Electronic Journal of Statistics
Vol. 18 (2024) 4882–4915
ISSN: 1935-7524
https://doi.org/10.1214/24-EJS2319

On a semiparametric estimation method for AFT mixture cure models

Ingrid Van Keilegom and Motahareh Parsa

ORSTAT, KU Leuven, Naamsestraat 69, 3000 Leuven, Belgium,
e-mail: ingrid.vankeilegom@kuleuven.be; motahareh.parsa@kuleuven.be

Abstract: When studying survival data in the presence of right censoring, it often happens that a certain proportion of the individuals under study do not experience the event of interest and are considered as cured. It is then common to model the data via a mixture cure model. It depends on a model for the conditional probability of being cured (called the incidence) and a model for the conditional survival function of the uncured individuals (called the latency). This work considers a logistic model for the incidence and a semiparametric accelerated failure time model for the latency part. The estimation of this model is obtained via the maximization of the semiparametric likelihood, in which the unknown error density is replaced by a kernel estimator based on the Kaplan-Meier estimator of the error distribution. Asymptotic theory for consistency and asymptotic normality of the parameter estimators is provided. Moreover, the proposed estimation method is compared with several competitors. Finally, the new method is applied to data coming from a cancer clinical trial. An R package, called *kmcure*, is developed to facilitate the use of the proposed methodology in practice.

MSC2020 subject classifications: Primary 62N01, 62N02; secondary 62G07, 62G20.
Keywords and phrases: Accelerated failure time model, cure models, Kaplan-Meier estimator, Kernel density estimator, semiparametric model.

Received January 2023.

1. Introduction

Cure models in survival analysis are nowadays standard in the modeling toolbox for situations in which a certain proportion of the subjects under study never experience the event of interest, i.e. their survival time will be equal to infinity. This kind of phenomenon often occurs in practice, when studying e.g. the time until death or recurrence of a certain disease, the time until an unemployed person finds a new job, the time until a bank goes bankrupt, or the time until a released prisoner is re-arrested. Basically, the population is then composed of two groups of subjects, the susceptible (or uncured) subjects and the non-susceptible (or cured) ones [8]. For book-long introductions to cure models we refer to [19] and the recent book by [25], while recent review papers on cure models are [24, 1, 12]. They all provide a comprehensive introduction to cure models in terms of modeling, estimation, inference, and software.

In the common case where the survival time is subject to right censoring and the censoring variable does not have a mass at infinity, all cured subjects will be censored, which makes the identifiability and estimation of this type of model challenging. It is clear that some assumptions will be needed to identify the cure fraction. A common assumption is on the duration of the experiment, which should be sufficiently long to distinguish cured from uncured subjects.

When covariates are present, a common class of cure regression models is the class of mixture cure models, which considers the population as a mixture of the susceptible and cured subpopulations, and which is determined by a model for the subpopulation of susceptible subjects (called the latency) and a model for the probability of being cured (called the cure fraction or the incidence), each time conditional on the covariates. Formally speaking, the survival function $S(t|x,z) = P(T > t \mid X = x, Z = z)$ of the survival time T given a set of real-valued covariates $(X, Z) = (x, z)$ is given by

$$S(t|x,z) = 1 - p(z) + p(z)S_u(t|x), \tag{1}$$

where $p(z) = P(B = 1 \mid Z = z)$ is the conditional probability of being uncured (called the incidence), $B = I(T < \infty)$ denotes the uncure status, and $S_u(t|x) = P(T > t \mid B = 1, X = x)$ is the conditional survival function for the uncured subjects (called the latency). The vectors of covariates X and Z are of dimension ℓ and $k+1$ respectively, and can contain the same covariates, but they can also be partially or completely different. The models for $p(z)$ and $S_u(t|x)$ can be parametric, semiparametric, or nonparametric in nature. We refer to [3, 2, 8] for fully parametric approaches, and [15, 14] for fully nonparametric approaches. For the middle category of semiparametric models, we like to mention [11, 23, 27, 7, 16, 4], who all proposed estimators for the semiparametric logistic/Cox mixture cure model, while [22] suggested an estimation strategy which is based on a parametric model for the incidence and a nonparametric model for the latency.

In this paper, we will focus on the case where the cure fraction follows a logistic model (which is common in the literature on cure models), and the conditional survival function $S_u(t|x)$ of the susceptible follows a semiparametric accelerated failure time (AFT) model. This model is a useful alternative to the Cox model thanks to its direct physical interpretation [10, 6]. When a cure fraction is present, the model has however not received much attention in the literature so far. As far as we know, the only papers that have proposed estimators for the semiparametric logistic/AFT mixture cure model are [31, 17, 26]. These papers differ in the way they estimate the nonparametric error survival function and in the likelihood they use to estimate the parameters in the model. A comparison between the logistic/Cox and the logistic/AFT mixture cure models in terms of their ability to estimate well the cure fraction is given in [21]. This comparison is especially relevant when the follow-up period of the experiment is insufficient since the AFT model is able to transfer tail information from regions in the covariate space where the follow-up is sufficient to regions where the follow-up is insufficient.

The goal of this paper is to propose a new estimation strategy for this semiparametric logistic/AFT mixture cure model. The estimation of the model is obtained via the maximization of a so-called semiparametric observed likelihood, in which the unknown error density in the AFT model is replaced by a kernel estimator based on the Kaplan-Meier estimator of the error distribution. We will develop rigorous asymptotic theory for the proposed estimator, and show via simulations and the analysis of real data how the estimator performs in practice, also compared to the existing estimators mentioned above.

This paper is organized as follows. In the next section we formally define the semiparametric logistic/AFT mixture cure model, and we introduce some notations. In Section 3 we explain what are the existing estimation procedures for this model, together with their pros and cons, we introduce our proposed estimation method and state the theorems for the consistency and asymptotic normality of the estimator. Section 4 is devoted to a finite sample study in which the proposed estimator is compared to the existing competitors. We also consider the drawbacks and benefits of each method. In Section 5 real data on the time to distant metastasis for lymph-node-negative breast cancer patients are analyzed. Finally, the Appendix contains the proofs of the asymptotic results and the results of additional simulations.

2. The AFT/logistic mixture cure model

We suppose that the uncure probability $p(z)$ follows a logistic model given by

$$p(z) = p_\gamma(z) = \frac{\exp(\gamma^t z)}{1 + \exp(\gamma^t z)}, \tag{2}$$

where the vector $\gamma = (\gamma_0, \ldots, \gamma_k)^t$ is associated with z and contains an intercept *i.e.* the first element of the vector z is 1. Note that other parametric models for $p(z)$ are also possible, as long as the parameters in the model are uniquely identified.

We can write $T = T^*B + \infty(1 - B)$, where T^* is the survival time of the susceptible subjects. For the latency part, we consider a semi-parametric accelerated failure time (AFT) model of the following form:

$$\log T^* = \beta^t X + \epsilon, \tag{3}$$

where the error ϵ is independent of (X, Z) and its distribution is unspecified, $\beta = (\beta_1, \ldots, \beta_\ell)^t$ is a vector of parameters associated with X, and the intercept is absorbed by the error term ϵ. Equivalently, we can define the AFT model by specifying the survival function:

$$S_u(t|x) = S_{u,\beta}(t|x) = S_0(t\exp(-\beta^t x)), \tag{4}$$

where $S_0(t) = P(\exp(\epsilon) > t)$ is the error survival function corresponding to the conditional survival function for $X = 0$.

Throughout the paper, we consider the AFT/logistic mixture cure model given by (1), (2) and (3). As is often the case with time-to-event data, the survival time T is subject to random right censoring, *i.e.* instead of observing T we observe the couple (Y, Δ), where $Y = \min(T, C)$ is the observed survival time, $\Delta = I(T \leq C)$ is the censoring indicator, and C is the censoring time. We assume that T and C are independent given the covariates (X, Z). Let $(Y_i, \Delta_i, X_i, Z_i)$, $i = 1, \ldots, n$, be i.i.d. realizations of (Y, Δ, X, Z).

The identifiability of model (1)-(3) has been shown in [21]. They showed that sufficient conditions for identifiability are

(A) (i) For all z, $0 < p(z) < 1$.
(ii) The matrices Var (X) and Var (Z) are positive definite.
(iii) The variable $\exp(\epsilon)$ has support $[0, \tau_0]$ for some $\tau_0 < \infty$.
(iv) $P(C > \tau_0 \exp(\beta^t X)|X, Z) > 0$ for all $(X, Z) \in S = S_X \times S_Z$, where S_X and S_Z are such that $P(X \in S_X, Z \in S_Z) > 0$, Var $(X|X \in S_X) > 0$ and Var $(Z|Z \in S_Z) > 0$.

Note that assumption (A)(iv) shows that the model is identified even if the follow-up period is insufficient for certain regions of the covariate space. This makes the AFT mixture cure model an attractive model in practice. [21] showed that this feature holds for the AFT but not for the Cox mixture cure model.

Under this data-generating process, the likelihood is given by

$$\mathcal{L}^O(\theta, S_0, f_0) = \prod_{i=1}^n \Big[p_\gamma(Z_i) e^{-\beta^t X_i} f_0(Y_i e^{-\beta^t X_i})\Big]^{\Delta_i} \times \Big[1 - p_\gamma(Z_i) + p_\gamma(Z_i) S_0(Y_i e^{-\beta^t X_i})\Big]^{1-\Delta_i}, \quad (5)$$

where $\theta = (\gamma, \beta)^t$ and $f_0(t) = -(d/dt)S_0(t)$. This likelihood is often called the observed likelihood, since it is based on the contributions of the uncensored and censored observations, which are observable. On the other hand, the complete likelihood is based on the contributions of uncensored subjects, censored and uncured subjects, and censored and cured subjects. The latter likelihood depends on the latent cure status B and is given by:

$$\mathcal{L}^C(\theta, S_0, f_0) = \prod_{i=1}^n \Big[p_\gamma(Z_i) e^{-\beta^t X_i} f_0(Y_i e^{-\beta^t X_i})\Big]^{B_i \Delta_i} \times \Big[1 - p_\gamma(Z_i)\Big]^{(1-B_i)(1-\Delta_i)} \Big[p_\gamma(Z_i) S_0(Y_i e^{-\beta^t X_i})\Big]^{B_i(1-\Delta_i)}. \quad (6)$$

3. The proposed estimator

3.1. Estimation procedure

We will now provide a semiparametric estimation method for the AFT/logistic mixture cure model defined in (1), (2) and (3). The estimation of this model has been studied already in the past. The different estimation approaches differ in the way in which they estimate S_0 and f_0, and in the likelihood they use (observed or complete). Note that the incidence $p_\gamma(z)$ is parametric and hence it is easy to estimate, whereas the latency $S_0(t\exp(-\beta^t x))$ is semiparametric and therefore more challenging. As far as we know only one approach is based on the observed likelihood, which is given in [26]. The estimation of the functions S_0 and f_0 is done based on a so-called SNP (semi-nonparametric) approach with exponential or normal basis functions. For selecting the number of basis functions an AIC criterion is exploited. Their method works both for right censored and interval censored data.

While the observed likelihood has the advantage of not depending on latent variables, the complete likelihood is computationally more attractive, since it can be decomposed in the product of two factors, one only depending on β, S_0 and f_0, and the other one only depending on γ. All existing approaches are based on the EM algorithm due to the unobserved B_i's. The first approach is the one by [31], who proposed a rank estimator for β. Their method was later included in the *smcure* R package (see [4]). Later, [17] used a kernel approach to maximize the profile likelihood in the M-step. In the E-step, the conditional expectation of the complete likelihood is computed given the observed data and the current parameter estimates. The proposed kernel estimation method is motivated by the work of [30], in which an efficient estimatior for the AFT model without cure fraction is introduced. The paper by [17] is the only one that developed asymptotic theory for the proposed estimators.

Our approach is based on the observed likelihood in (5) and on preliminary nonparametric estimators of the functions f_0 and S_0. Let $\theta_0 = (\gamma_0, \beta_0)^t$ be the true parameter vector. For fixed β, let $\epsilon_{i:n}$ $(i = 1, \dots, n)$ be the i-th order statistic of $\epsilon_1, \dots, \epsilon_n$, where $\epsilon_i = \log T_i^* - \beta^t X_i$, and let $\Delta_{i:n}$ be the corresponding censoring indicator. Then, assuming that the error distribution is smooth, the Kaplan-Meier estimator of $S_{0,\beta}(t) = P(T^* \exp(-\beta^t X) \le t)$ is given by

$$\hat{S}_{0,\beta}(t) = \frac{\hat{S}_\beta(t) - \hat{S}_\beta(\exp(\epsilon_{n:n}))}{1 - \hat{S}_\beta(\exp(\epsilon_{n:n}))}, \tag{7}$$

where

$$\hat{S}_\beta(t) = \prod_{i:\exp(\epsilon_{i:n}) \le t} \left(1 - \frac{1}{n-i+1}\right)^{\Delta_{i:n}}, \tag{8}$$

in which the estimator depends on β via the error terms $\epsilon_{i:n}$, $i = 1, \dots, n$. Note that when β equals the true parameter vector β_0, $\hat{S}_{0,\beta}(t)$ estimates the true error survival function $S_0(t)$. Standardization in (7) is necessary to make sure that $\hat{S}_{0,\beta}(t)$ is a proper survival function. For estimating the density of the error term, a kernel density estimator of $f_{0,\beta}(t) = -(d/dt)S_{0,\beta}(t)$ is used:

$$\hat{f}_{0,\beta}(t) = b^{-1} \int K\Big(\frac{t-s}{b}\Big) d\hat{F}_{0,\beta}(s), \tag{9}$$

where $\hat{F}_{0,\beta} = 1 - \hat{S}_{0,\beta}$, $b = b_n$ is a bandwidth parameter tending to zero as n tends to infinity, and K is a kernel function.

Define the vector of nuisance functions $h = (S_0, f_0, f_0')$, and let $M_n(\theta, h)$ be the vector of partial derivatives of the log-likelihood with respect to θ, *i.e.*

$$M_n(\theta,h) = \begin{pmatrix} \frac{\partial}{\partial\gamma}\log\mathcal{L}^O(\theta,h) \\ \frac{\partial}{\partial\beta}\log\mathcal{L}^O(\theta,h) \end{pmatrix} = \sum_{i=1}^n m(V_i,\theta,h) = \sum_{i=1}^n \begin{pmatrix} m_1(V_i,\theta,h) \\ m_2(V_i,\theta,h) \end{pmatrix},$$

where $V_i = (Y_i, \Delta_i, X_i, Z_i)$, $i = 1, \dots, n$,

$$m_1(V,\theta,h) = \frac{\Delta Z}{1+e^{\gamma^t Z}} - (1-\Delta)\frac{Ze^{\gamma^t Z}}{1+e^{\gamma^t Z}}\frac{1-S_0(Ye^{-\beta^t X})}{1+e^{\gamma^t Z}S_0(Ye^{-\beta^t X})},$$

and

$$\begin{aligned} m_2(V,\theta,h) = &-\Delta X\Big(1 + \frac{Ye^{-\beta^t X}f_0'(Ye^{-\beta^t X})}{f_0(Ye^{-\beta^t X})}\Big) \\ &+ (1-\Delta)X\frac{Ye^{-\beta^t X}e^{\gamma^t Z}f_0(Ye^{\beta^t X})}{1+e^{\gamma^t Z}S_0(Ye^{-\beta^t X})}. \end{aligned}$$

Moreover, let $M(\theta,h) = E[m(V,\theta,h)]$ be the score vector. Then, the true vector θ_0 satisfies $M(\theta_0,h_0) = 0$, where h_0 is the true vector of nuisance functions, and we define the estimator

$$\hat\theta = (\hat\gamma,\hat\beta)^t = \arg\min_{\gamma\in\Gamma,\beta\in\mathcal{B}} \|M_n(\theta,\hat h_\beta)\|, \tag{10}$$

where the parameter space $\Theta = \Gamma\times\mathcal{B}$ is a compact subspace of $\mathcal{R}^{k+\ell+1}$, $\|\cdot\|$ denotes the Euclidean norm, $\hat h_\beta = (\hat S_{0,\beta}, \hat f_{0,\beta}, \hat f_{0,\beta}')$, and

$$\hat f_{0,\beta}'(t) = b^{-2}\int K'\Big(\frac{t-s}{b}\Big)d\hat F_{0,\beta}(s).$$

Note that since $M_n(\theta,\hat h_\beta)$ is not smooth in θ (due to the non-smoothness of the Kaplan-Meier estimator $\hat S_{0,\beta}$), we minimize the norm of $M_n(\theta,\hat h_\beta)$ instead of solving the equation $M_n(\theta,\hat h_\beta) = 0$.

3.2. Asymptotic properties

Our criterion function $M_n(\theta,h)$ is semiparametric and is non-smooth in θ. We will therefore make use of the asymptotic theory for semiparametric Z-estimators based on non-smooth criterion functions, given in [5]. The latter paper provides high-level sufficient conditions under which consistency and asymptotic normality are guaranteed. We will check these high-level conditions for our estimation procedure.

This will be possible under assumption (A), which assures that there is a finite cure threshold $\tau_0 < \infty$, which is the upper bound of the support of $f_0(t)$. We start with the consistency of $\hat\theta$. For arbitrary $\beta\in\mathcal{B}$, note that $T^*\exp(-\beta^t X) = \exp(\epsilon)\exp(-(\beta-\beta_0)^t X)$ and hence the support of $f_{0,\beta}(t)$ is $[0,\tau(\beta)]$ with $\tau(\beta) = \tau_0\sup_{x\in R_X}\exp(-(\beta-\beta_0)^t x)$, since ϵ and X are independent, where R_X is the compact support of X.

Theorem 3.1. *Assume (A) and (C1)–(C8). Then,*

$$\hat{\theta} - \theta_0 \xrightarrow{P} 0.$$

The asymptotic normality of $\hat{\theta}$ can now be established.

Theorem 3.2. *Assume (A) and (C1)–(C8). Then,*

$$n^{1/2}(\hat{\theta} - \theta_0) \xrightarrow{d} N(0, \Sigma),$$

for some positive definite covariance matrix Σ*.*

As a by-product of our estimation procedure, we also obtain the following result regarding the estimators $\hat{S}_{0,\beta}$, $\hat{f}_{0,\beta}$ and $\hat{f}'_{0,\beta}$. Note that these results are well known in case β is fixed, so the challenge here is to show the stated rate of convergence uniformly in $\beta \in \mathcal{B}$.

Theorem 3.3. *Assume (A) and (C1)–(C8). Then,*

(i) $\sup_{\beta\in\mathcal{B}} \sup_{0\le t<\tau(\beta)} |\hat{S}_{0,\beta}(t) - S_{0,\beta}(t)| = O_P(n^{-1/2})$
(ii) $\sup_{\beta\in\mathcal{B}} \sup_{0\le t<\tau(\beta)} |\hat{f}_{0,\beta}(t) - f_{0,\beta}(t)| = O_P((nb_n)^{-1/2}(\log n)^{1/2}) + O(b_n^4)$
(iii) $\sup_{\beta\in\mathcal{B}} \sup_{0\le t<\tau(\beta)} |\hat{f}'_{0,\beta}(t) - f'_{0,\beta}(t)| = O_P((nb_n^3)^{-1/2}(\log n)^{1/2}) + O(b_n^4)$*.*

The proofs of all theorems are provided in Appendix A.

4. Simulation study

In this section we will carry out an extensive simulation study, in which we compare our proposed estimator with its competitors in the literature, namely the estimators of [31] (given by the R package *smcure*, see [4]), [17] and [26]. For our estimation procedure, we developed an R-package, called *kmcure*, which is available from `https://github.com/Motahareh-Parsa/kmcure`.

We consider the following simulation setup. The covariate X is generated under two scenarios: a Bernoulli distribution with success probability 0.5, or a uniform distribution on $[0, 1]$. We will concentrate below on the case where X is Bernoulli distributed. The uniform case is reported in Appendix B. Throughout this study, we set $Z = (1, X)$. The censoring time C is generated from a uniform distribution on $[0, \tau_C]$ and is independent of X and T, where τ_C equals either 20 or 100, corresponding to heavy or moderate right censoring.

The model for the latency is given by $\log T^* = \beta_1 X + \epsilon$, with $\beta_1 = 1$, whereas the model for the incidence is $p(z) = \exp(\gamma^t z)/(1+\exp(\gamma^t z))$ with γ_0 equal to 0.5 or 1 and $\gamma_1 = -0.5$, which means that the overall cure fraction is 0.44 or 0.32 respectively. The error term ϵ is generated from either a standard logistic distribution, a standard normal distribution, or a mixture $0.6\,\text{Weib}(6, 1) + 0.4\,\text{Weib}(2, 1)$ of two Weibull distributions. To satisfy the constraint that the support of the error distribution is bounded (see condition (A)(iii)), we truncate these error distributions at their 90 % percentile. Table 1 provides the cure fraction and right censoring rate which are produced by the different values of γ_0 and τ_C.

TABLE 1
Cure fraction (CF) and right censoring (RC) rate for each setting considered in the simulation (expressed in %).

	$\gamma_0 = 0.5$				$\gamma_0 = 1$			
	$\tau_C = 20$		$\tau_C = 100$		$\tau_C = 20$		$\tau_C = 100$	
Error	CF	RC	CF	RC	CF	RC	CF	RC
Logistic	44	56	44	48	32	47	32	37
Normal	44	52	44	45	32	42	32	34
Mix-Weibull	44	57	44	46	32	48	32	36

We compare our method with the kernel based approach of Lu (2010) [17], the rank-based method of Zhang and Peng (2007) [31], which was implemented in the R package *smcure* by [4], and the SNP approach of Scolas et al (2016) [26], based on two basis distributions (standard normal and standard exponential) and polynomials of order 0, 1 or 2. The AIC criterion is used to select the optimal choice of the error distribution.

In each scenario two sample sizes are considered, namely $n = 200$ and 400. Our method and the method of [17] require the selection of a bandwidth parameter. We follow the procedure proposed by [17], and work with a Gaussian kernel and with the bandwidth $b = (8\sqrt{2}/3)^{1/5}\hat{\sigma}n^{-1/5}$, where $\hat{\sigma}$ is the sample standard deviation for the uncensored error terms, in which β is substituted by the estimator derived by fitting the linear model to the uncensored data.

The simulation results are presented in Tables 2–4 and are based on 500 runs. Note that to calculate the bias, variance and MSE, we only use samples for which the four estimators could all be computed without errors. Specifically, Lu's method often encounters errors in parameter estimation, and such problematic samples are then excluded in the reported results. We will come back to these numerical problems later in this section.

The tables show that most of the time our method behaves slightly better than Lu's method, and is comparable to Zhang and Peng's method. This is true for all model scenarios, for both the incidence and the latency parameters, and for both the bias and variance.

To assess the normality of the estimated coefficients in our simulations, we employ Q-Q plots, which serve as a robust method for assessing normality. These plots provide a visual comparison between the observed quantiles of the estimated coefficients in 500 simulations and the expected quantiles under a normal distribution. The Q-Q plots are given in Figure 4 in the Appendix for one setting, and show that the normality is approximately satisfied for all methods. Furthermore, we provide in Table 5 the Pearson correlations (denoted by QQr) between the observed and expected quantiles as a metric for normality assessment.

Also, in the simulations we used 100 bootstrap samples to estimate the parameters' standard errors. Then, we used the asymptotic normality of the estimators to construct 95% confidence intervals (CI). Their coverage probabilities (CP), and the average length of these confidence intervals (CI L) are calculated based on 500 samples, and are presented in Table 5. The results show that the

Table 2. *Bias, variance and mean squared error (MSE) of the model parameters when X follows a Bernoulli distribution and the error has a truncated logistic distribution.*

				Ours			Lu (2010)			Zhang-Peng (2007)			Scolas et al. (2016)		
γ_0	τ_C	n	Par.	Bias	Var	MSE	Bias	Var	MSE	Bias	Var	MSE	Bias	Var	MSE
0.5	20	200	γ_0	.011	.054	.054	.035	.055	.056	.004	.051	.051	−.021	.052	.052
			γ_1	−.104	.110	.121	.226	.210	.261	.050	.167	.170	−.027	.127	.128
			β_1	−.221	.105	.154	.311	.374	.471	.058	.213	.216	.014	.153	.153
		400	γ_0	.015	.029	.029	.046	.028	.030	.011	.027	.027	−.015	.026	.026
			γ_1	−.115	.060	.073	.197	.110	.149	.020	.074	.074	−.032	.063	.064
			β_1	−.231	.061	.114	.270	.188	.261	.023	.099	.100	.002	.075	.075
	100	200	γ_0	−.003	.043	.043	.002	.043	.043	−.003	.042	.042	−.004	.042	.042
			γ_1	−.009	.083	.083	.003	.085	.085	−.005	.084	.084	−.008	.083	.083
			β_1	−.057	.117	.120	.041	.121	.123	.018	.099	.099	−.002	.112	.112
		400	γ_0	.010	.024	.024	.012	.024	.024	.008	.024	.024	.008	.024	.024
			γ_1	−.010	.044	.044	−.001	.044	.044	−.007	.044	.044	−.011	.044	.044
			β_1	−.081	.065	.072	.014	.053	.053	.002	.046	.046	−.001	.095	.095
1	20	200	γ_0	.024	.067	.068	.057	.068	.071	.016	.063	.063	−.017	.061	.061
			γ_1	−.093	.131	.140	.284	.275	.356	.092	.219	.227	.006	.172	.172
			β_1	−.216	.092	.139	.232	.246	.300	.038	.160	.161	.010	.125	.125
		400	γ_0	.015	.039	.039	.068	.035	.040	.023	.032	.033	−.007	.034	.034
			γ_1	−.142	.072	.092	.225	.129	.180	.017	.097	.097	−.040	.085	.087
			β_1	−.228	.049	.101	.243	.140	.199	.016	.081	.081	.001	.067	.067
	100	200	γ_0	.008	.049	.049	.013	.050	.050	.008	.049	.049	.008	.049	.049
			γ_1	.015	.092	.092	.031	.093	.094	.019	.091	.091	.015	.090	.090
			β_1	−.045	.109	.111	.024	.092	.093	.008	.080	.080	−.006	.109	.109
		400	γ_0	.018	.029	.029	.022	.028	.028	.017	.028	.028	.018	.028	.028
			γ_1	−.021	.049	.049	−.006	.049	.049	−.016	.048	.048	−.020	.047	.047
			β_1	−.055	.053	.056	.023	.043	.044	.003	.036	.036	.004	.090	.090

Table 3. *Bias, variance and mean squared error (MSE) of the model parameters when X follows a Bernoulli distribution and the error has a truncated normal distribution.*

				Ours			Lu (2010)			Zhang-Peng (2007)			Scolas et al. (2016)		
γ_0	τ_C	n	Par.	Bias	Var	MSE	Bias	Var	MSE	Bias	Var	MSE	Bias	Var	MSE
.5	20	200	γ_0	.001	.046	.046	.006	.046	.046	−.000	.046	.046	.023	.047	.048
			γ_1	−.029	.099	.100	.017	.104	.104	−.005	.102	.102	.024	.110	.111
			β_1	−.054	.052	.055	.028	.047	.048	−.000	.038	.038	.027	.043	.044
		400	γ_0	.009	.025	.025	.009	.024	.024	.003	.024	.024	.016	.025	.025
			γ_1	−.043	.047	.049	.024	.047	.048	.002	.046	.046	.030	.049	.050
			β_1	−.077	.026	.032	.030	.019	.020	−.001	.016	.016	.043	.018	.020
	100	200	γ_0	−.004	.042	.042	−.002	.041	.041	−.004	.041	.041	−.002	.043	.043
			γ_1	−.004	.082	.082	−.004	.081	.081	−.005	.082	.082	−.003	.084	.084
			β_1	−.120	.076	.090	−.001	.033	.033	−.001	.031	.031	.012	.034	.034
		400	γ_0	.003	.024	.024	.004	.023	.023	.002	.023	.023	.002	.023	.023
			γ_1	.005	.045	.045	.003	.043	.043	.002	.043	.043	.005	.045	.045
			β_1	−.102	.048	.058	.001	.014	.014	−.005	.013	.013	.027	.012	.013
1	20	200	γ_0	.010	.051	.051	.016	.051	.051	.008	.050	.050	.039	.056	.058
			γ_1	−.016	.102	.102	.052	.110	.113	.025	.106	.107	.063	.123	.127
			β_1	−.053	.036	.039	.035	.034	.035	.004	.028	.028	.038	.032	.033
		400	γ_0	.016	.030	.030	.020	.028	.028	.012	.027	.027	.030	.029	.030
			γ_1	−.070	.060	.065	.018	.056	.056	−.006	.055	.055	.030	.059	.060
			β_1	−.063	.024	.028	.024	.016	.017	−.003	.014	.014	.044	.015	.017
	100	200	γ_0	.003	.045	.045	.004	.045	.045	.003	.045	.045	.004	.046	.046
			γ_1	.013	.084	.084	.017	.084	.084	.014	.084	.084	.018	.088	.088
			β_1	−.063	.047	.051	.005	.024	.024	.002	.022	.022	.023	.023	.024
		400	γ_0	.010	.026	.026	.012	.026	.026	.011	.027	.027	.014	.027	.027
			γ_1	−.004	.049	.049	−.004	.048	.048	−.007	.048	.048	−.006	.049	.049
			β_1	−.059	.030	.033	.001	.011	.011	−.002	.011	.011	.032	.010	.011

Table 4. *Bias, variance and mean squared error (MSE) of the model parameters when X follows a Bernoulli distribution and the error distribution is a truncated mixture of Weibull distributions.*

				Ours			Lu (2010)			Zhang-Peng (2007)			Scolas et al. (2016)		
γ_0	τ_C	n	Par.	Bias	Var	MSE	Bias	Var	MSE	Bias	Var	MSE	Bias	Var	MSE
.5	20	200	γ_0	−.003	.058	.058	−.009	.047	.047	−.007	.049	.049	.222	.066	.115
			γ_1	−.075	.147	.153	.008	.115	.115	.009	.122	.122	1.749	6.589	9.647
			β_1	.003	.013	.013	.005	.003	.003	.004	.003	.003	.377	.009	.151
		400	γ_0	−.002	.034	.034	.003	.024	.024	.003	.023	.023	.234	.030	.085
			γ_1	−.128	.088	.105	−.005	.052	.052	.001	.052	.052	1.255	1.035	2.610
			β_1	−.018	.010	.010	.006	.001	.001	−.000	.002	.002	.364	.006	.139
	100	200	γ_0	−.004	.043	.043	.001	.040	.040	−.004	.043	.043	.014	.045	.045
			γ_1	−.010	.086	.086	−.010	.087	.087	−.005	.089	.089	.043	.094	.096
			β_1	−.012	.008	.008	.001	.002	.002	.002	.002	.002	.176	.018	.049
		400	γ_0	.002	.021	.021	−.002	.020	.020	.001	.020	.020	.016	.021	.021
			γ_1	−.008	.041	.041	−.001	.039	.039	−.001	.041	.041	.049	.043	.046
			β_1	−.015	.005	.006	.000	.001	.001	.000	.001	.001	.204	.003	.045
1	20	200	γ_0	.002	.068	.068	.017	.058	.058	.012	.059	.059	.330	.094	.203
			γ_1	−.114	.176	.189	−.002	.143	.143	.003	.134	.134	5.825	37.770	71.699
			β_1	−.003	.011	.011	.007	.003	.003	.003	.003	.003	.344	.006	.125
		400	γ_0	−.006	.046	.046	.008	.036	.036	.011	.031	.031	.326	.048	.154
			γ_1	−.153	.113	.136	.005	.079	.079	−.006	.071	.071	5.110	22.566	48.676
			β_1	−.017	.007	.007	.006	.001	.001	.001	.001	.001	.346	.004	.124
	100	200	γ_0	.007	.052	.052	.011	.051	.051	.009	.052	.052	.030	.056	.056
			γ_1	.004	.101	.101	.008	.103	.103	.009	.103	.103	.068	.109	.114
			β_1	−.007	.006	.006	.003	.002	.002	.004	.002	.002	.202	.012	.053
		400	γ_0	.008	.027	.027	.004	.026	.026	.008	.027	.027	.028	.028	.029
			γ_1	−.010	.052	.052	.006	.053	.053	−.003	.052	.052	.057	.057	.060
			β_1	−.012	.004	.004	.002	.001	.001	.001	.001	.001	.212	.003	.048

Table 5. *Coverage probabilities (CP) of* 95% *confidence intervals for the model parameters, the average length of these intervals (CI L), and the Pearson correlation (QQr) when X follows a Bernoulli distribution. All intervals are based on bootstrap standard errors except the column indicated as 'Lu Method (2010)', which is based on the method proposed in Lu (2010).*

				Ours			Lu (2010)			Lu Method (2010)		Zhang-Peng (2007)			Scolas et al. (2016)		
γ_0	τ_C	n	Par.	CP	CI L	QQr	CP	CI L	QQr	CP	CI L	CP	CI L	QQr	CP	CI L	QQr
				Logistic error													
.5	20	200	γ_0	.960	.990	.999	.753	.772	.999	.958	1.288	.960	.964	.999	.964	.939	.999
			γ_1	.950	1.407	.999	.870	1.318	.997	.922	2.586	.982	1.902	.985	.970	1.833	.997
			β_1	.900	1.274	.998	.701	1.497	.999	.738	1.937	.950	1.804	.997	.970	1.655	.998
.5	100	200	γ_0	.954	.842	.999	.917	.736	.999	.990	2.819	.954	.841	.999	.958	.843	.999
			γ_1	.966	1.187	.998	.944	1.147	.999	.984	3.274	.962	1.188	.999	.964	1.192	.999
			β_1	.942	1.314	.998	.929	1.328	.999	.882	1.363	.928	1.193	.998	.970	1.400	.998
				Normal error													
.5	20	200	γ_0	.960	.880	.997	.880	.688	.998	.980	1.304	.960	.874	.998	.962	.891	.998
			γ_1	.962	1.266	.998	.877	1.159	.998	.960	1.720	.964	1.279	.998	.964	1.323	.998
			β_1	.928	.863	.998	.880	.797	.998	.866	.755	.954	.765	.998	.938	.778	.998
.5	100	200	γ_0	.966	.843	.997	.942	.777	.997	.992	3.689	.968	.841	.997	.966	.846	.997
			γ_1	.968	1.178	.999	.958	1.159	.999	.996	4.177	.968	1.172	.999	.960	1.182	.999
			β_1	.860	.913	.987	.964	.746	.997	.970	.832	.944	.670	.997	.930	.686	.999

coverage probabilities are close to their nominal value 0.95, although for Lu's method the bootstrap standard errors often lead to too low coverage. Lu (2010) provides an alternative method, based on the inversion of the Fisher information matrix, which yields better results, but the intervals are considerably wider in that case than for the other methods.

The performance of the SNP method depends on the error distribution. For the standard normal distribution the method outperforms the three other methods, which is not surprising since in that case the true distribution belongs to the family of basis functions. For the other distributions the other methods have lower bias and variance. This is especially the case for the mixture of Weibull distributions, where the SNP approach has a very poor and sometimes even dramatic behavior, both in terms of bias and variance. This can be explained by the fact that this distribution cannot be well approximated by the basis functions, which are normal and exponential distributions enriched with polynomials.

While Tables 2–4 show that Lu's method performs well in practice, there is also a downside or weakness of this method. In the case of the mixture of two Weibulls, the method often has convergence problems, leading to errors or warnings when running the method in R. Table 6 shows the number of errors/warnings under each scenario when ϵ follows the mixture of Weibull distributions. Whenever an error in one of the estimation methods occurs (usually this happens with Lu's method, but occasionally also with one of the other methods), that sample is removed for all estimation methods and a new sample is taken to reach the required number of 500 simulation runs. The table shows that Lu's method faces indeed a lot of convergence issues, especially for large sample sizes and scenarios with heavy censoring. Note however that these convergence issues are almost absent in the case of the logistic or normal error distribution, so the results in the table cannot be generalized to other distributions.

TABLE 6

The frequency of errors that have occurred in the simulations (out of 500 samples), when the error distribution is a mixture of Weibull distributions.

γ_0	τ_C	n	Ours	Lu	Zhang-Peng	Scolas et al
0.5	20	200	0	37	0	0
		400	0	265	0	0
	100	200	0	12	0	0
		400	0	37	0	0
1	20	200	0	69	0	0
		400	2	261	0	0
	100	200	0	12	0	0
		400	0	34	0	0

Since our method and the method of [17] depend on a bandwidth, it is important to investigate the effect of the bandwidth on the performance of these two estimation methods. Table 7 shows the results when the error distribution is the mixture of two Weibull distributions, for $n = 200$ and for three choices of the bandwidth, namely $b/2$, b and $2b$, where b is selected as before. The table shows that both methods are robust to alterations of the bandwidth. However, when $\tau_C = 20$ (corresponding to the heavy censoring case), the results of our method are more stable than those of Lu's method. Also, note that the results for bandwidth b do not coincide with those in Table 4. This is because we replace a sample by another sample as soon as there is a convergence issue for at least one bandwidth or method (as for Table 6).

TABLE 7
Bias, variance and mean squared error (MSE) of the model parameters for $n = 200$ and for three values of the bandwidth, when X follows a Bernoulli distribution and the error distribution is a mixture of Weibull distributions.

				$b/2$			b			$2b$		
γ_0	τ_C	Method	Par.	Bias	Var	MSE	Bias	Var	MSE	Bias	Var	MSE
0.5	20	Ours	γ_0	.009	.051	.051	.009	.051	.051	.008	.051	.051
			γ_1	.000	.112	.112	.002	.113	.113	.007	.114	.114
			β_1	.000	.006	.006	.005	.004	.004	.010	.004	.004
		Lu	γ_0	.007	.049	.049	.010	.049	.049	.026	.051	.052
			γ_1	.005	.115	.115	.018	.118	.118	.068	.127	.132
			β_1	.006	.006	.006	.011	.004	.004	.025	.004	.005
	100	Ours	γ_0	−.002	.041	.041	−.002	.041	.041	−.003	.041	.041
			γ_1	.015	.075	.075	.015	.075	.075	.015	.075	.075
			β_1	.002	.005	.005	.001	.003	.003	.003	.003	.003
		Lu	γ_0	−.001	.041	.041	−.001	.041	.041	.001	.041	.041
			γ_1	.016	.075	.075	.017	.075	.075	.021	.075	.075
			β_1	.001	.004	.004	.003	.003	.003	.006	.003	.003
1	20	Ours	γ_0	.023	.066	.067	.023	.066	.067	.022	.067	.067
			γ_1	.029	.143	.144	.029	.141	.142	.034	.141	.142
			β_1	.002	.004	.004	.006	.003	.003	.011	.003	.003
		Lu	γ_0	.017	.064	.064	.023	.065	.066	.045	.067	.069
			γ_1	.026	.147	.148	.041	.149	.151	.116	.166	.179
			β_1	.005	.004	.004	.010	.003	.003	.023	.004	.005
	100	Ours	γ_0	.006	.050	.050	.006	.050	.050	.006	.050	.050
			γ_1	.007	.088	.088	.007	.089	.089	.007	.088	.088
			β_1	.001	.003	.003	.000	.002	.002	.001	.002	.002
		Lu	γ_0	.006	.050	.050	.006	.050	.050	.045	.067	.069
			γ_1	.009	.088	.088	.010	.089	.089	.002	.002	.002
			β_1	.009	.050	.050	.015	.089	.089	.004	.002	.002

The frequency of errors that occur also depends in a crucial way on the bandwidth used for our and Lu's method, as can be seen in Table 8. The table shows the number of samples that needs to be generated under a given scenario in order to obtain 500 samples for which no convergence problems exist. The table shows that such problems occur more often when the bandwidth is small.

TABLE 8
Number of needed simulations to obtain 500 successful fits in all methods, where the error distribution is a mixture of Weibull distributions and $n = 200$.

			$b/2$		b		$2b$	
γ_0	τ_C	Method	Success	Error	Success	Error	Success	Error
0.5	20	Ours	2285	4	2288	1	2289	0
		Lu	*500*	1789	2006	283	2288	1
	100	Ours	1054	1	1055	0	1055	0
		Lu	*500*	555	1031	24	1055	0
1	20	Ours	2309	2	2311	0	2311	0
		Lu	*500*	1811	2043	268	2310	1
	100	Ours	950	0	950	0	950	0
		Lu	*500*	450	930	20	950	0

Finally, we study the computation time of the four studied methods. Table 9 shows the average computation time in seconds (over 100 samples) in the case of the logistic error distribution with $\gamma = 0.5$. The table shows that the fastest method is the SNP approach, whereas the three others have more comparable computation times, with Lu's method being however the slowest of all methods.

Table 9
Computation time in seconds for the logistic error distribution with $\gamma_0 = 0.5$.

n	τ_C	Ours	Lu	Zhang-Peng	Scolas et al
200	20	1.65	3.89	1.37	0.43
	100	1.38	2.59	0.90	0.44
400	20	5.27	12.25	3.79	0.69
	100	3.87	10.87	2.31	0.67

We end this section with plots of the estimated error densities $\hat{f}_{0,\hat{\beta}}$ for 20 arbitrary samples of size $n = 400$ generated from a logistic, a normal and a mixture of Weibull densities. They are given in Figure 1 for $\gamma_0 = 0.5$, $\tau_C = 20$ or 100 and for a uniform covariate X. The plots show that the estimated curves are quite close to the true curves for all considered settings.

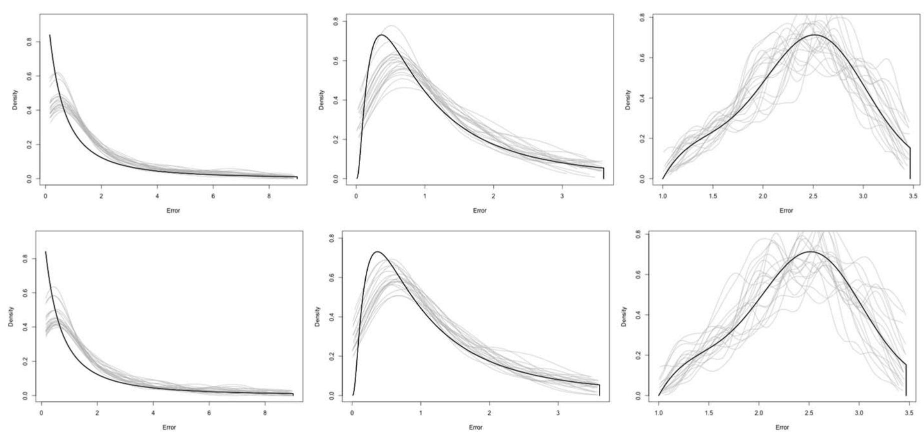

Fig 1. *Plots of the estimated error densities $\hat{f}_{0,\hat{\beta}}$ for 20 arbitrary samples of size $n = 400$ generated from a logistic density (first column), a normal density (second column), and a mixture of Weibull densities (third column). The first row corresponds to $\tau_C = 100$, the second row to $\tau_C = 20$. The covariate X follows a uniform distribution, and $\gamma_0 = 0.5$.*

To conclude, the simulations showed that the proposed method works well in practice under various model settings. It has the advantage of working well under all model settings (whereas the method of [26] does not work well for certain error distributions), it does not have any convergence problems (contrary to [17], which suffers sometimes from such problems), we developed rigorous asymptotic theory for the proposed estimator (which is not the case for the estimators of [26] and [31]), and it is the only method that has been used so far for variable selection in the AFT mixture cure model. For this we refer to

[20], who developed a penalized likelihood approach based on adaptive LASSO penalties to do variable selection both for the incidence and the latency.

5. Real data application

As an application of our estimation method, we study breast cancer data of 286 patients who experienced lymph-node-negative breast cancer between 1980 and 1995 [29]. The event of interest is distant metastasis, and the associated survival time is the time to distant metastasis (DM). Among the 286 patients, 107 experienced a distant recurrence from breast cancer. Figure 2 shows the Kaplan-Meier estimator of the survival function, from which it is clear that there is an overall cure fraction of about 60%. Moreover, the plateau is very long and contains 88% of the censored observations, which indicates that the follow-up period is sufficiently long [1].

The data set also contains four covariates: the age of the patient (ranging from 26 to 83), the estrogen receptor (ER) status (where 0 signifies ER $-$, defined as less than 10 fmol/mg protein, and 1 signifies ER $+$, defined as at least 10 fmol/mg protein), the size of the tumor (ranging from 1 to 4), and the menopausal status (where 0 means pre-menopausal defined as age ≤ 50, and 1 means post-menopausal meaning age > 50). We suppose that the AFT/logistic mixture cure model is valid for these data, and we estimate the model using the proposed approach, and also using the method of [31] (using the R package *smcure*), the kernel approach of [17] and the SNP method of [26]. The bandwidth is calculated in the same way as in the simulation study, and the initial values are obtained using the *survreg* function in *R* for the AFT model, and using the *glm* function for the logistic model.

Table 10 shows the estimated parameters, the estimated standard errors, the Wald statistics and the corresponding P-values for the four available methods. For all methods except for Lu's method, the standard errors are obtained from 500 bootstrap samples drawn with replacement from the original sample,

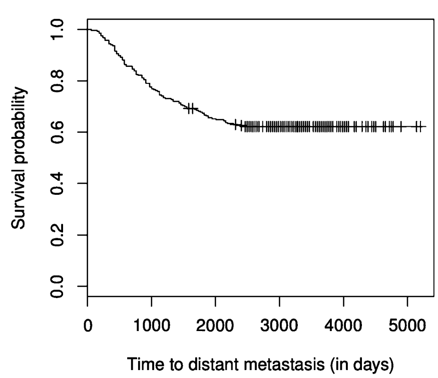

FIG 2. *Kaplan-Meier estimator of the survival function for the breast cancer data.*

whereas Lu's method uses the inverse Fisher information matrix to estimate the covariance matrix. The table shows that for all methods except the SNP approach of [26], the signs of the estimated coefficients are in agreement and the estimated parameters are close to each other. The coefficient of tumor size in the AFT model is significant according to these three methods. Finally, the SNP approach gives quite different results, both in terms of the significance of the coefficients, their size, and their sign. This can be explained by the fact that the estimated error density, given in Figure 3, is bimodal, and we know from the simulation study in Section 4 that the SNP approach is not able to approximate well bimodal densities.

Table 10

Estimated parameters, estimated standard errors (SE), Wald statistics and corresponding P-values using the four available methods for the breast cancer data. P-values that are significant at the 0.05 level are indicated by a $*$.

Method	Model	Variable	Est.par.	SE	Wald	P-value
Ours	Incidence	Intercept	.131	.607	.216	.829
		Age	−.012	.010	−1.18	.239
		ER	.230	.358	.643	.520
		Tumor size	−.085	.190	−.445	.656
		Menopausal	−.068	.330	−.207	.836
	Latency	Age	.002	.007	.265	.791
		ER	.309	.207	1.49	.136
		Tumor size	−.310	.158	−1.96	.050*
		Menopausal	.292	.205	1.43	.154
Lu	Incidence	Intercept	.231	.703	.328	.743
		Age	−.012	.039	−.310	.756
		ER	.237	.343	.689	.491
		Tumor size	−.133	.261	−.511	.610
		Menopausal	−.018	.594	−.030	.976
	Latency	Age	.004	.007	.567	.570
		ER	.278	.239	1.16	.245
		Tumor size	−.335	.168	−2.00	.046*
		Menopausal	.350	.140	2.50	.012*
Zhang-Peng	Incidence	Intercept	.157	.690	.228	.819
		Age	−.011	.011	−1.04	.299
		ER	.246	.363	.625	.532
		Tumor size	−.117	.221	−.528	.598
		Menopausal	−.044	.394	−.112	.911
	Latency	Age	.006	.006	.993	.320
		ER	.302	.248	1.21	.225
		Tumor size	−.340	.150	−2.27	.023*
		Menopausal	.364	.241	1.51	.131
Scolas et al	Incidence	Intercept	−3.59	.857	−4.19	.000*
		Age	.051	.015	3.44	.001*
		ER	.654	.694	.942	.346
		Tumor size	.792	.381	2.08	.037*
		Menopausal	−.047	.605	−.078	.937
	Latency	Age	.095	.016	5.80	.000*
		ER	.441	.517	.853	.394
		Tumor size	.653	.406	1.61	.108
		Menopausal	.771	.421	1.83	.067

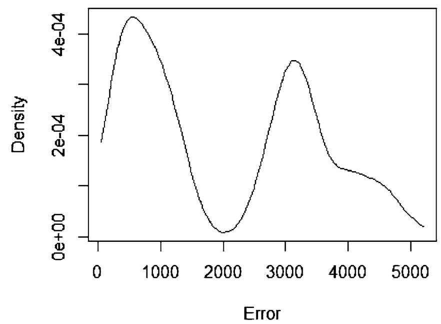

FIG 3. *Estimated density of* $\exp(\epsilon)$ *for the breast cancer data.*

Appendix A: Proofs

A.1. Definitions and assumptions

Here, we provide some necessary definitions and the conditions under which our asymptotic results are valid.

First of all, as explained already earlier, we will use the results in [5] to show the consistency and asymptotic normality of our estimators. The latter paper gives sufficient conditions under which Z-estimators in a semiparametric model based on a non-smooth criterion function, are consistent and asymptotically normal. We will suppose that the vector of nuisance functions $h_0 = (S_0, f_0, f_0')$ belongs to the space $\mathcal{H} = \mathcal{H}_1 \times \mathcal{H}_2 \times \mathcal{H}_3$, where

$$\mathcal{H}_1 = \Big\{ g : [0, \tau_{\max}] \to [0,1] : g \text{ is decreasing} \Big\}$$
$$\mathcal{H}_3 = \Big\{ g : [0, \tau_{\max}] \to \mathcal{R} : g \text{ is differentiable}, \sup_{t \le \tau_{\max}} |g^{(k)}(t)| \le M, k = 0, 1 \Big\}$$
$$\mathcal{H}_2 = \Big\{ g \in \mathcal{H}_3 : \inf_{t \le \tau_{\max}} g(t) > \zeta \Big\}$$

for some $M < \infty$ and some $\zeta > 0$, where $\tau_{\max} = \max_{\beta \in \mathcal{B}} \tau(\beta)$. For $h \in \mathcal{H}$, define $\|h\|_{\mathcal{H}} = \max(\|h_1\|_{\mathcal{H}}, \|h_2\|_{\mathcal{H}}, \|h_3\|_{\mathcal{H}})$, where $\|h_j\|_{\mathcal{H}} = \sup_{\beta \in \mathcal{B}} \sup_{t < \tau(\beta)} |h_j(t, \beta)|$, $h(t, \beta) = (h_1(t, \beta), h_2(t, \beta), h_3(t, \beta))$ and $h_0(t, \beta) = (S_{0,\beta}(t), f_{0,\beta}(t), f'_{0,\beta}(t))$. Finally, define $G_{0,\beta}(t) = P(C \exp(-\beta^t X) \le t)$ for any $\beta \in \mathcal{B}$, and let $S_{0,\beta}(\{t\}) = S_{0,\beta}(t-) - S_{0,\beta}(t)$ be the point mass of $S_{0,\beta}$ at t.

We will make use of the following theorems, which are Theorems 1 and 2 in [5]. They give high-level conditions under which $\hat{\theta}$ is respectively weakly consistent and asymptotically normal. In the next two subsections, we will check these high-level conditions for our estimator.

Theorem A.1. *Suppose that* $\theta_0 \in \Theta$ *satisfies* $M(\theta_0, h_0) = 0$*, and that:*

(1.1) $\|M_n(\hat{\theta}, \hat{h})\| \le \inf_{\theta \in \Theta} \|M_n(\theta, \hat{h})\| + o_P(1)$.

(1.2) *For all* $\delta > 0$*, there exists* $\epsilon(\delta) > 0$ *such that* $\inf_{\|\theta-\theta_0\|>\delta} \|M(\theta, h_0)\| \geq \epsilon(\delta) > 0$.

(1.3) *Uniformly for all* $\theta \in \Theta$*,* $M(\theta, h)$ *is continuous (w.r.t.* $\|\cdot\|_{\mathcal{H}}$*) in* h *at* $h = h_0$.

(1.4) $\|\hat{h} - h_0\|_{\mathcal{H}} = o_P(1)$.

(1.5) *For all sequences of positive numbers* δ_n *with* $\delta_n = o(1)$,

$$\sup_{\theta\in\Theta, \|h-h_0\|_{\mathcal{H}}\leq\delta_n} \|M_n(\theta, h) - M(\theta, h)\| = o_P(1),$$

Then, $\hat{\theta} - \theta_0 = o_P(1)$.

For the next result, we define the matrix of partial derivatives $\Gamma_1(\theta, h) = (\partial/\partial\theta)M(\theta, h(\cdot, \beta))$, which satisfies

$$\Gamma_1(\theta, h)(\bar{\theta} - \theta) = \lim_{\tau\to 0} \frac{1}{\tau}\Big[M\big(\theta + \tau(\bar{\theta} - \theta), h(\cdot, \beta + \tau(\bar{\beta} - \beta))\big) - M\big(\theta, h(\cdot, \beta)\big)\Big]$$

for $\bar{\theta} = (\bar{\gamma}, \bar{\beta})^t \in \Theta$, and we let $\Gamma_1 = \Gamma_1(\theta_0, h_0)$. For any $\theta \in \Theta$, we say that $M(\theta, h)$ is pathwise differentiable at $h \in \mathcal{H}$ in the direction $[\bar{h}-h]$ if $\{h+\tau(\bar{h}-h) : \tau \in [0, 1]\} \subset \mathcal{H}$ and if

$$\Gamma_2(\theta, h)[\bar{h} - h] = \lim_{\tau\to 0} \frac{1}{\tau}\Big[M\big(\theta, h(\cdot, \theta) + \tau(\bar{h}(\cdot, \theta) - h(\cdot, \theta))\big) - M\big(\theta, h(\cdot, \theta)\big)\Big]$$

exists. Also, for any $\delta > 0$, let $\Theta_\delta = \{\theta \in \Theta : \|\theta - \theta_0\| \leq \delta\}$ and $\mathcal{H}_\delta = \{h \in \mathcal{H} : \|h - h_0\|_{\mathcal{H}} \leq \delta\}$.

Theorem A.2. *Suppose that* $\theta_0 \in \Theta$ *satisfies* $M(\theta_0, h_0) = 0$*, that* $\hat{\theta} - \theta_0 = o_P(1)$*, and that:*

(2.1) $\|M_n(\hat{\theta}, \hat{h})\| = \inf_{\theta\in\Theta} \|M_n(\theta, \hat{h})\| + o_P(n^{-1/2})$.

(2.2) *For* $\theta \in \Theta$*, the matrix* $\Gamma_1(\theta, h_0)$ *exists and is continuous at* $\theta = \theta_0$*, and* Γ_1 *has full rank.*

(2.3) *For all* $\theta \in \Theta$ *the functional derivative* $\Gamma_2(\theta, h_0)[h - h_0]$ *exists in all directions* $[h - h_0] \in \mathcal{H}$*, and for all* $(\theta, h) \in \Theta_{\delta_n} \times \mathcal{H}_{\delta_n}$ *with a positive sequence* $\delta_n = o(1)$*:*
(i) $\|M(\theta, h) - M(\theta, h_0) - \Gamma_2(\theta, h_0)[h - h_0]\| \leq c\|h - h_0\|^2_{\mathcal{H}}$ *for some* $c < \infty$,
(ii) $\|\Gamma_2(\theta, h_0)[h - h_0] - \Gamma_2(\theta_0, h_0)[h - h_0]\| \leq o(1)\delta_n$.

(2.4) $P(\hat{h} \in \mathcal{H}) \to 1$*, and* $\|\hat{h} - h_0\|_{\mathcal{H}} = o_P(n^{-1/4})$.

(2.5) *For all sequences of positive numbers* $\{\delta_n\}$ *with* $\delta_n = o(1)$,

$$\sup_{\|\theta-\theta_0\|\leq\delta_n, \|h-h_0\|_{\mathcal{H}}\leq\delta_n} \|M_n(\theta, h) - M(\theta, h) - M_n(\theta_0, h_0)\| = o_P(n^{-1/2}).$$

(2.6) *For some finite matrix* S*,* $n^{1/2}\{M_n(\theta_0, h_0)+\Gamma_2(\theta_0, h_0)[\hat{h}-h_0]\} \xrightarrow{d} N(0, S)$.

Then, $n^{1/2}(\hat{\theta} - \theta_0) \xrightarrow{d} N(0, \Sigma)$*, where* $\Sigma = \Gamma_1^{-1}\, S\, \Gamma_1^{-1}$.

To establish the asymptotic results regarding our estimator $\hat{\theta}$ we need to impose the following assumptions:

(C1) The covariate vectors X and Z have compact support, denoted by R_X and R_Z. The true vector θ_0 belongs to the interior of Θ, and Θ is compact.
(C2) The kernel K is symmetric of order larger than 3, K is twice continuously differentiable with support $[-1,1]$, $K(\pm 1) = K'(\pm 1) = K''(\pm 1) = 0$.
(C3) The bandwidth b_n satisfies $nb_n^6(\log n)^{-2} \to \infty$ and $nb_n^8 \to 0$.
(C4) For all $\beta \in \mathcal{B}$, $S_{0,\beta}(t)$ is 6 times continuously differentiable in t for $t \in [0, \tau(\beta))$, $\sup_{\beta\in\mathcal{B}} \sup_{t<\tau(\beta)} |f_{0,\beta}^{(k)}(t)| < \infty$ for $k = 0, 1, \dots, 5$, and $\inf_{\beta\in\mathcal{B}} S_{0,\beta}(\{\tau(\beta)\}) > 0$.
(C5) For all $\beta \in \mathcal{B}$, $G_{0,\beta}(t)$ is continuous in t for $t \in [0, \tau(\beta))$, and $\inf_{\beta\in\mathcal{B}} \big(1 - G_{0,\beta}(\tau(\beta))\big) > 0$.
(C6) $\sup_{x,y} f_{Y|X}(y|x) < \infty$ and $\sup_x f_X(x) < \infty$.
(C7) For all $\delta > 0$, there exists $\epsilon(\delta) > 0$ such that $\inf_{\|\theta-\theta_0\|>\delta} \|M(\theta, h_0)\| \geq \epsilon(\delta) > 0$.
(C8) The matrix Γ_1 has full rank.

In the following subsections, we provide the proofs of Theorems 3.1, 3.2 and 3.3 under assumptions (C1)-(C8).

A.2. Proof of Theorem 3.1

We will verify conditions (1.1)-(1.5) of Theorem A.1, from which the stated result will follow. First, condition (1.1) holds true by definition of the estimator $\hat{\theta}$, and condition (1.2) is given in assumption (C7). The continuity of $M(\theta, h)$ is straightforward under the given assumptions, so (1.3) is also verified. Condition (1.4) is verified thanks to Theorem 3.3. Finally, condition (1.5) is satisfied if the class $\{v \to m(v, \theta, h) : \theta \in \Theta, h \in \mathcal{H}\}$ is Glivenko-Cantelli. We will show in the proof of Theorem 3.2 below that this class is even Donsker, which implies that it is Glivenko-Cantelli (see p. 80-81 in [28] for the definition of Glivenko-Cantelli and Donsker classes). $\square$

A.3. Proof of Theorem 3.2

We will now verify conditions (2.1)-(2.6) of Theorem A.2. First, condition (2.1) holds true by definition of the estimator $\hat{\theta}$, whereas for condition (2.2) the matrix $\Gamma_1(\theta, h_0)$ can be obtained using straightforward calculations. The continuity of $\Gamma_1(\theta, h_0)$ follows from assumptions (C1) and (C4), whereas the full rank condition is stated in assumption (C8).

For condition (2.3) tedious but straightforward calculations show that $\Gamma_2(\theta, h_0)[h - h_0]$ can be obtained by applying Taylor expansions of order one of the function m with respect to the nuisance functions S_0, f_0 and f_0'. This gives the following formula for $\Gamma_2(\theta, h_0)[h - h_0] = (\Gamma_{2,1}(\theta, h_0)[h - h_0], \Gamma_{2,2}(\theta, h_0)[h - h_0])^t$, where $\Gamma_{2,j}(\theta, h_0)[h - h_0]$ is the functional derivative of $E[m_j(\theta, h_0)]$ in the direction $[h - h_0]$, $j = 1, 2$:

$$\Gamma_{2,1}(\theta, h_0)[h - h_0] = E\Big\{ \frac{(1-\Delta) Z e^{\gamma^t Z} (S_\beta - S_{0,\beta})(Y e^{-\beta^t X})}{(1 + e^{\gamma^t Z} S_{0,\beta}(Y e^{-\beta^t X}))^2} \Big\} \tag{11}$$

$$\begin{aligned}\Gamma_{2,2}(\theta,h_0)[h-h_0] = E\Big\{ & -\Delta XYe^{-\beta^tX}\Big[\frac{(f'_\beta - f'_{0,\beta})(Ye^{-\beta^tX})}{f_{0,\beta}(Ye^{-\beta^tX})} \\ & - \frac{f'_{0,\beta}(Ye^{-\beta^tX})(f_\beta - f_{0,\beta})(Ye^{-\beta^tX})}{f^2_{0,\beta}(Ye^{-\beta^tX})}\Big]\Big\} \\ & + E\Big\{(1-\Delta)XYe^{-\beta^tX}e^{\gamma^tZ}\Big[\frac{(f_\beta - f_{0,\beta})(Ye^{-\beta^tX})}{1+e^{\gamma^tZ}S_{0,\beta}(Ye^{-\beta^tX})} \\ & - \frac{f_{0,\beta}(Ye^{-\beta^tX})(S_\beta - S_{0,\beta})(Ye^{-\beta^tX})}{(1+e^{\gamma^tZ}S_{0,\beta}(Ye^{-\beta^tX}))^2}\Big]\Big\}.\end{aligned} \tag{12}$$

The verification of (2.3) (i) and (ii) requires lengthy calculations, based however on simple algebraic manipulations and Taylor expansions of the functions $\Gamma_{2,j}(\theta,h_0)[h-h_0]$ $(j=1,2)$ given in (11) and (12).

The second part of condition (2.4) follows from Theorem 3.3 and assumption (C3) on the bandwidth. Indeed, we need that $O((nb_n^3)^{-1/2}(\log n)^{1/2}) + O(b_n^4) = o(n^{-1/4})$, which is satisfied if $nb_n^6(\log n)^{-2} \to \infty$ and $nb_n^{16} \to 0$. For the first part, we need to show that $(\hat{S}_{0,\beta}, \hat{f}_{0,\beta}, \hat{f}'_{0,\beta}) \in \mathcal{H}$ with probability tending to one. For $\hat{S}_{0,\beta}$ this is obvious. To show that $\hat{f}_{0,\beta} \in \mathcal{H}_2$ and $\hat{f}'_{0,\beta} \in \mathcal{H}_3$, we need to show that

$$\sup_{\beta\in\mathcal{B}} \sup_{0\le t<\tau(\beta)} |\hat{f}^{(k)}_{0,\beta}(t)| \le M$$

with probability tending to one, for $k=0,1,2$. For $k=0,1$ this follows from Theorem 3.3. For $k=2$ the proof is similar as for Theorem 3.3, and allows to show that $\sup_{\beta\in\mathcal{B}}\sup_{0\le t<\tau(\beta)} |\hat{f}''_{0,\beta}(t) - f''_{0,\beta}(t)| = O_P((nb_n^5)^{-1/2}(\log n)^{1/2}) + O(b_n^4) = o_P(1)$.

For condition (2.5) we apply Theorem 3 in [5], which says that (2.5) is satisfied if for each component $m_{1,j}$ $(j=1,\dots,k+1)$ of m_1 and each component $m_{2,j}$ $(j=1,\dots,\ell)$ of m_2, we have (with $i=1,2$)

$$|m_{i,j}(v,\theta,h) - m_{i,j}(v,\tilde{\theta},\tilde{h})| \le b_{i,j}(v)\{\|\theta-\tilde{\theta}\| + \|h-\tilde{h}\|\},$$

with $E[b^2_{i,j}(V)] < \infty$, and if

$$\int_0^\infty \sqrt{\log N(\varepsilon,\mathcal{H}_j,\|\cdot\|_{\mathcal{H}})}\,d\varepsilon < \infty, \tag{13}$$

for $j=1,2,3$, where $N(\varepsilon,\mathcal{H}_j,\|\cdot\|_{\mathcal{H}})$ is the ε-covering number of the class $\mathcal{H}_j$ with respect to the $\|\cdot\|_{\mathcal{H}}$-norm (see p. 83 in [28] for the definition of the covering number). The first requirement is easily seen to be satisfied thanks to the smoothness of the function m, whereas for the second one we apply Theorem 2.7.2 in [28] for $\mathcal{H}_2$ and $\mathcal{H}_3$, and Theorem 2.7.5 in [28] for $\mathcal{H}_1$, together with the fact that the covering number is bounded by the bracketing number (see p. 84 in [28]). This shows that $\log N(\varepsilon,\mathcal{H}_j,\|\cdot\|_{\mathcal{H}}) \le K\varepsilon^{-1}$, and hence the integral in (13) is bounded by $2(K\max\{2M,1\})^{1/2}$, since for $\varepsilon > \max\{2M,1\}$ one ε-ball suffices to cover the space $\mathcal{H}_j$.

It remains to verify condition (2.6). First note that it follows from (11) and (12) that $\Gamma_2(\theta_0, h_0)[\hat{h}-h_0]$ can be written as $E[G_1(V)\{\hat{S}_0-S_0\}(e^\epsilon)+G_2(V)\{\hat{f}_0-f_0\}(e^\epsilon)+G_3(V)\{\hat{f}_0'-f_0'\}(e^\epsilon)]$ for certain vectors of functions G_1, G_2 and G_3. We know from [13] and [18] that

$$\hat{S}_0(t) - S_0(t) = n^{-1}\sum_{i=1}^n \xi(Y_i e^{-\beta_0^t X_i}, \Delta_i, t) + O_P(n^{-1}\log n)$$

uniformly in $0 \le t < \tau_0$, where

$$\xi(e,\delta,t) = S_0(t)\Big\{\frac{I(e\le t, \delta=1)}{1-H_0(e)} - \int_0^{\min(e,t)} \frac{dH_0^1(s)}{(1-H_0(s))^2}\Big\},$$

$H_0(t) = P(Ye^{-\beta^t X} \le t)$ and $H_0^1(t) = P(Ye^{-\beta^t X} \le t, \Delta = 1)$. Using this i.i.d. representation, we can also decompose $\hat{f}_0(t) - f_0(t)$ in a sum of independent terms and a remainder term of smaller order:

$$\begin{aligned}
\hat{f}_0(t) - f_0(t) &= b^{-1}\int K\Big(\frac{t-s}{b}\Big) d(\hat{F}_0(s) - F_0(s)) + O(b_n^4) \\
&= -b^{-1}\int K(u) d(\hat{F}_0(t-ub) - F_0(t-ub)) + O(b_n^4) \\
&= b^{-1}\int (\hat{F}_0(t-ub) - F_0(t-ub))K'(u)du + O(b_n^4) \\
&= -(nb)^{-1}\sum_{i=1}^n \int \xi(Y_i e^{-\beta_0^t X_i}, \Delta_i, t-ub)K'(u)du \\
&\quad + O_P((nb_n)^{-1}\log n) + O(b_n^4) \\
&= -(nb)^{-1}\sum_{i=1}^n \int \xi(Y_i e^{-\beta_0^t X_i}, \Delta_i, t-ub)K'(u)du + o_P(n^{-1/2}),
\end{aligned}$$

since $nb_n^8 \to 0$ and $nb_n^2(\log n)^{-2} \to \infty$. Note that the order $O(b_n^4)$ of the bias term follows from the fact that the order of the kernel K is larger than 3 (see the proof of Theorem 3.3 (ii) for more details). Similarly we can show that

$$\hat{f}_0'(t) - f_0'(t) = -(nb^2)^{-1}\sum_{i=1}^n \int \xi(Y_i e^{-\beta_0^t X_i}, \Delta_i, t-ub)K''(u)du + o_P(n^{-1/2}),$$

since $nb_n^4(\log n)^{-2} \to \infty$. We can now write

$$\begin{aligned}
&M_n(\theta_0, h_0) + \Gamma_2(\theta_0, h_0)[\hat{h} - h_0] \\
&= n^{-1}\sum_{i=1}^n m(V_i, \theta_0, h_0) \\
&\quad + E\big[G_1(V)\{\hat{S}_0 - S_0\}(e^\epsilon) + G_2(V)\{\hat{f}_0 - f_0\}(e^\epsilon) + G_3(V)\{\hat{f}_0' - f_0'\}(e^\epsilon)\big] \\
&= n^{-1}\sum_{i=1}^n m(V_i, \theta_0, h_0) + n^{-1}\sum_{i=1}^n E\Big[G_1(V)\xi(Y_i e^{-\beta_0^t X_i}, \Delta_i, e^\epsilon)\Big]
\end{aligned}$$

$$- (nb)^{-1} \sum_{i=1}^n E\Big[G_2(V) \int \xi(Y_i e^{-\beta_0^t X_i}, \Delta_i, e^{\epsilon} - ub) K'(u) du\Big]$$
$$- (nb^2)^{-1} \sum_{i=1}^n E\Big[G_3(V) \int \xi(Y_i e^{-\beta_0^t X_i}, \Delta_i, e^{\epsilon} - ub) K''(u) du\Big] + o_P(n^{-1/2}).$$

Let $L_j(V_i, w) = E\big[G_j(V)\xi(Y_i e^{-\beta_0^t X_i}, \Delta_i, e^{\epsilon} - w)\big]$, $j = 1, 2, 3$, where the expected value is taken with respect to V, conditional on the i-th data point V_i. Then, with $L_j^{(k)}(V, w) = (\partial^k/\partial w^k) L_j(V, w)$,

$$\begin{aligned}
&M_n(\theta_0, h_0) + \Gamma_2(\theta_0, h_0)[\hat h - h_0] \\
&= n^{-1} \sum_{i=1}^n \Big\{ m(V_i, \theta_0, h_0) + L_1(V_i, 0) - b^{-1} \int L_2(V_i, ub) K'(u) du \\
&\quad - b^{-2} \int L_3(V_i, ub) K''(u) du \Big\} + o_P(n^{-1/2}) \\
&= n^{-1} \sum_{i=1}^n \big\{ m(V_i, \theta_0, h_0) + L_1(V_i, 0) \big\} \\
&\quad - (nb)^{-1} \sum_{i=1}^n \int \Big[\sum_{k=0}^4 \frac{1}{k!} L_2^{(k)}(V_i, 0)(ub)^k + \frac{1}{5!} L_2^{(5)}(V_i, \eta_2)(ub)^5 \Big] K'(u) du \Big\} \\
&\quad - (nb^2)^{-1} \sum_{i=1}^n \int \Big[\sum_{k=0}^5 \frac{1}{k!} L_3^{(k)}(V_i, 0)(ub)^k + \frac{1}{6!} L_3^{(6)}(V_i, \eta_3)(ub)^6 \Big] K''(u) du \Big\} \\
&\quad + o_P(n^{-1/2}),
\end{aligned}$$

for some values η_2 and η_3 between 0 and ub. We have that $\int u^k K'(u) du = 0$ for $k = 0, 2, 3, 4$, $\int u K'(u) du = -1$, $\int u^k K''(u) du = 0$ for $k = 0, 1, 3, 4, 5$, $\int u^2 K''(u) du = 2$. It follows that

$$\begin{aligned}
&M_n(\theta_0, h_0) + \Gamma_2(\theta_0, h_0)[\hat h - h_0] \\
&= n^{-1} \sum_{i=1}^n \Big\{ m(V_i, \theta_0, h_0) + L_1(V_i, 0) + L_2'(V_i, 0) + L_3''(V_i, 0) \Big\} + o_P(n^{-1/2}),
\end{aligned}$$

since $nb^8 \to 0$. Hence, $n^{1/2}(M_n(\theta_0, h_0) + \Gamma_2(\theta_0, h_0)[\hat h - h_0])$ converges to a zero mean normal vector with covariance matrix $S = E[s(V)s(V)^t]$, where $s(V) = m(V, \theta_0, h_0) + L_1(V, 0) + L_2'(V, 0) + L_3''(V, 0)$. It now follows from Theorem A.2 that $n^{1/2}(\hat\theta - \theta_0)$ converges to a zero mean normal vector with covariance matrix $\Gamma_1^{-1} S \Gamma_1^{-1}$. □

A.4. Proof of Theorem 3.3

In the proof we will show that the stated results are valid if certain results hold for the estimators

$$\hat{H}_{0,\beta}(t) = n^{-1}\sum_{i=1}^{n} I(Y_i e^{-\beta^t X_i} \le t)$$
$$\hat{H}^1_{0,\beta}(t) = n^{-1}\sum_{i=1}^{n} I(Y_i e^{-\beta^t X_i} \le t, \Delta_i = 1).$$

These are estimators of the distribution $H_{0,\beta}(t) = P(Ye^{-\beta^t X} \le t)$ of the observed survival times, and the subdistribution $H^1_{0,\beta}(t) = P(Ye^{-\beta^t X} \le t, \Delta = 1)$ of the uncensored survival times. Since these estimators are sums of i.i.d. terms, they are easier to handle than the estimators $\hat{F}_{0,\beta}(t)$, $\hat{f}_{0,\beta}(t)$ and $\hat{f}'_{0,\beta}(t)$.

A.4.1. Proof of Theorem 3.3 (i)

First, note that by Duhamel's identity (see [9]),

$$\hat{S}_{0,\beta}(t) - S_{0,\beta}(t) = -S_{0,\beta}(t)\int_0^t \frac{\hat{S}_{0,\beta}(t-)}{S_{0,\beta}(t)}\big(\hat{\Lambda}_{0,\beta}(ds) - \Lambda_{0,\beta}(ds)\big), \tag{14}$$

where

$$\hat{\Lambda}_{0,\beta}(t) = \int_0^t \frac{\hat{H}^1_{0,\beta}(ds)}{1-\hat{H}_{0,\beta}(s-)}$$

estimates the cumulative hazard given by

$$\Lambda_{0,\beta}(t) = \exp(-S_{0,\beta}(t)) = \int_0^t \frac{H^1_{0,\beta}(ds)}{1-H_{0,\beta}(s)}.$$

It can be easily seen that

$$\begin{aligned} &\hat{\Lambda}_{0,\beta}(t) - \Lambda_{0,\beta}(t) \qquad (15)\\ &= \int_0^t \Big[\frac{1}{1-\hat{H}_{0,\beta}(s-)} - \frac{1}{1-H_{0,\beta}(s)}\Big] d\hat{H}^1_{0,\beta}(s) + \int_0^t \frac{d(\hat{H}^1_{0,\beta}(s) - H^1_{0,\beta}(s))}{1-H_{0,\beta}(s)}. \end{aligned}$$

Hence, it follows from assumptions (C4)-(C5) that the stated result follows if we can show that $\sup_{\beta\in\mathcal{B}}\sup_{0\le t<\tau(\beta)}|\hat{H}_{0,\beta}(t)-H_{0,\beta}(t)| = O_P(n^{-1/2})$, and similarly with $H_{0,\beta}(t)$ replaced by $H^1_{0,\beta}(t)$.

Next, consider the class

$$\mathcal{F} = \Big\{(x,y) \to I(ye^{-\beta^t x} \le t) : \beta\in\mathcal{B}, 0 \le t \le \tau_{\max}\Big\}.$$

We suppose for notational simplicity that X is one-dimensional ($\ell = 1$). Divide $\mathcal{B}$ into small intervals $[b_{j-1}, b_j]$, $j = 1,\ldots,M$, with $M = O(\varepsilon^{-2})$ and $b_j = b_{j-1} + \varepsilon^2$, and similarly divide $[0,\tau_{\max}]$ into intervals $[t_{k-1}, t_k]$, $k = 1,\ldots,L$, with $L = O(\varepsilon^{-2})$ and $t_k = t_{k-1} + \varepsilon^2$. Then, for any $\beta\in\mathcal{B}$ and $t\in[0,\tau_{\max}]$ there exist a j and k such that $t_{k-1} < t \le t_k$ and $b_{j-1} < \beta \le b_j$. Hence,

$$I(ye^{-b^t_{j-1}x} \le t_{k-1}) < I(ye^{-\beta^t x} \le t) \le I(ye^{-b^t_j x} \le t_k)$$

(we suppose for simplicity that x is positive). Moreover,

$$\begin{aligned}
&E\Big[I(Ye^{-b_j^t X} \le t_k) - I(Ye^{-b_{j-1}^t X} \le t_{k-1})\Big]^2 \\
&= P\Big(Ye^{-b_j^t X} \le t_k\Big) - P\Big(Ye^{-b_{j-1}^t X} \le t_{k-1}\Big) \\
&= \int \Big[F_{Y|X}(t_k e^{b_j^t x}|x) - F_{Y|X}(t_{k-1}e^{b_{j-1}^t x}|x)\Big] f_X(x)\,dx \\
&\le \sup_{x,y} f_{Y|X}(y|x) \int \big[t_k e^{b_j^t x} - t_{k-1}e^{b_{j-1}^t x}\big] f_X(x)\,dx \le K\varepsilon^2,
\end{aligned}$$

where the last inequality follows from assumption (C6). Hence, $N_{[\,]}(\varepsilon, \mathcal{F}, L_2(P)) = O(\varepsilon^{-4})$, and

$$\int_0^1 \sqrt{\log N_{[\,]}(\varepsilon, \mathcal{F}, L_2(P))}d\varepsilon < \infty,$$

where $N_{[\,]}(\varepsilon, \mathcal{F}, L_2(P))$ is the ε-bracketing number of the class $\mathcal{F}$ with respect to the L_2-distance. This shows that the class $\mathcal{F}$ is Donsker (see p. 80-83 in [28] for the definition of a Donsker class and the bracketing number). It now follows from Theorem 2.5.6 in [28] that $\sup_{\beta\in\mathcal{B}} \sup_{0\le t<\tau(\beta)} |\hat{H}_{0,\beta}(t) - H_{0,\beta}(t)| = O_P(n^{-1/2})$, which shows the result. □

A.4.2. Proof of Theorem 3.3 (ii)

Write

$$\begin{aligned}
&\hat{f}_{0,\beta}(t) - f_{0,\beta}(t) \\
&= b^{-1}\int K\Big(\frac{t-s}{b}\Big) d(\hat{F}_{0,\beta}(s) - F_{0,\beta}(s)) + b^{-1}\int K\Big(\frac{t-s}{b}\Big) dF_{0,\beta}(s) - f_{0,\beta}(t) \\
&= T_1(t,\beta) + T_2(t,\beta).
\end{aligned}$$

We start with the bias term $T_2(t,\beta)$:

$$\begin{aligned}
&T_2(t,\beta) \\
&= b^{-1}\int K\Big(\frac{t-s}{b}\Big)\big[f_{0,\beta}(s) - f_{0,\beta}(t)\big]ds \\
&= \int K(u)\big[f_{0,\beta}(t-ub) - f_{0,\beta}(t)\big]du \\
&= \int K(u)\Big[-f'_{0,\beta}(t)ub + \frac{1}{2}f''_{0,\beta}(t)u^2b^2 - \frac{1}{6}f^{(3)}_{0,\beta}(t)u^3b^3 + \frac{1}{24}f^{(4)}_{0,\beta}(\xi)u^4b^4\Big]du \\
&= O(b_n^4),
\end{aligned}$$

uniformly in t and β, for some ξ between t and $t-ub$, since the order of K is larger than 3 (see assumption (C2)). Next, for the term $T_1(t,\beta)$, note that using

(14) and (15) we can decompose $T_1(t,\beta)$ into two terms. We will concentrate on the second one, since the first one is easier to handle:

$$
\begin{aligned}
& b^{-1} \int K\Big(\frac{t-s}{b}\Big) \frac{1-\hat{F}_{0,\beta}(s-)}{1-F_{0,\beta}(s)} \frac{f_{0,\beta}(s)}{1-H_{0,\beta}(s)} d(\hat{H}_{0,\beta}(s) - H_{0,\beta}(s)) \\
&= b^{-1} \int K(v) \frac{1-\hat{F}_{0,\beta}((t-vb)-)}{1-F_{0,\beta}(t-vb)} \frac{f_{0,\beta}(t-vb)}{1-H_{0,\beta}(t-vb)} d\big[(\hat{H}_{0,\beta} - H_{0,\beta})(t-vb)\big] \\
&= b^{-1} \int \big[(\hat{H}_{0,\beta} - H_{0,\beta})(t-vb) - (\hat{H}_{0,\beta} - H_{0,\beta})(t)\big] \\
& \qquad\qquad d\Big[K(v) \frac{1-\hat{F}_{0,\beta}((t-vb)-)}{1-F_{0,\beta}(t-vb)} \frac{f_{0,\beta}(t-vb)}{1-H_{0,\beta}(t-vb)}\Big],
\end{aligned}
$$

where the last equality holds since $K(\pm 1) = 0$. It follows that

$$
\sup_{t,\beta} |T_1(t,\beta)| \le K b^{-1} \sup_{t,\beta,v} \big|(\hat{H}_{0,\beta} - H_{0,\beta})(t-vb) - (\hat{H}_{0,\beta} - H_{0,\beta})(t)\big|.
$$

Let

$$
\begin{aligned}
\mathcal{F} = \Big\{ & (x,y) \to I(ye^{-\beta^t x} \le t - vb) - I(ye^{-\beta^t x} \le t) : \beta \in \mathcal{B}, 0 \le t \le \tau_{\max}, \\
& -1 \le v \le 1, 0 \le b \le 1 \Big\}.
\end{aligned}
$$

For any $f \in \mathcal{F}$, let $G_n(f) = n^{-1/2} \sum_{i=1}^n (f(X_i,Y_i) - Ef(X,Y)) = n^{1/2}\big[(\hat{H}_{0,\beta} - H_{0,\beta})(t-vb) - (\hat{H}_{0,\beta} - H_{0,\beta})(t)\big]$. It follows from Theorem 2.14.2 in [28] that

$$
\begin{aligned}
& E\Big(\sup_{f\in\mathcal{F}} |G_n(f)|\Big) \\
&= n^{1/2} E\Big(\sup_{t,\beta,v,b} \big|(\hat{H}_{0,\beta} - H_{0,\beta})(t-vb) - (\hat{H}_{0,\beta} - H_{0,\beta})(t)\big|\Big) \\
&\le J_{[\,]}(\delta, \mathcal{F}, L_2(P)) \|F\|_{P,2} + n^{1/2} E\big[F(X,Y) I\big(F(X,Y) > n^{1/2} a(\delta)\big)\big],
\end{aligned}
$$

provided $\|f\|_{P,2} \le \delta \|F\|_{P,2}$, where

$$
J_{[\,]}(\delta, \mathcal{F}, L_2(P)) = \int_0^\delta \sqrt{1 + \log N_{[\,]}(\varepsilon \|F\|_{P,2}, \mathcal{F}, L_2(P))} d\varepsilon,
$$

F is an envelope for the class $\mathcal{F}$, $\|F\|_{P,2}^2 = E[F^2(X,Y)]$, and

$$
a(\delta) = \frac{\delta \|F\|_{P,2}}{\sqrt{1 + \log N_{[\,]}(\delta \|F\|_{P,2}, \mathcal{F}, L_2(P))}}.
$$

Note that $F \equiv 1$ and hence $\|F\|_{P,2} = 1$. It follows from the proof of part (i) that $N_{[\,]}(\varepsilon, \mathcal{F}, L_2(P)) \le K\varepsilon^{-2(\ell+1)}$, where ℓ is the dimension of X. Moreover, for any $f \in \mathcal{F}$,

$$
\begin{aligned}
& \|f\|_{P,2}^2 \\
&= \int f^2(X,Y) dP
\end{aligned}
$$

$$
\begin{aligned}
&= E\Big[\Big\{I(Ye^{-\beta^t X} \le t - vb) - I(Ye^{-\beta^t X} \le t)\Big\}^2\Big] \\
&= P(Ye^{-\beta^t X} \le t - vb) + P(Ye^{-\beta^t X} \le t) - 2P(Ye^{-\beta^t X} \le t + \min(-vb, 0)) \\
&= \big|P(Ye^{-\beta^t X} \le t - vb) - P(Ye^{-\beta^t X} \le t)\big| \le Kb,
\end{aligned}
$$

since $\sup_{t,\beta} f_{Ye^{-\beta^t X}}(t) < \infty$. Hence, $\delta \propto b^{1/2}$, and for small δ,

$$
a(\delta) \ge \frac{\delta}{\sqrt{1 + \log(K\delta^{-2(m+1)})}} \ge \frac{\delta}{\sqrt{1+\delta^{-2}}} = \frac{\delta}{\sqrt{2\delta^{-2}}} = \frac{\delta^2}{\sqrt{2}} \propto b.
$$

It follows that $I(F(X,Y) > n^{1/2}a(\delta)) \le I(1 > (nb^2)^{1/2}) = 0$ for n large, since $nb_n^2 \to \infty$. Next, $J_{[\,]}(\delta, \mathcal{F}, L_2(P)) \le K\int_0^\delta \sqrt{\log(\varepsilon^{-1})}d\varepsilon$ and this is easily seen to be bounded by $K'\delta\sqrt{\log(\delta^{-1})}$ for some $K, K' < \infty$. It now follows that $E\big(\sup_{f\in\mathcal{F}} |G_n(f)|\big) = O(b_n^{1/2}(\log n)^{1/2})$, and hence $\sup_{t,\beta}|T_1(t,\beta)| = O_P((nb_n)^{-1/2}(\log n)^{1/2})$ thanks to Markov's inequality. □

A.4.3. Proof of Theorem 3.3 (iii)

Write

$$
\begin{aligned}
&\hat f'_{0,\beta}(t) - f'_{0,\beta}(t) \\
&= b^{-2}\int K'\Big(\frac{t-s}{b}\Big)d(\hat F_{0,\beta}(s) - F_{0,\beta}(s)) + b^{-2}\int K'\Big(\frac{t-s}{b}\Big)dF_{0,\beta}(s) - f'_{0,\beta}(t) \\
&= T_1(t,\beta) + T_2(t,\beta).
\end{aligned}
$$

We start again with the bias term $T_2(t,\beta)$:

$$
\begin{aligned}
&T_2(t,\beta) \\
&= b^{-1}\int K'(u) f_{0,\beta}(t-ub)du - f'_{0,\beta}(t) \\
&= \int K(u)\big[f'_{0,\beta}(t-ub) - f'_{0,\beta}(t)\big]du \\
&= \int K(u)\Big[-f''_{0,\beta}(t)ub + \frac{1}{2}f^{(3)}_{0,\beta}(t)u^2b^2 - \frac{1}{6}f^{(4)}_{0,\beta}(t)u^3b^3 + \frac{1}{24}f^{(5)}_{0,\beta}(\xi)u^4b^4\Big]du \\
&= O(b_n^4),
\end{aligned}
$$

uniformly in t and β, for some ξ between t and $t-ub$. For the term $T_1(t,\beta)$ we can follow a very similar development as in the proof of part (ii), provided $K'(\pm 1) = 0$. □

Appendix B: Further simulation results

Tables 11–13 show the simulations results when the covariate X follows a uniform distribution on $[0,1]$.

Table 11. *Bias, variance and mean squared error (MSE) of the model parameters when X follows a uniform distribution and the error has a logistic distribution.*

				Ours			Lu (2010)			Zhang-Peng (2007)			Scolas et al. (2016)		
γ_0	τ_C	n	Par.	Bias	Var	MSE	Bias	Var	MSE	Bias	Var	MSE	Bias	Var	MSE
0.5	20	200	γ_0	−.095	.126	.135	−.093	.131	.140	−.118	.120	.134	−.204	.095	.137
			γ_1	−.220	.358	.358	.199	.526	.566	.081	.460	.467	.046	.330	.332
			β_1	−.153	.518	.541	.200	.766	.806	.045	.506	.508	.159	.409	.434
		400	γ_0	−.091	.074	.082	−.059	.075	.078	−.109	.063	.075	−.213	.051	.096
			γ_1	.003	.209	.209	.202	.303	.344	.052	.216	.219	.048	.176	.178
			β_1	−.094	.433	.442	.199	.524	.564	−.004	.326	.326	.184	.278	.312
	100	200	γ_0	−.014	.093	.093	−.008	.093	.093	−.015	.091	.091	−.056	.077	.080
			γ_1	−.005	.265	.265	.040	.285	.287	.010	.269	.269	−.035	.241	.242
			β_1	−.028	.343	.344	.069	.397	.402	.025	.336	.337	−.136	.530	.548
		400	γ_0	−.016	.046	.046	.006	.050	.050	−.014	.045	.045	−.053	.040	.043
			γ_1	−.010	.119	.119	.032	.132	.133	−.003	.120	.120	−.057	.113	.116
			β_1	−.045	.223	.225	.031	.257	.258	−.016	.200	.200	−.267	.302	.373
1	20	200	γ_0	−.131	.154	.171	−.140	.149	.169	−.178	.133	.165	−.286	.096	.178
			γ_1	−.014	.398	.398	.296	.594	.682	.152	.514	.537	.098	.340	.350
			β_1	−.184	.398	.432	.194	.556	.594	.052	.404	.407	.149	.380	.402
		400	γ_0	−.112	.100	.113	−.082	.094	.101	−.146	.075	.096	−.284	.059	.140
			γ_1	−.009	.247	.247	.238	.366	.423	.072	.277	.282	.067	.206	.210
			β_1	−.138	.333	.352	.165	.360	.387	.013	.259	.259	.178	.215	.247
	100	200	γ_0	−.023	.098	.099	−.012	.096	.096	−.023	.097	.098	−.080	.078	.084
			γ_1	.022	.267	.267	.082	.287	.294	.040	.272	.274	−.019	.246	.246
			β_1	−.020	.288	.288	.069	.319	.324	.032	.271	.272	−.189	.470	.506
		400	γ_0	−.016	.058	.058	.008	.063	.063	−.017	.057	.057	−.074	.049	.054
			γ_1	−.013	.155	.155	.045	.172	.174	.000	.157	.157	−.061	.143	.147
			β_1	−.058	.178	.181	.018	.192	.192	−.025	.157	.158	−.303	.266	.358

Table 12. *Bias, variance and mean squared error (MSE) of the model parameters when X follows a uniform distribution and the error has a normal distribution.*

γ_0	τ_C	n	Par.	Ours Bias	Ours Var	Ours MSE	Lu (2010) Bias	Lu (2010) Var	Lu (2010) MSE	Zhang-Peng (2007) Bias	Zhang-Peng (2007) Var	Zhang-Peng (2007) MSE	Scolas et al. (2016) Bias	Scolas et al. (2016) Var	Scolas et al. (2016) MSE
0.5	20	200	γ_0	−.006	.110	.110	−.006	.113	.113	−.010	.110	.110	−.009	.105	.105
			γ_1	.021	.332	.332	.112	.378	.391	.051	.355	.358	.024	.331	.332
			β_1	−.015	.182	.182	.080	.197	.203	.021	.148	.148	.007	.083	.083
		400	γ_0	.023	.048	.049	.034	.051	.052	.021	.048	.048	.024	.045	.046
			γ_1	−.040	.138	.140	.040	.153	.155	−.024	.141	.142	−.038	.138	.139
			β_1	−.019	.083	.083	.062	.088	.092	−.001	.069	.069	−.014	.037	.037
	100	200	γ_0	−.017	.084	.084	−.014	.085	.085	−.016	.084	.084	−.018	.084	.084
			γ_1	.034	.244	.245	.038	.246	.247	.036	.244	.245	.034	.243	.244
			β_1	.008	.114	.114	.021	.120	.120	.009	.107	.107	.017	.061	.061
		400	γ_0	.026	.036	.037	.030	.037	.038	.026	.036	.037	.026	.036	.037
			γ_1	−.043	.114	.116	−.039	.115	.117	−.041	.114	.116	−.042	.115	.117
			β_1	−.016	.053	.053	−.003	.054	.054	−.007	.051	.051	−.013	.030	.030
1	20	200	γ_0	.021	.140	.140	.019	.143	.143	.013	.137	.137	.014	.130	.130
			γ_1	−.005	.406	.406	.109	.461	.473	.029	.422	.423	−.002	.385	.385
			β_1	−.001	.151	.151	.093	.148	.157	.030	.117	.118	.015	.061	.061
		400	γ_0	.023	.065	.066	.037	.068	.069	.019	.063	.063	.022	.059	.059
			γ_1	−.031	.188	.189	.072	.215	.220	−.010	.193	.193	−.022	.187	.187
			β_1	−.012	.064	.064	.058	.069	.072	.002	.057	.057	−.003	.034	.034
	100	200	γ_0	.006	.104	.104	.010	.104	.104	.007	.104	.104	.006	.104	.104
			γ_1	.004	.294	.294	.012	.297	.297	.007	.296	.296	.003	.295	.295
			β_1	.017	.091	.091	.032	.095	.096	.019	.086	.086	.016	.054	.054
		400	γ_0	.026	.046	.047	.029	.046	.047	.026	.046	.047	.024	.046	.047
			γ_1	−.034	.145	.146	−.026	.146	.147	−.031	.146	.147	−.032	.145	.146
			β_1	−.013	.043	.043	−.001	.044	.044	−.001	.042	.042	.001	.027	.027

TABLE 13. *Bias, variance and mean squared error (MSE) of the model parameters when X follows a uniform distribution and the error distribution is a mixture of Weibull distributions.*

				Ours			Lu (2010)			Zhang-Peng (2007)			Scolas et al. (2016)		
γ_0	τ_C	n	Par.	Bias	Var	MSE	Bias	Var	MSE	Bias	Var	MSE	Bias	Var	MSE
0.5	20	200	γ_0	.001	.106	.106	.002	.107	.107	.001	.107	.107	.231	.209	.262
			γ_1	−.007	.341	.341	.013	.351	.351	.006	.350	.350	.720	.978	1.50
			β_1	−.003	.011	.011	.007	.011	.011	−.001	.010	.010	.145	.159	.180
		400	γ_0	.016	.045	.045	.017	.046	.046	.015	.045	.045	.323	.106	.210
			γ_1	−.019	.144	.144	.005	.148	.148	−.007	.148	.148	.540	.428	.720
			β_1	.002	.004	.004	.007	.005	.005	.003	.005	.005	.038	.109	.110
	100	200	γ_0	−.021	.083	.083	−.020	.083	.083	−.021	.083	.083	−.016	.085	.085
			γ_1	.039	.246	.248	.040	.246	.248	.041	.247	.249	.097	.259	.268
			β_1	−.006	.009	.009	−.003	.009	.009	−.002	.009	.009	.418	.216	.391
		400	γ_0	.023	.036	.037	.025	.036	.037	.023	.036	.037	.017	.039	.039
			γ_1	−.036	.117	.118	−.036	.117	.118	−.035	.117	.118	.037	.125	.126
			β_1	.001	.004	.004	.002	.004	.004	.002	.004	.004	.561	.118	.433
1	20	200	γ_0	.032	.138	.139	.033	.139	.140	.032	.140	.141	.358	.440	.568
			γ_1	−.029	.409	.410	−.010	.428	.428	−.016	.429	.429	1.15	2.58	3.90
			β_1	.003	.008	.008	.008	.008	.008	−.001	.008	.008	.116	.140	.153
		400	γ_0	.012	.054	.054	.016	.054	.054	.012	.055	.055	.418	.197	.372
			γ_1	.006	.167	.167	.028	.174	.175	.018	.175	.175	.944	1.02	1.91
			β_1	.004	.004	.004	.008	.004	.004	.004	.004	.004	.022	.098	.098
	100	200	γ_0	.000	.102	.102	.001	.101	.101	.000	.102	.102	.005	.109	.109
			γ_1	.013	.292	.292	.015	.291	.291	.015	.293	.293	.090	.322	.330
			β_1	−.005	.007	.007	−.003	.007	.007	−.002	.007	.007	.464	.191	.406
		400	γ_0	.024	.045	.046	.025	.045	.046	.024	.045	.046	.019	.048	.048
			γ_1	−.029	.143	.144	−.026	.143	.144	−.027	.144	.145	.057	.155	.158
			β_1	.002	.003	.003	.003	.003	.003	.003	.004	.004	.585	.092	.434

We end this Appendix with Q-Q plots for the estimated parameters $\hat{\beta}_1$, $\hat{\gamma}_0$ and $\hat{\gamma}_1$ for the four methods, and for one setting, namely when $n = 400$, $\tau_C = 20$, $\gamma_0 = 0.5$, X follows a binomial distribution, and the error distribution is a mixture of two Weibull distributions.

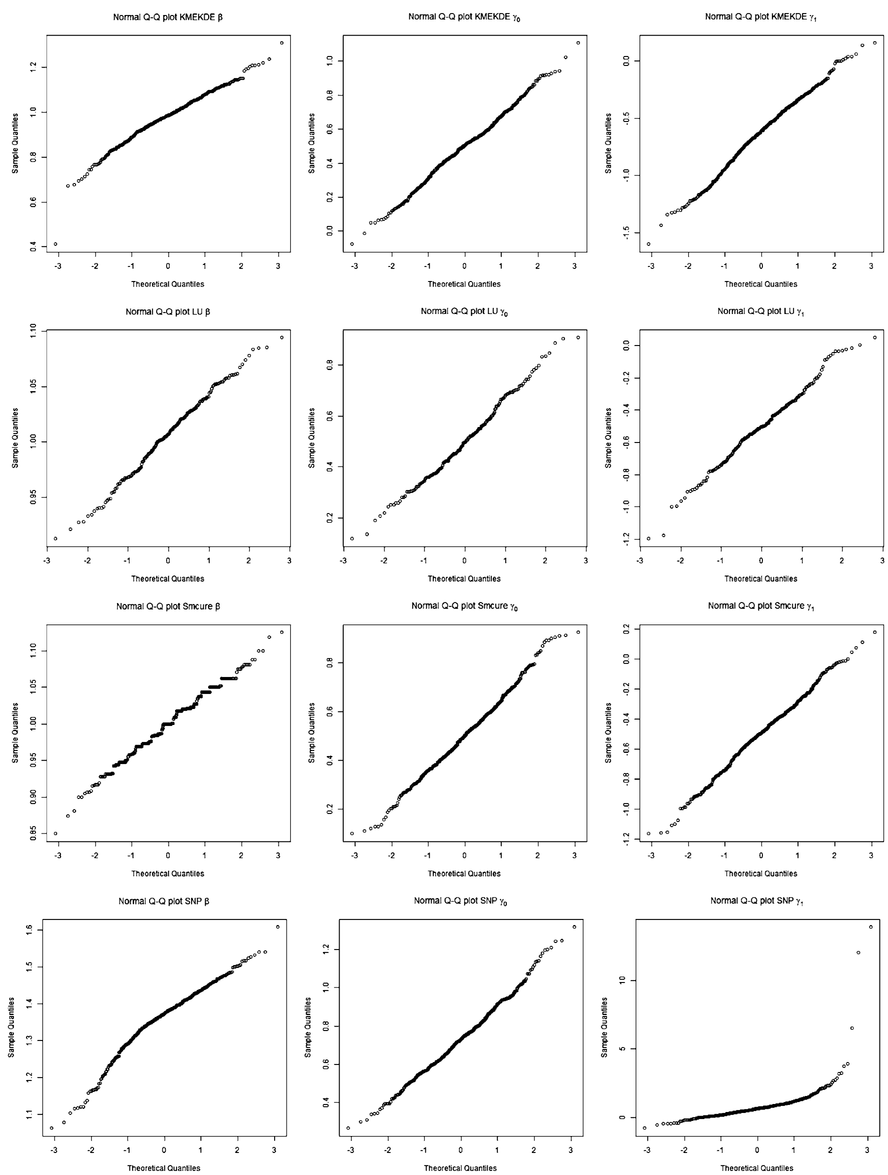

FIG 4. *Q-Q plots for the estimated parameters $\hat{\beta}_1$ (first column), $\hat{\gamma}_0$ (second column) and $\hat{\gamma}_1$ (third column) for the four methods: Our method (first row), Lu (2010)'s method (second row), Zhang-Peng (2007)'s method (third row), and Scolas et al (2016)'s method (fourth row).*

Acknowledgments

The authors are very grateful to Wenbin Lu and Sylvie Scolas for providing us the R code of their methods (given in [17] and [26] respectively), which we used in Sections 4 and 5 to compare our method with theirs.

Funding

Ingrid Van Keilegom acknowledges support from the FWO and F.R.S.-FNRS under the Excellence of Science (EOS) programme, project ASTeRISK (grant No. 40007517), and from the FWO (senior research projects fundamental research, grant no. G047524N).

References

[1] Amico, M. and Van Keilegom, I. (2018). Cure models in survival analysis. *Annual Review of Statistics and Its Application* **5** 311–342. MR3774750
[2] Berkson, J. and Gage, R. P. (1952). Survival curve for cancer patients following treatment. *Journal of the American Statistical Association* **47** 501–515.
[3] Boag, J. W. (1949). Maximum likelihood estimates of the proportion of patients cured by cancer therapy. *Journal of the Royal Statistical Society–Series B* **11** 15–44.
[4] Cai, C., Zou, Y., Peng, Y. and Zhang, J. (2012). smcure: An R-package for estimating semiparametric mixture cure models. *Computer Methods and Programs in Biomedicine* **108** 1255–1260.
[5] Chen, X., Linton, O. and Van Keilegom, I. (2003). Estimation of semiparametric models when the criterion function is not smooth. *Econometrica* **71** 1591–1608. MR2000259
[6] Cox, D. R. and Oakes, D. (1984). *Analysis of Survival Data*. Chapman and Hall, New York. MR0751780
[7] Fang, H.-B., Li, G. and Sun, J. (2005). Maximum likelihood estimation in a semiparametric logistic/proportional-hazards mixture model. *Scandinavian Journal of Statistics* **32** 59–75. MR2136802
[8] Farewell, V. T. (1982). The use of mixture models for the analysis of survival data with long-term survivors. *Biometrics* **38** 1041–1046.
[9] Gill, R. D. and Johansen, S. (1990). A survey of product-integration with a view toward application in survival analysis. *Annals of Statistics* **18** 1501–1555. MR1074422
[10] Kalbfleisch, J. D. and Prentice, R. L. (2011). *The Statistical Analysis of Failure Time Data*. John Wiley & Sons, New York. MR0570114
[11] Kuk, A. Y. and Chen, C.-H. (1992). A mixture model combining logistic regression with proportional hazards regression. *Biometrika* **79** 531–541. MR2400249

[12] Legrand, C. and Bertrand, A. (2019). Cure models in oncology clinical trials. *Textb Clin Trials Oncol Stat Perspect* **1** 465–492.

[13] Lo, S.-H. and Singh, K. (1986). The product-limit estimator and the bootstrap: some asymptotic representations. *Probability Theory and Related Fields* **71** 455–465. MR0824714

[14] López-Cheda, A., Cao, R., Jácome, M. A. and Van Keilegom, I. (2017). Nonparametric incidence estimation and bootstrap bandwidth selection in mixture cure models. *Computational Statistics & Data Analysis* **105** 144–165. MR3552194

[15] López-Cheda, A., Jácome, M. A. and Cao, R. (2017). Nonparametric latency estimation for mixture cure models. *TEST* **26** 353–376. MR3650531

[16] Lu, W. (2008). Maximum likelihood estimation in the proportional hazards cure model. *Annals of the Institute of Statistical Mathematics* **60** 545–574. MR2434411

[17] Lu, W. (2010). Efficient estimation for an accelerated failure time model with a cure fraction. *Statistica Sinica* **20** 661–674. MR2682635

[18] Major, P. and Rejto, L. (1988). Strong embedding of the estimator of the distribution function under random censorship. *Annals of Statistics* **16** 1113 – 1132. MR0959190

[19] Maller, R. A. and Zhou, X. (1996). *Survival Analysis with Long-Term Survivors*. Wiley, New York. MR1453117

[20] Parsa, M., Taghavi-Shahri, S. M. and Van Keilegom, I. (2024). On variable selection in a semiparametric AFT mixture cure model. *Lifetime Data Analysis* **30** 472–500. MR4721146

[21] Parsa, M. and Van Keilegom, I. (2023). Accelerated failure time vs Cox proportional hazards mixture cure models: David vs Goliath? *Statistical Papers* **64** 835–855. MR4588213

[22] Patilea, V. and Van Keilegom, I. (2020). A general approach for cure models in survival analysis. *Annals of Statistics* **48** 2323–2346. MR4134797

[23] Peng, Y. and Dear, K. B. (2000). A nonparametric mixture model for cure rate estimation. *Biometrics* **56** 237–243.

[24] Peng, Y. and Taylor, J. M. G. (2014). Cure models. *In: Handbook of Survival Analysis, Handbooks of Modern Statistical Methods series, ed. J Klein, H van Houwelingen, JG Ibrahim, TH Scheike, chapter 6, pages 113-134. Boca Raton, FL: Chapman & Hall.* MR3287588

[25] Peng, Y. and Yu, B. (2021). *Cure Models: Methods, Applications, and Implementation (1st ed.).* CRC Press. MR4596124

[26] Scolas, S., El Ghouch, A. and Legrand, C. (2016). The SNP representation in mixture cure models with interval-censoring: estimation and goodness-of-fit testing Technical Report, Université catholique de Louvain, Institute of Statistics, Biostatistics and Actuarial Sciences.

[27] Sy, J. P. and Taylor, J. M. (2000). Estimation in a Cox proportional hazards cure model. *Biometrics* **56** 227–236. MR1767631

[28] Van der Vaart, A. and Wellner, J. (1996). *Weak Convergence and Empirical Processes: with Applications to Statistics*. Springer. MR1385671

[29] Wang, Y., Klijn, J. G., Zhang, Y., Sieuwerts, A. M.,

Look, M. P., Yang, F., Talantov, D., Timmermans, M., Meijer-van Gelder, M. E., Yu, J. et al. (2005). Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. *The Lancet* **365** 671–679.

[30] Zeng, D. and Lin, D. (2007). Efficient estimation for the accelerated failure time model. *Journal of the American Statistical Association* **102** 1387–1396. MR2412556

[31] Zhang, J. and Peng, Y. (2007). A new estimation method for the semiparametric accelerated failure time mixture cure model. *Statistics in Medicine* **26** 3157–3171. MR2380509