NONPARAMETRIC ESTIMATION OF COMPETING RISKS MODELS WITH COVARIATES

JEAN-DAVID FERMANIAN

CREST, Malakoff, France

RUNNING HEAD : nonparametrically estimated competing risks.

ADDRESS FOR CORRESPONDENCE : J.-D. FERMANIAN
20 RUE DU 22 SEPTEMBRE
92400 COURBEVOIE. FRANCE.
E-MAIL : fermania@ensae.fr

Abstract

In competing risks model, several failure times arise potentially. The smallest failure time and its index only are observed. Without specific assumptions, the joint or even the marginal distribution functions of the underlying failure times are not identifiable (Tsiatis, 1975, [33]). Nonetheless, if each individual is characterized by a "sufficiently informative" set of covariates, these distributions are identifiable under some conditions of regularity (Heckman and Honoré, 1989, [17])). In this paper, nonparametric kernel estimators of the joint distribution function of failure times conditional on the covariates are proposed. Their weak and strong consistency are discussed.

AMS 1991 *subject classification.* Primary 62N02, 62G05, secondary 62H12.
Key words and phrases. Competing risks, nonparametric estimation, covariates, kernel method.

1. Introduction

Competing risks models arise very often in several fields : biostatistics, reliability, finance, labor economics... They are relevant when two or more causes of failure act simultaneously, but the smallest failure and its type only are observed. In other words, each failure time is potentially right censored by every other failure times. A key point is that all these failures are dependent a priori . Thus, they can not be dealt with the standard arguments of random censoring models. For instance, an unemployed worker can find a new job or quit the labor market (Flinn and Heckman [10]). Only one of these two events occurs for each individual. The longer the duration of unemployment, the longer the probability that he renounces finding a new job. Hence, we are faced surely with two non independent competing risks. In finance, before the full reimbursement of a loan, the creditor faces several risks concerning the borrower and disturbing eventually the reimbursement : unemployment, insolvency, or conversely advanced reimbursement. More generally, in most duration models with right censored data (e.g. the Cox model), the usual assumption of independence between failure and censoring processes is often doubtful and questionable. When this assumption is relaxed, we are faced with a competing risks model. In this case, without parametric assumptions on the joint d.f. of the underlying failure times, usual methods provide at best biased estimates.

More precisely, consider two failure times T_1 and T_2, viz positive a priori dependent random variables. Let $Y = T_1 \wedge T_2$ and $\delta = \mathbf{1}\{T_1 < T_2\}$. Observe an i.i.d sample $(Y_i, \delta_i)_{i=1,\ldots,n}$. The problem of estimating the joint distribution of (T_1, T_2) from these observations has been studied for a long time (see the survey of Crowder [5]). Briefly, Tsiatis [33] proved that for any distribution of (Y, δ), there exist independent r.v. T_1 and T_2 that provide such a distribution. Then, the joint distribution of (T_1, T_2) cannot be identified. For some authors, this impasse provided a key argument to concentrate no more on the latter joint distribution but on observed quantities only. Thus, they are rather estimate some "crude" or "specific" hazard rates

$$\tilde{\lambda}_j(t) = \lim_{\Delta t \to 0} \frac{1}{\Delta t} P(T_j \in [t, t + \Delta t], \delta_j = 1 | Y > t), \quad j = 1, 2, \tag{1.1}$$

instead of the "overall" or "latent" hazard rates

$$\lambda_j(t) = \lim_{\Delta t \to 0} \frac{1}{\Delta t} P(T_j \in [t, t + \Delta t] | T_j > t), \quad j = 1, 2. \tag{1.2}$$

Indeed, the "crude" hazard rates are easily estimated empirically, when the "overall" hazard rates are not identifiable. Even more, some of the authors argue that discussions about one risk only, as if the other risks do not exist, are nonsensical (see e.g. Prentice and Kalbfleisch [27], Kalbfleisch and Prentice [20], Crowder [5] and the opposite point of view in Slud [31]). Roughly, the main non technical argument is : "In the real world, all these risks are always present simultaneously".

Nonetheless, particularly in econometrics and finance, models with latent variables are commonly used. There exist several underlying "structural" variables (here the failure times) that are of interest. Some failures are key variables in their particular fields. That is why, most of the time, it is necessary to study a particular event independently from the other failure times. This is especially true when there are many other competing risks or when these risks are not the same from one experiment to another. In the case of competing risks, the goal is most often to estimate marginal and joint distributions of some failure times of interest. This will be our point of view.

Actually, the estimation of the distributions of the underlying latent variables has induced a large amount of literature in a lot of applied fields. In reliability, the behavior of individual components in complex engineering systems is crucial. For instance, Kwan and Singh [21] provide nonparametric estimates of the distribution function for every latent risks, when they are assumed mutually independent. In econometrics, the analysis of the labor market is faced with competing risks very usually. Particularly, Han and Hausman [15] estimate a proportional hazard model with heterogeneity in a competing risks framework. They apply their model to unemployment spells described in the Panel Study of Income Dynamics (PSID). In this study, the failures are due to recalls from unemployment to a previous job, new jobs or censorship. This methodology has been extended by Sueyoshi [32] and applied by McCall [23], among others.

Until now, authors have no choice to deal with the lack of identifiability in practice : either they impose some parametric assumptions on the joint distribution of failure times, or they suppose the independence between these random variables (even if it is not absolutely necessary, as proved by Langberg et al. [22]). For instance, the latter choice has been made by Giannelli [12] to study the dynamics of married women's labor market transitions in West Germany, especially transitions between full-time work, part-time work and non-employment.

These solutions are most of the time not fully satisfactory. We have seen that an independence assumption is often unrealistic. Moreover, parametric assumptions are not always necessary. Indeed, when covariates are inserted in the model, and under some conditions on their joint distribution, it is possible to identify the joint distribution of (T_1, T_2) (Heckman and Honoré [17]). We will consider this latter framework. Note that, in some other particular cases, the identifiability can be achieved : by incorporating prior information in a bayesian approach (Gasbarra and Karia [11], e.g.), when the copula function of the underlying risks is known (Zheng and Klein [34]), by assuming a type of independence conditional on a covariate (Slud [31])...

The goal of this paper is to provide explicit estimators of the joint distribution of all underlying failure times conditional on the covariates, and to discuss their statistical properties as far as possible. More generally we propose a global strategy to estimate non- or semi parametrically all the parameters of such models (written like in Heckman and Honoré [17]).

Therefore, the knowledge of the joint cdf will induce the knowledge of marginal distributions, what is obviously of the highest importance. Indeed, one source of failure is most often of primary interest for the researcher, and the others failures are considered as disturbances. More generally, when some of the latent risks are absent from the experiment, the previously estimated model can still be used. For instance, assume that a production line is faced with four different risks of failures. We have estimated the joint cdf of all these risks. Now, a new long-life component has been installed in the production line. The risk of failure caused by this component is now considered as zero. Thus, it remains three risks only from now on. The production manager could ask himself : now, what is the new probability of failures by one year ? And which of the three remainder risks is the highest ? Our methodology allows to answer these questions. A "crude" approach is not able to do the same.

After a description of our framework, we propose to estimate respectively all the unknown functions of the model. Then, we provide some simulations to get an idea of their performances, and we conclude with some comments. The proofs are expanded in the appendix of the paper.

2. The framework

For the sake of generality, it is relevant to consider p-dimensional distributions functions, $p \geq 2$. Let $T = (T_1, \ldots, T_p)$ be a p-dimensional vector of (a priori correlated) failure times, $Y = T_1 \wedge \ldots \wedge T_p$ and δ be the index in $\{1, \ldots, p\}$ of the smallest of these failure times. Let S and Q be the survival functions of $(T_1, \ldots, T_p)$ and Y respectively. Moreover, each individual or entity is characterized by a $\mathbb{R}^d$-valued covariate X, which is supposed to be time independent.

Extending Heckman and Honoré [17], suppose that S can be rewritten as

$$S(t_1, \ldots, t_p | X = x) = H\left(\exp\left(-Z_1(t_1)\phi_1(x)\right), \ldots, \exp\left(-Z_p(t_p)\phi_p(x)\right)\right), \tag{2.3}$$

where $H : [0,1]^p \longrightarrow [0,1]$ is a cumulative distribution function and $\phi_1, \dots, \phi_p$ are a priori unknown functions from $\mathbb{R}^d$ to $\mathbb{R}^+$. The Z_j, $j = 1, \dots, p$ are unknown non decreasing functions that satisfy $Z_1(0) = \dots = Z_p(0) = 0$.

This specification covers most of the commonly used models of multivariate survival analysis. Indeed, if each marginal failure time follows a proportional hazard or an accelerated time model, the model can be rewritten like (2.3). This latter result can be easily specified. To fix the ideas, assume that each failure time T_j, $j = 1, \dots, p$ follows a Cox model. Then, we can write for every j and x

$$\ln \Lambda_{0,j}(T_j) = -x'\beta_j + \varepsilon_j,$$

where $\Lambda_{0,j}$ denotes the integrated baseline hazard function. The ε_j are some r.v. whose distributions are doubly exponential. Moreover, assume the joint distribution of $(\varepsilon_1, \dots, \varepsilon_p)$ is independent from x. Thus, setting

$$Z_j(t) = \Lambda_{0,j}(t), \ \ \phi_j(x) = \exp(x'\beta_j), \ \text{ and}$$

$$H(u_1, \dots, u_p) = P(\varepsilon_1 > \ln(-\ln u_1), \dots, \varepsilon_p > \ln(-\ln u_p)),$$

we find easily the formula (2.3). Otherwise, in an accelerated time model, each failure time T_j can be written $T_j = T_{0,j}\exp(x'\beta_j)$, for some failure time $T_{0,j}$. By assumption, the law of $(T_{0,1}, \dots, T_{0,p})$ is independent from x. Setting

$$Z_j(t) = t, \ \ \phi_j(x) = \exp(-x'\beta_j), \ \text{ and}$$

$$H(u_1, \dots, u_p) = P(T_{0,1} > \ln(-u_1), \dots, T_{0,p} > \ln(-u_p)),$$

we get again the formula (2.3).

Nonetheless, it is possible to exhibit some distributions that cannot be rewritten like (2.3). For instance, assume that T_1 and T_2 are two independent failure times and that $P(T_k > t|x) = \exp(-t^2 - tx)$, $k = 1, 2$. Deriving the survival function $S(t, 0|x)$ with respects to t and x, it is easy to state that (2.3) cannot hold.

Particularly, the model (2.3) is well suited for frailty models (Clayton and Cuzick [3]). For instance, consider two failure times that share a parameter of heterogeneity. They are independent conditioned on this parameter. When each marginal distribution follows a proportional hazards models, the bivariate survival function, knowing the covariates x, can be written

$$S(t_1, t_2|x) = \int \exp\left(-Z_1(t_1)\exp(x'\beta_1 + c_1\omega)\right) \exp\left(-Z_2(t_2)\exp(x'\beta_2 + c_2\omega)\right) \, dG(\omega),$$

where G is the d.f. of the unoberved heterogeneity. In the latter example,

$$H(u_1, u_2) = \int u_1^{\exp(c_1\omega)} u_2^{\exp(c_2\omega)} \, dG(\omega).$$

We remind the reader that the at hand sample consists of i.i.d. random vectors (Y_i, δ_i, X_i), $i = 1, \dots, n$. If the model can be rewritten like in (2.3), it is well-known that, under the following conditions, the functions Z_j, ϕ_j, $j = 1, \dots, p$ and H are identified (Heckman and Honoré [17]) :

(H1) H is differentiable with continuous and strictly positive partial derivatives H_j, $j = 1, \dots, p$, and the limit of $H_j(u)$, $j = 1, \dots, p$ when $u \to 1$ is finite and strictly positive.

(H2) $\phi_j(x_0) = 1$ for some fixed point x_0 and all $j = 1, \dots, p$.

(H3) the image set of $x \mapsto (\phi_1(x), \dots, \phi_p(x))$ is $]0, +\infty[^p$.

(H4) the Z_j, $j = 1, \dots, p$, are nonnegative, differentiable, strictly increasing functions.

(H5) $Z_1(1) = \dots = Z_p(1) = 1$.

It is not necessary to specify the marginal distributions of H to prove the previous result. Particularly, these margins does not need to be uniform on $[0, 1]$, viz H is not necessarily a copula function.

Some conditions can be relaxed. Reading carefully Heckman and Honoré's proof, observe that it is sufficient to assume that, for all $j = 1, 2$, $H_j(u_n)$ tends to a finite limit for some sequence of positive numbers $(u_n)_{n\geq 1}$, $u_n \to 1$. Moreover, in $(H3)$, it can be allowed that $\phi_j(x)$ is zero for some values of x, and then the d.f. of T_j conditioned on $X = x$ is degenerate. Moreover, the choice of point 1 in $(H5)$ is convenient and partly arbitrary. In fact, it is possible to suppose that $Z_1(t_1) = \dots = Z_p(t_p) = 1$ for some point $(t_1, \dots, t_p)$ in $]0, +\infty[^p$ instead of $(H5)$. In this latter case, each function ϕ_j is multiplied by an extra factor.

Actually, conditions $(H1) - (H5)$ has been weaken by various authors (Omori [24], Abbring and van den Berg [1], among others). Some of them are regularity assumptions that hold usually. The key condition is $(H3)$, or one of its variations. In practice, covariates are introduced by some univariate linear indices $\beta_j' x$, $j = 1, \dots, p$. Also, it is sufficient that there exist as many different continuous covariates as latent variables, viz at least p. Moreover, the parameters β_j (more precisely, the β_j's subvectors corresponding to the continuous covariates) have to be different, for every $j = 1, \dots, p$. Usually, this holds easily when the number of underlying risks is reasonable. Thus, the set described by $(\phi_1(x), \dots, \phi_p(x))$ is an open subset of $\mathbb{R}^p$. This is sufficient to our purpose. Indeed, Abbring and van den Berg [1] proved it is usually not necessary to describe all the space $\mathbb{R}^p$ to obtain the identifiability in such types of models.

3. Estimation of the function H

We seek to estimate all the previous unknown quantities without specific assumptions (other than regularity assumptions) on the joint distribution of the random vector $(T_1, \dots, T_p)$. We use some usual kernel methods, even if other nonparametric methods are surely relevant (nearest neighbors, local polynomials, wavelets....). In all the paper, we assume that (2.3) and $(H1) - (H5)$ are satisfied.

Before to introduce the nonparametric estimators, denote K, L, M, N some kernel functions, and h, h_0, h_1, h_2 some bandwidth sequences, viz some sequences of positive numbers that tend to zero when n tends to infinity. As usual, denote for each kernel function, say K, $K_h(x) = K(x/h)/h^r$ if $x \in \mathbb{R}^r$. For practical purpose, set $Z(t) = (Z_1(t), \dots, Z_p(t))$ and

$$\begin{aligned} \Phi : \mathbb{R}^d &\longrightarrow \mathbb{R}^p \\ x &\longmapsto (\exp(-\phi_1(x)), \dots, \exp(-\phi_p(x))). \end{aligned}$$

By assumption $(H3)$, Φ maps $\mathbb{R}^d$ into $]0, 1[^p$.

Moreover, suppose we know an estimator $\widehat{\Phi}$ of Φ for every compact subset C of $\mathbb{R}^d$ (eventually reduced to one point) such that a.e.

$$\sup_{x\in C} \|\widehat{\Phi}(x) - \Phi(x)\| = O(a_n). \tag{3.4}$$

Here, a_n denotes a sequence of positive numbers that tends to 0 when $n \to 0$. Assumption (3.4) is justified hereafter.

Note that $P(Y > t|X = x) \equiv Q(t|X = x) = S(t, \dots, t|X = x)$ and that, because of $(H5)$,

$$Q(1|X = x) = H \circ \Phi(x). \tag{3.5}$$

The key point is to approximate $H(y)$ by $E[Q(1|X = x)|\Phi(x) = y]$, from an i.i.d. sample $(Y_i, \delta_i, X_i)_{i=1,\dots,n}$.

Since the probability of the previous conditional event is zero if the density of $\Phi(X)$ is continuous, it is straightforward to consider some neighborhoods of y and some weights that decrease with the distance from y. Then we propose to approximate $H(y)$ by

$$\frac{\sum_{i=1}^n Q(1|X_i) K_h(y - \Phi(X_i))}{\sum_{i=1}^n K_h(y - \Phi(X_i))}. \tag{3.6}$$

Since we do not know the quantities $Q(1|X_i)$, they can be estimated by

$$\hat{Q}(1|X_i) = \frac{\sum_{j=1, j\neq i}^n \mathbf{1}\{Y_j > 1\} L_{h_0}(X_i - X_j)}{\sum_{j \neq i} L_{h_0}(X_i - X_j)}. \tag{3.7}$$

Hence $H(y)$ can be estimated by

$$\tilde{H}(y) = \frac{\sum_{i=1}^n \hat{Q}(1|X_i) K_h(y - \widehat{\Phi}(X_i))}{\sum_{i=1}^n K_h(y - \widehat{\Phi}(X_i))}. \tag{3.8}$$

Let C be some compact subset of $\mathbb{R}^d$. By continuity, $\Phi(C)$ is a compact subset of $\mathbb{R}^p$. We seek to prove the strong consistency of such an estimator of H uniformly on $\Phi(C)$. For technical reasons, we need to restrict ourselves to observations $(Y_i, \delta_i, X_i)_{i=1\dots,n}$ such that X_i lies into a compact subset of $\mathbb{R}^d$, say C_0. Then the previous estimator has to be calculated using this subsample. The data set is then truncated by the event $\{X \in C_0\}$. The reason of this loss of information is simple : to estimate $H(y)$ at some point y in $\Phi(C)$, we can use every point X_i such that $\Phi(X_i)$ is near from y, viz approximately such that $X_i \in \Phi^{-1}(\Phi(C))$. The latter set is generally strictly larger than C, and a priori not bounded. This causes a lack of control about the conditional probabilities of $\{Y > 1\}$ knowing $X = X_i$ when they are estimated by $\hat{Q}(1|X_i)$.

Theoretically, it could be sufficient to restrict ourselves to some points X_i that belong to a neighborhood of C if we seek to estimate $H(y)$, $y \in \Phi(C)$. Since Φ (Φ^{-1} in fact) is unknown and since this would restrict excessively the subsample that will be used in our formulas, we are rather to keep a larger compact set C_0. For practical and theoretical reasons, it would be convenient to impose that C is a subset of C_0, but it is not a necessity (nonetheless, it is important that $y \in \Phi(C_0)$ when $y \in \Phi(C)$). Adding some technical complications, it would be even possible to choose $C_0 = C_{0,n}$, building a sequence of compact subsets $(C_{0,n})_n$ that is increasing towards $\mathbb{R}^d$. This way is left to the reader. In practice, a rough rule should eliminate the outliers of the sample set.

The truncation effect needs to be corrected to obtain an asymptotically unbiased estimator of H. That is why we consider from now on

$$\hat{H}(y) = \frac{1}{\hat{P}(C_0, y)} \hat{H}_0(y), \tag{3.9}$$

$$\hat{H}_0(y) = \frac{\sum_{i=1}^n \mathbf{1}\{X_i \in C_0\} \hat{Q}(1|X_i) K_h(y - \widehat{\Phi}(X_i))}{\sum_{i=1}^n K_h(y - \widehat{\Phi}(X_i))}, \tag{3.10}$$

$$\hat{P}(C_0, y) = \frac{\sum_{i=1}^n \mathbf{1}\{X_i \in C_0\} K_h(y - \widehat{\Phi}(X_i))}{\sum_{i=1}^n K_h(y - \widehat{\Phi}(X_i))}. \tag{3.11}$$

Indeed, under some forthcoming conditions, $\hat{H}_0(y)$ tends to

$$E\left[\mathbf{1}\{X \in C_0\}Q(1|X)|\Phi(X) = y\right] = H(y)P(X \in C_0|\Phi(X) = y) \equiv H(y)P(C_0, y).$$

For each random variable Z, denote f_Z the density of Z with respects to the Lebesgue measure. For each subset $A \subset \mathbb{R}^d$, denote $\tilde{A}$ a compact ε-neighborhood of A for some $\varepsilon > 0$ (e.g. $\tilde{A} = \{x|d(x, A) \leq \varepsilon\}$). As usual, denote C a compact subset of $\mathbb{R}^d$.

The technical assumptions we need can be summarized. First, the considered kernel functions need to be sufficiently smooth.

Assumption K.1 : *L is a d-dimensional bounded kernel of order k, $\|x\|^d L(x)$ tends to 0 when $\|x\|$ tends to ∞, and L is γ-Lipschitz continuous. Similarly, K is a p-dimensional bounded kernel of order k', $\|x\|^p K(x) \to 0$ when $\|x\|$ tends to ∞ and K is γ'-Lipschitz continuous.*

Second, the underlying density and distribution functions have to be sufficiently regular.

Assumption R.1 : *$H \circ \Phi$ and f_X are k-times differentiable on $\mathbb{R}^d$, $k \in \mathbb{N}$, and their derivatives of order k are γ-Lipschitz continuous on $\mathbb{R}^d$ for some $\gamma \in]0,1[$. H, $P(C_0, \cdot)$ and $f_{\Phi(X)}$ are k'-times differentiable on $\mathbb{R}^p$, $k' \in \mathbb{N}$, and their derivatives of order k' are γ'-Lipschitz continuous on $\mathbb{R}^p$ for some $\gamma' \in]0,1[$.*

Third, as usual in kernel regression, the denominators have to be bounded from above on the considered subsets.

Assumption B.1 : $\inf_{x\in\tilde{C}_0} f_X(x) > 0$, $\inf_{y\in\tilde{\Phi}(C)} f_{\Phi(X)}(y) > 0$ *and* $\inf_{y\in\tilde{\Phi}(C)} P(C_0, y) > 0$.

Theorem 1. *If (3.4),* K.1, R.1 *and* B.1 *are satisfied, if $nh^p/\ln n \underset{n\to\infty}{\longrightarrow} \infty$ and $a_n/h^{1+p} \underset{n\to\infty}{\longrightarrow} 0$, then a.e.*

$$\sup_{y\in\Phi(C)} |\hat{H}(y) - H(y)| = O\left(h^{-p^+}h_0^{k+\gamma} + h^{-p^+}\left(\frac{\ln n}{nh_0^d}\right)^{1/2} + h^{k'+\gamma'} + \left(\frac{\ln n}{nh^p}\right)^{1/2} + \frac{a_n}{h^{1+p}}\right),$$

where $p^+ = 0$ if $K \geq 0$, and $p^+ = p$ else.

See the proof in the appendix. It is possible to maximize the previous upper bound with respects to h_0 and h. It is particularly easy when $p^+ = 0$. The following sections are dealing with the estimation of the other unknown quantities, viz the functions ϕ_j and Z_j, $j = 1, \ldots, p$ respectively.

4. Estimation of the functions ϕ_j

We need to precise the rate of convergence a_n appearing in (3.4). Before dealing with the non-parametric estimation of Φ, it is useful to discuss two particularly frequent regression models in the competing risks framework in details.

First, some authors impose that the "crude" or "specific" hazard functions (1.1) satisfy some proportional hazards model (see e.g. Prentice and Kalbfleisch [27], Cox and Oakes [4]). A commonly used proportional hazards assumption specifies that, for every (t, x) and $j = 1, \ldots, p$,

$$\tilde{\lambda}_j(t|X = x) = \tilde{\lambda}_{0,j}(t)\psi_j(x'\beta_j), \tag{4.12}$$

where β_j is the parameter associated with the j-th failure time, $\tilde{\lambda}_{0,j}$ is an unknown baseline hazard function and ψ_j is a known strictly positive function. The most commonly used assumption is the Cox model, for which ψ_j is the exponential function. Under the previous assumptions $(H1)$-$(H5)$ and (4.12), it can be proved that the p durations T_j are mutually independent.

Lemma 1. *Under $(H1)$-$(H5)$, if a competing risks model (2.3) satisfies (4.12) (in other words if each "crude" hazard function follows a proportional hazard model) then each component T_j, $j = 1, \ldots, p$, is independent from the others.*

See the proof of lemma 1. in the appendix. Deduce that $\tilde{\lambda}_j(\cdot|X = x)$ is in fact the true hazard function of T_j conditional on $X = x$ and that $\phi_j(x)$ is equal to $\psi_j(x'\beta_j)\phi_j(x_0)/\psi_j(x_0'\beta_j)$.

Then, in this case, we can deal with each component T_j independently from the others. More precisely, each failure time T_j is independently right-censored by the other failure times (see Kalbfleisch and Prentice [27] e.g.). Then, the usual Cox's partial likelihood provides a strongly consistent and asymptotically normal estimator of each parameter β_j. Therefore, $\psi_j(x'\hat{\beta}_j)\phi_j(x_0)/\psi_j(x_0'\hat{\beta}_j)$ provides a consistent and asymptotically normal estimator of $\phi_j(x)$. Since x is contained in a bounded set C, every sequences $(a_n)_{n\geq 1}$ such that $1 >> a_n >> n^{-1/2}$ are suitable.

Remark **1.** *It should be possible to estimate nonparametrically the functions ψ_j if they are unknown : see O'Sullivan [28] or Fan et al. [7].*

Second, it could be attractive to specify some proportional hazards assumptions no more on the crude but on the true hazard functions. More precisely, suppose that each component of T follows a proportional hazards model. Then, for each $j = 1, \ldots, p$, the hazard function $\lambda_j(\cdot|X = x)$ of T_j conditional on $X = x$ satisfies

$$\lambda_j(t|X = x) = \lambda_{0,j}(t)\psi_j(x'\beta_j), \tag{4.13}$$

where ψ_j is known and $\lambda_{0,j}$ is the baseline hazard function of T_j. Clearly

$$\begin{aligned}\lambda_j(t_j|x) &= -\frac{\partial_j S(0,\ldots,0,t_j,0,\ldots|x)}{S(0,\ldots,0,t_j,0,\ldots|x)} \\ &= \frac{\partial_j H[1,\ldots,1,\exp(-Z_j(t)\phi_j(x)),1,\ldots,1]}{H[1,\ldots,1,\exp(-Z_j(t)\phi_j(x)),1,\ldots,1]}\cdot\exp(-Z_j(t)\phi_j(x))\phi_j(x)Z_j'(t).\end{aligned}$$

Using the proportional hazards specification (4.13), the same reasoning like in lemma 1. can be lead. Deduce that $\phi_j(x)$ is proportional to $\psi_j(x'\beta_j)$ and is fully specified due to $(H2)$. Thus, we know ϕ_j when we know β_j. Moreover, there exist some nonnegative constants c_j such that for all $u_j \in]0,1]$,

$$H_j[1,\ldots,1,u_j,1,\ldots,1] = u_j^{c_j}. \tag{4.14}$$

Note that the latter relation is satisfied when the components of T are mutually independent, but the converse is false. Indeed, in a model without covariates, consider the bivariate exponential distribution

$$S(t_1,t_2) = \exp(-\lambda_1 t_1 - \lambda_2 t_2 - \nu t_1 t_2), \lambda_1 > 0, \lambda_2 > 0, \nu > 0,$$

which provides

$$H(u_1,u_2) = u_1^{\lambda_1}u_2^{\lambda_2}\exp(-\nu\ln u_1 \ln u_2).$$

This bivariate distribution satisfy (4.14) even if the two marginal failure times are not independent. To our knowledge, there are no simple ways to estimate the parameters $(\beta_1,\ldots,\beta_p)$ in this case without parametric assumptions on S or H.

Remark **2.** *In the independent case, the crude hazard function $\tilde{\lambda}_j(\cdot|X = x)$ is equal to $\lambda_j(\cdot|X = x)$. Then, equations (4.12) and (4.13) holds simultaneously, but the latter is a consequence of the former and assumptions $(H1)$-$(H5)$.*

Remark **3.** *Note that, under (4.13), $E[T_j|X=x]$ is a function of $x'\beta_j$ only. Then each T_j belongs to the class of single-index models for which some estimators of β_j have been proposed (Han [14], Härdle and Stoker [16], Horowitz and Härdle [18], Ichimura [19], Powell et al. [26], Sherman [29]...). Nonetheless, we do not observe a sample from T_j's trials knowing $X=x$ but only such a sample that is right-censored by the other failure times. The distribution of this sample depends on other indices $x'\beta_k, k\neq j$. To connect this feature with the single-index literature seems to be messy and to be an open problem.*

Without specific assumption about the functional form of Φ, it is possible to estimate the functions ϕ_j, $j=1,\ldots,p$. Indeed, it is the case through a multidimensional nonparametric estimation of each ϕ_j. As usual, fix an index $j\in\{1,\ldots,p\}$. Consider

$$\hat{R}_j(t|x)=\frac{\sum_{i=1}^n \mathbf{1}\{\delta_i=j\}M_{h_1}(t-Y_i)N_{h_2}(x-X_i)}{\sum_{i=1}^n N_{h_2}(x-X_i)}, \tag{4.15}$$

where M, N denote some kernel functions as usual, and h_1, h_2 denote some bandwidth sequences.

We have approximately

$$\begin{aligned}
\hat{R}_j(t|x) &\simeq \frac{E[\mathbf{1}\{T_j<T_k,\forall k\neq j\}M_{h_1}(t-T_j)N_{h_2}(x-X)]}{f_X(x)}\\
&\simeq \frac{(-1)^p}{f_X(x)}\int \mathbf{1}\{v_j<v_k,\forall k\neq j\}M_{h_1}(t-v_j)N_{h_2}(x-u)f_X(u)\,S(dv|X=u)\,du\\
&\simeq \frac{(-1)}{f_X(x)}\int M_{h_1}(t-v_j)N_{h_2}(x-u)f_X(u)\,\partial_j S(v_j,\cdots,v_j|X=u)\,du\,dv_j\\
&\simeq -\int M_{h_1}(t-v_j)\partial_j S(v_j,\cdots,v_j|X=x)\,dv_j\\
&\simeq -\partial_j S(t,\cdots,t|X=x),
\end{aligned} \tag{4.16}$$

when f_X is continuous at x and $\partial_j S(t,\cdots,t|x)$ is continuous at t. The latter continuity assumptions can be weakened, because some Bochner's lemma is still available when the density function is right continuous with left-hand limits (see Fermanian [8]). Hereafter, the term $\hat{R}_j(0|X=x)$ only will be considered. Thus, let $\lim_{t\to 0,t>0}\partial_j S(t,\ldots,t|X=x)\equiv l_0^+$. If $l_0^+\neq 0$, then the right-hand side of (4.16) becomes $-l_0^+\int_{u<0}M(u)\,du$ at point $t=0$. In fact, as we will deal with fractions, the multiplicative extra factor will disappear in all of our results and it will be not necessary to modify our estimators.

Since $\lim_{t\to 0,t>0}\partial_j S(t,\cdots,t|X=x)=-H_j[1,\ldots,1]\phi_j(x)Z_j'(0^+)$, define an estimator of $\phi_j(x)$ by

$$\hat{\phi}_j^{(0)}(x)=\frac{\hat{R}_j(0,x)}{\hat{R}_j(0,x_0)}, \tag{4.17}$$

for all $j=1,\ldots,p$. The properties of $\hat{\phi}_j^{(0)}(x)$ should be relatively easy to state, because this estimator is similar to usual density and regression kernel estimators.

In fact, the estimator (4.17) is sufficient for our purpose if $\partial_j S(t,\ldots,t|x)$ is nonzero at the origin on the right. If it is not the case, some undesirable problems of unboundedness appear. That is why we propose to slightly modify (4.17) if $\partial_j S(t,\ldots,t|x)$ tends to zero when $t\to 0^+$.

To do this, consider the previous kernel estimators at some points α_n where $(\alpha_n)_{n\geq 1}$ is a sequence of positive numbers which tends to 0 "not too quickly" when $n\to\infty$ (see assumption Z.0). The sequence $(\alpha_n)_{n\geq 1}$ depends on the index j, but we omit it for simplicity. Assume there exists such a sequence. Then, we can define an alternative estimator of ϕ_j by replacing 0 by α_n in definition (4.17), providing

$$\hat{\phi}_j^{(1)}(x) = \frac{\hat{R}_j(\alpha_n|x)}{\hat{R}_j(\alpha_n|x_0)}\cdot \tag{4.18}$$

To state the strong uniform consistency of the previous estimators of ϕ_j, we need to impose some precise regularity conditions. Particularly, the considered kernels have to be positive and compactly supported, e.g. Epanechnikov kernels.

ASSUMPTION K.2 : M and N are compactly supported 2-order kernels, $M(t) = m(|t|)$ and $N(z) = n(\|z\|)$, where m and n are some decreasing functions on $[0, +\infty[$.

ASSUMPTION R.2 : H belongs to $C^3(]0,1]^p)$, ϕ belongs to $C^2(\tilde{C})$ and Z belongs to $C^3(]0,\eta[)$, for some $\eta > 0$, $f_{(Y,X)}$ is bounded, f_X is 2-times continuously differentiable on $\tilde{C}$ and $\inf_{x\in\tilde{C}} f_X(x) > 0$.

For each positive function f of t, we have denoted $f^*(t) = \sup_{\{u,|u-t|\leq h_1\}} f(u)$. Similarly, for each positive function g of x, we will denote $g^*(x) = \sup_{\{v,\|v-x\|\leq h_2\}} g(v)$. When the upper bound is taken with respects to u and v simultaneously, we will use the same notation. Moreover, for each real number t, denote

$$\mathcal{Z}_j(t) \equiv |Z_j'''|^*(t) + \|Z'\|^*(t)|Z_j''|^*(t) + |Z_j'|^*(t)\left(\|Z'\|^2\right)^*(t) + |Z_j'|^*(t)\|Z''\|^*(t). \tag{4.19}$$

ASSUMPTION Z.0 : there exists a sequence of positive real numbers $(\alpha_n)_{n\geq 1}$ such that $\alpha_n \to 0$, $h_1 << \alpha_n$, and

$$\left(\frac{\ln^2 n}{nh_1h_2^d}\right)^{1/2} + h_1^2\mathcal{Z}(\alpha_n) << Z_j'(\alpha_n).$$

THEOREM 2. *Under* K.2 *and* R.2*, if* $x_0 \in C$ *and if* $\lim_{t\to 0,t>0} Z_j'(t) > 0$*, then a.e.*

$$\sup_{x\in C} |\hat{\phi}_j^{(0)}(x) - \phi_j(x)| = O\left(h_1 + h_2^2 + \left(\frac{\ln^2 n}{nh_1h_2^d}\right)^{1/2}\right) \equiv O(a_n). \tag{4.20}$$

Moreover, assume $\lim_{t\to 0,t>0} Z_j'(t) = 0$ *and* Z.0*. Then a.e.*

$$\sup_{x\in C} |\hat{\phi}_j^{(1)}(x) - \phi_j(x)| =$$

$$O\left(\|Z\|(\alpha_n) + h_2^2 + \frac{h_1^2\mathcal{Z}_j(\alpha_n)}{Z_j'(\alpha_n)} + \frac{1}{Z_j'(\alpha_n)}\left(\frac{\ln^2 n}{nh_1h_2^d}\right)^{1/2}\right) \equiv O(a_n).$$

With a suitable choice of (α_n), the previous estimator of ϕ_j is strongly uniformly consistent, and a_n tends to 0 when $n \to \infty$.

Theoretically, it is always possible to build a convenient sequence $(\alpha_n)_{n\geq 1}$. The problem is that we do not know Z and its derivatives in a neighborhood of zero. Hence, the convergence to zero of the latter sequence (a_n) is not ensured. Nonetheless, under some reasonable conditions of regularity on Z_j', it is easy to find such a convenient sequence $(\alpha_n)_{n\geq 1}$. Suppose there exists an integer m such that Z_j is $m+1$-times continuously differentiable on $]0,\eta]$, $\eta > 0$ and at 0 on the right. Moreover, suppose that $Z_j(0^+) = Z_j'(0^+) = \cdots = Z_j^{(m)}(0^+) = 0$ and $Z_j^{(m+1)}(0^+) \neq 0$. Then, as n tends to infinity, we have the equivalences

$$Z_j(\alpha_n) \sim \frac{\alpha_n^{m+1}}{(m+1)!} Z_j^{(m+1)}(0^+), \;\; Z_j'(\alpha_n) \sim \frac{\alpha_n^m}{m!} Z_j^{(m+1)}(0^+).$$

Hence, after specifying the order of $\mathcal{Z}_j(\alpha_n)$, it is possible to optimize an upper bound of the uniform rate of strong convergence of $\hat{\phi}_j^{(1)}$, say (4.21), with respects to α_n. For instance, when $m = 1$ or 2 and when $\|Z\|(\alpha_n)$ is of order α_n^q with some $q \geq 1$, set

$$\alpha_n = Cst. \left[h_1^2 + \left(\frac{\ln^2 n}{n h_1 h_2^d} \right)^{1/2} \right]^{1/(q+2)}.$$

Hence, in the latter case,

$$\sup_{x \in C} |\hat{\phi}_j^{(1)}(x) - \phi_j(x)| = O\left(h_2^2 + \left[h_1^2 + \left(\frac{\ln^2 n}{n h_1 h_2^d} \right)^{1/2} \right]^{q/(q+2)} \right).$$

Here, it is easy to optimize the latter bound with respects to h_1 and h_2.

5. Estimation of the functions Z_j

To solve completely our competing risks problem, it remains to exhibit some estimators of Z_j, $j = 1, \ldots, p$. We remind that the data set at hand provides the knowledge of the d.f. of Y, viz the function

$$Q(t|x) = S(t, \ldots, t|x) = H[\exp(-Z_1(t)\phi_1(x)), \ldots, \exp(-Z_p(t)\phi_p(x))].$$

The previously studied estimators of H and ϕ_j, $j = 1, \ldots, p$ can now be used. Fix an index j, say $j = 1$. We seek to estimate Z_1. It seems to be relevant to deal mainly with the first component of H. To cancel the influence of $Z_k, k > 1$, the so-called condition $(H3)$ is necessary. It allows us to use mainly observations i such that $\phi_k(X_i)$ is near from zero for some k, and to replace the variability of Q with respects to t by its variability with respects to x.

Set $S_1(z|x) \equiv H\left(\exp(-z\phi_1(x)), 1, \ldots, 1\right)$. It could be estimated by

$$\hat{S}_1(z|x) = \hat{H}(\exp(-z\hat{\phi}_1(x)), 1, \ldots, 1),$$

where $\hat{\phi}_1$ is some estimator of ϕ_1, e.g. $\hat{\phi}_1^{(k)}$, $k = 0, 1$. Note that, for these x such that $\phi_k(x) \simeq 0$, $k > 1$, we have

$$S_1(Z_1(t)|x) \simeq Q(t|x), \text{ or } Z_1(t) \simeq S_1^{-1}[Q(t|x)|x].$$

Denote $\bar{K}$ and $\bar{L}$ some kernel functions of dimensions 1 and $p-1$ respectively, that are nonzero at the origin. Moreover, denote $\bar{h}$ and $\bar{h}_0$ some bandwidth sequences. When $\bar{h} \to 0$, we obtain approximately

$$\begin{aligned} S_1^{-1}(t|x) &\simeq \int S_1^{-1}(v|x)\bar{K}_{\bar{h}}(v-t)\,dv \\ &\simeq \int u\bar{K}_{\bar{h}}[S_1(u|x) - t]\frac{\partial S_1}{\partial u}(u|x)\,du, \end{aligned}$$

because $S_1(\cdot|x)$ is monotone.

Weighting the observations with respects to the distance from $(\phi_2(X_i), \ldots, \phi_p(X_i))$ to zero provides

$$\begin{aligned} Z_1(t) &\simeq \sum_{i=1}^n \frac{\bar{L}_{\bar{h}_0}(\phi_2(X_i), \ldots, \phi_p(X_i))}{\sum_{i=1}^n \bar{L}_{\bar{h}_0}(\phi_2(X_i), \ldots, \phi_p(X_i))} \int u\bar{K}_{\bar{h}}[S_1(u|X_i) - Q(t|X_i)]\frac{\partial S_1}{\partial u}(u|X_i)\,du \\ &\simeq \sum_{i=1}^n \frac{\bar{L}_{\bar{h}_0}(\hat{\phi}_2(X_i), \ldots, \hat{\phi}_p(X_i))}{\sum_{i=1}^n \bar{L}_{\bar{h}_0}(\hat{\phi}_2(X_i), \ldots, \hat{\phi}_p(X_i))} \int u\bar{K}_{\bar{h}}[\hat{S}_1(u|X_i) - \hat{Q}(t|X_i)]\frac{\partial \hat{S}_1}{\partial u}(u|X_i)\,du \\ &\equiv \hat{Z}_1^{(1)}(t), \end{aligned}$$

where $\hat{\phi}_k$ denotes a consistent estimator of ϕ_k and where, extending (3.7),

$$\hat{Q}(t|X_i) = \frac{\sum_{j=1,j\neq i}^n \mathbf{1}\{Y_j > t\} L_{h_0}(X_i - X_j)}{\sum_{j=1,j\neq i}^n L_{h_0}(X_i - X_j)}.$$

Similarly, define some estimators $\hat{Z}_j^{(1)}$ of Z_j, $j > 1$. The statistical properties of these complicated estimators has not been done in this paper, because $\hat{Z}_j^{(1)}$ can not be rewritten easily like a sufficiently regular functional of the other relatively simpler quantities $\hat{H}$ and $\hat{\phi}_j$. This is partly due to the necessity of inverting S_1 or an estimator of S_1. That is why we propose now another analytically simpler solution to estimate the functions Z_j, $j = 1, \ldots, p$.

Let a covariable vector x such that $\phi_k(x) \simeq 0$, for all $k = 1, \ldots, p$. Such vectors exist through to assumption $(H3)$. Then, for each t and $j = 1, \ldots, p$,

$$\begin{aligned} P(Y > t, \delta = j|X = x) &= -\int \mathbf{1}\{u > t\} \partial_j S(\cdot|x)_{|t_1 = \cdots = t_p = u} \, du \\ &= \int \mathbf{1}\{u > t\} \partial_j H[\exp(-Z_1(u)\phi_1(x)), \ldots, \exp(-Z_p(u)\phi_p(x))] \\ &\cdot \quad \exp(-Z_j(u)\phi_j(x))\phi_j(x) Z_j'(u) \, du \\ &\simeq \int \mathbf{1}\{u > t\} \partial_j H[1, \ldots, 1] \exp(-Z_j(u)\phi_j(x))\phi_j(x) Z_j'(u) \, du \\ &\simeq \exp(-Z_j(t)\phi_j(x)) \partial_j H[1, \ldots, 1]. \end{aligned}$$

Normalizing with the value at $t = 1$, deduce for such x,

$$Z_j(t) \simeq 1 + \frac{1}{\phi_j(x)} \ln \left[\frac{P(Y > 1, \delta = j|X = x)}{P(Y > t, \delta = j|X = x)} \right]. \tag{5.22}$$

Since simple estimators of the unknown quantities of the latter expression exist, we obtain the relatively simple estimator of Z_j at point t

$$\hat{Z}_j^{(2)}(t) = 1 + \sum_{i=1}^n R_n(\hat{\phi}_1(X_i), \ldots, \hat{\phi}_p(X_i)) \frac{1}{\hat{\phi}_j(X_i) + e_n} \ln \left[\frac{P_1(X_i)}{P_t(X_i)} \right], \tag{5.23}$$

$$R_n(\hat{\phi}_1(X_i), \ldots, \hat{\phi}_p(X_i) \equiv \frac{n^{-1} \bar{M}_l(\hat{\phi}_1(X_i), \ldots, \hat{\phi}_p(X_i))}{n^{-1} \sum_{k=1}^n \bar{M}_l(\hat{\phi}_1(X_k), \ldots, \hat{\phi}_p(X_k)) + b_n}, \tag{5.24}$$

where $\bar{M}$ is an even compactly supported kernel functions in $\mathbb{R}^p$, $l = l_n$ is a bandwidth sequence and for all x and t,

$$P_t(x) = \frac{1}{n} \sum_{k=1}^n \mathbf{1}\{Y_k > t, \delta_k = j\} L_{h_0}(x - X_k) + d_n.$$

For technical reasons, we have introduced some sequences of nonnegative real numbers (b_n), (d_n) and (e_n), which tend to 0 when $n \to \infty$. Note that the previous ratio is approximated by $\hat{Q}_j(1|x)/\hat{Q}_j(t|x)$, where

$$\hat{Q}_j(t|x) \equiv \frac{n^{-1} \sum_{k=1}^n \mathbf{1}\{Y_k > t, \delta_k = j\} L_{h_0}(x - X_k)}{n^{-1} \sum_{k=1}^n L_{h_0}(x - X_k)} \simeq P(Y > t, \delta = j|X = x).$$

Moreover, we need to use a slightly modified estimator of each function ϕ_j by adding the sequence of constants ρ_n to the denominator of $\hat{R}_j(t|x)$ (see Eq. (7.31)).

The new estimator $\hat{Z}_j^{(2)}$ is simpler but surely less efficient than $\hat{Z}_j^{(1)}(t)$. Indeed it uses the observations X_k s.t. $\phi(X_k)$ belongs to a neighborhood of the origin in $\mathbb{R}^p$ only, when they belong in a neighborhood of 0 in $\mathbb{R}^{p-1}$ for $\hat{Z}_j^{(1)}$. Moreover, it seems to be necessary to introduce the positive real sequences (b_n), (d_n) and (e_n) because of the lack of control on the quantities $\hat{\phi}_j(X_i)$ and on the location of the X_i.

The difficulty to prove the weak consistency of $\hat{Z}^{(1)}$ or $\hat{Z}^{(2)}$ can be easily understood. We consider some points X_i in the sample s.t. $\phi_k(X_i) = o(1)$ for some indices k. Thus, these points cannot belong to some fixed compact subset of $\mathbb{R}^d$. Since we want to control the distance between $\hat{\phi}(X_i)$ and $\phi(X_i)$ for these X_i, this implies strong regularity conditions on the tails of these distributions. Nonetheless, we provide a theoretical positive result, whose proof can be found in the appendix. The tedious conditions Q.0-Q.6 have not been rewritten in this section. They can be found in the appendix. Nonetheless, we precise below the technical assumptions on the kernels and on the underlying distribution functions. They are commonly used in such problems.

Assumption K.3 : L and $\bar{M}$ are compactly supported 2-order positive kernels of dimensions d and p respectively. $L(x) = l(\|x\|)$ and $\bar{M}(\phi) = \bar{m}(\|\phi\|)$, where l and $\bar{m}$ are some decreasing functions on $[0, +\infty[$. Moreover, $\bar{m}$ is continuously differentiable.

Assumption R.3 : f_X and ϕ belong to $C^2(\mathbb{R}^d)$, $d^k\phi$ and $d^k f_X$, $k = 0, 1, 2$ are bounded, and ϕ belongs to $C^2(\mathbb{R}^d)$.

Theorem 3. *Assume the assumptions K.3, R.3 and Q.0-Q.6. Then for each* $j = 1, \ldots, p$ *and* $T \subset \mathbb{R}^+$ *compact,* $\sup_{t \in T} |\hat{Z}_j(t) - Z_j(t)|$ *tends to* 0 *in probability when* n *tends to infinity.*

6. A short simulation study

To assess the performances of our estimators, we choose a particular case of the model proposed by Clayton and Cuzik [3], viz

$$S(t_1, t_2|x) = [\exp(\,\gamma Z_1(t_1)\phi_1(x)\,) + \exp(\,\gamma Z_2(t_2)\phi_2(x)\,) - 1]^{-1/\gamma}.$$

Here, we assume simply that $\gamma = 1$, $p = d = 2$, $Z_1(t) = Z_2(t) = t$, $x_0 = (0, 0)$, and that

$$\phi_1(x) = \exp(x_1 + x_2), \quad \phi_2(x) = \exp(x_1 - x_2).$$

Thus, $H(y_1, y_2) = [y_1^{-1} + y_2^{-1} - 1]^{-1}$ when y_1 and y_2 belong to $]0, 1]$. The distribution of X is a standard bivariate normal $\mathcal{N}(0, Id)$.

We simulate 500 samples of size 1000. For convenience and for computational purpose, all the bandwidths are selected following Silverman's rule [30] for density estimation, in the univariate and multivariate cases. For convenience too, we use some product of normal kernels, even if they need to be compactly supported theoretically. When we choose Epanechnikov kernels, our results have changed very slightly. Roughly, we set $b_n = d_n = e_n = \rho_n = n^{-1}$, so that these terms are negligible with respect to kernel estimates (at least a.e.). The empirical standard deviations are denoted by the letter "σ".

Table A provides the results for the estimations of the functions ϕ_1 and ϕ_2. They are not as good as we could hope. Larger are the values of ϕ, lower is the quality of the results. This is not a surprise : the computation of $\hat{\phi}$ is based on observations whose dates Y_i are very near 0. And when ϕ_1 or ϕ_2 is (relatively) large, such points become very sparse.

Let us now turn to the estimation of H, Z_1 and Z_2. In one case, we assume that the functions ϕ_1 and ϕ_2 were known. This provides some estimates denoted by the subscript $\tilde{}$ (e.g. $\tilde{Z}_1$ or $\tilde{H}$).

In the other case, we compute the estimators described in the paper. They are denoted by their subscript $\hat{}$. Dealing with $\hat{Z}^{(2)} = (\hat{Z}_1^{(2)}, \hat{Z}_2^{(2)})$, we get table B. The estimations of H are put in table C.

Obviously, the standard deviations obtained using estimated functions ϕ are larger than in the first case. Nonetheless, the bias are not always worse, especially when $Z(t) = t$ is not "too large" (less than 3). Globally, the results are not so bad, especially considering all the steps and relatively arbitrary choices that have been made to compute then. A fully nonparametric point of view should be surely difficult to practice. Since, in a lot of case, some of the functions ϕ, Z or H are known or even belong to some known parametric family, the task is not always so hard. Especially when ϕ is assumed to be known, this simulation study seems to provide reasonable results, particularly in the "center" of the distributions and not near the frontiers of the domains of definition.

7. Conclusion and Comments

The quantity of interest was $S(\cdot|x)$ initially. Now, it is possible to estimate this one by using the previous statistics, viz for every p-uples $(t_1, \ldots, t_p)$ and x, set

$$\hat{S}(t_1, \ldots, t_p|x) = \hat{H}\left(\exp(-\hat{Z}_1(t_1)\hat{\phi}_1(x)), \ldots, \exp(-\hat{Z}_p(t_p)\hat{\phi}_p(x))\right),$$

where the $\hat{\phi}_j$ (respectively $\hat{Z}_j$) are, for each j, one of the previous estimators of ϕ_j (resp. Z_j).

It would be better to exhibit directly an estimator of $S(\cdot|x)$ without estimating first other unknown functions, but we have not succeeded in finding a simple direct estimator of S. That is why we have adopted the parametric approach proposed by Heckman and Honoré [17], which allows to separate formally the influence of time and the influence of covariates.

The hardest task was to estimate nonparametrically the functions Z_j. Probably, some competitive estimators of Z could be found too. Particularly, the use of other functional techniques like wavelets, local polynomials or more generally sieves could be fruitful. Indeed, these techniques allow a better control on the tail behaviors and on the rates of uniform convergence, for the considered regression functions. Nevertheless, they are not as simple as the kernel method.

It remains to do a lot of work. As usual in kernel estimation, the results depend widely on the choice of the bandwidth sequences. Some theoretical results about "optimal" bandwidth choices would be valuable. Moreover, the variance of all the previous estimators should be estimated. It would be surely impossible to find exact formulas, and asymptotic expansions would be even cumbersome. Some numerically approximated variances could be provided by bootstrap techniques. Finally, we do not have discussed the asymptotic normality nor the exact rates of convergence, due to technical complexity.

APPENDIX

Proof of theorem 1.

Let us remind a preliminary theorem, that will be used hereafter. Let $(X_i, Y_i)_{i=1,\ldots,n}$ be an i.i.d. sample of a r.v. $(X, Y) \in \mathbb{R}^d \times \mathbb{R}$. Denote $r(x) = E[Y|X = x]$. Let the standard kernel regression estimator of Y on $X = x$ be

$$r_n(x) = \frac{\sum_{i=1}^n Y_i K_h(x - X_i)}{\sum_{i=1}^n K_h(x - X_i)}.$$

Applying Györfi et al.'s theorem 3.3.2 ([13]) with independent observations and bounded variables, we get for every compact subset A of $\mathbb{R}^d$:

Lemma 2. *If*

i. r *and* f_X *are* k*-times differentiable on* $\mathbb{R}^d$, $k \in \mathbb{N}$, *and their derivatives of order* k *are* γ*-Lipschitz continuous on* $\mathbb{R}^d$ *for some* $\gamma \in]0,1[$,

ii. $\inf_{x\in\tilde{A}} f_X(x) > 0$,

iii. K *is a bounded* d*-dimensional kernel of order* k, $\|x\|^d K(x) \to 0$ *when* $\|x\|$ *tends to* ∞ *and* K *is* γ*-Lipschitz continuous, ,*

then a.e.

$$\sup_{x\in A} |r_n(x) - r(x)| = O\left(h^{k+\gamma} + \left(\frac{\ln n}{nh^d}\right)^{1/2} \right).$$

Therefore, according to the previous lemma,

Lemma 3. *If*

i. $H \circ \Phi$ *and* f_X *are* k*-times differentiable on* $\mathbb{R}^d$, $k \in \mathbb{N}$, *and their derivatives of order* k *are* γ*-Lipschitz continuous on* $\mathbb{R}^d$ *for some* $\gamma \in]0,1[$,

ii. $\inf_{x\in\tilde{C}_0} f_X(x) > 0$,

iii. L *is a bounded* d*-dimensional kernel of order* k, $\|x\|^d L(x) \to 0$ *when* $\|x\|$ *tends to* ∞ *and* L *is* γ*-Lipschitz continuous,*

then a.e.

$$\sup_{X_i\in C_0} |\hat{Q}(1|X_i) - Q(1|X_i)| = O\left(h_0^{k+\gamma} + \left(\frac{\ln n}{nh_0^d}\right)^{1/2} \right). \tag{7.25}$$

Now let us prove theorem 1. To simplify the notations, denote $K_i = K_h(y - \Phi(X_i))$, $\hat{K}_i = K_h(y - \widehat{\Phi}(X_i))$, $\bar{K}_i = K_h(y - \Phi(X_i))\mathbf{1}\{X_i \in C_0\}$, $\tilde{K}_i = K_h(y - \widehat{\Phi}(X_i))\mathbf{1}\{X_i \in C_0\}$, $Q_i = \mathbf{1}\{X_i \in C_0\}Q(1|X_i)$ and $\hat{Q}_i = \mathbf{1}\{X_i \in C_0\}\hat{Q}(1|X_i)$. Note that $\sum_i \bar{K}_i / \sum_i K_i$ is the usual kernel estimator of $P(C_0, y)$. Then $\hat{H}(y)$ can be split into several terms. More precisely

$$\begin{aligned}
\hat{H}(y) &= \frac{\sum_{i=1}^n \hat{Q}_i\hat{K}_i}{\sum_{i=1}^n \tilde{K}_i} \\
&= \frac{\sum_i K_i}{\sum_i \bar{K}_i} \cdot \frac{\sum_i \hat{Q}_i\hat{K}_i}{\sum_i K_i} + \frac{\sum_i \hat{Q}_i\hat{K}_i}{\sum_i \tilde{K}_i} \cdot \frac{\sum_i(\bar{K}_i - \tilde{K}_i)}{\sum_i \bar{K}_i} \\
&= \frac{1}{P(C_0,y)} \left\{ \frac{\sum_i Q_iK_i}{\sum_i K_i} + \frac{\sum_i(\hat{Q}_i - Q_i)K_i}{\sum_i K_i} + \frac{\sum_i \hat{Q}_i(\hat{K}_i - K_i)}{\sum_i K_i} \right\} \\
&+ \left[\frac{\sum_i K_i}{\sum_i \bar{K}_i} - \frac{1}{P(C_0,y)} \right] \frac{\sum_i \hat{Q}_i\hat{K}_i}{\sum_i K_i} + \frac{\sum_i \hat{Q}_i\hat{K}_i}{\sum_i \tilde{K}_i} \cdot \frac{\sum_i(\bar{K}_i - \tilde{K}_i)}{\sum_i \bar{K}_i} \\
&= \frac{1}{P(C_0,y)} \left\{ \hat{H}_1(y) + \hat{H}_2(y) + \hat{H}_3(y) \right\} + \hat{H}(y)\frac{\sum_i(\bar{K}_i - \tilde{K}_i)}{\sum_i \bar{K}_i} \\
&+ \left[\frac{\sum_i K_i}{\sum_i \bar{K}_i} - \frac{1}{P(C_0,y)} \right] \frac{\sum_i \hat{Q}_i\hat{K}_i}{\sum_i K_i}.
\end{aligned}$$

Here, y denotes some point of the compact subset $\Phi(C)$. First, applying lemma 2., we obtain the rate of convergence of $\hat{H}_1(y)$, that is the kernel regression estimator of $\mathbf{1}\{X_i \in C_0\}Q(1|X_i)$ knowing $\Phi(X_i) = y$, viz $P(C_0,y)H(y)$.

LEMMA 4. *If*

i. *H, $P(C_0,\cdot)$ and $f_{\Phi(X)}$ are k'-times differentiable on $\mathbb{R}^p$, $k' \in \mathbb{N}$, and their derivatives of order k are γ'-Lipschitz continuous on $\mathbb{R}^p$ for some $\gamma' \in]0,1[$,*

ii. $\inf_{y\in\tilde{\Phi}(C)} f_{\Phi(X)}(y) > 0$,

iii. *K is a bounded p-dimensional kernel of order k', $\|x\|^p K(x) \to 0$ and K is γ'-Lipschitz continuous,*

then a.e.

$$\sup_{y\in\Phi(C)} |\hat{H}_1(y) - P(C_0,y)H(y)| = O\left(h^{k'+\gamma'} + (\ln n/(nh^p))^{1/2}\right) \text{, and}$$

$$\begin{aligned}\sup_{y\in\Phi(C)} &|\frac{\sum_{i=1}^n \mathbf{1}\{X_i\in C_0\}K_h(y-\Phi(X_i))}{\sum_{i=1}^n K_h(y-\Phi(X_i))} - P(C_0,y)| \\ &= O\left(h^{k'+\gamma'} + (\ln n/(nh^p))^{1/2}\right).\end{aligned}$$

Second, invoking lemma 3., a.e.

$$\begin{aligned}\sup_{y\in\Phi(C)} |\hat{H}_2(y)| &\le \sup_{i|X_i\in C_0} |\hat{Q}_i - Q_i| \cdot \sup_{y\in\Phi(C)} \frac{\sum_{i=1}^n |K_i|}{|\sum_{i=1}^n K_i|} \\ &= O\left(h_0^{k+\gamma} + \left(\ln n/\left(nh_0^d\right)\right)^{1/2}\right).h^{-p}\end{aligned}$$

Third, using the boundedness of the quantities $\hat{Q}(1|X_i)$ that has been deduced from lemma 3., we have a.e.

$$\begin{aligned}\sup_{y\in\Phi(C)} |\hat{H}_3(y)| &= \sup_{y\in\Phi(C)} |\frac{\sum_i \hat{Q}_i(\hat{K}_i - K_i)}{\sum_i K_i}| \\ &\le \sup_{X_i\in C_0} |\hat{Q}(1|X_i)| \frac{\sum_{i=1}^n \mathbf{1}\{X_i\in C_0\}|K_h'(y-\Phi^*(X_i))\cdot(\Phi-\hat{\Phi})(X_i)|}{h|\sum_{i=1}^n K_i|} \\ &\le \frac{h^{-1-p}\|K'\|_\infty a_n.Cst}{n^{-1}|\sum_i K_i|}.\end{aligned}$$

Since K is of bounded variation and $nh^p/\ln n \to \infty$, then $n^{-1}\sum_i K_i$ tends to $f_{\Phi(X)}(y)$ a.e. uniformly on every compact set of $\mathbb{R}^d$ where $f_{\Phi(X)}$ is continuous (see e.g. Bosq and Lecoutre [2]). Since $\inf_{y\in\Phi(C)} f_{\Phi(X)}(y) > 0$, we have a.e.

$$\sup_{y\in C} |\hat{H}_3(y)| = O(a_n h^{-1-p}). \tag{7.26}$$

Fourth, similarly to $\hat{H}_3$, we deduce a.e.

$$\begin{aligned}\sup_{y\in\Phi(C)} &|\hat{H}(y)\frac{\sum_i(\bar{K}_i - \tilde{K}_i)}{\sum_i \bar{K}_i}| \\ &\le \sup_{y\in\Phi(C)} Cst\left\{|\hat{H}(y)|.\sup_{X_i\in C_0}|\hat{\Phi}(X_i) - \Phi(X_i)|h^{-p-1}\cdot\frac{n}{\sum_i \bar{K}_i}|\right\}.\end{aligned}$$

Since $\sum_i \bar{K}_i/\sum_i K_i$ converges uniformly on $y\in\Phi(C)$ towards $P(C_0,y)$, and since the latter function is bounded from below by a strictly positive constant on $\Phi(C)$, we obtain

$$\sup_{y\in\Phi(C)} |\hat{H}(y)\frac{\sum_i(\bar{K}_i-\tilde{K}_i)}{\sum_i \bar{K}_i}| = O\left(a_n h^{-p-1}\sup_{y\in\Phi(C)}|\hat{H}(y)|\right). \quad (7.27)$$

Fifth, for each $y\in\Phi(C)$, a.e.

$$\left|\left[\frac{\sum_i K_i}{\sum_i \bar{K}_i} - \frac{1}{P(C_0,y)}\right]\frac{\sum_i \hat{Q}_i\hat{K}_i}{\sum_i K_i}\right|$$
$$\leq \left|\hat{H}(y)\cdot\left[\frac{\sum_i(\tilde{K}_i-\bar{K}_i)}{\sum_i K_i}+\frac{\sum_i\bar{K}_i}{\sum_i K_i}\right]\cdot\left[\frac{\sum_i K_i}{\sum_i\bar{K}_i}-\frac{1}{P(C_0,y)}\right]\right|.$$

Since $\sum_i \bar{K}_i/\sum_i K_i$ tends to $P(C_0,y)$ uniformly on $y\in\Phi(C)$, making the same reasoning as for the previous term and invoking lemma 2., we obtain a.e., if a_n/h^{1+p} tends to 0,

$$\sup_{y\in\Phi(C)}\left|\left[\frac{\sum_i K_i}{\sum_i \bar{K}_i} - \frac{1}{P(C_0,y)}\right]\frac{\sum_i \hat{Q}_i\hat{K}_i}{\sum_i K_i}\right|$$
$$= \sup_{y\in\Phi(C)}|\hat{H}(y)|.O\left(h^{k'+\gamma'}+(\ln n/(nh^p))^{1/2}\right).$$

Then, we have a.e.

$$\left(1+O\left(h^{k'+\gamma'}+(\ln n/(nh^p))^{1/2}+\frac{a_n}{h^{1+p}}\right)\right)\sup_{y\in\Phi(C)}|\hat{H}(y)-H(y)|$$
$$= O\left(h^{k'+\gamma'}+(\ln n/(nh^p))^{1/2}+(h_0^{k+\gamma}+(\ln n/\left(nh_0^d\right))^{1/2}).h^{-p}+\frac{a_n}{h^{1+p}}\right),$$

proving the result. □

Remark **4.** *If we suppose that K is positive, the proof is simpler, because the rate of convergence of $\sup_y|\hat{H}_2(y)|$ to zero is obviously the rate of convergence obtained in lemma 3.. Thus, we win a factor h^{-p}. The price would be to set $k'=2$, k' denoting the order of the kernel function K.*

Proof of lemma 1.

Tsiatis [33] proved that, for each j, t and x,

$$\tilde{\lambda}_j(t|x) = \frac{1}{S(t,\ldots,t|x)}\cdot\partial_j S(t_1,\ldots,t_p|x)_{|t_1=\cdots=t_p=t} \quad (7.28)$$

Deduce from Eq. (2.3) that

$$\tilde{\lambda}_j(t|x) = \frac{\partial_j H[\exp(-Z_1(t)\phi_1(x)),\ldots,\exp(-Z_p(t)\phi_p(x))]}{H[\exp(-Z_1(t)\phi_1(x)),\ldots,\exp(-Z_p(t)\phi_p(x))]}\cdot\frac{\phi_j(x)Z_j'(t)}{\exp(Z_j(t)\phi_j(x))}. \quad (7.29)$$

From (4.12) this can be rewritten

$$\frac{\partial_j H[\exp(-Z_1(t)\phi_1(x)),\ldots,\exp(-Z_p(t)\phi_p(x))]}{H[\exp(-Z_1(t)\phi_1(x)),\ldots,\exp(-Z_p(t)\phi_p(x))]\exp(Z_j(t)\phi_j(x))} = \frac{\psi_j(x'\beta_j)}{\phi_j(x)}A(t), \quad (7.30)$$

for some function A. Taking the limit as $t\to 0$ proves that $\psi_j(x'\beta_j)$ is proportional to $\phi_j(x)$. Then, due to condition $(H2)$, ϕ_j is entirely known.

Set $t=1$ in (7.30) and define the change of variables from x to $(u_1,\ldots,u_p)$ in $\mathbb{R}^p$ by $(u_1,\ldots,u_p)=\Phi(x)$. By assumption $(u_1,\ldots,u_p)$ describes all the set $]0,1[^p$, and there exists some constants c_j such that, for all $j=1,\ldots,p$ and u_j,

$$u_j\partial_j H[u_1,\ldots,u_p] = c_j H[u_1,\ldots,u_p].$$

It is easy to solve this differential equation. Thus, reminding that H is bounded and $\lim_{u\to 1} H(u,\dots,u) = 1$, there exist some nonnegative constants $c_1,\dots,c_p$ such that

$$H[u_1,\dots,u_p] = u_1^{c_1}\dots u_p^{c_p}.$$

proving the assertion.□

Proof of theorem 2.
Consider, for each $t\geq 0$, $\rho_n \geq 0$ and $x\in\mathbb{R}^d$, the quantity

$$\hat{\phi}_j(t,x) \equiv \frac{\hat{N}_j(t|x)}{\hat{D}_j(x)+\rho_n}\cdot\frac{\hat{D}_j(x_0)}{\hat{N}_j(t|x_0)}, \tag{7.31}$$

$$\hat{N}_j(t|x) = \frac{1}{n}\sum_{i=1}^n \mathbf{1}\{\delta_i = j\} M_{h_1}(t-Y_i)N_{h_2}(x-X_i),$$

$$\hat{D}_j(t|x) = \frac{1}{n}\sum_{i=1}^n N_{h_2}(x-X_i).$$

When $\rho_n\equiv 0$, note that $\hat{\phi}_j(0,x) = \hat{\phi}_j^{(0)}(x)$ and $\hat{\phi}_j(\alpha_n,x) = \hat{\phi}_j^{(1)}(x)$. The extra factor can be useful to obtain similar results uniformly on x belonging to an unbounded compact subset of $\mathbb{R}^d$. Since this is necessary in the discussion of $\hat{Z}_j^{(2)}(t)$ (see the next proof), we keep the extra term ρ_n. this To simplify, assume that $supp(M)\subset[-1,1]$ and that $supp(N)\subset[-1,1]^d$. First, example 2.38 and exercise 2.28 from Pollard [25] provide

$$\sup_{t\geq 0, x\in\mathbb{R}^d} |\hat{D}_j(t|x) - E[\hat{D}_j(t|x)]| = O\left(\frac{\ln n}{(nh_2^d)1/2}\right) \equiv v_n \quad \text{a.e.} \tag{7.32}$$

Moreover, consider the family of real valued functions $\{\phi_{t,x,h_1,h_2}, t\geq 0, x\in\mathbb{R}^d, h_1>0, h_2>0\}$, where for all $\vec{t} = (t_1,\dots,t_p)\in\mathbb{R}^p$ and $z\in\mathbb{R}^d$,

$$\phi_{t,x,h_1,h_2}(\vec{t},z) = \mathbf{1}\{t_k\geq t_j, k=1,\dots,p\}M_{h_1}(t-t_j)N_{h_2}(x-z).$$

Since the graphs associated to this family have polynomial discrimination, deduce from example 2.38 of Pollard (1984) that for all j,

$$\sup_{t\geq 0, x\in\mathbb{R}^d} |\hat{N}_j(t|x) - E[\hat{N}_j(t|x)]| = O\left(\frac{\ln n}{(nh_1h_2^d)^{1/2}}\right) \equiv u_n \quad \text{a.e.} \tag{7.33}$$

Note that

$$\begin{aligned} E[\hat{N}_j(t|x)] &= E[\mathbf{1}\{\delta = j\}M_{h_1}(t-Y)N_{h_2}(x-X)] \\ &= -\int \partial_j S(v,\dots,v|u) f_X(u) M_{h_1}(t-v)N_{h_2}(x-u)\,dv\,du \\ &= -\int \psi_j(t-h_1v|x-h_2u) f_X(x-h_2u)M(v)N(u)\,dv\,du, \end{aligned}$$

denoting

$$\begin{aligned} \psi_j(t|x) &\equiv \partial_j S(t,\dots,t|x) \\ &= -\partial_j H[\exp(-Z_1(t)\phi_1(x)),\dots,\exp(-Z_p(t)\phi_p(x))]\frac{Z_j'(t)\phi_j(x)}{\exp(Z_j(t)\phi_j(x))}. \end{aligned}$$

Note that the domain of integration can be restricted to real numbers v such that $t-h_1v\geq 0$, since $\psi_j(\cdot|x)$ is zero on $\mathbb{R}_-^*$. By assumption, for all x, the function $\psi_j(\cdot|x)$ is differentiable on $]0,\eta[$, for some $\eta > 0$. Moreover, there exists a unique continuous prolongation of $\psi_j(\cdot|x)$ at 0 on the right, denoted by $\psi_j(0^+|x)$.

For each t, we have

$$
\begin{aligned}
E[\hat{N}_j(t|x)] &= -\psi_j(t|x) f_X(x) \int_{h_1 v \le t} M(v)\, d\, v \\
&- \frac{h_2^2}{2} \int d^2[f_X(\cdot)\psi_j(t|\cdot)]_{|x-\bar{\theta}h_2 u}.u^{(2)} N(u)\, du \cdot \int_{h_1 v\le t} M(v)\, d\, v \\
&+ h_1 \int_{h_1 v \le t} v M(v) N(u) \partial \psi_j(t|x - h_2 u) f_X(x - h_2 u)\, dv\, du \\
&- \frac{h_1^2}{2} \int_{h_1 v\le t} v^2 M(v) N(u) \partial^2 \psi_j(t - \theta h_1 v | x - h_2 u) f_X(x - h_2 u)\, dv\, du \\
&\equiv A_n(t) f_X(x)\psi_j(t|x) + h_2^2 B_n(t,x) + h_1 \bar{C}_n(t,x) + h_1^2 C_n(t,x),
\end{aligned}
$$

where θ and $\bar{\theta}$ denote real numbers between 0 and 1. We have used the fact that N is a 2-order kernel.

Therefore, we have for every real number t and every $x \in C$,

$$
\begin{aligned}
A_n(t) &= -\int_{t \ge h_1 v} M(v)\, dv, \\
|B_n(t,x)| &\le Cst. \left\| d^2[f_X(\cdot)\psi_j(t|\cdot)] \right\|^* (x), \\
|\bar{C}_n(t,x)| &\le Cst. f_X^*(x) \left\| \partial \psi_j(t|\cdot)] \right\|^* (x), \\
|C_n(t,x)| &\le Cst. f_X^*(x) \left\| \partial^2 \psi_j(\cdot|\cdot)] \right\|^* (t,x).
\end{aligned}
$$

Note that $A_n(\alpha_n) = 1$ and $\bar{C}_n(\alpha_n, x) = 0$ when $\alpha_n \ge h_1$, which is assumed to deal with $\hat{\phi}_j^{(1)}$. Thus,

$$
E\hat{N}_j(\alpha_n|x) = -\psi_j(\alpha_n|x) f_X(x) + O(h_1^2 + h_2^2), \text{ and} \tag{7.34}
$$

$$
E\hat{N}_j(0|x) = -\frac{1}{2}\psi_j(0^+|x) f_X(x) + O(h_1 + h_2^2). \tag{7.35}
$$

Similarly, we obtain easily

$$
\begin{aligned}
E[\hat{D}_j(x)] &= f_X(x) + h_2^2 D_n(x), \\
|D_n(x)| &\le Cst. \left\| d^2 f_X \right\|^* (x).
\end{aligned}
$$

Thus,

$$
\begin{aligned}
&\hat{\phi}_j(x) - \phi_j(x) = \\
&\quad \left(\frac{A_n(\alpha_n) f_X(x)\psi_j(\alpha_n|x) + h_2^2 B_n(\alpha_n, x) + h_1^2 C_n(\alpha_n, x) + u_n}{A_n(\alpha_n) f_X(x_0)\psi_j(\alpha_n|x_0) + h_2^2 B_n(\alpha_n, x_0) + h_1^2 C_n(\alpha_n, x_0) + u_n} \right. \\
&\quad \left. \cdot \quad \frac{f_X(x_0) + h_2^2 D_n(x_0) + v_n}{f_X(x) + h_2^2 D_n(x) + v_n + \rho_n} - \frac{\psi_j(\alpha_n|x)}{\psi_j(\alpha_n|x_0)} \right) - \left(\phi_j(x) - \frac{\psi_j(\alpha_n|x)}{\psi_j(\alpha_n|x_0)} \right) \\
&\equiv \quad \Delta\phi_1 + \Delta\phi_2.
\end{aligned} \tag{7.36}
$$

Simple calculations show that $\Delta\phi_1$ is the ratio Δ_N/Δ_D, with

$$
\begin{aligned}
\Delta_N &= \psi_j(\alpha_n|x_0) \left[h_2^2 f_X(x_0) B_n(\alpha_n, x) + h_2^2 f_X(x)\psi_j(\alpha_n|x) D_n(x_0) \right. \\
&+ \quad h_1^2 f_X(x_0) C_n(\alpha_n, x) + h_2^4 D_n(x_0) B_n(\alpha_n, x) + h_1^2 h_2^2 D_n(x_0) C_n(\alpha_n, x) \\
&+ \quad \left. u_n f_X(x_0) + u_n h_2^2 D_n(x_0) + u_n v_n \right] - \psi_j(\alpha_n|x) \left[h_2^2 f_X(x) B_n(\alpha_n, x_0) \right. \\
&+ \quad h_2^2 f_X(x_0) D_n(x)\psi_j(\alpha_n|x_0) + h_1^2 f_X(x) C_n(\alpha_n, x_0) + h_2^4 D_n(x) B_n(\alpha_n, x_0) \\
&+ \quad \left. h_1^2 h_2^2 D_n(x) C_n(\alpha_n, x_0) + u_n f_X(x) + u_n h_2^2 D_n(x) + u_n(\rho_n + v_n) \right] \\
&- \quad (\rho_n + v_n)\psi_j(\alpha_n|x) \left[f_X(x_0)\psi_j(\alpha_n|x_0) + h_2^2 B_n(\alpha_n, x_0) + h_1^2 C_n(\alpha_n, x_0) \right] \\
&+ \quad v_n \psi_j(\alpha_n|x_0) \left[f_X(x)\psi_j(\alpha_n|x) + h_2^2 B_n(\alpha_n, x) + h_1^2 C_n(\alpha_n, x) \right],
\end{aligned}
$$

$$\Delta_D = \psi_j(\alpha_n|x_0)\left(f_X(x_0)\psi_j(\alpha_n|x_0) + h_2^2 B_n(\alpha_n,x_0) + h_1^2 C_n(\alpha_n,x_0) + u_n\right)$$
$$\cdot \quad \left(f_X(x) + h_2^2 D_n(x) + \rho_n + v_n\right).$$

Let us assume that x belongs to some compact subset C. Hence $\phi(x)$ and $f_X(x)$ are uniformly bounded on C. Denoting (f_n) and (g_n) some sequences of real functions, we write $f_n \asymp g_n$, $f_n, g_n \geq 0$ when there exist some positive constants a and b such that $af_n(x) \leq g_n(x) \leq bf_n(x)$ for all the considered x and each n.

Since $\psi_j(\alpha_n|x)/Z_j'(\alpha_n)$ is bounded from above and below from zero, when $x \in C$, the behavior of $\psi_j(\alpha_n|x)$ depends on $Z_j'(\alpha_n)$'s one. It follows

$$\psi_j(\alpha_n|x) \asymp Z_j'(\alpha_n)\phi_j(x),$$

$$B_n(\alpha_n,x) \leq Z_j'(\alpha_n)\left[\phi_j^*(x)\|d^2 f_X\|^*(x) + \|df_X\|^*(x).\|d\phi\|^*(x)\right.$$
$$\left. + \quad f_X^*(x)\left(\left(\|d\phi\|^2\right)^*(x) + \|d^2\phi\|^*(x)\right)\right],$$

$$C_n(\alpha_n,x) \leq f_X^*(x)\phi_j^*(x)\left[|Z_j'''|^*(\alpha_n) + \|Z'\|^*(\alpha_n)|Z_j''|^*(\alpha_n)\|\phi\|^*(x)\right.$$
$$\left. + \quad |Z_j'|^*(\alpha_n)\left(\|Z'\|^2\right)^*(\alpha_n)\left(\|\phi\|^2\right)^*(x) + |Z_j'|^*(\alpha_n)\|Z''\|^*(\alpha_n)\|\phi\|^*(x)\right]$$
$$\asymp \quad f_X^*(x)\mathcal{Z}(\alpha_n).$$

Thus, we get

$$\Delta_D \asymp Z_j'(\alpha_n)^2 f_X(x).$$

Deduce that

$$\sup_{x\in C}|\Delta\phi_1| \leq \sup_{x\in C}\left|\frac{\Delta_N}{\Delta_D}\right| \leq \sup_{x\in C}\frac{Cst}{f_X(x)}\left|h_2^2 + h_1^2\frac{C_n(\alpha_n,x)}{Z_j'(\alpha_n)} + \frac{u_n}{Z_j'(\alpha_n)} + \rho_n + v_n\right|. \qquad (7.37)$$

Remark **5.** *Note that, if $\rho_n >> h_2^2 + v_n$, the same inequality is true for every $x \in \mathbb{R}^d$, changing the denominator into $f_X(x) + \rho_n$.*

For sake of simplicity, we have assumed that the partial second derivatives of H are bounded uniformly on every compact subset, particularly on the image set of the application $\bar\Phi : [0,\eta]\times C \to]0,1]^p$, $\eta > 0$, $\bar\Phi(t,x) \equiv (\exp(-Z_1(t)\phi_1(x)),\ldots,\exp(-Z_p(t)\phi_p(x)))$. Then, with obvious notations, we have

$$\left|\frac{\psi_j(\alpha_n|x)}{\psi_j(\alpha_n|x_0)} - \phi_j(x)\right| \leq \frac{\phi_j(x)}{\phi_j(x_0)}\frac{\exp(Z_j(\alpha_n)\phi_j(x_0))}{|\partial_j H(\bar\Phi(\alpha_n,x_0)|)}$$
$$\cdot \quad \left\{\sum_{k=1}^p \frac{\partial_{jk}^2 H[\bar\Phi(\alpha_n,x^*)]}{\exp(Z_k(\alpha_n)\phi_k(x^*))} d\phi_k(x^*).(x-x_0)Z_k(\alpha_n)\exp(-Z_j(\alpha_n)\phi_j(x))\right.$$
$$\left. + \quad \partial_j H(\bar\Phi(\alpha_n,x))\exp(-Z_j(\alpha_n)\phi_j(x^*))Z_j(\alpha_n)d\phi_j(x^*).(x-x_0)\right\},$$

where $\|x^* - x_0\| \leq \|x - x_0\|$. Hence,

$$\sup_{x\in C}|\Delta\phi_2| \leq \sup_{x\in C}\left|\frac{\psi_j(\alpha_n|x)}{\psi_j(\alpha_n|x_0)} - \phi_j(x)\right| \leq Cst.\left(\sup_{k=1,\ldots,p}|Z_k|(\alpha_n)\right). \qquad (7.38)$$

Deduce from (7.36), (7.37) and (7.38) that

$$\sup_{x\in C}|\hat\phi_j(x) - \phi_j(x)| = O\left(\|Z\|(\alpha_n) + \left[h_2^2 + h_1^2\frac{\mathcal{Z}_j(\alpha_n)}{Z_j'(\alpha_n)} + \frac{u_n}{Z_j'(\alpha_n)} + \rho_n + v_n\right]\right). \qquad (7.39)$$

In the particular case $\rho_n \equiv 0$, we obtain the result for $\hat{\phi}_j^{(1)}$.

If $Z'(0+) \neq 0$, take $\alpha_n \equiv 0$. We can now deal with $\hat{\phi}_j^{(0)}$ in the same way. Since $A_n(0) = 1/2 \neq 0$, we have to replace $h_1^2 C_n(\alpha_n, x)$ by $h_1 \bar{C}_n(0, x)$ in the previous inequalities. Therefore,

$$\begin{aligned}\sup_{x\in C}\left|\frac{\Delta_N}{\Delta_D}\right| &\leq \sup_{x\in C}\frac{Cst}{f_X(x)+\rho_n}\left|h_2^2 + h_1\frac{\bar{C}_n(0,x)}{Z_j'(0+)} + \frac{u_n}{Z_j'(0+)} + \rho_n + v_n\right| \\ &\leq O\left(h_2^2 + h_1 + u_n\right), \end{aligned} \tag{7.40}$$

choosing $\rho_n \equiv 0$. Since $\Delta\phi_2 = 0$ in this case, we have stated the result for $\hat{\phi}_j^{(0)}$. □

Remark **6.** *It is possible to replace the factor h_1 in (7.40) by h_1^2, using local polynomials or sieves instead of kernels. Indeed, these methods allow asymptotically unbiased estimators near the boundaries (see Fan and Gijbels [6] e.g.).*

Proof of theorem 3.

Our goal is to prove the weak consistency of $\hat{Z}^{(2)}(t)$. We do not precise the rates of convergence. These rates are connected with the tail behavior of $f_X(X)$ and $\phi(X)$ and are difficult to exhibit. Unfortunately, the technical assumptions are numerous, often messy and most of them cannot be verified easily. They will appear during the progress of the proof. We will not try to provide the weakest assumptions but rather to simplify at most some sufficient conditions.

Applying Pollard [25], we get easily that a.e.

$$\begin{aligned}\sup_{x\in\mathbb{R}^d, t\in\mathbb{R}} & |n^{-1}\sum_{k=1}^n \mathbf{1}\{Y_k > t, \delta_k = j\} L_{h_0}(x - X_k) \\ - & E[\mathbf{1}\{Y_k > t, \delta_k = j\} L_{h_0}(x - X_k)]| = O\left(\frac{\ln n}{(nh_0^d)^{1/2}}\right).\end{aligned}$$

It is necessary to precise the previous expectation. Denote $P(Y > t, \delta = j | X = x) \equiv Q_j(t|x)$. Since L is even and $d^2 f$ is bounded, we get

$$\begin{aligned} E[\mathbf{1}\{Y_k > t, \delta_k = 1\} L_{h_0}(x - X_k)] &= \int Q_j(t|u) L_{h_0}(x-u) f_X(u)\, du \\ \simeq\ & Q_j(t|x) f_X(x) + O\left(h_0^2\right).\end{aligned}$$

Thus, a.e.

$$\sup_{x\in\mathbb{R}^d, t\in\mathbb{R}} |\frac{1}{n}\sum_{k=1}^n \mathbf{1}\{Y_k > t, \delta_k = j\} L_{h_0}(x - X_k) - Q_j(t|x) f_X(x)| = O\left(\frac{\ln n}{(nh_0^d)^{1/2}} + h_0^2\right).$$

Moreover,

$$\begin{aligned} Q_j(t|x) = & \int \mathbf{1}\{u > t\} \partial_j H[\exp(-Z_1(u)\phi_1(x)), \ldots, \exp(-Z_p(u)\phi_p(x))] \\ \cdot & \exp(-Z_j(u)\phi_j(x))\phi_j(x) Z_j'(u)\, du.\end{aligned}$$

The function $\partial_j H(u)$ can be extended continuously at 1 by assumption. Denote $\partial_j H(1) \equiv \lim_{u\to 1} \partial_j H(u)$. Therefore,

$$\begin{aligned} & |\partial_j H\left[\exp(-Z_1(u)\phi_1(x)), \ldots, \exp(-Z_p(u)\phi_p(x)\right] - \partial_j H(1)| \\ \leq\ & \sum_{k=1}^p \|\partial_{jk}^2 H\|_\infty \exp(-\theta_k Z_k(u)\phi_k(x))\phi_k(x) Z_k(u) \\ \leq\ & Cst. \sum_{k=1}^p Z_k(u)\phi_k(x), \end{aligned}$$

where θ_k denotes some real numbers between 0 and 1. Assume for instance that

Assumption Q.0 :

$$\sup_{\lambda>0,k=1,\ldots,p} \int \exp(-\lambda Z_j(u))\lambda Z_k(u) Z_j'(u) < \infty.$$

Particularly, this is satisfied when the Z_k, $k=1,\ldots,p$ are some polynomials. Then, for each i,

$$\sup_{t>0} |Q_j(t|X_i) - \partial_j H(1)\exp(-Z_j(t)\phi_j(X_i))| \leq Cst. \sum_{k=1}^p \phi_k(X_i).$$

We obtain for every X_i,

$$\begin{aligned} \ln\left[\frac{P_1(X_i)}{P_t(X_i)}\right] &= \ln\left[\partial_j H(1)\exp(-\phi_j(X_i)) f_X(X_i)\right. \\ +& \quad d_n + O\left((\sum_{k=1}^p \phi_k(X_i)) f_X(X_i) + h_0^2 + \ln n/(nh_0^d)^{1/2}\right)\Bigg] \\ -& \quad \ln\left[\partial_j H(1)\exp(-Z_j(t)\phi_j(X_i)) f_X(X_i) + d_n\right. \\ +& \quad O\left((\sum_{k=1}^p \phi_k(X_i)) f_X(X_i) + h_0^2 + \ln n/(nh_0^d)^{1/2}\right)\Bigg] = (Z_j(t)-1)\phi_j(X_i) + r_i, \end{aligned}$$

where r_i denotes some remainder term associated with X_i. To simplify, assume that the support of $\bar{M}$ is included into $[-1,1]^p$. We will consider mainly the points X_i such that

$$\sup_{k=1,\ldots,p} \phi_k(X_i) \leq 2l. \tag{7.41}$$

Because of the definition of R_n, the points X_i that provide non zero terms in the sum (5.23) satisfy such a constraint (at least approximately and in probability, for n sufficiently large).

For the points X_i satisfying (7.41) and for n sufficiently large, simple calculations provide that $|r_i|$ is less than the ratio

$$\frac{d_n[\exp(-\phi_j(X_i)) - \exp(-Z_j(t)\phi_j(X_i))] + O(h_0^2 + \ln n/(nh_0^d)^{1/2} + l f_X(X_i))}{\exp(-\phi_j(X_i) - Z_j(t)\phi_j(X_i)) f_X(X_i) + d_n + 0(h_0^2 + \ln n/(nh_0)^{1/2} + l f_X(X_i))}.$$

Thus, for each constant $T>0$ and uniformly on the points X_i satisfying (7.41), we have

$$\begin{aligned} \sup_{|t|\leq T} |r_i| &\leq \frac{1}{f_X(X_i) + d_n + o(d_n)} O\left(d_n l + h_0^2 + \frac{\ln n}{(nh_0^d)^{1/2}} + l f_X(X_i)\right) \\ &\leq O\left(l + \frac{h_0^2}{d_n} + \frac{\ln n}{d_n (nh_0^d)^{1/2}}\right) \equiv \tilde{l}. \end{aligned} \tag{7.42}$$

Moreover, it will be necessary to control the distance between $\hat{\phi}(X_i)$ and $\phi(X_i)$ for the points X_i such that $\phi_k(X_i) \leq 2l$, $k=1,\ldots,p$. In the proof of theorem 2., we have shown that for all $x \in \mathbb{R}^d$ and j, there exists a function $\chi_j(x,n)$, which tends to zero when n tends to infinity, such that a.e.

$$|\hat{\phi}_j(x) - \phi_j(x)| \leq \frac{\chi_j(x,n)}{f_X(x) + \rho_n}. \tag{7.43}$$

To simplify, we assume that $Z_j'(0+) \neq 0$. Thus, we are dealing with $\hat{\phi}_j = \hat{\phi}_j^{(0)}$. Using the same calculations and notations as in the proof of theorem 2., it can be proved that for each x, and each

j,

$$
\begin{aligned}
|\chi_j(x,n)| &\leq O\left(h_2^2 \left\| d^2[f_X(\cdot)\psi_j(t|\cdot)]\right\|^* (x) + h_2^2 f_X(x)\phi_j(x) + h_2^2\phi_j(x)\left\|d^2 f_X\right\|^*(x)\right.\\
&+ \quad h_1 f_X^*(x)\left\|\partial\psi_j(\cdot|\cdot)\right\|^*(0,x) + u_n + h_1\phi_j(x)f_X(x) + u_n\phi_j(x)f_X(x)\\
&+ \quad \left.(\rho_n + v_n)\phi_j(x)\right)\\
&\leq \quad O\left(h_2^2\left[\phi_j^*(x)\|d^2 f_X\|^*(x) + \|df_X\|^*(x).\|d\phi\|^*(x)\right.\right.\\
&+ \quad \left. f_X^*(x)\left(\left(\|d\phi\|^2\right)^*(x) + \|d^2\phi\|^*(x)\right) + f_X(x)\phi_j(x) + \phi_j(x)\left\|d^2 f_X\right\|^*(x)\right]\\
&+ \quad h_1\left[f_X^*(x).\|\phi\|^*(x)\phi_j^*(x) + \phi_j(x)f_X(x)\right]\\
&+ \quad \left. u_n\left[1 + \phi_j(x)f_X(x)\right] + (\rho_n + v_n)\phi_j(x)\right), \qquad (7.44)
\end{aligned}
$$

where $u_n = O(\ln n/(nh_1h_2^d)^{1/2})$ and $v_n = O(\ln n/(nh_2^d)^{1/2})$.

Denoting $\hat{\phi}_i = (\hat{\phi}_1(X_i),\ldots,\hat{\phi}_p(X_i))$ and $\phi_i = (\phi_1(X_i),\ldots,\phi_p(X_i))$, we have

$$
\begin{aligned}
\hat{Z}_j^{(2)}(t) &= 1 + \sum_{i=1}^n R_n(\hat{\phi}_i)\frac{\mathbf{1}\{\phi_i \leq 2l\}}{\hat{\phi}_j(X_i) + e_n}\ln\left[\frac{P_1(X_i)}{P_t(X_i)}\right]\\
&+ \quad \sum_{i=1}^n R_n(\hat{\phi}_i)\frac{\mathbf{1}\{\phi_i > 2l\}}{\hat{\phi}_j(X_i) + e_n}\ln\left[\frac{P_1(X_i)}{P_t(X_i)}\right] \equiv \tilde{Z}_j^{(2)}(t) + T_0.
\end{aligned}
$$

The first step is to prove that T_0 is negligible.

Study of T_0. For the points X_i which do not satisfy (7.41), we need to control the quantities $\ln P_t(X_i)$. Since L is positive, $P_t(X_i) \geq d_n$. Moreover,

$$
\begin{aligned}
P_t(X_i) &\leq \partial_j H(1)\exp(-\phi_j(X_i)Z_j(t))f_X(X_i) + d_n\\
&+ \quad Cst.\left(\sum_{k=1}^p \phi_k(X_i)\right)f_X(X_i) + h_0^2 + \frac{\ln n}{(nh_0^d)^{1/2}}.
\end{aligned}
$$

Thus, for each t and n sufficiently large,

$$
\begin{aligned}
|\ln(P_t(X_i)| &\leq 2|\ln d_n| + \left|\ln\left(d_n + f_X(X_i)(1 + \sum_{k=1}^p\phi_k(X_i)) + h_0^2 + \frac{\ln n}{(nh_0^d)^{1/2}}\right)\right|\\
&\leq \quad 2|\ln d_n| + \left|\ln\left(1 + f_X(X_i)(1 + \sum_{k=1}^p\phi_k(X_i))\right)\right| \equiv B(X_i).
\end{aligned}
$$

Moreover, applying Pollard [25], we get a.e.

$$
n^{-1}\sum_{i=1}^n \bar{M}_l(\phi_i) = E[\bar{M}_l(\phi_i)] + \left(\frac{\ln n}{(nl^p)^{1/2}}\right), \qquad (7.45)
$$

$$
\begin{aligned}
P\left(\left|n^{-1}\sum_{i=1}^n\bar{M}_l(\hat{\phi}_i) - n^{-1}\sum_{i=1}^n\bar{M}_l(\phi_i)\right| > \varepsilon w_n\right) &\leq P\left(\sum_{i=1}^n\|\hat{\phi}_i - \phi_i\| > nl^{p+1}\varepsilon w_n\right)\\
\leq \quad \frac{1}{nl^{p+1}\varepsilon w_n}E\left[\frac{\|\chi(X_i,n)\|}{f_X(X_i) + \rho_n}\right]
\end{aligned}
$$

for every positive sequence $(w_n)_n$. The latter term tends to zero for some sequence (w_n) satisfying

ASSUMPTION Q.1 : For some $\alpha > 0$,

$$E\left[\|\chi(X_i,n)\|\mathbf{1}\{f_X(X_i) \leq \rho_n^{1-\alpha}\}\right] + \rho_n^{\alpha} E\left[\|\chi(X_i,n)\|\right] = o\left(l^{p+1}\rho_n w_n\right).$$

Therefore, the denominator of T_0 is

$$\begin{aligned} n^{-1}\sum_{i=1}^{n} \bar{M}_l(\hat{\phi}_i) + b_n &= E[\bar{M}_l(\phi_i)] + b_n + o_P\left(\frac{\ln n}{(nl^p)^{1/2}} + w_n\right) \\ &\equiv E[\bar{M}_l(\phi_i)] + \tilde{b}_n. \end{aligned} \tag{7.46}$$

Thus, since $R_n(\phi_i)\mathbf{1}\{\phi_i > 2l\}$ is zero, we get

$$\begin{aligned} & P(\sup_{t\leq T} |T_0| > \varepsilon) \\ \leq\ & P\left(\frac{2}{n}\sum_{i=1}^{n} \frac{\|\hat{\phi}_i - \phi_i\|.\mathbf{1}\{\phi_i > 2l, \hat{\phi}_i \leq l\}B(X_i)}{(\hat{\phi}_j(X_i) + e_n)} > \varepsilon l^{p+1}\left(E[\bar{M}_l(\phi_i)] + \tilde{b}_n\right)\right) \\ \leq\ & \frac{2E\left[\|\hat{\phi}_i - \phi_i\|.\mathbf{1}\{\|\hat{\phi}_i - \phi_i\| > l\}B(X_i)\right]}{\varepsilon e_n l^{p+1}(E[\bar{M}_l(\phi_i)] + \tilde{b}_n)} \\ \leq\ & \frac{Cst}{\varepsilon e_n l^{p+1}(E[\bar{M}_l(\phi_i)] + \tilde{b}_n)} \cdot E\left[\frac{\|\chi(X_i,n)\|}{(f_X(X_i)+\rho_n)}.\mathbf{1}\{\|\frac{\chi(X_i,n)\|}{f_X(X_i)+\rho_n} > l\}B(X_i)\right] \\ \leq\ & \frac{Cst|\ln d_n|}{\varepsilon e_n l^{p+1}E[\bar{M}_l(\phi_i)]} E\left[\frac{\|\chi(X_i,n)\|}{(f_X(X_i)+\rho_n)}\mathbf{1}\{\|\frac{\chi(X_i,n)\|}{(f_X(X_i)+\rho_n)} > l\}\right] \\ +\ & \frac{Cst}{\varepsilon e_n l^{p+1}E[\bar{M}_l(\phi_i)]} E\left[\|\chi(X_i,n)\|\mathbf{1}\{\|\frac{\chi(X_i,n)\|}{f_X(X_i)+\rho_n} > l\}\left(1 + \sum_{k=1}^{p}\phi_k(X_i))\right)\right]. \end{aligned}$$

It is not easy to find simple conditions which ensure $\sup_{t\leq T}|T_0| = o_P(1)$. Nonetheless, it is the case under

ASSUMPTION Q.2 : For some $\alpha > 0$,

$$E\left[\|\chi(X_i,n)\|\mathbf{1}\{\|\chi(X_i,n)\| > l\rho_n, f_X(X_i) \leq \rho_n^{1-\alpha}\}\right] = o\left(e_n l^{p+1}\rho_n \ln(d_n) E[\bar{M}_l(\phi_i)]\right),$$

$$E\left[\|\chi(X_i,n)\|\mathbf{1}\{\|\chi(X_i,n)\| > l f_X(X_i)\}\right] = o\left(e_n l^{p+1}\rho_n^{1-\alpha}\ln(d_n)E[\bar{M}_l(\phi_i)]\right),$$

$$E\left[\|\chi(X_i,n)\|\mathbf{1}\{\|\chi(X_i,n)\| > l f_X(X_i)\}(1 + \sum_{k=1}^{p}\phi_k(X_i))\right] = o\left(e_n l^{p+1}E[\bar{M}_l(\phi_i)]\right),$$

$$E[\bar{M}_l(\phi_i)] >> b_n + w_n + \ln n/(nl^p)^{1/2}.$$

Thus it is sufficient to discuss

$$
\begin{aligned}
\tilde{Z}_j^{(2)}(t) &= 1 + \sum_{i=1}^n R_n(\hat{\phi}_i) \frac{\mathbf{1}\{\phi_i \le 2l\}}{\hat{\phi}_j(X_i) + e_n} \ln\left[\frac{P_1(X_i)}{P_t(X_i)}\right] \\
&= 1 + \sum_{i=1}^n R_n(\hat{\phi}_i) \frac{\mathbf{1}\{\phi_i \le 2l\}}{\hat{\phi}_j(X_i) + e_n} \left[(Z_j(t) - 1)\phi_j(X_i) + r_i\right] \\
&= 1 + \sum_{i=1}^n R_n(\hat{\phi}_i) \frac{r_i \mathbf{1}\{\phi_j(X_i) \le 2l\}}{\hat{\phi}_j(X_i) + e_n} + (Z_j(t) - 1). \left(\sum_{i=1}^n R_n(\hat{\phi}_i)\mathbf{1}\{\phi_j(X_i) \le 2l\} \right. \\
&+ \left. \sum_{i=1}^n R_n(\hat{\phi}_i)\mathbf{1}\{\phi_j(X_i) \le 2l\} \frac{(\phi_j(X_i) - \hat{\phi}_j(X_i))}{\hat{\phi}_j(X_i) + e_n} - e_n \sum_{i=1}^n \frac{R_n(\hat{\phi}_i)\mathbf{1}\{\phi_j(X_i) \le 2l\}}{\hat{\phi}_j(X_i) + e_n} \right) \\
&= Z_j(t) + \sum_{i=1}^n R_n(\hat{\phi}_i) \frac{r_i \mathbf{1}\{\phi_j(X_i) \le 2l\}}{\hat{\phi}_j(X_i) + e_n} + (Z_j(t) - 1) \\
&\cdot \left(\sum_{i=1}^n R_n(\hat{\phi}_i)\mathbf{1}\{\phi_j(X_i) \le 2l\} \frac{(\phi_j(X_i) - \hat{\phi}_j(X_i))}{\hat{\phi}_j(X_i) + e_n} - \frac{b_n}{n^{-1}\sum_{k=1}^n \bar{M}_l(\hat{\phi}_k) + b_n} \right. \\
&- \left. \sum_{i=1}^n R_n(\hat{\phi}_i)\mathbf{1}\{\phi_j(X_i) > 2l\} - e_n \sum_{i=1}^n \frac{R_n(\hat{\phi}_i)\mathbf{1}\{\phi_j(X_i) \le 2l\}}{\hat{\phi}_j(X_i) + e_n} \right) \\
&\equiv Z_j(t) + T_1 + (Z_j(t) - 1)[T_2 + T_3 + T_4 + T_5].
\end{aligned}
$$

Then, the goal would be to show that T_k, $k = 1, \ldots, 5$ tends to zero in probability under some convenient assumptions.

Study of T_1 and T_5. These two terms are very similar. Thus, to prove that T_1 and T_5 are $o_P(1)$, it is sufficient to show (with the notation of (7.42) and (7.46)) that

$$
\frac{e_n + \tilde{l}}{n} \sum_{i=1}^n \bar{M}_l(\hat{\phi}_i) \frac{\mathbf{1}\{\phi_j \le 2l\}}{\hat{\phi}_j(X_i) + e_n} = o_P\left(E[\bar{M}_l(\phi_i)] + \tilde{b}_n\right).
$$

But we get easily, for each $\eta > 0$,

$$
\begin{aligned}
P\left(\frac{1}{n}\sum_{i=1}^n \bar{M}_l(\hat{\phi}_i)\frac{\mathbf{1}\{\phi_j \le 2l\}}{\hat{\phi}_j(X_i) + e_n} > \eta\right) &\le \frac{1}{\eta} E\left[\bar{M}_l(\hat{\phi}_i)\frac{|\phi_j(X_i) - \hat{\phi}_j(X_i)|\mathbf{1}\{\phi_i \le 2l\}}{(\phi_j(X_i) + e_n)(\hat{\phi}_j(X_i) + e_n)}\right] \\
&+ \frac{1}{\eta} E\left[|\bar{M}_l(\hat{\phi}_i) - \bar{M}_l(\phi_i)|\frac{\mathbf{1}\{\phi_i \le 2l\}}{(\phi_j(X_i) + e_n)}\right] + \frac{1}{\eta} E\left[\bar{M}_l(\phi_i)\frac{\mathbf{1}\{\phi_i \le 2l\}}{(\phi_j(X_i) + e_n)}\right] \\
&\le \frac{1}{\eta e_n} E\left[\bar{M}_l(\hat{\phi}_i)\frac{\chi_j(X_i, n)\mathbf{1}\{\phi_i \le 2l\}}{(f_X(X_i) + \rho_n)(\phi_j(X_i) + e_n)}\right] \\
&+ \frac{1}{\eta l^{p+1}} E\left[\frac{\|\chi(X_i, n)\|\mathbf{1}\{\phi_i \le 2l\}}{(f_X(X_i) + \rho_n)(\phi_j(X_i) + e_n)}\right] + \frac{1}{\eta} E\left[\frac{\bar{M}_l(\phi_i)\mathbf{1}\{\phi_i \le 2l\}}{(\phi_j(X_i) + e_n)}\right] \\
&\le \frac{(l + e_n)}{\eta e_n l^{p+1}} E\left[\frac{\|\chi(X_i, n)\|\mathbf{1}\{\phi_i \le 2l\}}{(f_X(X_i) + \rho_n)(\phi_j(X_i) + e_n)}\right] + \frac{1}{\eta} E\left[\frac{\bar{M}_l(\phi_i)\mathbf{1}\{\phi_i \le 2l\}}{(\phi_j(X_i) + e_n)}\right].
\end{aligned}
$$

Set $\eta = \varepsilon(E[\bar{M}_l(\phi_i)] + \tilde{b}_n)/(e_n + \tilde{l})$. Thus, so that T_1 or T_5 could be $o_P(1)$, it is sufficient that the following assumption be satisfied.

Assumption Q.3 : For some $\beta > 0$, we have

$$
E\left[\frac{\|\chi(X_i, n)\|}{(f_X(X_i) + \rho_n)}\mathbf{1}\{\phi_i \le 2l, \phi_j(X_i) \le e_n^{1-\beta}\}\right] = o\left(\frac{l^{p+1}e_n^2 E[\bar{M}_l(\phi_i)]}{(l + e_n)(\tilde{l} + e_n)}\right),
$$

$$E\left[\frac{\|\chi(X_i,n)\|\mathbf{1}\{\phi_i\leq 2l\}}{(f_X(X_i)+\rho_n)}\right]=o\left(\frac{l^{p+1}e_n^{2-\beta}E[\bar{M}_l(\phi_i)]}{(l+e_n)(\tilde{l}+e_n)}\right),$$

$$E\left[\bar{M}_l(\phi_i)\mathbf{1}\{\phi_i\leq \inf(2l,e_n^{1-\beta})\}\right]+e_n^{\beta}E\left[\bar{M}_l(\phi_i)\mathbf{1}\{\phi_i\leq 2l\}\right]=o\left(\frac{e_nE[\bar{M}_l(\phi_i)]}{e_n+\tilde{l}}\right).$$

Study of T_2. A rough upper bound for T_2 could be

$$\begin{aligned}
P(|T_2|>\varepsilon) &\leq P\left(\left|\frac{Cst}{n}\sum_{i=1}^n\bar{M}_l(\phi_i)\frac{\chi_j(X_i,n)\mathbf{1}\{\phi_i\leq 2l\}}{\rho_n+f_X(X_i)}\right|>\varepsilon e_n\left(E[\bar{M}_l(\phi_i)+\tilde{b}_n\right)\right)\\
&+\quad P\left(\frac{Cst}{n}\sum_{i=1}^n\frac{\|\chi(X_i,n)\|^2\mathbf{1}\{\phi_i\leq 2l\}}{(\rho_n+f_X(X_i))^2}>e_nl^{p+1}\varepsilon\left(E[\bar{M}_l(\phi_i)]+\tilde{b}_n\right)\right)\\
&\leq \quad \frac{Cst}{\varepsilon e_n(E[\bar{M}_l(\phi_i)]+\tilde{b}_n)}E[\bar{M}_l(\phi_i)\frac{\chi_j(X_i,n)\mathbf{1}\{\phi_i\leq 2l\}}{\rho_n+f_X(X_i)}]\\
&+\quad \frac{Cst}{\varepsilon l^{p+1}e_n(E[\bar{M}_l(\phi_i)]+\tilde{b}_n)}E[\frac{\|\chi(X_i,n)\|^2\mathbf{1}\{\phi_i\leq 2l\}}{(\rho_n+f_X(X_i))^2}]\\
&\leq \quad \frac{Cst}{\varepsilon\rho_ne_n(E[\bar{M}_l(\phi_i)]+\tilde{b}_n)}E[\bar{M}_l(\phi_i)\mathbf{1}\{f_X(X_i)\leq\rho_n^{1-\alpha},\phi_i\leq 2l\}\chi_j(X_i,n)]\\
&+\quad \frac{Cst}{\varepsilon\rho_n^{1-\alpha}e_n(E[\bar{M}_l(\phi_i)]+\tilde{b}_n)}E[\bar{M}_l(\phi_i)\mathbf{1}\{\phi_i\leq 2l\}\chi_j(X_i,n)]\\
&+\quad \frac{Cst}{\varepsilon\rho_n^2l^{p+1}e_n(E[\bar{M}_l(\phi_i)]+\tilde{b}_n)}E[\mathbf{1}\{f_X(X_i)\leq\rho_n^{1-\alpha},\phi_i\leq 2l\}\|\chi(X_i,n)\|^2]\\
&+\quad \frac{Cst}{\varepsilon\rho_n^{2(1-\alpha)}l^{p+1}e_n(E[\bar{M}_l(\phi_i)]+\tilde{b}_n)}E[\mathbf{1}\{\phi_i\leq 2l\}\|\chi(X_i,n)\|^2].
\end{aligned}$$

Under assumption $R.3$, it is possible to improve the latter upper bound. Indeed, in this case, Eq. (7.44) provides

$$\sup_i\|\chi(X_i,n)\|\mathbf{1}\{\phi_i\leq 2l\}\leq Cst.\mathbf{1}\{\phi_i\leq 2l\}(h_2^2+h_1+u_n+l\rho_n+lv_n)\equiv g_n\mathbf{1}\{\phi_i\leq 2l\}.$$

Thus, it is easily to verify that T_2 tends to zero in probability under these sufficient conditions :

Assumption Q.4. For some $\alpha>0$,

$$E[\bar{M}_l(\phi_i)\mathbf{1}\{f_X(X_i)\leq\rho_n^{1-\alpha},\phi_i\leq 2l\}]+\rho_n^{\alpha}E[\bar{M}_l(\phi_i)\mathbf{1}\{\phi_i\leq 2l\}]=o\left(\frac{\rho_ne_n}{g_n}E[\bar{M}_l(\phi_i)]\right),$$

$$\left(E[\mathbf{1}\{f_X(X_i)\leq\rho_n^{1-\alpha},\phi_i\leq 2l\}]+\rho_n^{2\alpha}P(\phi_i\leq 2l)\right)=o\left(\frac{\rho_n^2e_n}{g_n^2}l^{p+1}E[\bar{M}_l(\phi_i)]\right).$$

Study of T_3. It is sufficient to show that

$$b_n=o_P\left(n^{-1}\sum_{i=1}^n\bar{M}_l(\phi_i)\right),\text{ and}$$

$$n^{-1}\sum_{i=1}^n\left(\bar{M}_l(\hat{\phi}_k)-\bar{M}_l(\phi_k)\right)=o_P\left(E[\bar{M}_l(\phi_k)]\right).$$

The first point is clearly satisfied under

Assumption Q.5. $E[\bar{M}_l(\phi_i)]/b_n$ tends to infinity with n, and $\ln n/(nl^p)^{1/2}$ is $O(b_n)$.

Moreover, with the same reasoning as previously, we get easily that the second term tends to zero in probability under

Assumption Q.6.

$$E\left[\frac{\|\chi(X_i,n)\|\mathbf{1}\{\phi_i \leq 2l\}}{(f_X(X_i)+\rho_n)}\right] = o(l^{p+1}E[\bar{M}_l(\phi_i)).$$

Study of T_4. This term can be dealt exactly like T_0. In fact, assumption Q.2 is sufficient so that T_4 tends to zero in probability, so the result.□.

ACKNOWLEDGMENT

I thank D. Bosq, E. Guerre and M.T. Wells for their comments. Remaining errors and imperfections are mine.

References

[1] J. H. Abbring and G. J. van den Berg, The non-parametric identification of treatment effects in duration models, mimeo (2000), available on the web.

[2] D. Bosq and J.-P. Lecoutre, "Théorie de l'estimation fonctionnelle", Economica, Paris, 1987.

[3] D. Clayton and J. Cuzick, Multivariate generalizations of the proportional hazard model, *J. Roy. Statist. Soc. Ser. A* **148** (1985), 82 − 117.

[4] D. R. Cox and D. Oakes, "Analysis of Survival Data", Chapman and Hall, New-York, 1984.

[5] M. Crowder, Identifiability Crises in Competing Risks, *Int. Statist. Rev.* **62** (1994), 379 − 391.

[6] J. Fan and I. Gijbels, "Local polynomial modelling and its applications", Chapman & Hall, 1996.

[7] J. Fan, I. Gijbels and M. King, Local likelihood and local partial likelihood in hazard estimation, *Ann. Statist.* **25** (1997), 1661 − 1690.

[8] J.-D. Fermanian, Multivariate hazard rates under random censorship, Working Paper CREST **9603**, 1996.

[9] J.-D. Fermanian, Multivariate hazard rates under random censorship, *J. Multivariate Anal.* **62** (1997), 273 − 309.

[10] C. J. Flinn and J. J. Heckman, Are unemployment and out of labor force behaviorally distinct labor force states ? *J. Labor Econ.* **1** (1983), 28 − 42.

[11] D. Gasbarra and S. R. Karia, Analysis of competing risks by using bayesian smoothing, *Scand. J. Statist.* **27** (2000), 605 − 617.

[12] G. C. Giannelli, Women's transitions in the labour market : A competing risks analysis on German Panel Data, *J. Popul. Econom.* **9** (1996), 287 − 300.

[13] L. Györfi, W. Härdle, P. Sarda and P. Vieu, "Nonparametric Curve Estimation from Time Series", Lecture Notes in Statistics 60 Springer Verlag, 1989.

[14] A. K. Han, Non-parametric analysis of a generalized regression model, *J. Econometrics* **35** (1987), 303 − 316.

[15] A. Han and J. A. Hausman, Flexible parametric estimation of duration and competing risk models, *J. Applied Econometrics* **5** (1990), 1 − 28.

[16] W. Härdle and T. M. Stoker, Investing smooth multiple regression by the method of average derivatives, *J. Amer. Statist. Assoc.* **84** (1989), 986 − 995.

[17] J. J. Heckman and B. Honoré, The identifiability of the competing risks model, *Biometrika* **76** (1989), 325 − 330.

[18] J. L. Horowitz and W. Härdle, Direct semiparametric estimation of single-index models with discrte covariates, *J. Amer. Statist. Assoc.* **91** (1996), 1632 − 1640.

[19] H. Ichimura, Semiparametric least squares (SLS) and weighted SLS estimation of single index models, *J. Econometrics* **58** (1993), 71 − 120.

[20] J. D. Kalbfleisch and R. L. Prentice, "The Statistical Analysis of Failure Time Data", Wiley, 1980.

[21] P. H. Kwan and H. Singh, On nonparametric estimation of the survival function with competing risks, *Scand. J. Statist.* **28** (2001), 751 − 724.

[22] N. Langberg, F. Proschan and A. J. Quinzi, Estimating dependent life lengths, with applications to the theory of competing risks, *Ann. Statist.* **9** (1981), 157 − 167.

[23] B. P. McCall, Unemployment insurance rules, joblessness, and part-time work, *Econometrica* **64** (1996), 647 − 682.

[24] Y. Omori, The identifiability of independent competing risks models with multiple spells, *Oxford Bull. Econom. Statist.* **60** (1998), 107 − 117.

[25] D. Pollard, "Convergence of Stochastic Processes", Springer, 1984.

[26] J. L. Powell, J. H. Stock and T. M. Stoker, Semiparametric estimation of index coefficients, *Econometrica* **57** (1989), 474 − 523.

[27] R. L. Prentice and J. D. Kalbfleisch, The analysis of failure times in the presence of competing risks, *Biometrics* **34** (1978), 541 − 554.

[28] F. O'Sullivan, Nonparametric estimation in the Cox model, *Ann. Statist.* **21** (1993), 124−145.

[29] R. P. Sherman, The limiting distribution of the maximum rank correlation estimator, *Econometrica* **61** (1993), 123 − 137.

[30] B. W. Silverman, "Density Estimation for Statistics and Data Analysis", Chapman & Hall, 1986.

[31] E. Slud, Nonparametric identifiability of marginal survival distributions in the presence of dependent competing risks and a prognostic covariate, in "Survival analysis : State of the Art", J-P. Klein and P. K. Goel (eds), Kluwer (1992), 355 − 368.

[32] G. T. Sueyoshi, Semiparametric proportional hazards estimation of competing risks model with time-varying covariates, *J. Econometrics* **51** (1992), 25 − 58.

[33] A. Tsiatis, A nonidentifiability aspect of the problem of competing risks, *Proc. Nat. Acad. Sci. USA* **72** (1975), 20 − 22.

[34] M. Zheng and J. Klein, Identifiability and estimation of marginal survival functions for dependent competing risks assuming the copula is known, in "Lifetime Data : Models in Reliability and Survival Analysis", N. Jewell et al. (eds), Kluwer (1996), 401 − 408

Table A : Estimation of the functions ϕ_1 and ϕ_2.

x_1	x_2	$\phi_1(x)$	$\hat{\phi}_1(x)$	$\sigma\hat{\phi}_1(x)$	$\phi_2(x)$	$\hat{\phi}_2(x)$	$\sigma\hat{\phi}_2(x)$
−1.0	−0.8	0.165	0.213	0.139	0.818	1.061	0.382
−1.0	−0.3	0.272	0.358	0.160	0.496	0.674	0.224
−1.0	0.2	0.449	0.591	0.247	0.301	0.417	0.191
−1.0	0.7	0.740	0.922	0.283	0.182	0.232	0.141
−0.5	−0.8	0.272	0.300	0.155	1.349	1.524	0.395
−0.5	−0.3	0.449	0.509	0.165	0.818	0.975	0.269
−0.5	0.2	0.740	0.844	0.232	0.496	0.609	0.194
−0.5	0.7	1.221	1.369	0.369	0.301	0.355	0.189
0.0	−0.8	0.449	0.406	0.220	2.225	2.169	0.462
0.0	−0.3	0.740	0.733	0.143	1.349	1.363	0.195
0.0	0.2	1.221	1.215	0.142	0.818	0.802	0.109
0.0	0.7	2.013	1.901	0.431	0.496	0.451	0.186
0.5	−0.8	0.740	0.477	0.214	3.669	2.882	0.617
0.5	−0.3	1.221	0.933	0.225	2.225	1.889	0.386
0.5	0.2	2.013	1.671	0.388	1.349	1.086	0.297
0.5	0.7	3.320	2.514	0.558	0.818	0.554	0.214
1.0	−0.8	1.221	0.553	0.334	6.049	3.584	0.796
1.0	−0.3	2.013	1.140	0.394	3.669	2.475	0.581
1.0	0.2	3.320	2.096	0.528	2.225	1.415	0.405
1.0	0.7	5.473	3.198	0.734	1.349	0.639	0.293

Table B : Estimation of the functions Z : $Z_1(t) = Z_2(t) = t$.

t	$\tilde{Z}_1(t)$	$\sigma\tilde{Z}_1(t)$	$\tilde{Z}_2(t)$	$\sigma\tilde{Z}_2(t)$	$\hat{Z}_1(t)$	$\sigma\hat{Z}_1(t)$	$\hat{Z}_2(t)$	$\sigma\hat{Z}_2(t)$
0.5	0.09	0.13	0.09	0.13	0.27	0.17	0.28	0.16
0.8	0.60	0.18	0.61	0.18	0.73	0.13	0.73	0.13
1.1	1.18	0.13	1.18	0.13	1.11	0.11	1.12	0.08
1.4	1.67	0.27	1.67	0.24	1.45	0.23	1.45	0.20
1.7	2.09	0.32	2.10	0.31	1.74	0.46	1.73	0.29
2.0	2.47	0.38	2.48	0.36	1.99	0.52	1.98	0.43
2.3	2.80	0.41	2.83	0.41	2.20	0.68	2.21	0.54
2.6	3.10	0.44	3.12	0.44	2.38	0.73	2.39	0.63
2.9	3.37	0.45	3.39	0.47	2.57	1.15	2.53	0.66
3.2	3.62	0.46	3.64	0.51	2.70	1.19	2.68	0.75
3.5	3.84	0.49	3.87	0.52	2.84	1.44	2.81	0.83
3.8	4.04	0.52	4.07	0.53	2.94	1.48	2.92	0.86
4.1	4.22	0.53	4.25	0.55	3.03	1.49	3.02	0.91
4.4	4.38	0.56	4.42	0.56	3.11	1.60	3.11	0.93
4.7	4.54	0.57	4.57	0.58	3.19	1.62	3.19	0.97
5.0	4.67	0.58	4.72	0.60	3.27	1.67	3.27	1.00

Table C : Estimation of the function H.

y_1	y_2	$H(y_1, y_2)$	$\tilde{H}(y_1, y_2)$	$\sigma\tilde{H}(y_1, y_2)$	$\hat{H}(y_1, y_2)$	$\sigma\hat{H}(y_1, y_2)$
0.2	0.1	0.071	0.057	0.026	0.003	0.008
0.2	0.3	0.136	0.133	0.042	0.035	0.023
0.2	0.5	0.167	0.164	0.044	0.105	0.040
0.2	0.7	0.184	0.183	0.051	0.168	0.052
0.2	0.9	0.196	0.200	0.067	0.202	0.085
0.4	0.1	0.087	0.072	0.028	0.014	0.012
0.4	0.3	0.207	0.198	0.490	0.172	0.044
0.4	0.5	0.286	0.280	0.058	0.357	0.061
0.4	0.7	0.341	0.334	0.059	0.480	0.083
0.4	0.9	0.383	0.381	0.079	0.549	0.157
0.6	0.1	0.094	0.081	0.031	0.031	0.020
0.6	0.3	0.250	0.246	0.052	0.296	0.062
0.6	0.5	0.375	0.365	0.063	0.525	0.081
0.6	0.7	0.477	0.467	0.069	0.667	0.106
0.6	0.9	0.562	0.551	0.083	0.730	0.185
0.8	0.1	0.098	0.090	0.038	0.042	0.027
0.8	0.3	0.279	0.278	0.064	0.385	0.094
0.8	0.5	0.444	0.438	0.072	0.637	0.129
0.8	0.7	0.596	0.582	0.072	0.768	0.159
0.8	0.9	0.735	0.708	0.084	0.774	0.298