, Cai and Betensky [9] and Joly et al. [4] for respectively penalized linear spline and M-spline based MPL estimation, and Ma, Heritier, and Lô [12] for constrained MPL estimation.

Using the notations introduced above, the log-likelihood for observation i is

$$\begin{aligned} l_i(\boldsymbol{\beta}, \boldsymbol{\theta}) = {} & \delta_i(\log\, h_0(t_i) + \mathbf{x}_i\boldsymbol{\beta} + \log\, S_i(t_i)) + \delta_i^{\mathrm{R}}\, \log\, S_i(t_i) \\ & + \delta_i^{\mathrm{L}}\, \log(1 - S_i(t_i)) + \delta_i^{\mathrm{I}}\, \log(S_i(t_i^L) - S_i(t_i^{\mathrm{R}})). \end{aligned} \tag{3}$$

The log-likelihood from the entire data set is then

$$l(\boldsymbol{\beta}, \boldsymbol{\theta}) = \sum_{i=1}^{n} l_i(\boldsymbol{\beta}, \boldsymbol{\theta}). \tag{4}$$

In this paper we introduce an efficient algorithm to compute the penalized likelihood estimate of $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ where a penalty function is used to regularize/smooth the $h_0(t)$ estimate. Another reason for a penalty is that it will dampen non-important θ_u to 0 and hence relaxes the potential influence of the number and location of knots on the estimates. In fact, a penalty plays an important role for facilitating a stable computational algorithm for our problem.

Thus, the MPL estimation of $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ is a constrained optimization problem:

$$(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\theta}}) = \operatorname*{argmax}_{\boldsymbol{\beta}, \boldsymbol{\theta}} \{\Phi(\boldsymbol{\beta}, \boldsymbol{\theta}) = l(\boldsymbol{\beta}, \boldsymbol{\theta}) - \lambda J(\boldsymbol{\theta})\}, \tag{5}$$

subject to $\boldsymbol{\theta} \geq 0$. In (5), $\lambda \geq 0$ is the smoothing parameter and $J(\boldsymbol{\theta})$ is the penalty function to regularize the $h_0(t)$ estimate. For example, the well known roughness penalty [13] is $J(\boldsymbol{\theta}) = \int h_0''(t)^2 \mathrm{d}t = \boldsymbol{\theta}^{\mathrm{T}}\mathbf{R}\boldsymbol{\theta}$, where matrix $\mathbf{R}$ has the dimension of $m \times m$ with its (u, v)th element $r_{uv} = \int \psi_u''(t)\psi_v''(t)\mathrm{d}t$. Once $\widehat{\boldsymbol{\theta}}$ is obtained, the MPL estimate of $h_0(t)$ is $\widehat{h}_0(t) = \sum_{u=1}^{m} \widehat{\theta}_u \psi_u(t)$.

We propose an algorithm similar to the one developed in [12] to find the required estimates for $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$, where $\boldsymbol{\theta} \geq 0$. This is an alternating iterative method where each iteration involves two steps: firstly, $\boldsymbol{\beta}$ is updated using the Newton algorithm, and then $\boldsymbol{\theta}$ is computed from the multiplicative iterative (MI) algorithm (e.g. [14]) which produces estimates satisfying the non-negativity constraint.

The Karush–Kuhn–Tucker (KKT) conditions for the constrained optimization (5) are $\partial\Phi/\partial\beta_j = 0$, and $\partial\Phi/\partial\theta_u = 0$ if $\theta_u > 0$ and $\partial\Phi/\partial\theta_u < 0$ if $\theta_u = 0$, where $j = 1, \ldots, p$ and $u = 1, \ldots, m$.

In our algorithm, the vector $\boldsymbol{\beta}$ is first updated by the Newton algorithm at each iteration as follows:

$$\boldsymbol{\beta}^{(k+1)} = \boldsymbol{\beta}^{(k)} + \omega_1^{(k)} \left[-\frac{\partial^2\Phi(\boldsymbol{\beta}^{(k)}, \boldsymbol{\theta}^{(k)})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^T}\right]^{-1} \frac{\partial\Phi(\boldsymbol{\beta}^{(k)}, \boldsymbol{\theta}^{(k)})}{\partial\boldsymbol{\beta}}, \tag{6}$$

where $\omega_1^{(k)} \in (0, 1]$ is a line search step size for $\Phi(\boldsymbol{\beta}^{(k+1)}, \boldsymbol{\theta}^{(k)}) \geq \Phi(\boldsymbol{\beta}^{(k)}, \boldsymbol{\theta}^{(k)})$. Expressions for the first and the second derivatives in (6) are available in Supplementary Material. Next, $\boldsymbol{\theta}$ is updated by the MI algorithm to give

$$\boldsymbol{\theta}^{(k+1)} = \boldsymbol{\theta}^{(k)} + \omega_2^{(k)} \mathbf{D}^{(k)} \frac{\partial \Phi(\boldsymbol{\beta}^{(k+1)}, \boldsymbol{\theta}^{(k)})}{\partial \boldsymbol{\theta}}, \tag{7}$$

where $\mathbf{D}^{(k)}$ is a diagonal matrix with diagonals $\theta_u^{(k)}/d_u^{(k)}$ for $u = 1, \ldots, m$, where

$$d_u^{(k)} = \sum_{i=1}^{n} \delta_i \Psi_u(t_i) \exp(\mathbf{x}_i \boldsymbol{\beta}^{(k+1)}) + \sum_{i=1}^{n} \delta_i^{\mathrm{R}} \Psi_u(t_i) \exp(\mathbf{x}_i \boldsymbol{\beta}^{(k+1)})$$
$$+ \sum_{i=1}^{n} \delta_i^{\mathrm{I}} \frac{S_i^{(k)}(t_i^{\mathrm{L}}) \Psi_u(t_i^{\mathrm{L}})}{S_i^{(k)}(t_i^{\mathrm{L}}) - S_i^{(k)}(t_i^{\mathrm{R}})} \exp(\mathbf{x}_i \boldsymbol{\beta}^{(k+1)}) + \lambda \left[\frac{\partial J(\boldsymbol{\theta}^{(k)})}{\partial \theta_u} \right]^{+} + \xi_u.$$

Here, $\Psi_u(t) = \int_0^t \psi_u(w) \mathrm{d}w$, $[a]^+ = \max\{0, a\}$ and ξ_u is a small non-negative constant used to avoid the possibility of zero d_u; the size of ξ_u will not affect the final solution of this algorithm but will affect its convergence speed. In (7), $\omega_2^{(k)} \in (0, 1]$ is a line search step size. Since this algorithm involves both Newton and MI steps, we call it the Newton–MI algorithm. Step sizes $\omega_1^{(k)}$ and $\omega_2^{(k)}$ can be determined efficiently by, for example, the Armijo rule [15]; see also [16].

Following the same argument as in [14] we can show that (i) if $\boldsymbol{\theta}^{(k)}$ is non-negative then $\boldsymbol{\theta}^{(k+1)}$ is also non-negative, and (ii) under certain regularity conditions, this Newton–MI algorithm converges to a solution satisfying the KKT conditions.

3 Smoothing parameter estimation

Automatic smoothing parameter selection is pivotal for successful implementation of the penalized likelihood parameter estimation, particularly for users who are less experienced with manual selection of smoothing values.

Akaike's information criterion (AIC), cross validation (CV) or generalized cross validation (GCV) are all possible choices for estimating smoothing parameter λ, but in this section we particular apply the marginal likelihood approach (e.g. [9]).

We adopt a quadratic penalty function $J(\boldsymbol{\theta}) = \boldsymbol{\theta}^{\mathrm{T}} \mathbf{R} \boldsymbol{\theta}$, which can be related to the normal prior distribution for $\boldsymbol{\theta}$, $N(0_{m\times 1}, \sigma_\theta^2 \mathbf{R}^{-1})$, where $\sigma_\theta^2 = 1/2\lambda$. Thus, after omitting the terms independent of $\boldsymbol{\beta}$, $\boldsymbol{\theta}$ and σ_θ^2, the log-posterior is

$$l_p(\boldsymbol{\beta}, \boldsymbol{\theta}) = -\frac{m}{2} \log \sigma_\theta^2 + l(\boldsymbol{\beta}, \boldsymbol{\theta}) - \frac{1}{2\sigma_\theta^2} \boldsymbol{\theta}^{\mathrm{T}} \mathbf{R} \boldsymbol{\theta}. \tag{8}$$

The log-marginal likelihood for σ_θ^2 (after integrating out $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$) is

$$l_m(\sigma_\theta^2) = -\frac{m}{2} \log \sigma_\theta^2 + \log \int \exp\left(l(\boldsymbol{\beta}, \boldsymbol{\theta}) - \frac{1}{2\sigma_\theta^2} \boldsymbol{\theta}^{\mathrm{T}} \mathbf{R} \boldsymbol{\theta} \right) \mathrm{d}\boldsymbol{\beta} \mathrm{d}\boldsymbol{\theta}. \tag{9}$$

After applying Laplace's approximation and plugging-in the MPL estimates for $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$, we have

$$l_m(\sigma_\theta^2) \approx -\frac{m}{2} \log \sigma_\theta^2 + l(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\theta}}) - \frac{1}{2\sigma_\theta^2} \widehat{\boldsymbol{\theta}}^{\mathrm{T}} R \widehat{\boldsymbol{\theta}} - \frac{1}{2} \log \left| \widehat{\mathbf{G}} + \mathbf{Q}(\sigma_\theta^2) \right|, \tag{10}$$

where $\mathbf{G}$ is the negative Hessian from the log-likelihood $l(\boldsymbol{\beta}, \boldsymbol{\theta})$, $\widehat{\mathbf{G}}$ is $\mathbf{G}$ evaluated at $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\theta}}$, and

$$\mathbf{Q}(\sigma_\theta^2) = \begin{pmatrix} 0 & 0 \\ 0 & \frac{1}{\sigma_\theta^2} \mathbf{R} \end{pmatrix}.$$

The solution of σ_θ^2 maximizing (10) satisfies

$$\widehat{\sigma}_\theta^2 = \frac{\widehat{\boldsymbol{\theta}}^{\mathrm{T}} \mathbf{R} \widehat{\boldsymbol{\theta}}}{m - \nu}, \tag{11}$$

where $\nu = \mathrm{tr}\{(\widehat{\mathbf{G}} + \mathbf{Q}(\widehat{\sigma}_\theta^2))^{-1} \mathbf{Q}(\widehat{\sigma}_\theta^2)\}$, which can be conceived as the model degrees of freedom. Active constraints of $\boldsymbol{\theta} \geq 0$ may bring difficulties to the matrix inversion in ν. In this case, a modification similar to the covariance matrix in Theorem 2 (using matrix $\mathbf{U}$) can be adopted. A recommended procedure for identifying active constraints is available in Remark 2 after Theorem 2.

Since $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ depend on σ_θ^2, expression (11) naturally suggests an iterative procedure: with σ_θ^2 being fixed at the its current estimate, the corresponding MPL estimates of $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ are obtained, and then σ_θ^2 is updated by (11) with the new $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\theta}}$. This process is continued until the degree-of-freedom ν is stabilized. Results in Section 5 reveal that this iterative procedure usually converges quickly.

4 Asymptotic properties

In this section we present asymptotic results; these results are useful for conducting large sample inferences without resorting to computer-intensive methods such as bootstrapping.

Let $(\boldsymbol{\beta}_0, h_{00}(t))$ be the true parameters and $(\widehat{\boldsymbol{\beta}}, \widehat{h}_0(t))$ the MPL estimates. Theorem 1 below provides asymptotic consistency for $(\widehat{\boldsymbol{\beta}}, \widehat{h}_0(t))$ when the number of basis functions $m \to \infty$ but $m/n \to 0$ and the scaled smoothing value $\mu_n = \lambda/n \to 0$ when $n \to \infty$. Let a and b be respectively the minimum and maximum of all the observed survival times (including interval censoring but excluding 0 and ∞).

Theorem 1. *Assume Assumptions A1–A4 in* Supplementary Material *hold. Assume $h_0(t)$ is bounded and has up to $r \geq 1$ derivatives over $[a, b]$. Assume $m = n^\upsilon$, where $0 < \upsilon < 1$, and $\mu_n \to 0$ as $n \to \infty$. Then, when $n \to \infty$,*

(1) $\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\| \to 0$ *almost surely, and*
(2) $\sup_{t \in [a,b]} |\widehat{h}_0(t) - h_{00}(t)| \to 0$ *almost surely.*

Proof. See Supplementary Material Section S2.1. □

If following, for example [3, 8], or [17]; asymptotic normality results for $\widehat{\boldsymbol{\beta}}$ can be formulated from the efficient score function. An efficient score can be interpreted as the orthogonal projection of the score for $\boldsymbol{\beta}$ onto the space spanned by the score for $h_0(t)$ (for example, [18, 19] and [20, Chapter 2]). However, one issue with this method is that efficient scores are generally difficult to compute. Moreover, people may also wish to make inferences on quantities such as survival probabilities, thus it can be useful if the approximated $h_0(t)$ is included in the asymptotic normality. This motivates us to develop asymptotic normality results for $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\theta}}$ jointly.

One noticeable problem is that, if m increases with n, the length of $\boldsymbol{\theta}$ (i.e. m) can go to ∞ when $n \to \infty$. We avoid this by fixing m to a finite number. This strategy also appears in [21] for semi-parametric partially linear single index models. Yu and Ruppert [21] comment that this approach lies somewhere between parametric and nonparametric modeling. A fixed m still allows the approximation given in (2) flexible enough to adapt to different baseline hazard functions. The penalty function is used to avoid over fitting. Therefore, m remains in the asymptotic normal distribution. The MPL estimates $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\theta}}$ can still achieve $\sqrt{n}$ convergence rate in this context. The simulation results reported in Table 5 demonstrate that the error in the estimation of $h_0(t)$ is small when compared with its competitors.

The MPL estimate of $\boldsymbol{\theta}(\geq 0)$ often encounters active constraints: i.e. $\theta_u = 0$ for some u. This happens particularly when the number of knots is bigger than it should be. In this case, the penalty function will dampen unnecessary θ_u's to zero. In Remark 2 below, we explain how to identify active constraints. Active constraints must be taken into consideration in the asymptotic covariance matrix as otherwise unpleasant results, such as negative variances, can be produced [22]. In the following discussions, we follow [23] to develop our asymptotic normality results for $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\theta}}$.

Let $\boldsymbol{\eta} = (\boldsymbol{\theta}^{\mathrm{T}}, \boldsymbol{\beta}^{\mathrm{T}})^{\mathrm{T}}$, whose length is a finite number $m + p$. The penalized likelihood in (5) can be expressed in $\boldsymbol{\eta}$ as

$$\Phi(\boldsymbol{\eta}) = l(\boldsymbol{\eta}) - \lambda J(\boldsymbol{\eta}),$$

where $J(\boldsymbol{\eta}) = J(\boldsymbol{\theta})$. The MPL estimate of $\boldsymbol{\eta}$, denoted by $\widehat{\boldsymbol{\eta}}$, is obtained by maximizing $\Phi(\boldsymbol{\eta})$ with the constraint $\boldsymbol{\theta} \geq 0$. Let $\boldsymbol{\eta}_0$ represent the 'true value' of parameter $\boldsymbol{\eta}$.

We frequently experience active constraints when estimating $\boldsymbol{\theta}$ so this fact must be allowed when developing asymptotic results. To elucidate discussions we assume, without loss of generality, that the first q of $\boldsymbol{\theta} \geq 0$ constraints are active in the MPL solution. Correspondingly, define

$$\mathbf{U} = [\mathbf{O}_{(m-q+p)\times q}, \mathbf{I}_{(m-q+p)\times(m-q+p)}]^{\mathrm{T}}, \tag{12}$$

which satisfies $\mathbf{U}^{\mathrm{T}}\mathbf{U} = \mathbf{I}_{(m-q+p)\times(m-q+p)}$. In (12), $\mathbf{O}$ denotes a matrix of zeros and $\mathbf{I}$ represents an identity matrix.

Theorem 2. *Assume Assumptions B1–B5 in* Supplementary Material *hold and the number of basis functions m is fixed. Assume the scaled smoothing value* $\mu_n = o(n^{1/2})$. *Assume there are q active constraints in the MPL estimate of* $\boldsymbol{\theta}$. *Let* $\mathbf{F}(\boldsymbol{\eta}) = -E_{\boldsymbol{\eta}_0}[\lim_{n\to\infty} n^{-1}\partial^2 l(\boldsymbol{\eta})/\partial\boldsymbol{\eta}\partial\boldsymbol{\eta}^{\mathrm{T}}]$. *Then, when* $n \to \infty$, $\sqrt{n}(\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta}_0)$ *converges in distribution to* $N(\mathbf{0}, \widetilde{\mathbf{F}}(\boldsymbol{\eta}_0)^{-1})$, *where* $\widetilde{\mathbf{F}}(\boldsymbol{\eta})^{-1} = \mathbf{U}(\mathbf{U}^{\mathrm{T}}\mathbf{F}(\boldsymbol{\eta})\mathbf{U})^{-1}\mathbf{U}^{\mathrm{T}}$.

Proof. See Supplementary Materials Section S2.2. □

Remark 1. We comment that matrix $\widetilde{\mathbf{F}}(\boldsymbol{\eta})^{-1}$ is in fact very easy to compute. Firstly, $\mathbf{U}^{\mathrm{T}}\mathbf{F}(\boldsymbol{\eta})\mathbf{U}$ is obtained simply by deleting the rows and columns of $\mathbf{F}(\boldsymbol{\eta})$ associated with the active constraints. Then $(\mathbf{U}^{\mathrm{T}}\mathbf{F}(\boldsymbol{\eta})\mathbf{U})^{-1}$ is calculated. Finally, $\widetilde{\mathbf{F}}(\boldsymbol{\eta})^{-1}$ is obtained by padding $(\mathbf{U}^{\mathrm{T}}\mathbf{F}(\boldsymbol{\eta})\mathbf{U})^{-1}$ with zeros in the deleted rows and columns.

Remark 2. To implement the results in Theorem 2 (and also for Corollary 1), one must identify active constraints. We suggest the following process for this task. At the end of running the Newton–MI algorithm, often some $\widehat{\theta}_u$ are exactly zero with negative gradients, so they are active. Also, some $\widehat{\theta}_u$ may be close to, but not exactly, zero. In this case, we check the corresponding gradient values to see if they are negative. In our R program, we adopt the criterion '$\widehat{\theta}_u < 10^{-3}$ and the corresponding gradient less than -10^{-2}' to define active constraints.

Finite sample (i.e. n is finite) inferences are required in practice. Corollary 1 below provides an approximate distribution for $\widehat{\boldsymbol{\eta}}$ when n is large.

Corollary 1. *Assume Assumption B1–B5 in* Supplementary Material *hold and assume the smoothing parameter* $\lambda \ll n$. *When n is large, the distribution for the MPL estimate* $\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta}_0$ *can be approximated by a multivariate normal distribution with mean zero and covariance matrix*

$$\widehat{var}(\widehat{\boldsymbol{\eta}}) = \mathbf{A}(\widehat{\boldsymbol{\eta}})^{-1}\frac{\partial^2 l(\widehat{\boldsymbol{\eta}})}{\partial\boldsymbol{\eta}\partial\boldsymbol{\eta}^T}\mathbf{A}(\widehat{\boldsymbol{\eta}})^{-1}, \tag{13}$$

where

$$\mathbf{A}(\widehat{\boldsymbol{\eta}})^{-1} = \mathbf{U}(\mathbf{U}^{\mathrm{T}}(\partial^2 l(\widehat{\boldsymbol{\eta}})/\partial\boldsymbol{\eta}\partial\boldsymbol{\eta}^{\mathrm{T}} + \lambda\partial^2 J(\widehat{\boldsymbol{\eta}})/\partial\boldsymbol{\eta}\partial\boldsymbol{\eta})\mathbf{U})^{-1}\mathbf{U}^{\mathrm{T}}$$

.

The results in Corollary 1 are useful in practice as they accommodate nonzero smoothing values and active constraints. These results allow us to make inferences with respect to, for example, regression coefficients, baseline hazard, cumulative baseline hazard and survival probabilities. Also, inferences on some predictive values can be made. The simulation results reported in Section 5 demonstrate that biases in the MPL estimates are usually negligible when smoothing values are smaller than $\sqrt{n}$.

5 Results

5.1 Simulation results

To assess the performance of our estimator, we conducted two simulations respectively considering Weibull and log-logistic baseline hazard functions. Observed survival times $(t_i^{\mathrm{L}}, t_i^{\mathrm{R}})$, including event and interval censoring times, were generated as follows. Let Y_i denote a simulated event time, π^{E} represents the chosen proportion for event times, U_i^{L}, and U_i^{E} are uniform $u(0,1)$ random numbers and U_i^{R} is a uniform $u(U_i^{\mathrm{L}},1)$ random number. Let γ_{L} and γ_{R} (with $\gamma_{\mathrm{L}} \le \gamma_{\mathrm{R}}$) be two positive scalars. For given π^{E}, γ_{L} and γ_{R}, we generated partly interval-censored survival times using the following procedure. For each i, we first generated uniform numbers U_i^{L}, U_i^{R} and U_i^{E} (uniform limits are specified above) and event time Y_i. If $U_i^{\mathrm{E}} < \pi^{\mathrm{E}}$ then we had an event time so that $t_i^{\mathrm{L}} = t_i^{\mathrm{R}} = Y_i$; otherwise, if $\gamma_{\mathrm{L}} U_i^{\mathrm{L}} \le Y_i \le \gamma_{\mathrm{R}} U_i^{\mathrm{R}}$ we had an interval censoring with $t_i^{\mathrm{L}} = \gamma_{\mathrm{L}} U_i$ and $t_i^{\mathrm{R}} = \gamma_{\mathrm{R}} U_i^{\mathrm{R}}$, if $Y_i < \gamma_{\mathrm{L}} U_i^{\mathrm{L}}$ we had a left censoring with $t_i^{\mathrm{L}} = 0$ and $t_i^{\mathrm{R}} = \gamma_{\mathrm{L}} U_i^{\mathrm{L}}$ and if $\gamma_{\mathrm{R}} U_i^{\mathrm{R}} < Y_i$ we had a right censoring with $t_i^{\mathrm{L}} = \gamma_{\mathrm{R}} U_i^{\mathrm{R}}$ and $t_i^{\mathrm{R}} = \infty$.

Table 1 provides more information for the simulations (associated results are in Tables 2–5 of this paper), including the regression coefficients, **X** matrix, baseline hazard function, sample size and event proportion for each simulation. The censoring proportion per censoring type is also indicated: Simulation 1 shows smaller left censoring proportions and Simulation 2 shows larger interval censoring proportions.

Table 1: Parameters used for the simulations, where $\mathbf{b}_1$ is a vector of Bernoulli variates with probability of success $\pi = 0.5$ (treatment predictor), and $\mathbf{u}_1$ and $\mathbf{u}_2$ are vectors for standard uniform random numbers.

	Simulation 1	**Simulation 2**
	Simulation parameters	
$\boldsymbol{\beta}$ vector	$\boldsymbol{\beta} = [0.75, -0.50, 0.25]^{\mathrm{T}}$	$\boldsymbol{\beta} = [0.25, 0.25]^{\mathrm{T}}$
X matrix	$\mathbf{X} = [\mathbf{b}_1, 5\mathbf{u}_1, 7\mathbf{u}_2]$	$\mathbf{X} = [\mathbf{b}_1, 7\mathbf{u}_1]$
Y distribution	Weibull	Log logistic
Baseline hazard	$h_0(t) = 3t^2$	$h_0(t) = 4e^2/[t(e^2 + t^{-4})]$
γ_{L} and γ_{R}	$\gamma_{\mathrm{L}} = 0.9, \gamma_{\mathrm{R}} = 1.3$	$\gamma_{\mathrm{L}} = 0.5, \gamma_{\mathrm{R}} = 1.1$
	Simulation scenarios	
Sample sizes	$n = 100, 500, 2000$	$n = 100, 500, 2000$
Percentages of events	$\pi^{\mathrm{E}} = 0\%, 25\%, 50\%$	$\pi^{\mathrm{E}} = 0\%, 25\%, 50\%$
	Repartition of the censored observations by censoring type	
Left censoring	17.90%	17.90%
Interval censoring	43.70%	60.80%
Right censoring	38.40%	21.40%
	Specific estimator parameters	
EM I-spline	3rd order I-splines	3rd order I-splines
	AIC optimized n_α for $n = 100, 500, 2000$	AIC optimized n_α for $n = 100, 500, 2000$
MPL M-spline	3rd order M-splines	3rd order M-splines
	$n_\alpha = 7, 9, 11$ for $n = 100, 500, 2000$	$n_\alpha = 7, 9, 11$ for $n = 100, 500, 2000$
MPL Gaussian	$\zeta_1 = 0.35, \zeta_2 = 0.4$	$\zeta_1 = 0.35, \zeta_2 = 0.4$
	$n_\alpha = 7, 9, 11$ for $n = 100, 500, 2000$	$n_\alpha = 7, 9, 11$ for $n = 100, 500, 2000$

Table 2: Simulation 1 results for $\boldsymbol{\beta}$, where $\boldsymbol{\beta} = [\beta_1, \beta_2, \beta_3]^\top = [0.75, -0.50, 0.25]^\top$. Asymptotic standard errors for the *convex minorant estimator* are missing due to unavailable inference for this estimator.

		$\pi^E = 0\%$			$\pi^E = 25\%$			$\pi^E = 50\%$		
		$n = 100$	$n = 500$	$n = 2000$	$n = 100$	$n = 500$	$n = 2000$	$n = 100$	$n = 500$	$n = 2000$
		Biases								
β_1	PL	−0.226	−0.240	−0.237	−0.160	−0.177	−0.174	−0.093	−0.115	−0.118
	CM	0.150	−0.001	−0.045	0.126	0.065	−0.065	0.027	−0.076	−0.199
	EM-I	0.047	−0.006	−0.008	0.042	−0.003	−0.001	0.048	0.000	−0.003
	MPL-M	−0.044	−0.029	−0.012	−0.078	−0.030	−0.010	−0.072	−0.027	−0.011
	MPL-G	−0.016	−0.029	−0.013	−0.033	−0.032	−0.013	−0.028	−0.027	−0.013
β_2	PL	−0.203	−0.226	−0.231	−0.141	−0.165	−0.170	−0.094	−0.109	−0.114
	CM	−0.119	−0.372	−0.487	−0.377	−0.470	−0.481	−0.353	−0.400	−0.394
	EM-I	0.065	0.009	−0.004	0.056	0.008	0.001	0.036	0.004	−0.002
	MPL-M	−0.032	−0.016	−0.009	−0.067	−0.019	−0.007	−0.081	−0.022	−0.009
	MPL-G	−0.003	−0.017	−0.010	−0.021	−0.020	−0.010	−0.036	−0.023	−0.011
β_3	PL	−0.207	−0.225	−0.235	−0.149	−0.168	−0.174	−0.090	−0.109	−0.117
	CM	0.142	−0.059	−0.153	0.066	−0.062	−0.180	−0.047	−0.161	−0.255
	EM-I	0.063	0.011	−0.008	0.053	0.007	−0.002	0.046	0.008	−0.003
	MPL-M	−0.027	−0.012	−0.011	−0.067	−0.020	−0.011	−0.074	−0.019	−0.011
	MPL-G	0.002	−0.012	−0.013	−0.021	−0.021	−0.014	−0.029	−0.020	−0.013
		Mean asymptotic and Monte Carlo (in brackets) standard errors								
β_1	PL	0.273	0.117	0.058	0.257	0.110	0.055	0.244	0.104	0.052
		(0.286)	(0.123)	(0.060)	(0.271)	(0.115)	(0.057)	(0.257)	(0.113)	(0.054)
	CM	–	–	–	–	–	–	–	–	–
		(0.357)	(0.121)	(0.059)	(0.269)	(0.108)	(0.060)	(0.230)	(0.103)	(0.052)
	EM-I	0.344	0.154	0.073	0.289	0.131	0.063	0.261	0.117	0.056
		(0.358)	(0.149)	(0.072)	(0.313)	(0.130)	(0.064)	(0.282)	(0.118)	(0.056)
	MPL-M	0.325	0.142	0.071	0.281	0.124	0.062	0.252	0.111	0.055
		(0.334)	(0.145)	(0.072)	(0.280)	(0.126)	(0.063)	(0.249)	(0.114)	(0.056)
	MPL-G	0.335	0.143	0.071	0.291	0.124	0.062	0.260	0.112	0.055
		(0.333)	(0.145)	(0.072)	(0.287)	(0.126)	(0.063)	(0.257)	(0.115)	(0.056)
β_2	PL	0.100	0.042	0.021	0.096	0.041	0.020	0.092	0.039	0.019
		(0.108)	(0.045)	(0.023)	(0.106)	(0.043)	(0.022)	(0.099)	(0.040)	(0.020)
	CM	–	–	–	–	–	–	–	–	–
		(0.140)	(0.066)	(0.032)	(0.104)	(0.043)	(0.021)	(0.100)	(0.040)	(0.020)
	EM-I	0.139	0.057	0.028	0.117	0.049	0.024	0.102	0.044	0.022
		(0.144)	(0.057)	(0.028)	(0.127)	(0.050)	(0.025)	(0.110)	(0.044)	(0.022)
	MPL-M	0.118	0.053	0.027	0.102	0.046	0.024	0.091	0.042	0.021
		(0.133)	(0.054)	(0.028)	(0.109)	(0.046)	(0.024)	(0.097)	(0.041)	(0.021)
	MPL-G	0.125	0.054	0.027	0.109	0.047	0.024	0.097	0.042	0.021
		(0.127)	(0.055)	(0.028)	(0.109)	(0.047)	(0.024)	(0.098)	(0.042)	(0.021)
β_3	PL	0.069	0.029	0.015	0.065	0.028	0.014	0.062	0.026	0.013
		(0.076)	(0.030)	(0.015)	(0.070)	(0.028)	(0.014)	(0.067)	(0.026)	(0.014)
	CM	–	–	–	–	–	–	–	–	–
		(0.082)	(0.031)	(0.017)	(0.061)	(0.026)	(0.014)	(0.059)	(0.024)	(0.013)
	EM-I	0.086	0.043	0.019	0.076	0.037	0.016	0.068	0.032	0.015
		(0.092)	(0.035)	(0.018)	(0.079)	(0.031)	(0.016)	(0.072)	(0.028)	(0.014)
	MPL-M	0.082	0.036	0.018	0.071	0.031	0.016	0.063	0.028	0.014
		(0.085)	(0.034)	(0.018)	(0.071)	(0.030)	(0.015)	(0.064)	(0.026)	(0.014)
	MPL-G	0.086	0.036	0.018	0.074	0.032	0.016	0.066	0.028	0.014
		(0.086)	(0.034)	(0.018)	(0.072)	(0.030)	(0.015)	(0.065)	(0.027)	(0.014)

Table 2: (continued)

		$\pi^E = 0\%$			$\pi^E = 25\%$			$\pi^E = 50\%$		
		$n = 100$	$n = 500$	$n = 2000$	$n = 100$	$n = 500$	$n = 2000$	$n = 100$	$n = 500$	$n = 2000$
						Root mean square errors				
β_1	PL	**0.576**	0.467	0.433	0.544	0.419	0.377	0.516	0.377	0.322
	CM	0.612	**0.348**	**0.261**	**0.533**	**0.344**	0.277	0.480	0.343	0.397
	EM-I	0.599	0.386	0.269	0.561	0.360	0.253	0.533	0.344	**0.237**
	MPL-M	0.579	0.383	0.269	0.535	0.357	**0.252**	**0.504**	**0.340**	**0.237**
	MPL-G	0.577	0.383	0.269	0.537	0.358	**0.252**	0.507	0.341	0.238
β_2	PL	0.385	0.348	0.343	0.357	0.305	0.296	0.332	0.261	0.246
	CM	0.390	0.444	0.495	0.464	0.489	0.491	0.450	0.451	0.445
	EM-I	0.384	0.239	0.169	0.360	0.223	0.157	0.334	0.209	**0.148**
	MPL-M	0.366	**0.234**	**0.168**	0.337	**0.217**	**0.155**	0.325	**0.207**	**0.148**
	MPL-G	**0.356**	0.235	0.169	**0.331**	0.219	0.156	**0.315**	0.209	0.149
β_3	PL	0.303	0.252	0.246	0.282	0.224	0.214	0.265	0.195	0.180
	CM	0.299	0.184	0.204	0.251	0.173	0.217	0.245	0.217	0.255
	EM-I	0.306	0.188	**0.135**	0.283	0.176	**0.125**	0.270	0.167	**0.119**
	MPL-M	**0.293**	**0.185**	0.136	0.270	**0.173**	**0.125**	0.259	**0.164**	**0.119**
	MPL-G	**0.293**	0.186	0.136	**0.269**	0.174	**0.125**	**0.256**	0.165	0.120
						95% coverage probabilities				
β_1	PL	**0.889**	**0.647**	**0.141**	0.907	**0.755**	**0.344**	0.919	**0.840**	**0.591**
	CM	–	–	–	–	–	–	–	–	–
	EM-I	0.911	0.945	0.952	0.900	0.935	0.950	**0.896**	0.938	0.951
	MPL-M	0.951	0.945	0.953	0.955	0.951	0.945	0.953	0.938	0.943
	MPL-G	0.952	0.943	0.952	0.959	0.948	0.946	0.951	0.941	0.944
β_2	PL	**0.786**	**0.242**	**0.002**	**0.841**	**0.459**	**0.018**	**0.887**	**0.692**	**0.182**
	CM	–	–	–	–	–	–	–	–	–
	EM-I	0.953	0.949	0.948	0.923	0.948	0.940	0.935	0.940	0.960
	MPL-M	0.925	0.945	0.944	0.927	0.939	0.934	0.901	0.937	0.945
	MPL-G	0.951	0.946	0.944	0.950	0.935	0.935	0.942	0.938	0.948
β_3	PL	**0.849**	**0.501**	**0.029**	**0.888**	**0.667**	**0.124**	0.923	**0.819**	**0.400**
	CM	–	–	–	–	–	–	–	–	–
	EM-I	**0.836**	0.956	0.949	**0.856**	0.965	0.971	**0.872**	0.961	0.963
	MPL-M	0.945	0.959	0.953	0.946	0.952	0.958	0.933	0.956	0.949
	MPL-G	0.952	0.961	0.948	0.956	0.953	0.959	0.956	0.954	0.949

Smallest root mean square errors per scenario as well as coverage probabilities below 0.9 appear in bold.

For MPL estimations in the simulations, we approximate $h_0(t)$ by either third order M-spline or Gaussian basis functions, and their expressions are given in Supplementary Material. The smoothing parameter was selected automatically as described in Section 3. We chose to place (interior) knots at quantiles of the finite end points (excluding 0 and ∞) of the observed survival intervals boundaries or events (when available), as this strategy performs better than equally spaced knots to our experience. The optimal number of knots can be selected using a criterion such as AIC or BIC or GCV. However, as a result of the penalty function, the MPL approach introduces certain degree of robustness to the knot selection. Therefore, in this simulation, we did not optimize the number of knots (denoted by n_α) for the MPL estimates as n_α has a rather low impact on the $\boldsymbol{\beta}$ and $h_0(t)$ estimates as long as n_α is not too large. The last portion of Table 1 exhibits parameters used for M-spline and Gaussian basis. To assess the influence of n_α on the performance of the MPL estimator, we also

Table 3: Simulation 2 results for $\boldsymbol{\beta}$, where $\boldsymbol{\beta} = [\beta_1, \beta_2]^\top = [0.25, 0.25]^\top$. Asymptotic standard errors for the *convex minorant estimator* are missing due to unavailable inference for this estimator. Smallest root mean square errors per scenario as well as.

		$\pi^E = 0\%$			$\pi^E = 25\%$			$\pi^E = 50\%$		
		$n = 100$	$n = 500$	$n = 2000$	$n = 100$	$n = 500$	$n = 2000$	$n = 100$	$n = 500$	$n = 2000$
					Biases					
β_1	PL	−0.426	−0.431	−0.439	−0.332	−0.357	−0.366	−0.275	−0.273	−0.282
	CM	0.464	0.702	0.824	1.199	1.354	0.972	0.807	0.604	0.122
	EM-I	−0.052	−0.057	−0.093	0.016	0.012	−0.007	0.006	0.023	0.002
	MPL-M	−0.016	0.044	0.018	0.002	0.045	0.017	−0.010	0.034	0.011
	MPL-G	−0.028	0.019	0.005	0.022	0.025	0.011	−0.003	0.017	0.004
β_2	PL	−0.444	−0.462	−0.463	−0.369	−0.387	−0.389	−0.263	−0.294	−0.299
	CM	0.055	−0.107	−0.176	0.119	−0.042	−0.193	−0.076	−0.238	−0.387
	EM-I	−0.051	−0.080	−0.106	−0.004	−0.016	−0.023	0.024	0.004	−0.007
	MPL-M	−0.038	0.017	0.006	−0.020	0.013	0.002	−0.004	0.014	0.001
	MPL-G	−0.043	−0.006	−0.005	0.001	−0.005	−0.004	0.012	−0.001	−0.004
					Mean asymptotic and Monte Carlo (in brackets) standard errors					
β_1	PL	0.236	0.102	0.051	0.228	0.099	0.049	0.221	0.096	0.048
		(0.257)	(0.104)	(0.053)	(0.249)	(0.101)	(0.051)	(0.234)	(0.101)	(0.050)
	CM	–	–	–	–	–	–	–	–	–
		(0.343)	(0.114)	(0.054)	(0.255)	(0.096)	(0.059)	(0.207)	(0.089)	(0.055)
	EM-I	0.333	0.140	0.079	0.285	0.125	0.065	0.235	0.109	0.058
		(0.326)	(0.132)	(0.065)	(0.290)	(0.124)	(0.061)	(0.255)	(0.108)	(0.054)
	MPL-M	0.334	0.146	0.072	0.287	0.124	0.061	0.252	0.109	0.054
		(0.333)	(0.147)	(0.074)	(0.284)	(0.128)	(0.062)	(0.250)	(0.110)	(0.055)
	MPL-G	0.337	0.145	0.072	0.291	0.124	0.061	0.255	0.109	0.054
		(0.325)	(0.143)	(0.073)	(0.288)	(0.125)	(0.062)	(0.251)	(0.108)	(0.054)
β_2	PL	0.060	0.026	0.013	0.058	0.025	0.012	0.057	0.024	0.012
		(0.062)	(0.026)	(0.013)	(0.063)	(0.026)	(0.013)	(0.062)	(0.026)	(0.013)
	CM	–	–	–	–	–	–	–	–	–
		(0.077)	(0.026)	(0.014)	(0.054)	(0.024)	(0.015)	(0.050)	(0.021)	(0.014)
	EM-I	0.085	0.048	0.029	0.077	0.049	0.021	0.057	0.040	0.020
		(0.077)	(0.034)	(0.017)	(0.075)	(0.033)	(0.017)	(0.066)	(0.029)	(0.015)
	MPL-M	0.084	0.039	0.020	0.073	0.033	0.017	0.064	0.029	0.015
		(0.088)	(0.039)	(0.020)	(0.074)	(0.033)	(0.017)	(0.064)	(0.029)	(0.015)

Table 3: (continued)

		$\pi^E = 0\%$			$\pi^E = 25\%$			$\pi^E = 50\%$		
		$n = 100$	$n = 500$	$n = 2000$	$n = 100$	$n = 500$	$n = 2000$	$n = 100$	$n = 500$	$n = 2000$
	MPL-G	0.087	0.039	0.020	0.076	0.033	0.017	0.067	0.029	0.015
		(0.084)	(0.038)	(0.020)	(0.076)	(0.033)	(0.017)	(0.066)	(0.028)	(0.015)
Root mean square errors										
β_1	PL	0.527	0.386	0.349	0.512	0.367	0.324	0.494	0.350	0.294
	CM	0.601	0.458	0.461	0.627	0.593	0.500	0.538	0.418	0.251
	EM-I	0.571	**0.364**	**0.263**	0.539	**0.352**	**0.246**	0.505	0.329	**0.233**
	MPL-M	0.577	0.384	0.272	**0.533**	0.359	0.249	**0.500**	0.332	0.234
	MPL-G	**0.570**	0.379	0.269	0.537	0.354	0.248	0.501	**0.328**	**0.233**
β_2	PL	0.356	0.344	0.341	0.334	0.316	0.313	0.300	0.279	0.275
	CM	**0.280**	0.193	0.215	**0.248**	**0.162**	0.225	0.230	0.252	0.313
	EM-I	**0.280**	0.200	0.177	0.273	0.181	0.132	0.257	0.170	**0.123**
	MPL-M	0.298	0.198	0.141	0.273	0.182	**0.129**	**0.252**	0.170	**0.123**
	MPL-G	0.291	**0.196**	**0.140**	0.275	0.181	**0.129**	0.256	**0.169**	**0.123**
95% coverage probabilities										
β_1	PL	0.910	**0.813**	**0.403**	0.912	**0.850**	**0.520**	0.920	**0.880**	**0.664**
	CM	–	–	–	–	–	–	–	–	–
	EM-I	0.927	0.936	0.941	0.910	0.920	0.956	0.905	0.914	0.926
	MPL-M	0.963	0.945	0.936	0.954	0.948	0.941	0.953	0.953	0.944
	MPL-G	0.968	0.948	0.936	0.954	0.949	0.944	0.956	0.957	0.949
β_2	PL	**0.516**	**0.009**	**0.000**	**0.610**	**0.037**	**0.000**	**0.746**	**0.163**	**0.000**
	CM	–	–	–	–	–	–	–	–	–
	EM-I	**0.771**	**0.857**	**0.827**	**0.747**	0.952	0.963	**0.739**	**0.891**	0.946
	MPL-M	0.938	0.943	0.955	0.940	0.948	0.952	0.951	0.949	0.944
	MPL-G	0.956	0.950	0.955	0.949	0.950	0.947	0.953	0.952	0.938

coverage probabilities below 0.9 appear in bold.

applied different number of knots on some of the simulated data. The corresponding results are available in Table S4 of the Supplementary Material.

We compared the performance of MPL with M-spline basis (MPL-M) and with Gaussian basis (MPL-G) against some other semi-parametric competitors. In particular, we considered the following methods: (1) the partial likelihood (PL) estimator with the middle point to replace left or interval censoring; (2) the convex minorant (CM) estimator of [5]; which also provides piecewise constant estimation to the cumulative baseline hazard function; and (3) the more recent expectation-maximization I-spline (EM-I) estimator of [7]; which consists in a two-stage data augmentation algorithm. The partial likelihood, convex minorant and EM-I estimates were respectively obtained by means of the 'survival', 'intcox' and 'ICsurv' R packages. For EM-I, in order to achieve a fair comparison with our MPL results, we adopted the AIC based knots selection strategy indicated in [7, Section 2.2] to define the optimal number of knots and knot placement strategy. More specifically, for each simulated sample, we estimated the model parameters using 40 different knot sequences, where 20 different number of interior knots (ranged from 1 to 20) were used for both the equal- and quantile-spaced knot schemes. For each simulated sample, this computationally intensive procedure thus considered 40 sets of parameters, among which the one corresponding to the smallest AIC estimate was finally selected. The EM-I and CM estimators were developed for cases with 100% interval censoring, so that we used observed event time '$+/-$' a small epsilon (10^{-4}), as advised by the ICsurv package maintainer, to handle non-censored times. Results obtained using different (but small) epsilon values (like 10^{-3}) were very similar.

We generated 1000 samples for each combination of sample size and proportion of events. Our MPL estimator proved very reliable as we noted numerical issues in only 1 case (over 36,000 fits, i.e., 3 censoring levels $\times$ 3 sample sizes $\times$ 2 spline functions $\times$ 2 simulations $\times$ 1000 Monte Carlo samples) for a sample with $n = 100$ and 100% censoring when considering the Gaussian basis. This numerical issue was caused by a negative degree-of-freedom in the smoothing parameter formula (11). Reducing the number of knots from 7 to 6 solved the issue. The EM I-spline method suffers from numerical instability. On average, among 720,000 fits (i.e., 3 censoring levels $\times$ 3 sample sizes $\times$ 40 knot sequence $\times$ 2 simulations $\times$ 1000 Monte Carlo samples), no estimates could be obtained in 7.5% of the cases for reasons unclear to us. After selecting the best knot sequence based on the (non-missing) AIC criterion (i.e., 18,000 cases), parameter estimates were obtained in all cases but inference was missing in 22.9% of the cases for at least 1 parameter due to non-positive definite covariance matrices. This percentage was especially large with small sample sizes. In such cases, we understand that the penalty used by the MPL is of a great help in stabilizing the estimates.

Tables 2 and 3 report, for all the simulations, biases, mean of the asymptotic and Monte Carlo (in brackets) standard errors, root mean squared errors (RMSE) and coverage probabilities of the $\boldsymbol{\beta}$ estimates. For the RMSEs we highlighted the smallest values and for the coverage probabilities we highlighted the values smaller than 0.9, indicating a poor coverage. The method of partial likelihood with mid-point imputation displays large biases in all scenarios of both simulations, often leading to extremely poor coverage probabilities, such as for β_2 in Simulation 2. The fact that coverage probabilities decrease when the sample sizes increase is due to the combination of persistent biases and smaller standard errors, suggesting that this estimator is not (always) consistent with interval censored data. As asymptotic inference was not developed for the convex minorant method, no asymptotic standard errors are reported here. Similarly to the PL method, the CM estimator often showed large biases compensated by small Monte Carlo standard errors sometimes leading to the smallest RMSE in some cases. The presence of large biases and small Monte Carlo standard errors suggests that computationally intensive bootstrap-based confidence intervals would also lead to poor coverage probabilities. The EM I-spline estimator appears to have slightly larger standard errors for β_1 and β_3 when considering small sample sizes in Simulation 1, and shows poor coverage probabilities for β_2 for almost all sample sizes in Simulation 2 (when focusing on cases with available standard errors). We believe that the coverage probabilities of the EM-I estimator could be improved by adopting the strategy described in Section 4 when active constraints are present.

In general, the coverage probabilities of MPL confidence intervals tend to be close to the 95% nominal value in all simulations except for β_2 in Simulation 1 when sample sizes are small. Furthermore, with both small biases and standard errors, the MPL often obtained the smallest RMSE per scenario in both simulations. Compared to the EM-I, our proposed MPL method proved numerically more stable with inference available in all cases (through knot reduction in one case). It also proved faster as the penalty function automatically evokes a type of knots selection so that the computer intensive AIC-based knots sequence optimization was is not required. For the scenario of Simulation 1 considering a sample size $n = 100$ and 100% censoring, the median run times of the MPL and EM-I estimators were respectively equal to 1.8 and 6.8 CPU-seconds when considering the same 7 quantile knot sequence and tolerance value for convergence. When considering the AIC knots selection described above (and considering 40 different knot sequences), the EM-I median running time jumped to 771.4 CPU-seconds.

Results on $\widehat{h}_0(t)$ are available in Tables 4 and 5. These results compare the MPL estimates of baseline hazard with the Nelson–Aalen estimate for the partial likelihood method, the piecewise constant estimate for the convex minorant method and the estimate for the EM I-spline method. Estimates of $h_0(t)$ were not provided directly *per se* in the R packages of the competing methods where only cumulative hazard estimates were provided. Estimates of $h_0(t)$ based on the Nelson–Aalen and convex minorant piecewise constant cumulative hazard estimates were deduced by computing the difference in cumulative hazard estimates between consecutive time points. For the EM-I estimator, estimates of $h_0(t)$ were obtained by using the relationship between the M- and I-splines. These tables report biases, mean of the asymptotic standard errors and mean of Monte Carlo standard errors (in bracket), RMSE values and 95% confidence intervals for $\widehat{h}_0(t)$ at the 25th (t_1), 50th (t_2) and 75th (t_3) percentile, as well as the integrated discrepancy between the estimated and the true $h_0(t)$ over an interval $[0, t^*]$, defined as

$$D[\widehat{h}_0(t^*), h_0(t^*)] = \int_0^{t^*} \left|\widehat{h}_0(t) - h_0(t)\right| \mathrm{d}t,$$

where $t^\star$ corresponds to the 90th percentile. The smallest RMSE and integrated discrepancy values per scenario and poor coverage probabilities (smaller than 0.9) were highlighted.

In general the Nelson–Aalen and convex minorant estimators obtained larger biases and standard errors, leading to large RMSE and integrated discrepancy estimates, while the EM-I, MPL-M and MPL-G estimators led to small biases and small errors in most cases. Focusing on the EM-I and MPL estimators, we can note that the biases of the MPL were greater that the ones of the EM-I when estimating $h_0(t_3)$ while the standard errors of the EM-I appeared a bit larger than the ones of the MPL. The MPL-M estimator often got the smallest RMSE and integrated discrepancy estimates. Regarding coverage probabilities, no estimator performed well in all the cases. As expected, the coverage probabilities tend to improve when the sample size increases and/or the percentage of censoring decreases. In Simulation 1, the coverage levels of our MPL estimators are poor for $h_0(t_3)$ especially when the sample sizes are small. Simulation 2 shows again poor coverage levels for the MPL estimators at 75th percentile for small sample sizes. These can be caused by the fact that the numbers of events at t_3 are extremely small.

Figure S1 in Supplementary Material displays the baseline hazard estimates obtained by competing methods for the scenario considering of $\pi^{\mathrm{E}} = 0$ and $n = 500$. In both simulations, the PL and CM methods provided biased noisy estimates in agreement with the results of Tables 4 and 5. For the PL estimator, the bump observed around the quantile 0.1 of the simulated survival times is caused by the midpoint strategy to left censored data. More specifically, left censored events are shrunk towards zero with a maximum around 0.45 (which is half the maximum of 0.9 chosen for this type of censoring. Evidence of this is provided in Figure S2, which shows a comparison of the interval and left censored times with the real ones, as well as the estimated hazard and cumulative hazard estimates for a simulated dataset of Simulation 1). In contrast, both the EM-I and MPL methods provided reliable estimates. EM-I was the least biased method and the MPL $h_0(t)$ estimates appeared biased for large survival times. The MPL still obtained the smallest RMSE and integrated discrepancy statistics in most cases due to its smaller variance.

Finally, to assess the impact of the number of knots n_α on $\boldsymbol{\beta}$ and $h_0(t)$ estimates, we estimated the parameters of the data sets generated in Simulation 1 by considering two different sets of knots for $n = 100$

Table 4: Simulation 1 results for $h_0(t)$ for the 25th (t_1), 50th (t_2), and 75th (t_3) percentiles of T. Some results are missing for the *convex minorant estimator* as no inference was developed for this estimator, and for the *Nelson–Aalen estimator* as we did not estimate the variance for this estimator.

		$\pi^E = 0\%$			$\pi^E = 25\%$			$\pi^E = 50\%$		
		$n = 100$	$n = 500$	$n = 2000$	$n = 100$	$n = 500$	$n = 2000$	$n = 100$	$n = 500$	$n = 2000$
Biases										
$h_0(t_1)$	Nelson–Aalen	−0.211	−0.256	−0.269	−0.090	−0.037	−0.114	−0.052	−0.142	−0.030
	CM	−0.019	−0.155	−0.371	−0.420	−0.464	−0.438	−0.290	−0.259	−0.284
	EM-I	0.056	−0.002	−0.007	0.052	0.011	−0.002	0.025	−0.004	0.001
	MPL-M	0.133	0.042	0.020	0.126	0.029	0.012	0.107	0.016	0.009
	MPL-G	0.114	0.041	0.009	0.078	0.024	0.012	0.048	0.012	0.005
$h_0(t_2)$	Nelson–Aalen	0.134	−0.143	−0.305	0.138	−0.107	−0.178	0.120	−0.047	−0.128
	CM	0.300	−0.224	−0.371	−0.429	−0.315	−0.250	−0.226	−0.224	−0.258
	EM-I	0.164	0.012	0.001	0.100	0.009	0.001	0.056	−0.001	−0.003
	MPL-M	0.060	0.032	0.026	0.051	0.034	0.016	0.039	0.025	0.010
	MPL-G	0.071	0.064	0.033	0.071	0.036	0.011	0.044	0.018	0.009
$h_0(t_3)$	Nelson–Aalen	0.537	0.096	0.032	0.312	0.053	0.046	0.210	0.102	−0.068
	CM	0.416	−0.411	−0.601	−0.424	−0.594	−0.618	−0.384	−0.504	−0.550
	EM-I	0.468	0.042	0.000	0.229	0.026	0.007	0.148	0.018	0.000
	MPL-M	−0.072	−0.044	−0.014	−0.060	−0.003	0.008	−0.054	0.003	0.008
	MPL-G	0.002	−0.035	0.009	0.042	0.015	0.012	0.029	0.014	0.006
Mean asymptotic (Monte Carlo) standard errors										
$h_0(t_1)$	Nelson–Aalen	–	–	–	–	–	–	–	–	–
		(1.474)	(2.076)	(1.385)	(1.872)	(4.729)	(1.927)	(1.677)	(1.413)	(2.746)
	CM	–	–	–	–	–	–	–	–	–
		(1.436)	(1.326)	(0.920)	(0.680)	(0.551)	(0.552)	(1.050)	(0.850)	(0.682)
	EM-I	0.600	0.260	0.123	0.494	0.224	0.103	0.439	0.197	0.092
		(0.636)	(0.249)	(0.135)	(0.586)	(0.232)	(0.112)	(0.473)	(0.190)	(0.095)
	MPL-M	0.547	0.224	0.115	0.478	0.196	0.101	0.425	0.176	0.092
		(0.592)	(0.233)	(0.121)	(0.488)	(0.206)	(0.102)	(0.428)	(0.178)	(0.093)
	MPL-G	0.567	0.236	0.123	0.491	0.213	0.116	0.437	0.194	0.104
		(0.619)	(0.253)	(0.127)	(0.517)	(0.227)	(0.116)	(0.448)	(0.203)	(0.103)

Table 4: (continued)

		$\pi^E = 0\%$			$\pi^E = 25\%$			$\pi^E = 50\%$		
		$n = 100$	**$n = 500$**	**$n = 2000$**	**$n = 100$**	**$n = 500$**	**$n = 2000$**	**$n = 100$**	**$n = 500$**	**$n = 2000$**
$h_0(t_2)$	Nelson–Aalen	–	–	–	–	–	–	–	–	–
		(4.479)	(3.868)	(1.830)	(4.226)	(3.795)	(2.754)	(4.627)	(3.935)	(2.977)
	CM	–	–	–	–	–	–	–	–	–
		(4.134)	(2.766)	(2.980)	(1.618)	(1.827)	(2.316)	(2.361)	(1.677)	(2.492)
	EM-I	1.563	0.614	0.267	1.133	0.496	0.219	0.976	0.428	0.194
		(1.782)	(0.540)	(0.294)	(1.280)	(0.504)	(0.244)	(1.014)	(0.401)	(0.210)
	MPL-M	1.061	0.454	0.231	0.914	0.397	0.201	0.812	0.354	0.180
		(1.292)	(0.482)	(0.233)	(0.971)	(0.412)	(0.199)	(0.816)	(0.358)	(0.180)
	MPL-G	1.109	0.481	0.247	0.982	0.424	0.222	0.866	0.379	0.200
		(1.254)	(0.507)	(0.252)	(1.057)	(0.439)	(0.225)	(0.896)	(0.380)	(0.200)
$h_0(t_3)$	Nelson–Aalen	–	–	–	–	–	–	–	–	–
		(12.260)	(18.085)	(7.227)	(8.644)	(17.343)	(7.641)	(8.047)	(13.021)	(5.996)
	CM	–	–	–	–	–	–	–	–	–
		(9.568)	(3.735)	(3.773)	(3.341)	(2.569)	(2.154)	(3.437)	(2.215)	(2.276)
	EM-I	4.307	1.278	0.562	2.717	1.014	0.444	2.202	0.860	0.387
		(5.293)	(1.309)	(0.575)	(3.441)	(1.036)	(0.483)	(2.568)	(0.882)	(0.406)
	MPL-M	1.828	0.854	0.465	1.603	0.769	0.404	1.447	0.691	0.359
		(2.412)	(0.946)	(0.482)	(1.848)	(0.803)	(0.392)	(1.586)	(0.711)	(0.353)
	MPL-G	2.090	0.892	0.494	1.901	0.829	0.445	1.684	0.751	0.398
		(2.379)	(0.997)	(0.518)	(2.117)	(0.868)	(0.434)	(1.803)	(0.779)	(0.390)
					Root mean square errors					
$h_0(t_1)$	Nelson–Aalen	1.221	1.447	1.189	1.369	2.174	1.389	1.295	1.192	1.657
	CM	1.198	1.156	1.002	0.904	0.863	0.852	1.046	0.946	0.865
	EM-I	0.799	0.498	0.367	0.767	0.482	0.335	0.688	0.436	0.309
	MPL-M	**0.780**	**0.487**	**0.350**	**0.711**	**0.456**	**0.321**	**0.665**	**0.423**	**0.305**
	MPL-G	0.794	0.506	0.357	0.723	0.478	0.341	0.671	0.451	0.321
$h_0(t_2)$	Nelson–Aalen	2.118	1.969	1.397	2.058	1.950	1.668	2.152	1.983	1.729
	CM	2.046	1.676	1.758	1.370	1.398	1.543	1.553	1.322	1.599
	EM-I	1.348	0.735	0.542	1.139	0.710	0.494	1.010	0.633	0.458
	MPL-M	**1.139**	**0.698**	**0.490**	**0.988**	**0.646**	**0.449**	**0.906**	**0.602**	**0.425**
	MPL-G	1.124	0.725	0.513	1.034	0.668	0.476	0.949	0.618	0.449

Table 4: (continued)

		$\pi^E = 0\%$			$\pi^E = 25\%$			$\pi^E = 50\%$		
		$n = 100$	$n = 500$	$n = 2000$	$n = 100$	$n = 500$	$n = 2000$	$n = 100$	$n = 500$	$n = 2000$
$h_0(t_3)$	Nelson–Aalen	3.531	4.252	2.688	2.957	4.164	2.764	2.845	3.609	2.449
	CM	3.119	2.031	2.137	1.949	1.901	1.847	1.952	1.758	1.810
	EM-I	2.378	1.149	0.758	1.891	1.021	0.695	1.626	0.941	0.637
	MPL-M	1.559	**0.982**	**0.697**	**1.366**	**0.896**	**0.627**	**1.265**	**0.843**	**0.595**
	MPL-G	**1.542**	1.004	0.720	1.457	0.933	0.661	1.344	0.884	0.625
Integrated discrepancy between $\widehat{h}_0(t)$ and $h_0(t)$ defined between 0 and the 90th percentile of T										
	Nelson–Aalen	21.841	6.543	4.776	4.945	3.011	2.866	4.026	2.970	2.854
	CM	5.241	2.752	2.935	2.618	2.630	2.734	2.573	2.523	2.668
	EM-I	3.611	0.987	0.461	2.246	0.748	0.348	1.715	0.618	0.298
	MPL-M	1.615	**0.727**	**0.389**	**1.352**	**0.591**	**0.292**	**1.176**	**0.518**	**0.265**
	MPL-G	**1.581**	0.808	0.443	1.398	0.641	0.323	1.232	0.565	0.294
95% coverage probabilities										
$h_0(t_1)$	EM-I	**0.864**	0.926	0.938	**0.830**	0.920	0.948	**0.839**	0.924	0.941
	MPL-M	0.922	0.945	0.941	0.926	0.935	0.951	0.932	0.931	0.950
	MPL-G	0.908	0.928	0.948	0.906	0.928	0.952	0.925	0.929	0.946
$h_0(t_2)$	EM-I	**0.869**	0.949	0.941	**0.847**	0.927	0.934	**0.860**	0.932	0.938
	MPL-M	**0.881**	0.933	0.957	0.908	0.936	0.952	0.915	0.948	0.945
	MPL G	**0.884**	0.946	0.948	0.917	0.939	0.943	0.916	0.940	0.947
$h_0(t_3)$	EM-I	0.913	0.939	0.946	**0.875**	0.931	0.939	**0.887**	0.927	0.941
	MPL-M	**0.735**	**0.868**	0.925	**0.793**	0.919	0.949	**0.821**	0.935	0.953
	MPL-G	**0.829**	**0.855**	0.930	**0.872**	0.924	0.951	**0.875**	0.939	0.949

Smallest root mean square errors and integrated discrepancy between per scenario as well as coverage probabilities below 0.9 appear in bold.

Table 5: Simulation 2 results for $h_0(t)$ for the 25th (t_1), 50th (t_2), and 75th (t_3) percentiles of T. Some results are missing for the *convex minorant estimator* as no inference was developed for this estimator, and for the *Nelson–Aalen estimator* as we did not estimate the variance for this estimator.

		$\pi^E = 0\%$			$\pi^E = 25\%$			$\pi^E = 50\%$		
		$n = 100$	$n = 500$	$n = 2000$	$n = 100$	$n = 500$	$n = 2000$	$n = 100$	$n = 500$	$n = 2000$
Biases										
$h_0(t_1)$	Nelson–Aalen	−0.584	−0.670	−0.687	−0.522	−0.553	−0.603	0.428	−0.390	−0.501
	CM	0.194	0.170	0.114	−0.201	−0.167	−0.239	0.086	0.128	0.059
	EM-I	−0.012	0.037	0.176	−0.012	−0.003	0.033	−0.035	−0.008	0.017
	MPL-M	−0.004	−0.020	0.010	−0.018	−0.014	0.013	−0.029	−0.025	0.014
	MPL-G	0.050	0.032	0.025	0.017	0.020	0.013	0.017	0.016	0.011
$h_0(t_2)$	Nelson–Aalen	−0.611	−0.617	−0.638	−0.554	−0.549	−0.603	−0.555	−0.540	−0.543
	CM	0.591	0.150	0.120	−0.221	−0.111	−0.154	−0.012	0.074	−0.004
	EM-I	−0.049	0.005	0.199	−0.019	0.022	0.032	−0.025	−0.002	0.011
	MPL-M	−0.123	−0.100	−0.050	−0.110	−0.067	−0.021	−0.107	−0.049	−0.010
	MPL-G	−0.084	−0.058	−0.027	−0.061	−0.030	−0.012	−0.037	−0.018	−0.004
$h_0(t_3)$	Nelson–Aalen	−0.530	−0.580	−0.494	−0.524	−0.563	−0.517	−0.546	−0.599	−0.567
	CM	1.024	0.149	0.011	−0.342	−0.597	−0.624	−0.399	−0.439	−0.423
	EM-I	0.040	−0.034	−0.019	0.032	0.017	0.040	0.021	0.008	0.013
	MPL-M	−0.139	−0.115	−0.078	−0.112	−0.085	−0.043	−0.105	−0.065	−0.027
	MPL-G	−0.166	−0.104	−0.055	−0.104	−0.062	−0.031	−0.077	−0.041	−0.017
Mean asymptotic (Monte Carlo) standard errors										
$h_0(t_1)$	Nelson–Aalen	–	–	–	–	–	–	–	–	–
		(1.274)	(0.491)	(0.452)	(0.860)	(0.641)	(0.516)	(31.683)	(4.300)	(1.615)
	CM	–	–	–	–	–	–	–	–	–
		(1.409)	(1.817)	(2.620)	(1.019)	(0.825)	(0.817)	(1.478)	(1.391)	(1.259)
	EM-I	0.504	0.312	0.223	0.428	0.271	0.146	0.320	0.220	0.129
		(0.452)	(0.262)	(0.162)	(0.450)	(0.247)	(0.147)	(0.405)	(0.229)	(0.128)
	MPL-M	0.397	0.190	0.110	0.349	0.170	0.096	0.314	0.154	0.087
		(0.424)	(0.201)	(0.115)	(0.373)	(0.185)	(0.098)	(0.329)	(0.156)	(0.090)
	MPL-G	0.451	0.211	0.113	0.398	0.184	0.098	0.370	0.170	0.088
		(0.447)	(0.216)	(0.117)	(0.418)	(0.195)	(0.102)	(0.375)	(0.170)	(0.091)

Table 5: (continued)

		$\pi^E = 0\%$			$\pi^E = 25\%$			$\pi^E = 50\%$		
		$n = 100$	$n = 500$	$n = 2000$	$n = 100$	$n = 500$	$n = 2000$	$n = 100$	$n = 500$	$n = 2000$
$h_0(t_2)$	Nelson–Aalen	–	–	–	–	–	–	–	–	–
		(1.596)	(1.356)	(1.135)	(1.605)	(1.563)	(1.061)	(1.656)	(1.448)	(1.608)
	CM	–	–	–	–	–	–	–	–	–
		(5.265)	(4.801)	(3.890)	(1.664)	(1.709)	(1.655)	(1.856)	(2.125)	(1.704)
	EM-I	0.770	0.470	0.346	0.685	0.468	0.257	0.536	0.386	0.228
		(0.878)	(0.517)	(0.340)	(0.817)	(0.459)	(0.258)	(0.745)	(0.377)	(0.232)
	MPL-M	0.632	0.293	0.169	0.558	0.268	0.151	0.500	0.249	0.138
		(0.686)	(0.319)	(0.180)	(0.603)	(0.301)	(0.156)	(0.538)	(0.263)	(0.146)
	MPL-G	0.676	0.316	0.173	0.616	0.286	0.154	0.577	0.266	0.141
		(0.708)	(0.329)	(0.178)	(0.661)	(0.306)	(0.157)	(0.603)	(0.267)	(0.147)
$h_0(t_3)$	Nelson–Aalen	–	–	–	–	–	–	–	–	–
		(1.854)	(1.560)	(4.388)	(2.154)	(1.832)	(3.267)	(2.185)	(1.602)	(2.558)
	CM	–	–	–	–	–	–	–	–	–
		(10.753)	(5.271)	(5.261)	(2.779)	(1.310)	(0.966)	(1.755)	(1.450)	(1.769)
	EM-I	1.321	0.637	0.410	1.045	0.633	0.348	0.832	0.546	0.314
		(1.982)	(0.570)	(0.255)	(1.314)	(0.587)	(0.323)	(1.085)	(0.552)	(0.296)
	MPL-M	0.938	0.422	0.227	0.832	0.376	0.205	0.741	0.345	0.188
		(1.064)	(0.459)	(0.238)	(0.949)	(0.420)	(0.216)	(0.839)	(0.373)	(0.201)
	MPL-G	0.912	0.429	0.233	0.856	0.390	0.210	0.790	0.362	0.192
		(0.999)	(0.455)	(0.235)	(0.968)	(0.416)	(0.215)	(0.847)	(0.380)	(0.201)
					Root mean square error					
$h_0(t_1)$	Nelson–Aalen	1.191	0.934	0.930	1.013	0.935	0.910	5.628	2.078	1.304
	CM	1.193	1.351	1.619	1.020	0.919	0.925	1.217	1.182	1.122
	EM-I	0.672	0.515	0.499	0.671	0.496	0.389	0.638	0.478	0.360
	MPL-M	**0.651**	**0.449**	**0.340**	**0.611**	**0.430**	**0.315**	**0.574**	**0.398**	**0.302**
	MPL-G	0.671	0.467	0.346	0.647	0.443	0.320	0.613	0.413	0.303
$h_0(t_2)$	Nelson–Aalen	1.414	1.350	1.303	1.393	1.379	1.263	1.409	1.341	1.390
	CM	2.322	2.193	1.974	1.311	1.312	1.297	1.362	1.459	1.305
	EM-I	0.939	0.719	0.721	0.904	0.679	0.515	0.864	0.614	0.482
	MPL-M	**0.853**	0.613	0.452	**0.800**	0.573	0.401	**0.760**	0.530	**0.384**
	MPL-G	**0.853**	**0.590**	**0.428**	0.820	**0.558**	**0.398**	0.779	**0.519**	**0.384**

Table 5: (continued)

		$\pi^{E} = 0\%$			$\pi^{E} = 25\%$			$\pi^{E} = 50\%$		
		$n = 100$	$n = 500$	$n = 2000$	$n = 100$	$n = 500$	$n = 2000$	$n = 100$	$n = 500$	$n = 2000$
$h_0(t_3)$	Nelson–Aalen	1.562	1.528	2.151	1.632	1.577	1.901	1.651	1.551	1.752
	CM	3.343	2.299	2.293	1.720	1.488	1.450	1.456	1.398	1.475
	EM-I	1.409	0.761	**0.511**	1.147	0.768	0.587	1.042	0.743	0.547
	MPL-M	1.069	0.758	0.577	**1.003**	0.701	0.500	0.946	0.649	0.465
	MPL-G	**1.056**	**0.742**	0.535	1.008	**0.675**	**0.483**	**0.937**	**0.632**	**0.456**
	Integrated discrepancy between $\widehat{h}_0(t)$ and $h_0(t)$ defined between 0 and the 90th percentile of T									
	Nelson–Aalen	0.802	0.775	0.770	0.779	0.750	0.747	0.764	0.741	0.737
	CM	1.289	0.849	0.837	0.714	0.645	0.648	0.665	0.624	0.671
	EM-I	0.415	0.179	0.139	0.345	0.170	0.098	0.304	0.155	0.088
	MPL-M	0.326	0.161	0.089	0.287	0.139	0.070	0.259	0.121	0.063
	MPL-G	**0.317**	**0.153**	**0.079**	**0.285**	**0.130**	**0.068**	**0.255**	**0.116**	**0.060**
	Empirical 95% coverage probabilities									
$h_0(t_1)$	EM-I	**0.882**	**0.892**	**0.872**	**0.782**	**0.890**	0.930	**0.769**	**0.812**	**0.890**
	MPL-M	**0.859**	0.917	0.942	**0.881**	0.909	0.947	**0.879**	0.934	0.938
	MPL-G	**0.899**	0.947	0.940	0.903	0.935	0.938	0.905	0.950	0.942
$h_0(t_2)$	EM-I	**0.784**	**0.845**	**0.693**	**0.741**	**0.866**	0.938	**0.706**	**0.854**	0.907
	MPL-M	**0.786**	**0.810**	**0.859**	**0.802**	**0.849**	0.920	**0.807**	**0.890**	0.934
	MPL-G	**0.828**	**0.896**	0.922	**0.851**	**0.898**	0.928	**0.871**	0.939	0.938
$h_0(t_3)$	EM-I	**0.818**	**0.874**	0.950	**0.795**	**0.893**	0.953	**0.776**	**0.821**	0.927
	MPL-M	**0.757**	**0.775**	**0.772**	**0.793**	**0.801**	**0.877**	**0.784**	**0.845**	**0.893**
	MPL-G	**0.731**	**0.801**	**0.850**	**0.805**	**0.863**	0.910	**0.823**	**0.882**	0.914

Smallest root mean square errors and integrated discrepancy between per scenario as well as coverage probabilities below 0.9 appear in bold.

and $n = 500$. In fact, we increased n_α to $n_\alpha = 9$ for $n = 100$ and $n_\alpha = 11$ for $n = 500$ and these results are denoted as MPL-M(+2). We also doubled the number of knots, namely $n_\alpha = 14$ and $n_\alpha = 18$ for respectively $n = 100$ and $n = 500$, and their results are denoted by MPL-M(×2). The estimates were obtained from MPL with M-splines. The results from these new n_α are reported in Table S4 in Supplementary Material. They suggest that a change in the number of knots has a low impact on the estimates. However, a large number of knots can induce bias to the $\boldsymbol{\beta}$ estimates when n is small but their impact on the standard errors of these $\boldsymbol{\beta}$ estimates are minor. A large number of knots can also impact $\widehat{h}_0(t)$ when n is small, especially for large values of t, leading to poor coverage probabilities.

5.2 Application in a melanoma study

In this section, we apply the MPL estimator with M-spline bases to fit a Cox model for the time of first local melanoma recurrence for patients who were diagnosed with melanoma between 1998 and 2016 in Australia; see [24] for some further information about a similar data set. Our data set, kindly provided by the Melanoma Institute Australia, indicates the date of melanoma diagnosis (t_d) and the date of last follow-up (t_f) with recurrence status for 2175 patients. If a melanoma recurrence was observed, it also indicates when the first recurrence was diagnosed (t_r) as well as the date of the last negative check before recurrence (t_n), if available.

Melanoma recurrence was observed for 37% of the patients. At time of last follow-up, 70.5% of the patients were alive and 29.5% dead. Among the alive patients, 95% were with no melanoma, 4% with melanoma and 1% with unknown melanoma status. Among the dead patients, 18% were with no melanoma, 71% with melanoma and 11% with unknown melanoma status. We set the melanoma diagnosis time as the time origin for each patient. Times of first recurrence are typically interval censored as they occurred between patient visits to the doctor. For a patient with non-missing t_n and t_r, the first melanoma recurrence is censored in $[t_\mathrm{n} - t_\mathrm{d}, t_\mathrm{r} - t_\mathrm{d}]$. If a patient whose t_n is missing but t_r is available, then melanoma recurrence is censored in $[0, t_\mathrm{r} - t_\mathrm{d}]$. If t_r is missing and the patient had melanoma at t_f, then the recurrence time is censored in $[0, t_\mathrm{f} - t_\mathrm{d}]$. If t_r is missing and the patient had no melanoma at t_f, the recurrence time is (right) censored in $[t_\mathrm{f} - t_\mathrm{d}, \infty)$. Cases with no observed recurrence and no known status at time of last follow up were considered as missing.

We considered the following covariates in our model: (1) melanoma location at first diagnostic, which is a categorical variable with levels 'Head and neck' (19.1%), 'Arm' (14.4%), 'Leg' (28.7%), 'Trunk' (37.8%); (2) melanoma stage at first diagnostic according to Nelson–Aalen's thickness scale, and this is an ordinal variable with levels '[0,1) mm' (15.2%), '[1,2) mm' (42.5%), '[2,4) mm' (29.2%) and '4 mm and more' (13.2%); (3) gender, which is a categorical variable with levels 'Men' (58.1%), 'Women' (41.9%); (4) (centered) age in years at first diagnostic, where the range of the non centred ages is [5, 94] and the mean of non centered ages equals 55.7 years. The contrasts were chosen so that the baseline hazard corresponds to the instantaneous risk to have a first melanoma recurrence on the head/neck for a male of 55.7 years old who was initially diagnosed with a melanoma of small size (<1 mm). We chose to model the baseline hazard function using 10 M-spline bases (again no effort was made to optimize this number). Two of them were placed at the extremities of the time range of interest and the others were placed at equidistant interval mid-points.

The hazard ratio estimates are exhibited in Table 6. Compared with melanoma that were first diagnosed at the head & neck, melanoma at arm or trunk have significantly lower risk of recurrence. Initial melanoma thickness is another strong risk factor for melanoma recurrence. A 10-year increase of age corresponds to a significant (9–21%) risk increase of melanoma recurrence. Gender is marginally significant, with women having a lower risk of melanoma recurrence than men.

The hazard ratio estimates from the competitors are available in Table S2 in Supplementary Material. In this application example, all the methods, even the PL method with middle point imputation, lead to very similar estimates, likely due to relative narrow censoring intervals (for both left and interval censoring).

The estimated baseline hazard function, together with its 95% pointwise confidence interval, is displayed in Figure 1. This plot indicates that when the covariates are all set to their baseline values, the risk of melanoma

Table 6: Hazard ratio estimates ($e^{\hat{\beta}}$), hazard ratio 95% confidence intervals, and *p*-values of the significant tests.

		HR estimates	HR 95% CI	p-Value
Location	Arm	0.570	[0.429;0.757]	0.0001
	Leg	1.008	[0.816;1.244]	0.9430
	Trunk	0.802	[0.658;0.977]	0.0283
Thickness	1–2 mm	1.245	[0.937;1.653]	0.1303
	2–4 mm	2.390	[1.804;3.166]	<0.0001
	4 mm and more	3.108	[2.295;4.208]	<0.0001
Gender	Female	0.843	[0.717;0.990]	0.0375
Centered age (10 years)	–	1.148	[1.093;1.205]	<0.0001

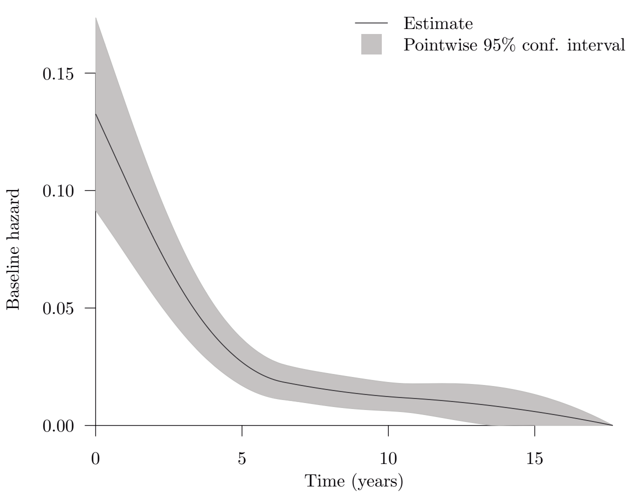

Figure 1: Plots of baseline hazard estimate and pointwise 95% CI.

recurrence strongly and monotonically decreases during the first 5 years. After that, the risk decreases at a slow rate to a level close to 0 during the next decade.

6 Conclusions

This paper develops a new computational procedure for the semi-parametric proportional hazards model where survival time observations include left, interval and right censoring times as well as event times. Since the baseline hazard is non-parametric and subject to a non-negativity constraint, we approximate this function using a finite number of non-negative basis functions with corresponding coefficients constrained to be non-negative. An efficient Newton–MI algorithm is developed. The asymptotic consistency results for the estimates of regression coefficients and baseline hazard are established. Asymptotic normality results were further developed to accommodate active constraints, which is of practical importance.

The simulation results reveal that our MPL estimator often produces more satisfactory results than its competitors for both the regression coefficients and baseline hazard estimates. In particular, the MPL estimates of regression coefficients can usually achieve smaller biases and standard errors than the competitors. The 95% confidence intervals from the MPL estimates also usually achieve better coverage probabilities (i.e. closer to 95%) than the competitors. Other benefits of our MPL method include its lower sensitivity to choice of the knot sequence, including spacing and number of knots, as well as its ability to lead to positive definite asymptotic covariance matrices for the estimated parameters and therefore enables inferences on hazard, cumulative hazard and survival functions.

The Cox model fitting method described in this paper is implemented in an R package '`survivalMPL`', which is available at CRAN.

Acknowledgments: The authors wish to thank the referees and the associate editor for their thoughtful comments and suggestions which helped improve the quality of this paper. Heritier's research was partially supported by a Centre of Research Excellence grant from the National Health and Medical Research Council, ID# 1171422, to the Australian Trials Methodology Research Network (AusTriM).
Author contribution: All the authors have accepted responsibility for the entire content of this submitted manuscript and approved submission.
Research funding: None declared.
Conflict of interest statement: The authors declare no conflicts of interest regarding this article.

References

1. Kim JS. Maximum likelihood estimation for the proportional hazard model with partly interval-censored data. J Roy Stat Soc B 2003;65:489−502.
2. Finkelstein DM. A proportional hazard model for interval-censored failure time data. Biometrics 1986;42:845−54.
3. Huang J. Efficient estimation for the proportional hazard model with interval censoring. Ann Stat 1996;24:540−68.
4. Joly P, Commenges D, Letenneur L. A penalized likelihood approach for arbitrarily censored and truncated data: application to age-specific incidence of dementia. Biometrics 1998;54:185−94.
5. Pan W. Extending the iterative convex minorant algorithm to the Cox model for interval-censored data. J Comput Graph Stat 1999;8:109−20.
6. Sun J. The statistical analysis of interval-censored failure time data. New York: Springer; 2006.
7. Wang L, McMahan C, Hudgens M, Qureshi Z. A flexible, computationally efficient method for fitting the proportional hazards model to interval-censored data. Biometrics 2016;72:222−31.
8. Zhang Y, Hua L, Huang J. A spline-based semiparametric maximum likelihood estimation method for the Cox model with interval-censored data. Scand J Stat 2010;37:338−54.
9. Cai T, Betensky RA. Hazard regression for interval-censored data with penalized spline. Biometrics 2003;59:570−9.
10. Chen D-G, Sun J, Peace KE. Interval-censored time-to-event data: methods and applications. New York: Chapman and Hall/CRC; 2012.
11. Bouaziz O, Lauridsen E, Nuel G. Regression modelling of interval censored data based on the adaptive ridge procedure. J Appl Stat 2021. https://doi.org/10.1080/02664763.2021.1944996.
12. Ma J, Heritier S, Lô S. On the maximum penalized likelihood approach for proportional hazard models with right censored survival data. Comput Stat Data Anal 2014;74:142−56.
13. Green PJ, Silverman BW. Nonparametric regression and generalized linear models - a roughness penealty approach. London: Chapman & Hall; 1994.
14. Lindsay BG. Efficiency of the conditional score in a mixture setting. Ann Stat 1983;11:486−97.
15. Armijo L. Minimization of functions having lipschitz-continuous first partial derivatives. Pac J Math 1966;16:1−3.
16. Luenberger D. Linear and nonlinear programming, 2nd ed. New York: J. Wiley; 1984.
17. Xue H, Lam KF, Li G. Sieve maximum likelihood estimator for semiparametric regression models with current status data. J Am Stat Assoc 2004;99:346−56.
18. Begun JM, Hall WJ, Huang WM, Wellner JA. Information and asymptotic efficiency in parametric-nonparametric models. Ann Stat 1983;11:432−52.
19. Chan RH, Ma J. A multiplicative iterative algorithm for box-constrained penalized likelihood image restoration. IEEE Trans Image Process 2012;21:3168−81.
20. Bickel PJ, Klaasen CAJ, Ritov Y, Wellner JA. Efficient and adaptive estimation for semiparametric models. Baltimore: The Johns Hopkins University Press; 1993.
21. Yu Y, Ruppert D. Penalized spline estimation for partially linear single index models. J Am Stat Assoc 2002;97:1042−54.
22. Crowder M. On constrained maximum likelihood estimation with non-i.i.d. observations. Ann Inst Stat Math 1984;36:239−49.

23. Moore TJ, Sadler BM, Kozick RJ. Maximum-likelihood estimation, the Cramér — Rao bound, and the method of scoring with parameter constraints. IEEE Trans Signal Process 2008;56:895—908.
24. Morton D, Thompson J, Cochran A, Mozzillo N, Nieweg O, Roses D, et al. Final trial report of sentinel-node biopsy versus nodal observation in melanoma. N Engl J Med 2014;307:599—609.

Supplementary Material: The online version of this article offers supplementary material (https://doi.org/10.1515/ijb-2020-0104).