A general approach for cure models in survival analysis

Patilea V, Van Keilegom I.

A GENERAL APPROACH FOR CURE MODELS IN SURVIVAL ANALYSIS

By Valentin Patilea and Ingrid Van Keilegom

CREST, Ensai, France
ORSTAT, KU Leuven, Belgium

In survival analysis it often happens that some subjects under study do not experience the event of interest; they are considered to be 'cured'. The population is thus a mixture of two subpopulations : the one of cured subjects, and the one of 'susceptible' subjects. In this paper we propose a novel approach to estimate a mixture cure model when covariates are present and the lifetime is subject to random right censoring. We work with a parametric model for the cure proportion (like e.g. a logistic model), while the conditional survival function of the uncured subjects is unspecified. The approach is based on an inversion which allows to write the survival function as a function of the distribution of the observable random variables. This leads to a very general class of models, which allows a flexible and rich modeling of the conditional survival function. We show the identifiability of the proposed model, as well as the weak consistency and the asymptotic normality of the model parameters. We also consider in more detail the case where kernel estimators are used for the nonparametric part of the model. The new estimators are compared with the estimators from a Cox mixture cure model via finite sample simulations. Finally, we apply the new model and estimation procedure on two medical data sets.

1. Introduction. Driven by emerging applications, over the last two decades there has been an increasing interest for time-to-event analysis models allowing the situation where a fraction of the right censored observed lifetimes corresponds to subjects who will never experience the event. In biostatistics such models including covariates are usually called *cure models* and they allow for a positive *cure fraction*, or *cure rate*, that corresponds to the proportion of patients cured of their disease. For a review of these models in survival analysis, see for instance Peng & Taylor (2014) or Amico

& Van Keilegom (2018). Economists sometimes call such models *split population models* (see Schmidt & Witte 1989), while the reliability engineers refer to them as *limited-failure population life models* (Meeker 1987).

At first sight, a cure regression model is nothing but a binary outcome, cured versus uncured, regression problem. The difficulty comes from the fact that the cured subjects are unlabeled observations among the censored data. Then one has to use all the observations, censored and uncensored, to complete the missing information and thus to identify, estimate and make inference on the cure fraction regression function. We propose a general approach for this task, a tool that provides a general ground for cure regression models. The idea is to start from the laws of the observed variables and to express the quantities of interest, such as the cure rate and the conditional survival of the uncured subjects, as functionals of these laws. These general expressions, that we call inversion formulae and that we derive with no particular constraint on the space of the covariates, are the vehicles that allow for a wide modeling choice, parametric, semiparametric and nonparametric, for both the law of the lifetime of interest and the cure rate. Indeed, the inversion formulae allow to express the likelihood of the binary outcome model as a function of the laws of the observed variables. The maximum likelihood estimator of the parameter vector of the cure fraction function is then simply the maximizer of the likelihood obtained by replacing the laws of the observations by some estimators. With at hand the estimate of the parameter of the cure fraction, the inversion formulae will provide an estimate for the conditional survival of the uncured subjects. For the sake of clarity, we focus on the so-called mixture cure models with a parametric cure fraction function, the type of model that is most popular among practitioners. Meanwhile, the law of the lifetime of interest is left unspecified.

The paper is organized as follows. In Section 2 we provide a general description of the mixture cure model and next we introduce the needed notation and present the inversion formulae on which our approach is built. We finish Section 2 by a discussion of the identification issue and some new insight in the existing approaches in the literature on cure models. Section 3 introduces the general maximum likelihood estimator, while in Section 4 we derive the general asymptotic results. A simple bootstrap procedure for making feasible inference is proposed. Section 4 ends with an illustration of the general approach in the case where the conditional law of the observations is estimated by kernel smoothing. In Section 5 we calculate the efficient score, which is needed for obtaining an efficient estimator of the parameters in the model. In Sections 6 and 7 we report some empirical results obtained with simulated and two real data sets. Our estimator performs well in sim-

ulations and provides similar or more interpretable results in applications compared with a competing logistic/proportional hazards mixture approach. The proofs of the main results are relegated to the Appendix, while some technical details are collected in an online Supplement.

2. The model.

2.1. *A general class of mixture cure models.* Let T denote (a possible monotone transformation of) the lifetime of interest that takes values in $(-\infty, \infty]$. A cured observation corresponds to the event $\{T = \infty\}$, and in the following this event is allowed to have a positive probability. Let X be a covariate vector belonging to $\mathcal{X}$ a general covariate space. The covariate vector could include discrete and continuous components. The survival function $F_T((t, \infty] \mid x) = \mathbb{P}(T > t \mid X = x)$, can be written as

$$(2.1) \quad F_T((t, \infty] \mid x) = 1 - \phi(x) + \phi(x) F_{T,0}((t, \infty) \mid x), \quad t \in \mathbb{R}, \ x \in \mathcal{X},$$

where $1 - \phi(x) = \mathbb{P}(T = \infty \mid X = x)$ is the cure rate and $F_{T,0}((t, \infty) \mid x) = \mathbb{P}(T > t \mid X = x, T < \infty)$. Depending on which model is used for $\phi(x)$ and $F_{T,0}(\cdot \mid x)$, one obtains a parametric, semiparametric or nonparametric model, called a 'mixture cure model'. In the literature, one often assumes that $\phi(x)$ follows a logistic model, i.e. $\phi(x) = \exp(a + x^\top b)/[1 + \exp(a + x^\top b)]$ for some $(a, b^\top)^\top \in \mathbb{R}^{1+d}$. (Here and in the following, a vector is a column matrix and for any matrix A, $A^\top$ denotes its transpose.) Recently, semiparametric models (like a single-index model as in Amico *et al.* 2018) or nonparametric models (as in Xu & Peng 2014 or López-Cheda *et al.* 2017) have been proposed. As for the survival function $F_{T,0}(\cdot \mid x)$ of the susceptible subjects, a variety of models have been proposed, including parametric models (see e.g. Boag 1949, Farewell 1982), semiparametric models based on a proportional hazards assumption (see e.g. Kuk & Chen 1992, Sy & Taylor 2000, Fang *et al.* 2005, Lu 2008; see also Othus *et al.* 2009) or nonparametric models (see e.g. Taylor 1995, Xu & Peng 2014).

In this paper we propose to model $\phi(x)$ parametrically, i.e. we assume that $\phi(\cdot)$ belongs to the family of conditional probability functions $\{\phi(\cdot, \beta) : \beta \in B\}$, where $\phi(\cdot, \beta)$ takes values in the interval $(0, 1)$, β is the parameter vector of the model and B is the parameter set. This family could be the logistic family or any other parametric family. For the survival function $F_{T,0}(\cdot \mid x)$ we do not impose any assumptions in order to have a flexible and rich class of models for $F_T(\cdot \mid x)$ to choose from. Later on we will see that for the estimation of $F_{T,0}(\cdot \mid x)$ any estimator that satisfies certain minimal conditions can be used, and hence we allow for a large variety of parametric, semiparametric and nonparametric estimation methods.

As is often the case with time-to-event data, we assume that the lifetime T is subject to random right censoring, i.e. instead of observing T, we only observe the pair (Y, δ), where $Y = T \wedge C$, $\delta = \mathbf{1}\{T \leq C\}$ and C is a real-valued random variable, called the censoring time. Some identification assumptions are required to be able to identify the conditional law of T from the observed variables Y and δ. Let us assume that

$$(2.2) \qquad C \perp T \mid X \qquad \text{and} \qquad \mathbb{P}(C < \infty) = 1.$$

The conditional independence between T and C is an usual identification assumption in survival analysis in the presence of covariates. The zero probability at infinity condition for C implies that $\mathbb{P}(C < \infty \mid X) = 1$ almost surely (a.s.). This latter mild condition is required if we admit that the observations Y are finite, which is the case in the common applications. For the sake of simplicity, let us also consider the commonly used condition

$$(2.3) \qquad \mathbb{P}(T = C) = 0,$$

which implies that $\mathbb{P}(T = C \mid X) = 0$ a.s.

2.2. *Some notations and preliminaries.* We start with some preliminary arguments, which are valid in general without assuming any model on the functions ϕ, F_T and F_C.

The observations are characterized by the conditional sub-probabilities

$$
\begin{aligned}
H_1((-\infty, t] \mid x) &= \mathbb{P}(Y \leq t, \delta = 1 \mid X = x) \\
H_0((-\infty, t] \mid x) &= \mathbb{P}(Y \leq t, \delta = 0 \mid X = x), \qquad t \in \mathbb{R}, \ x \in \mathcal{X}.
\end{aligned}
$$

Then $H((-\infty, t] \mid x) \stackrel{def}{=} \mathbb{P}(Y \leq t \mid X = x) = H_0((-\infty, t] \mid x) + H_1((-\infty, t] \mid x)$. Since we assume that Y is finite, we have

$$(2.4) \qquad H((-\infty, \infty) \mid x) = 1, \qquad \forall x \in \mathcal{X}.$$

For $j \in \{0, 1\}$ and $x \in \mathcal{X}$, let $\tau_{H_j}(x) = \inf\{t : H_j([t, \infty) \mid x) = 0\}$ denote the right endpoint of the support of the conditional sub-probability H_j. Let us define $\tau_H(x)$ in a similar way and note that $\tau_H(x) = \max\{\tau_{H_0}(x), \tau_{H_1}(x)\}$. Note that $\tau_{H_0}(x)$, $\tau_{H_1}(x)$ and $\tau_H(x)$ can equal infinity, even though Y is finite. For $x \in \mathcal{X}$ and $-\infty < t \leq \infty$, we define the conditional probabilities

$$F_C((-\infty, t] \mid x) = \mathbb{P}(C \leq t \mid X = x) \quad \text{and} \quad F_T((-\infty, t] \mid x) = \mathbb{P}(T \leq t \mid X = x).$$

Let us show how the probability of being cured could be identified from the observations without any reference to a model for this probability. Under conditions (2.2)-(2.3) we can write

$$H_1(dt \mid x) = F_C([t, \infty) \mid x) F_T(dt \mid x), \quad H_0(dt \mid x) = F_T([t, \infty] \mid x) F_C(dt \mid x),$$

and $H([t,\infty) \mid x) = F_T([t,\infty] \mid x)F_C([t,\infty) \mid x)$. These equations could be solved and thus they allow to express the functions $F_T(\cdot \mid x)$ and $F_C(\cdot \mid x)$ in an unique way as explicit transformations of $H_0(\cdot \mid x)$ and $H_1(\cdot \mid x)$. For this purpose, let us consider the conditional cumulative hazard measures

$$\Lambda_T(dt \mid x) = \frac{F_T(dt \mid x)}{F_T([t,\infty] \mid x)} \quad \text{and} \quad \Lambda_C(dt \mid x) = \frac{F_C(dt \mid x)}{F_C([t,\infty) \mid x)}, \quad x \in \mathcal{X}.$$

The model equations yield

$$(2.5) \quad \Lambda_T(dt \mid x) = \frac{H_1(dt \mid x)}{H([t,\infty) \mid x)} \quad \text{and} \quad \Lambda_C(dt \mid x) = \frac{H_0(dt \mid x)}{H([t,\infty) \mid x)}.$$

Then, we can write the following functionals of $H_0(\cdot \mid x)$ and $H_1(\cdot \mid x)$:

$$F_T((t,\infty] \mid x) = \prod_{s \leq t}\{1 - \Lambda_T(ds \mid x)\},$$

$$(2.6) \qquad F_C((t,\infty) \mid x) = \prod_{s \leq t}\{1 - \Lambda_C(ds \mid x)\}, \qquad t \in \mathbb{R},$$

where $\prod_{s \in A}$ stands for the product-integral over the set A (see Gill and Johansen 1990). Let us point out that

$$(2.7) \qquad \mathbb{P}(T = \infty \mid x) = \prod_{t \in \mathbb{R}}\{1 - \Lambda_T(dt \mid x)\} = \prod_{t \in \mathbb{R}}\left\{1 - \frac{H_1(dt \mid x)}{H([t,\infty) \mid x)}\right\}.$$

Moreover, if $\tau_{H_1}(x) < \infty$, then

$$\mathbb{P}(T > \tau_{H_1}(x) \mid x) = \prod_{t \in (-\infty, \tau_{H_1}(x)]}\{1 - \Lambda_T(dt \mid x)\},$$

but there is no way to identify the conditional law of T beyond $\tau_{H_1}(x)$. Therefore, we will impose

$$(2.8) \qquad \qquad \mathbb{P}(T > \tau_{H_1}(x) \mid x) = \mathbb{P}(T = \infty \mid x),$$

i.e. $\prod_{t \in \mathbb{R}}\{1 - \Lambda_T(dt \mid x)\} = \prod_{-\infty < t \leq \tau_{H_1}(x)}\{1 - \Lambda_T(dt \mid x)\}$. Note that if $\tau_{H_1}(x) = \infty$, condition (2.8) is no longer an identification restriction, but just a simple consequence of the definition of $\Lambda_T(\cdot \mid x)$. Finally, the condition that $\mathbb{P}(C < \infty) = 1$ in (2.2) can be re-expressed by saying that we assume that $H_0(\cdot \mid x)$ and $H_1(\cdot \mid x)$ are such that, $\forall x \in \mathcal{X}$,

$$(2.9) \;\; \mathbb{P}(C = \infty \mid x) = \prod_{t \in \mathbb{R}}\{1 - \Lambda_C(dt \mid x)\} = \prod_{t \in \mathbb{R}}\left\{1 - \frac{H_0(dt \mid x)}{H([t,\infty) \mid x)}\right\} = 0.$$

This condition is satisfied only if $\tau_{H_1}(x) \leq \tau_{H_0}(x)$. Indeed, if $\tau_{H_1}(x) > \tau_{H_0}(x)$ then necessarily $\tau_{H_0}(x) < \tau_H(x)$ and so $H([\tau_{H_0}(x), \infty) \mid x) > 0$. Hence, $\Lambda_C(\mathbb{R} \mid x) = \Lambda_C((-\infty, \tau_{H_0}(x)] \mid x) < \infty$, and thus $\mathbb{P}(C = \infty \mid x) > 0$, which contradicts (2.9).

It is important to understand that *any* two conditional sub-probabilities $H_0(\cdot \mid x)$ and $H_1(\cdot \mid x)$ such that $H_0(\mathbb{R} \mid x) + H_1(\mathbb{R} \mid x) = 1$, $\forall x \in \mathcal{X}$, define uniquely $F_T(\cdot \mid x)$ and $F_C(\cdot \mid x)$. If $H_0(\cdot \mid x)$ and $H_1(\cdot \mid x)$ are such that conditions (2.2) and (2.3) hold true, then $F_T(\cdot \mid x)$ is precisely the probability distribution of T given $X = x$, with all the mass beyond $\tau_{H_1}(x)$ concentrated at infinity provided condition (2.8) holds true. In general, $F_T(\cdot \mid x)$ and $F_C(\cdot \mid x)$ are only functionals of $H_0(\cdot \mid x)$ and $H_1(\cdot \mid x)$.

We will assume conditions (2.2), (2.3) and (2.8) throughout the paper.

2.3. *A key point for the new approach: the inversion formulae.* Write

$$
\begin{aligned}
H([t, \infty) \mid x) &= F_T([t, \infty] \mid x) F_C([t, \infty) \mid x) \\
&= F_T([t, \infty) \mid x) F_C([t, \infty) \mid x) + \mathbb{P}(T = \infty \mid x) F_C([t, \infty) \mid x),
\end{aligned}
$$

and thus

$$
(2.10) \qquad F_T([t, \infty) \mid x) = \frac{H([t, \infty) \mid x) - \mathbb{P}(T = \infty \mid x) F_C([t, \infty) \mid x)}{F_C([t, \infty) \mid x)}.
$$

Consider the conditional cumulative hazard measure for the finite values of the lifetime of interest:

$$
\Lambda_{T,0}(dt \mid x) \stackrel{def}{=} \frac{F_{T,0}(dt \mid x)}{F_{T,0}([t, \infty) \mid x)} = \frac{F_T(dt \mid x)}{F_T([t, \infty) \mid x)}
$$

for $t \in \mathbb{R}$. Since $H_1(dt \mid x) = F_C([t, \infty) \mid x) F_T(dt \mid x)$, using relationship (2.10) we obtain

$$
(2.11) \quad \Lambda_{T,0}(dt \mid x) = \frac{H_1(dt \mid x)}{H([t, \infty) \mid x) - \mathbb{P}(T = \infty \mid x) F_C([t, \infty) \mid x)}.
$$

Next, using the product-integral we can write

$$
(2.12) \qquad F_{T,0}((t, \infty) \mid x) = \prod_{s \leq t} \{1 - \Lambda_{T,0}(ds \mid x)\}, \quad t \in \mathbb{R}, \ x \in \mathcal{X}.
$$

Note that, by construction, $F_{T,0}(\mathbb{R} \mid x) = 1$.

Let us recall that $F_C(\cdot \mid x)$ can be written as a transformation of $H_0(\cdot \mid x)$ and $H_1(\cdot \mid x)$, see equations (2.5) and (2.6). This representation is not surprising since we can consider C as a lifetime of interest and hence T

plays the role of a censoring variable. Hence, estimating the conditional distribution function $F_C(\cdot \mid x)$ should not be more complicated than in a classical conditional Kaplan-Meier setup, since the fact that T could be equal to infinity with positive probability is irrelevant when estimating $F_C(\cdot \mid x)$.

Finally, the representations of $F_C(\cdot \mid x)$ and $\mathbb{P}(T = \infty \mid x)$ given in equations (2.6) and (3.3), respectively, plugged into equation (2.11), allows to express *in a unique way* $\Lambda_{T,0}(\cdot|x)$, and thus $F_{T,0}(\cdot|x)$, as maps of the measures $H_0(\cdot|x)$ and $H_1(\cdot|x)$. This will be the key element for providing more insight in existing approaches and the starting point of our new approach.

2.4. *Model identification issues.* Recall that our model involves the functions $F_{T,0}(\cdot \mid x)$, $F_C(\cdot \mid x)$ and $\phi(\cdot, \beta)$, and the assumptions (2.2), (2.3), (2.8).

Let $F_{Y,\delta}(\cdot, \cdot \mid x)$ denote the conditional law of (Y, δ) given $X = x$. Moreover, let $F_{Y,\delta}^{\beta}(dt, 1 \mid x) = \phi(x, \beta) F_C((t, \infty) \mid x) F_{T,0}(dt \mid x)$, and

$$F_{Y,\delta}^{\beta}(dt, 0 \mid x) = [F_{T,0}([t, \infty) \mid x)\phi(x, \beta) + 1 - \phi(x, \beta)] F_C(dt \mid x).$$

These equations define a conditional law for the observations (Y, δ) based on the model. More precisely, for a choice of $F_T(\cdot \mid x)$, $F_C(\cdot \mid x)$ and β, the model yields a conditional law for (Y, δ) given $X = x$. If the mixture cure model is correctly specified, there exists a value β_0 such that

$$(2.13) \qquad F_{Y,\delta}(\cdot, \cdot \mid x) = F_{Y,\delta}^{\beta_0}(\cdot, \cdot \mid x), \quad \forall x \in \mathcal{X}.$$

The next question is whether $F_{T,0}(\cdot \mid x)$, $F_C(\cdot \mid x)$ and β_0 are identifiable.

PROPOSITION 2.1. *Consider a cure mixture model as in equation (2.1). Assume that conditions (2.2), (2.3), (2.8) are met. Moreover, assume that the cure rate model $\{\phi(\cdot, \beta) : \beta \in B\}$ is correct and satisfies the condition*

$$(2.14) \quad \forall \beta, \widetilde{\beta} \in B \text{ such that } \mathbb{P}(\phi(X, \beta) = \phi(X, \widetilde{\beta})) = 1, \text{ we have } \beta = \widetilde{\beta}.$$

Then $F_{T,0}(\cdot \mid x)$, $F_C(\cdot \mid x)$ and β_0 are identifiable.

The proof is given in the Appendix.

2.5. *Interpreting the previous modeling approaches.* We suppose here that the function $\phi(x)$ follows a logistic model, and comment on several models for $F_{T,0}$ that have been considered in the literature.

2.5.1. *Parametric and proportional hazards mixture model.* In a parametric modeling, one usually supposes that $\tau_{H_0}(x) = \tau_{H_1}(x) = \infty$ and $\Lambda_{T,0}(\cdot \mid x)$ belongs to a parametric family of cumulative hazard functions, like for instance the Weibull model; see Farewell (1982). Several contributions proposed a more flexible semiparametric proportional hazards (PH) approach; see Fang *et al.* (2005), Lu (2008) and the references therein. In such a model one imposes a PH structure for the $\Lambda_{T,0}(\cdot \mid x)$ measure. More precisely, it is supposed that

$$\Lambda_{T,0}(dt \mid x) = \frac{H_1(dt \mid x)}{H([t,\infty) \mid x) - \mathbb{P}(T = \infty \mid x) F_C([t,\infty) \mid x)} = \exp(x^\top \gamma) \Lambda_0(dt),$$

where γ is some parameter to be estimated and $\Lambda_0(\cdot)$ is an unknown baseline cumulative hazard function. Our inversion formulae reveal that the parameters γ and Λ_0 depend on the conditional measures $H_0(\cdot \mid x)$, $H_1(\cdot \mid x)$, but these parameters are also connected to the parameter β used to model the cure rate $\mathbb{P}(T = \infty \mid x)$. The same is true for the parametric models.

2.5.2. *Kaplan-Meier mixture cure model.* Taylor (1995) suggested to estimate $F_{T,0}$ using a Kaplan-Meier type estimator. With such an approach one implicitly assumes that the law of T given X and given that $T < \infty$ does not depend on X. This is equivalent to supposing $\Lambda_{T,0}(\cdot \mid x) = \Lambda_{T,0}(\cdot)$. Next, to estimate $\Lambda_{T,0}(\cdot)$ one has to modify the unconditional version of the inversion formulae (2.5) to take into account the conditional probability of the event $\{T = \infty\}$. Following Taylor's approach we rewrite (2.11) as

$$\Lambda_{T,0}(dt) = \frac{H_1(dt \mid x)}{H_1([t,\infty) \mid x) + \int_{[t,\infty)} \left\{1 - \frac{1-\phi(x,\beta)}{\phi(x,\beta) F_{T,0}([s,\infty)) + 1 - \phi(x,\beta)}\right\} H_0(ds \mid x)}.$$

Next, assume that the last equality remains true if $H_0(dt \mid x)$ and $H_1(dt \mid x)$ are replaced by their unconditional versions, that is assume that

$$(2.15) \quad \Lambda_{T,0}(dt) = \frac{H_1(dt)}{H_1([t,\infty)) + \int_{[t,\infty)} \left\{1 - \frac{1-\phi(x,\beta)}{\phi(x,\beta) F_{T,0}([s,\infty)) + 1 - \phi(x,\beta)}\right\} H_0(ds)}.$$

See equations (2) and (3) in Taylor (1995). The equation above could be solved iteratively by a EM-type procedure: for a given β and an iteration $F_{T,0}^{(m)}(\cdot)$, build $\Lambda_{T,0}^{(m+1)}(dt)$ and the updated estimate $F_{T,0}^{(m+1)}(\cdot)$; see Taylor (1995) for the details. Let us point out that even if (T,C) is independent of X given that $T < \infty$, the subdistributions $H_0(\cdot \mid x)$ and $H_1(\cdot \mid x)$ still depend on x, since

$$H_0(dt \mid x) = F_T([t,\infty] \mid x) F_C(dt)$$

$$= \{\mathbb{P}(T < \infty \mid x)F_{T,0}([t,\infty)) + \mathbb{P}(T = \infty \mid x)\}F_C(dt),$$

and

$$H_1(dt \mid x) = F_C([t,\infty))F_T(dt \mid x) = \mathbb{P}(T < \infty \mid x)F_C([t,\infty))F_{T,0}(dt).$$

Hence, a more natural form of equation (2.15) is

$$\Lambda_{T,0}(dt) = \frac{H_1(dt \mid x)}{H_1([t,\infty)\mid x) + \int_{[t,\infty)}\left\{1 - \frac{1-\phi(x,\beta)}{\phi(x,\beta)F_{T,0}([s,\infty)) + 1 - \phi(x,\beta)}\right\}H_0(ds \mid x)}.$$

The investigation of an iterative procedure based on this alternative identity will be considered elsewhere.

3. Maximum likelihood estimation. Let (Y_i, δ_i, X_i) $(i = 1, \ldots, n)$ be a sample of n i.i.d. copies of the vector (Y, δ, X). We use a likelihood approach based on formulae (2.6), (2.11) and (2.12) to build an estimator of $\phi(\cdot, \beta)$. To build the likelihood we use estimates $\widehat{H}_k(\cdot \mid x)$ of the subdistributions $H_k(\cdot \mid x)$, $k \in \{0, 1\}$. We consider that for each x, $\widehat{H}(\cdot \mid x) = \widehat{H}_0(\cdot \mid x) + \widehat{H}_1(\cdot \mid x)$ is a proper distribution. These estimates are constructed with the sample of (Y, δ, X), without reference to any model for the variables T, C or for the conditional probability $\mathbb{P}(T < \infty \mid x)$. At this stage it is not necessary to impose a particular form for $\widehat{H}_k(\cdot \mid x)$. One could, for instance, use nonparametric estimators as proposed by Stone (1977) separately for $\delta = 1$ and $\delta = 0$. To derive the asymptotic results we will only impose that these estimators satisfy some mild conditions.

Let $\widehat{F}_{T,0}(\cdot \mid x)$ be defined as in equations (2.11) and (2.12) with $\widehat{H}_0(\cdot \mid x)$ and $\widehat{H}_1(\cdot \mid x)$ instead of $H_0(\cdot \mid x)$ and $H_1(\cdot \mid x)$, that is $\forall t \in \mathbb{R}, \forall x \in \mathcal{X}$,

$$(3.1) \quad \widehat{F}_{T,0}((t,\infty)\mid x) = \prod_{s \leq t}\left\{1 - \frac{\widehat{H}_1(ds \mid x)}{\widehat{H}([s,\infty)\mid x) - \mathbb{P}(\widehat{T = \infty}\mid x)\widehat{F}_C([s,\infty)\mid x)}\right\},$$

where

$$(3.2) \qquad \widehat{F}_C((t,\infty) \mid x) = \prod_{s \leq t}\left\{1 - \frac{\widehat{H}_0(ds \mid x)}{\widehat{H}([s,\infty) \mid x)}\right\},$$

and the estimator of the cure rate is derived from equation (3.3), i.e.

$$(3.3) \qquad \mathbb{P}(\widehat{T = \infty}\mid x) = \prod_{t \in \mathbb{R}}\left\{1 - \frac{\widehat{H}_1(dt \mid x)}{\widehat{H}([t,\infty) \mid x)}\right\}.$$

By construction, for each x, $\widehat{F}_{T,0}(\cdot \mid x) = 0$ is a proper distribution function and $\widehat{F}_{T,0}((t, \infty) \mid x) = 0$ for any t such that $\widehat{H}_1((t, \infty) \mid x) = 0$, provided $\widehat{H}_1(\cdot \mid x)$ puts a positive mass at the right endpoint of its support. See Lemma 8.1 in the Appendix for the proof of this result.

Let f_X denote the density of the covariate vector with respect to some dominating measure. The contribution of the observation (Y_i, δ_i, X_i) to the likelihood when $\delta_i = 1$ is then

$$\phi(X_i, \beta)\widehat{F}_{T,0}(\{Y_i\} \mid X_i)f_X(X_i)\widehat{F}_C((Y_i, \infty) \mid X_i),$$

while the contribution when $\delta_i = 0$ is

$$\widehat{F}_C(\{Y_i\} \mid X_i)f_X(X_i)[\phi(X_i, \beta)\widehat{F}_{T,0}([Y_i, \infty) \mid X_i) + 1 - \phi(X_i, \beta)].$$

Since the law of the covariate vector does not carry information on the parameter β, and since $\widehat{F}_C(\cdot \mid x)$ and $\widehat{F}_{T,0}(\cdot \mid x)$ could be directly computed from $\widehat{H}_0(\cdot \mid x)$ and $\widehat{H}_1(\cdot \mid x)$, we can drop the factors $\widehat{F}_{T,0}(\{Y_i\} \mid X_i)f_X(X_i)\widehat{F}_C((Y_i, \infty) \mid X_i)$ and $\widehat{F}_C(\{Y_i\} \mid X_i)f_X(X_i)$. Hence the criterion to be maximized with respect to $\beta \in B$ is $\widehat{L}_n(\beta)$ where

$$\widehat{L}_n(\beta) = \prod_{i=1}^{n} \phi(X_i, \beta)^{\delta_i}\Big\{\phi(X_i, \beta)\widehat{F}_{T,0}([Y_i, \infty) \mid X_i) + 1 - \phi(X_i, \beta)\Big\}^{1-\delta_i}.$$

The estimator we propose is

(3.4) $$\widehat{\beta} = \arg\max_{\beta \in B} \log \widehat{L}_n(\beta).$$

The existence and the unicity of $\widehat{\beta}$ could be guaranteed using the same conditions one would impose when considering the binary outcome regression model $\{\phi(\cdot, \beta) : \beta \in B\}$ and when observing the outcomes.

4. General asymptotic results. Little assumptions were needed so far. To proceed further with the asymptotic results we need to be more specific with respect to several aspects. In order to prove consistency, we have to control the asymptotic behavior of $\widehat{L}_n(\beta)$ along sequences of values of the parameter β. Such a control requires a control of denominators like

$$\widehat{H}([t, \infty) \mid x) - \mathbb{P}(\widehat{T = \infty} \mid x)\widehat{F}_C([t, \infty) \mid x)$$

on the support of $H_1(\cdot \mid x)$, uniformly with respect to x. A usual way to deal with this technical difficulty is to consider a finite threshold $\tau(x)$, that

could depend on the covariate vector value x, beyond which no uncensored lifetime is observed, i.e., for each x,

$$(4.1) \qquad \tau(x) \stackrel{def}{=} \tau_{H_1}(x) = \inf_t\{H_1((t, \infty) \mid x) = 0\} < \infty.$$

Moreover, to be able to keep denominators, such as in equation (3.1), away from zero, we require the condition

$$(4.2) \qquad \inf_{x \in \mathcal{X}} H_1(\{\tau(x)\} \mid x) H_0((\tau(x), \infty) \mid x) > 0.$$

In particular, this condition implies $\tau_{H_0}(x) > \tau(x)$, $\forall x \in \mathcal{X}$. Moreover, given condition (2.3), necessarily $H_0(\{\tau(x)\} \mid x) = 0$, $\forall x$. This means that $F_C(\{\tau(x)\} \mid x) = 0$, $\forall x$. This constraint on $H_0(\{\tau(x)\} \mid x)$ could be relaxed at the expense of suitable adjustments of the inversion formulae. For simplicity, we keep condition (2.3). Let us also notice that condition (4.2) implies $\inf_x F_C((\tau(x), \infty) \mid x) > 0$, and $\inf_x F_{T,0}(\{\tau(x)\} \mid x) > 0$.

Conditions like in equations (4.1)-(4.2) are more or less explicitly used in the literature of cure models, where so far the threshold τ was always considered independent of the covariate values. Sometimes τ is justified as representing a total follow-up of the study. For instance, Lu (2008) supposes that $Y = \min\{T, \min(C, \tau)\}$ and $\delta = \mathbf{1}\{T \le \min(C, \tau)\}$, where $T = \eta T^* + (1 - \eta)\infty$, with $T^* < \infty$ and $\eta \in \{0, 1\}$. The conditional probability of being cured is precisely the conditional probability of the event $\{\eta = 0\}$. Next, Lu (2008) supposes that $\inf_x \mathbb{P}(\tau \le T \le C \mid x) > 0$, and $\Lambda_0(\tau) < \infty$, where $\Lambda_0(\cdot)$ is the cumulative hazard function of T^*. All these conditions together clearly imply our conditions (4.1)-(4.2).

Fang *et al.* (2005) implicitly restrict the uncensored lifetimes to some compact interval $[0, \tau]$ and suppose $\mathbb{E}(\delta \mathbf{1}\{Y \ge \tau\}) > 0$. This could be possible only if $H_1(\{\tau\} \mid x) > 0$ for a set of values x with positive probability. In a proportional hazards context with the covariates taking values in a bounded set, as assumed by Fang *et al.* (2005), this is equivalent to $H_1(\{\tau\} \mid x) \ge c > 0$ for almost all x, for some constant c.

The fact that technical conditions similar to our conditions (4.1)-(4.2) could be traced in the cure models literature is not unexpected in view of our Section 2.5. Indeed, the existing approaches could be interpreted through our inversion formulae and thus the technical problems we face in the asymptotic investigation are expected to be also present in the alternative approaches.

4.1. *Consistency.* Let us sketch the arguments we use for proving the consistency of $\widehat{\beta}$. On one hand, if the conditional subdistributions $H_k(\cdot \mid x)$

are given, one can build the purely parametric likelihood

$$(4.3) \quad \mathcal{L}_n(\beta) = \prod_{i=1}^n \phi(X_i, \beta)^{\delta_i} \left\{ \phi(X_i, \beta) F_{T,0}([Y_i, \infty) \,|\, X_i) + 1 - \phi(X_i, \beta) \right\}^{1-\delta_i}.$$

By construction, $\mathcal{L}_n(\beta)$ is a functional of $H_0(\cdot \mid x)$ and $H_1(\cdot \mid x)$, $x \in \mathcal{X}$, while $\widehat{L}_n(\beta)$ is a functional of the estimated versions of $H_0(\cdot \mid x)$ and $H_1(\cdot \mid x)$. Hence, for deriving the consistency of $\widehat{\beta}$, first we show that β_0 defined by equation (2.13) is a well-separated maximum of the function $\beta \mapsto \mathbb{E}[\log \mathcal{L}_n(\beta)]$ and that $\sup_{\beta \in B} |\log \mathcal{L}_n(\beta) - \mathbb{E}[\log \mathcal{L}_n(\beta)]| = o_{\mathbb{P}}(n)$. Next, we check that

$$(4.4) \qquad\qquad \sup_{\beta \in B} |\log \widehat{L}_n(\beta) - \log \mathcal{L}_n(\beta)| = o_{\mathbb{P}}(n).$$

We then have that $\log \mathcal{L}_n(\widehat{\beta}) \geq \sup_{\beta \in B} \log \mathcal{L}_n(\beta) - o_{\mathbb{P}}(n)$. From this we will derive the consistency of $\widehat{\beta}$ using results from Section 5.2 in van der Vaart (1998). To prove condition (4.4), we have to guarantee the uniform convergence of $\widehat{H}_k - H_k$, $k \in \{0, 1\}$, as stated in Assumption (AC1) below.

Formally, to prove the consistency of $\widehat{\beta}$, we use the following assumptions:

(AC1) For $\tau(x)$ appearing in conditions (4.1)-(4.2),

$$\sup_{x \in \mathcal{X}} \sup_{t \in (-\infty, \tau(x)]} |\widehat{H}_k([t, \infty) \mid x) - H_k([t, \infty) \mid x)| = o_{\mathbb{P}}(1), \quad k \in \{0, 1\}.$$

(AC2) The parameter set $B \subset \mathbb{R}^p$ is compact.

(AC3) There exist some constants $a > 0$ and $c_1 > 0$ such that

$$\left| \phi(x, \beta) - \phi(x, \beta') \right| \leq c_1 \|\beta - \beta'\|^a, \qquad \forall \beta, \beta' \in B, \; \forall x, x' \in \mathcal{X},$$

where $\|A\|$ is the Euclidean norm of any vector A.

(AC4) $\inf_{\beta \in B} \inf_{x \in \mathcal{X}} \phi(x, \beta) > 0$.

THEOREM 4.1. *Assume that (AC1)-(AC4) and (2.2), (2.3), (2.8), (2.14), (4.1) and (4.2) hold true. Moreover, assume that there exists a unique value β_0 in the parameter set B such that (2.13) is true. Then, $\widehat{\beta} - \beta_0 = o_{\mathbb{P}}(1)$.*

Let us point out that the consistency result is stated in terms of the subdistributions of the observations and the conditional probability model $\{\phi(\cdot, \beta) : \beta \in B\}$. If the model is correctly specified, $\phi(x, \widehat{\beta})$ consistently estimates the cure probability $\mathbb{P}(T = \infty \mid x)$ for all x in the support of X. Let us also notice that condition (AC3) guarantees the Glivenko-Cantelli property we use in the proof and it will be satisfied in the common modeling situations. Condition (AC4) is a weak condition on the model $\phi(x, \beta)$ and is e.g. satisfied for the logistic model if $\mathcal{X}$ and B are compact.

4.2. *Asymptotic normality.* For the asymptotic normality we use the approach in Chen *et al.* (2003), Section 3.2. For this purpose we use the derivative of $\log \widehat{L}_n(\beta)$ with respect to β and we embed the nuisance functions $H_k([\cdot, \infty) \mid \cdot)$ ($k = 0, 1$) in a functional space $\mathcal{H}$ equipped with a pseudo-norm $\|\cdot\|_{\mathcal{H}}$. Both the space $\mathcal{H}$ and its pseudo-norm $\|\cdot\|_{\mathcal{H}}$ will be chosen depending on the estimators $\widehat{H}_k([\cdot, \infty) \mid \cdot)$, $k = 0, 1$, and have to satisfy certain conditions, given below. The true vector of nuisance functions is

$$\eta_0(t, x) = (\eta_{01}(t, x), \eta_{02}(t, x)) = (H_0([t, \infty) \mid x), H_1([t, \infty) \mid x)).$$

For each $x \in \mathcal{X}$ and for each $\eta_1, \eta_2 \in \mathcal{H}$, let $\eta(dt, x) = (\eta_1(dt, x), \eta_2(dt, x))$ be the measures associated to the functions $\eta(\cdot, x) = (\eta_1(\cdot, x), \eta_2(\cdot, x))$.

Let ∇_β denote the vector-valued partial differentiation operator with respect to the components of β. First, note that the vector of partial derivatives of the log-likelihood $\log \widehat{L}_n(\beta)$ with respect to the components of β equals

$$\nabla_\beta \log \widehat{L}_n(\beta) = \frac{1}{n} \sum_{i=1}^{n} m(Y_i, \delta_i, X_i; \beta, \widehat{\eta}), \quad \widehat{\eta} = (\widehat{H}_0, \widehat{H}_1),$$

and

$$(4.5) \qquad m(t, d, x; \beta, \eta) = \left\{ \frac{d}{\phi(x, \beta)} - \frac{(1 - d)T(\eta)(t, x)}{1 - \phi(x, \beta)T(\eta)(t, x)} \right\} \nabla_\beta \phi(x, \beta).$$

The map $\eta \mapsto T(\eta)(\cdot, \cdot)$ is given by

$$(4.6) \qquad\qquad\qquad T(\eta)(t, x) = F_{T,0}((-\infty, t] \mid x)$$

and defines $F_{T,0}$ as a functional of H_0 and H_1 via the composition

$$(H_0, H_1)(\cdot \mid \cdot) \mapsto (\mathbb{P}(T = \infty \mid \cdot), (F_C, H_0, H_1)(\cdot \mid \cdot)) \mapsto \Lambda_{T,0}(\cdot \mid \cdot) \mapsto F_{T,0}(\cdot \mid \cdot).$$

The maps in the above composition are defined in equations (3.3)-(2.5)-(2.6), (2.11) and (2.12), respectively. We also define

$$M_n(\beta, \eta) = \frac{1}{n} \sum_{i=1}^{n} m(Y_i, \delta_i, X_i; \beta, \eta) \quad \text{and} \quad M(\beta, \eta) = \mathbb{E}[m(Y, \delta, X; \beta, \eta)].$$

Hence, we have that $M(\beta_0, \eta_0) = 0$ and $\|M_n(\widehat{\beta}, \widehat{\eta})\| = \inf_{\beta \in B} \|M_n(\beta, \widehat{\eta})\|$, where $\widehat{\eta}(t, x) = (\widehat{\eta}_1(t, x), \widehat{\eta}_2(t, x)) = (\widehat{H}_0([t, \infty) \mid x), \widehat{H}_1([t, \infty) \mid x))$.

Next, we have to investigate the derivatives of $M(\beta, \eta)$ with respect to β and η. We will show in the proof of Theorem 4.2 below that

$$(4.7) \qquad \nabla_\beta M(\beta_0, \eta_0) = -\mathbb{E}\left\{ W(X) \nabla_\beta \phi(X, \beta_0) \nabla_\beta \phi(X, \beta_0)^\top \right\},$$

with the positive function $W(\cdot)$ given by

$$W(x) = \frac{\int_{(-\infty,\tau(x)]} F_C((t,\infty) \mid x) F_{T,0}(dt \mid x)}{\mathbb{P}(T < \infty \mid x)}$$
$$+ \int_{(-\infty,\tau(x)]} \frac{F_{T,0}^2((-\infty,t] \mid x) F_C(dt \mid x)}{1 - \mathbb{P}(T < \infty \mid x) F_{T,0}((-\infty,t] \mid x)}.$$

Further, define the Gâteaux derivative of $M(\beta, \eta_0)$ in the direction $[\eta - \eta_0]$

$$\nabla_\eta M(\beta, \eta_0)[\eta - \eta_0] = \lim_{\tau \to 0} \frac{1}{\tau} \Big[M(\beta, \eta_0 + \tau(\eta - \eta_0)) - M(\beta, \eta_0) \Big].$$

We show in the proof of Theorem 4.2 below how to compute this derivative. We make the following assumptions:

(AN1) The value β_0 is an interior point of the compact set $B \subset \mathbb{R}^p$. For any $x \in \mathcal{X}$, the function $\beta \to \phi(x, \beta)$, $\beta \in B$, is twice continuously differentiable and the first and second order partial derivatives are bounded uniformly in $(x, \beta) \in \mathcal{X} \times B$.

(AN2) $H_k([\cdot, \infty) \mid \cdot) \in \mathcal{H}$ for $k = 0, 1$.

(AN3) The matrix $\mathbb{E}\left\{ \nabla_\beta \phi(X, \beta_0) \nabla_\beta \phi(X, \beta_0)^\top \right\}$ is positive definite.

(AN4) For $k = 0, 1$, the estimator $\widehat{H}_k([\cdot, \infty) \mid \cdot)$ satisfies the following :

    (i) $\mathbb{P}(\widehat{H}_k([\cdot, \infty) \mid \cdot) \in \mathcal{H}) \to 1$;

    (ii) $\|(\widehat{H}_k - H_k)([\cdot, \infty) \mid \cdot)\|_\mathcal{H} = o_\mathbb{P}(n^{-1/4})$;

    (iii) For some functions ψ_{jk}, $j = 1, \dots 5, k = 0, 1$, defined in (8.3) in the Appendix, there exist functions Ψ_1 and Ψ_2, such that

$$\sum_{k=0}^{1} \mathbb{E}^*\Big[\psi_{1k}(Y, X) \int_{(-\infty,Y)} \psi_{2k}(u, X) d\left((\widehat{H}_k - H_k)([u, \infty) \mid X) \right) \Big]$$
$$= \frac{1}{n} \sum_{i=1}^{n} \Psi_1(Y_i, \delta_i, X_i) + R_{1n},$$

and

$$\sum_{k,\ell=0}^{1} \mathbb{E}^*\Big[\psi_{3k}(Y, X) \int_{(-\infty,Y)} \psi_{4k}(u, X) \psi_{5k}\left((\widehat{H}_k - H_k)([u, \infty) \mid X), X \right)$$
$$\times dH_\ell(u \mid X) \Big] = \frac{1}{n} \sum_{i=1}^{n} \Psi_2(Y_i, \delta_i, X_i) + R_{2n},$$

where $\mathbb{E}^*$ denotes conditional expectation given the sample, taken with respect to the generic variables Y, δ, X. Moreover,

$$\mathbb{E}[\Psi_1(Y, \delta, X)] = \mathbb{E}[\Psi_2(Y, \delta, X)] = 0, \quad \|R_{1n}\| + \|R_{2n}\| = o_{\mathbb{P}}(n^{-1/2}).$$

(AN5) The class $\mathcal{H}$ satisfies $\int_0^\infty \sqrt{\log N(\varepsilon, \mathcal{H}, \|\cdot\|_{\mathcal{H}})} d\varepsilon < \infty$, where $N(\varepsilon, \mathcal{H}, \|\cdot\|_{\mathcal{H}})$ is the ε-covering number of the space $\mathcal{H}$ with respect to the norm $\|\cdot\|_{\mathcal{H}}$, i.e. the smallest number of balls of $\|\cdot\|_{\mathcal{H}}$-radius ε needed to cover the space $\mathcal{H}$.

THEOREM 4.2. *Assume that $\widehat{\beta} - \beta_0 = o_{\mathbb{P}}(1)$ and that (AN1)-(AN5) and (2.2), (2.3), (2.8), (2.14), (4.1) and (4.2) hold true. Then,*

$$n^{1/2}\left(\widehat{\beta} - \beta_0\right) \Rightarrow \mathcal{N}\left(0, \{\nabla_\beta M(\beta_0, \eta_0)\}^{-1} V \{\nabla_\beta M(\beta_0, \eta_0)\}^{-1}\right),$$

where $V = Var\big(m(Y, \delta, X; \beta_0, \eta_0) + \Psi_1(Y, \delta, X) + \Psi_2(Y, \delta, X)\big)$, $\nabla_\beta M(\beta_0, \eta_0)$ is given in (4.7), and Ψ_1 and Ψ_2 are defined in Assumption (AN4)(iii).

The structure of the asymptotic variance of $\widehat{\beta}$ reveals the differences with a standard binary outcome regression. The expression of the functions $\nabla_\beta M$ and m include terms depending on H_0 and H_1 which account for the uncertainty on the censored observations, which could correspond to cured or uncured subjects. The expression of the variance V includes the functions Ψ_1 and Ψ_2 which account for the fact that H_0 and H_1 are unknown and have been estimated from the data.

4.3. *Bootstrap consistency.* Although in principle one can use Theorem 4.2 above for making inference, the asymptotic variance of $\widehat{\beta}$ has a complicated structure, and estimating it would not only be cumbersome, but the precision of the estimate for small samples could moreover be rather poor. In this section we show that a bootstrap procedure can be used to estimate the asymptotic variance of $\widehat{\beta}$, to approximate the whole distribution of $\widehat{\beta}$ or to construct confidence intervals or test hypotheses regarding β_0.

Here, we propose to use a naive bootstrap procedure, consisting in drawing triplets $(Y_i^*, \delta_i^*, X_i^*)$, $1 \le i \le n$, randomly with replacement from the data (Y_i, δ_i, X_i), $1 \le i \le n$. Let $\widehat{H}_k^*$ be the same estimator as $\widehat{H}_k$ ($k = 0, 1$) but based on the bootstrap data, and for each (β, η) let $M_n^*(\beta, \eta) = n^{-1} \sum_{i=1}^n m(Y_i^*, \delta_i^*, X_i^*; \beta, \eta)$. Define the bootstrap estimator $\widehat{\beta}^*$ to be any sequence satisfying $\|M_n^*(\widehat{\beta}^*, \widehat{\eta}^*) - M_n(\widehat{\beta}, \widehat{\eta})\| = \inf_{\beta \in B} \|M_n^*(\beta, \widehat{\eta}^*) - M_n(\widehat{\beta}, \widehat{\eta})\|$, where $\widehat{\eta}^*(t, x) = (\widehat{\eta}_1^*(t, x), \widehat{\eta}_2^*(t, x)) = (\widehat{H}_0^*([t, \infty) \mid x), \widehat{H}_1^*([t, \infty) \mid x))$.

The following result shows that the bootstrap works, in the sense that it allows to recover correctly the distribution of $n^{1/2}(\widehat{\beta} - \beta_0)$.

THEOREM 4.3. *Assume that $\widehat{\beta} - \beta_0 = o_{\mathbb{P}}(1)$ and that (AN1)-(AN5) hold true. Moreover, assume that $\nabla_\beta M(\beta, \eta)$ is continuous in η (with respect to $\|\cdot\|_{\mathcal{H}}$) at $(\beta, \eta) = (\beta_0, \eta_0)$, and that (AN4) holds true with $\widehat{H}_k - H_k$ replaced by $\widehat{H}_k^* - \widehat{H}_k$ ($k = 0, 1$) in $\mathbb{P}^*$-probability. Then,*

$$\sup_{u \in \mathbb{R}^p} \left| \mathbb{P}^*\big(n^{1/2}(\widehat{\beta}^* - \widehat{\beta}) \leq u\big) - \mathbb{P}\big(n^{1/2}(\widehat{\beta} - \beta_0) \leq u\big) \right| = o_{\mathbb{P}}(1),$$

where $\mathbb{P}^*$ denotes probability conditionally on the data, and where the inequality sign means the component-wise inequality for vectors.

4.4. *Verification of the assumptions for kernel estimators.* Next, we illustrate the verification of the assumptions of our asymptotic results when the conditional subdistributions H_k are estimated by means of kernel smoothing.

Consider the case where X is composed of continuous and discrete components, that is $X = (X_c, X_d) \in \mathcal{X}_c \times \mathcal{X}_d \subset \mathbb{R}^{d_c} \times \mathbb{R}^{d_d}$, with $d_c + d_d = d \geq 1$. For simplicity, assume that the support $\mathcal{X}_d$ of the discrete subvector X_d is finite. We also assume that the lifetime T has not been transformed by a logarithmic or other transformation, so that its support is $(0, \infty]$. The subdistributions $H_k([t, \infty) \mid x)$ are assumed continuous in both arguments t and x and are estimated by means of a kernel estimator:

$$(4.8) \qquad \widehat{H}_k([t, \infty) \mid x) = \sum_{i=1}^n \frac{\widetilde{K}_{h_n}(X_i - x)}{\sum_{j=1}^n \widetilde{K}_{h_n}(X_j - x)} I(Y_i \geq t, \delta_i = k),$$

where for any $(x_c, x_d) \in \mathcal{X}_c \times \mathcal{X}_d$, $\widetilde{K}_{h_n}(X_i - x) = K_{h_n}(X_{c,i} - x_c)I(X_{d,i} = x_d)$, h_n is a bandwidth sequence, $K_h(\cdot) = K(\cdot/h)/h^{d_c}$, $K(u) = k(u_1) \cdot \ldots \cdot k(u_{d_c})$ and $k(\cdot)$ is a probability density function.

Nonparametric smoothing of continuous covariates is possible for dimensions d_c larger than 1. However, the technical arguments necessary to verify the assumptions used for the asymptotic results are tedious. Therefore, in the following we consider $d_c = 1$. The discrete covariates do not contribute to the curse of dimensionality, and therefore d_d could be larger than 1. However, for simplicity, below we do not consider discrete covariates.

To satisfy assumption (AN4), we impose the following conditions:

(C1) The sequence h_n satisfies $nh_n^4 \to 0$ and $nh_n^{3+\zeta}/\log n \to \infty$ for some $\zeta > 0$.

(C2) The support $\mathcal{X}$ of X is a compact subset of $\mathbb{R}$ and the density $f_X(\cdot)$ of X is bounded and bounded away from zero.

(C3) The probability density function K has compact support, $\int uK(u)du = 0$ and K is twice continuously differentiable.

(C4) There exists τ such that $\sup_x \tau(x) < \tau < \inf_x \tau_{H_0}(x) \leq \infty$.

Further, let $\mathcal{F}_1$ be the space of functions from $[0, \tau]$ to $[0, 1]$ with variation bounded by $0 < M < \infty$, and let $\mathcal{F}_2$ be the space of continuously differentiable functions f from $\mathcal{X}$ to $[-M, M]$ that satisfy $\sup_{x \in \mathcal{X}} |f'(x)| \leq M$ and $\sup_{x_1, x_2 \in \mathcal{X}} |f'(x_1) - f'(x_2)|/|x_1 - x_2|^\epsilon \leq M$ for some $0 < \epsilon < 1$. Let

$$
\mathcal{H} = \left\{ (t, x) \rightarrow \eta(t, x) : \eta(\cdot, x) \in \mathcal{F}_1, \frac{\partial}{\partial x}\eta(\cdot, x) \in \mathcal{F}_1 \text{ for all } x \in \mathcal{X}, \right.
$$
$$
\left. \text{and } \eta(t, \cdot) \in \mathcal{F}_2 \text{ for all } 0 \leq t \leq \tau \right\}.
$$

We define the following norm associated with the space $\mathcal{H}$: for $\eta \in \mathcal{H}$, let $\|\eta\|_\mathcal{H} = \sup_{0 \leq t \leq \tau} \sup_{x \in \mathcal{X}} |\eta(t, x)|$. Then, it follows from Propositions 1 and 2 in Akritas and Van Keilegom (2001) that $\mathbb{P}(\widehat{H}_k \in \mathcal{H}) \rightarrow 1$ provided $nh_n^{3+\zeta}(\log n)^{-1} \rightarrow \infty$, with $\zeta > 0$ as in condition (C1). Moreover, $\sup_{x,t} |\widehat{H}_k([t, \infty) \mid x) - H_k([t, \infty) \mid x)| = O_\mathbb{P}((nh_n)^{-1/2}(\log n)^{1/2}) = o_\mathbb{P}(n^{-1/4})$ (see Proposition 1 in Akritas and Van Keilegom 2001). The class $\mathcal{H}$ satisfies assumption (AN5) thanks to Lemma 6.1 in Lopez (2011). It remains to show the validity of assumption (AN4)(iii). We will show the first statement, the second one can be shown in a similar way. Note that the left hand side equals

$$
\sum_{k=0}^{1} \mathbb{E}_i \left[\psi_{1k}(Y, X) \int_{(0,Y)} \psi_{2k}(u, X) \frac{1}{n} f_X^{-1}(X) \right.
$$
$$
\left. \times \sum_{i=1}^{n} K_{h_n}(X_i - X) d\Big(I(Y_i \geq u, \delta_i = k) - H_k([u, \infty) \mid X) \Big) \right] + o_\mathbb{P}(n^{-1/2})
$$
$$
= \sum_{k=0}^{1} \mathbb{E}_i \left[\frac{\psi_{1k}(Y, X)}{f_X(X)} \frac{1}{n} \sum_{i=1}^{n} K_{h_n}(X_i - X) \Big\{ -\psi_{2k}(Y_i, X) I(Y_i \leq Y, \delta_i = k) \right.
$$
$$
\left. - \int_{(0,Y)} \psi_{2k}(u, X) \, dH_k([u, \infty) \mid X) \Big\} \right] + o_\mathbb{P}(n^{-1/2})
$$
$$
= \frac{1}{n} \Psi_1(Y_i, \delta_i, X_i) + o_\mathbb{P}(n^{-1/2} + h_n^2),
$$

where $\mathbb{E}_i[\cdot] = \mathbb{E}[\cdot \mid Y_i, \delta_i, X_i]$, $\mathbb{E}_i[\cdot \mid X = X_i] = \mathbb{E}[\cdot \mid Y_i, \delta_i, X_i, X = X_i]$ and

$$
\Psi_1(Y_i, \delta_i, X_i) = -\sum_{k=0}^{1} \Big\{ \mathbb{E}_i(\psi_{1k}(Y, X_i) I(Y_i \leq Y, \delta_i = k) \mid X = X_i) \psi_{2k}(Y_i, X_i)
$$
$$
- \int \mathbb{E}_i(\psi_{1k}(Y, X_i) I(u \leq Y, \delta_i = k) \mid X = X_i) \psi_{2k}(u, X_i) dH((-\infty, u] \mid X_i) \Big\}.
$$

which is of the required form. Moreover, the $o_\mathbb{P}(\cdot)$ rates hold uniformly with respect to i, and we have $\mathbb{E}[\Psi_1(Y_i, \delta_i, X_i)] = 0$.

Plugging the estimators $\widehat{H}_0$, $\widehat{H}_1$ and $\widehat{\beta}$ into the equations (2.11) and (2.12), we obtain an estimator for $F_{T,0}$. Then, under the additional conditions (C1) to (C4), with $\mathbb{P}(\widehat{T = \infty} \mid x)$ defined in equation (3.3), uniformly in $0 < t \leq \tau$ and $x \in \mathcal{X}$, we have the following i.i.d. representation:

$$
\begin{aligned}
(4.9) \quad & \widehat{F}_{T,0}((0,t] \mid x) - F_{T,0}((0,t] \mid x) \\
& = \frac{F_{T,0}((t,\tau] \mid x)}{f_X(x)} \\
& \quad \times \frac{1}{n}\sum_{i=1}^{n} K_{h_n}(h_n^{-1}(x - X_i))\left\{\frac{\delta_i I(Y_i \leq t)}{H([Y_i,\infty) \mid x)} - \int_{(0,t\wedge Y_i]}\frac{H_1(ds \mid x)}{H^2([s,\infty) \mid x)}\right\} \\
& \quad - \left\{\mathbb{P}(\widehat{T = \infty} \mid x) - [1 - \phi(x,\beta_0)]\right\}\Upsilon(t,x;H_0,H_1,\beta_0) \\
& \quad + o_{\mathbb{P}}((nh_n)^{-1/2}),
\end{aligned}
$$

uniformly in x and t. The expression of the functional Υ, as well as the i.i.d. representation of the nonparametric estimator $\mathbb{P}(\widehat{T = \infty} \mid x)$, are given in the Supplementary Material. As a consequence of the i.i.d representation (4.9), weak convergence theorems for $\sqrt{nh_n}\left(\widehat{F}_{T,0} - F_{T,0}\right)$ could be derived by standard arguments, that we omit for the sake of brevity.

5. Efficiency aspects. Our modeling situation belongs to the general information loss model framework, see van der Vaart (1998), section 25.5.2. Here the observed variables (Y,δ,X) are a measurable transformation of the (partially) unobservable variables (T,C,X). In order to separate the parametric and nonparametric parts of our semiparametric model, let us introduce a latent finite lifetime $T_0 \in \mathbb{R}$ and a latent binary variable Δ such that $T = T_0$ when $\Delta = 1$ and $T = \infty$ when $\Delta = 0$. Thus we have $\{T < \infty\} = \{\Delta = 1\}$. Let us assume $\Delta \perp (T_0,C) \mid X$ and $T_0 \perp C \mid X$, which guarantees the conditional independence assumption (2.2). The infinite dimensional parameters of the model are given by the laws of C, X and T_0. Given the structure of the likelihood, and assuming that the laws of C and X do not carry any information on the parameter of interest β, we could only consider submodels in the direction of $F_{T,0}$. Then the score function for the path induced in our model by a submodel in the direction of $F_{T,0}$, with score function $b(\cdot)$, is given by

$$
\begin{aligned}
(5.1) \quad g(Y,\delta,X) &= \mathbb{E}\left[b(T_0,X) \mid Y,\delta,X\right] \\
&= \delta b(Y,X) + (1-\delta)\phi(X,\beta)\frac{\int_{(Y,\infty)} b(t,X)dF_{T,0}(t \mid X)}{1 - \phi(X,\beta)F_{T,0}((-\infty,Y] \mid X)}.
\end{aligned}
$$

The tangent set, say $\mathcal{T}$, in the nonparametric direction of $F_{T,0}$ is composed of vector-valued functions $g(\cdot)$ depending on the observations and defined by squared norm integrable centered vector-valued functions $b(T_0, X)$. To project the score of the parametric part of our model

$$m_0(Y, \delta, X) = \left\{ \frac{\delta}{\phi(X, \beta_0)} - \frac{(1-\delta)F_{T,0}((-\infty, Y] \mid X)}{1 - \phi(X, \beta_0)F_{T,0}((-\infty, Y] \mid X)} \right\} \nabla_\beta \phi(X, \beta_0)$$

onto the tangent set $\mathcal{T}$, we have to find a function $b_0(T_0, X)$, yielding $g_0(Y, \delta, X)$ defined as in equation (5.1), such that

$$\mathbb{E}\left[\{ m_0(Y, \delta, X) - g_0(Y, \delta, X) \} \, g_*(Y, \delta, X) \right] = 0$$

for all $g_*(Y, \delta, X)$ defined by score functions $b_*(T_0, X)$. In the Supplementary Material, we show that this equation is an integral equation with a solution that could be approximated numerically.

With this solution we then obtain the efficient score

$$S_{\text{eff}}(Y, \delta, X; \beta_0, \eta_0) = m_0(Y, \delta, X) - g_0(Y, \delta, X).$$

A semi-parametrically efficient estimator could be obtained by solving the equation

$$n^{-1} \sum_{i=1}^{n} S_{\text{eff}}(Y_i, \delta_i, X_i; \beta, \widehat{\eta}) = o_{\mathbb{P}}(n^{-1/2})$$

for β. The solution $\widehat{\beta}_{\text{eff}}$ will have the following i.i.d. representation:

$$\widehat{\beta}_{\text{eff}} - \beta_0 = n^{-1} \sum_{i=1}^{n} \left[\mathbb{E}(S_{\text{eff},i} S_{\text{eff},i}^\top) \right]^{-1} S_{\text{eff},i} + o_{\mathbb{P}}(n^{-1/2})$$

where $S_{\text{eff},i} = S_{\text{eff}}(Y_i, \delta_i, X_i; \beta_0, \eta_0)$.

6. Simulations. In this section we will investigate the small sample performance of our estimation method. We consider the following model. The covariate X is generated from a uniform distribution on $[-1, 1]$, and the conditional probability $\phi(x, \beta)$ of not being cured follows a logistic model : $\phi(x, \beta) = \text{logit}(\beta_1 + \beta_2 x)$, for any $-1 \le x \le 1$. We will work with $\beta_0 = (\beta_{01}, \beta_{02}) = (1.75, 2)$ and $(1.1, 2)$, corresponding to an average cure rate of 20% respectively 30%. The conditional distribution function $F_{T,0}(\cdot|x)$ of the uncured individuals is constructed as follows. For a given $X = x$, we draw T from a Weibull distribution: $F_{T,0}([t, \infty)|x) = \exp(-t^{k(x)} \exp(\gamma_0 + \gamma_1 x))$, where $k(x) = k + a(1/(1+x) - 1/2)$. We take $k = 0.75$, $\gamma_0 = 0.5$ and $\gamma_1 = 1$,

and $a \in \{0, 1, 2\}$. Next, in order to respect condition (4.2), we truncate this distribution at $\tau(x)$, where $\tau(x) \equiv \tau$ equals the quantile of order 0.97 of $F_{T,0}(y|0)$ when $a = 0$, and the quantile of order 0.97 of $F_{T,0}(y|x)$ when $a = 1$ or 2. Note that the Cox model is only verified if $a = 0$, and in that case we have a baseline cumulative hazard function given by $\exp(\gamma_0)t^k$ for $t \leq \tau$ and given by infinity for $t > \tau$.

Next, we generate the censoring variable C as follows: $F_C([t, \infty)|x) = \exp(-t^{k(x)} \exp(\gamma_0 + \gamma_1 x)/p_C)$, where $p_C = (1 - p_{\text{cens}})/(p_{\text{cens}} - p_{\text{cure}})$, with p_{cens} the proportion of censored subjects in the population when $a = 0$ and p_{cure} the proportion of cured subjects in the population when $a = 0$. When $p_{\text{cure}} = 0.20$ we take $p_{\text{cens}} = 0.25$ (model 1) and 0.35 (model 2) and when $p_{\text{cure}} = 0.30$ we take $p_{\text{cens}} = 0.35$ (model 3) and 0.45 (model 4).

In what follows we will compare our estimator of β with the estimator proposed by Lu (2008) which assumes a Cox model for the uncured individuals. The estimated β coefficients under the Cox model are obtained using the R package smcure.

For our estimation procedure we used the kernel estimators given in Section 4.4, and we programmed $\widehat{\beta}$ using the optimization procedure optim in R. The details of the maximization procedure we used are provided in the Supplementary Material. The results of our maximization procedure are given in Table 1 for the case where $\beta_0 = (1.75, 2)$ (model 1 and 2), and in Table 2 for the case where $\beta_0 = (1.1, 2)$ (model 3 and 4). A total of 1020 samples of size $n = 200$ and $n = 400$ are generated, and the tables show the bias and the variance of the estimators $\widehat{\beta}_1$ and $\widehat{\beta}_2$ obtained under the Cox model and from our procedure. For stability reasons, the results are truncated in the sense that the lowest and highest 1% of the estimators are omitted, and so the actual number of samples is 1000. The kernel function K is taken equal to the Epanechnikov kernel : $K(u) = (3/4)(1 - u^2)I(|u| \leq 1)$. For the bandwidth h of the kernel estimators $\widehat{H}_k(\cdot|X_i)$ ($k = 0, 1$), defined as in equation (4.8), we used the cross-validation (CV) procedure proposed by Li *et al.* (2013) for kernel estimators of conditional distribution functions. The CV procedure is implemented in the package np in R. For each sample in our simulation, we calculated these bandwidths for $\widehat{H}_0$ and $\widehat{H}_1$ and used the average of these two bandwidths, truncated to the interval $[n^{-2/7}, 5n^{-2/7}]$ in order to verify regularity condition (C1).

The tables show that our estimator outperforms the one that is based on the Cox model when the Cox model is not verified ($a = 1$ or $a = 2$), whereas when the model is satisfied (so when $a = 0$) the two estimators behave rather similarly.

Next, we look at the estimation of the quartiles of the distribution $F_{T,0}(\cdot|x)$

| | | | Model 1 | | | | Model 2 | | | |
| | | | New | | Cox | | New | | Cox | |
n	a	Par.	Bias	Var	Bias	Var	Bias	Var	Bias	Var
200	0	β_1	.010	.059	.044	.063	-.001	.092	.080	.111
		β_2	-.034	.175	.061	.186	-.038	.258	.116	.305
	1	β_1	.025	.058	.063	.053	.032	.088	.087	.082
		β_2	-.068	.180	-.350	.175	.093	.273	-.590	.245
	2	β_1	.027	.058	.093	.053	.030	.079	.085	.074
		β_2	-.076	.180	-.489	.172	-.097	.247	-.593	.220
400	0	β_1	.005	.032	.027	.034	-.001	.045	.039	.051
		β_2	-.039	.088	.032	.091	-.045	.125	.051	.134
	1	β_1	.020	.031	.054	.029	.022	.044	.067	.039
		β_2	-.082	.090	-.372	.086	-.133	.129	-.637	.113
	2	β_1	.021	.031	.086	.029	.025	.045	.135	.042
		β_2	-.087	.090	-.515	.083	-.142	.131	-.842	.111

TABLE 1

Bias and Var of $\widehat{\beta}_1$ and $\widehat{\beta}_2$ for two sample sizes and three values of a. Here, $\mathbb{P}(cured) = 0.2$ and $\mathbb{P}(censoring)$ is either 0.25 (model 1) or 0.35 (model 2) for $a = 0$. The Cox model is satisfied for $a = 0$.

when $x = -0.5$. We estimate these quartiles by means of our nonparametric estimator $\widehat{F}_{T,0}(\cdot|x)$ and by means of the Cox model studied in Lu (2008). The results given in Figure 1 for $n = 200$ show that, as could be expected, when the Cox model is satisfied (i.e. when $a = 0$) our quantile estimators

| | | | Model 3 | | | | Model 4 | | | |
| | | | New | | Cox | | New | | Cox | |
n	a	Par.	Bias	Var	Bias	Var	Bias	Var	Bias	Var
200	0	β_1	.002	.039	.024	.040	-.009	.059	.053	.069
		β_2	-.039	.136	.039	.138	-.026	.195	.096	.220
	1	β_1	.017	.038	.057	.036	.017	.059	.076	.058
		β_2	-.049	.136	-.255	.140	-.052	.212	-.427	.198
	2	β_1	.018	.038	.085	.037	.020	.060	.129	.063
		β_2	-.052	.136	-.366	.138	-.056	.214	-.599	.200
400	0	β_1	-.006	.020	.008	.021	-.004	.030	.023	.032
		β_2	-.040	.064	.024	.067	-.034	.099	.047	.102
	1	β_1	.006	.020	.044	.019	.014	.029	.066	.028
		β_2	-.065	.066	-.266	.067	-.094	.100	-.459	.096
	2	β_1	.007	.020	.073	.019	.015	.029	.123	.031
		β_2	-.069	.066	-.377	.066	-.100	.100	-.632	.096

TABLE 2

Bias and Var of $\widehat{\beta}_1$ and $\widehat{\beta}_2$ for two sample sizes and three values of a. Here, $\mathbb{P}(cured) = 0.3$ and $\mathbb{P}(censoring)$ is either 0.35 (model 3) or 0.45 (model 4) for $a = 0$. The Cox model is satisfied for $a = 0$.
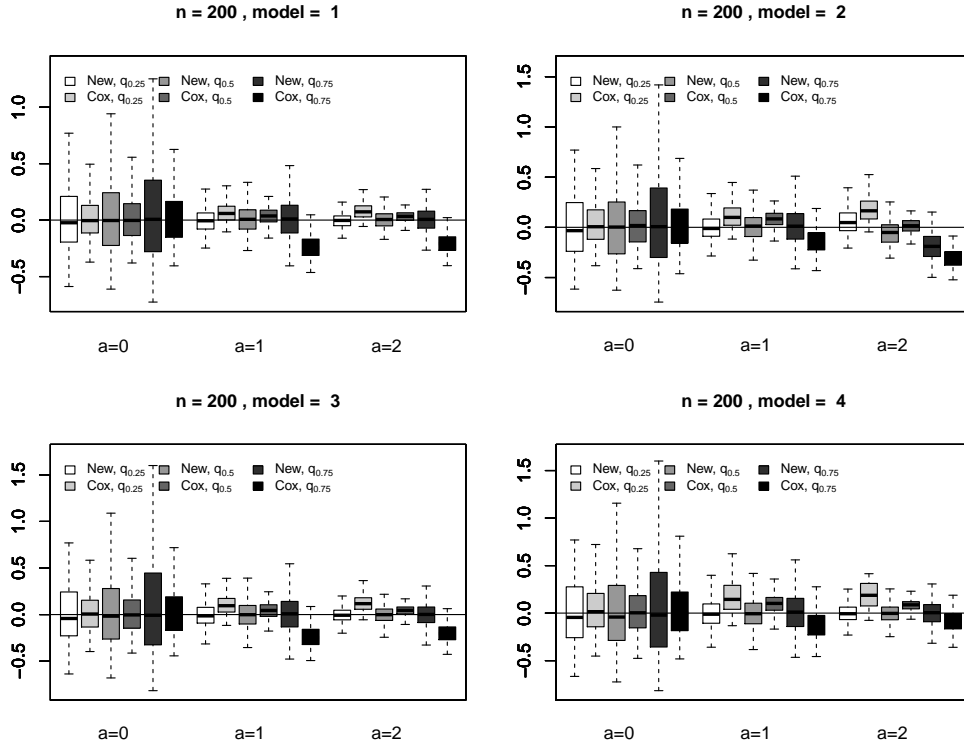
FIG 1. *Boxplots for quantile estimators.*

have a larger variance than those obtained under the Cox model, whereas the bias is similar for both estimators. On the other hand, when the Cox model is not satisfied (i.e. when $a = 1$ or 2), the estimated quartiles obtained under the Cox model are heavily biased. This shows the importance of having a model that does not impose any assumptions on the distribution of the uncured individuals and which still provides very accurate estimators for the logistic part of the model.

We also verify how close the distributions of $\widehat{\beta}_1$ and $\widehat{\beta}_2$ are to a normal distribution. We know thanks to Theorem 4.2 that the estimators converge to a normal limit when n tends to infinity. In the Supplementary Material we provide simulation results showing that the asymptotic normal approximation is already quite accurate for $n = 200$.

Next, we verify the accuracy of the naive bootstrap proposed in Section 4.3. Figure 3 shows boxplots of the variance of $\widehat{\beta}_1$ and $\widehat{\beta}_2$, centered around the empirical variance of the estimators of β_1 and β_2. The boxplots are obtained from 500 bootstrap resamples for each of 1000 samples of size $n = 200$.

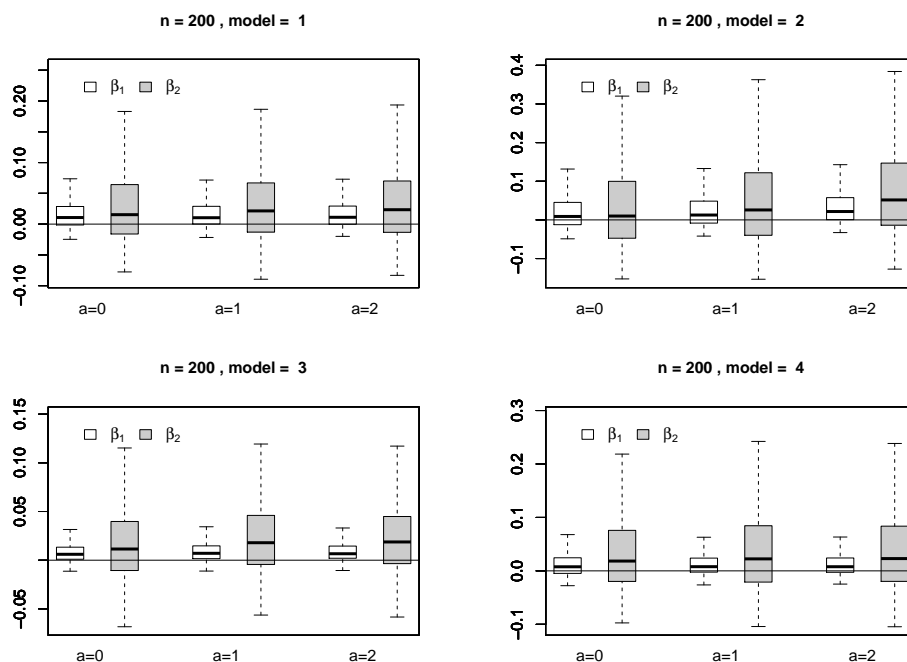The boxplots show that the bootstrap variance is well centered around the corresponding empirical variance.

FIG 2. *Boxplots of the bootstrap variance of $\widehat{\beta}_1$ and $\widehat{\beta}_2$, centered around the empirical variance of the estimators of β_1 and β_2. The boxplots are obtained from 500 bootstrap resamples for each of 1000 samples of size $n = 200$, using the new estimation method.*

We also obtain percentile bootstrap confidence intervals for β_1 and β_2, and calculate the coverage probability and average length of these intervals. The results are given in Table 3 and show that the coverage is very close to the desired 95% and that the results are better for $n = 400$ than for $n = 200$.

In the last simulation study, we investigate what happens when the model is misspecified. More precisely, we generate data from a model containing two covariates, and we want to know how our estimator and the estimator under the Cox model behave when fitting a mixture cure model containing only one of the two covariates. The results provided in the Supplementary Material show that our nonparametric model on $F_{T,0}$ is able to compensate better for the forgotten covariate than the Cox model. This is what is expected given that our model on $F_{T,0}$ is nonparametric and hence more flexible than the Cox model.

			$n = 200$				$n = 400$			
			Model				Model			
a	Par.		1	2	3	4	1	2	3	4
0	β_1	Cov	0.97	0.96	0.96	0.96	0.95	0.96	0.95	0.96
		Len	1.07	1.31	0.85	1.05	0.73	0.90	0.58	0.73
	β_2	Cov	0.95	0.97	0.96	0.97	0.96	0.96	0.96	0.95
		Len	1.77	2.12	1.53	1.87	1.21	1.46	1.04	1.29
1	β_1	Cov	0.97	0.96	0.95	0.96	0.95	0.96	0.95	0.97
		Len	1.07	1.31	0.85	1.05	0.73	0.89	0.58	0.73
	β_2	Cov	0.95	0.96	0.96	0.96	0.95	0.95	0.95	0.95
		Len	1.81	2.23	1.56	1.96	1.25	1.52	1.07	1.34
2	β_1	Cov	0.97	0.97	0.95	0.96	0.95	0.96	0.95	0.96
		Len	1.07	1.31	0.85	1.05	0.73	0.89	0.58	0.73
	β_2	Cov	0.95	0.97	0.95	0.96	0.95	0.95	0.95	0.96
		Len	1.81	2.22	1.56	1.97	1.24	1.52	1.07	1.34

TABLE 3

Coverage probability (Cov) and average length (Len) of bootstrap confidence intervals for $\widehat{\beta}_1$ and $\widehat{\beta}_2$ for two sample sizes and three values of a, using the new estimation method.

7. Data analysis. Let us now apply our estimation procedure on two medical data sets. The first one is about 286 breast cancer patients with lymph-node-negative breast cancer treated between 1980 and 1995 (Wang *et al.* (2005)). The event of interest is distant-metastasis, and the associated survival time is the distant metastasis-free survival time (defined as the time to first distant progression or death, whichever comes first). 107 of the 286 patients experience a relapse from breast cancer. The plot of the Kaplan-Meier (KM) estimator of the data is given in Figure 3(a) and shows a large plateau at about 0.60. Furthermore, a large proportion of the censored observations is in the plateau, which suggests that a cure model is appropriate for these data. As a covariate we use the age of the patients, which ranges from 26 to 83 years and the average age is about 54 years.

We estimate β using our estimator and using the estimator based on the Cox model. The bandwidth h is selected using cross-validation, as in the simulation section. The estimated intercept is -0.168 (with standard deviation equal to 0.502 obtained using a naive bootstrap procedure), and the estimated slope parameter is -0.006 (with standard deviation equal to 0.009). Under the Cox model the estimated intercept and slope are respectively 0.063 and -0.010. A 95% confidence interval is given by $(-1.152, 0.815)$ for the intercept and $(-0.024, 0.012)$ for the slope, where the variance is again based on the naive bootstrap procedure. The graph of the two estimators of the function $\phi(x)$ is given in Figure 3(b). The estimated coefficients and curves are quite close to each other, suggesting that the Cox model might be
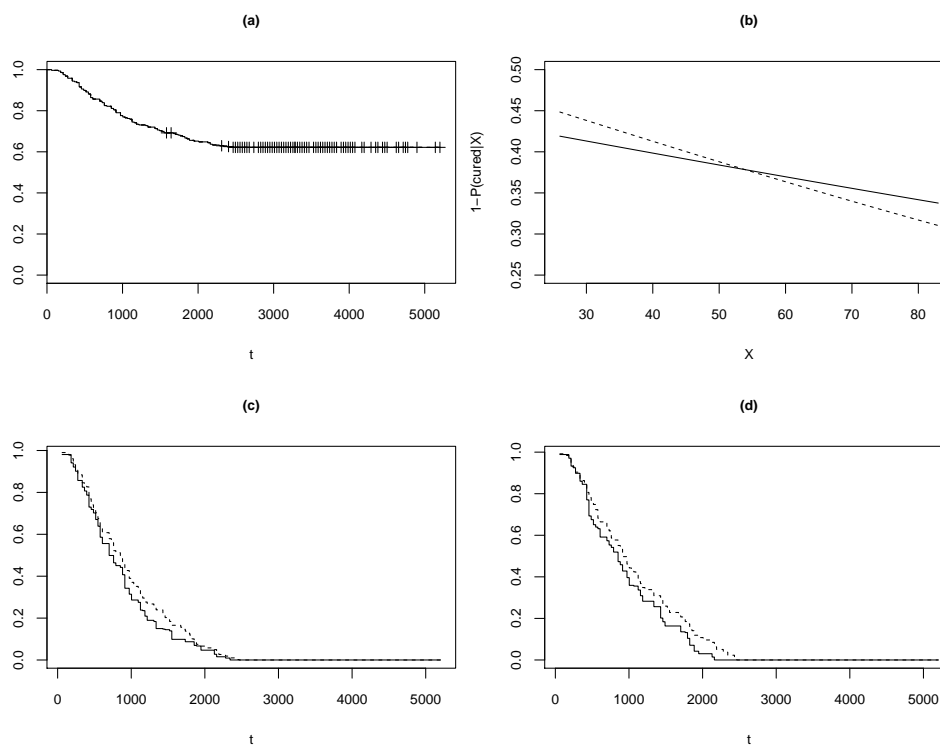
FIG 3. *Analysis of the breast cancer data : (a) KM estimator; (b) Graph of the proposed estimator of $\phi(x)$ (solid curve) and of the estimator based on the Cox model (dashed curve); (c) Estimation of $1 - F_{T,0}(\cdot|x)$ using the proposed estimator (solid curve) and using the estimator based on the Cox model (dashed curve) when $x = 48$; (d) Idem when $x = 60$.*

valid. This is also confirmed by Figure 3(c)-(d), which shows the estimation of the survival function of the uncured patients for $x = 48$ and $x = 60$ based on our estimation procedure and the one based on the Cox model. The figure shows that the two estimators are close for both values of x.

Next, we analyse data provided by the Medical Birth Registry of Norway (see http://folk.uio.no/borgan/abg-2008/data/data.html). The data set contains information on births in Norway since 1967, related to a total of 53,558 women. We are interested in the time between the birth of the first and the second child, for those mothers whose first child died within the first year ($n = 262$). The covariate of interest is age (X), which is the age of the mother at the birth of the first child. The age ranges from 16.8 to 29.8 years, with an average of 23.2 years. The cure rate is the fraction of women who gave birth only once. Figure 4(a) shows the Kaplan-Meier estimator, and suggests that a cure fraction is present.
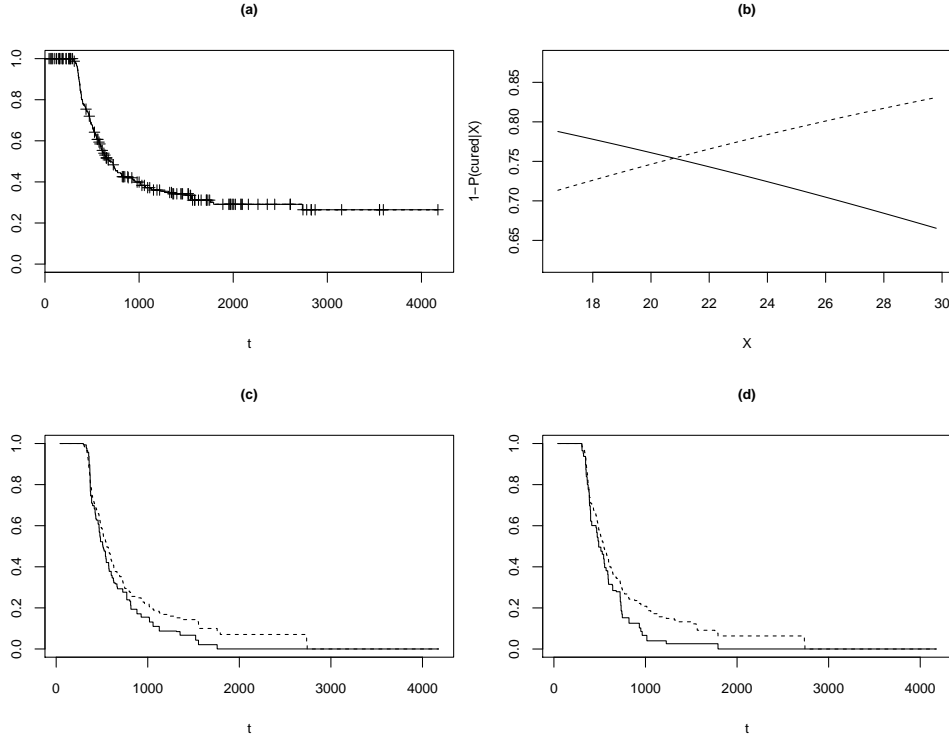
FIG 4. *Analysis of the second birth data : (a) KM estimator; (b) Graph of the proposed estimator of $\phi(x)$ (solid curve) and of the estimator based on the Cox model (dashed curve); (c) Estimation of $1 - F_{T,0}(\cdot|x)$ using the proposed estimator (solid curve) and using the estimator based on the Cox model (dashed curve) when $x = 21$; (d) Idem when $x = 25$.*

Again, we analyse these data using the approach proposed in this paper, and also using the Cox mixture cure model. The estimated intercept equals 2.119 using our model and 0.034 using the Cox model. The bootstrap confidence interval for the intercept is $(-1.306, 5.544)$ (the estimated standard deviation equals 1.747). The estimated slope equals -0.048 respectively 0.052 using the two models. For our estimation procedure the confidence interval is given by $(-0.195, 0.099)$ (with estimated standard deviation equal to 0.075). Figure 4(b) shows that the two estimators of the function $\phi(x)$ are quite different, and have opposite slopes. Moreover, the survival function $1 - F_{T,0}(\cdot|x)$ of the uncured patients is given in Figure 4(c)-(d) for $x = 21$ and $x = 25$. We see that the estimator based on the Cox model is quite different from ours, suggesting that the Cox model might not be valid for these data, although a formal test would need to confirm this. A possible test statistic could be the L_∞ or L_2 distance between our nonparametric estimator of $F_{T,0}$ and the

estimator under the Cox model or under any other parametric or semiparametric model. The development of this testing procedure is however beyond the scope of this paper. Also note that the estimator of the cure proportion is increasing under our model and decreasing under the Cox model. It seems however natural to believe that the probability of having no second child (so the cure proportion) is increasing with age, which is again an indication that the Cox model is not valid for these data.

8. Appendix: Proofs.

LEMMA 8.1. *Assume that conditions (2.2), (2.3) and (2.8) are met. Fix $x \in \mathcal{X}$ arbitrarily. Assume that $H_1(\cdot \mid x)$ puts a positive mass at the right endpoint of its right-bounded support. Then, $F_{T,0}$ is a proper distribution function (df) and $F_{T,0}((t, \infty) \mid x) = 0$ if $H_1((t, \infty) \mid x) = 0$. If $\widehat{H}_1(\cdot \mid x)$ puts a positive mass at the right endpoint of its right-bounded support, then $\widehat{F}_{T,0}$ is a proper df and $\widehat{F}_{T,0}((t, \infty) \mid x) = 0$ if $\widehat{H}_1((t, \infty) \mid x) = 0$.*

PROOF OF LEMMA 8.1. We show the statement for $F_{T,0}$, the case of $\widehat{F}_{T,0}$ follows by the same arguments. Let $\tau(x)$ be the right endpoint of the support of $H_1(\cdot \mid x)$. Then, $H_1(\{\tau(x)\} \mid x) > 0$ and hence condition (2.3) implies that $F_C([\tau(x), \infty) \mid x) = F_C((\tau(x), \infty) \mid x)$. Then we have

$$H([\tau(x), \infty) \mid x) - \mathbb{P}(T = \infty \mid x) F_C([\tau(x), \infty) \mid x)$$
$$= H([\tau(x), \infty) \mid x) - \mathbb{P}(T = \infty \mid x) F_C((\tau(x), \infty) \mid x)$$
$$= H([\tau(x), \infty) \mid x) - H((\tau(x), \infty) \mid x)$$
$$= H(\{\tau(x)\} \mid x) = H_1(\{\tau(x)\} \mid x).$$

By virtue of formula (2.12), $F_{T,0}((\tau(x), \infty) \mid x) = 0$ and hence $F_{T,0}((t, \infty) \mid x) = 0$ when $H_1((t, \infty) \mid x) = 0$. Moreover, $F_{T,0}$ is a proper distribution function by Theorem 11, case (a), of Gill & Johansen (1990). $\qquad \square$

PROOF OF PROPOSITION 2.1. Equations (2.6) and (2.5) show that $F_C(\cdot \mid x)$ is uniquely recovered from $F_{Y,\delta}(\cdot, \cdot \mid x)$. From condition (2.14) and the identity (3.3) it follows that β_0 is identifiable. Finally, the identities (2.11) and (2.12) guarantee that $F_{T,0}(\cdot \mid x)$ is identifiable. $\qquad \square$

The following lemma is needed for the proof of Theorem 4.1. The proof is provided in the Supplementary Material.

LEMMA 8.2. *Let conditions (4.2), (AC1) and (AC4) hold true. Then,*

$$\sup_{x \in \mathcal{X}} \sup_{t \in (-\infty, \tau(x)]} |\widehat{F}_C([t, \infty) \mid x) - F_C([t, \infty) \mid x)| = o_{\mathbb{P}}(1),$$

$$\sup_{x \in \mathcal{X}} \sup_{t \in (-\infty, \tau(x)]} |T_1(\widehat{H}_0, \widehat{H}_1)(t, x) - T_1(H_0, H_1)(t, x)| = o_{\mathbb{P}}(1),$$

where $T_1(H_0, H_1)(t, x) = H([t, \infty) \mid x) - \mathbb{P}(T = \infty \mid x) F_C([t, \infty) \mid x)$, and $T_1(\widehat{H}_0, \widehat{H}_1)$ is defined similarly, but with H_0, H_1 replaced by $\widehat{H}_0$ and $\widehat{H}_1$, respectively. Moreover,

$$\sup_{x \in \mathcal{X}} \sup_{t \in (-\infty, \tau(x)]} |\widehat{F}_{T,0}([t, \infty) \mid x) - F_{T,0}([t, \infty) \mid x)| = o_{\mathbb{P}}(1).$$

PROOF OF THEOREM 4.1. First, we show the consistency of $\widetilde{\beta}$ defined as the maximizer of $\mathcal{L}_n(\beta)$, see equation (4.3). Let

$$\log p(t, \delta, x; \beta) = \delta \log p_1(t, x; \beta) + (1 - \delta) \log p_0(t, x; \beta),$$

with $p_1(t, x; \beta) = \phi(x, \beta) F_{T,0}(dt \mid x) F_C((t, \infty) \mid x)$ and

$$p_0(t, x; \beta) = F_C(dt \mid x)[\phi(x, \beta) F_{T,0}([t, \infty) \mid x) + 1 - \phi(x, \beta)].$$

Following a common notational convention, see Gill (1994), here we treat dt not just as the length of a small time interval $[t, t + dt)$ but also as the name of the interval itself.

Let us notice that $\log p(t, \delta, x; \beta_0) = \delta \log H_1(dt \mid x) + (1 - \delta) \log H_0(dt \mid x)$ and the difference $\mathbb{E}[\log p(Y, \delta, X; \beta)] - \mathbb{E}[\log \mathcal{L}_n(\beta)]$ does not depend on β. Hence, a minimal condition for guaranteeing the consistency of the likelihood estimation approach is that β_0 is the maximizer of the limit likelihood criterion $\beta \mapsto \mathbb{E}[\log p(Y, \delta, X; \beta)]$, or equivalently of the log-likelihood ratio $\beta \mapsto \mathbb{E}[\log\{p(Y, \delta, X; \beta)/p(Y, \delta, X; \beta_0)\}]$. This is proved in the following. By the properties of the likelihood of a Bernoulli random variable given that $Y \in dt$ and $X = x$, for any $\beta \in B$ we have (with the convention $0/0 = 1$)

$$\mathbb{E}\left[\delta \log \frac{\phi(x, \beta) F_{T,0}(dt \mid x) F_C((t, \infty) \mid x)}{H_1(dt \mid x)}\right.$$
$$\left. + (1 - \delta) \log \frac{F_C(dt \mid x)[\phi(x, \beta) F_{T,0}([t, \infty) \mid x) + 1 - \phi(x, \beta)]}{H_0(dt \mid x)} \right| Y \in dt, X = x\right] \le 0.$$

Integrate with respect to Y and X and deduce that

$$\mathbb{E}[\log p(Y, \delta, X; \beta)] \le \mathbb{E}[\log p(Y, \delta, X; \beta_0)].$$

If there exists some $\beta \ne \beta_0$ such that the last inequality becomes an equality, then necessarily $p_1(Y, X; \beta) = 1$ almost surely. Then, by the identity $H_1(dt \mid x) = F_C([t, \infty) \mid x) F_T(dt \mid x)$ and condition (2.14), necessarily $\beta = \beta_0$.

Assumptions (AC3) and (AC4) and condition (2.14) guarantee that β_0 is a well-separated maximum of the continuous map $\beta \mapsto \mathbb{E}[\log \mathcal{L}_n(\beta)]$, $\beta \in B$.

Next, let us rewrite $\log \mathcal{L}_n(\beta) = \sum_{i=1}^n q(Y_i, \delta_i, X_i; H_0, H_1, \beta)$, where

$$q(t, d, x; H_0, H_1, \beta) = d \log \phi(x, \beta)$$
$$+ (1 - d) \log\{\phi(x, \beta) F_{T,0}([t, \infty) \,|\, x) + 1 - \phi(x, \beta)\}.$$

Given our assumptions, it is easy to check that for any $t \in (-\infty, \tau(x)]$, $d \in \{0, 1\}$, $x \in \mathcal{X}$,

$$\left| q(t, d, x; H_0, H_1, \beta) - q(t, d, x; H_0, H_1, \beta') \right| \le C \|\beta - \beta'\|^a, \quad \forall \beta, \beta' \in B,$$

with $a > 0$ from Assumption (AC3) and some constant C depending only on c_1 from Assumption (AC3) and the positive values $\inf_{x \in \mathcal{X}} H_1(\{\tau(x)\} \,|\, x)$ and $\inf_{x \in \mathcal{X}} H_0((\tau(x), \infty) \,|\, x)$. It follows that the class of functions $\{(t, d, x) \to q(t, d, x; H_0, H_1, \beta) : \beta \in B\}$ is Glivenko-Cantelli. Hence,

$$\sup_{\beta \in B} \left| n^{-1} \log \mathcal{L}_n(\beta) - \mathbb{E}\left[n^{-1} \log \mathcal{L}_n(\beta) \right] \right| = o_{\mathbb{P}}(1).$$

Finally, Lemma 8.2 guarantees that $\sup_{\beta \in B} n^{-1} \left| \log \widehat{L}_n(\beta) - \log \mathcal{L}_n(\beta) \right| = o_{\mathbb{P}}(1)$. Gathering facts, by standard asymptotic results, see Section 5.2 in van der Vaart (1998), we deduce that $\widehat{\beta} - \beta_0 = o_{\mathbb{P}}(1)$. $\qquad\square$

PROOF OF THEOREM 4.2. We show the asymptotic normality of our estimator by verifying the high-level conditions in Theorem 2 in Chen *et al.* (2003). First of all, for the consistency we refer to Section 4.1, whereas condition (2.1) in Chen *et al.* (2003) is satisfied by construction. Next, we compute the matrix $\nabla_\beta M(\beta_0, \eta_0)$. Note that for each $x \in \mathcal{X}$,

$$\mathbb{E}(\delta \mid X = x) = \phi(x, \beta_0) \int_{(-\infty, \tau(x)]} F_C((t, \infty) \mid x) F_{T,0}(dt \mid x),$$

$\mathbb{E}(1 - \delta \mid Y = t, X = x) = 1 - \phi(x, \beta_0) F_{T,0}((-\infty, t] \mid x)$ and

$$\int_{(-\infty, \tau(x)]} F_C((t, \infty) \mid x) F_{T,0}(dt \mid x) = \int_{(-\infty, \tau(x)]} F_{T,0}(-\infty, t] \mid x) F_C(dt \mid x).$$

Hence, we have

$$\mathbb{E}\left[\frac{\delta}{\phi(X, \beta_0)} - \frac{(1 - \delta) T(\eta_0)(Y, X)}{1 - \phi(X, \beta_0) T(\eta_0)(Y, X)} \mid X \right] = 0 \qquad \text{almost surely.}$$

It is now easy to see that $\nabla_\beta M(\beta_0, \eta_0)$ is given by (4.7). Hence, the matrix is negative definite, and thus condition (2.2) in Chen *et al.* (2003) is satisfied, as long as the matrix $-\mathbb{E}\left\{\nabla_\beta \phi(X, \beta_0) \nabla_\beta \phi(X, \beta_0)^\top\right\}$ has this property, which is the case thanks assumption (AN3). Thus, we do not need more assumptions than the ones usually considered in with observed binary outcomes.

Next, for condition (2.3), note that using the equation (4.5), we have

$$
\begin{aligned}
&\nabla_\eta M(\beta, \eta_0)[\eta - \eta_0] \\
&= -\lim_{\tau \to 0} \frac{1}{\tau} \mathbb{E}\Bigg[\frac{1 - \delta}{\{1 - \phi(X, \beta) T(\eta_0)(Y, X)\}\{1 - \phi(X, \beta) T(\eta_0 + \tau(\eta - \eta_0))(Y, X)\}} \\
&\qquad\qquad \times \{T(\eta_0 + \tau(\eta - \eta_0))(Y, X) - T(\eta_0)(Y, X)\} \nabla_\beta \phi(X, \beta) \Bigg].
\end{aligned}
$$

By equation (4.6) and the formulae in section (2.1), we can also write

$$
\begin{aligned}
T(\eta_0)(t, x) &= 1 - \prod_{-\infty < s \le t} \left\{ 1 - \frac{H_1(ds \mid x)}{H([s, \infty) \mid x) - \mathbb{P}(T = \infty \mid x) F_C([s, \infty) \mid x)} \right\} \\
&= 1 - \prod_{-\infty < s \le t} \left\{ 1 - \frac{\eta_{02}(ds, x)}{T_1(\eta_0)(s, x)} \right\}
\end{aligned}
$$

with $T_1(\eta)(t, x) = \eta_1(t, x) + \eta_2(t, x) - T_2(\eta)(x) T_3(\eta)(t, x)$,

$$
\begin{aligned}
T_2(\eta)(x) &= \prod_{t \in \mathbb{R}} \left\{ 1 - \frac{\eta_2(dt, x)}{\eta_1(t, x) + \eta_2(t, x)} \right\}, \\
T_3(\eta)(t, x) &= \prod_{-\infty < s \le t} \left\{ 1 - \frac{\eta_1(ds, x)}{\eta_1(s, x) + \eta_2(s, x)} \right\}.
\end{aligned}
$$

Using Duhamel's formula (see Gill and Johansen 1990) and the dominated convergence theorem, we have

$$
\begin{aligned}
\nabla_\eta M(\beta, \eta_0)[\eta - \eta_0] &= -\mathbb{E}\left[\frac{(1 - \delta) \nabla_\eta T(\eta_0)[\eta - \eta_0](Y, X)}{\{1 - \phi(X, \beta) T(\eta_0)(Y, X)\}^2} \nabla_\beta \phi(X, \beta) \right] \\
&= \mathbb{E}\Bigg[\int_{(-\infty, Y)} \frac{1 - T(\eta_0)(Y, X)}{T_1(\eta_0)(u, X) - \eta_{02}(\{u\}, X)} \\
&\quad \times \left\{ (\eta_2 - \eta_{02})(du, X) - \frac{\nabla_\eta T_1(\eta_0)[\eta - \eta_0](u, X)}{T_1(\eta_0)(u, X)} \eta_{02}(du, X) \right\} \\
&\qquad\qquad \times \frac{1 - \delta}{\{1 - \phi(X, \beta) T(\eta_0)(Y, X)\}^2} \nabla_\beta \phi(X, \beta) \Bigg],
\end{aligned}
$$

where

(8.2) $\quad \nabla_\eta T_1(\eta_0)[\eta - \eta_0](y,x) = (\eta_1 - \eta_{01})(y,x) + (\eta_2 - \eta_{02})(y,x)$
$\quad\quad - \nabla_\eta T_2(\eta_0)[\eta - \eta_0](x)T_3(\eta_0)(y,x) - T_2(\eta_0)(x)\nabla_\eta T_3(\eta_0)[\eta - \eta_0](y,x),$

and the expressions of $\nabla_\eta T_2(\eta_0)[\eta - \eta_0](x)$ and $\nabla_\eta T_3(\eta_0)[\eta - \eta_0](y,x)$ are provided in the Supplementary Material.

Note that the map $y \mapsto T_1(\eta_0)(y,x)$ is decreasing on $(-\infty, \tau(x)]$. Moreover, by condition (4.2), $\inf_{x \in \mathcal{X}} T_1(\eta_0)(\tau(x),x) > 0$. Finally, let us note that by construction for any $y \in (-\infty, \tau(x))$, $H(\{y\} \mid x) = H_0(\{y\} \mid x) + H_1(\{y\} \mid x) \geq \mathbb{P}(T = \infty \mid x)F_C(\{y\} \mid x) + H_1(\{y\} \mid x)$, and thus

$$H([y,\infty) \mid x) - \mathbb{P}(T = \infty \mid x)F_C([y,\infty) \mid x) - H_1(\{y\} \mid x)$$
$$\geq H((y,\infty) \mid x) - \mathbb{P}(T = \infty \mid x)F_C((y,\infty) \mid x).$$

Then, $\inf_{x \in \mathcal{X}} \inf_{y \in (-\infty, \tau(x))}[T_1(\eta_0)(y,x) - H_1(\{y\} \mid x)] > 0$. Hence, all denominators in $\nabla_\eta M(\beta, \eta_0)[\eta - \eta_0]$ are bounded away from zero. By tedious but rather elementary Taylor expansions, it now follows that $\nabla_\eta M(\beta, \eta_0)[\eta - \eta_0]$ satisfies the second property in condition (2.3) of Theorem 2 in Chen *et al.* (2003). Similarly, by decomposing $T_j(\eta) - T_j(\eta_0) - \nabla_\eta T_j(\eta_0)[\eta - \eta_0]$ $(j = 1, 2, 3)$ using Taylor-type arguments (in η), the first property in condition (2.3) is easily seen to hold true.

Next, conditions (2.4) and (2.6) of Theorem 2 in Chen *et al.* (2003) are satisfied thanks to Assumption (AN4) and because it follows from the above calculations of $\nabla_\eta T_j(\eta_0)[\eta - \eta_0]$ $(j = 1, 2, 3)$ that

(8.3)
$$\nabla_\eta M(\beta_0, \eta_0)[\eta - \eta_0] = \sum_{k=0}^{1} \mathbb{E}\left[\psi_{1k}(Y,X)\int_{(-\infty,Y)} \psi_{2k}(u,X)d\left((\eta_k - \eta_{0k})(u,X)\right)\right]$$
$$+ \sum_{k,\ell \in \{0,1\}} \mathbb{E}\left[\psi_{3k}(Y,X)\int_{(-\infty,Y)} \psi_{4k}(u,X)\psi_{5k}\left((\eta_k - \eta_{0k})(u,X)\right)dH_\ell(u \mid X)\right]$$

for certain measurable functions ψ_{jk} $(j = 1, \ldots, 5; k = 0, 1)$.

It remains to verify condition (2.5) of Theorem 2 in Chen *et al.*. Note that

$$|m(t,\delta,x;\beta_2,\eta_2) - m(t,\delta,x;\beta_1,\eta_1)| \leq C_1(t,\delta,x)\|\beta_2 - \beta_1\| + C_2(t,\delta,x)\|\eta_2 - \eta_1\|_{\mathcal{H}}$$

for some C_j satisfying $\mathbb{E}[C_j^2(Y,\delta,X)] < \infty$ $(j = 1, 2)$, and hence (2.5) follows from our assumption (AN5) and Theorem 3 in Chen *et al.* (2003). $\quad\quad\square$

PROOF OF THEOREM 4.3. To prove this theorem we will check the conditions of Theorem B in Chen *et al.* (2003), which gives high level conditions under which the naive bootstrap is consistent. The only difference between their setting and our setting is that we are proving bootstrap consistency in $\mathbb{P}$-probability, whereas their result holds true a.s. $[\mathbb{P}]$. As a consequence, in their high level conditions we can replace all a.s. $[\mathbb{P}]$ statements by the corresponding statements in $\mathbb{P}$-probability.

First of all, it follows from assumption (AN2) that condition (2.2) in Chen *et al.* (2003) holds with η_0 replaced by any η in a neighbourhood of η_0, and from the proof of Theorem 4.2 it follows that the same holds true for condition (2.3). Next, conditions (2.4B) and (2.6B) in Chen *et al.* follow from the fact that we impose that assumption (AN4) continues to hold true if we replace $\widehat{H}_k - H_k$ by $\widehat{H}_k^* - \widehat{H}_k$ ($k = 0, 1$). It remains to verify condition (2.5'B) in Chen *et al.* This follows from Theorem 3 in Chen *et al.*, whose conditions have been verified already for our Theorem 4.2. $\square$

SUPPLEMENT

The supplement contains five appendices, and is organized as follows. Appendix A contains the proof of the i.i.d. representation (4.9), Appendix B shows the calculation of the efficient score function, Appendix C contains details on the maximization procedure used in the simulations, Appendix D collects additional simulation results, and Appendix E collects proofs of technical lemmas and equations.

REFERENCES

[1] AKRITAS, M.G. & VAN KEILEGOM, I. (2001). Nonparametric estimation of the residual distribution. *Scand. J. Statist.* **28**, 549–568.

[2] AMICO, M., LEGRAND, C. & VAN KEILEGOM, I. (2018). The single-index/Cox mixture cure model *Biometrics* (to appear).

[3] AMICO, M. & VAN KEILEGOM, I. (2018). Cure models in survival analysis. *Ann. Rev. Statist. Applic.*, **5**, 311-342.

[4] BOAG, J.W. (1949). Maximum likelihood estimates of the proportion of patients cured by cancer therapy. *J. Roy. Statist. Soc. - Series B* **11**, 15–53.

[5] CHEN, X., LINTON, O. & VAN KEILEGOM, I. (2003). Estimation of semiparametric models when the criterion function is not smooth. *Econometrica* **71**, 1591–1608.

[6] FANG, H.B., LI, G. & SUN, J. (2005). Maximum likelihood estimation in a semiparametric logistic/proportional-hazards mixture model. *Scand. J. Statist.* **32**, 59–75.

[7] FAREWELL, V.T. (1982). The use of mixture models for the analysis of survival data with long-term survivors. *Biometrics* **38**, 1041–1046.

[8] GILL, R.D. (1994). *Lectures on survival analysis.* Lectures on probability theory: Ecole d'été de probabilités de Saint-Flour XXII. Lecture notes in mathematics 1581. Springer.

[9] GILL, R.D. & JOHANSEN, S. (1990). A survey of product-integration with a view toward application in survival analysis. *Ann. Statist.* **18**, 1501–1555.

[10] KUK, A.Y.C. & CHEN, C.-H. (1992). A mixture model combining logistic regression with proportional hazards regression. *Biometrika* **79**, 531–541.

[11] LI, Q., LIN, J. & RACINE, J.S. (2013). Optimal bandwidth selection for nonparametric conditional distribution and quantile functions. *J. Buss. Econ. Statist.* **31**, 57–65.

[12] LÓPEZ-CHEDA, A., CAO, R., JÁCOME M.A. & VAN KEILEGOM, I. (2017). Nonparametric incidence estimation and bootstrap bandwidth selection in mixture cure models. *Comput. Statist. Data Anal.* **105**, 144–165.

[13] LOPEZ, O. (2011). Nonparametric estimation of the multivariate distribution function in a censored regression model with applications. *Commun. Stat. - Theory Meth.* **40**, 2639–2660.

[14] LU, W. (2008). Maximum likelihood estimation in the proportional hazards cure model. *Ann. Inst. Stat. Math.* **60**, 545–574.

[15] MEEKER, W.Q. (1987). Limited failure population life tests: Application to integrated circuit reliability. *Technometrics* **29**, 51–65.

[16] OTHUS, M., LI, Y. & TIWARI, R.C. (2009). A class of semiparametric mixture cure survival models with dependent censoring. *J. Amer. Statist. Assoc.* **104**, 1241–1250.

[17] PENG, Y. & TAYLOR, J.M.G. (2014). Cure models. In: Klein, J., van Houwelingen, H., Ibrahim, J. G., and Scheike, T. H., editors, *Handbook of Survival Analysis*, Handbooks of Modern Statistical Methods series, chapter 6, pages 113-134. Chapman & Hall, Boca Raton, FL, USA.

[18] SCHMIDT, P. & WITTE, A.D. (1989). Predicting criminal recidivism using split population survival time models. *J. Econometrics* **40**, 141–159.

[19] STONE, C.J. (1977). Consistent Nonparametric Regression. *Ann. Statist.* **5**, 595–620.

[20] SY, J.P. & TAYLOR, J.M.G. (2000). Estimation in a Cox proportional hazards cure model. *Biometrics* **56**, 227–236.

[21] TAYLOR, J.M.G. (1995). Semi-parametric estimation in failure time mixture models. *Biometrics* **51**, 899–907.

[22] VAN DER VAART, A.D. (1998). *Asymptotic Statistics.* Cambridge University Press.

[23] WANG, Y., KLIJN, J.G.M., SIEUWERTS, A.M., LOOK, M.P., YANG, F., TALANTOV, D., TIMMERMANS, M., MEIJET-VAN GELDER, M.E.M., YU, J., JATKOE, T., BERNS, E.M.J.J., ATKINS, D. & FOEKENS, J.A. (2005). Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. *The Lancet* **365**, 671–679.

[24] XU, J. & PENG, Y. (2014). Nonparametric cure rate estimation with covariates. *Canad. J. Statist.* **42**, 1–17.

CREST UMR 9194
ENSAI
CAMPUS DE KER-LANN
35172 BRUZ, FRANCE
E-MAIL: patilea@ensai.fr

ORSTAT, KU LEUVEN
NAAMSESTRAAT 69
3000 LEUVEN, BELGIUM
E-MAIL: ingrid.vankeilegom@kuleuven.be