Identification and Estimation in Mixture Cure Models Without Sufficient Follow-Up

Sander G.H. van Beek *

This version: November 12, 2026
Click here for the latest version

Abstract

Duration data often involve a fraction of individuals who never experience the event of interest. Without assumptions on the tail of the duration distribution, the probability of experiencing the event and the duration distribution of those who do are not separately identified. Identification is typically achieved by assuming that follow-up outlasts all finite durations, but this is often implausible in economic applications. I show that if the two depend on the covariates through linear indices that are not proportional, the indeterminacy reduces to a single constant with identified bounds, without any tail assumption. I propose estimators for the indices, for both functions up to this constant, and for its bounds. Monte Carlo simulations show that the proposed estimators perform well when follow-up is short, whereas estimators that assume long follow-up remain biased as the sample size grows. I apply the method to Swedish register data on unemployment spells of low-skilled workers and bound the share of these workers who never return to regular work.

*Department of Economics, University of Groningen.

Keywords: cure models, multiple index models, sufficient follow-up, split population models, average derivatives, survival analysis

Acknowledgements: I would like to thank Gerard van den Berg and Tom Boot for their helpful comments and discussions. I am grateful to IFAU-Uppsala, in particular to Linus Liljeberg, Jörgen Moen, Martin Söderström and Johan Vikström, for access to the data. Parts of this paper were written while I was visiting Princeton University. I am grateful for the hospitality during my visit and to seminar participants there for their helpful comments and suggestions. In particular, I would like to thank Bo Honoré.

1 Introduction

In many duration settings a fraction of the population never experiences the event of interest. In such settings, whether people experience the event at all is often as important as when they do. For instance, a labor market policy is judged not only by whether it brings people back to work sooner, but by whether it brings them back at all. Cure models address such questions by explicitly splitting the population into those who will eventually experience the event and those who never will.

The most popular cure model is the mixture cure model, which writes the survival function of the duration Y given covariates X as

$$S_Y(t|X) = \pi(X) + (1 - \pi(X))\tilde{S}_Y(t|X), \tag{1}$$

where $1 - \pi(X)$ is the probability of being susceptible and $\tilde{S}_Y$ is the survival function of the susceptible. In economics this is also known as a split population model (Schmidt and Witte, 1989). The components $1 - \pi(X)$ and $\tilde{S}_Y(t|X)$ are referred to as incidence and latency, respectively, and the goal is to identify them separately.

The cure fraction $\pi(x)$ describes the behavior of $S_Y(t|x)$ as $t \to \infty$. In practice, however, durations are only observed over a finite window, as they are right-censored at censoring times that are typically bounded. The central difficulty is then separating a genuine cure fraction from long but finite durations, as any spell censored beyond the last observed event is consistent with both. Point identification of the incidence and latency therefore requires a claim about the time-dependence of the latency distribution.

In the semi- and nonparametric literature, the standard such claim is the assumption of sufficient follow-up, which states that no susceptible duration can exceed the largest possible censoring time (Maller and Zhou, 1992; Laska and Meisner, 1992). The cure fraction then equals the height of the plateau of the survival function, i.e. $\pi(x) = S_Y(c^*|x)$. With few covariates, the incidence can be estimated directly from a conditional Kaplan-Meier estimator (Xu and Peng, 2014; López-Cheda et al., 2017). With more covariates, structure is typically imposed on the way they enter the incidence and latency to avoid the curse of dimensionality (Kuk and Chen, 1992; Taylor, 1995; Peng and Dear, 2000; Sy and Taylor, 2000; Amico, Van Keilegom and Legrand, 2019; Patilea and Van Keilegom, 2020).

In practice, the observation window often closes before all susceptible spells have ended (Escobar-Bach and Van Keilegom, 2019; Tai et al., 2005). Spells that

are still ongoing at that point are then counted as cured, so the cure fraction is overestimated. Tests for sufficient follow-up exist, but they themselves rest on assumptions about the tail, such as a known parametric family or a known domain of attraction (Maller and Zhou, 1996; Xie, Escobar-Bach and Van Keilegom, 2024).

Without sufficient follow-up, the cure fraction can still be approximated if the survival function can be extrapolated beyond the observation window. To this end, the latency tail is assumed to lie in a known class from extreme value theory, such as Pareto-type or Gumbel-type tails (Escobar-Bach and Van Keilegom, 2019, 2023; Escobar-Bach et al., 2022; Beirlant, Bladt and Van Keilegom, 2026). Within such a class, the tail is governed by a few parameters, which can be estimated from the longest observed durations. While this reduces the bias when sufficient follow-up fails, the resulting estimators remain inconsistent for a fixed observation window.

In this paper, I leave the latency beyond the observation window unrestricted, and use the covariates to separate the incidence from the latency. Specifically, I assume that the incidence and the latency each depend on the covariates through a linear index:

$$S_Y(t|X) = \pi(X'\theta) + (1 - \pi(X'\theta))\tilde{S}_Y(t|X'\beta). \tag{2}$$

Apart from this index structure, I make no parametric assumptions. Instead, I impose the restriction that θ and β are not proportional. This is a weak version of an exclusion restriction: rather than excluding a covariate from one of the indices, it only requires the covariates to affect the indices in different proportions.

Under this non-proportionality, observationally equivalent pairs of incidence and latency can differ only by a constant scale factor, as a rescaling that varies with the covariates would break one of the index structures. I first show that both index vectors are identified. The covariates can then move one component while holding the other fixed, which identifies incidence and latency up to the unknown scale factor. Only the levels of the incidence and latency functions remain unknown. I identify bounds on these levels that exploit the fact that incidence and latency are both probabilities and therefore cannot exceed one.

Identification through the covariates is a familiar idea in duration models with unobserved heterogeneity. For instance, covariate variation separates unobserved heterogeneity from duration dependence in the mixed proportional hazards model, which cannot be done from the durations alone (Elbers and Ridder, 1982; Heckman and Singer, 1984; Ridder, 1990). Similarly, dependent competing risk models are identified when the covariates enter the risks in a restricted way, which

are not identified otherwise (Tsiatis, 1975; Heckman and Honoré, 1989; Abbring and van den Berg, 2003a). The same idea underlies the identification of treatment effects in duration models and of hitting time models with heterogeneity (Abbring and van den Berg, 2003b; Abbring, 2012). This paper carries this principle over to cure models, where the cure status itself can be seen as a form of unobserved heterogeneity.

The paper also connects to the literature on multiple index models, which summarize the covariates through a few linear combinations and an unknown link function. In such models, the index vectors are in general identified only up to an invertible linear transformation, as any such transformation can be absorbed into the unknown link (Ichimura and Lee, 1991; Li, 1991; Hristache et al., 2001; Xia et al., 2002; Donkers and Schafghans, 2008). The mixture model ties the link to a specific form, which allows both index vectors to be identified.

I propose estimators for the index vectors based on weighted average derivatives in the spirit of Powell, Stock and Stoker (1989), and estimate the remaining objects by local linear regression on the estimated indices. Upon normalization, the index vectors are $\sqrt{n}$-consistent and asymptotically normal, and the shapes of the incidence and the latency converge at the optimal nonparametric rates. The endpoints of the interval for κ converge at the same rate, which yields consistent bounds on both the incidence and the latency.

In the main analysis, I focus on the case where all covariates are continuous. As an extension, I show how discrete covariates can be included, following Horowitz and Härdle (1996), as long as at least two continuous covariates remain. I also show that the identification argument carries over to the Box-Cox family of Yin and Ibrahim (2005). This family nests both the mixture model and the promotion time model, the leading alternative to the mixture model in the literature.

I use Monte Carlo simulations to assess the small-sample performance of the estimators, with the estimator of Amico, Van Keilegom and Legrand (2019) as a benchmark. In the simulated data some susceptible spells are still ongoing when follow-up ends, whereas the benchmark imposes sufficient follow-up. In general, the new estimators perform well with moderate sample sizes. Not surprisingly, the precision of the estimated cure index is strongly affected by how close the two index vectors are to proportionality. However, the estimators are centered on the true values throughout all experiments, whereas the benchmark estimate of the cure index is biased, and this bias does not shrink as the sample size grows.

I apply the method to Swedish administrative register data on unemployment durations of low-skilled workers. Among young, elderly and low-skilled individuals, median durations of joblessness can reach seven years, and many have ac-

commodated to a life without regular work and may feel excluded from society (Crpon and van den Berg, 2016). Durations of this length make it implausible that follow-up outlasts all finite durations. The registers record all unemployment spells in the country, together with the subsequent employment of the workers involved. I use them to bound the share of low-skilled unemployed workers who eventually return to regular work.

The remainder of the paper is structured as follows. Section 2 sets up the model and states the main assumptions. Section 3 establishes identification. Section 4 proposes the estimators and derives their asymptotic properties. Section 5 presents the results from Monte Carlo simulations. Section 6 discusses extensions, and Section 7 presents the empirical application. Section 8 concludes.

2 Setup and main assumptions

Let Y denote the time until an event of interest and X a vector of continuous covariates. The population consists of two types of individuals. Cured individuals never experience the event and therefore have an infinite duration; susceptible individuals have a finite duration. The survival function of Y is the mixture

$$S_Y(t|X) = \pi(X'\theta) + (1 - \pi(X'\theta))\tilde{S}_Y(t|X'\beta), \tag{3}$$

where $\pi(X'\theta)$ is the cure probability and $\tilde{S}_Y(t|X'\beta)$ is the survival function of the susceptible population. Durations are right censored by C, whose survival function $S_C(t|X)$ is proper. Cured individuals are therefore always censored.

The functions π, $\tilde{S}_Y$, and S_C are unknown, as are the parameter vectors β and θ. The observed sample consists of $\{T, 1(Y \leq C), X\}$, where $T = \min(Y, C)$. The objects of interest are the parameter vectors, the function S_Y, the cure probability π, and the latency distribution $\tilde{F}_Y = 1 - \tilde{S}_Y$. Their identification rests on several assumptions, of which I introduce the main ones here and the rest where they are needed.

As stated in the introduction, identification comes from the covariates rather than from the tail of the data. To this end, I impose the following identifying assumption.

Assumption 1 (**Non-proportionality**). *It holds that* $\theta \not\propto \beta$.

To see why this assumption is needed, consider the more general model

$$S_Y(t|X) = \pi(X) + (1 - \pi(X))\tilde{S}_Y(t|X), \tag{4}$$

and suppose follow-up ends at some finite time c^*. Then the pair $(1-\pi(x), \tilde{F}_Y(t|x))$ is observationally equivalent to $(\kappa^*(x)(1-\pi(x)), \tilde{F}_Y(t|x)/\kappa^*(x))$ for any positive function κ^* that keeps both bounded by 1. Consequently, the cure probability is only subject to the trivial bound $0 \leq \pi(x) \leq S_Y(c^*|x)$. However, if incidence and latency depend on non-proportional indices, any κ^* that preserves both index structures must be constant. If $\theta \propto \beta$, any function of $x'\beta$ does, and the trivial bound returns. Proportionality of θ and β is testable, as shown in Section 4.3.

The remaining assumptions are stated below, and further conditions are introduced where they are needed. First, the average derivative arguments require continuous covariates and a region of the support on which X is well behaved.

Assumption 2 (**Continuous covariates**). *The distribution of X is absolutely continuous with respect to Lebesgue measure. The support $\mathcal{X}$ of X is compact. There exists a compact rectangle $\mathcal{X}_0$ in the interior of $\mathcal{X}$ such that the density f_X of X is positive and continuously differentiable on an open set $\mathcal{X}_0^+$ containing $\mathcal{X}_0$.*

Section 6.1 discusses how this may be weakened to allow for discrete covariates. For readability, here and below I rule out degenerate cases, such as an empty $\mathcal{X}_0$, without further comment.

Next, I assume there exists a point in time at which the probabilities of having failed, of not yet being censored, and of being susceptible are all positive.

Assumption 3 (**Informative follow-up**). *There exist $\bar{\tau} > 0$ and $\varepsilon > 0$ such that $\tilde{F}_Y(\bar{\tau}|x'\beta) \geq \varepsilon$, $S_C(\bar{\tau}|x) \geq \varepsilon$, and $1 - \pi(x'\theta) \geq \varepsilon$ for all $x \in \mathcal{X}_0^+$.*

The restriction on the latency is much weaker than under sufficient follow-up, which requires $\tilde{F}_Y(\bar{\tau}|x'\beta) = 1$ for all $x \in \mathcal{X}_0^+$. Here, only some susceptible spells need to end within the window, and the latency distribution beyond $\bar{\tau}$ is left unrestricted.

In addition, censoring must be independent of the durations given the covariates. Without this, the model is not identified, by the same argument as in Tsiatis (1975).

Assumption 4 (**Conditional random censoring**). *It holds that $Y \perp\!\!\!\perp C|X$.*

I require π to be smooth, both π and $\tilde{F}_Y(t|v)$ to vary with the indices, and $\tilde{F}_Y(t|v)$ not to factor into a function of t times a function of v. To state this, let $\mathcal{U} = \{x'\theta : x \in \mathcal{X}\}$ and $\mathcal{V} = \{x'\beta : x \in \mathcal{X}\}$. Let $\mathcal{U}_0$ and $\mathcal{V}_0$ be closed intervals such that $\mathcal{U}_0 \times \mathcal{V}_0$ is contained in the interior of $\{(x'\theta, x'\beta) : x \in \mathcal{X}_0\}$.

Assumption 5 (**Covariate effects**). *The function π is twice continuously differentiable on $\mathcal{U}$ and not constant on $\mathcal{U}_0$. The function $\tilde{F}_Y(t|v)/\tilde{F}_Y(\bar{\tau}|v)$ is not constant in v on $\mathcal{V}_0$ for some $t \in (0, \bar{\tau})$.*

Finally, I assume the durations and censoring times are continuously distributed.

Assumption 6 (**Continuous outcomes**). *The distribution of the susceptible duration given $X = x$ and the distribution of C given $X = x$ are absolutely continuous with respect to Lebesgue measure for all $x \in \mathcal{X}$. The derivatives $\partial \tilde{F}_Y(t|v)/\partial t$, $\partial \tilde{F}_Y(t|v)/\partial v$ and $\partial^2 \tilde{F}_Y(t|v)/\partial t \partial v$ exist and are continuous on $[0, \bar{\tau}] \times \mathcal{V}$.*

3 Identification strategy

This section establishes identification, under slightly stronger assumptions than strictly necessary so as to obtain identifying equations with natural plug-in counterparts. All conditional expectations in this section are conditional on $X \in \mathcal{X}_0$ unless otherwise stated; the conditioning is suppressed for readability.

The covariates affect the observed data through three channels: the censoring distribution, the cure probability and the latency distribution. Identification requires separating these. The censoring distribution is the easiest to recover; the cure fraction plays no role for censoring, so S_C satisfies the usual identity

$$S_C(t|x) = \exp\left(- \int_0^t \frac{\frac{\partial}{\partial s} P(T \leq s, Y > C | X = x)}{P(T \geq s | X = x)} \mathrm{d}s \right). \tag{5}$$

The index structure separates the remaining two channels. Decompose $\theta = \theta_\perp + \alpha\beta$ such that $\theta_\perp \perp \beta$, and define

$$W = \frac{1(Y \leq C) 1(T \leq \bar{\tau})}{S_C(T|X)}. \tag{6}$$

Both β and $\theta_\perp$ can be recovered from conditional moments of W. Let $\psi : \mathbb{R} \to \mathbb{R}$ be a known function bounded and non-constant on $(0, \bar{\tau})$, and define

$$m_\psi(x) \equiv E[W\psi(T)|X = x] = (1 - \pi(x'\theta)) \int_0^{\bar{\tau}} \psi(s) \tilde{f}_Y(s|x'\beta) \mathrm{d}s \tag{7}$$

and

$$m_1(x) \equiv E[W|X = x] = (1 - \pi(x'\theta)) \tilde{F}_Y(\bar{\tau}|x'\beta). \tag{8}$$

As both functions depend on θ only through the factor $1 - \pi(x'\theta)$, their ratio

$$\frac{m_\psi(x)}{m_1(x)} = \frac{\int_0^{\bar{\tau}} \psi(s) \tilde{f}_Y(s|x'\beta) \mathrm{d}s}{\tilde{F}_Y(\bar{\tau}|x'\beta)} \tag{9}$$

is a function of $x'\beta$ alone, and hence $\beta \propto \frac{\partial}{\partial x}[m_\psi(x)/m_1(x)]$. The function m_1 itself depends on both indices, so its gradient

$$\frac{\partial}{\partial x} m_1(x) = -\pi'(x'\theta)\tilde{F}_Y(\bar{\tau}|x'\beta)\theta + (1-\pi(x'\theta))\frac{\partial}{\partial v}\tilde{F}_Y(\bar{\tau}|v)\Big|_{v=x'\beta}\beta \tag{10}$$

is a linear combination of θ and β. Projecting out β gives

$$\theta_\perp \propto \left(I - \frac{\beta\beta'}{\|\beta\|^2}\right)\frac{\partial}{\partial x} m_1(x), \tag{11}$$

where I denotes the identity matrix.

For estimation, it is preferable to integrate across multiple values of x rather than holding it fixed. Equation (9) implies that

$$\beta \propto \int \omega^*(x)\frac{\partial}{\partial x}\left[\frac{m_\psi(x)}{m_1(x)}\right]\mathrm{d}x \tag{12}$$

for an appropriately chosen weight function ω^*. I take $\omega^*(x) = \omega(x)f_X(x)^2 m_1(x)^2$, where ω is a nonnegative weight function with support $\mathcal{X}_0$. This gives

$$\beta \propto \int \omega(x) f_X(x)^2 \left(m_1(x)\frac{\partial}{\partial x} m_\psi(x) - m_\psi(x)\frac{\partial}{\partial x} m_1(x)\right)\mathrm{d}x. \tag{13}$$

This expression is convenient as it avoids random denominators in the corresponding plug-in estimation formula. Similarly, equation (11) implies

$$\theta_\perp \propto \left(I - \frac{\beta\beta'}{\|\beta\|^2}\right)\int \omega(x) f_X(x)^2 \frac{\partial}{\partial x} m_1(x)\mathrm{d}x. \tag{14}$$

Both integrals are nonzero under a non-degeneracy condition stated below.

The scales of β and θ are arbitrary, as any rescaling can be absorbed into $\tilde{F}_Y$ and π, respectively. However, the scalar α is meaningful only once these scales are fixed. Therefore, I fix both scales by replacing proportionality with equality in equations (13) and (14), which is without loss of generality.

Assumption 7 (**Normalization**). *It holds that*

$$\beta = \int \omega(x) f_X(x)^2 \left(m_1(x)\frac{\partial}{\partial x} m_\psi(x) - m_\psi(x)\frac{\partial}{\partial x} m_1(x)\right)\mathrm{d}x \tag{15}$$

and

$$\theta_\perp = \left(I - \frac{\beta\beta'}{\|\beta\|^2}\right)\int \omega(x) f_X(x)^2 \frac{\partial}{\partial x} m_1(x)\mathrm{d}x. \tag{16}$$

It remains to identify α. Let $U_\perp = X'\theta_\perp$ and $V = X'\beta$, and define

$$g(u_\perp, v) \equiv \log E[W|U_\perp = u_\perp, V = v] = \log(1-\pi(u_\perp+\alpha v))+\log \tilde{F}_Y(\bar{\tau}|v). \quad (17)$$

Write g_1 for the derivative of g with respect to its first argument, and g_{11} and g_{12} for the derivatives of g_1 with respect to its first and second argument, respectively. The latency term does not depend on $u_\perp$, so

$$g_1(u_\perp, v) = -\frac{\pi'(u_\perp + \alpha v)}{1 - \pi(u_\perp + \alpha v)} \quad (18)$$

is a function of $u_\perp + \alpha v$ alone. Hence $g_{12}(u_\perp, v) = \alpha g_{11}(u_\perp, v)$. Let $\rho : \mathbb{R}^2 \to \mathbb{R}$ be a known twice continuously differentiable weight function that is zero outside a compact subset of the interior of $\{(x'\theta_\perp, x'\beta) : x \in \mathcal{X}_0\}$. Then

$$\int \rho(u_\perp, v) g_{12}(u_\perp, v) \mathrm{d}u_\perp \mathrm{d}v = \alpha \int \rho(u_\perp, v) g_{11}(u_\perp, v) \mathrm{d}u_\perp \mathrm{d}v. \quad (19)$$

Integrate by parts twice to obtain

$$\alpha = \frac{\int \frac{\partial^2}{\partial u_\perp \partial v} \rho(u_\perp, v) g(u_\perp, v) \mathrm{d}u_\perp \mathrm{d}v}{\int \frac{\partial^2}{\partial u_\perp^2} \rho(u_\perp, v) g(u_\perp, v) \mathrm{d}u_\perp \mathrm{d}v}. \quad (20)$$

The following assumption ensures that the integrals defining β and $\theta_\perp$ and the denominator of α are nonzero.

Assumption 8 (**Non-degeneracy**). *Define*

$$H(v) = \tilde{F}_Y(\bar{\tau}|v)^{-1} \int_0^{\bar{\tau}} \psi(s) \tilde{f}_Y(s|v) \mathrm{d}s. \quad (21)$$

It holds that

$$\int \omega(x) f_X(x)^2 (1 - \pi(x'\theta))^2 \tilde{F}_Y(\bar{\tau}|x'\beta)^2 H'(x'\beta) \mathrm{d}x \neq 0, \quad (22)$$

$$\int \omega(x) f_X(x)^2 \pi'(x'\theta) \tilde{F}_Y(\bar{\tau}|x'\beta) \mathrm{d}x \neq 0, \quad (23)$$

and

$$\int \rho(u - \alpha v, v) \frac{\partial^2}{\partial u^2} \log(1 - \pi(u)) \mathrm{d}u \mathrm{d}v \neq 0. \quad (24)$$

This assumption is weak. The first two conditions fail only if H' or π' averages exactly to zero, as the remaining factors in each integrand are positive. They hold for every weight function if H and π are monotone, and a weight function satisfying both always exists. The third condition fails only if the curvature of $\log(1-\pi)$ averages to zero, and holds whenever $\log(1-\pi)$ is strictly concave or strictly convex, as for the logit and probit links. It fails for every weight function only if $\log(1-\pi)$ is affine, which is testable.

Once the indices are identified, the improper survival function S_Y is point-identified on $\mathcal{X}_0$, and the functions $1-\pi$ and $\tilde{F}_Y$ are identified up to scale on $\mathcal{U}_0$ and $\mathcal{V}_0$. To see this, define

$$G(t, u_\perp, v) \equiv E[W1(T \leq t)|U_\perp = u_\perp, V = v] = (1 - \pi(u_\perp + \alpha v))\tilde{F}_Y(t|v) \quad (25)$$

for $(u_\perp, v) \in \{(x'\theta_\perp, x'\beta) : x \in \mathcal{X}_0\}$. Since $1 - S_Y(t|x) = (1 - \pi(x'\theta))\tilde{F}_Y(t|x'\beta)$, this immediately identifies

$$S_Y(t|x) = 1 - G(t, x'\theta_\perp, x'\beta) \quad (26)$$

for all $t \leq \bar{\tau}$ and $x \in \mathcal{X}_0$.

Averaging G over $v \in \mathcal{V}_0$ with $u = u_\perp + \alpha v$ held fixed gives the function

$$r(u) \equiv \frac{1}{|\mathcal{V}_0|} \int_{\mathcal{V}_0} G(\bar{\tau}, u - \alpha v, v)\mathrm{d}v = (1 - \pi(u))\kappa \quad (27)$$

for $u \in \mathcal{U}_0$, where $|\mathcal{V}_0|$ denotes the length of $\mathcal{V}_0$. This identifies $1 - \pi(u)$ on $\mathcal{U}_0$ up to the unknown multiplicative constant

$$\kappa \equiv \frac{1}{|\mathcal{V}_0|} \int_{\mathcal{V}_0} \tilde{F}_Y(\bar{\tau}|v)\mathrm{d}v. \quad (28)$$

This means the shape of the cure fraction is identified: given a reference point $u_0 \in \mathcal{U}_0$, the ratio $(1 - \pi(u))/(1 - \pi(u_0))$ is known, as is the semi-elasticity $\pi'(u)/(1 - \pi(u))$. It can therefore be established, for instance, whether the cure fraction is monotone in the index, and by what factor a change in $x'\theta$ scales the probability of being susceptible.

In a similar way, the function

$$q(t, v) \equiv \frac{\int_{\mathcal{U}_0} G(t, u - \alpha v, v)\mathrm{d}u}{\int_{\mathcal{U}_0} r(u)\mathrm{d}u} = \frac{\tilde{F}_Y(t|v)}{\kappa} \quad (29)$$

identifies $\tilde{F}_Y(t|v)$ on $\mathcal{V}_0$ up to the constant $1/\kappa$ for any $t \leq \bar{\tau}$. The shape of the latency distribution follows through the ratio $q(t, v)/q(t_0, v_0)$ for some reference points $t_0 \leq \bar{\tau}$ and $v_0 \in \mathcal{V}_0$.

Without additional assumptions, κ is not identified. This is the usual ambiguity between genuine cure fractions and long durations: the pair $(1-\pi, \tilde{F}_Y)$ is observationally equivalent to $(\kappa^*(1-\pi), \tilde{F}_Y/\kappa^*)$ for any $\kappa^* > 0$ for which $\kappa^*(1-\pi(u))$ and $\tilde{F}_Y(\bar{\tau}|v)/\kappa^*$ remain bounded from above by 1. The admissible values of κ form the interval

$$\kappa_- \equiv \max_{u \in \mathcal{U}_0} r(u) \leq \kappa \leq 1/\max_{v \in \mathcal{V}_0} q(\bar{\tau}, v) \equiv \kappa_+. \tag{30}$$

As both endpoints are identified, the interval is as well.

4 Estimation and asymptotic theory

4.1 Estimators

This section turns the identifying equations into plug-in estimators. Throughout, I assume random samples.

Assumption 9 (**Sampling**). *$\{T_i, 1(Y_i \leq C_i), X_i : i = 1, \ldots, n\}$ is a random sample of $(T, 1(Y \leq C), X)$.*

The weight W depends on the unknown censoring distribution, so all estimators require an estimator $\hat{S}_C$ of S_C. I do not propose a new one, as the literature offers many. I only require a uniform rate of convergence and an asymptotically linear representation of averages weighted by the estimation error.

Assumption 10 (**Censoring estimator**).

(a) There exists $1/4 < \mu \leq 1/2$ such that

$$\sup_{t \in [0,\bar{\tau}]} \sup_{x \in \mathcal{X}_0^+} |\hat{S}_C(t|x) - S_C(t|x)| = O_p(n^{-\mu}). \tag{31}$$

(b) There exists a mapping $\xi \mapsto \Omega_C(T, 1(Y \leq C), X; \xi)$, such that for any bounded function $\xi(t, x)$ supported on $\mathbb{R} \times \mathcal{X}_0$, it holds that

$$\begin{aligned} &\frac{1}{n}\sum_{i=1}^{n} \frac{1(Y_i \leq C_i)1(T_i \leq \bar{\tau})}{S_C(T_i|X_i)^2} \xi(T_i, X_i) \left(\hat{S}_C(T_i|X_i) - S_C(T_i|X_i)\right) \\ &\quad = \frac{1}{n}\sum_{i=1}^{n} \Omega_C(T_i, 1(Y_i \leq C_i), X_i; \xi) + o_p(n^{-1/2}), \end{aligned} \tag{32}$$

and $\Omega_C(T, 1(Y \leq C), X; \xi)$ has mean zero and finite variance.

(c) There exists an estimator $\hat{\Omega}_C$ of Ω_C such that, for any ξ as in part (b) and any $\hat{\xi}$ supported on $\mathbb{R} \times \mathcal{X}_0$ with $\sup_{t,x} \|\hat{\xi}(t,x) - \xi(t,x)\| \xrightarrow{p} 0$, it holds that

$$\sup_{t,c,x} \|\hat{\Omega}_C(t,c,x;\hat{\xi}) - \Omega_C(t,c,x;\xi)\| \xrightarrow{p} 0. \tag{33}$$

Assumption 10 holds for standard specifications, such as a Cox model for the censoring times, by routine arguments. Proposition 5 in Appendix B shows that it also holds when the censoring distribution depends on a single index of the covariates and is estimated by kernel methods. The index itself can be estimated by, for instance, the estimators of Gørgens (2004, 2006).

Let $\hat{W}$ denote W with $\hat{S}_C$ plugged in. Let d denote the dimension of X and let $K_\beta : \mathbb{R}^d \to \mathbb{R}$ be a symmetric kernel with bandwidth h_β, and write K'_β for its gradient. Rewriting equation (13) as an expectation and replacing population quantities by kernel estimators gives

$$\hat{\beta} = \frac{1}{n^2 h_\beta^{d+1}} \sum_{i \neq j} \omega(X_i) \hat{W}_i \hat{W}_j \left(\psi(T_j) - \psi(T_i)\right) K'_\beta \left(\frac{X_i - X_j}{h_\beta}\right). \tag{34}$$

Similarly, let $K_\delta : \mathbb{R}^d \to \mathbb{R}$ be a symmetric kernel with bandwidth h_δ and gradient K'_δ, and write ω' for the gradient of ω. Write δ for the integral in equation (14), so that

$$\theta_\perp = \left(I - \frac{\beta\beta'}{\|\beta\|^2}\right)\delta. \tag{35}$$

Integrating equation (14) by parts and proceeding in the same way gives

$$\hat{\delta} = -\frac{1}{n^2 h_\delta^d} \sum_{i \neq j} \hat{W}_i \left(\frac{2}{h_\delta}\omega(X_i) K'_\delta \left(\frac{X_i - X_j}{h_\delta}\right) + \omega'(X_i) K_\delta \left(\frac{X_i - X_j}{h_\delta}\right)\right); \tag{36}$$

the estimator of $\theta_\perp$ is then

$$\hat{\theta}_\perp = \left(I - \frac{\hat{\beta}\hat{\beta}'}{\|\hat{\beta}\|^2}\right)\hat{\delta}. \tag{37}$$

The estimators $\hat{\beta}$ and $\hat{\delta}$ are computable in $O(n^2 d)$ time.

Let $\hat{U}_\perp = X'\hat{\theta}_\perp$ and $\hat{V} = X'\hat{\beta}$. Let K_α be a one-dimensional kernel with associated bandwidth h_α. For a point $(u_\perp, v)$, run the local linear regression with logarithmic link

$$\begin{aligned}\hat{g}(u_\perp, v) = \arg\max_{b \in \mathbb{R}^3} \sum_{i=1}^n &(\hat{W}_i(1, \hat{U}_{\perp i} - u_\perp, \hat{V}_i - v)b - \exp((1, \hat{U}_{\perp i} - u_\perp, \hat{V}_i - v)b)) \\ &\times K_\alpha\left(\frac{\hat{U}_{\perp i} - u_\perp}{h_\alpha}\right) K_\alpha\left(\frac{\hat{V}_i - v}{h_\alpha}\right) 1(X_i \in \mathcal{X}_0).\end{aligned} \tag{38}$$

The intercept $\hat{g}_0(u_\perp, v)$ estimates $g(u_\perp, v)$. The estimator of α is given by

$$\hat{\alpha} = \frac{\int \frac{\partial^2}{\partial u_\perp \partial v} \rho(u_\perp, v) \hat{g}_0(u_\perp, v) \mathrm{d}u_\perp \mathrm{d}v}{\int \frac{\partial^2}{\partial u_\perp^2} \rho(u_\perp, v) \hat{g}_0(u_\perp, v) \mathrm{d}u_\perp \mathrm{d}v}; \tag{39}$$

the estimator of θ is subsequently given by

$$\hat{\theta} = \hat{\theta}_\perp + \hat{\alpha}\hat{\beta}. \tag{40}$$

The estimator $\hat{g}$ is a local linear quasi-likelihood estimator of the type of Fan, Heckman and Wand (1995).

The remaining functions are estimated along the same lines as α. Let K_S be a one-dimensional kernel with associated bandwidth h_S. For points $u_\perp$, v and $t \leq \bar{\tau}$, run the local linear regression

$$\begin{aligned} \hat{G}(t, u_\perp, v) = \underset{b \in \mathbb{R}^3}{\arg\min} \sum_{i=1}^{n} (\hat{W}_i 1(T_i \leq t) - (1, \hat{U}_{\perp i} - u_\perp, \hat{V}_i - v) b)^2 \\ \times K_S\left(\frac{\hat{U}_{\perp i} - u_\perp}{h_S}\right) K_S\left(\frac{\hat{V}_i - v}{h_S}\right) 1(X_i \in \mathcal{X}_0). \end{aligned} \tag{41}$$

The intercept $\hat{G}_0(t, u_\perp, v)$ estimates $G(t, u_\perp, v)$. The survival function is estimated by

$$\hat{S}_Y(t|x) = 1 - \hat{G}_0(t, x'\hat{\theta}_\perp, x'\hat{\beta}) \tag{42}$$

for $t \leq \bar{\tau}$ and $x \in \mathcal{X}_0$. Importantly, this does not require $\hat{\alpha}$. The functions r and q are estimated by

$$\hat{r}(u) = \frac{1}{|\mathcal{V}_0|} \int_{\mathcal{V}_0} \hat{G}_0(\bar{\tau}, u - \hat{\alpha}v, v) \mathrm{d}v \tag{43}$$

and

$$\hat{q}(t, v) = \frac{\int_{\mathcal{U}_0} \hat{G}_0(t, u - \hat{\alpha}v, v) \mathrm{d}u}{\int_{\mathcal{U}_0} \hat{r}(u) \mathrm{d}u}, \tag{44}$$

respectively. Finally, the admissible interval for κ is estimated by $[\hat{\kappa}_-, \hat{\kappa}_+]$, with

$$\hat{\kappa}_- = \max_{u \in \mathcal{U}_0} \hat{r}(u) \tag{45}$$

and

$$\hat{\kappa}_+ = 1 / \max_{v \in \mathcal{V}_0} \hat{q}(\bar{\tau}, v). \tag{46}$$

Evaluating $1 - \hat{r}(u)/\kappa$ and $\kappa \hat{q}(t, v)$ at both endpoints bounds $\pi(u)$ and $\tilde{F}_Y(t|v)$.

4.2 Asymptotic properties

This subsection derives the asymptotic properties of the estimators. Their influence functions and asymptotic variances are given in Appendix A, and the proofs in Appendix B.

The asymptotic results require conditions on smoothness, kernels and bandwidths. For the bias of $\hat{\beta}$ and $\hat{\delta}$ to be $o(n^{-1/2})$, the bandwidths must vanish sufficiently fast and the kernels must be of sufficiently high order. The required smoothness increases as μ decreases, since a less precise $\hat{S}_C$ forces h_β and h_δ to vanish more slowly, which must be compensated by higher-order kernels.

In the assumptions below, let $p_\beta, p_\delta > \max\{(d+2)/4\mu, 1/(4\mu-1)\}$ be even integers denoting the orders of the kernels K_β and K_δ. In addition, let $\mathcal{R}$ be a compact subset of the interior of $\{(x'\theta_\perp, x'\beta) : x \in \mathcal{X}_0\}$ that contains the support of ρ and $\{(u - \alpha v, v) : u \in \mathcal{U}_0, v \in \mathcal{V}_0\}$. Finally, for vectors $k \in \mathbb{N}^d$ and $b \in \mathbb{R}^d$, write $|k| = k_1 + \ldots + k_d$ and $b^k = b_1^{k_1} \cdots b_d^{k_d}$.

Assumption 11 (**Smoothness**).

(a) The derivatives $\partial^{|k|} f_X(x)/\partial x^k$ exist and are bounded on $\mathcal{X}_0^+$ for all $k \in \mathbb{N}^d$ with $0 \leq |k| \leq \max\{p_\beta, p_\delta\} + 1$.

(b) There exists an open set $\mathcal{R}^+$ containing $\mathcal{R}$ on which the density of $(U_\perp, V)$ given $X \in \mathcal{X}_0$ is twice continuously differentiable.

(c) The derivatives $\partial^{|k|} S_C(t|x)/\partial x^k$ exist and are bounded on $[0, \bar{\tau}] \times \mathcal{X}_0^+$ for all $k \in \mathbb{N}^d$ with $0 \leq |k| \leq 2$.

(d) The derivatives $\partial^j \tilde{f}_Y(t|v)/\partial v^j$ exist and are bounded on $[0, \bar{\tau}] \times \mathcal{V}$ for all $j \leq \max\{p_\beta, p_\delta\} + 1$.

(e) The derivatives $\partial^j \pi(u)/\partial u^j$ exist and are bounded on $\mathcal{U}$ for all $j \leq \max\{p_\beta, p_\delta\} + 1$.

(f) The derivatives $\partial^{|k|} \omega(x)/\partial x^k$ exist and are bounded on $\mathcal{X}_0^+$ for all $k \in \mathbb{N}^d$ with $0 \leq |k| \leq \max\{p_\beta, p_\delta\} + 1$.

Assumption 12 (**Kernels**).

(a) The kernel K_β is bounded, symmetric, continuously differentiable with bounded gradient, and has compact support. It satisfies $\int K_\beta(\nu)\mathrm{d}\nu = 1$ and $\int \nu^k K_\beta(\nu)\mathrm{d}\nu = 0$ for all $k \in \mathbb{N}^d$ with $1 \leq |k| \leq p_\beta - 1$.

(b) The kernel K_δ is bounded, symmetric, continuously differentiable with bounded gradient, and has compact support. It satisfies $\int K_\delta(\nu)\mathrm{d}\nu = 1$ and $\int \nu^k K_\delta(\nu)\mathrm{d}\nu = 0$ for all $k \in \mathbb{N}^d$ with $1 \leq |k| \leq p_\delta - 1$.

(c) The kernel K_α is bounded, symmetric, Lipschitz continuous, and has compact support. It satisfies $\int K_\alpha(\nu)\mathrm{d}\nu = 1$.

(d) The kernel K_S is bounded, symmetric, twice continuously differentiable, and has compact support. It satisfies $\int K_S(\nu)\mathrm{d}\nu = 1$.

Assumption 13 (**Bandwidths**). *As $n \to \infty$, the bandwidths satisfy*

$$\begin{aligned}
&nh_\beta^{2p_\beta} \to 0, && \frac{\log n}{n^{2\mu}h_\beta^{d+2}} \to 0, && \frac{1}{n^{4\mu-1}h_\beta^2} \to 0,\\
&nh_\delta^{2p_\delta} \to 0, && \frac{\log n}{n^{2\mu}h_\delta^{d+2}} \to 0, && \frac{1}{n^{4\mu-1}h_\delta^2} \to 0,\\
&\frac{\log n}{n^{1/2}h_\alpha^2} \to 0, && nh_\alpha^6 \to 0, &&\\
&\frac{(\log n)^2}{nh_S^3} \to 0, && nh_S^5 \to 0, && n^{1-2\mu}h_S^2 \to 0.
\end{aligned} \tag{47}$$

The first result is for the index vectors. The estimators $\hat{\beta}$ and $\hat{\delta}$ are second order U-statistics of the same form as in Powell, Stock and Stoker (1989), except that the weights $\hat{W}_i$ are estimated. Although $\hat{g}_0$ converges at a nonparametric rate, $\hat{\alpha}$ integrates it over both arguments and is therefore $\sqrt{n}$-consistent.

Theorem 1 (**Index vectors**). *Assume Assumptions 1–13 hold.*

(a) As $n \to \infty$, it holds that

$$\hat{\beta} - \beta = \frac{1}{n}\sum_{i=1}^{n} \Omega_\beta(T_i, 1(Y_i \le C_i), X_i) + o_p(n^{-1/2}), \tag{48}$$

where $\Omega_\beta(T, 1(Y \le C), X)$ has mean zero and finite variance.

(b) As $n \to \infty$, it holds that

$$\hat{\theta} - \theta = \frac{1}{n}\sum_{i=1}^{n} \Omega_\theta(T_i, 1(Y_i \le C_i), X_i) + o_p(n^{-1/2}), \tag{49}$$

where $\Omega_\theta(T, 1(Y \le C), X)$ has mean zero and finite variance.

The next theorem establishes uniform expansions and pointwise asymptotic normality of the function estimators. Because $\hat{r}$ and $\hat{q}$ average $\hat{G}_0$ over one direction, they converge at the one-dimensional rate, which is faster than that of $\hat{S}_Y$. The estimated indices, $\hat{\alpha}$ and $\hat{S}_C$ converge faster than $\hat{G}_0$, so they do not affect the limit distributions.

Theorem 2 (**Function estimators**). *Fix $\underline{\tau} \in (0, \bar{\tau})$. Assume Assumptions 1–13 hold.*

(a) As $n \to \infty$, it holds that

$$\hat{S}_Y(t|x) = S_Y(t|x) + L_S(t,x) + o_p((nh_S^2)^{-1/2}) \tag{50}$$

uniformly over $t \in [\underline{\tau}, \bar{\tau}]$ and $x \in \mathcal{X}_0$ with $(x'\theta, x'\beta) \in \mathcal{U}_0 \times \mathcal{V}_0$, and, for each such t and x,

$$\sqrt{nh_S^2}(\hat{S}_Y(t|x) - S_Y(t|x)) \xrightarrow{d} N(0, \sigma_S^2(t,x)). \tag{51}$$

(b) As $n \to \infty$, it holds that

$$\hat{r}(u) = (1 - \pi(u))\kappa + L_r(u) + o_p((nh_S)^{-1/2}) \tag{52}$$

uniformly over $u \in \mathcal{U}_0$, and, for each $u \in \mathcal{U}_0$,

$$\sqrt{nh_S}(\hat{r}(u) - (1 - \pi(u))\kappa) \xrightarrow{d} N(0, \sigma_r^2(u)). \tag{53}$$

(c) As $n \to \infty$, it holds that

$$\hat{q}(t,v) = \tilde{F}_Y(t|v)/\kappa + L_q(t,v) + o_p((nh_S)^{-1/2}) \tag{54}$$

uniformly over $t \in [\underline{\tau}, \bar{\tau}]$ and $v \in \mathcal{V}_0$, and, for each such t and v,

$$\sqrt{nh_S}(\hat{q}(t,v) - \tilde{F}_Y(t|v)/\kappa) \xrightarrow{d} N(0, \sigma_q^2(t,v)). \tag{55}$$

Finally, the bounds on κ inherit the uniform rate of $\hat{r}$ and $\hat{q}$.

Proposition 1 (**Bounds on κ**). *Under Assumptions 1–13, as $n \to \infty$, it holds that*

$$|\hat{\kappa}_- - \kappa_-| + |\hat{\kappa}_+ - \kappa_+| = O_p((\log(n)/nh_S)^{1/2}). \tag{56}$$

Since taking a maximum is not a smooth operation, the bounds are generally not asymptotically normal. The maxima of $\hat{r}$ and $\hat{q}$ are also biased upward by Jensen's inequality, which pushes the two bounds toward each other Bounds of this type are known as intersection bounds, and Chernozhukov, Lee and Rosen (2013) propose bias corrections and confidence intervals for them.

4.3 Testing the assumptions

In this subsection I propose tests of the main assumptions behind the identification and estimation results. The most important one is non-proportionality of θ and β, without which only trivial bounds on the incidence and the latency remain. I also test the more technical condition that $\log(1-\pi)$ is not affine, without which α is not identified. Finally, I test the mixture structure itself, which implies an exclusion restriction that the promotion time model violates. The proofs are given in Appendix B.

Non-proportionality is equivalent to $\theta_\perp \neq 0$, which can be tested by a Wald test. As $\hat{\theta}_\perp$ is orthogonal to $\hat{\beta}$ by construction, the test uses only its coordinates in the orthogonal complement of $\hat{\beta}$.

Proposition 2 (**Test of non-proportionality**). *Define*

$$\Sigma_{\theta_\perp} = E[\Omega_{\theta_\perp}(T, 1(Y \leq C), X)\Omega_{\theta_\perp}(T, 1(Y \leq C), X)'] \tag{57}$$

and

$$\hat{\Sigma}_{\theta_\perp} = \frac{1}{n}\sum_{i=1}^{n} \hat{\Omega}_{\theta_\perp}(T_i, 1(Y_i \leq C_i), X_i)\hat{\Omega}_{\theta_\perp}(T_i, 1(Y_i \leq C_i), X_i)'. \tag{58}$$

Let $\hat{B}$ be a $d \times (d-1)$ matrix whose columns form an orthonormal basis of the orthogonal complement of $\hat{\beta}$, and set

$$J_1 = n\hat{\theta}_\perp'\hat{B}(\hat{B}'\hat{\Sigma}_{\theta_\perp}\hat{B})^{-1}\hat{B}'\hat{\theta}_\perp. \tag{59}$$

Suppose Assumptions 2–13 hold and that $b'\Sigma_{\theta_\perp}b > 0$ for every nonzero vector b with $b'\beta = 0$. The statistic J_1 does not depend on the choice of $\hat{B}$. As $n \to \infty$, if $\theta \propto \beta$, then $J_1 \xrightarrow{d} \chi^2(d-1)$. If $\theta \not\propto \beta$, then $J_1 \xrightarrow{p} \infty$.

Affinity of $\log(1-\pi)$ means that $g_1(x'\theta_\perp, x'\beta)$ is constant on $\mathcal{X}_0$. To test this, project the gradient of m_1 from Section 3 onto the orthogonal complement of β. Since $-\pi'(x'\theta)\tilde{F}_Y(\bar{\tau}|x'\beta) = m_1(x)g_1(x'\theta_\perp, x'\beta)$, this gives

$$\left(I - \frac{\beta\beta'}{\|\beta\|^2}\right)\frac{\partial}{\partial x}m_1(x) = m_1(x)g_1(x'\theta_\perp, x'\beta)\theta_\perp. \tag{60}$$

Let $\omega_1, \ldots, \omega_L$ be linearly independent weight functions with support $\mathcal{X}_0$ and $\omega_1 = \omega$. For $l = 1, \ldots, L$, let δ_l denote δ with ω_l in place of ω, and define

$$e_l = \int \omega_l(x) f_X(x)^2 m_1(x)\mathrm{d}x = E[\omega_l(X)f_X(X)W]. \tag{61}$$

Then

$$\frac{\theta_\perp'\delta_l}{\|\theta_\perp\|^2 e_l} = \frac{\int \omega_l(x) f_X(x)^2 m_1(x) g_1(x'\theta_\perp, x'\beta)\mathrm{d}x}{\int \omega_l(x) f_X(x)^2 m_1(x)\mathrm{d}x}. \tag{62}$$

If $g_1(x'\theta_\perp, x'\beta)$ is constant on $\mathcal{X}_0$, the right-hand side does not depend on l. For $l = 1$, the left-hand side equals $1/e_1$ by Assumption 7. Hence

$$\theta_\perp'\delta_l e_1 - \|\theta_\perp\|^2 e_l = 0 \tag{63}$$

for all $l = 2, \ldots, L$. The test below checks whether these conditions hold.

Proposition 3 (**Test of affinity**). *Let $\hat{\delta}_l$ denote $\hat{\delta}$ with ω_l in place of ω. Define, for $l = 1, \ldots, L$,*

$$\hat{e}_l = \frac{1}{n^2 h_\delta^d} \sum_{i \neq j} \omega_l(X_i) \hat{W}_i K_\delta \left(\frac{X_i - X_j}{h_\delta} \right), \tag{64}$$

and

$$\hat{R}_l = \hat{\theta}_\perp' \hat{\delta}_l \hat{e}_1 - \|\hat{\theta}_\perp\|^2 \hat{e}_l. \tag{65}$$

Finally, define $\hat{R} = (\hat{R}_2, \ldots, \hat{R}_L)'$,

$$\Sigma_R = E[\Omega_R(T, 1(Y \leq C), X)\Omega_R(T, 1(Y \leq C), X)'] \tag{66}$$

and

$$\hat{\Sigma}_R = \frac{1}{n} \sum_{i=1}^{n} \hat{\Omega}_R(T_i, 1(Y_i \leq C_i), X_i)\hat{\Omega}_R(T_i, 1(Y_i \leq C_i), X_i)', \tag{67}$$

and set

$$J_2 = n\hat{R}'\hat{\Sigma}_R^{-1}\hat{R}. \tag{68}$$

Suppose Assumptions 1–7 and 9–13 hold, that the conditions on ω hold for all ω_l, and that Σ_R is nonsingular. If $\log(1 - \pi(u))$ is affine on $\{x'\theta : x \in \mathcal{X}_0\}$, then $J_2 \xrightarrow{d} \chi^2(L-1)$. If $\theta_\perp'\delta_l e_1 \neq \|\theta_\perp\|^2 e_l$ for some l, then $J_2 \xrightarrow{p} \infty$.

Under the mixture model, the cure probability affects how many spells complete, but not how long completed spells take. In the identification argument, this shows up in the ratio $m_\psi(x)/m_1(x)$, which depends on $x'\beta$ but not on $x'\theta$,

whatever the choice of ψ. This exclusion restriction does not hold in the promotion time model $S_Y(t|x) = \exp(-\eta(x'\theta)F(t|x'\beta))$, where the cure index scales the hazard of every spell. For instance, for $\psi(t) = 1(t \leq \tau)$,

$$\frac{m_\psi(x)}{m_1(x)} = \frac{1 - \exp(-\eta(x'\theta)F(\tau|x'\beta))}{1 - \exp(-\eta(x'\theta)F(\bar{\tau}|x'\beta))}, \tag{69}$$

which varies with $x'\theta$ whenever η does. As a result, the average derivative in Assumption 7 points in the direction of β for every ψ under the mixture model, but not under the promotion time model.

The test below compares this direction for two choices of ψ. Let $\psi^* : \mathbb{R} \to \mathbb{R}$ be a known function bounded on $[0, \bar{\tau}]$ that is not an affine function of ψ on $[0, \bar{\tau}]$, and define

$$\beta^* = \int \omega(x) f_X(x)^2 \left(m_1(x) \frac{\partial}{\partial x} m_{\psi^*}(x) - m_{\psi^*}(x) \frac{\partial}{\partial x} m_1(x) \right) \mathrm{d}x. \tag{70}$$

Under the mixture model,

$$\beta^* = \beta \int \omega(x) f_X(x)^2 m_1(x)^2 H^{*\prime}(x'\beta) \mathrm{d}x, \tag{71}$$

where H^* denotes H with ψ^* in place of ψ, so that

$$\left(I - \frac{\beta\beta'}{\|\beta\|^2} \right) \beta^* = 0. \tag{72}$$

Proposition 4 (Test of the mixture structure). *Let $\hat{\beta}^*$ denote $\hat{\beta}$ with ψ^* in place of ψ, and let Ω_{β^*} denote Ω_β with ψ^* in place of ψ. Let Σ_{β^*} denote the variance of $\Omega_{\beta^*}(T, 1(Y \leq C), X) - \frac{\beta'\beta^*}{\|\beta\|^2}\Omega_\beta(T, 1(Y \leq C), X)$ and $\hat{\Sigma}_{\beta^*}$ its plug-in estimator. With $\hat{B}$ as in Proposition 2, set*

$$J_3 = n\hat{\beta}^{*\prime}\hat{B}(\hat{B}'\hat{\Sigma}_{\beta^*}\hat{B})^{-1}\hat{B}'\hat{\beta}^*. \tag{73}$$

Suppose Assumptions 1–13 hold, also with ψ^* in place of ψ, and that $b'\Sigma_{\beta^*}b > 0$ for every nonzero vector b with $b'\beta = 0$. As $n \to \infty$, if $\beta^* \propto \beta$, then $J_3 \xrightarrow{d} \chi^2(d-1)$. If $\beta^* \not\propto \beta$, then $J_3 \xrightarrow{p} \infty$.

The tests J_2 and J_3 are not omnibus, as they only detect departures that show up through the chosen weights. Omnibus kernel tests in the spirit of Zheng (1996) are possible, at the cost of a slower rate against local alternatives.

5 Monte Carlo results

I conduct a Monte Carlo study to assess the finite-sample behavior of the estimators. As a benchmark I use the estimator of Amico, Van Keilegom and Legrand (2019), the closest alternative in the literature. It models the incidence by a single index with an unknown link and the latency by a Cox model, and identifies the cure fraction by imposing sufficient follow-up.

The Monte Carlo design has three independent standard normal covariates. Both the censoring and latency distributions follow a Weibull proportional hazards model, given by

$$S_C(t|x) = \exp\left(-t\exp(-2.5 + x'\gamma)\right) 1(t < 6), \tag{74}$$

and

$$\tilde{S}_Y(t|x'\beta) = \exp\left(-t^{1.5}\exp(-1.5 + x'\beta)\right), \tag{75}$$

respectively. The indicator in the censoring distribution sets the end of follow-up at 6. The cure probability follows the logistic link function

$$\pi(x'\theta) = \frac{1}{1 + \exp(2 + 2x'\theta)}. \tag{76}$$

In all experiments, I fix $\gamma = (1, 1, 0)'$ and $\beta = (1, 0, 1)'$. For θ, I consider three values, chosen so that the two indices move closer to proportionality from one configuration to the next. In the first, I set $\theta = (0, 1, 1)'$, which gives strong identification. In the second, I set $\theta = (\sqrt{1/2}, \sqrt{1/2}, 1)'$, which gives moderate identification. In the third, I set $\theta = (\sqrt{2/3}, \sqrt{1/3}, 1)'$, which leaves identification weak.

For expositional reasons, the values of β and θ above do not satisfy the normalizations in Assumption 7. This is without loss of generality, as the scales of the indices are arbitrary. I rescale all index estimators and their influence functions so that they estimate the values above.

Across all configurations I consider two cutoffs, $\bar{\tau} = 3$ and $\bar{\tau} = 4$. The three configurations and two thresholds give six experiments in total. Each experiment consists of 1,000 Monte Carlo replications and is conducted with sample sizes $n = 4{,}000$ and $n = 8{,}000$.

Table 1 reports summary statistics of the design for each configuration and each threshold. As the censoring times are bounded by 6 while the latency has unbounded support, sufficient follow-up fails in all experiments. Specifically, the probability that a susceptible spell outlasts follow-up ranges from 0.08 to 0.13 across configurations.

Table 1: Summary statistics

$\bar{\tau}$	$E[\tilde{Y}]$	$Q_{25}(\tilde{Y})$	$Q_{50}(\tilde{Y})$	$Q_{75}(\tilde{Y})$	$E[\pi(X'\theta)]$	$P(Y > C)$	$P(\tilde{Y} \leq \bar{\tau})$	$P(Y \leq \min(C, \bar{\tau}))$
Configuration 1: $\theta = (0, 1, 1)'$								
3	3.05	0.71	1.65	3.60	0.274	0.511	0.696	0.406
4	3.05	0.71	1.65	3.60	0.274	0.511	0.779	0.446
Configuration 2: $\theta = (\sqrt{1/2}, \sqrt{1/2}, 1)'$								
3	2.46	0.66	1.47	3.06	0.274	0.490	0.744	0.434
4	2.46	0.66	1.47	3.06	0.274	0.490	0.825	0.473
Configuration 3: $\theta = (\sqrt{3/4}, 1/2, 1)'$								
3	2.37	0.65	1.45	2.98	0.274	0.485	0.752	0.439
4	2.37	0.65	1.45	2.98	0.274	0.485	0.833	0.478

Note: $\tilde{Y}$ denotes Y conditional on the spell being susceptible, that is, $\tilde{Y} = Y|(Y < \infty)$.

I estimate the censoring distribution by Cox partial likelihood with a Breslow baseline, which is correctly specified here and satisfies Assumption 10 with $\mu = 1/2$.

To estimate β and δ, I choose the weight functions $\psi(t) = -t$ and $\omega(x) = 1(x \in \mathcal{X}_0)$ with $\mathcal{X}_0 = [-3, 3]^3$, which also serves as the trimming set in the estimation of α and of the functions. Although this ω does not satisfy the smoothness requirement in Assumption 11, this is immaterial in finite samples, as the discontinuity is contained in an interval in which there are no observations of X with probability 1. For both estimators, the kernel function is the product kernel $K_4(x_1)K_4(x_2)K_4(x_3)$, where K_4 denotes the univariate fourth-order kernel

$$K_4(z) = \frac{105}{64}(1 - 5z^2 + 7z^4 - 3z^6)\, 1(|z| \leq 1). \tag{77}$$

The bandwidths are $h_\beta = h_\delta = 5n^{-1/6}$.

To estimate α, I use the weight function

$$\rho(u_\perp, v) = \left(1 - \frac{u_\perp^2}{(1.5\|\hat{\theta}_\perp\|)^2}\right)^3 \left(1 - \frac{v^2}{(1.5\|\hat{\beta}\|)^2}\right)^3 \times 1(|u_\perp| \leq 1.5\|\hat{\theta}_\perp\|)1(|v| \leq 1.5\|\hat{\beta}\|). \tag{78}$$

The kernel function is

$$K_\alpha(z) = \frac{15}{16}(1 - z^2)^2\, 1(|z| \leq 1), \tag{79}$$

with bandwidths $5\,\mathrm{sd}(X'\hat{\theta}_\perp) \times n^{-2/11}$ in the kernel factor for $\hat{U}_\perp$ and $5\,\mathrm{sd}(X'\hat{\beta}) \times n^{-2/11}$ for that of $\hat{V}$, where $\mathrm{sd}(\cdot)$ denotes the sample standard deviation.

For the remaining estimators, I set $K_S(z) = K_\alpha(z)$ with bandwidths $4 \times \mathrm{sd}(X'\hat{\theta}_\perp) \times n^{-2/9}$ and $4 \times \mathrm{sd}(X'\hat{\beta}) \times n^{-2/9}$. The sets $\mathcal{U}_0$ and $\mathcal{V}_0$ shrink from one configuration to the next, since the function estimators need observations at every combination of $u \in \mathcal{U}_0$ and $v \in \mathcal{V}_0$, and such combinations become rare as the indices move closer to proportionality. I set $\mathcal{U}_0 = \mathcal{V}_0 = [-1.75, 1.75]$ in the first configuration, $[-1, 1]$ in the second and $[-0.5, 0.5]$ in the third.

The benchmark is computed by the EM algorithm of Amico, Van Keilegom and Legrand (2019), to which I refer for details. I set its bandwidth to $n^{-1/5}$ times the sample standard deviation of the estimated cure index.

Tables 2 and 3 report the coefficient estimates. The new estimator recovers the latency index accurately throughout, with precision that does not seem to depend on the configuration. The cure index is accurate in the first configuration, but loses precision as the indices approach proportionality. The loss enters through $\hat{\alpha}$, which at $n = 4{,}000$ is biased toward zero and more variable than its estimated standard error suggests in the third configuration. Both improve with the sample size and with a higher value of $\bar{\tau}$. For the benchmark only the root mean squared error is reported, as the method does not come with a closed-form variance estimator. The benchmark is more precise than the new estimator, but its cure index is biased toward β in every configuration, and the bias does not shrink with n. This is the mechanics of insufficient follow-up: long susceptible spells are counted as cured, so covariates that lengthen spells appear to raise the cure probability.

Table 4 reports the bounds on κ. Since the maxima of $\hat{r}$ and $\hat{q}$ are biased upward, $\hat{\kappa}_-$ can exceed $\hat{\kappa}_+$ when the identified interval is narrow; in that case I report the two in reversed order, which leaves the rate of convergence unchanged. The identified interval is narrow in the first configuration and widens as $\mathcal{U}_0$ and $\mathcal{V}_0$ shrink. The median estimates lie just outside the identified interval in every experiment, so the estimated interval leans conservative. In the first configuration, the sampling variation in the bounds is much larger than the width of the interval itself.

Figures 1 and 2 show the estimated incidence and latency shapes for $n = 8{,}000$ and $\bar{\tau} = 3$. The median of the new estimator tracks the truth in all configurations, with bands that widen toward the ends of $\mathcal{U}_0$ and $\mathcal{V}_0$ and as the indices approach proportionality. The benchmark varies little across replications but is inconsistent in every configuration: its incidence shape is too steep and its latency shape too flat, for the same reason as above.

The overall picture is as expected. When the indices are well separated, the

Table 2: Coefficient estimates for $n = 4{,}000$

	θ_1	θ_2	θ_3	β_1	β_2	β_3	α
Configuration 1: $\theta = (0, 1, 1)'$							
True value	0.000	1.000	1.000	1.000	0.000	1.000	0.500
$\bar{\tau} = 3$	0.007	0.964	1.008	1.003	0.002	0.989	0.498
	(0.173)	(0.136)	(0.130)	(0.066)	(0.098)	(0.066)	(0.082)
	[0.159]	[0.130]	[0.119]	[0.061]	[0.090]	[0.061]	[0.087]
$\bar{\tau} = 4$	-0.003	0.977	1.003	1.003	0.002	0.991	0.493
	(0.143)	(0.107)	(0.104)	(0.057)	(0.086)	(0.057)	(0.072)
	[0.137]	[0.106]	[0.100]	[0.053]	[0.082]	[0.053]	[0.081]
Benchmark	0.393	0.741	1.138	1.004	0.014	0.995	-
	[0.394]	[0.261]	[0.139]	[0.023]	[0.037]	[0.024]	-
Configuration 2: $\theta = (\sqrt{1/2}, \sqrt{1/2}, 1)'$							
True value	0.707	0.707	1.000	1.000	0.000	1.000	0.854
$\bar{\tau} = 3$	0.680	0.640	0.977	1.003	0.002	0.991	0.821
	(0.213)	(0.225)	(0.168)	(0.060)	(0.089)	(0.060)	(0.089)
	[0.265]	[0.222]	[0.244]	[0.058]	[0.082]	[0.056]	[0.219]
$\bar{\tau} = 4$	0.684	0.674	0.991	1.003	0.002	0.992	0.833
	(0.167)	(0.187)	(0.120)	(0.053)	(0.079)	(0.053)	(0.076)
	[0.188]	[0.191]	[0.162]	[0.049]	[0.074]	[0.049]	[0.143]
Benchmark	0.828	0.508	1.027	1.002	0.015	0.997	-
	[0.123]	[0.201]	[0.035]	[0.020]	[0.034]	[0.020]	-
Configuration 3: $\theta = (\sqrt{2/3}, \sqrt{1/3}, 1)'$							
True value	0.816	0.577	1.000	1.000	0.000	1.000	0.908
$\bar{\tau} = 3$	0.686	0.494	0.875	1.002	0.002	0.992	0.773
	(0.273)	(0.308)	(0.224)	(0.060)	(0.088)	(0.059)	(0.124)
	[0.498]	[0.273]	[0.487]	[0.058]	[0.080]	[0.055]	[0.464]
$\bar{\tau} = 4$	0.729	0.535	0.933	1.002	0.002	0.993	0.826
	(0.182)	(0.247)	(0.140)	(0.052)	(0.078)	(0.052)	(0.078)
	[0.372]	[0.239]	[0.363]	[0.049]	[0.072]	[0.048]	[0.350]
Benchmark	0.895	0.412	1.014	1.001	0.015	0.998	-
	[0.081]	[0.168]	[0.025]	[0.019]	[0.033]	[0.020]	-

Note: Averages over 1,000 replications. Estimated standard errors in parentheses and root mean squared errors in square brackets.

Table 3: Coefficient estimates for $n = 8{,}000$

	θ_1	θ_2	θ_3	β_1	β_2	β_3	α
Configuration 1: $\theta = (0, 1, 1)'$							
True value	0.000	1.000	1.000	1.000	0.000	1.000	0.500
$\bar{\tau} = 3$	0.004	0.980	1.006	1.002	0.000	0.994	0.499
	(0.123)	(0.096)	(0.093)	(0.047)	(0.070)	(0.047)	(0.060)
	[0.112]	[0.091]	[0.085]	[0.043]	[0.067]	[0.042]	[0.062]
$\bar{\tau} = 4$	-0.003	0.988	1.001	1.002	0.000	0.995	0.495
	(0.105)	(0.078)	(0.076)	(0.042)	(0.062)	(0.042)	(0.054)
	[0.103]	[0.077]	[0.072]	[0.038]	[0.060]	[0.038]	[0.060]
Benchmark	0.392	0.741	1.139	1.004	0.012	0.996	-
	[0.393]	[0.260]	[0.140]	[0.016]	[0.027]	[0.017]	-
Configuration 2: $\theta = (\sqrt{1/2}, \sqrt{1/2}, 1)'$							
True value	0.707	0.707	1.000	1.000	0.000	1.000	0.854
$\bar{\tau} = 3$	0.703	0.669	1.001	1.002	0.001	0.995	0.848
	(0.143)	(0.157)	(0.107)	(0.043)	(0.064)	(0.042)	(0.062)
	[0.141]	[0.159]	[0.111]	[0.040]	[0.061]	[0.039]	[0.087]
$\bar{\tau} = 4$	0.695	0.691	0.999	1.002	0.000	0.995	0.844
	(0.118)	(0.135)	(0.081)	(0.038)	(0.057)	(0.038)	(0.056)
	[0.120]	[0.138]	[0.095]	[0.035]	[0.055]	[0.035]	[0.082]
Benchmark	0.826	0.509	1.029	1.001	0.014	0.999	-
	[0.120]	[0.200]	[0.032]	[0.014]	[0.026]	[0.014]	-
Configuration 3: $\theta = (\sqrt{2/3}, \sqrt{1/3}, 1)'$							
True value	0.816	0.577	1.000	1.000	0.000	1.000	0.908
$\bar{\tau} = 3$	0.787	0.525	0.973	1.001	0.001	0.996	0.876
	(0.152)	(0.202)	(0.118)	(0.042)	(0.062)	(0.042)	(0.062)
	[0.254]	[0.204]	[0.240]	[0.040]	[0.059]	[0.038]	[0.226]
$\bar{\tau} = 4$	0.803	0.552	0.991	1.002	0.001	0.996	0.895
	(0.122)	(0.177)	(0.086)	(0.037)	(0.056)	(0.037)	(0.055)
	[0.150]	[0.175]	[0.125]	[0.034]	[0.053]	[0.034]	[0.117]
Benchmark	0.893	0.412	1.015	1.001	0.014	0.999	-
	[0.078]	[0.166]	[0.021]	[0.013]	[0.025]	[0.013]	-

Note: Averages over 1,000 replications. Estimated standard errors in parentheses and root mean squared errors in square brackets.

Figure 1: Estimated incidence shapes for $n = 8{,}000$ and $\bar{\tau} = 3$.

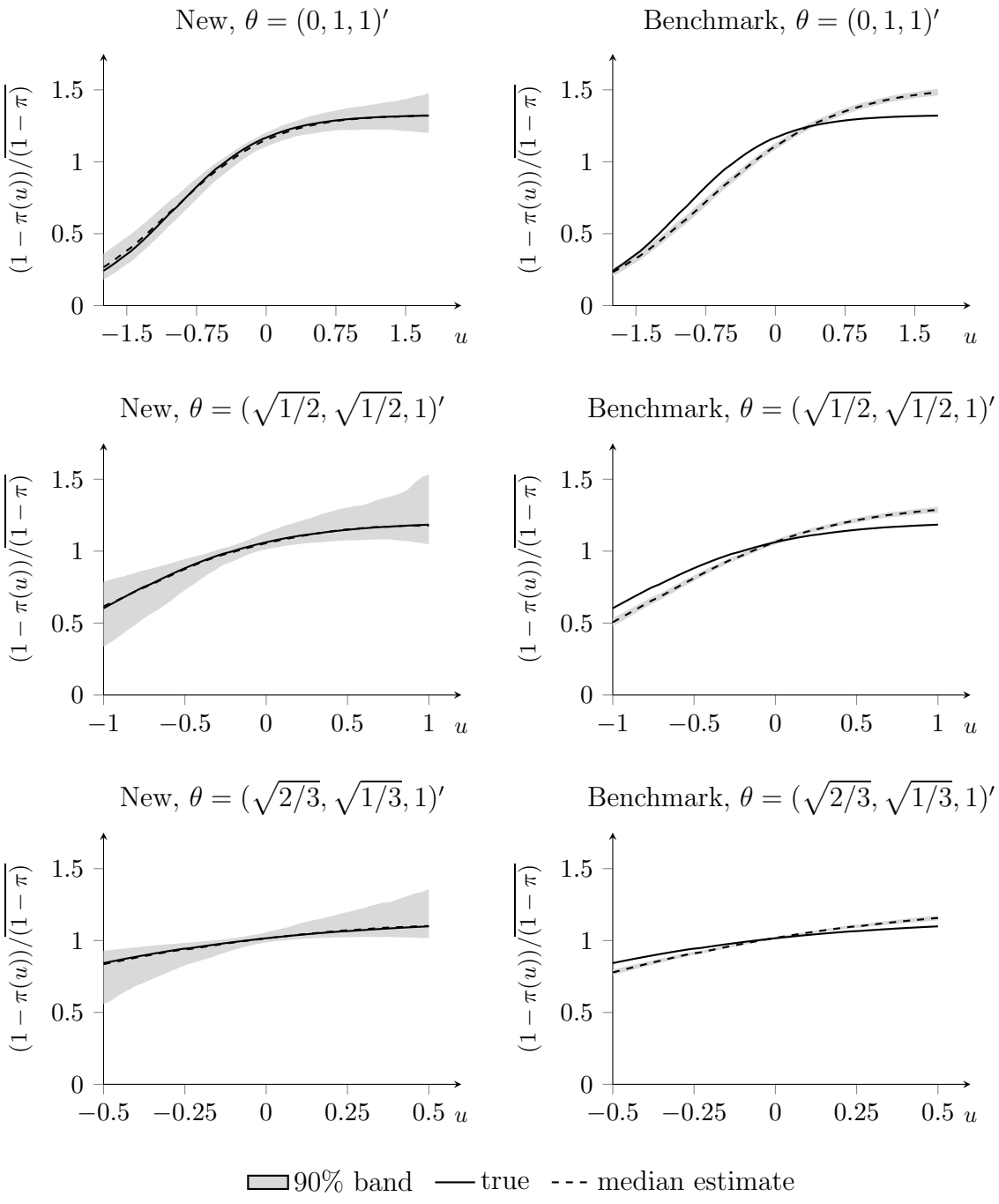


Figure 2: Estimated latency shapes for $n = 8{,}000$ and $\bar{\tau} = 3$.

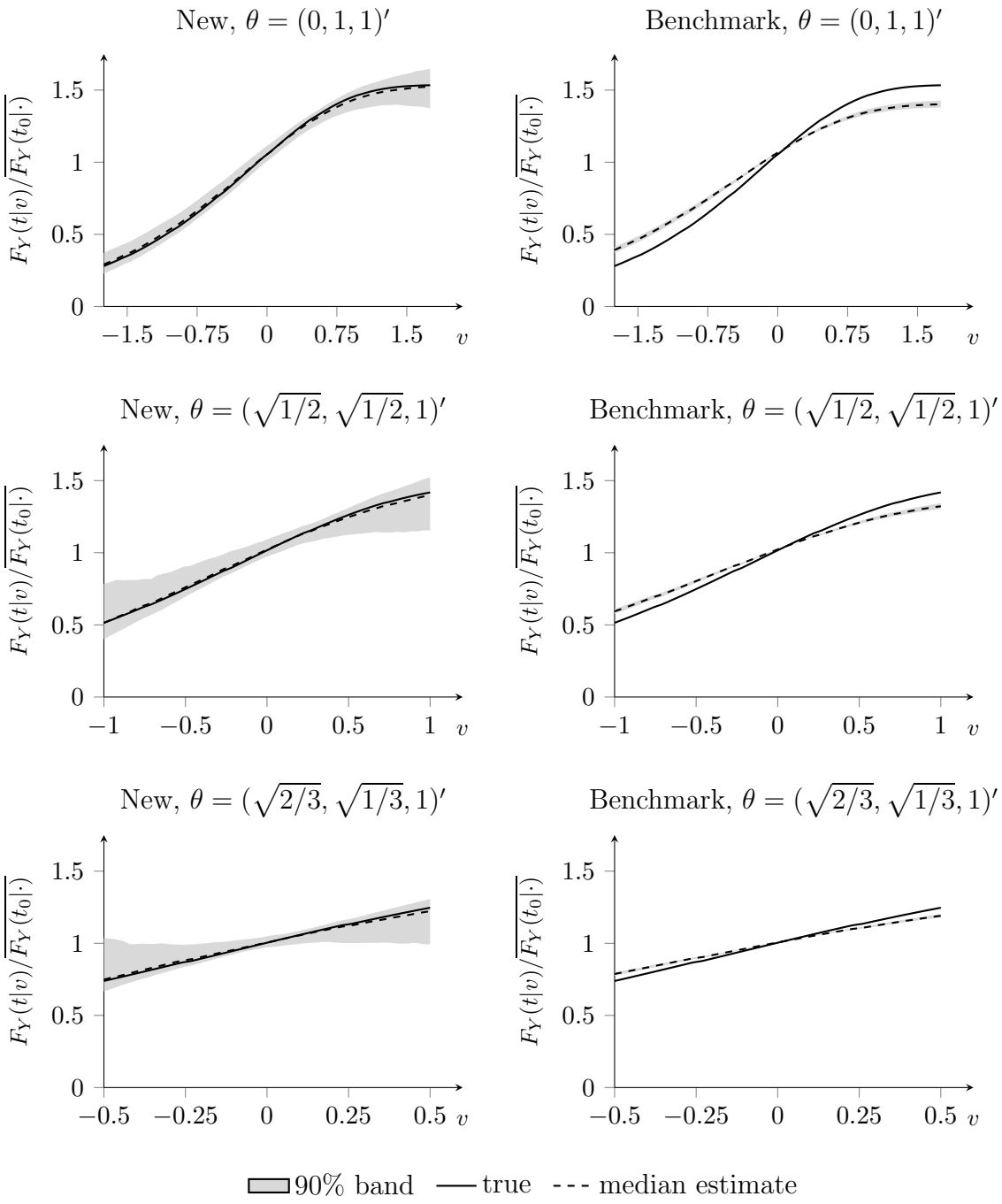


Table 4: Bounds on κ and their estimates

				$n = 4{,}000$		$n = 8{,}000$	
	κ_-	κ	κ_+	$\hat{\kappa}_-$	$\hat{\kappa}_+$	$\hat{\kappa}_-$	$\hat{\kappa}_+$
Configuration 1: $\theta = (0, 1, 1)'$							
$\bar{\tau} = 3$	0.649	0.651	0.652	0.640	0.673	0.642	0.665
				[0.590, 0.712]	[0.621, 0.762]	[0.602, 0.692]	[0.629, 0.731]
$\bar{\tau} = 4$	0.743	0.746	0.746	0.727	0.757	0.728	0.755
				[0.672, 0.783]	[0.707, 0.857]	[0.682, 0.779]	[0.713, 0.835]
Configuration 2: $\theta = (\sqrt{1/2}, \sqrt{1/2}, 1)'$							
$\bar{\tau} = 3$	0.663	0.675	0.705	0.660	0.721	0.662	0.717
				[0.586, 0.796]	[0.651, 0.942]	[0.609, 0.764]	[0.660, 0.877]
$\bar{\tau} = 4$	0.782	0.796	0.803	0.761	0.808	0.766	0.807
				[0.688, 0.861]	[0.735, 1.037]	[0.711, 0.857]	[0.749, 0.945]
Configuration 3: $\theta = (\sqrt{2/3}, \sqrt{1/3}, 1)'$							
$\bar{\tau} = 3$	0.651	0.684	0.802	0.648	0.821	0.648	0.817
				[0.582, 0.759]	[0.742, 0.941]	[0.597, 0.739]	[0.763, 0.921]
$\bar{\tau} = 4$	0.783	0.822	0.868	0.764	0.874	0.770	0.874
				[0.685, 0.861]	[0.809, 0.972]	[0.719, 0.852]	[0.828, 0.953]

Note: Medians over 1,000 replications. The 5th and 95th percentiles are in square brackets.

new estimators are accurate at moderate sample sizes and their standard errors are reliable. As the indices approach proportionality, the cure index loses precision and its standard errors become optimistic, while the latency index is unaffected. The benchmark is precise but biased throughout, and the bias does not shrink with the sample size.

6 Extensions

6.1 Allowing for discrete covariates

The requirement that all covariates be continuous can be relaxed, provided at least two continuous covariates remain. What is needed is that the continuous covariates alone move the two indices in different directions.

Partition $X = (\mathring{X}', \ddot{X}')'$, where $\mathring{X}$ and $\ddot{X}$ denote the continuous and discrete components of X, respectively. Partition $\beta = (\mathring{\beta}', \ddot{\beta}')'$ and $\theta = (\mathring{\theta}', \ddot{\theta}')'$ analogously,

and strengthen Assumption 1 to $\mathring{\theta} \not\propto \mathring{\beta}$.

To identify $\mathring{\beta}$, write $m_\psi(\mathring{x}, \ddot{x})$ and $m_1(\mathring{x}, \ddot{x})$ for $m_\psi(x)$ and $m_1(x)$ and fix $\ddot{x}$ in the support of $\ddot{X}$. The ratio

$$\frac{m_\psi(\mathring{x}, \ddot{x})}{m_1(\mathring{x}, \ddot{x})} \equiv H(\mathring{x}'\mathring{\beta}, \ddot{x}) \tag{80}$$

then depends on $\mathring{x}$ through $\mathring{x}'\mathring{\beta}$ only, so the same weighted average derivative as before gives

$$\mathring{\beta} \propto \int \left(m_1(\mathring{x}, \ddot{x}) \frac{\partial}{\partial \mathring{x}} m_\psi(\mathring{x}, \ddot{x}) - m_\psi(\mathring{x}, \ddot{x}) \frac{\partial}{\partial \mathring{x}} m_1(\mathring{x}, \ddot{x}) \right) \mathrm{d}\mathring{x} \tag{81}$$

for each $\ddot{x}$. A weighted average over the support of $\ddot{X}$ thus identifies $\mathring{\beta}$ up to scale. The scale can be fixed by an arbitrary normalization.

For the discrete part of β, I follow Horowitz and Härdle (1996). Since the index is $\mathring{x}'\mathring{\beta} + \ddot{x}'\ddot{\beta}$, a change in $\ddot{x}$ shifts H horizontally:

$$H(\mathring{v}, \ddot{x}_2) = H(\mathring{v} + (\ddot{x}_2 - \ddot{x}_1)'\ddot{\beta}, \ddot{x}_1). \tag{82}$$

Let $\mathring{V} = \mathring{X}'\mathring{\beta}$ and let $[\mathring{v}_-, \mathring{v}_+]$ be an interval contained in the support of $\mathring{V}|\ddot{X} = \ddot{x}$ for every $\ddot{x}$ in the support of $\ddot{X}$. Moreover, let $H_- < H_+$ be such that, for every $\ddot{x}$ in the support of $\ddot{X}$, $H(\mathring{v}, \ddot{x}) \leq H_-$ for $\mathring{v} \leq \mathring{v}_-$ and $H(\mathring{v}, \ddot{x}) \geq H_+$ for $\mathring{v} \geq \mathring{v}_+$. A change of variables then gives, for any two $\ddot{x}_1$ and $\ddot{x}_2$ in the support of $\ddot{X}$,

$$\begin{aligned} &\int_{\mathring{v}_-}^{\mathring{v}_+} \max(H_-, \min(H_+, H(\mathring{v}, \ddot{x}_1)))\mathrm{d}\mathring{v} \\ &\quad - \int_{\mathring{v}_-}^{\mathring{v}_+} \max(H_-, \min(H_+, H(\mathring{v}, \ddot{x}_2)))\mathrm{d}\mathring{v} = (H_+ - H_-)(\ddot{x}_1 - \ddot{x}_2)'\ddot{\beta}, \end{aligned} \tag{83}$$

which identifies $\ddot{\beta}$.

Next, decompose $\theta = \theta_\perp + \alpha\beta$ with $\theta_\perp = (\mathring{\theta}_\perp', \ddot{\theta}_\perp')'$, where now $\mathring{\theta}_\perp \perp \mathring{\beta}$. At each $\ddot{x}$, the gradient of $m_1(\mathring{x}, \ddot{x})$ with respect to $\mathring{x}$ is a linear combination of $\mathring{\theta}$ and $\mathring{\beta}$, so

$$\mathring{\theta}_\perp \propto \left(I - \frac{\mathring{\beta}\mathring{\beta}'}{\|\mathring{\beta}\|^2} \right) \int \frac{\partial}{\partial \mathring{x}} m_1(\mathring{x}, \ddot{x}) \mathrm{d}\mathring{x} \tag{84}$$

for each $\ddot{x}$. A weighted average over the support of $\ddot{X}$ identifies $\mathring{\theta}_\perp$ up to scale, which can again be fixed by an arbitrary normalization. Let $\mathring{U}_\perp = \mathring{X}'\mathring{\theta}_\perp$. Since $X'\theta = \mathring{U}_\perp + \alpha V + \ddot{X}'\ddot{\theta}_\perp$, it holds that

$$\log E[W|\mathring{U}_\perp = \mathring{u}_\perp, V = v, \ddot{X} = \ddot{x}] = \log(1 - \pi(\mathring{u}_\perp + \alpha v + \ddot{x}'\ddot{\theta}_\perp)) + \log \tilde{F}_Y(\bar{\tau}|v). \tag{85}$$

At fixed v, the right-hand side varies with $\ddot{x}$ only through the shift $\ddot{x}'\ddot{\theta}_\perp$ in the first term, so the same clipped integral argument as for $\ddot{\beta}$ identifies $\ddot{\theta}_\perp$. Hence $\theta_\perp$ is identified.

The remaining identification steps require no further changes.

6.2 Extension to other classes of cure models

In the literature, there exist two major classes of cure models: mixture models, as used in this paper, and promotion time models. Both are nested in the Box-Cox transformation family of Yin and Ibrahim (2005),

$$S_Y(t|X) = (1 - \lambda\eta(X'\theta)F(t|X'\beta))^{1/\lambda}, \qquad \lambda \in [0,1], \tag{86}$$

where η is a positive function with $\lambda\eta(X'\theta) \leq 1$. The case $\lambda = 1$ yields the mixture model of this paper, with $\eta(x'\theta) = 1 - \pi(x'\theta)$ and $F(t|v) = \tilde{F}_Y(t|v)$. The limit $\lambda \to 0$ gives the promotion time model $S_Y(t|X) = \exp(-\eta(X'\theta)F(t|X'\beta))$ (Yakovlev and Tsodikov, 1996; Chen, Ibrahim and Sinha, 1999).

The identification strategy extends to any known λ. Since the generalized framework concerns the distribution of Y only, the censoring part is unaffected: $S_C(t|X)$ remains identified, and the weight W is defined exactly as before. This gives

$$E[W1(T \leq t)|X = x] = 1 - S_Y(t|x). \tag{87}$$

Since λ is known, this means the transformed function

$$\Gamma(t,x) = \frac{1 - S_Y(t|x)^\lambda}{\lambda} \tag{88}$$

is identified with $\Gamma(t,x) = -\log S_Y(t|x)$ for $\lambda = 0$. Substituting equation (86) into this definition gives

$$\Gamma(t,x) = \eta(x'\theta)F(t|x'\beta). \tag{89}$$

The ratio $\Gamma(t,x)/\Gamma(\bar{\tau},x) = F(t|x'\beta)/F(\bar{\tau}|x'\beta)$ depends on x through $x'\beta$ only, so its derivative with respect to x is proportional to β.

Since $\log\Gamma(\bar{\tau},x) = \log\eta(x'\theta) + \log F(\bar{\tau}|x'\beta)$ has the same additive form as in the mixture model, the remaining arguments apply verbatim to identify θ and the shapes of η and F. The only substantive change is the interval for κ, where the bound $\eta \leq 1/\lambda$ replaces $1 - \pi \leq 1$. This gives

$$\lambda \max_{u\in\mathcal{U}_0} r(u) \leq \kappa \leq 1/\max_{v\in\mathcal{V}_0} q(\bar{\tau},v). \tag{90}$$

For $\lambda = 0$, the lower bound is trivial. Moreover, the canonical promotion time model sets $\eta(u) = \exp(u)$, so that $\log \eta$ is affine and α is not identified.

When λ is unknown, the function Γ in equation (88) can still be constructed from the data for any candidate λ^*. However, provided η is not constant, $\log \Gamma(t, x)$ is additive in the two index functions at the true λ only. This suggests λ itself is identified; this is left for future work.

7 Empirical application

I illustrate the estimators on Swedish administrative data on unemployment durations. Some job seekers never return to regular work, and whether a given job seeker does is only settled at retirement. Sufficient follow-up would therefore require following everyone until then, which is far beyond the horizon of typical data.

The data come from Swedish administrative registers. The main source is the unemployment register, which covers every registration at the public employment service between 1996 and 2021 and records the start and end date of each spell together with a coded exit destination. I link it to the birth register for sex and country of birth, and to the LISA register for yearly information on educational attainment, annual earnings and municipality of residence.

I restrict the sample to Swedish-born men who entered unemployment between January 2000 and December 2015, were aged 25 to 40 at entry, and whose highest schooling is compulsory school or two-year upper secondary, according to the first two positions of the SUN 2020 code. These restrictions hold fixed the dimensions that would otherwise enter as discrete covariates. To avoid dependence between spells of the same individual, only the first spell per individual is kept.

The duration is the time from entry into unemployment until the first regular job, defined as a permanent or temporary contract or continued work for the previous employer, excluding subsidized employment. For spells that end in such a job according to the register, the duration is the time until the exit date. Spells that end for other or unknown reasons are followed further in the LISA register, and the event occurs in the first later year in which the individual is recorded as employed in November. It is dated at the exit from the register if that falls in the same calendar year and at mid-year otherwise. If a later register spell ends in a regular job, its exit date is used instead. If no employment is recorded through 2021, the individual remains at risk until the end of the window. Cure therefore means never returning to regular employment, and not merely leaving the register.

Spells are right-censored at the close of the window in December 2021 and at death and emigration. Since entries stop in December 2016, every spell is followed for at least five years, and censoring within the first five years is due to death and emigration only, which affect about 0.3 percent of spells.

The covariates are age at entry in years, the logarithm of one plus average annual earnings over the three years before entry, and the local unemployment rate at entry, defined as the share of the county labor force aged 25 to 55 that was registered as unemployed in the month of entry, in percent.

Table 5 gives summary statistics.

Table 5: Summary statistics

	Mean	Std. dev.	Median
Observed duration (days)	1,177.1	1,721.0	365
Spell outcome			
Regular job within 5 years	77.4		
Regular job after 5 years	16.6		
No regular job through 2021	5.1		
Censored before 5 years (death, emigration)	0.3		
Age at entry (years)	31.7	4.7	31.5
Prior annual earnings (100s of kr.)	1,546.5	1,007.2	1,587.7
Unemployment rate at entry (%)	13.7	3.6	13.7
Number of spells	180,856		

Note: Spell outcomes are reported as percentages.

[Work in progress.]

8 Conclusion

In this paper I consider mixture cure models of the form

$$S_Y(t|X) = \pi(X'\theta) + (1 - \pi(X'\theta))\tilde{S}_Y(t|X'\beta), \tag{91}$$

where the incidence and the latency each depend on the covariates through a linear index with an unknown link. Rather than assuming sufficient follow-up, which is implausible in economic applications, I use the covariates to separate the two components.

If θ and β are not proportional, the index vectors, the incidence and the latency are all identified up to scale. The scales of the incidence and the latency are governed by a single constant, for which I provide bounds. I propose estimators that are $\sqrt{n}$-consistent for the index vectors, and that converge at the typical nonparametric rates for the incidence, the latency and the bounds.

The results of the Monte Carlo simulations illustrate the robustness of the estimators to insufficient follow-up, in settings where estimators that impose sufficient follow-up, such as that of Amico, Van Keilegom and Legrand (2019), are biased. Not surprisingly, the precision of the estimators are strongly affected by the degree of non-proportionality between the two indices.

I show that the identification argument also covers the Box-Cox family of Yin and Ibrahim (2005), which nests the mixture and the promotion time model, as long as its transformation parameter is known. The structure of the model suggests that this parameter itself is identified as well, in which case the type of cure model can be determined from the data. I leave this for future research.

References

Abbring, J.H. (2012), "Mixed hitting-time models", *Econometrica*, 80, 783–819.

Abbring, J.H. and G.J. van den Berg (2003a), "The identifiability of the mixed proportional hazards competing risks model", *Journal of the Royal Statistical Society, Series B*, 65, 701–710.

Abbring, J.H. and G.J. van den Berg (2003b), "The nonparametric identification of treatment effects in duration models", *Econometrica*, 71, 1491–1517.

Amico, M., I. Van Keilegom, and C. Legrand (2019), "The single-index/Cox mixture cure model", *Biometrics*, 75, 452–462.

Beirlant, J., M. Bladt, and I. Van Keilegom (2026), "Nonparametric cure models through extreme-value tail estimation", *Scandinavian Journal of Statistics*, 53, 984–1000.

Chen, M.-H., J.G. Ibrahim, and D. Sinha (1999), "A new Bayesian model for survival data with a surviving fraction", *Journal of the American Statistical Association*, 94, 909–919.

Crépon, B. and G.J. van den Berg (2016), "Active labor market policies", *Annual Review of Economics*, 8, 521–546.

Donkers, B. and M. Schafgans (2008), "Specification and estimation of semiparametric multiple-index models", *Econometric Theory*, 24, 1584–1606.

Einmahl, U. and D.M. Mason (2005), "Uniform in bandwidth consistency of kernel-type function estimators", *The Annals of Statistics*, 33, 1380–1403.

Elbers, C. and G. Ridder (1982), "True and spurious duration dependence: the identifiability of the proportional hazard model", *Review of Economic Studies*, 49, 403–409.

Escobar-Bach, M., R. Maller, I. Van Keilegom, and M. Zhao (2022), "Estimation of the cure rate for distributions in the Gumbel maximum domain of attraction under insufficient follow-up", *Biometrika*, 109, 243–256.

Escobar-Bach, M. and I. Van Keilegom (2019), "Non-parametric cure rate estimation under insufficient follow-up by using extremes", *Journal of the Royal Statistical Society, Series B*, 81, 861–880.

Escobar-Bach, M. and I. Van Keilegom (2023), "Nonparametric estimation of conditional cure models for heavy-tailed distributions and under insufficient follow-up", *Computational Statistics & Data Analysis*, 183, 107728.

Fan, J., N.E. Heckman, and M.P. Wand (1995), "Local polynomial kernel regression for generalized linear models and quasi-likelihood functions", *Journal of the American Statistical Association*, 90, 141–150.

Gørgens, T. (2004), "Average derivatives for hazard functions", *Econometric Theory*, 20, 437–463.

Gørgens, T. (2006), "Semiparametric estimation of single-index hazard functions without proportional hazards", *The Econometrics Journal*, 9, 1–22.

Heckman, J.J. and B.E. Honoré (1989), "The identifiability of the competing risks model", *Biometrika*, 76, 325–330.

Heckman, J.J. and B. Singer (1984), "The identifiability of the proportional hazard model", *Review of Economic Studies*, 51, 231–241.

Horowitz, J.L. and W. Härdle (1996), "Direct semiparametric estimation of single-index models with discrete covariates", *Journal of the American Statistical Association*, 91, 1632–1640.

Hristache, M., A. Juditsky, J. Polzehl, and V. Spokoiny (2001), "Structure adaptive approach for dimension reduction", *The Annals of Statistics*, 29, 1537–1566.

Ichimura, H. and L.-F. Lee (1991), "Semiparametric least squares estimation of multiple index models: single equation estimation", in W.A. Barnett, J. Powell, and G. Tauchen (eds.), *Nonparametric and Semiparametric Methods in Econometrics and Statistics*, Cambridge University Press, Cambridge, 3–49.

Kuk, A.Y.C. and C.-H. Chen (1992), "A mixture model combining logistic regression with proportional hazards regression", *Biometrika*, 79, 531–541.

Laska, E.M. and M.J. Meisner (1992), "Nonparametric estimation and testing in a cure model", *Biometrics*, 48, 1223–1234.

Li, K.-C. (1991), "Sliced inverse regression for dimension reduction", *Journal of the American Statistical Association*, 86, 316–327.

Linton, O. and J.P. Nielsen (1995), "A kernel method of estimating structured nonparametric regression based on marginal integration", *Biometrika*, 82, 93–100.

Lopez, O., V. Patilea, and I. Van Keilegom (2013), "Single index regression models in the presence of censoring depending on the covariates", *Bernoulli*, 19, 721–747.

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

Maller, R.A. and S. Zhou (1992), "Estimating the proportion of immunes in a censored sample", *Biometrika*, 79, 731–739.

Maller, R.A. and X. Zhou (1996), *Survival Analysis with Long-Term Survivors*, Wiley, Chichester.

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

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

Powell, J.L., J.H. Stock, and T.M. Stoker (1989), "Semiparametric estimation of index coefficients", *Econometrica*, 57, 1403–1430.

Ridder, G. (1990), "The non-parametric identification of generalized accelerated failure-time models", *Review of Economic Studies*, 57, 167–181.

Schmidt, P. and A.D. Witte (1989), "Predicting criminal recidivism using 'split population' survival time models", *Journal of Econometrics*, 40, 141–159.

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

Tai, P., E. Yu, G. Cserni, G. Vlastos, M. Royce, I. Kunkler, and V. Vinh-Hung (2005), "Minimum follow-up time required for the estimation of statistical cure of cancer patients: verification using data from 42 cancer sites in the SEER database", *BMC Cancer*, 5, 48.

Taylor, J.M.G. (1995), "Semi-parametric estimation in failure time mixture models", *Biometrics*, 51, 899–907.

Tsiatis, A. (1975), "A nonidentifiability aspect of the problem of competing risks", *Proceedings of the National Academy of Sciences*, 72, 20–22.

Xia, Y., H. Tong, W.K. Li, and L.-X. Zhu (2002), "An adaptive estimation of dimension reduction space", *Journal of the Royal Statistical Society, Series B*, 64, 363–410.

Xie, P., M. Escobar-Bach, and I. Van Keilegom (2024), "Testing for sufficient follow-up in censored survival data by using extremes", *Biometrical Journal*, 66, e202400033.

Xu, J. and Y. Peng (2014), "Nonparametric cure rate estimation with covariates", *The Canadian Journal of Statistics*, 42, 1–17.

Yakovlev, A.Y. and A.D. Tsodikov (1996), *Stochastic Models of Tumor Latency and Their Biostatistical Applications*, World Scientific, Singapore.

Yin, G. and J.G. Ibrahim (2005), "Cure rate models: a unified approach", *The Canadian Journal of Statistics*, 33, 559–570.

A Variance definitions

A.1 Index vectors

The influence function of $\hat{\beta}$ in Theorem 1 is

$$\Omega_\beta(t,c,x) = \frac{c1(t \leq \bar{\tau})}{S_C(t|x)}\xi_\beta(t,x) - \Omega_C(t,c,x;\xi_\beta) - 2\beta, \tag{92}$$

where

$$\begin{aligned}\xi_\beta(t,x) = \omega(x)\frac{\partial}{\partial x}\left[m_\psi(x)f_X(x)\right] - \omega(x)\psi(t)\frac{\partial}{\partial x}\left[m_1(x)f_X(x)\right] \\ + \frac{\partial}{\partial x}\left[\omega(x)m_\psi(x)f_X(x)\right] - \psi(t)\frac{\partial}{\partial x}\left[\omega(x)m_1(x)f_X(x)\right].\end{aligned} \tag{93}$$

The influence function of $\hat{\delta}$ in Lemma 1 is

$$\Omega_\delta(t,c,x) = \frac{c1(t \leq \bar{\tau})}{S_C(t|x)}\xi_{\delta 1}(x) + \xi_{\delta 2}(x) - \Omega_C(t,c,x;\xi_{\delta 1}) - 2\delta, \tag{94}$$

where

$$\xi_{\delta 1}(x) = -\left(2\omega(x)\frac{\partial}{\partial x}f_X(x) + \omega'(x)f_X(x)\right) \tag{95}$$

and

$$\xi_{\delta 2}(x) = 2\frac{\partial}{\partial x}\left[\omega(x)m_1(x)f_X(x)\right] - \omega'(x)m_1(x)f_X(x). \tag{96}$$

Combining the two, the influence function of $\hat{\theta}_\perp$ in Corollary 1 is

$$\Omega_{\theta_\perp}(t,c,x) = \left(I - \frac{\beta\beta'}{\|\beta\|^2}\right)\Omega_\delta(t,c,x) - \frac{1}{\|\beta\|^2}\left(\beta'\delta I + \beta\delta' - \frac{2\beta'\delta}{\|\beta\|^2}\beta\beta'\right)\Omega_\beta(t,c,x). \tag{97}$$

The influence function of $\hat{\alpha}$ in Lemma 3 is

$$\Omega_\alpha(t,c,x) = \frac{\left(\frac{c1(t\leq\bar{\tau})}{S_C(t|x)} - e^{g(x'\theta_\perp,x'\beta)}\right)\xi_{\alpha 1}(x) - \Omega_C(t,c,x;\xi_{\alpha 1}) - \xi'_{\alpha 2}\Omega_{\theta_\perp}(t,c,x) - \xi'_{\alpha 3}\Omega_\beta(t,c,x)}{\int \frac{\partial^2}{\partial u_\perp^2}\rho(u_\perp,v)g(u_\perp,v)\mathrm{d}u_\perp\mathrm{d}v}, \tag{98}$$

where $f_{U_\perp V}$ denotes the density of $(U_\perp, V)$ given $X \in \mathcal{X}_0$,

$$\xi_{\alpha 1}(x) = \frac{\left(\frac{\partial^2}{\partial u_\perp \partial v}\rho(x'\theta_\perp, x'\beta) - \alpha\frac{\partial^2}{\partial u_\perp^2}\rho(x'\theta_\perp, x'\beta)\right)1(x \in \mathcal{X}_0)}{P(X \in \mathcal{X}_0)e^{g(x'\theta_\perp,x'\beta)}f_{U_\perp V}(x'\theta_\perp, x'\beta)}, \tag{99}$$

$$\xi_{\alpha 2} = \int\left(\frac{\partial^2}{\partial u_\perp \partial v}\rho(u_\perp,v) - \alpha\frac{\partial^2}{\partial u_\perp^2}\rho(u_\perp,v)\right)g_1(u_\perp,v)E[X|U_\perp = u_\perp, V = v, X \in \mathcal{X}_0]\mathrm{d}u_\perp\mathrm{d}v, \tag{100}$$

$$\xi_{\alpha 3} = \int\left(\frac{\partial^2}{\partial u_\perp \partial v}\rho(u_\perp,v) - \alpha\frac{\partial^2}{\partial u_\perp^2}\rho(u_\perp,v)\right)g_2(u_\perp,v)E[X|U_\perp = u_\perp, V = v, X \in \mathcal{X}_0]\mathrm{d}u_\perp\mathrm{d}v, \tag{101}$$

and g_1 and g_2 denote the derivatives of g with respect to its first and second argument, respectively. Finally, the influence function of $\hat{\theta}$ in Theorem 1 is

$$\Omega_\theta(t,c,x) = \Omega_{\theta_\perp}(t,c,x) + \alpha\Omega_\beta(t,c,x) + \Omega_\alpha(t,c,x)\beta. \tag{102}$$

Each influence function can be estimated uniformly consistently by replacing unknown quantities with sample analogs using standard methods.

A.2 Function estimators

The leading term in Lemma 4 is

$$L_G(t,u_\perp,v) = \frac{1}{nh_S^2 P(X\in\mathcal{X}_0) f_{U_\perp V}(u_\perp,v)} \sum_{i=1}^{n} \left(W_i 1(T_i\le t) - G(t, X_i'\theta_\perp, X_i'\beta)\right) \\ \times K_S\left(\frac{U_{\perp i}-u_\perp}{h_S}\right) K_S\left(\frac{V_i - v}{h_S}\right) 1(X_i\in\mathcal{X}_0). \tag{103}$$

For the survival function in Theorem 2, the leading term and the asymptotic variance are

$$L_S(t,x) = -L_G(t, x'\theta_\perp, x'\beta) \tag{104}$$

and

$$\sigma_S^2(t,x) = \left(\int K_S(s)^2 \mathrm{d}s\right)^2 \frac{\mathrm{Var}(W1(T\le t)|U_\perp = x'\theta_\perp, V = x'\beta, X\in\mathcal{X}_0)}{P(X\in\mathcal{X}_0) f_{U_\perp V}(x'\theta_\perp, x'\beta)}. \tag{105}$$

For the incidence, they are

$$L_r(u) = \frac{1}{|\mathcal{V}_0|}\int_{\mathcal{V}_0} L_G(\bar\tau, u-\alpha v, v)\mathrm{d}v \tag{106}$$

and

$$\sigma_r^2(u) = \frac{1}{|\mathcal{V}_0|^2}\int \left(\int K_S(z+\alpha s)K_S(s)\mathrm{d}s\right)^2 \mathrm{d}z \\ \times \int_{\mathcal{V}_0} \frac{\mathrm{Var}(W1(T\le\bar\tau)|U_\perp = u-\alpha v, V = v, X\in\mathcal{X}_0)}{P(X\in\mathcal{X}_0) f_{U_\perp V}(u-\alpha v, v)}\mathrm{d}v. \tag{107}$$

For the latency, they are

$$L_q(t,v) = \frac{\int_{\mathcal{U}_0} L_G(t, u-\alpha v, v)\mathrm{d}u}{\int_{\mathcal{U}_0} r(u)\mathrm{d}u} \tag{108}$$

and

$$\sigma_q^2(t,v) = \frac{\int K_S(z)^2\mathrm{d}z}{\left(\int_{\mathcal{U}_0} r(u)\mathrm{d}u\right)^2}\int_{\mathcal{U}_0} \frac{\mathrm{Var}(W1(T\le t)|U_\perp = u-\alpha v, V = v, X\in\mathcal{X}_0)}{P(X\in\mathcal{X}_0) f_{U_\perp V}(u-\alpha v, v)}\mathrm{d}u. \tag{109}$$

The variances can be estimated consistently by replacing unknown quantities with sample analogs using standard methods.

A.3 Test statistics

The variance in J_1 only involves $\Omega_{\theta_\perp}$, and the variance in J_3 only involves Ω_β and Ω_{β^*}. For J_2, let Ω_{δ_l} denote Ω_δ with ω_l in place of ω. The influence function of $\hat e_l$ in Lemma 5 is

$$\Omega_{e_l}(t,c,x) = \frac{c1(t\le\bar\tau)}{S_C(t|x)}\omega_l(x) f_X(x) + \omega_l(x) m_1(x) f_X(x) - \Omega_C(t,c,x;\xi_{e_l}) - 2e_l, \tag{110}$$

where

$$\xi_{e_l}(x) = \omega_l(x) f_X(x). \tag{111}$$

The influence function of $\hat{R}_l$ is

$$\Omega_{R_l}(t,c,x) = (e_1\delta_l - 2e_l\theta_\perp)' \Omega_{\theta_\perp}(t,c,x) + e_1\theta_\perp' \Omega_{\delta_l}(t,c,x) + \theta_\perp'\delta_l \Omega_{e_1}(t,c,x) - \|\theta_\perp\|^2 \Omega_{e_l}(t,c,x), \tag{112}$$

and that of $\hat{R}$ is

$$\Omega_R(t,c,x) = (\Omega_{R_2}(t,c,x), \ldots, \Omega_{R_L}(t,c,x))'. \tag{113}$$

As for the index vectors, these influence functions can be estimated uniformly consistently by replacing unknown quantities with sample analogs, which yields consistent estimators of $\Sigma_{\theta_\perp}$, Σ_R and Σ_{β^*}.

B Proofs

Assumption C (Single-index censoring).

(a) There exists γ such that $S_C(t|x) = S_0(t|x'\gamma)$ for all $t \in [0, \bar{\tau}]$ and $x \in \mathcal{X}$.

(b) Let $Z = X'\gamma$. There exist an open set Γ containing γ and an open set $\mathcal{Z}_0$ containing $\{x'\gamma^* : x \in \mathcal{X}_0^+, \gamma^* \in \Gamma\}$ such that the density of $X'\gamma^*$ at z and the functions $P(T \le t, Y > C|X'\gamma^* = z)$ and $P(T \ge t|X'\gamma^* = z)$ are twice differentiable in (γ^*, z) with bounded derivatives on $\Gamma \times \mathcal{Z}_0$. Moreover, the density of $X'\gamma^*$ at z and $P(T \ge \bar{\tau}|X'\gamma^* = z)$ are bounded away from zero on $\Gamma \times \mathcal{Z}_0$, and the conditional distribution of $(T, 1(Y \le C), X)$ given $Z = z$ is twice differentiable in z with bounded derivatives on $\mathcal{Z}_0$.

(c) There exists an estimator $\hat{\gamma}$ of γ that satisfies

$$\hat{\gamma} - \gamma = \frac{1}{n}\sum_{i=1}^{n} \Omega_\gamma(T_i, 1(Y_i \le C_i), X_i) + o_p(n^{-1/2}), \tag{114}$$

where Ω_γ is bounded and $\Omega_\gamma(T, 1(Y \le C), X)$ has mean zero. Moreover, there exists an estimator $\hat{\Omega}_\gamma$ of Ω_γ such that $\sup_{t,c,x} \|\hat{\Omega}_\gamma(t, c, x) - \Omega_\gamma(t, c, x)\| \xrightarrow{p} 0$.

(d) The kernel K_C is nonnegative, symmetric, twice continuously differentiable, and has compact support. It satisfies $\int K_C(\nu)\mathrm{d}\nu = 1$. The bandwidth satisfies $h_C \propto n^{-a}$ for some $a \in (1/4, 1/3)$.

Proposition 5 (Single-index censoring). *Suppose Assumption 4 and Assumption C hold. Then Assumption 10 is satisfied for any $\mu \in (1/4, (1-a)/2)$ and*

$$\hat{S}_C(t|x) = \prod_{i:T_i \le t} \left(1 - \frac{1(Y_i > C_i)K_C((X_i'\hat{\gamma} - x'\hat{\gamma})/h_C)}{\sum_{j=1}^n 1(T_j \ge T_i)K_C((X_j'\hat{\gamma} - x'\hat{\gamma})/h_C)}\right). \tag{115}$$

Proof. By Assumptions 4 and C, the conditions of Lopez, Patilea and Van Keilegom (2013). Part (a) follows from equation (A.10) in their Lemma A.2.

For part (b), let ξ be as in Assumption 10b. Part (a) gives

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n} \frac{1(Y_i \le C_i)1(T_i \le \bar{\tau})}{S_C(T_i|X_i)^2}\xi(T_i, X_i)(\hat{S}_C(T_i|X_i) - S_C(T_i|X_i)) \\
&\quad = -\frac{1}{n}\sum_{i=1}^{n} 1(Y_i \le C_i)1(T_i \le \bar{\tau})\xi(T_i, X_i)\left(\frac{1}{\hat{S}_C(T_i|X_i)} - \frac{1}{S_C(T_i|X_i)}\right) + o_p(n^{-1/2}).
\end{aligned} \tag{116}$$

Theorems 3.1(ii) and 3.2(ii) of Lopez, Patilea and Van Keilegom (2013) apply to this sum and give part (b) with

$$\begin{aligned}
\Omega_C(t, c, x; \xi) = &-\frac{(1-c)E[W\xi(T, X)1(T \ge t)|Z = x'\gamma]}{P(T \ge t|Z = x'\gamma)} \\
&- \int_0^t \frac{E[W\xi(T, X)1(T \ge s)|Z = x'\gamma]}{P(T \ge s|Z = x'\gamma)}\mathrm{d}\log S_C(s|x) \\
&- E[W\xi(T, X)\Lambda_\gamma(T, X)]'\Omega_\gamma(t, c, x),
\end{aligned} \tag{117}$$

where

$$\Lambda_\gamma(t,x) = \frac{\partial}{\partial \gamma^*} \int_0^t \frac{\frac{\partial}{\partial s} P(T \leq s, Y > C | X'\gamma^* = x'\gamma^*)}{P(T \geq s | X'\gamma^* = x'\gamma^*)} \mathrm{d}s \Big|_{\gamma^* = \gamma}. \tag{118}$$

For part (c), estimate Ω_C by replacing the conditional expectations given Z with kernel regressions on $X_i'\hat{\gamma}$, $\Lambda_\gamma(t,x)$ with the gradient of $-\log \hat{S}_C(t|x)$ with respect to $\hat{\gamma}$, and the remaining unknown quantities with sample analogs. Uniform consistency follows from Lemma A.1 and equations (A.8) to (A.10) of Lopez, Patilea and Van Keilegom (2013) together with Assumption C. □

Proof of Theorem 1a. Define

$$p_n(X_i, X_j) = \frac{1}{2} \left(\omega(X_i) + \omega(X_j)\right) K_\beta' \left(\frac{X_i - X_j}{h_\beta} \right).$$

Decompose

$$\hat{W}_i \hat{W}_j = W_i W_j + (\hat{W}_i - W_i) W_j + W_i (\hat{W}_j - W_j) + (\hat{W}_i - W_i)(\hat{W}_j - W_j) \tag{119}$$

and note that

$$\hat{W}_i - W_i = -\frac{W_i (\hat{S}_C(T_i|X_i) - S_C(T_i|X_i))}{S_C(T_i|X_i)} + \frac{W_i (\hat{S}_C(T_i|X_i) - S_C(T_i|X_i))^2}{S_C(T_i|X_i) \hat{S}_C(T_i|X_i)}. \tag{120}$$

Use these identities to write

$$\hat{\beta} = \frac{n-1}{n} (\hat{\beta}_0 - \hat{\beta}_1 + \hat{\beta}_2), \tag{121}$$

where

$$\hat{\beta}_0 = \binom{n}{2}^{-1} \sum_{i<j} \frac{1}{h_\beta^{d+1}} W_i W_j \left(\psi(T_j) - \psi(T_i)\right) p_n(X_i, X_j), \tag{122}$$

$$\begin{aligned} \hat{\beta}_1 = \binom{n}{2}^{-1} \sum_{i<j} \frac{1}{h_\beta^{d+1}} & \left(\frac{\hat{S}_C(T_i|X_i) - S_C(T_i|X_i)}{S_C(T_i|X_i)} + \frac{\hat{S}_C(T_j|X_j) - S_C(T_j|X_j)}{S_C(T_j|X_j)} \right) \\ & \times W_i W_j \left(\psi(T_j) - \psi(T_i)\right) p_n(X_i, X_j), \end{aligned} \tag{123}$$

and

$$\begin{aligned} \hat{\beta}_2 = \binom{n}{2}^{-1} \sum_{i<j} \frac{1}{h_\beta^{d+1}} & \left(\frac{(\hat{S}_C(T_i|X_i) - S_C(T_i|X_i))^2}{S_C(T_i|X_i) \hat{S}_C(T_i|X_i)} + \frac{(\hat{S}_C(T_j|X_j) - S_C(T_j|X_j))^2}{S_C(T_j|X_j) \hat{S}_C(T_j|X_j)} \right. \\ & \left. + \frac{(\hat{S}_C(T_i|X_i) - S_C(T_i|X_i))(\hat{S}_C(T_j|X_j) - S_C(T_j|X_j))}{\hat{S}_C(T_i|X_i) \hat{S}_C(T_j|X_j)} \right) \\ & \times W_i W_j \left(\psi(T_j) - \psi(T_i)\right) p_n(X_i, X_j). \end{aligned} \tag{124}$$

The three terms are handled in turn. Consider $\hat{\beta}_0$. As

$$E[(h_\beta^{-(d+1)} p_n(X_i, X_j))^2] = O(h_\beta^{-(d+2)}) = o(n), \tag{125}$$

by Lemma 3.1 of Powell, Stock and Stoker (1989) it holds that

$$\begin{aligned}\hat{\beta}_0 = \frac{1}{n}\sum_{i=1}^{n} \frac{2}{h_\beta^{d+1}} E[W_i W_j \left(\psi(T_j) - \psi(T_i)\right) p_n(X_i, X_j) | T_i, 1(Y_i \leq C_i), X_i] \\ - \frac{1}{h_\beta^{d+1}} E[W_i W_j \left(\psi(T_j) - \psi(T_i)\right) p_n(X_i, X_j)] + o_p(n^{-1/2}).\end{aligned} \tag{126}$$

The summand in equation (126) can be expressed as

$$\frac{2W_i}{h_\beta^{d+1}} \int m_\psi(\zeta) f_X(\zeta) p_n(X_i, \zeta) \mathrm{d}\zeta - \frac{2W_i \psi(T_i)}{h_\beta^{d+1}} \int m_1(\zeta) f_X(\zeta) p_n(X_i, \zeta) \mathrm{d}\zeta. \tag{127}$$

Substitute $\zeta = x - h_\beta \nu$ to obtain

$$\frac{2}{h_\beta^{d+1}} \int m_\psi(\zeta) f_X(\zeta) p_n(x, \zeta) \mathrm{d}\zeta = \frac{1}{h_\beta} \int (\omega(x) + \omega(x - h_\beta \nu)) m_\psi(x - h_\beta \nu) f_X(x - h_\beta \nu) K'_\beta(\nu) \mathrm{d}\nu. \tag{128}$$

Expand the integrand in a Taylor series about $h_\beta = 0$ through order p_β and use Assumption 12 to obtain

$$\frac{2}{h_\beta^{d+1}} \int m_\psi(\zeta) f_X(\zeta) p_n(x, \zeta) \mathrm{d}\zeta = \omega(x) \frac{\partial}{\partial x} \left[m_\psi(x) f_X(x)\right] + \frac{\partial}{\partial x} \left[\omega(x) m_\psi(x) f_X(x)\right] + O(h_\beta^{p_\beta}) \tag{129}$$

uniformly over $x \in \mathcal{X}$. Identical steps yield

$$\frac{2}{h_\beta^{d+1}} \int m_1(\zeta) f_X(\zeta) p_n(x, \zeta) \mathrm{d}\zeta = \omega(x) \frac{\partial}{\partial x} \left[m_1(x) f_X(x)\right] + \frac{\partial}{\partial x} \left[\omega(x) m_1(x) f_X(x)\right] + O(h_\beta^{p_\beta}) \tag{130}$$

uniformly over $x \in \mathcal{X}$. Hence

$$\frac{2}{h_\beta^{d+1}} E[W_i W_j \left(\psi(T_j) - \psi(T_i)\right) p_n(X_i, X_j) | T_i, 1(Y_i \leq C_i), X_i] = W_i \xi_\beta(T_i, X_i) + O(h_\beta^{p_\beta}). \tag{131}$$

Moreover,

$$E[W \xi_\beta(T, X)] = 2 \int \omega(x) f_X(x)^2 \left(m_1(x) \frac{\partial}{\partial x} m_\psi(x) - m_\psi(x) \frac{\partial}{\partial x} m_1(x) \right) \mathrm{d}x = 2\beta. \tag{132}$$

Since $n h_\beta^{2p_\beta} \to 0$ by Assumption 13, combining equations (126) and (132) yields

$$\hat{\beta}_0 - \beta = \frac{1}{n} \sum_{i=1}^{n} \left(W_i \xi_\beta(T_i, X_i) - 2\beta\right) + o_p(n^{-1/2}). \tag{133}$$

Next, consider $\hat{\beta}_1$. Rewrite the summand to

$$\hat{\beta}_1 = \frac{1}{n} \sum_{i=1}^{n} \frac{\hat{S}_C(T_i | X_i) - S_C(T_i | X_i)}{S_C(T_i | X_i)} W_i \hat{\xi}_{\beta 1 i}, \tag{134}$$

where

$$\hat{\xi}_{\beta 1 i} = \frac{2}{(n-1) h_\beta^{d+1}} \sum_{j \neq i} W_j \left(\psi(T_j) - \psi(T_i)\right) p_n(X_i, X_j). \tag{135}$$

Note that

$$E[W_i\hat{\xi}_{\beta 1i}|T_i, 1(Y_i \leq C_i), X_i] = W_i\xi_\beta(T_i, X_i) + O(h_\beta^{p_\beta}). \tag{136}$$

By Bernstein's inequality and a union bound,

$$\max_{i\in\mathcal{I}} \left| W_i\hat{\xi}_{\beta 1i} - E[W_i\hat{\xi}_{\beta 1i}|T_i, 1(Y_i \leq C_i), X_i] \right| = O_p((\log n/(nh_\beta^{d+2}))^{1/2}) \tag{137}$$

where $\mathcal{I} = \{i : (T_i, X_i) \in [0, \bar{\tau}] \times \mathcal{X}_0^+\}$.

Substitute equation (137) into (134) and use Assumptions 10 and 13 to obtain

$$\hat{\beta}_1 = \frac{1}{n}\sum_{i=1}^n \frac{\hat{S}_C(T_i|X_i) - S_C(T_i|X_i)}{S_C(T_i|X_i)} W_i\xi_\beta(T_i, X_i) + o_p(n^{-1/2}). \tag{138}$$

Another application of Assumption 10 now yields

$$\hat{\beta}_1 = \frac{1}{n}\sum_{i=1}^n \Omega_C(T_i, 1(Y_i \leq C_i), X_i; \xi_\beta) + o_p(n^{-1/2}). \tag{139}$$

Finally, consider $\hat{\beta}_2$. Since $S_C(T_i|X_i) \geq \varepsilon$ and $\hat{S}_C(t|x) \xrightarrow{p} S_C(t|x)$ uniformly over $[0, \bar{\tau}] \times \mathcal{X}_0^+$, by Assumption 10 it holds that

$$\max_{i\in\mathcal{I}} \left| \frac{(\hat{S}_C(T_i|X_i) - S_C(T_i|X_i))^2}{S_C(T_i|X_i)\hat{S}_C(T_i|X_i)} \right| = O_p(n^{-2\mu}) \tag{140}$$

and

$$\max_{i,j\in\mathcal{I}} \left| \frac{(\hat{S}_C(T_i|X_i) - S_C(T_i|X_i))(\hat{S}_C(T_j|X_j) - S_C(T_j|X_j))}{\hat{S}_C(T_i|X_i)\hat{S}_C(T_j|X_j)} \right| = O_p(n^{-2\mu}). \tag{141}$$

Since W and ψ are bounded and $E[|p_n(X_i, X_j)|] = O(h_\beta^d)$, it follows that

$$|\hat{\beta}_2| = O_p\left(n^{-2\mu}\binom{n}{2}^{-1} \sum_{i<j} \frac{1}{h_\beta^{d+1}} W_iW_j \left|\psi(T_j) - \psi(T_i)\right| |p_n(X_i, X_j)| \right) = O_p(h_\beta^{-1}n^{-2\mu}) \tag{142}$$

which is in $o_p(n^{-1/2})$ by Assumption 13. The proof follows by combining the results on $\hat{\beta}_0$, $\hat{\beta}_1$ and $\hat{\beta}_2$. □

Lemma 1. *Under Assumptions 1–13, as $n \to \infty$, it holds that*

$$\hat{\delta} - \delta = \frac{1}{n}\sum_{i=1}^n \Omega_\delta(T_i, 1(Y_i \leq C_i), X_i) + o_p(n^{-1/2}). \tag{143}$$

The function $\Omega_\delta(T, 1(Y \leq C), X)$ has mean zero and finite variance.

The proof of Lemma 1 is similar to that of Theorem 1a and therefore omitted.

Corollary 1. *Under Assumptions 1–13, as $n \to \infty$, it holds that*

$$\hat{\theta}_\perp - \theta_\perp = \frac{1}{n}\sum_{i=1}^{n} \Omega_{\theta_\perp}(T_i, 1(Y_i \le C_i), X_i) + o_p(n^{-1/2}). \tag{144}$$

The function $\Omega_{\theta_\perp}(T, 1(Y \le C), X)$ has mean zero and finite variance.

Proof of Corollary 1. Decompose

$$\hat{\theta}_\perp - \theta_\perp = \left(I - \frac{\beta\beta'}{\|\beta\|^2}\right)(\hat{\delta} - \delta) + \left(\frac{\beta\beta'}{\|\beta\|^2} - \frac{\hat{\beta}\hat{\beta}'}{\|\hat{\beta}\|^2}\right)\delta + \left(\frac{\beta\beta'}{\|\beta\|^2} - \frac{\hat{\beta}\hat{\beta}'}{\|\hat{\beta}\|^2}\right)(\hat{\delta} - \delta). \tag{145}$$

A first-order expansion of $b \mapsto bb'\delta/\|b\|^2$ around β gives

$$\left(\frac{\beta\beta'}{\|\beta\|^2} - \frac{\hat{\beta}\hat{\beta}'}{\|\hat{\beta}\|^2}\right)\delta = -\frac{1}{\|\beta\|^2}\left(\beta'\delta I + \beta\delta' - \frac{2\beta'\delta}{\|\beta\|^2}\beta\beta'\right)(\hat{\beta} - \beta) + O_p(\|\hat{\beta} - \beta\|^2). \tag{146}$$

The third term is $O_p(\|\hat{\beta} - \beta\|\|\hat{\delta} - \delta\|)$. The result follows from Theorem 1a and Lemma 1. □

Lemma 2. *Let $\chi : \mathbb{R}^2 \to \mathbb{R}$ be a known twice continuously differentiable function that is zero outside $\mathcal{R}$. Under Assumptions 1–13, as $n \to \infty$, it holds that*

$$\begin{aligned}
&\int \chi(u_\perp, v)\left(\hat{g}_0(u_\perp, v) - g(u_\perp, v)\right)\mathrm{d}u_\perp \mathrm{d}v \\
&= \frac{1}{n}\sum_{i=1}^{n}\left(W_i - e^{g(U_{\perp i}, V_i)}\right)\frac{\chi(U_{\perp i}, V_i)1(X_i \in \mathcal{X}_0)}{P(X \in \mathcal{X}_0)e^{g(U_{\perp i}, V_i)}f_{U_\perp V}(U_{\perp i}, V_i)} \\
&\quad + \frac{h_\alpha^2}{2}\int s^2 K_\alpha(s)\mathrm{d}s \int \chi(u_\perp, v)\left(g_{11}(u_\perp, v) + g_{22}(u_\perp, v)\right)\mathrm{d}u_\perp \mathrm{d}v \\
&\quad - \left(\int \chi(u_\perp, v) g_1(u_\perp, v) E[X | U_\perp = u_\perp, V = v, X \in \mathcal{X}_0]\mathrm{d}u_\perp \mathrm{d}v\right)'(\hat{\theta}_\perp - \theta_\perp) \\
&\quad - \left(\int \chi(u_\perp, v) g_2(u_\perp, v) E[X | U_\perp = u_\perp, V = v, X \in \mathcal{X}_0]\mathrm{d}u_\perp \mathrm{d}v\right)'(\hat{\beta} - \beta) \\
&\quad - \frac{1}{n}\sum_{i=1}^{n}\Omega_C(T_i, 1(Y_i \le C_i), X_i; \xi_\chi) + o_p(n^{-1/2}),
\end{aligned} \tag{147}$$

where

$$\xi_\chi(x) = \chi(x'\theta_\perp, x'\beta)1(x \in \mathcal{X}_0)/(P(X \in \mathcal{X}_0)e^{g(x'\theta_\perp, x'\beta)}f_{U_\perp V}(x'\theta_\perp, x'\beta)). \tag{148}$$

Proof of Lemma 2. Let $\check{g}$ denote $\hat{g}$ with W_i in place of $\hat{W}_i$, and $\tilde{g}$ with W_i, $U_{\perp i}$ and V_i in place of $\hat{W}_i$, $\hat{U}_{\perp i}$ and $\hat{V}_i$. Decompose

$$\begin{aligned}
\int \chi(u_\perp, v)(\hat{g}_0(u_\perp, v) - g(u_\perp, v))\mathrm{d}u_\perp \mathrm{d}v = &\int \chi(u_\perp, v)(\tilde{g}_0(u_\perp, v) - g(u_\perp, v))\mathrm{d}u_\perp \mathrm{d}v \\
&+ \int \chi(u_\perp, v)(\check{g}_0(u_\perp, v) - \tilde{g}_0(u_\perp, v))\mathrm{d}u_\perp \mathrm{d}v \\
&+ \int \chi(u_\perp, v)(\hat{g}_0(u_\perp, v) - \check{g}_0(u_\perp, v))\mathrm{d}u_\perp \mathrm{d}v.
\end{aligned} \tag{149}$$

The estimators $\hat{g}(u_\perp, v)$, $\check{g}(u_\perp, v)$ and $\tilde{g}(u_\perp, v)$ solve the first-order conditions

$$\begin{aligned}\sum_{i=1}^{n} &\left(\hat{W}_i - e^{(1,\hat{U}_{\perp i}-u_\perp,\hat{V}_i-v)b}\right)(1, \hat{U}_{\perp i}-u_\perp, \hat{V}_i - v)' \\ &\times K_\alpha\left(\frac{\hat{U}_{\perp i}-u_\perp}{h_\alpha}\right) K_\alpha\left(\frac{\hat{V}_i - v}{h_\alpha}\right) 1(X_i \in \mathcal{X}_0) = 0,\end{aligned} \tag{150}$$

$$\begin{aligned}\sum_{i=1}^{n} &\left(W_i - e^{(1,\hat{U}_{\perp i}-u_\perp,\hat{V}_i-v)b}\right)(1, \hat{U}_{\perp i}-u_\perp, \hat{V}_i - v)' \\ &\times K_\alpha\left(\frac{\hat{U}_{\perp i}-u_\perp}{h_\alpha}\right) K_\alpha\left(\frac{\hat{V}_i - v}{h_\alpha}\right) 1(X_i \in \mathcal{X}_0) = 0,\end{aligned} \tag{151}$$

and

$$\begin{aligned}\sum_{i=1}^{n} &\left(W_i - e^{(1,U_{\perp i}-u_\perp,V_i-v)b}\right)(1, U_{\perp i}-u_\perp, V_i - v)' \\ &\times K_\alpha\left(\frac{U_{\perp i}-u_\perp}{h_\alpha}\right) K_\alpha\left(\frac{V_i - v}{h_\alpha}\right) 1(X_i \in \mathcal{X}_0) = 0,\end{aligned} \tag{152}$$

respectively.

Consider the first term on the right-hand side of equation (149). Expand the first-order condition for $\tilde{g}$ around $b = (g(u_\perp, v), g_1(u_\perp, v), g_2(u_\perp, v))'$ and solve for the intercept to obtain

$$\begin{aligned}&\int \chi(u_\perp, v)(\tilde{g}_0(u_\perp, v) - g(u_\perp, v))\mathrm{d}u_\perp \mathrm{d}v \\ &= \int \frac{\chi(u_\perp, v)}{nh_\alpha^2 P(X \in \mathcal{X}_0) e^{g(u_\perp,v)} f_{U_\perp V}(u_\perp, v)} \sum_{i=1}^{n} \left(W_i - e^{g(U_{\perp i}, V_i)}\right) \\ &\quad \times K_\alpha\left(\frac{U_{\perp i}-u_\perp}{h_\alpha}\right) K_\alpha\left(\frac{V_i - v}{h_\alpha}\right) 1(X_i \in \mathcal{X}_0)\mathrm{d}u_\perp \mathrm{d}v \\ &\quad + \int \frac{\chi(u_\perp, v)}{nh_\alpha^2 P(X \in \mathcal{X}_0) e^{g(u_\perp,v)} f_{U_\perp V}(u_\perp, v)} \sum_{i=1}^{n} \left(e^{g(U_{\perp i}, V_i)} - e^{g(u_\perp,v) + g_1(u_\perp,v)(U_{\perp i}-u_\perp) + g_2(u_\perp,v)(V_i - v)}\right) \\ &\quad \times K_\alpha\left(\frac{U_{\perp i}-u_\perp}{h_\alpha}\right) K_\alpha\left(\frac{V_i - v}{h_\alpha}\right) 1(X_i \in \mathcal{X}_0)\mathrm{d}u_\perp \mathrm{d}v + o_p(n^{-1/2}).\end{aligned} \tag{153}$$

Substitute $u_\perp = U_{\perp i} - h_\alpha s_1$ and $v = V_i - h_\alpha s_2$ and expand around $h_\alpha = 0$ to write the first term on the right-hand side of equation (153) as

$$\frac{1}{n}\sum_{i=1}^{n} \left(W_i - e^{g(U_{\perp i}, V_i)}\right) \frac{\chi(U_{\perp i}, V_i) 1(X_i \in \mathcal{X}_0)}{P(X \in \mathcal{X}_0) e^{g(U_{\perp i}, V_i)} f_{U_\perp V}(U_{\perp i}, V_i)} + o_p(n^{-1/2}). \tag{154}$$

Similarly, through a Taylor expansion of g around $(u_\perp, v)$, the second term on the right-hand side of equation (153) equals

$$\frac{h_\alpha^2}{2} \int s^2 K_\alpha(s)\mathrm{d}s \int \chi(u_\perp, v)(g_{11}(u_\perp, v) + g_{22}(u_\perp, v))\mathrm{d}u_\perp \mathrm{d}v + o_p(n^{-1/2}). \tag{155}$$

Next, consider the second term on the right-hand side of equation (149). Since K_α is Lipschitz and $\hat{\theta}_\perp - \theta_\perp$ and $\hat{\beta} - \beta$ are $O_p(n^{-1/2})$ by Theorem 1a and Corollary 1, by an expansion of the first-order condition of $\check{g}$ around $\tilde{g}$ it equals

$$\begin{aligned}
&\int \frac{\chi(u_\perp, v)}{nh_\alpha^2 P(X \in \mathcal{X}_0) e^{g(u_\perp, v)} f_{U_\perp V}(u_\perp, v)} \sum_{i=1}^n 1(X_i \in \mathcal{X}_0) \\
&\quad \times \left(\left(W_i - e^{g(\hat{U}_{\perp i}, \hat{V}_i)} \right) K_\alpha \left(\frac{\hat{U}_{\perp i} - u_\perp}{h_\alpha} \right) K_\alpha \left(\frac{\hat{V}_i - v}{h_\alpha} \right) \right. \\
&\quad \left. - \left(W_i - e^{g(U_{\perp i}, V_i)} \right) K_\alpha \left(\frac{U_{\perp i} - u_\perp}{h_\alpha} \right) K_\alpha \left(\frac{V_i - v}{h_\alpha} \right) \right) \mathrm{d}u_\perp \mathrm{d}v + o_p(n^{-1/2}).
\end{aligned} \tag{156}$$

Substitute $u_\perp = \hat{U}_{\perp i} - h_\alpha s_1$ and $v = \hat{V}_i - h_\alpha s_2$ and expand around $h_\alpha = 0$ and around $(\hat{U}_{\perp i}, \hat{V}_i) = (U_{\perp i}, V_i)$ to write the right-hand side as

$$\begin{aligned}
&\frac{1}{n} \sum_{i=1}^n \left(W_i - e^{g(U_{\perp i}, V_i)} \right) 1(X_i \in \mathcal{X}_0) \frac{\partial}{\partial u_\perp} \left[\frac{\chi(u_\perp, v)}{P(X \in \mathcal{X}_0) e^{g(u_\perp, v)} f_{U_\perp V}(u_\perp, v)} \right]_{(u_\perp, v) = (U_{\perp i}, V_i)} X_i'(\hat{\theta}_\perp - \theta_\perp) \\
&\quad + \frac{1}{n} \sum_{i=1}^n \left(W_i - e^{g(U_{\perp i}, V_i)} \right) 1(X_i \in \mathcal{X}_0) \frac{\partial}{\partial v} \left[\frac{\chi(u_\perp, v)}{P(X \in \mathcal{X}_0) e^{g(u_\perp, v)} f_{U_\perp V}(u_\perp, v)} \right]_{(u_\perp, v) = (U_{\perp i}, V_i)} X_i'(\hat{\beta} - \beta) \\
&\quad - \frac{1}{n} \sum_{i=1}^n \frac{\chi(U_{\perp i}, V_i) 1(X_i \in \mathcal{X}_0)}{P(X \in \mathcal{X}_0) f_{U_\perp V}(U_{\perp i}, V_i)} \left(g_1(U_{\perp i}, V_i) X_i'(\hat{\theta}_\perp - \theta_\perp) + g_2(U_{\perp i}, V_i) X_i'(\hat{\beta} - \beta) \right) \\
&\quad + o_p(n^{-1/2}).
\end{aligned} \tag{157}$$

The first two terms are $O_p(n^{-1})$. The third term equals

$$\begin{aligned}
&- \left(\int \chi(u_\perp, v) g_1(u_\perp, v) E[X | U_\perp = u_\perp, V = v, X \in \mathcal{X}_0] \mathrm{d}u_\perp \mathrm{d}v \right)' (\hat{\theta}_\perp - \theta_\perp) \\
&\quad - \left(\int \chi(u_\perp, v) g_2(u_\perp, v) E[X | U_\perp = u_\perp, V = v, X \in \mathcal{X}_0] \mathrm{d}u_\perp \mathrm{d}v \right)' (\hat{\beta} - \beta) \\
&\quad + o_p(n^{-1/2}).
\end{aligned} \tag{158}$$

Finally, consider the third term on the right-hand side of equation (149). Subtract the first-order conditions for $\hat{g}(u_\perp, v)$ and $\check{g}(u_\perp, v)$ and solve for the intercept to write this term as

$$\begin{aligned}
&\int \chi(u_\perp, v) (\hat{g}_0(u_\perp, v) - \check{g}_0(u_\perp, v)) \mathrm{d}u_\perp \mathrm{d}v \\
&\quad = \int \frac{\chi(u_\perp, v)}{nh_\alpha^2 P(X \in \mathcal{X}_0) e^{g(u_\perp, v)} f_{U_\perp V}(u_\perp, v)} \sum_{i=1}^n (\hat{W}_i - W_i) \\
&\quad \times K_\alpha \left(\frac{\hat{U}_{\perp i} - u_\perp}{h_\alpha} \right) K_\alpha \left(\frac{\hat{V}_i - v}{h_\alpha} \right) 1(X_i \in \mathcal{X}_0) \mathrm{d}u_\perp \mathrm{d}v + o_p(n^{-1/2}).
\end{aligned} \tag{159}$$

Substitute $u_\perp = \hat{U}_{\perp i} - h_\alpha s_1$ and $v = \hat{V}_i - h_\alpha s_2$, expand around $h_\alpha = 0$ and around $(\hat{U}_{\perp i}, \hat{V}_i) = (U_{\perp i}, V_i)$, and use the identities

$$\hat{W}_i - W_i = -W_i (\hat{S}_C(T_i | X_i) - S_C(T_i | X_i)) / S_C(T_i | X_i) + O_p(n^{-2\mu}) \tag{160}$$

and

$$W_i 1(T_i \leq \bar{\tau})/S_C(T_i|X_i) = 1(Y_i \leq C_i)1(T_i \leq \bar{\tau})/S_C(T_i|X_i)^2 \tag{161}$$

to obtain

$$\int \chi(u_\perp, v)(\hat{g}_0(u_\perp, v) - \check{g}_0(u_\perp, v))\mathrm{d}u_\perp \mathrm{d}v = -\frac{1}{n}\sum_{i=1}^{n} \Omega_C(T_i, 1(Y_i \leq C_i), X_i; \xi_\chi) + o_p(n^{-1/2}), \tag{162}$$

where the second equality follows from Assumption 10b.

The proof follows by combining the results on the three terms. □

Lemma 3. *Under Assumptions 1–13, as $n \to \infty$, it holds that*

$$\hat{\alpha} - \alpha = \frac{1}{n}\sum_{i=1}^{n} \Omega_\alpha(T_i, 1(Y_i \leq C_i), X_i) + o_p(n^{-1/2}). \tag{163}$$

The function $\Omega_\alpha(T, 1(Y \leq C), X)$ has mean zero and finite variance.

Proof of Lemma 3. Define

$$\hat{\alpha}_N = \int \frac{\partial^2}{\partial u_\perp \partial v}\rho(u_\perp, v)\hat{g}_0(u_\perp, v)\mathrm{d}u_\perp \mathrm{d}v, \tag{164}$$

$$\hat{\alpha}_D = \int \frac{\partial^2}{\partial u_\perp^2}\rho(u_\perp, v)\hat{g}_0(u_\perp, v)\mathrm{d}u_\perp \mathrm{d}v, \tag{165}$$

$$\alpha_N = \int \frac{\partial^2}{\partial u_\perp \partial v}\rho(u_\perp, v)g(u_\perp, v)\mathrm{d}u_\perp \mathrm{d}v \tag{166}$$

and

$$\alpha_D = \int \frac{\partial^2}{\partial u_\perp^2}\rho(u_\perp, v)g(u_\perp, v)\mathrm{d}u_\perp \mathrm{d}v, \tag{167}$$

so that $\hat{\alpha} = \hat{\alpha}_N/\hat{\alpha}_D$ and $\alpha = \alpha_N/\alpha_D$. It holds that

$$g_{11}(u_\perp, v) + g_{22}(u_\perp, v) = (1+\alpha^2)\frac{\partial^2}{\partial u^2}\log(1-\pi(u))\Big|_{u=u_\perp+\alpha v} + \frac{\partial^2}{\partial v^2}\log \tilde{F}_Y(\bar{\tau}|v). \tag{168}$$

Substitute $u_\perp = s - \alpha v$ and integrate over v first to obtain

$$\int \left(\frac{\partial^2}{\partial u_\perp \partial v}\rho(u_\perp, v) - \alpha\frac{\partial^2}{\partial u_\perp^2}\rho(u_\perp, v)\right)\frac{\partial^2}{\partial u^2}\log(1-\pi(u))\Big|_{u=u_\perp+\alpha v}\mathrm{d}u_\perp \mathrm{d}v = 0, \tag{169}$$

and integrate over $u_\perp$ first to obtain

$$\int \left(\frac{\partial^2}{\partial u_\perp \partial v}\rho(u_\perp, v) - \alpha\frac{\partial^2}{\partial u_\perp^2}\rho(u_\perp, v)\right)\frac{\partial^2}{\partial v^2}\log \tilde{F}_Y(\bar{\tau}|v)\mathrm{d}u_\perp \mathrm{d}v = 0. \tag{170}$$

Therefore, applying Lemma 2 with $\chi = \partial^2\rho/\partial u_\perp \partial v$ and $\chi = \partial^2\rho/\partial u_\perp^2$, together with Theorem 1a, Corollary 1 and Assumption 10b, yields

$$\begin{aligned}\hat{\alpha}_N - \alpha_N - \alpha\left(\hat{\alpha}_D - \alpha_D\right) = \frac{1}{n}\sum_{i=1}^{n}\Big(\Big(W_i - e^{g(U_{\perp i}, V_i)}\Big)\xi_{\alpha 1}(X_i) - \Omega_C(T_i, 1(Y_i \le C_i), X_i; \xi_{\alpha 1}) \\ -\xi'_{\alpha 2}\Omega_{\theta_\perp}(T_i, 1(Y_i \le C_i), X_i) - \xi'_{\alpha 3}\Omega_\beta(T_i, 1(Y_i \le C_i), X_i)\Big) \\ + o_p(n^{-1/2}).\end{aligned} \tag{171}$$

Since $\hat{\alpha}_D \xrightarrow{p} \alpha_D$, it follows that

$$\hat{\alpha} - \alpha = \frac{\hat{\alpha}_N - \alpha_N - \alpha(\hat{\alpha}_D - \alpha_D)}{\hat{\alpha}_D} = \frac{1}{n}\sum_{i=1}^{n}\Omega_\alpha(T_i, 1(Y_i \le C_i), X_i) + o_p(n^{-1/2}). \tag{172}$$

□

Proof of Theorem 1b. Decompose

$$\hat{\theta} - \theta = (\hat{\theta}_\perp - \theta_\perp) + \alpha(\hat{\beta} - \beta) + (\hat{\alpha} - \alpha)\beta + (\hat{\alpha} - \alpha)(\hat{\beta} - \beta). \tag{173}$$

The result follows then from Theorem 1a, Lemma 3 and Corollary 1. □

Lemma 4. *Under Assumptions 1–13, as $n \to \infty$, it holds that*

$$\hat{G}_0(t, u_\perp, v) = G(t, u_\perp, v) + L_G(t, u_\perp, v) + o_p((nh_S^2)^{-1/2}) \tag{174}$$

and

$$L_G(t, u_\perp, v) = O_p((\log(n)/nh_S^2)^{1/2}) \tag{175}$$

uniformly over $t \in [0, \bar{\tau}]$ and $(u_\perp, v) \in \mathcal{R}$.

Proof of Lemma 4. Let $\check{G}$ denote $\hat{G}$ with W_i in place of $\hat{W}_i$, and $\tilde{G}$ with W_i, $U_{\perp i}$ and V_i in place of $\hat{W}_i$, $\hat{U}_{\perp i}$ and $\hat{V}_i$. Decompose

$$\begin{aligned}\hat{G}_0(t, u_\perp, v) = G(t, u_\perp, v) + L_G(t, u_\perp, v) + \Big(\hat{G}_0(t, u_\perp, v) - \check{G}_0(t, u_\perp, v)\Big) \\ + \Big(\check{G}_0(t, u_\perp, v) - \tilde{G}_0(t, u_\perp, v)\Big) + \Big(\tilde{G}_0(t, u_\perp, v) - G(t, u_\perp, v) - L_G(t, u_\perp, v)\Big).\end{aligned} \tag{176}$$

Note that $\hat{G} - \check{G}$ satisfies

$$\begin{aligned}\hat{G}(t, u_\perp, v) - \check{G}(t, u_\perp, v) = \underset{b \in \mathbb{R}^3}{\arg\min} \sum_{i=1}^{n}((\hat{W}_i - W_i)1(T_i \le t) - (1, \hat{U}_{\perp i} - u_\perp, \hat{V}_i - v)b)^2 \\ \times K_S\left(\frac{\hat{U}_{\perp i} - u_\perp}{h_S}\right) K_S\left(\frac{\hat{V}_i - v}{h_S}\right) 1(X_i \in \mathcal{X}_0);\end{aligned} \tag{177}$$

therefore, Assumption 10a gives

$$\sup_{t \in [0, \bar{\tau}], (u_\perp, v) \in \mathcal{R}} \left|\hat{G}_0(t, u_\perp, v) - \check{G}_0(t, u_\perp, v)\right| = o_p((nh_S^2)^{-1/2}). \tag{178}$$

An expansion of the first order condition of $\check{G}$ around $\tilde{G}$ as in the proof of Lemma 2 gives

$$\sup_{t\in[0,\bar{\tau}],(u_\perp,v)\in\mathcal{R}} \left|\check{G}_0(t,u_\perp,v) - \tilde{G}_0(t,u_\perp,v)\right| = o_p((nh_S^2)^{-1/2}). \tag{179}$$

Define

$$\begin{aligned}\hat{Q}_S(u_\perp,v) = \frac{1}{nh_S^2}\sum_{i=1}^n &\left(1, \frac{U_{\perp i}-u_\perp}{h_S}, \frac{V_i - v}{h_S}\right)' \left(1, \frac{U_{\perp i}-u_\perp}{h_S}, \frac{V_i - v}{h_S}\right) \\ &\times K_S\left(\frac{U_{\perp i}-u_\perp}{h_S}\right) K_S\left(\frac{V_i - v}{h_S}\right) 1(X_i \in \mathcal{X}_0)\end{aligned} \tag{180}$$

and

$$\begin{aligned}Q_S(u_\perp,v) &= P(X\in\mathcal{X}_0) f_{U_\perp V}(u_\perp,v) \\ &\quad\times \int (1,s_1,s_2)'(1,s_1,s_2) K_S(s_1) K_S(s_2) \mathrm{d}s.\end{aligned} \tag{181}$$

The intercept of the local linear fit does not change when $U_{\perp i}-u_\perp$ and $V_i - v$ are divided by h_S, so the first order condition of $\tilde{G}$ gives

$$\begin{aligned}\tilde{G}_0(t,u_\perp,v) = (1,0,0)\hat{Q}_S(u_\perp,v)^{-1}&\frac{1}{nh_S^2}\sum_{i=1}^n W_i 1(T_i\le t) 1(X_i\in\mathcal{X}_0) \\ &\times \left(1, \frac{U_{\perp i}-u_\perp}{h_S}, \frac{V_i - v}{h_S}\right)' K_S\left(\frac{U_{\perp i}-u_\perp}{h_S}\right) K_S\left(\frac{V_i - v}{h_S}\right).\end{aligned} \tag{182}$$

Since

$$\begin{aligned}&G(t,u_\perp,v) + (U_{\perp i}-u_\perp)\frac{\partial}{\partial u_\perp}G(t,u_\perp,v) + (V_i - v)\frac{\partial}{\partial v}G(t,u_\perp,v) \\ &\quad = (1,(U_{\perp i}-u_\perp)/h_S, (V_i-v)/h_S)(G(t,u_\perp,v), h_S\frac{\partial}{\partial u_\perp}G(t,u_\perp,v), h_S\frac{\partial}{\partial v}G(t,u_\perp,v))'\end{aligned} \tag{183}$$

the function $G(t,u_\perp,v)$ equals

$$\begin{aligned}&(1,0,0)\hat{Q}_S(u_\perp,v)^{-1}\frac{1}{nh_S^2}\sum_{i=1}^n (G(t,u_\perp,v) + (U_{\perp i}-u_\perp)\frac{\partial}{\partial u_\perp}G(t,u_\perp,v) \\ &\quad + (V_i-v)\frac{\partial}{\partial v}G(t,u_\perp,v)) \\ &\quad\times \left(1, \frac{U_{\perp i}-u_\perp}{h_S}, \frac{V_i - v}{h_S}\right)' K_S\left(\frac{U_{\perp i}-u_\perp}{h_S}\right) K_S\left(\frac{V_i - v}{h_S}\right) 1(X_i\in\mathcal{X}_0).\end{aligned} \tag{184}$$

Subtract the two displays to obtain

$$\begin{aligned}\tilde{G}_0(t,u_\perp,v) - G(t,u_\perp,v) = (1,0,0)\hat{Q}_S(u_\perp,v)^{-1}&\frac{1}{nh_S^2}\sum_{i=1}^n (W_i 1(T_i\le t) - G(t,u_\perp,v) \\ &- (U_{\perp i}-u_\perp)\frac{\partial}{\partial u_\perp}G(t,u_\perp,v) - (V_i-v)\frac{\partial}{\partial v}G(t,u_\perp,v)) 1(X_i\in\mathcal{X}_0) \\ &\times \left(1, \frac{U_{\perp i}-u_\perp}{h_S}, \frac{V_i - v}{h_S}\right)' K_S\left(\frac{U_{\perp i}-u_\perp}{h_S}\right) K_S\left(\frac{V_i - v}{h_S}\right).\end{aligned} \tag{185}$$

Theorem 4 of Einmahl and Mason (2005) gives

$$\sup_{(u_\perp,v)\in\mathcal{R}} \left|\hat{Q}_S(u_\perp,v) - Q_S(u_\perp,v)\right| = O_p(h_S + (\log(n)/nh_S^2)^{1/2}) \tag{186}$$

and

$$\begin{aligned}&\frac{1}{nh_S^2}\sum_{i=1}^n (W_i 1(T_i \le t) - G(t,U_{\perp i},V_i)) \left(1, \frac{U_{\perp i}-u_\perp}{h_S}, \frac{V_i - v}{h_S}\right)' \\ &\quad \times K_S\left(\frac{U_{\perp i}-u_\perp}{h_S}\right) K_S\left(\frac{V_i-v}{h_S}\right) 1(X_i \in \mathcal{X}_0) = O_p((\log(n)/nh_S^2)^{1/2})\end{aligned} \tag{187}$$

uniformly over $t \in [0,\bar{\tau}]$ and $(u_\perp, v) \in \mathcal{R}$. The first element of the latter average equals

$$\begin{aligned}&\frac{1}{nh_S^2}\sum_{i=1}^n (W_i 1(T_i \le t) - G(t,U_{\perp i},V_i)) K_S\left(\frac{U_{\perp i}-u_\perp}{h_S}\right) K_S\left(\frac{V_i-v}{h_S}\right) 1(X_i \in \mathcal{X}_0) \\ &\quad = P(X \in \mathcal{X}_0) f_{U_\perp V}(u_\perp, v) L_G(t, u_\perp, v),\end{aligned} \tag{188}$$

which gives the second result. A Taylor expansion gives

$$\begin{aligned}&\frac{1}{nh_S^2}\sum_{i=1}^n (G(t,U_{\perp i},V_i) - G(t,u_\perp,v) - (U_{\perp i}-u_\perp)\frac{\partial}{\partial u_\perp}G(t,u_\perp,v) \\ &\quad - (V_i - v)\frac{\partial}{\partial v}G(t,u_\perp,v)) \left(1, \frac{U_{\perp i}-u_\perp}{h_S}, \frac{V_i - v}{h_S}\right)' \\ &\quad \times K_S\left(\frac{U_{\perp i}-u_\perp}{h_S}\right) K_S\left(\frac{V_i-v}{h_S}\right) 1(X_i \in \mathcal{X}_0) \\ &= o_p((nh_S^2)^{-1/2})\end{aligned} \tag{189}$$

uniformly over $t \in [0,\bar{\tau}]$ and $(u_\perp, v) \in \mathcal{R}$. The first row of $Q_S(u_\perp, v)$ is $(P(X \in \mathcal{X}_0) f_{U_\perp V}(u_\perp, v), 0, 0)$. Substitute the three rates into equation (185) to obtain

$$\tilde{G}_0(t,u_\perp,v) = G(t,u_\perp,v) + L_G(t,u_\perp,v) + o_p((nh_S^2)^{-1/2}) \tag{190}$$

uniformly over $t \in [0,\bar{\tau}]$ and $(u_\perp, v) \in \mathcal{R}$. The first result follows by combining the three terms. □

Proof of Theorem 2. A Taylor expansion of $\hat{G}_0(t, x'\hat{\theta}_\perp, x'\hat{\beta}) - \hat{G}_0(t, x'\theta_\perp, x'\beta)$ together with Lemma 4, Theorem 1a and Corollary 1 yields

$$\hat{S}_Y(t|x) = 1 - \hat{G}_0(t, x'\theta_\perp, x'\beta) + o_p((nh_S^2)^{-1/2}) = S_Y(t|x) + L_S(t,x) + o_p((nh_S^2)^{-1/2}) \tag{191}$$

uniformly over $t \in [\underline{\tau}, \bar{\tau}]$ and $x \in \mathcal{X}_0$ with $(x'\theta, x'\beta) \in \mathcal{U}_0 \times \mathcal{V}_0$, which is the expansion in part (a). For each such t and x, the term $L_S(t,x)$ has mean zero and satisfies

$$nh_S^2 \operatorname{Var}(L_S(t,x)) = \sigma_S^2(t,x) + o(1), \tag{192}$$

so the limit distribution in part (a) follows from the Lindeberg-Feller central limit theorem.

Similar Taylor expansions, together with Lemmas 3 and 4, yield

$$\hat{r}(u) = r(u) + L_r(u) + o_p((nh_S)^{-1/2}) \tag{193}$$

and

$$\hat{q}(t,v) = q(t,v) + L_q(t,v) + o_p((nh_S)^{-1/2}). \tag{194}$$

Parts (b) and (c) now follow from analogous arguments as for part (a). □

Proof of Proposition 1. By Theorem 2b, it holds that

$$\hat{r}(u) = r(u) + L_r(u) + o_p((nh_S)^{-1/2}) \tag{195}$$

uniformly over $u \in \mathcal{U}_0$. Substituting $v = V_i - h_S s$ in $L_r(u)$ and applying Theorem 4 of Einmahl and Mason (2005) gives

$$\sup_{u\in\mathcal{U}_0} |L_r(u)| = O_p((\log(n)/nh_S)^{1/2}) \tag{196}$$

and

$$\sup_{u\in\mathcal{U}_0} |L_r'(u)| = O_p((\log(n)/nh_S^3)^{1/2}) = o_p(1). \tag{197}$$

The rate on $\hat{\kappa}_-$ follows from equations (195) and (196) and

$$|\hat{\kappa}_- - \kappa_-| \leq \sup_{u\in\mathcal{U}_0} |\hat{r}(u) - r(u)|. \tag{198}$$

The results for $\hat{\kappa}_+$ follow from identical steps. □

Proof of Proposition 2. Any other orthonormal basis of the orthogonal complement of $\hat{\beta}$ can be written as $\hat{B}Q$ with Q an orthogonal $(d-1)\times(d-1)$ matrix. Since $\hat{B}Q(Q'\hat{B}'\hat{\Sigma}_{\theta_\perp}\hat{B}Q)^{-1}Q'\hat{B}' = \hat{B}(\hat{B}'\hat{\Sigma}_{\theta_\perp}\hat{B})^{-1}\hat{B}'$, the statistic does not depend on the choice of $\hat{B}$.

Let A be a fixed $d\times(d-1)$ matrix such that (β, A) is nonsingular, and let $\hat{B}$ be obtained by orthonormalizing the columns of $(I - \hat{\beta}\hat{\beta}'/\|\hat{\beta}\|^2)A$. The orthonormalization is continuous in $\hat{\beta}$ in a neighborhood of β, so $\hat{\beta} \xrightarrow{p} \beta$ implies $\hat{B} \xrightarrow{p} B$, where the columns of B form an orthonormal basis of the orthogonal complement of β.

Write

$$\hat{B}'\hat{\theta}_\perp = \hat{B}'(\hat{\theta}_\perp - \theta_\perp) + \hat{B}'\theta_\perp. \tag{199}$$

By Corollary 1, the influence function $\Omega_{\theta_\perp}(T, 1(Y \leq C), X)$ has mean zero and finite variance, so the central limit theorem gives

$$\sqrt{n}(\hat{\theta}_\perp - \theta_\perp) \xrightarrow{d} N(0, \Sigma_{\theta_\perp}), \tag{200}$$

and the law of large numbers gives $\hat{\Sigma}_{\theta_\perp} \xrightarrow{p} \Sigma_{\theta_\perp}$. Combined with $\hat{B} \xrightarrow{p} B$, this yields

$$\sqrt{n}\hat{B}'(\hat{\theta}_\perp - \theta_\perp) \xrightarrow{d} N(0, B'\Sigma_{\theta_\perp}B) \tag{201}$$

and $\hat{B}'\hat{\Sigma}_{\theta_\perp}\hat{B} \xrightarrow{p} B'\Sigma_{\theta_\perp}B$, which is nonsingular.

If $\theta \propto \beta$, then $\theta_\perp = 0$, so $\sqrt{n}\hat{B}'\hat{\theta}_\perp \xrightarrow{d} N(0, B'\Sigma_{\theta_\perp}B)$, and the continuous mapping theorem yields $J_1 \xrightarrow{d} \chi^2(d-1)$. If $\theta \not\propto \beta$, then $\hat{B}'\hat{\theta}_\perp \xrightarrow{p} B'\theta_\perp$, which is nonzero because $\theta_\perp \neq 0$ lies in the span of the columns of B. Hence $J_1/n \xrightarrow{p} \theta_\perp' B(B'\Sigma_{\theta_\perp}B)^{-1}B'\theta_\perp > 0$, so $J_1 \xrightarrow{p} \infty$. □

Lemma 5. *Let $\hat{e}_l$ be as in Proposition 3. Suppose Assumptions 1–7 and 9–13 hold, and that the conditions on ω hold for all ω_l. As $n \to \infty$, it holds that*

$$\hat{e}_l - e_l = \frac{1}{n}\sum_{i=1}^{n} \Omega_{e_l}(T_i, 1(Y_i \le C_i), X_i) + o_p(n^{-1/2}). \tag{202}$$

The function $\Omega_{e_l}(T, 1(Y \le C), X)$ has mean zero and finite variance.

The proof of Lemma 5 is similar to that of Theorem 1a and therefore omitted.

Proof of Proposition 3. Lemma 1 with ω_l in place of ω gives

$$\hat{\delta}_l - \delta_l = \frac{1}{n}\sum_{i=1}^{n} \Omega_{\delta_l}(T_i, 1(Y_i \le C_i), X_i) + o_p(n^{-1/2}). \tag{203}$$

Write $R_l = \theta_\perp' \delta_l \, e_1 - \|\theta_\perp\|^2 e_l$ and $R = (R_2, \ldots, R_L)'$. Expand $\hat{R}_l$ around $(\theta_\perp, \delta_l, e_1, e_l)$ to obtain

$$\hat{R}_l - R_l = e_1\delta_l'(\hat{\theta}_\perp - \theta_\perp) + e_1\theta_\perp'(\hat{\delta}_l - \delta_l) + \theta_\perp'\delta_l(\hat{e}_1 - e_1) - 2e_l\theta_\perp'(\hat{\theta}_\perp - \theta_\perp) - \|\theta_\perp\|^2(\hat{e}_l - e_l) + O_p(n^{-1}), \tag{204}$$

and therefore,

$$\hat{R} - R = \frac{1}{n}\sum_{i=1}^{n} \Omega_R(T_i, 1(Y_i \le C_i), X_i) + o_p(n^{-1/2}), \tag{205}$$

where $\Omega_R(T, 1(Y \le C), X)$ has mean zero and finite variance.

The central limit theorem gives $\sqrt{n}(\hat{R} - R) \xrightarrow{d} N(0, \Sigma_R)$, and the law of large numbers gives $\hat{\Sigma}_R \xrightarrow{p} \Sigma_R$, which is nonsingular. If $\log(1 - \pi(u))$ is affine on $\{x'\theta : x \in \mathcal{X}_0\}$, then $R = 0$, so the continuous mapping theorem yields $J_2 \xrightarrow{d} \chi^2(L-1)$. If $R \neq 0$, then $\hat{R} \xrightarrow{p} R$, so $J_2/n \xrightarrow{p} R'\Sigma_R^{-1}R > 0$ and $J_2 \xrightarrow{p} \infty$. □

Proof of Proposition 4. Theorem 1a with ψ^* in place of ψ gives

$$\hat{\beta}^* - \beta^* = \frac{1}{n}\sum_{i=1}^{n} \Omega_{\beta^*}(T_i, 1(Y_i \le C_i), X_i) + o_p(n^{-1/2}). \tag{206}$$

Decompose

$$\hat{B}'\hat{\beta}^* = \hat{B}'(\hat{\beta}^* - \beta^*) + \hat{B}'\beta^*. \tag{207}$$

Under the null, $\beta^* = (\beta'\beta^*/\|\beta\|^2)\beta$. Since $\hat{B}'\hat{\beta} = 0$,

$$\hat{B}'\beta^* = \frac{\beta'\beta^*}{\|\beta\|^2}\hat{B}'\beta = -\frac{\beta'\beta^*}{\|\beta\|^2}\hat{B}'(\hat{\beta} - \beta), \tag{208}$$

and therefore

$$\hat{B}'\hat{\beta}^* = \hat{B}'\frac{1}{n}\sum_{i=1}^{n}\left(\Omega_{\beta^*}(T_i, 1(Y_i \le C_i), X_i) - \frac{\beta'\beta^*}{\|\beta\|^2}\Omega_\beta(T_i, 1(Y_i \le C_i), X_i)\right) + o_p(n^{-1/2}). \tag{209}$$

As in the proof of Proposition 2, $\hat{B} \xrightarrow{p} B$ and $\hat{\Sigma}_{\beta^*} \xrightarrow{p} \Sigma_{\beta^*}$, so the central limit theorem gives $\sqrt{n}\hat{B}'\hat{\beta}^* \xrightarrow{d} N(0, B'\Sigma_{\beta^*}B)$ and $\hat{B}'\hat{\Sigma}_{\beta^*}\hat{B} \xrightarrow{p} B'\Sigma_{\beta^*}B$, which is nonsingular. The continuous mapping theorem yields $J_3 \xrightarrow{d} \chi^2(d-1)$. If $\beta^* \not\propto \beta$, then $\hat{B}'\hat{\beta}^* \xrightarrow{p} B'\beta^* \neq 0$, so

$$J_3/n \xrightarrow{p} \beta^{*\prime}B(B'\Sigma_{\beta^*}B)^{-1}B'\beta^* > 0 \tag{210}$$

and $J_3 \xrightarrow{p} \infty$. □