Nonparametric Estimation of Cumulative Incidence Functions for Competing Risks Data with Missing Cause of Failure

by

Georgios Effraimidis

and

Christian M. Dahl

Discussion Papers on Business and Economics
No. 21/2013

FURTHER INFORMATION
Department of Business and Economics
Faculty of Business and Social Sciences
University of Southern Denmark
Campusvej 55
DK-5230 Odense M
Denmark

Tel.: +45 6550 3271
Fax: +45 6550 3237
E-mail: lho@sam.sdu.dk
http://www.sdu.dk/ivoe

Nonparametric Estimation of Cumulative Incidence Functions for Competing Risks Data with Missing Cause of Failure*

Georgios Effraimidis †
Department of Business and Economics and COHERE
University of Southern Denmark

Christian M. Dahl
Department of Business and Economics, COHERE and CREATES
University of Southern Denmark

December 15, 2013

Abstract

In this paper, we develop a fully nonparametric approach for the estimation of the cumulative incidence function with Missing At Random right-censored competing risks data. We obtain results on the pointwise asymptotic normality as well as the uniform convergence rate of the proposed nonparametric estimator. A simulation study that serves two purposes is provided. First, it illustrates in details how to implement our proposed nonparametric estimator. Secondly, it facilitates a comparison of the nonparametric estimator to a parametric counterpart based on the estimator of Lu and Liang (2008). The simulation results are generally very encouraging.

Keywords: Cumulative incidence function; Inverse probability weighting; Kernel estimation; Local linear estimation; Martingale central limit theorem.

JEL Codes: C14, C41.

*The second author acknowledges support from CREATES, Center for Research in Econometric Analysis of Time Series (DNRF78), funded by the Danish National Research Foundation.

†Corresponding author. Adress: Department of Business and Economics, Campusvej 55, 5230 Odense M, Denmark. Email: georgiosnl@gmail.com

1 Introduction

Competing risks models are widely used in biostatistics, empirical health economics and labor economics, for example, when analyzing data for onset of diseases, mortality due to mutually exclusive causes of death or unemployment where each individual is faced with competing exits (full-time employment, part-time employment). Hence, studying estimators of the cumulative incidence function within this modelling framework is of great importance. The goal of this paper is to derive asymptotic results for the nonparametric estimator of the cumulative incidence function in cases where continuous covariates affect the realization of the failure time and the cause of failure is Missing At Random (MAR) for some observations. The proposed nonparametric estimator is complementary to *i*) the developed (semi)-parametric procedures with right-censored data and continuous explanatory covariates (e.g., Andersen et al., 1993; Jeong and Fine, 2007; Scheike et al., 2008) and *ii*) the suggested parametric methods for right-censored data where the cause of failure is sometimes missing (Lu and Liang, 2008). Finally, we compare our results on uniform convergence rates with the results of Bordes and Gneyou (2011) who discuss the uniform convergence rate for the nonparametric estimator with right-censored competing risks data.

2 The Nonparametric Estimator

For expositional convenience, we will focus on two risks, 1 and 2. Let Y be the (actual) failure time and $\gamma \in \{1, 2\}$ be a failure type indicator. Let $X \in \mathcal{X} \subset \mathcal{R}^d$ be a vector of observed covariates and denote by x its realization. Define for each risk $j = 1, 2$ and $(t, x) \in \mathcal{R}_+ \times \mathcal{X}$ the cumulative incidence function

$$F_j(t|x) := \mathbb{P}(Y \leq t, \gamma = j|x). \tag{1}$$

We introduce the cause-specific hazard rate

$$\lambda_j(t,x) := \lim_{t\to 0} \frac{\mathbb{P}(t \le Y < t+dt, \gamma = j | Y \ge t, x)}{t}. \tag{2}$$

The cumulative cause-specific hazard rate is defined as follows: $\Lambda_j(t,x) := \int_0^t \lambda_j(u,x)du$. Also, consider the overall hazard rate $\lambda(t,x) := \lambda_1(t,x) + \lambda_2(t,x)$, the corresponding cumulative overall hazard rate $\Lambda(t,x) := \int_0^t \lambda(u,x)du$, and the survival function $S(t-|x) := \mathbb{P}(Y \ge t|x)$. By using (1) and (2) we get for $j = 1, 2$

$$F_j(t|x) = \int_0^t S(u-|x) d\Lambda_j(u,x), \tag{3}$$

where

$$S(t-|x) = \prod_{u<t} \{1 - d\Lambda(u,x)\}. \tag{4}$$

Denote by Z the censoring variable with $Z \perp\!\!\!\perp Y, \gamma \mid X$, where the symbol $\perp\!\!\!\perp$ implies independence between the underlying random variables. Also, $T := \min(Y, Z)$, $\tilde{\gamma} := \gamma 1\{Y \le Z\}$. We observe n independently and identically distributed copies $(T_i, X_i, 1\{\tilde{\gamma}_i > 0\}, R_i, R_i\tilde{\gamma}_i,)$, where R_i is the missing indicator variable and the missing data mechanism satisfies the MAR assumption (Rubin, 1976; Little and Rubin, 1987). The value of R_i equals 0 if $T_i = Y_i$ and the cause of failure is not observed. On the other hand, the indicator variable R_i is equal to 1 if $T_i = Y_i$ and the cause of failure is observed or if $T_i = Z_i$. The MAR scheme that we adopt is described as follows:

$$\mathbb{P}(R = 1|\tilde{\gamma}, \tilde{\gamma} > 0, T, x) = \mathbb{P}(R = 1|\tilde{\gamma} > 0, x) =: \pi(x). \tag{5}$$

The independence of the probability on T has as its consequence the predictability of all integrands of the proposed estimator. We also assume that

$$R \perp\!\!\!\perp T|X. \tag{6}$$

The latter is necessary in order to ensure that the underlying martingale processes are zero-mean. In the above discussion we assume that the covariates are time-invariant. This setup is adopted only for notational convenience as all the results in the sequel are true if X is predictable.

We will study the two following estimators for the cumulative incidence function,

$$\hat{F}_j^C(t|x) = \int_0^t \hat{S}^C(u-|x)d\hat{\Lambda}_j^C(u,x), \qquad j = 1,2 \tag{7}$$

and

$$\hat{F}_j^L(t|x) = \int_0^t \hat{S}^L(u-|x)d\hat{\Lambda}_j^L(u,x), \qquad j = 1,2, \tag{8}$$

where the superscripts C and L refer to the type of smoothing with respect to vector x. In particular, C is used for the local constant smoothing, whereas L is used for the local linear smoothing.

Let $\omega = (\omega_1, ..., \omega_d) \in \mathcal{R}^d$ and $\mathcal{K}_h(\omega) = \frac{1}{h^d}\prod_{p=1}^{d} K\left(\frac{\omega_p}{h}\right)$, where K is a kernel with compact support $\mathbb{K}$ and $h = o(n)$. Introduce the quantity $\mathcal{L}_{h,x}(\omega) = \frac{\mathcal{K}_h(\omega)-\mathcal{K}_h(\omega)\omega^{\mathrm{T}}\bar{D}^{-1}\bar{c}_1}{\bar{c}_0-c_1^{\mathrm{T}}\bar{D}^{-1}\bar{c}_1}$, with $\bar{c}_0 = \frac{1}{n}\sum_{i=1}^n \mathcal{K}_h(x-X_i)$, $\bar{c}_{1\rho} = \frac{1}{n}\sum_{i=1}^n \mathcal{K}_h(x-X_i)(x_\rho-X_{i\rho})$, $\bar{d}_{\rho\kappa} = \frac{1}{n}\sum_{i=1}^n \mathcal{K}_h(x-X_i)(x_\rho-X_{i\rho})(x_\kappa-X_{i\kappa})$, $\bar{c}_1 = (\bar{c}_{1\rho})_{\rho=1}^d$, and $\bar{D} = (\bar{d}_{\rho\kappa})_{\rho,k=1}^d$. The notations x_ρ and $X_{i\rho}$ refer to the $\rho-$th element of the corresponding row vector. The quantity $\mathcal{K}_h(.)$ will be used for the construction of the weights for the local constant estimator. On the other hand, the quantity $\mathcal{L}_{h,x}(.)$, which is also commonly referred to as the equivalent kernel, will be used for the construction of the weights for the local linear estimator.

First, we will describe the nonpamatetric estimator for the probability of having an observation with a missing cause of failure. The estimator of this probability is needed for the estimator of $\Lambda_j(t,x)$. Define $\pi(x,\tilde{\gamma}) := \mathbb{P}(R = 1|x,\tilde{\gamma}) = 1\{\tilde{\gamma} > 0\}\pi(x) + 1\{\tilde{\gamma} = 0\}$. That is, $\pi(x,\tilde{\gamma})$ specifies the probability of having an observation with a missing cause of failure given the observed characteristics x and the value of the indicator $\tilde{\gamma}$. This probability is independent of the exact value of $\tilde{\gamma}$ in case the latter is strictly positive (i.e., 1 or 2), whereas it is equal to one if $\tilde{\gamma} > 0$ (i.e., the observation is censored). For the local constant smoothing we have $\hat{\pi}^C(x) = \frac{\sum_{i=1}^n \mathcal{K}_h(x-X_i)1\{\tilde{\gamma}_i>0\}R_i}{\sum_{i=1}^n \mathcal{K}_h(x-X_i)1\{\tilde{\gamma}_i>0\}}$, whereas for the local linear smoothing we have $\hat{\pi}^L(x) = \frac{\sum_{i=1}^n \mathcal{L}_{h,x}(x-X_i)1\{\tilde{\gamma}_i>0\}R_i}{\sum_{i=1}^n \mathcal{L}_{h,x}(x-X_i)1\{\tilde{\gamma}_i>0\}}$.

Next, we proceed with the description of the estimator for the function $\Lambda_j(t,x)$. For this purpose, consider the counting process $N_{ji}(t) = 1\{T_i \le t, \tilde{\gamma}_i = j, R_i = 1\}$. This process describes whether subject i has failed due to risk j in the time interval $[0,t]$, and the cause of failure is not missing. Moreover, consider the "at risk" predictable process $Y_i(t) = 1\{T_i \ge t\}$, which describes whether subject i has survived and has not been censored up to $t-$. Furthermore, we will make use of the following weights for the local constant and local linear smoothing:

$$w_i^C(x) = \frac{R_i}{\hat{\pi}^C(X_i,\tilde{\gamma}_i)}\mathcal{K}_h(x-X_i)\Big/\sum_{i=1}^n \frac{R_i}{\hat{\pi}^C(X_i,\tilde{\gamma}_i)}\mathcal{K}_h(x-X_i),$$
$$w_i^L(x) = \frac{R_i}{\hat{\pi}^L(X_i,\tilde{\gamma}_i)}\mathcal{L}_{h,x}(x-X_i)\Big/\sum_{i=1}^n \frac{R_i}{\hat{\pi}^L(X_i,\tilde{\gamma}_i)}\mathcal{L}_{h,x}(x-X_i).$$

Denote by $H(t|x)$ the conditional survival function of the random variable T. The estimator of the cumulative hazard rate $\Lambda_j(t,x)$ is given by

$$\hat{\Lambda}_j^\nu(t,x) = \frac{1}{n}\sum_{i=1}^n \int_0^t \frac{w_i^\nu(x)}{\hat{H}^\nu(u-|x)} dN_{ji}(u), \tag{9}$$

where $\hat{H}^\nu(t-|x) = \sum_{i=1}^n w_i^\nu(x)1(T_i \ge t)$ for $\nu = C, L$. Note that we employ an inverse probability weighting (either local constant or local linear smoothing) scheme. A similar approach is commonly used in the standard regression context for dealing with MAR observations (Hu et al., 2010).

Finally, it remains to present the estimator for the survival function $S(t|x)$. We introduce the counting process $\bar{N}_i(t) = 1\{T_i \le t, \tilde{\gamma}_i > 0\}$, which specifies whether subject i has failed due to either risk 1 or risk 2 in the time interval $[0,t]$. Additionally, we will use the following weights for the two different smoothing techniques:

$$b_i^C(x) = \mathcal{K}_h(x - X_i) \Big/ \sum_{i=1}^{n} \mathcal{K}_h(x - X_i),$$

$$b_i^L(x) = \mathcal{L}_{h,x}(x - X_i) \Big/ \sum_{i=1}^{n} \mathcal{L}_{h,x}(x - X_i)$$

We have

$$\hat{\Lambda}^\nu(t, x) = \frac{1}{n} \sum_{i=1}^{n} \int_0^t \frac{b_i^\nu(x)}{\bar{H}^\nu(u - |x)} d\bar{N}_i(u),$$

where $\bar{H}^\nu(t - |x) = \sum_{i=1}^{n} b_i^\nu(x) 1(T_i \geq t)$ for $\nu = C, L$. The estimator of $S(t - |x)$ is given by

$$\hat{S}^\nu(t - |x) = \prod_{u<t} \left\{1 - d\hat{\Lambda}^\nu(u, x)\right\}. \tag{10}$$

Note that for the latter estimator, we adopt the conventional smoothing (either local constant or local linear) techniques (i.e., without considering the missing cause of failure). The reason that we do not need inverse probability weights for the estimation is that we always observe variable T_i and the stochastic variable $1\{\tilde{\gamma}_i > 0\}$ irrespective of whether we observe the cause of failure.

3 Asymptotic results

We assume that the support of X is of the form $\mathcal{X} = \bigotimes_{p=1}^{p=d} [x_{lp}, x_{up}] \subset \mathcal{R}^d$, with $x_{lp} < x_{up}$ for any $p = 1, ..., d$. We also define the internal region $\mathcal{X}_h := \left\{x \in \mathcal{X} : \left\{x - h\omega : \omega \in \mathbb{K}^d\right\} \subset \mathcal{X}\right\}$. For $x \in \mathcal{X}_h$, let $\tau(x)$ be some real positive number such that $\tau(x) < \sup\{t \in \mathcal{R}_+ : H(t|x) > 0\}$. Finally, $H(t, x) := H(t|x)f(x)$, where $f(x)$ is the probability density function of X, $H(t, x, \tilde{\gamma} > 0) := H(t, x|\tilde{\gamma} > 0)\mathbb{P}(\tilde{\gamma} > 0), H(t, x, \tilde{\gamma} = 0) := H(t, x|\tilde{\gamma} = 0)\mathbb{P}(\tilde{\gamma} = 0)$ and $||K||_2^2 := \int K^2(u)du$. We will employ the following assumptions to derive the asymptotic normality of the proposed estimator. All the results are proved in the appendix.

Assumption 1 *The derivatives of $\lambda_j(t,x)$ $(j=1,2)$ and $H(t|x)$ with respect to x are continuously differentiable up to order 2 on the interior of $[0,\tau(x)]$ for any $x \in \mathcal{X}_h$, and the corresponding derivatives are uniformly bounded. Moreover, the probability density function, $f(x)$, is strictly positive on $\mathcal{X}_h$.*

Assumption 2 *It holds that $\pi(x) \geq \epsilon > 0$ for any $x \in \mathcal{X}$.*

Assumption 3 *The univariate kernel, K, is (i) a continuous probability density function with compact support $\mathbb{K}$=$[-\mathcal{S}_k, \mathcal{S}_k]$, where $0 < \mathcal{S}_k < \infty$, and (ii) of order 2.*

Assumption 4 *For the bandwidth sequence it holds that $nh^{d+4} = O(1)$.*

Define for $j = 1, 2$,

$$g(t,x) = \frac{S(t|x)}{H(t,x)}, \qquad \rho_j(t,u,x) = -\frac{\int_u^t S(\epsilon|x)\lambda_j(\epsilon,x)d\epsilon}{H(u,x)}.$$

Additionally, for $\xi = 1, 2$, with $\xi \neq j$,

$$b_{jA}^C(t,x) = \sum_{l=0}^{1} \frac{\mu_2(K)h^2}{(2-l)!l!} \sum_{p=1}^{d} \int_0^t \frac{\partial^{2-l}\lambda_j(u,x)}{\partial x_p^{2-l}} \frac{\partial^l H(u,x)}{\partial x_p^l} \left[g(u,x) + \rho_j(t,u,x)\right] du$$
$$b_{jB}^C(t,x) = \sum_{l=0}^{1} \frac{\mu_2(K)h^2}{(2-l)!l!} \sum_{p=1}^{d} \int_0^t \frac{\partial^{2-l}\lambda_\xi(u,x)}{\partial x_p^{2-l}} \frac{\partial^l H(u,x)}{\partial x_p^l} \rho_j(t,u,x)du,$$

and

$$b_{jA}^L(t,x) = \frac{\mu_2(K)h^2}{2} \sum_{p=1}^{d} \int_0^t \frac{\partial^2\lambda_j(u,x)}{\partial x_p^2} H(u,x) \left[g(u,x) + \rho_j(t,u,x)\right] du,$$
$$b_{jB}^L(t,x) = \frac{\mu_2(K)h^2}{2} \sum_{p=1}^{d} \int_0^t \frac{\partial^2\lambda_\xi(u,x)}{\partial x_p^2} H(u,x)\rho_j(t,u,x)du.$$

Moreover,

$$
\begin{aligned}
v_{jA}(t,x) &= ||K||_2^2 \int_0^t \left[\frac{1}{\pi(x)} H(u,x,\tilde{\gamma}>0) + H(u,x,\tilde{\gamma}=0)\right] g^2(u,x)\lambda_j(u,x)du, \\
v_{jB}(t,x) &= ||K||_2^2 \int_0^t H(u,x)\lambda(u,x)\rho_j^2(t,u,x)du, \\
v_{jAB}(t,x) &= 2\,||K||_2^2 \int_0^t H(u,x)g(u,x)\rho_j(t,u,x)\lambda_j(u,x)du, \\
\varsigma_1(t,x) &= ||K||_2^2 \int_0^t H(u,x)\left[g(u,x)+\rho_1(t,u,x)\right]\rho_2(t,u,x)\lambda_1(u,x)du, \\
\varsigma_2(t,x) &= ||K||_2^2 \int_0^t H(u,x)\left[g(u,x)+\rho_2(t,u,x)\right]\rho_1(t,u,x)\lambda_2(u,x)du.
\end{aligned}
$$

Let $\mathcal{D}\left[0,\tau(x)\right]$ denote the space of cadlag functions endowed with the Skorohod topology. Additionally, the symbol $\Longrightarrow$ will imply weak convergence. We now state the main result of the paper.

Theorem 1 *Suppose that Assumptions 1-4 hold. Then, for each $x \in \mathcal{X}_h$, we have, as $n \to \infty$*

$$
\sqrt{nh^d}\begin{bmatrix} \hat{F}_1^{\nu}(t|x) - F_1(t|x) - b_{1A}^{\nu}(t,x) - b_{1B}^{\nu}(t,x) \\ \hat{F}_2^{\nu}(t|x) - F_2(t|x) - b_{2A}^{\nu}(t,x) - b_{2B}^{\nu}(t,x) \end{bmatrix} \Longrightarrow \mathcal{N}(0,V(t,x))
$$

over $\mathcal{D}\left[0,\tau(x)\right]^2$, where

$$
V(t,x) = \begin{bmatrix} v_{1A}(t,x)+v_{1B}(t,x)+v_{1AB}(t,x) & \varsigma_1(t,x)+\varsigma_2(t,x) \\ \varsigma_1(t,x)+\varsigma_2(t,x) & v_{2A}(t,x)+v_{2B}(t,x)+v_{2AB}(t,x) \end{bmatrix}
$$

is a positive semidefinite matrix on $[0,\tau(x)]$ for each $x \in \mathcal{X}_h$.

Theorem 1 is obtained by applying the martingale central theorem (Andersen et al., 1993; Nielsen and Linton, 1995; Linton et al., 2011). To digest the above result, recall that

$$
\hat{F}_j^{\nu}(t|x) = \int_0^t \hat{S}^{\nu}(u-|x)d\hat{\Lambda}_j^{\nu}(u,x).
$$

The bias and variance due to $\hat{\Lambda}_j^{\nu}(t,x)$ are captured by the terms $b_{jA}^{\nu}(t,x)$ and $v_{jA}(t,x)$. On the other hand, the bias and variance due to $\hat{S}^{\nu}(t,x)$ are captured by the terms $b_{jB}^{\nu}(t,x)$ and $v_{jB}(t,x)$. Moreover, the term $v_{jAB}(t,x)$ refers to the covariance of the estimators $\hat{\Lambda}_j^{\nu}(t,x)$ and $\hat{S}^{\nu}(t,x)$. The next result gives the asympotic distribution in case the cause of failure is observed for the uncensored observations.

Corollary 1 *Suppose that Assumptions 1-4 hold, and $\pi(x) = 1$ for all $x \in \mathcal{X}_h$. Then, for each $x \in \mathcal{X}_h$, we have, as $n \to \infty$*

$$\sqrt{nh^d}\begin{bmatrix} \hat{F}_1^{\nu}(t|x) - F_1(t|x) - b_{1A}^{\nu}(t,x) - b_{1B}^{\nu}(t,x) \\ \hat{F}_2^{\nu}(t|x) - F_2(t|x) - b_{2A}^{\nu}(t,x) - b_{2B}^{\nu}(t,x) \end{bmatrix} \Longrightarrow \mathcal{N}(0, \ddot{V}(t,x))$$

over $\mathcal{D}\left[0,\tau(x)\right]^2$, where

$$\ddot{V}(t,x) = \begin{bmatrix} \dot{v}_{jA}(t,x) + v_{1B}(t,x) + v_{1AB}(t,x) & \varsigma_1(t,x) + \varsigma_2(t,x) \\ \varsigma_1(t,x) + \varsigma_2(t,x) & \dot{v}_{jA} + v_{2B}(t,x) + v_{2AB}(t,x) \end{bmatrix}$$

with $\ddot{v}_{jA}(t,x) = ||K||_2^2 \int_0^t H(u,x) g^2(u,x) \lambda_j(u,x) du$.

In case there is no censoring, we have that $S(t|x) = H(t|x)$ for each $(t,x) \in \mathcal{R}_+ \times \mathcal{X}$, and we can show by using the Duhamel equation (Gill, 1994) that,

$$\begin{aligned} \hat{F}_j^C(t|x) - F_j(t|x) &= \frac{1}{\sum_{i=1}^n \mathcal{K}_h(x - X_i)} \sum_{i=1}^n \int_0^t \mathcal{K}_h(x - X_i) dN_{ji}(u) - \int_0^t S(u - |x) \lambda_j(u,x) du \\ &= \left[\mathcal{V}_{jj}(t,x) + \mathcal{V}_{j\xi}(t,x) + \mathcal{B}_{jj}(t,x) + \mathcal{B}_{j\xi}(t,x)\right]\left[1 + o_p(1)\right], \end{aligned}$$

where the quantities $\mathcal{V}_{jj}(t,x), \mathcal{V}_{j\xi}(t,x), \mathcal{B}_{jj}(t,x), \mathcal{B}_{j\xi}(t,x)$ are defined in the appendix by setting $\pi(X_i, \tilde{\gamma}_i) = 1$. To derive the above equation, we also use the fact that $\inf_{x \in \mathcal{X}_h} \sum_{i=1}^n \mathcal{K}_h(x - X_i)/n \geq \epsilon + o_p(1)$ for large n, which is obtained by combining standard results in nonparametric density estimation, see, e.g., Hansen (2008) and Assumption 1. Using arguments similar to the ones applied in the proof of Theorem 1, we get the distribution of Corollary 1. Finally, results about the rate of uniform convergence of $\hat{F}_j^{\nu}(t|x)$ to $F_1(t|x)$ would be interesting. In particular, let $\alpha_n \equiv \left(\frac{\ln n}{nh^d}\right)^{\frac{1}{2}} + h^2$, $\Xi \equiv [0,\tau] \times \mathcal{X}_h$ and replace Assumption 4, which is concerned with the bandwidth, with the following Assumption

Assumption 4* *For the bandwidth sequence it that holds* $\frac{\ln n}{nh^d} = o(1)$.

Then the following result emerges:

Theorem 2 *Suppose Assumptions 1-3, and 4* hold. Then, for* $j = 1, 2$ *and* $\nu = C, L$, *we have, as* $n \to \infty$

$$\sup_{(t,x)\in\Xi} \left| \hat{F}_j^{\nu}(t|x) - F_j(t|x) \right| = O_p(\alpha_n).$$

The convergence rate is almost identical to the rate of Bordes and Gneyou (2011) who study the uniform convergence rate just for right-censored competing risks data. In their result, the variance term is of the same order whereas their bias term goes faster to zero as it is of order h^{2d}.

4 Simulation studies

In this section, the main focus will be on evaluating the performance of the proposed nonparametric estimators of the cumulative incidence function $F_j(t|x)$. The design of the numerical study will be similar to Lu and Liang (2008), which makes benchmarking to a parametric estimator of $F_j(t|x)$ straightforward. The (cause-specific) hazards model generating failure time from the first course, Y_1, is given by $\lambda_1(t|x) = \lambda_1(t) + \beta' x$, where $x = (x_1, x_2)'$, $\beta = (\beta_1, \beta_2)' = (1, -1)'$ and the baseline hazard function is defined as $\lambda_1(t) = 1.3$. Furthermore, x_1 is assumed standard normally distributed, while x_2 is following a binomial distribution with a probability of success equal to 0.5. The (cause-specific) hazards model generating failure time due to the second course, Y_2, is specified as $\lambda_2(t|x) = \exp(a + bt)$ for $(a, b) = (-1, 1)$. Censoring time Z is generated from a uniform distribution on $(0, c)$. By choosing $c = 3.6$, the censoring level equals about 15 percent, and in this case 55 percent of all failures are of type 1. The case $c = 0.75$ is also considered, and in this setting the censoring level equals about 40 percent, while approximately 42 percent of all failures are of type 1. The missing cause of failure indicator R is generated from a logistic distribution, i.e., $\pi(x) = \exp(-2.5 + \gamma' x)/(1 + \exp(-2.5 + \gamma' x))$, where $\gamma = (2, 2)'$. Consequently, about 44 percent of all failures have missing causes when the censoring level is 15 percent. When the censoring level is about 40 percent, approximately 50 percent of the failures have missing causes. We will report only the results based on the kernel-based estimator of the cumulative incidence function. The results for the local constant smoother are similar.[1] Within the described simulations setup, the estimator given by equation (7) is computed numerically as

$$\hat{F}_j^C(t|x_1, x_2) = \sum_{i=1}^{n} \hat{S}^C(T_i - |x_1, x_2) \frac{w_i^C(x_1, x_2) 1\{T_i \leq t, R_i \tilde{\gamma}_i = j\}}{\sum_{\rho=1}^{n} w_\rho^C(x_1, x_2) 1(T_\rho \geq T_i)}, \qquad j = 1, 2, \tag{11}$$

for

$$\hat{S}^C(t - |x_1, x_2) = \prod_{i=1}^{n} \left\{ 1 - \frac{b_i^C(x_1, x_2)}{\sum_{\rho=1}^{n} 1\{T_\rho \geq T_i\} b_\rho^C(x_1, x_2)} \right\}^{1\{T_i < t, \tilde{\gamma}_i > 0\}},$$

where T_i, R_i, $\tilde{\gamma}_i$ and the functionals $w_i^C(x_1, x_2)$, $\hat{\pi}^C(X_{1i}, X_{2i}, \tilde{\gamma}_i)$ and $b_i^C(x_1, x_2)$ are defined as in Section 1. In all of the computations involving smoothing, the product kernel for mixed continuous and discrete data types, described in Li and Racine (2008), page 424, is applied. To the best of our knowledge, there is no existing theory currently available regarding datadriven bandwidth selection procedures within our setup. Consequently, we have used an m-fold cross validation approach similar to Nielsen and Linton (1995). Specifically, the selected bandwidth is obtained as $h = \arg\min_b \sum_{i=1}^{n} \left(\hat{F}_{j,-m_i}^C(T_i|X_{i1}, X_{i2}; b) - \hat{F}_j^C(T_i|X_{i1}, X_{i2}; b) \right)^2$, where m_i denotes a subset of observations (including observation i) randomly drawn from the sample. $\hat{F}_{j,-m_i}^C(T_i|X_{i1}, X_{i2}; b)$ denotes the nonparametric estimator using bandwidth b and is computed using all observations in the sample except observations included in subset m_i. In the simulations, m_i contains 20 percent of the observations in the sample. The parametric

[1] The results for the estimator of the cumulative incidence function based on local constant smoothing is available from the authors upon request.

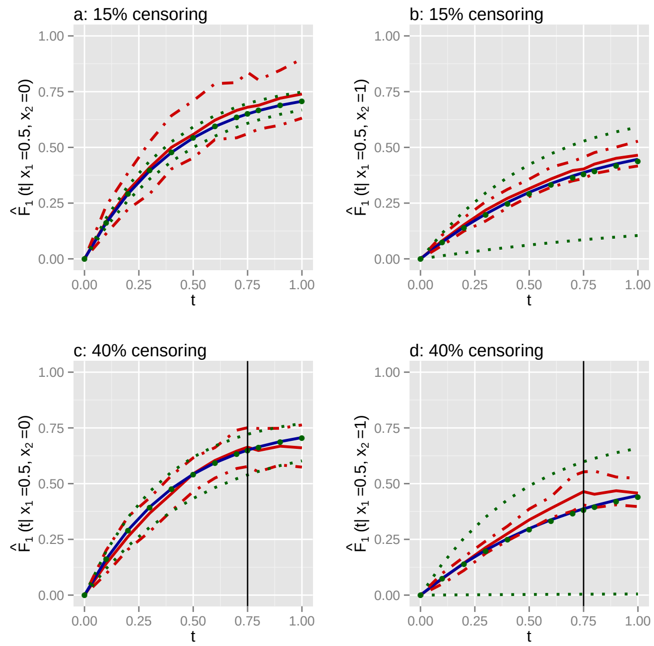

Figure 1: Nonparametric and parametric estimators of the cumulative incidence function (CIF): Blue line represents the true CIF; red solid line represents the nonparametric CIF estimator (the median of the estimator over the 499 runs and a sample size of 2000 observations); red dashed-dotted lines represent the 95 percent confidence intervals associated with the nonparametric CIF estimator; green solid line represents the parametric CIF estimator (median over 499 runs and a sample size of 2000 observations), while the green dotted lines are the 95 percent confidence bands. The vertical lines in Panels **c** and **d** indicate the censoring point (no observations in the sample exceeding this point).

benchmark estimator of the cumulative incidence function, denoted $\hat{F}_j^{IPW}(t|x_1, x_2; \widehat{\beta}_{IPW})$, is computed using the inverse probability weighted estimator suggested by Lu and Liang (2008). In order to make the benchmark as efficient as possible, the parametric data generating process and the parametric inverse probability weighting scheme applied in the simulations are always correctly specified.

The simulations are based on 499 runs with a sample size of $n = 2000$ observation under each of the four settings which are considered in Figure 1, Panels **a-d**.[2] In Panel **a**, the CIF functions $\hat{F}_1^C(|x_1 = 0.5, x_2 = 0)$ (red solid line), $\hat{F}^{IPW}(t|x_1 = 0.5, x_2 = 0; \widehat{\beta}_{IPW})$ (green solid line), and the corresponding 95 percent confidence bands are depicted when

[2] All simulations are done in R, and the R-scripts are available upon request.

the censoring level is 15 percent. The nonparametric CIF estimator performs quite well, although it appears to be slightly upward biased for larger values of t. In comparison to the parametric counterpart the nonparametric CIF estimators 95 percent confidence bands are also wider. The precision of the nonparametric estimator increases substantially in absolute terms but also relatively compared to the parametric estimator when evaluated at $x_2 = 1$, as illustrated in Panel **b**. In the absence of any form of parametric misspecification, this finding, in favor of the nonparametric CIF estimator, is somewhat surprising. Additional parameter estimation uncertainty related to $\widehat{\beta}_{2,IPW}$ might be one possible explanation for the relatively poor precision of the parametric CIF estimator in Panel **b**. The scenarios in Panels **c** and **d** are identical to Panels **a** and **b,** respectively, except that the censoring level has now been raised to **40** percent. The increased level of censoring as well as the increased level of failures with missing cause tend to decrease the precision of the parametric CIF estimator, whereas the confidence bands of the nonparametric CIF estimator remain almost unchanged. However, in Panel **d** the upward bias in the nonparametric CIF estimator is increasing but is still insignificant. Additional simulation results (not reported) suggest that by a different choice of bandwidth the bias could have been reduced at the expense of wider confidence bands. It should be noted that the vertical lines in Panels **c** and **d** indicate the censoring point and that there are no observations in the sample exceeding this point. This explains why the nonparametric CIF estimator becomes "flat" after the censoring point.

5 Conclusion

Studying estimators of cumulative incidence functions is important as these quantities are powerful tools for analyzing competing risks data which arise very often in different scientific fields such as demography, biostatistics, health economics and labor economics. This paper proposes a nonparametric method for estimating, for each risk, the corresponding cumulative incidence function in competing risks models, when continuous covariates affect the latent failure outcomes and the cause of failure is missing at random for some observations. The pointwise asymptotic normality as well as uniform convergence rate of the proposed estimators are derived. Existing estimation procedures, which account for covariates and missing cause if failure, are either fully parametric or semiparametric. In contrast to these estimation methods, the proposed estimator does not make any functional assumptions and thus it is robust under any specification of the underlying model. A simulation study shows that the proposed nonparametric estimator performs quite well relatively to a parametric benchmark.

Appendix 1

In this appendix a technical proof of Theorem 1 is provided. To obtain the asymptotic results, we need to introduce some extra notation. Consider the counting processes $N_{oi}(t) = 1\{T_i \leq t, \tilde{\gamma}_i > 0, R_i = 0\}$, which specify whether subject i has failed due to either risk 1 or risk 2 in the time interval $[0, t]$ and whether the cause of failure is missing. It is straightforward to see that $\bar{N}_i(t) = N_{1i}(t) + N_{2i}(t) + N_{oi}(t)$. For any $t > 0$, we have the filtration

$$\mathcal{F}_t = \sigma(N_{1i}(u), N_{2i}(u), N_{oi}(u), X_i, R_i, Y_i(u) : 0 \leq u \leq t, 1 \leq i \leq n),$$

where the notation σ specifies the sigma algebra generated by the events within the parenthesis. For each $j = 1, 2$, the counting processes $N_{ji}(t)$ and $N_{oi}(t)$ have stochastic intensities $\lambda_j(t, X_i)R_iY_i(t)$ and $\lambda(t, X_i)(1 - R_i)Y_i(t)$, respectively. That is,

$$\lambda_j(t, X_i)R_iY_i(t)dt = \mathbb{E}((N_{ji}(t + dt)_-) - N_{ji}(t)_-|\mathcal{F}_{t^-}),$$

$$\lambda(t, X_i)(1 - R_i)Y_i(t)dt = \mathbb{E}((N_{oi}(t + dt)_-) - N_{oi}(t)_-|\mathcal{F}_{t^-}),$$

where $N_{ji}(t)_- = \lim_{u\uparrow t} N_{ji}(u)$ and $N_{oi}(t)_- = \lim_{u\uparrow t} N_{oi}(u)$. The stochastic intensities $\lambda_j(t, X_i)R_iY_i(t)$ and $\lambda(t, X_i)(1 - R_i)Y_i(t)$ are predictable with respect to $\mathcal{F}_t$.

For each $t > 0$, $j = 1, 2$, and $i = 1, ..., n$, consider the $\mathcal{F}_t$-measurable processes

$$M_{ji}(t) = N_{ji}(t) - \int_0^t \lambda_j(u, X_i)R_iY_i(u)du, \qquad M_{oi}(u) = N_{oi}(t) - \int_0^t \lambda(u, X_i)(1 - R_i)Y_i(u)du.$$

Working analogously to the proof of Theorem 1 (page 311) of Shorack and Wellner (2009), we can show that $M_{ji}(t)$ and $M_{oi}(t)$ are zero-mean (by using (6) for this) local square integrable martingales with respect to filtration $\mathcal{F}_t$.

For $\xi = 1, 2$, with $\xi \neq j$, introduce the quantities

$$\mathcal{V}_{jj}(t,x) = \frac{1}{n}\sum_{i=1}^{n}\int_0^t \frac{\mathcal{K}_h(x - X_i)}{\pi(X_i, \tilde{\gamma}_i)} \left[g(u,x) + \pi(X_i, \tilde{\gamma}_i)\rho_j(t,u,x)\right] dM_{ji}(u)$$
$$\mathcal{V}_{j\xi}(t,x) = \frac{1}{n}\sum_{i=1}^{n}\int_0^t \mathcal{K}_h(x - X_i)\rho_j(t,u,x) dM_{\xi i}(u)$$
$$\mathcal{V}_{jo}(t,x) = \frac{1}{n}\sum_{i=1}^{n}\int_0^t \mathcal{K}_h(x - X_i)\rho_j(t,u,x) dM_{oi}(u)$$

and

$$\mathcal{B}_{jj}(t,x) = \frac{1}{n}\sum_{i=1}^{n}\int_0^t \mathcal{K}_h(x - X_i)Y_i(u)\left[g(u,x)\frac{R_i}{\pi(X_i, \tilde{\gamma}_i)} + \rho_j(t,u,x)\right] \times \left[\lambda_j(u, X_i) - \lambda_j(u,x)\right] du,$$
$$\mathcal{B}_{j\xi}(t,x) = \frac{1}{n}\sum_{i=1}^{n}\int_0^t \rho_j(t,u,x)\mathcal{K}_h(x - X_i)Y_i(u)\left[\lambda_\xi(u, X_i) - \lambda_\xi(u,x)\right] du.$$

Proof of Theorem 1. Similar to Linton et al. (2011), we will show the asymptotic normality for the estimator $\hat{F}_j^C(t|x)$. The asymptotic distribution for $\hat{F}_j^L(t|x)$ can be derived by following similar arguments. For ease of notation we skip the superscript C.

We write

$$\begin{aligned}\hat{F}_j(t|x) - F_j(t|x) &= \int_0^t \hat{S}(u - |x) d\left[\hat{\Lambda}_j(u,x) - \Lambda_j(u,x)\right] \\ &\quad + \int_0^t \left[\hat{S}(u - |x) - S(u - |x)\right]\lambda_j(u,x) du \\ &=: \hat{\Upsilon}_j(t,x) + \hat{\Omega}_j(t,x). \end{aligned} \tag{A-1}$$

By property $M_{ji}(t) = N_{ji}(t) - \int_0^t \lambda_j(u, X_i)Y_i(u)R_i du$, it follows that

$$\begin{aligned}
\hat{\Upsilon}_j(t,x) &= \frac{1}{n}\sum_{i=1}^n \int_0^t \hat{S}(u-|x)\frac{R_i}{\hat{\pi}(X_i,\tilde{\gamma}_i)} \frac{\mathcal{K}_h(x-X_i)dM_{ji}(u)}{\frac{1}{n}\sum_{i=1}^n \frac{R_i}{\hat{\pi}(X_i,\tilde{\gamma}_i)}\mathcal{K}_h(x-X_i)Y_i(u)} \\
&+ \frac{1}{n}\sum_{i=1}^n \int_0^t \hat{S}(u-|x)\frac{R_i}{\hat{\pi}(X_i,\tilde{\gamma}_i)} \frac{\mathcal{K}_h(x-X_i)Y_i(u)\left[\lambda_j(u,X_i)-\lambda_j(u,x)\right]}{\frac{1}{n}\sum_{i=1}^n \frac{R_i}{\hat{\pi}(X_i,\tilde{\gamma}_i)}\mathcal{K}_h(x-X_i)Y_i(u)}du. \qquad \text{(A-2)}
\end{aligned}$$

Next, we work on $\hat{\Omega}_j(t,x)$ by making use of the Duhamel equation (Gill, 1994) $\hat{S}(t|x) - S(t|x) = -S(t|x)\int_0^t \frac{\hat{S}(u-|x)}{S(u|x)} d\left[\hat{\Lambda}(u,x) - \Lambda(u,x)\right]$. By using the equalities $\bar{N}_i(t) = N_{1i}(t) + N_{2i}(t) + N_{oi}(t)$, $M_{ji}(t) = N_{ji}(t) - \int_0^t \lambda_j(u,X_i)R_iY_i(u)du$ and $M_{oi}(t) = N_{oi}(t) - \int_0^t \lambda(u,X_i)(1-R_i)Y_i(u)du$, the Duhamel formula, and the fact that the mapping $t \mapsto S(t|x)$ is continuous for all $x \in \mathcal{X}_h$, as well as doing some algebra, we obtain

$$\begin{aligned}
\hat{\Omega}_j(t,x) = &-\frac{1}{n}\sum_{i=1}^n \int_0^t \left[\int_u^t S(\epsilon|x)\lambda_j(\epsilon,x)d\epsilon\right]\frac{\hat{S}(u-|x)}{S(u|x)} \\
&\times \frac{\mathcal{K}_h(x-X_i)}{\frac{1}{n}\sum_{i=1}^n \mathcal{K}_h(x-X_i)Y_i(u)} d(M_{1i}(u)+M_{2i}(u)+M_{oi}(u)) \\
&- \frac{1}{n}\sum_{i=1}^n \int_0^t \frac{\hat{S}(u-|x)}{S(u|x)} \frac{\mathcal{K}_h(x-X_i)Y_i(u)\left[\lambda(u,X_i)-\lambda(u,x)\right]}{\frac{1}{n}\sum_{i=1}^n \mathcal{K}_h(x-X_i)Y_i(u)} \\
&\times \left[\int_u^t S(\epsilon|x)\lambda_j(\epsilon,x)d\epsilon\right] du. \qquad \text{(A-3)}
\end{aligned}$$

By continuity of the mapping $t \mapsto S(t|x)$ and using uniform convergence results (Lecoutre and Ould-Said, 1995), we obtain $\frac{\hat{S}(t-|x)}{S(t|x)} = 1 + o_p(1)$ uniformly over $t \in [0, \tau(x)]$ for each $x \in \mathcal{X}_h$. Moreover, $\hat{\pi}(X_i, \tilde{\gamma}_i) = \pi(X_i, \tilde{\gamma}_i) + o_p(1)$ uniformly over $= 1, ..., n$ (e.g., Hansen, 2008) with $\pi(X_i, \tilde{\gamma}_i)$ to be bounded away from zero. Similarly, we can show pointwise in x, $\sup_{t\in[0,\tau(x)]}\left|\frac{1}{n}\sum_{i=1}^n \frac{R_i}{\pi(X_i,\tilde{\gamma}_i)}\mathcal{K}_h(x-X_i)Y_i(t)\right| = H(t,x) + o_p(1)$ by noting that $\mathbb{E}(R_i|\tilde{\gamma}_i, X_i, T_i) = \pi(X_i, \tilde{\gamma}_i)$, and also $\sup_{t\in[0,\tau(x)]}\left|\frac{1}{n}\sum_{i=1}^n \mathcal{K}_h(x-X_i)Y_i(t)\right| = H(t,x) + o_p(1)$, where we use continuity of the map $t \mapsto H(t,x)$ for the two latter results.

Recall that $\lambda(t,x) = \lambda_1(t,x) + \lambda_2(t,x)$ for any $(t,x) \in \mathcal{R}_+ \times \mathcal{X}_h$. Combining (A-1)-(A-3) and the previous uniform convergence results, we get

$$\hat{F}_j(t|x) - F_j(t|x) = [\mathcal{V}_{j1}(t,x) + \mathcal{V}_{j\xi}(t,x) + \mathcal{V}_{oj}(t,x) + \mathcal{B}_{jj}(t,x) + \mathcal{B}_{j\xi}(t,x)]\,[1 + o_p(1)]\,. \quad \text{(A-4)}$$

for each $t \in [0, \tau(x)]$ and $x \in \mathcal{X}_h$. The expression is equivalent to

$$\begin{aligned}\hat{F}_j(t|x) - F_j(t|x) = \{&\mathcal{V}_{j1}(t,x) + \mathcal{V}_{j2}(t,x) + \mathcal{V}_{jo}(t,x) + \mathbb{E}\mathcal{B}_{jj}(t,x) + \mathbb{E}\mathcal{B}_{j\xi}(t,x) \\ &+ [\mathcal{B}_{jj}(t,x) - \mathbb{E}\mathcal{B}_{jj}(t,x)] + [\mathcal{B}_{j\xi}(t,x) - \mathbb{E}\mathcal{B}_{j\xi}(t,x)]\}\,[1 + o_p(1)]\,. \quad \text{(A-5)}\end{aligned}$$

To derive the asymptotic distribution of $\hat{F}_j(t|x) - F_j(t|x)$, it suffices to consider the term { }. Application of the martingale central limit theorem (Andersen et al., 1993; Nielsen and Linton, 1995; Linton et al., 2011) yields $\sqrt{nh^d}\mathcal{V}_{j1}(t,x) \Longrightarrow \mathcal{N}(0, v_{jj}(t,x))$, $\sqrt{nh^d}\mathcal{V}_{j\xi}(t,x) \Longrightarrow \mathcal{N}(0, v_{j\xi}(t,x))$, and $\sqrt{nh^d}\mathcal{V}_{jo}(t,x) \Longrightarrow \mathcal{N}(0, v_{jo}(t,x))$, where

$$\begin{aligned} v_{jj}(t,x) &= p\lim_{n\to\infty}\left[\frac{h^d}{n}\sum_{i=1}^{n}\int_0^t \frac{\mathcal{K}_h^2(x-X_i)}{\pi^2(X_i,\tilde{\gamma}_i)}\left[g(u,x) + \pi(X_i,\tilde{\gamma}_i)\rho_j(t,u,x)\right]^2 d < M_{ji} > (u)\right] \\ &= p\lim_{n\to\infty}\left[\frac{h^d}{n}\sum_{i=1}^{n}\int_0^t \frac{\mathcal{K}_h^2(x-X_i)}{\pi^2(X_i,\tilde{\gamma}_i)}\left[g(u,x) + \pi(X_i,\tilde{\gamma}_i)\rho_j(t,u,x)\right]^2 \lambda_j(u,X_i)R_iY_i(u)du\right] \\ &= \frac{||K||_2^2}{\pi(x)}\int_0^t H(u,x,\tilde{\gamma}>0)\left[g(u,x) + \pi(x)\rho_j(t,u,x)\right]^2\lambda_j(u,x)du \\ &+ ||K||_2^2\int_0^t H(u,x,\tilde{\gamma}=0)\left[g(u,x) + \rho_j(t,u,x)\right]^2\lambda_j(u,x)du, \end{aligned} \quad \text{(A-6)}$$

$$\begin{aligned} v_{j\xi}(t,x) &= p\lim_{n\to\infty}\left[\frac{h^d}{n}\sum_{i=1}^{n}\int_0^t \mathcal{K}_h^2(x-X_i)\rho_j^2(t,u,x)d < M_{\xi i} > (u)\right] \\ &= p\lim_{n\to\infty}\left[\frac{h^d}{n}\sum_{i=1}^{n}\int_0^t \mathcal{K}_h^2(x-X_i)\rho_j^2(t,u,x)\lambda_\xi(u,X_i)R_iY_i(u)du\right] \\ &= ||K||_2^2\,\pi(x)\int_0^t H(u,x,\tilde{\gamma}>0)\rho_j^2(t,u,x)\lambda_\xi(u,x)du \\ &+ ||K||_2^2\int_0^t H(u,x,\tilde{\gamma}=0)\rho_j^2(t,u,x)\lambda_\xi(u,x)du, \end{aligned} \quad \text{(A-7)}$$

and

$$
\begin{aligned}
v_{jo}(t,x) &= p\lim_{n\to\infty}\left[\frac{h^d}{n}\sum_{i=1}^{n}\int_0^t \mathcal{K}_h^2(x-X_i)\rho_j^2(t,u,x)d<M_{oi}>(u)\right]\\
&= p\lim_{n\to\infty}\left[\frac{h^d}{n}\sum_{i=1}^{n}\int_0^t \mathcal{K}_h^2(x-X_i)\rho_j^2(t,u,x)\lambda(u,X_i)(1-R_i)Y_i(u)du\right]\\
&= ||K||_2^2\,[1-\pi(x)]\int_0^t H(u,x,\tilde{\gamma}>0)\rho_j^2(t,u,x)\lambda(u,x)du, \qquad \text{(A-8)}
\end{aligned}
$$

where the equalities in (A-6), (A-7), and (A-8) follow by using the MAR property, the fact that $\pi(x,\tilde{\gamma}) = 1\{\tilde{\gamma} > 0\}\pi(x) + 1\{\tilde{\gamma} = 0\}$, the definition of $\mathcal{K}_h(x-X_i)$, the change of variables, and the dominated convergence theorem. By construction, as soon as one of the counting processes $N_{1i}(t)$, $N_{2i}(t), N_{oi}(t)$ jumps, the other ones cannot jump. The latter, combined with the equality $\lambda(t,x) = \lambda_1(t,x) + \lambda_2(t,x)$ and some algebra, implies $\sqrt{nh^d}\,[\mathcal{V}_{jj}(t,x) + \mathcal{V}_{j\xi}(t,x) + \mathcal{V}_{jo}(t,x)] \Longrightarrow \mathcal{N}(0, v_{jA}(t,x) + v_{jB}(t,x) + v_{jAB}(t,x))$. Next, we proceed with the stable parts in a similar manner to the approach of Nielsen and Linton (1995). Let $\omega = (\omega_1, \omega_2, ..., \omega_d)$. It is easy to check that

$$
\mathbb{E}\mathcal{B}_{jj}(t,x) = \sum_{l=0}^{1}\frac{\mu_2(K)h^2}{(2-l)!l!}\sum_{p=1}^{d}\int_0^t \frac{\partial^{2-l}\lambda_j(u,x)}{\partial x_p^{2-l}}\frac{\partial^l H(u,x)}{\partial x_p^l}\left[g(u,x)+\rho_j(t,u,x)\right]du + o(h^2), \qquad \text{(A-9)}
$$

$$
\mathbb{E}\mathcal{B}_{j\xi}(t,x) = \sum_{l=0}^{1}\frac{\mu_2(K)h^2}{(2-l)!l!}\sum_{p=1}^{d}\int_0^t \frac{\partial^{2-l}\lambda_\xi(u,x)}{\partial x_p^{2-l}}\frac{\partial^l H(u,x)}{\partial x_p^l}\rho_j(t,u,x)du + o(h^2), \qquad \text{(A-10)}
$$

where the two above equations are obtained by using the MAR property, the definition of $\mathcal{K}_h(x-X_i)$ and $r-$th Taylor series expansion with a Lagrange remainder for the difference $\lambda_j(u,x-h\omega) - \lambda_j(u,x)$ $(j = 1,2)$ and the quantity $H(u,x-h\omega)$, along with the fact that the K is of order 2. Also, by working in a completely analogous way it is straightforward to show that $\mathbb{E}\mathcal{B}_{j\iota}^2(t,x) = O(nh^{d-2})^{-1}$ for $\iota = j,\xi$, which in turn gives $\mathbb{E}(\mathcal{B}_{j\iota}(t,x) - \mathbb{E}\mathcal{B}_{j\iota}(t,x))^2 = O(nh^{d-2})^{-1} = o(nh^d)^{-1}$ and consequently, by Chebyshev's inequality, $|\mathcal{B}_{j\iota}(t,x) - \mathbb{E}\mathcal{B}_{j\iota}(t,x)| = o_p(nh^d)^{-\frac{1}{2}}$.

Application of the martingale central limit theorem for the covariance terms yields

$$\varsigma_{11}(t,x) := p\lim_{n\to\infty}\left[\frac{h^d}{n}\sum_{i=1}^{n}\int_0^t \frac{\mathcal{K}_h^2(x-X_i)}{\pi(X_i,\tilde{\gamma}_i)}[g(u,x)+\pi(X_i,\tilde{\gamma}_i)\rho_1(t,u,x)]\rho_2(t,u,x)d<M_{1i}>(u)\right]$$
$$= ||K||_2^2\int_0^t H(u,x,\gamma>0)[g(u,x)+\pi(x)\rho_1(t,u,x)]\rho_2(t,u,x)\lambda_1(u,x)du$$
$$+ ||K||_2^2\int_0^t H(u,x,\gamma=0)[g(u,x)+\rho_1(t,u,x)]\rho_2(t,u,x)\lambda_1(u,x)du, \tag{A-11}$$

$$\varsigma_{22}(t,x) := p\lim_{n\to\infty}\left[\frac{h^d}{n}\sum_{i=1}^{n}\int_0^t \frac{\mathcal{K}_h^2(x-X_i)}{\pi(X_i,\tilde{\gamma}_i)}[g(u,x)+\pi(X_i,\tilde{\gamma}_i)\rho_2(t,u,x)]\rho_1(t,u,x)d<M_{2i}>(u)\right]$$
$$= ||K||_2^2\int_0^t H(u,x,\gamma>0)[g(u,x)+\pi(x)\rho_2(t,u,x)]\rho_1(t,u,x)\lambda_2(u,x)du$$
$$+ ||K||_2^2\int_0^t H(u,x,\gamma=0)[g(u,x)+\rho_2(t,u,x)]\rho_1(t,u,x)\lambda_2(u,x)du, \tag{A-12}$$

and

$$\varsigma_o(t,x) := p\lim_{n\to\infty}\left[\frac{h^d}{n}\sum_{i=1}^{n}\int_0^t \mathcal{K}_h^2(x-X_i)\rho_1(t,u,x)\rho_2(t,u,x)d<M_{oi}>(u)\right]$$
$$= ||K||_2^2\,[1-\pi(x)]\int_0^t H(u,x,\gamma>0)\rho_1(t,u,x)\rho_2(t,u,x)\lambda(u,x)du, \tag{A-13}$$

where we make use of the MAR property, the equality $\pi(x,\tilde{\gamma}) = 1\{\tilde{\gamma}>0\}\pi(x) + 1\{\tilde{\gamma}=0\}$, the definition of $\mathcal{K}_h(x-X_i)$, a change of variables, and the dominated convergence theorem. By simple algebra, we have $\varsigma_1(t,x)+\varsigma_2(t,x) = \varsigma_{11}(t,x)+\varsigma_{22}(t,x)+\varsigma_o(t,x)$. ■

Appendix 2

In this appendix a technical proof of Theorem 2 is provided. We restrict our attention to the local constant estimator. Similar algebraic calculations can be carried out for the local linear estimator and therefore we will skip this part. To keep the notation simple, we omit the superscript C.

Recall that

$$\hat{\Lambda}_j(t,x) = \frac{1}{n}\sum_{i=1}^{n}\int_0^t \frac{w_i(x)}{\hat{H}(u-|x)} dN_{ji}(u), \;\; j = 1,2$$

Define $F_{\tilde{\gamma}=j}(t,x) = F_{\tilde{\gamma}=j}(t|x)f(x)$, where $F_{\tilde{\gamma}=j}(t|x) = \mathbb{P}(T \leq t, \tilde{\gamma} = j|x)$. Moreover,

$$\hat{F}_{\tilde{\gamma}=j}(t,x) = \frac{1}{n}\sum_{i=1}^{n}\frac{R_i}{\hat{\pi}(X_i,\tilde{\gamma}_i)}\mathcal{K}_h(x-X_i)N_{ji}(t) \tag{A-14}$$

and

$$\hat{H}(t-,x) = \frac{1}{n}\sum_{i=1}^{n}\frac{R_i}{\hat{\pi}(X_i,\tilde{\gamma}_i)}\mathcal{K}_h(x-X_i)Y_i(t), \tag{A-15}$$

It is then easy to see (by making use of the definition for the weights $w_i(x)$) that

$$\hat{\Lambda}_j(t,x) = \int_0^t \frac{d\hat{F}_{\tilde{\gamma}=j}(u,x)}{\hat{H}(u-,x)}$$

The following lemma deals with the uniform convergence rates of $\hat{F}_{\tilde{\gamma}=j}(t,x)$ and $\hat{H}(t-,x)$.

Lemma 1 *Suppose Assumptions 1-3, and 4* hold. Then, it holds for $j = 1,2,$ as $n \to \infty$,*

$$\sup_{(t,x)\in\Xi}\left|\hat{F}_{\tilde{\gamma}=j}(t,x) - F_{\tilde{\gamma}=j}(t,x)\right| = O_p(\alpha_n)$$

and

$$\sup_{(t,x)\in\Xi}\left|\hat{H}(t-,x) - H(t-,x)\right| = O_p(\alpha_n).$$

Proof. Let $\beta_n \equiv \left(\frac{\ln n}{nh^d}\right)^{\frac{1}{2}} + h$. By Hansen (2008), $\hat{\pi}(X_i) - \pi(X_i) = O_p(\alpha_n)$ uniformly in $i = 1,...,n$, if $X_i \in \mathcal{X}_h$, and $\hat{\pi}(X_i) - \pi(X_i) = O_p(\beta_n)$ uniformly in $i = 1,...,n$, if $X_i \in \mathcal{X}\backslash\mathcal{X}_h$.

Consequently,

$$\hat{F}_{\tilde{\gamma}=j}(t,x) = \frac{1}{n}\sum_{i=1}^{n}\frac{R_i}{\pi(X_i,\tilde{\gamma}_i)}\mathcal{K}_h(x-X_i)N_{ji}(t)\left[1+O_p(\alpha_n)1\{X_i\in\mathcal{X}_h\}+O_p(\beta_n)1\{X_i\in\mathcal{X}\backslash\mathcal{X}_h\}\right]$$

uniformly over Ξ. Given that $1\{X_i \in \mathcal{X}_h\} = O_p(1)$ and $1\{X_i \in \mathcal{X}\backslash\mathcal{X}_h\} = O_p(h)$, we have uniformly over Ξ

$$\hat{F}_{\tilde{\gamma}=j}(t,x) = \frac{1}{n}\sum_{i=1}^{n}\frac{R_i}{\pi(X_i,\tilde{\gamma}_i)}\mathcal{K}_h(x-X_i)N_{ji}(t)\left[1+O_p(\alpha_n)\right]. \tag{A-16}$$

Using results developed by Lecoutre and Ould-Said (1995) (proof Theorem 2) we can deduce that

$$\frac{1}{n}\sum_{i=1}^{n}\frac{R_i}{\pi(X_i,\tilde{\gamma}_i)}\mathcal{K}_h(x-X_i)N_{ji}(t) = F_{\tilde{\gamma}=j}(t,x)+O_p(\alpha_n). \tag{A-17}$$

Combining (A-16) and (A-17), it follows uniformly over Ξ

$$\hat{F}_{\tilde{\gamma}=j}(t,x) = F_{\tilde{\gamma}=j}(t,x)+O_p(\alpha_n).$$

The proof of the second statement of the Lemma is similar to the above procedure and thus omitted. ■

To prove the main result, we need also the following result.

Lemma 2 *Suppose Assumptions 1-3, and 4* hold. Then, it holds for* $j=1,2,$ *as* $n\to\infty$,

$$\sup_{(t,x)\in\Xi}\left|\hat{\Lambda}_j(t,x)-\Lambda_j(t,x)\right| = O_p(\alpha_n).$$

Proof. Using similar arguments as in the proof of Theorem 1 of Lecoutre and Ould-Said (1995) and employing Lemma 1, we can obtain the desired result. ■

Now we proceed with the proof of Theorem 2.

Proof of Theorem 2. Clearly,

$$\hat{F}_j(t|x) - F_j(t|x) = \int_0^t \hat{S}(u-|x) d\left[\hat{\Lambda}_j(u,x) - \Lambda_j(u,x)\right] + \int_0^t \left[\hat{S}(u-|x) - S(u-|x)\right] \lambda_j(u,x) du$$
$$=: \hat{\Upsilon}_j(t,x) + \hat{\Omega}_j(t,x), \tag{A-18}$$

Triangle inequality entails

$$\sup_{(t,x)\in\Xi} \left|\hat{F}_j(t|x) - F_j(t|x)\right| \leq \sup_{(t,x)\in\Xi} \left|\hat{\Upsilon}_j(t,x)\right| + \sup_{(t,x)\in\Xi} \left|\hat{\Omega}_j(t,x)\right|. \tag{A-19}$$

Partial integration, triangle inequality and use of Lemma 2 yields for the first term

$$\sup_{(t,x)\in\Xi} \left|\hat{\Upsilon}_j(t,x)\right| \leq \sup_{(t,x)\in\Xi} \left|\hat{\Lambda}_j(t,x) - \Lambda_j(t,x)\right| \sup_{(t,x)\in\Xi} \left|\hat{S}(t-|x)\right|$$
$$+ \sup_{(t,x)\in\Xi} \left|\hat{\Lambda}_j(t,x) - \Lambda_j(t,x)\right| \sup_{(t,x)\in\Xi} \left|\int_0^t d\hat{S}(u-|x)\right| = O_p(\alpha_n), \tag{A-20}$$

as $\sup_{(t,x)\in\Xi} \left|\hat{S}(t-|x)\right| = O_p(1)$ (Lecoutre and Ould-Said, 1995). Regarding the second term, it is straightforward to check that

$$\sup_{(t,x)\in\Xi} \left|\hat{\Omega}_j(t,x)\right| \leq \sup_{(t,x)\in\Xi} \left|\hat{S}(t-|x) - S(t-|x)\right| \sup_{(t,x)\in\Xi} |\Lambda_j(t,x)| = O_p(\alpha_n), \tag{A-21}$$

where we apply Theorem 2 of Lecoutre and Ould-Said (1995). Combining (A-19)-(A-21) completes the proof. ■

Acknowledgements

We thank Gerard J. van den Berg, Christoph Breunig, Enno Mammen, and Geert Ridder for insightful comments and suggestions. We are also grateful to the two referees for a number of suggestions which helped us to improve the presentation. Finally, we would like to thank Wenbin Lu for sharing his code with us. All errors are our own responsibility.

References

Andersen, P. K., Borgan, Ø., Gill, R. D., and Keiding, N. (1993), *Statistical models based on counting processes*, Springer Verlag.

Bordes, L. and Gneyou, K. E. (2011), "Uniform convergence of nonparametric regressions in competing risk models with right censoring," *Statistics and Probability Letters*, 81, 1654 – 1663.

Gill, R. (1994), "Lectures on survival analysis," *Lectures on Probability Theory*, 115–241.

Hansen, B. E. (2008), "Uniform convergence rates for kernel estimation with dependent data," *Econometric Theory*, 24, 726–746.

Hu, Z., Follmann, D. A., and Qin, J. (2010), "Semiparametric dimension reduction estimation for mean response with missing data," *Biometrika*, 97, 305–319.

Lecoutre, J.-P. and Ould-Said, E. (1995), "Convergence of the conditional Kaplan-Meier estimate under strong mixing," *Journal of statistical planning and inference*, 44, 359–369.

Li, Q. and Racine, J. S. (2008), "Nonparametric estimation of eonditional cdf and quantile functions with mixed categorial and continous data," *Journal of Business and Economic Statistics*, 26, 423–434.

Linton, O., Mammen, E., Nielsen, J. P., and Van Keilegom, I. (2011), "Nonparametric regression with filtered data," *Bernoulli*, 17, 60–87.

Little, R. J. A. and Rubin, D. B. (1987), *Statistical analysis with missing data*, vol. 4, Wiley New York.

Lu, W. and Liang, Y. (2008), "Analysis of competing risks data with missing cause of failure under addtive hazards model," *Statistica Sinica*, 19, 219–234.

Nielsen, J. P. and Linton, O. B. (1995), "Kernel estimation in a nonparametric marker dependent hazard model," *Annals of Statistics*, 23, 1735–1748.

Rubin, D. (1976), "Inference and missing data," *Biometrika*, 63, 581–592.

Shorack, G. R. and Wellner, J. A. (2009), *Empirical processes with applications to statistics*, vol. 59, Society for Industrial Mathematics.