Nonparametric estimation of the conditional survival function with double smoothing

Rebeca Peláez Suárez[a,*], Ricardo Cao Abad[b], Juan M. Vilar Fernández[b]

[a] *Research Group MODES, Department of Mathematics, CITIC, University of A Coruña, A Coruña, Spain*

[b] *Research Group MODES, Department of Mathematics, CITIC, University of A Coruña and ITMATI, A Coruña, Spain*

Abstract

In this paper a conditional survival function estimator for censored data is proposed. It is based on a double smoothing technique: both the covariate and the variable of interest (usually, the time) are smoothed. Asymptotic expressions for the bias and the variance and the asymptotic normality of the smoothed survival estimator derived from Beran's estimator are found. A simulation study shows the performance of some doubly smoothed estimators of the conditional survival function and compares them with the smoothed ones only in the covariate. The influence of the two smoothing parameters involved in the proposed estimators is also studied.

Keywords: Censored data, Conditional survival function, Survival analysis, Nonparametric estimator

---

1. Introduction

Let T be the time until the occurrence of an event and let X be a covariate related to the time. For instance, the survival time of an individual involved in a clinical study and his physiological features or the time until a debtor goes into default and his credit scoring in a credit risk context. In several cases the time variable, T, is subject to random right censoring. In this scenario it is of

*Corresponding author

Email address: `rebeca.pelaez@udc.es` (Rebeca Peláez Suárez)

Preprint submitted to ... *November 23, 2020*

interest to estimate the distribution function of T conditional to $X = x$, that is, $F(t|x) = P(T \leq t|X = x)$ or, equivalently, to estimate the conditional survival function $S(t|x) = 1 - F(t|x)$.

The most commonly used nonparametric estimator of $F(t|x)$ was introduced by [1]. This estimator turns out to be the Kaplan-Meier estimator (see [2]) in absence of covariates. Asymptotic properties of this estimator have been widely studied in the literature [3], [4], [5] and [6], among others. Another nonparametric estimator of the conditional distribution function with censored data is proposed by [7], [8]. It presents a better behaviour than Beran's estimator when estimating the ditribution function in the right tail with heavy censoring. In [9] and [10] an alternative estimator based on the local linear method proposed in [11] is studied. All these nonparametric distribution estimators are based on a covariate smoothing. Here, a double smoothing both in the covariate and in the time variable is proposed. Although this idea can be applied to any nonparametric estimator of the distribution (or survival) function, this paper focuses on Beran's estimator with double smoothing.

Survival analysis and, in particular, survival function estimation have traditionally been used for medical problems. However, in recent years they have also been successfully applied to risk problems, especially credit risk. See, among others, the works of [12], [13], [14] and [15].

The remainder of this paper is organized as follows. In Section 2, the nonparametric estimator of the conditional survival function with double smoothing is defined. Asymptotic properties of the nonparametric estimator with double smoothing associated with Beran's estimator [1] are presented in Section 3. In Section 4 a simulation study shows the improvement obtained by using the double smoothing in several nonparametric estimators of the conditional survival function for censored data. Section 5 contains some concluding remarks. Finally, in Section 6 sketches of the proofs of the theoretical results presented in Section 3 are shown.

2. Doubly smoothed conditional survival estimator

Let $\{(X_i, Z_i, \delta_i)\}_{i=1}^n$ be a simple random sample of (X, Z, δ) with X being the covariate, $Z = \min\{T, C\}$ the observed variable and $\delta = I_{T \leq C}$ the uncensoring indicator. Usually, T is the time until the occurrence of an event and C is the censoring time. The distribution function of T is denoted by $F(t)$ and its survival function by $S(t)$. The functions $F(t|x)$ and $S(t|x)$ are the conditional distribution and survival functions of T given $X = x$ evaluated at t. Let $\widehat{S}_h(t|x)$ be a nonparametric estimator of the conditional survival function with $h = h_n$ being the smoothing parameter for the covariate. Then, the expression of the smoothed survival estimator is as follows:

$$\widetilde{S}_{h,g}(t|x) \quad = \quad 1 - \sum_{i=1}^{n} s_{(i)} \mathbb{K}\left(\frac{t - Z_{(i)}}{g} \right) \tag{1}$$

where $s_{(i)} = \widehat{S}_h(Z_{(i-1)}|x) - \widehat{S}_h(Z_{(i)}|x)$ with $Z_{(i)}$ the i-th element of the sorted sample of Z, $\mathbb{K}(t)$ is the distribution function of a kernel K, $\mathbb{K}(t) = \int_{-\infty}^{t} K(u)du$, and $g = g_n$ is the smoothing parameter for the time variable.

This survival estimator is not only smoothed in the covariate but also in the time variable. It is based on the idea of estimating the survival function in a point t conditional to x by means of a weighted mean of the values that the estimator $\widehat{S}_h(t|x)$ takes in points near t so that a smoothed estimation is obtained.

According to the nonparametric estimator $\widehat{S}_h(t|x)$ used in (1), the corresponding doubly smoothed estimator of $S(t|x)$ is obtained. Therefore, the proposed estimator, $\widetilde{S}_{h,g}(t|x)$, is very general. In this paper, the study focuses on Beran's estimator given by

$$\widehat{S}_h^B(t|x) \quad = \quad \prod_{i=1}^{n} \left(1 - \frac{I_{\{Z_i \leq t,\ \delta_i = 1\}} w_{n,i}(x)}{1 - \sum_{j=1}^{n} I_{\{Z_j < Z_i\}} w_{n,j}(x)} \right) \tag{2}$$

where

$$w_{n,i}(x) = \frac{K\big((x - X_i)/h\big)}{\sum_{j=1}^{n} K\big((x - X_j)/h\big)}$$

with $i = 1, ..., n$ and $h = h_n$ is the smoothing parameter for the covariable. The smoothed survival function estimator based on Beran's estimator, $\widetilde{S}^B_{h,g}(t|x)$, is obtained by replacing $\widehat{S}_h(t|x)$ with $\widehat{S}^B_h(t|x)$ in (1).

The idea of a time variable smoothing is used in [16] to propose a smoothed Kaplan-Meier estimator and the doubly smoothed Beran's estimator, $\widetilde{S}^B_{h,g}(t|x)$, was considered in [17]. In these two papers the behaviour of the smoothed estimator is compared with the original one by simulation, but the asymptotic properties of the estimator are not studied. This issue will be addressed in the next section.

3. Asymptotic results of the smoothed Beran's estimator

Asymptotic properties of the smoothed Beran's estimator of the conditional survival function are studied here. The following assumptions will be required in the results.

A.1. X, T, C are absolutely continuous random variables.

A.2. The density function of X, m, has support $[0,1]$.

A.3. Let $H(t|x)$ be the conditional distribution function of $Z|X=x$,

(a) Let $I=[x_1,x_2]$ be an interval contained in the support of m such that,

$$0<\gamma=\inf\{m(x):x\in I_c\}<\sup\{m(x):x\in I_c\}=\Gamma<\infty$$

for some $I_c=[x_1-c,x_2+c]$ with $c>0$ and $0<c\,\Gamma<1$.

(b) For any $x\in I$, the random variables $T|_{X=x}$ and $C|_{X=x}$ are independent.

(c) Denoting $a_{H(\cdot|x)}=\inf\{t/H(t|x)>0\}$ and $b_{H(\cdot|x)}=\inf\{t/H(t|x)=1\}$, for any $x\in I_c$, $0\leq a_{H(\cdot|x)}$, $0\leq b_{H(\cdot|x)}$

(d) There exist $a,b,\theta\in\mathbb{R}$ with $a<b$, satisfying $\inf\{1-H(b|x):x\in I_c\}\geq\theta>0$. Therefore $1-H(t|x)\geq\theta>0$ for every $(t,x)\in[a,b]\times I_c$.

A.4. The first and second derivatives of m, $m'(x)$ and $m''(x)$, respectivaly, exist and are continuous in I_c.

A.5. Let $H_1(t|x)=P(T\leq t,\delta=1|X=x)$ be the conditional subdistribution function when $\delta=1$. The corresponding density functions of $H(t)$ and $H_1(t)$ are bounded away from 0 in $[a,b]$.

A.6. The first and second derivatives with respect to t of the functions $H(t|x)$ and $H_1(t|x)$, i.e. $H'(t|x)$, $H_1'(t|x)$, $H''(t|x)$ and $H_1''(t|x)$, exist and are continuous in $[a,b] \times I_c$.

A.7. The second partial derivatives first with respect to x and second with respect to t of the functions $H(t|x)$ and $H_1(t|x)$, i.e. $\dot{H}'(t|x)$ and $\dot{H}_1'(t|x)$ respectively, exist and are continuous in $[a,b] \times I_c$.

A.8. The kernel, K, is a symmetric, continuous and differentiable density function with compact support $[-1,1]$.

These assumptions are standard in the literature and affordable in this context. They were previously required in [3] and [6]. Conditions A.2, A.3a, A.3b and A.4 are assumed in [3] to obtain exponential bounds for the tails of the distribution of $\widehat{S}_h^B(t|x)$ and, from them, to obtain the weak and strong convergence of this estimator. Assumptions given in A.3c and A.3d are necessary to estimate the tails of the distribution functions involved. Conditions A.5, A.6 and A.7 along with those imposed on the kernel function ensure asymptotic unbiasedness of $\widehat{S}_h^B(t|x)$.

The following notation will be used. Let $R : \mathbb{R} \longrightarrow \mathbb{R}$ be any function, the constants c_R and d_R are defined as follows

$$c_R = \int R(t)^2 dt, \quad d_R = \int t^2 R(t) dt.$$

In particular, one can consider the kernel K and its distribution function $\mathbb{K}$ to define these constants. In this case, Assumption A.8. guarantees that c_K and d_K are finite. Being that,

$$c_K = \int K(t)^2 dt \leq 2\|K\|_\infty^2 < \infty$$

$$d_K = \int t^2 K(t) dt \leq \int_{-1}^{1} t^2 \|K\|_\infty dt \leq \frac{2}{3}\|K\|_\infty < \infty$$

From A.8. it follows that $\mathbb{K}(u) = 1 - \mathbb{K}(-u)$. Therefore

$$\begin{aligned} c_{\mathbb{K}} &= \int_{-1}^{1} \mathbb{K}^2(u) du = \int_{-1}^{0} \big(1 - \mathbb{K}(-u)\big)^2 du + \int_0^1 \mathbb{K}^2(u) du \\ &= \int_0^1 \Big(1 - 2\mathbb{K}(u)\big(1 - \mathbb{K}(u)\big)\Big) du \leq 1 \end{aligned}$$

The following functions are also defined,

$$K_l(u) = u^l K(u), \quad \mathbb{K}_l(u) = \int_{-\infty}^{u} K_l(t)dt. \tag{3}$$

Given any function $f : \mathbb{R}^k \longrightarrow \mathbb{R}$, its first derivatives with respect to the first and second variables are denoted as follows:

$$f'(x_1, ..., x_k) = \frac{\partial f(x_1, ..., x_k)}{\partial x_1}, \quad \dot{f}(x_1, ..., x_k) = \frac{\partial f(x_1, ..., x_k)}{\partial x_2}$$

Correspondingly, the second derivatives with respect to the first or second variable are denoted by $f''(x_1, ..., x_k)$ and $\ddot{f}(x_1, ..., x_k)$. Finally, let $f * g$ be the convolution of any two functions f and g.

In [6], an almost sure representation is found for a generalized Beran's estimator of the conditional survival function when the data are subject to random left truncation and right censoring. Taking into account we do not consider truncation but only right censoring, an almost sure representation of Beran's estimator can be obtained from the results shown in [6].

Theorem 1 (Almost sure representation for Beran's estimator of the conditional survival function)**.** *Under assumptions A.1-A.8, if* $a < a_{H(\cdot|x)}$ *for any* $x \in I$*, then*

$$\widehat{S}_h^B(t|x) - S(t|x) = (1 - F(t|x)) \sum_{i=1}^{n} w_{n,i}(x)\xi(Z_i, \delta_i, t, x) + R_n(t|x)$$

for $t \in [a, b]$*,* $x \in I$*, where*

$$\xi(Z, \delta, t, x) = \frac{1_{\{Z \le t, \delta = 1\}}}{1 - H(Z|x)} - \int_0^t \frac{1_{\{u \le Z\}}}{\big(1 - H(u|x)\big)^2} dH_1(u|x)$$

and

$$\sup_{[a,b]\times I} |R_n(t|x)| = O\left(\frac{\ln n}{nh}\right)^{3/4} \quad a.s.$$

Theorem 1 is a direct consequence of Theorem 2(c) in [6] by just assuming a degenerated in zero distribution for the left truncation time variable. A similar result is obtained below for the smoothed Beran's estimator.

Theorem 2 (Almost sure representation for the smoothed Beran's estimator of the conditional survival function)**.** *Under assumptions A.1-A.8, if* $a < a_{H(\cdot|x)}$ *for any* $x \in I$*, then*

$$\widetilde{S}^{B}_{h,g}(t|x) - S(t|x) = \sum_{i=1}^{n} w_{n,i}(x)\eta(Z_i,\delta_i,t,x) - \frac{1}{2}d_K F''(t|x)g^2 + R^1_n(t|x) + R^2_n(t|x)$$

for $t \in [a', b']$*,* $x \in I$*, where* $a' = a + \varepsilon$*,* $b' = b - \varepsilon$ *for* $\varepsilon > 0$*,*

$$\eta(Z,\delta,t,x) = \int K(u)\big(1 - F(t - gu|x)\big)\xi(Z,\delta,t-gu,x)du,$$

$$\sup_{(t,x)\in[a',b']\times I} \left|R^1_n(t|x)\right| = O\left(\frac{\ln n}{nh}\right)^{3/4} \quad a.s.,$$

and

$$R^2_n(t|x) = o(g^2).$$

Applying Theorem 2, the asymptotic bias and covariance of the smoothed Beran's estimator of the conditional survival function are obtained. Firstly, the smoothed Beran's estimator $\widetilde{S}^{B}_{h,g}(t|x)$ is split into two terms: one dominant term and some insignificant summands. This is shown in Lemma 1.

Lemma 1. *Under the assumptions of Theorem 2, the smoothed Beran's estimator* $\widetilde{S}^{B}_{h,g}(t|x)$ *can be split into the following terms*

$$\widetilde{S}^{B}_{h,g}(t|x) = \widetilde{S}^{AB}_{h,g}(t|x) + R^1_n(t|x) + R^2_n(t|x) + R^3_n(t|x)$$

where

$$\widetilde{S}^{AB}_{h,g}(t|x) = S(t|x) + \sum_{i=1}^{n} w^A_{n,i}(x)\eta(Z_i,\delta_i,t,x) - \frac{1}{2}d_K F''(t|x)g^2,$$

with

$$w^A_{n,i}(x) = \frac{1}{nh}\frac{K\big((x - X_i)/h\big)}{m(x)}$$

for all $i = 1, ..., n$*,* $R^1_n(t|x)$ *and* $R^2_n(t|x)$ *are the terms in Theorem 2 and*

$$R^3_n(t|x) = O_p\left(h^2 + \frac{1}{\sqrt{nh}}\right)\sum_{i=1}^{n} w^A_{n,i}(x)\eta(Z_i,\delta_i,t,x).$$

Theorem 3 (Bias and covariance of $\widetilde{S}^{AB}_{h,g}(t|x)$)**.** *Under the assumptions of Theorem 2, the asymptotic expressions for the bias and the covariance of* $\widetilde{S}^{B}_{h,g}(t|x)$ *are the following:*

$$
\begin{aligned}
Bias\big(\widetilde{S}^{AB}_{h,g}(t|x)\big) &= \frac{d_K}{2m(x)}\Big(2\Phi'_\eta(x,t,x)m'(x)+\Phi''_\eta(x,t,x)m(x)\Big)h^2 \\
&\quad -\frac{1}{2}d_K F''(t|x)g^2+o(h^2),
\end{aligned}
$$

$$
\begin{aligned}
Cov\big(\widetilde{S}^{AB}_{h,g}(t_1|x),\widetilde{S}^{AB}_{h,g}(t_2|x)\big) &= \frac{c_K}{m(x)}V_1(x,t_1,t_2)\frac{1}{nh}+\frac{c_K}{m(x)}V_2(x,t_1,t_2)\frac{g}{nh} \\
&\quad +\frac{c_K}{m(x)}V_3(x,t_1,t_2)\frac{g^2}{nh}+\frac{d_{K^2}}{m^2(x)}V_4(x,t_1,t_2)\frac{h}{n} \\
&\quad +o\left(\frac{g^2}{nh}+\frac{h}{n}\right).
\end{aligned}
$$

where

$$
\begin{aligned}
&\Phi_\eta(u,t,x)=\textstyle\int K(v)\big(1-F(t-gv|x)\big)\Phi_\xi(u,t-gv,x)dv,\\
&\Phi_\xi(u,t,x)=E\big[\xi(Z_1,\delta_1,t,x)|X_1=u\big] \text{ with } \xi(u,t,x) \text{ defined in Theorem 1,}\\
&V_1(x,t_1,t_2)=2J(t_1|x)\big(1-F(t_2|x)\big)\mathbb{K}*K\left(\frac{t_2-t_1}{g}\right),\\
&V_2(x,t_1,t_2)=2J(t_1|x)f(t_2|x)\mathbb{K}*K_1\left(\frac{t_1-t_2}{g}\right)\\
&\qquad\qquad +2J'(t_1|x)\big(1-F(t_2|x)\big)\mathbb{K}*K_1\left(\frac{t_2-t_1}{g}\right),\\
&V_3(x,t_1,t_2)=J''(t_1|x)\big(1-F(t_2|x)\big)\mathbb{K}*K_2\left(\frac{t_2-t_1}{g}\right)\\
&\qquad\qquad -J(t_1|x)f'(t_2|x)\left(d_K-\mathbb{K}*K_2\left(\frac{t_1-t_2}{g}\right)\right)\\
&\qquad\qquad +2J'(t_1|x)f(t_2|x)\mathbb{K}_1*K_1\left(\frac{t_2-t_1}{g}\right),\\
&V_4(x,t_1,t_2)=m(x)\big(1-F(t_1|x)\big)\big(1-F(t_2|x)\big)\Phi'_\xi(x,t_1,x)\Phi'_\xi(x,t_2,x)\\
&\qquad\qquad +\frac{1}{2}D''(x,t_1,t_2,x),\\
&D(u,t_1,t_2,x)=Cov\big[\eta(Z_1,\delta_1,t_1,x),\eta(Z_1,\delta_1,t_2,x)|X_1=u\big]m(u),\\
&J(t|x)=\big(1-F(t|x)\big)L(t|x),\\
&L(t|x)=\int_0^t\frac{dH_1(z|x)}{\big(1-H(z|x)\big)^2}.
\end{aligned}
$$

Finally, the asymptotic distribution of the smoothed Beran's estimator of the conditional survival function is obtained.

Theorem 4 (Limit distribution of $\widetilde{S}^{B}_{h,g}(t|x)$)**.** *Under the assumptions of Theorem 2, the limit distribution of $\widetilde{S}^{B}_{h,g}(t|x)$ is given by*

$$\frac{\widetilde{S}^{B}_{h,g}(t|x) - S(t|x) - \mu_n}{\sigma_n} \xrightarrow{d} N(0,1)$$

where $\mu_n = Bias\big(\widetilde{S}^{AB}_{h,g}(t|x)\big)$ given in Theorem 3 and $\sigma_n^2 = Var\big(\widetilde{S}^{AB}_{h,g}(t|x)\big)$ with

$$\begin{aligned}
Var\big(\widetilde{S}^{AB}_{h,g}(t|x)\big) = \quad & \frac{c_K}{m(x)}\big(1-F(t|x)\big)^2 L(t|x)\frac{1}{nh} \\
& + \frac{c_K(c_{\mathbb{K}}-1)}{m(x)}\big(1-F(t|x)\big)^2 L'(t|x)\frac{g}{nh} \\
& - \frac{c_K}{m(x)}\left(d_K\big(1-F(t|x)\big)f(t|x)L(t|x) - \left(\frac{1}{2}-\mu_1(\mathbb{K}^2)\right)L''(t|x)\right. \\
& \left. -2\big(\mu_1(\mathbb{K}^2)-1\big)\big(1-F(t|x)\big)f(t|x)L'(t|x)\right)\frac{g^2}{nh} \\
& + \frac{d_{K^2}}{m^2(x)}\left(\big(1-F(t|x)\big)^2\big(\Phi'_{\xi}(x,t,x)\big)^2 + \frac{1}{2}D''(x,t,t,x)\right)\frac{h}{n} \\
& + o\left(\frac{g^2}{nh}+\frac{h}{n}\right).
\end{aligned}$$

The asymptotic properties of Beran's estimator for the conditional survival function were proven in both [3] and [6]. It is worth noting that the asymptotic bias of Beran's estimator and the smoothed Beran's estimator have the same order as long as g is negligible with respect to h, i.e., $g = o(h)$. On the other hand, assuming $h \to 0$ and $g \to 0$, the asymptotic variance of Beran's estimator and the smoothed Beran's estimator have the same order since the terms g/nh and h/n are negligible compared to $1/nh$.

3.1. Bandwidth ratio

In this section, a discussion about the smoothing parameters of the smoothed Beran's survival estimator takes place in order to find the asymptotic optimal bandwidths defined as those that minimize the mean square error (MSE).

Considering only the dominant terms of the bias and the variance of the asymptotic estimator $\tilde{S}^{AB}_{h,g}(t|x)$ from the expressions given in Theorems 3 and 4, it follows that

$$\mathrm{Var}\big(\tilde{S}^{AB}_{h,g}(t|x)\big) = c_1\frac{1}{nh} - c_2\frac{g}{nh} + c_3\frac{h}{n} + o\left(\frac{g}{nh}\right),$$

$$\text{Bias}\big(\tilde{S}^{AB}_{h,g}(t|x)\big) = c_4 h^2 + c_5 g^2 + o(h^2) + o(g^2),$$

where the constants c_1, c_2, c_3, c_4 and c_5 are defined by

$$\begin{aligned}
c_1 &= c_K \frac{\big(1-F(t|x)\big)^2 L(t|x)}{m(x)} > 0,\\
c_2 &= c_K(1-c_{\mathbb{K}})\frac{\big(1-F(t|x)\big)^2 L'(t|x)}{m(x)} > 0,\\
c_3 &= d_{K^2}\left(\frac{1}{m(x)}\big(1-F(t|x)\big)^2\big(\Phi'_\xi(x,t,x)\big)^2 + \frac{1}{2m^2(x)}D''(x,t,t,x)\right),\\
c_4 &= \frac{d_K}{2m(x)}\Big(2\Phi'_\eta(x,t,x)m'(x) + \Phi''_\eta(x,t,x)m(x)\Big),\\
c_5 &= \frac{1}{2}d_K F''(t|x).
\end{aligned}$$

Then, the asymptotic bandwidths that minimize the dominant terms of the MSE can be obtained by minizing the formula:

$$\Psi(h,g) = c_1\frac{1}{nh} - c_2\frac{g}{nh} + c_3\frac{h}{n} + c_4^2h^4 + c_5^2g^4 + 2c_4c_5h^2g^2.$$

In order to obtain the asymptotically optimal bandwidths, it is necessary to consider the partial derivatives of Ψ with respect to both h and g, equal them to zero and distinguish three different cases depending on the relative asymptotic behaviour of h and g. The partial derivative of Ψ with respect to h is

$$\frac{\partial\Psi}{\partial h} = -c_1\frac{1}{nh^2} + c_2\frac{g}{nh^2} + c_3\frac{1}{n} + 4c_4^2h^3 + 4c_4c_5hg^2,$$

but, the terms $c_2\frac{g}{nh^2}$ and $c_3\frac{1}{n}$ are negligible with respect to the term $\frac{c_1}{nh^2}$. Similarly,

$$\frac{\partial\Psi}{\partial g} = -c_2\frac{1}{nh} + 4c_5^2g^3 + 4c_4c_5h^2g.$$

Therefore, the equations to be taken into account are the following ones

$$-c_1\frac{1}{nh^2} + 4c_4^2h^3 + 4c_4c_5hg^2 = 0, \tag{4}$$

$$-c_2\frac{1}{nh} + 4c_5^2g^3 + 4c_4c_5h^2g = 0. \tag{5}$$

There are three possible cases for the asymptotic behaviour of $\frac{g}{h}$.

Case 1. $g = o(h)$

Equations equivalent to (4) and (5) in this case are

$$-c_1 \frac{1}{nh^2} + 4c_4^2 h^3 = 0,$$
$$-c_2 \frac{1}{nh} + 4c_4 c_5 h^2 g = 0.$$

Then, the optimal bandwidths are $h_{opt} = c_0 n^{-1/5}$ and $g_{opt} = d_0 n^{-2/5}$ with $c_0 = \left(\frac{c_1}{4c_4^2}\right)^{1/5}$ and $d_0 = \frac{c_2 c_4^{1/5}}{4^{2/5} c_1^{3/5} c_5}$. In this case,

$$\Psi(h_{opt}, g_{opt}) = \left(\frac{c_1}{c_0} + c_4^2 c_0^4\right) n^{-4/5} + \left(c_3 c_0 + 2c_4 c_5 c_0^2 d_0^2 - \frac{c_2 d_0}{c_0}\right) n^{-6/5} + c_5^2 d_0^4 n^{-8/5}.$$

Case 2. $h = o(g)$

When $h = o(g)$, asymptotically equivalent versions of Equations (4) and (5) are

$$-\frac{c_1}{nh^2} + 4c_4 c_5 h g^2 = 0,$$
$$-\frac{c_2}{nh} + 4c_5^2 g^3 = 0.$$

and the solution of this system is $h_{opt} = e_0 n^{-1/7}$ and $g_{opt} = \left(\frac{c_2}{4c_5^2 e_0}\right)^{1/3} n^{-2/7}$ with $e_0 = \left(\frac{c_1 (4c_5^2)^{2/3}}{4c_4 c_5 c_2^{2/3}}\right)^{3/7}$. So, $g_{opt} = o(h_{opt})$ which contradicts the initial hypothesis. Case 2 is discarded.

Case 3. $\lim_{h\to\infty} \frac{h}{g} = \alpha > 0$, $\alpha \in \mathbb{R}$.

In this case, $\frac{h}{g} = \alpha$ asymptotically and the expression for Ψ becomes

$$\Psi(h, g) \quad = \quad \frac{c_1}{n\alpha g} - \frac{c_2}{n\alpha} + c_6 g^4$$

with $c_6 = c_4^2 \alpha^4 + c_5^2 + 2c_4 c_5 \alpha^2 = (c_4 \alpha^2 + c_5)^2$.

The option $c_6 = 0$ is discarded because it leads to an optimal bandwidth g which does not tend to zero. Therefore, $c_6 = (c_4\alpha^2 + c_5)^2 > 0$ and the minimun is reached at $h_{opt} = \alpha l_0 n^{-1/5}$ and $g_{opt} = l_0 n^{-1/5}$ with $l_0 = \left(\frac{c_1}{4c_6\alpha}\right)^{1/5}$. This means that the minimal value of Ψ is attained at $\alpha = \infty$ which contradicts Case 3.

From the arguments above it follows that $g = o(h)$ is the only feasible case, obtaining the corresponding optimal bandwidths for the estimator $\widetilde{S}_{h,g}^{AB}(t|x)$.

4. Simulation study

A simulation study was conducted in order to compare the performance of the proposed estimator of the conditional survival function. The study considers two models, one with Weibull lifetime and censoring time distributions and another one with exponential distributions.

For Model 1, the covariate X follows a $U(0,1)$ distribution. The time to occurrence of the event conditional to the covariate, $T|_{X=x}$, follows a Weibull distribution with parameters d and $A(x)^{-1/d}$ where $d=2$ and $A(x)=1+5x$, and the censoring time conditional to the covariate, $C|_{X=x}$, follows a Weibull distribution with parameters d and $B(x)^{-1/d}$ where $d=2$ and $B(x)=10+b_1x+20x^2$. In this case, the conditional survival function and the censoring conditional probability are given by:

$$S(t|x) = e^{-A(x)t^d},$$

$$P(\delta=0|X=x) = \frac{B(x)}{A(x)+B(x)}.$$

Having set the value of the covariate, $x=0.6$ the value of b_1 is chosen so that the censoring conditional probability is 0.2, 0.5 and 0.8. These values are $b_1=-27$, $b_1=-22$ and $b_1=-2$, respectively. The conditional survival function for this model is estimated in a time grid of size n_t, $0<t_1<\cdots<t_{n_t}$, where $t_{n_t}+b=F^{-1}(0.95|x)$ for the value of the covariable $x=0.6$.

Model 2 considers a $U(0,1)$ distribution for X. The time to occurrence of the event conditional to the covariate, $T|_{X=x}$, follows an exponential distribution with parameter $\Gamma(x)=2+58x-160x^2+107x^3$, and the censoring time conditional to the covariate, $C|_{X=x}$, follows an exponential distribution with parameter $\Delta(x)=10+d_1x+20x^2$. In this scenario, the conditional survival function and the censoring conditional probability are the following:

$$S(t|x) = e^{-\Gamma(x)t},$$

$$P(\delta=0|X=x) = \frac{\Delta(x)}{\Gamma(x)+\Delta(x)}.$$

Having set the value of the covariate, $x = 0.8$, the value of b_1 is chosen so that the censoring conditional probability is 0.2, 0.5 and 0.8. These values are $d_1 = -113/4$, $d_1 = -55/2$ and $d_1 = -123/5$, respectively. The conditional survival function is estimated in a time grid of size n_t, $0 < t_1 < \cdots < t_{n_t}$, where $t_{n_t} + b = F^{-1}(0.95|x)$ for the value of the covariable $x = 0.8$.

It can be proved that Model 1 is close to a proportional hazards model, while Model 2 moves away from this parametric model. These two models were used in the simulation study by [15]. The standard Gaussian kernel truncated in the range $[-50, 50]$ is used for both the covariate and the time variable smoothing. The sample size is $n = 400$, and the size of the lifetime grid is $n_t = 100$. In addition, the boundary effect is corrected using the reflexion principle proposed in [18].

The optimal bandwidth for $\widehat{S}_h^B(t|x)$, h_1, is taken as the value which minimizes a Monte Carlo approximation of the MISE given by

$$MISE_x(h) = E\left(\int \left(\widehat{S}_h^B(t|x) - S(t|x)\right)^2 dt\right)$$

based on $N = 100$ simulated samples. The value of $MISE$ using this smoothing parameter is approximated from $N = 1000$ simulated samples and used, along with its square root ($RMISE$), as a measure of the estimation error which is committed by $\widehat{S}_h^B(t|x)$.

The smoothed survival estimator $\widetilde{S}_{h,g}^B(t|x)$ depends on two bandwidths. Two strategies are be used in order to obtain these smoothing parameters.

Strategy 1. It consists in fixing the covariate smoothing parameter to the optimal one, h_1, and approximating the time variable smoothing parameter. The error to minimize is

$$MISE_x(h_1, g) = E\left(\int \left(\widetilde{S}_{h_1,g}^B(t|x) - S(t|x)\right)^2 dt\right)$$

considered as a function of the bandwidth g. It is approximated from $N = 100$ simulated samples in a grid of 50 values of g and the bandwidth which provides the smaller error is chosen as g_1. Then, $N = 1000$ samples are simulated to approximate $MISE_x(h_1, g_1)$ which is the measure of the estimation error of $\widetilde{S}^B_{h,g}(t|x)$. The main advantage of using this strategy is its relatively low computational cost.

Strategy 2. The optimal bandwidth (h_2, g_2) is chosen (from a meshgrid of 50 values of h and 50 values of g) as the pair which minimizes some Monte Carlo approximations of

$$MISE_x(h,g) = E\left(\int \left(\widetilde{S}^B_{h,g}(t|x) - S(t|x)\right)^2 dt\right)$$

based on $N = 100$ simulated samples. Then, the value of the $MISE$ committed by $\widetilde{S}^B_{h_2,g_2}(t|x)$ is approximated from $N = 1000$ simulated samples.

Neither the bandwidth obtained with Strategy 1 nor Strategy 2 can be used in practice but their choice produces a fair comparison since the estimators are built using the best possible smoothing parameters.

Figure 1 shows the function $MISE_x(h_1, g)$ for each level of censoring conditional probability and each model. These graphs show the error curve to minimize in order to obtain the optimal time smoothing parameter. It follows from this that the optimal bandwidth g is easily approximated by Strategy 1.

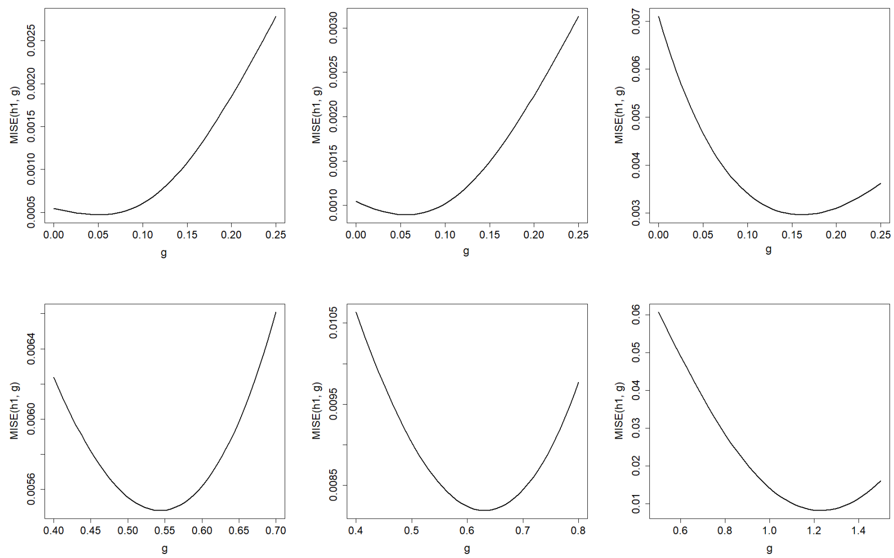

Figure 1: $MISE_x(h_1, g)$ function approximated via Monte Carlo for the smoothed Beran's estimator using $N = 100$ simulated samples from Model 1 (top) and Model 2 (bottom) with $P(\delta = 0|x) = 0.2$ (left), $P(\delta = 0|x) = 0.5$ (center) and $P(\delta = 0|x) = 0.8$ (right).

Figure 2 shows the function $MISE_x(h, g)$ of Model 1 and Model 2 for the lowest and highest censoring level. These graphs show the two-dimensional functions to be minimized in Strategy 2. The red zone is where this minimum is reached and its coordinates provide the optimal smoothing bandwidths.

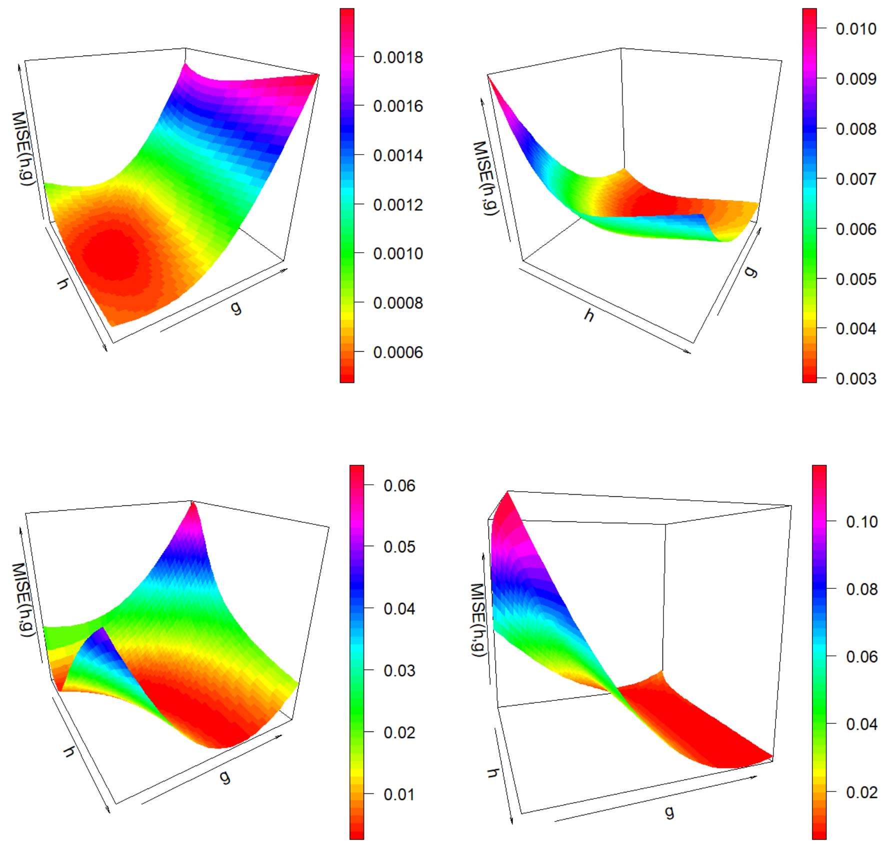

Figure 2: $MISE_x(h,g)$ function approximated via Monte Carlo for the smoothed Beran's estimator using $N = 100$ simulated samples from Model 1 (top) and Model 2 (bottom) with $P(\delta = 0|x) = 0.2$ (left) and $P(\delta = 0|x) = 0.5$ (right).

The graphs for both strategies show that the smoothing parameters can be well approximated and bandwidths slightly larger or smaller than the true optimum value do not greatly affect the estimation error, so the different estimators can be reasonably compared.

It is clear that the magnitude of the estimation error is notably affected by the choice of the time smoothing bandwidth (g). However, for a fixed value of h, the value of g for which the smallest error is committed does not seem to vary too much depending on the value of the covariate smoothing bandwidth

(h). This can be seen in Figure 3. There, $MISE_x(h, g)$ is shown as a function of g for some fixed values of h within the interval where the optimum is reached. The obtained curves are similar and close for all the h values and every model.

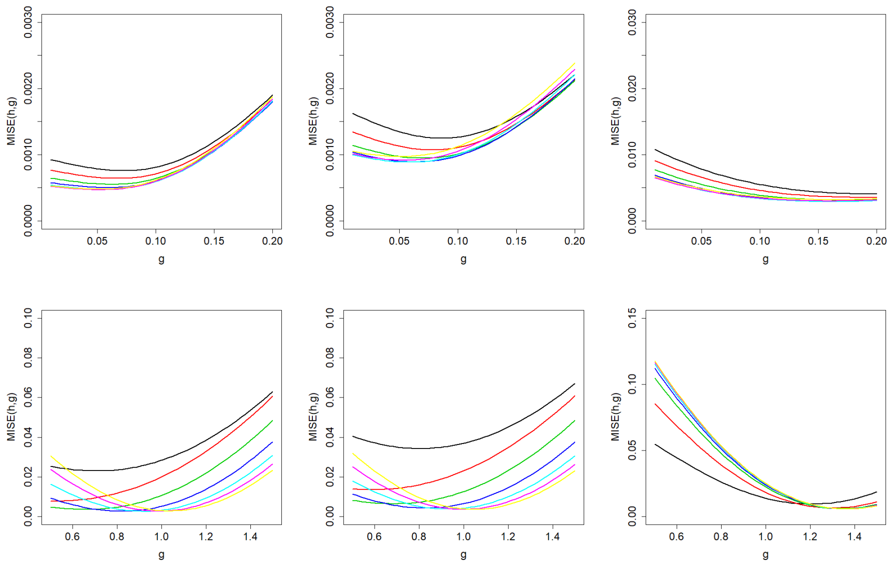

Figure 3: $MISE_x(h, g)$ as a function of g approximated via Monte Carlo for the smoothed Beran's estimator using $N = 100$ simulated samples for some fixed equispaced values of $h \in [0.1, 0.4]$ for Model 1 (top) and $h \in [0.01, 0.18]$ for Model 2 (bottom) with $P(\delta = 0|x) = 0.2$ (left), $P(\delta = 0|x) = 0.5$ (center) and $P(\delta = 0|x) = 0.8$ (right).

Table 1 shows the optimal bandwidths and the estimation errors that are committed by Beran's estimator and the smoothed Beran's estimator with both Strategies 1 and 2 for each model. In order to compare the behaviour of the estimators and quantify the improvement of the smoothing over the original estimator, the ratio R_i is defined

$$R_i(t|x) \quad = \quad \frac{RMISE\big(\widetilde{S}^B_{h_i,g_i}(t|x)\big)}{RMISE\big(\widehat{S}^B_{h_1}(t|x)\big)}$$

with $i = 1, 2$ depending on the smoothing strategy used. The closer to 0 the value of R_1 or R_2, the greater the improvement with respect to Beran's estimator. The relation between R_1 and R_2 (R_1 greater than R_2 or viceversa) also informs which of the two strategies reduces the error most.

In all cases, $RMISE$ values are lower for the smoothed Beran's estimator than for Beran's estimator and this difference becomes bigger when increasing the censoring conditional probability. Moreover, the values of R_1 and R_2 satisfy $0 < R_2 < R_1 < 1$ in all cases, so the estimator $\widetilde{S}^B_{h,g}(t|x)$ with optimal bandwidth (h_2, g_2) (Strategy 2) provides the most accurate estimation.

When the censoring conditional probability is 0.2 or 0.5, the time smoothing, with either Strategy 1 or 2, reduces the error by about 8% in Model 1, this improvement is about 40% when the probability of conditional censoring is 0.8. The error reduction in Model 2 with respect to the nonsmoothed survival estimator is more significant, reaching 30% and 70% when censoring is moderate or heavy, respectively.

		Model 1			**Model 2**		
$\mathbf{P(\delta = 0 \mid x)}$		**0.2**	**0.5**	**0.8**	**0.2**	**0.5**	**0.8**
$\widehat{\mathbf{S}}^{\mathbf{B}}_{\mathbf{h_1}}$	$\mathbf{h_1}$	0.25918	0.22857	0.23469	0.04490	0.05265	0.12837
	RMISE	0.02304	0.03186	0.08641	0.11112	0.14644	0.28914
$\widetilde{\mathbf{S}}^{\mathbf{B}}_{\mathbf{h_1},\mathbf{g_1}}$	$\mathbf{h_1}$	0.25918	0.22857	0.23469	0.04490	0.05265	0.12837
	$\mathbf{g_1}$	0.05110	0.05620	0.16330	0.54082	0.62857	1.23469
	RMISE	0.02144	0.02943	0.05185	0.07817	0.10091	0.08886
	$\mathbf{R_1}$	0.93055	0.92373	0.60005	0.70347	0.68909	0.30733
$\widetilde{\mathbf{S}}^{\mathbf{B}}_{\mathbf{h_2},\mathbf{g_2}}$	$\mathbf{h_2}$	0.24082	0.20408	0.20408	0.11061	0.16980	2
	$\mathbf{g_2}$	0.05265	0.06041	0.16510	0.88776	1.03061	1.35714
	RMISE	0.02129	0.02907	0.05067	0.05248	0.06379	0.07550
	$\mathbf{R_2}$	0.92405	0.91243	0.58639	0.47228	0.43561	0.26112

Table 1: Optimal bandwidths, $RMISE$, R_1 and R_2 of the survival estimation for Beran's estimator, the smoothed Beran's estimator with Strategy 1 and the smoothed Beran's estimator with Strategy 2 in each level of censoring conditional probability for Models 1 and 2.

Table 2 shows the computation time (in seconds) of Beran's estimator and smoothed Beran's estimator when estimating the conditional survival curve in a

100-point time grid and a fixed value of x for different values of the sample size in Model 1. The smoothing parameters are fixed to the optimal values before estimating the curve.

n	50	100	200	400	1200
Beran	0.01	0.01	0.01	0.02	0.02
SBeran	0.01	0.01	0.02	0.03	0.07

Table 2: CPU time (in seconds) for estimating $S(t|x)$ in a time grid of size 100 for each estimator and different sample sizes (n).

Time variable smoothing results in an increase of the CPU time. The smoothed Beran's estimator is the most affected by the increase of the sample size and its CPU times are higher than those of the nonsmoothed estimator.

It is also interesting to compare the computational efficiency of the two strategies used to find the optimal bandwidths, since Strategy 1 seems to be faster but Strategy 2 provides smaller estimation errors. Table 3 shows the CPU time (in minutes) for each strategy and several number of trials.

Strategy 1 consists in looking for the optimal bandwidth for $\widehat{S}_h^B(t|x)$ as the value which minimizes $MISE_x(h)$ out of $n_h = 50$ possible values. It is called h_1 and it is fixed when lookig for the optimal bandwidth for $\widetilde{S}_{h_1,g}^B(t|x)$ out of $n_g = 50$ possible values, g_1. Strategy 2 cosists in obtanining the pair (h_2, g_2) as the value that minimizes $MISE_x(h, g)$ for $\widetilde{S}_{h,g}^B(t|x)$ from a two-dimensional grid of size $n_h \times n_g = 50 \times 50$.

In both strategies the sample size is $n = 400$ and the conditional survival function is estimated in a time grid of size $n_t = 100$. The number of simulated samples (N) used to approximate the $MISE$ by Monte Carlo is the parameter that varies to compare the time each strategy takes to obtain the optimal bandwidths. The results clearly show the computational advantage of using Strategy 1, since Strategy 2 is significantly slower.

N	50	100	150	200
Strategy 1	1.12	1.95	2.99	3.86
Strategy 2	37.58	79.83	117.79	159.99

Table 3: CPU time (in minutes) for approximating the optimal bandwidth (h,g) for $\widetilde{S}^B_{h,g}(t|x)$ with Strategies 1 and 2 and different numbers of trials (N).

The advantages or disadvantages of the smoothed Beran's survival estimator have been discussed. Now, it is interesting to compare the behaviour of other survival estimators with their smoothed versions. In a second simulation study, three nonparametric estimators of the conditional survival function are considered: Beran's estimator, the Weighted Nadaraya-Watson estimator (WNW) and the Van Keilegom-Akritas estimator (VKA). The WNW estimator was built following the idea of [11], where the survival estimator is based on a local lineal regression. Since the weighted local lineal estimator presents problems when estimating probabilities, a constant fit is proposed in [15]. The VKA estimator was defined in [7]. Their expressions are shown in Section 2 of [15] and they are denoted by $\widehat{S}^{WNW}_h(t|x)$ and $\widehat{S}^{VKA}_h(t|x)$. Their smoothed versions (SWNW and SVKA) are built according to Equation (1), obtaining the following smoothed survival estimators: $\widetilde{S}^{WNW}_{h,g}(t|x)$ and $\widetilde{S}^{VKA}_{h,g}(t|x)$.

Since the computational cost of these estimators is pretty high, only Strategy 1 is used to look for the optimal smoothing parameters. Strategy 2 would further increase the computation time of the simulations.

Table 4 shows the values of the optimal smoothing parameters and the error committed by each estimator. In order to quantify the improvement that the time smoothing brings to the survival estimation and compare the performance of the three estimators, the ratios $R^\bullet_S$ and $R^\bullet_c$ are defined as follows and included in Table 4:

$$R^\bullet_S(t|x) = \frac{RMISE\big(\widetilde{S}^\bullet_{h_1,g_1}(t|x)\big)}{RMISE\big(\widehat{S}^\bullet_{h_1}(t|x)\big)}$$

$$R^\bullet_c(t|x) = \frac{RMISE\big(\widetilde{S}^\bullet_{h_1,g_1}(t|x)\big)}{RMISE\big(\widehat{S}^B_{h_2,g_2}(t|x)\big)}$$

being $\bullet = B, WNW, VKA$.

The values of $R_S^\bullet$ report the influence of the smoothing. The smaller the value, the better the estimation obtained with the doubly smoothed estimator compared to the corresponding single smoothed estimator. Since its value is less than 1 in all cases of Models 1 and 2, the time smoothing seems to improve the estimators. Moreover, the further away from 1 the value of $R_S^\bullet$, the greater the improvement. In general, the estimator whose error is reduced the most is the WNW estimator, especially when censoring is heavy.

The value of $R_c^\bullet$ is used to compare the behaviour of the three estimators with the behaviour of the smoothed Beran's estimator (Strategy 2). Almost all the $R_c^\bullet$ values obtained are greater than 1. Therefore, the smoothed Beran's estimator with Strategy 2 provides more accurate survival estimations. The smoothed Beran's estimator with Strategy 1 is the second best option for estimating the survival function, since its $R_c^\bullet$ values are the closest to 1. When using Strategy 1, the smoothed WNW estimator appears to be competitive with Beran's estimator.

	$\mathbf{P(\delta=0\|x)=0.2}$			$\mathbf{P(\delta=0\|x)=0.5}$			$\mathbf{P(\delta=0\|x)=0.8}$		
	SBeran	**SWNW**	**SVKA**	**SBeran**	**SWNW**	**SVKA**	**SBeran**	**SWNW**	**SVKA**
Model 1									
$\mathbf{h_1}$	0.2592	1.5000	0.2592	0.2286	0.3755	0.2286	0.2347	1.5000	0.2347
$\mathbf{g_1}$	0.0511	0.0751	0.0294	0.0562	0.1402	0.0896	0.1633	0.4530	0.3225
RMISE	0.0214	0.0196	0.0489	0.0294	0.0350	0.0857	0.0519	0.0653	0.1025
$\mathbf{R_S^\bullet}$	0.9306	0.8602	0.9817	0.9237	0.8064	0.9666	0.6001	0.4507	0.5737
$\mathbf{R_c^\bullet}$	1.0066	0.9188	2.2944	1.0124	1.2037	2.9481	1.0233	1.2895	2.0229
Model 2									
$\mathbf{h_1}$	0.0449	0.0410	0.0457	0.0527	0.0449	0.0538	0.1284	0.0737	0.1284
$\mathbf{g_1}$	0.5408	0.0193	0.1994	0.6286	1.1531	0.5469	1.2347	1.4123	1.2000
RMISE	0.0782	0.0840	0.4171	0.1009	0.0883	0.2062	0.0889	0.0756	0.0558
$\mathbf{R_S^\bullet}$	0.7035	0.5757	0.8858	0.6891	0.3401	0.4427	0.3073	0.1512	0.1104
$\mathbf{R_c^\bullet}$	1.4895	1.6014	7.9482	1.5819	1.3839	3.2319	1.1770	1.0014	0.7393

Table 4: Optimal bandwidths, $RMISE$, $R_S^\bullet$ and $R_c^\bullet$ of the survival estimation for the smoothed Beran's estimator, the smoothed WNW estimator and the smoothed VKA estimator with Strategy 1 in each level of censoring conditional probability for Models 1 and 2.

The computation time of these techniques should be considered in the comparison. Table 5 shows the CPU times (in seconds) that each of the estimators spends in estimating the conditional survival function in a 100-point time's grid and a fixed value of x for different values of the sample size in Model 1.

Time variable smoothing always results in an increase of the CPU time since smoothed versions of Beran's, the WNW and the VKA estimators have higher CPU times than Beran's estimator. The smoothed Beran's estimator is barely affected by the increase of the sample size and it is the fastest of the three smoothed estimators. The following one is the smoothed VKA, although its CPU time increases very fast with the sample size. The slowest and most affected by the sample size increase is the smoothed WNW estimator.

n	50	100	200	400	1200
Beran	0.01	0.01	0.01	0.02	0.02
SBeran	0.01	0.01	0.02	0.03	0.07
SWNW	1.5	3.15	10.80	47.40	710.13
SVKA	0.40	1.65	7.33	36.56	608.90

Table 5: CPU time (in seconds) for estimating $S(t|x)$ in a time grid of size 100 for every estimator and different sample sizes (n).

5. Conclusions

A general doubly smoothed estimator of the conditional survival function is proposed in this paper. Asymptotic properties of the smoothed survival estimator based on Beran's estimator are proved. The asymptotic expressions of the bias and the variance of this estimator are complex but the rate of convergence is equal to that of Beran's estimator.

The first simulation study carried out shows that the estimation error is reduced when using the smoothed Beran's estimator of the conditional survival function, most notably with heavy censoring. However, the time variable smoothing increases the computation time, with Strategy 2 being much less efficient than Strategy 1. The second simulation study revealed that the double smoothing also implies an improvement of other nonparametric survival estimators. As a general conclusion, the good behaviour of the doubly smoothed Beran's estimator in the different analysed contexts is remarkable.

This work could be extended to the case of having a multidimensional covariate $\mathbf{x} = (x_1, ..., x_q)$ where each x_i is a feature of the individual. In the credit risk context, all these features are usually sumarized in the covariate called credit scoring.

6. Proofs

Proofs of the results shown in Section 3 are complex. In this section a sketch of them is given. The following lemma will be used in the proofs.

Lemma 2 (Integration by parts formula for Riemann-Stiltjes integral with a piecewise-defined function)**.** *Let* $u : [0, L] \longrightarrow \mathbb{R}$ *be a differentiable function in* $[0, L]$ *and let* $v : [0, L] \longrightarrow \mathbb{R}$ *be a nondecreasing piecewise function, i.e.,*

$$v(x) = \sum_{j=1}^{k-1} b_j 1_{[a_{j-1}, a_j)}(x) + b_k 1_{[a_{k-1}, a_k]}(x)$$

where $0 = a_0 < a_1 < \cdots < a_k = L$ *and* $b_i \in \mathbb{R}$ *for all* $i = 1, ..., k$*,* $b_1 < b_2 < \cdots < b_k$*. Then,*

$$\int_0^L u(x)v(dx) = \Big[u(x)v(x)\Big]_{x=0}^{x=L} - \int_0^L u'(x)v(x)dx.$$

Proof of Lemma 2.

On the one hand,

$$\int_0^L u(x)v(dx) = \sum_{i=1}^{k-1} u(a_i)\big(v(a_i) - v(a_{i-1})\big) = \sum_{i=1}^{k-1} u(a_i)\big(b_{i+1} - b_i\big) \qquad (6)$$

On the other hand,

$$
\begin{aligned}
&\Big[u(x)v(x)\Big]_{x=0}^{x=L} - \int_0^L u'(x)v(x)dx = \\
&= \quad u(L)v(L) - u(0)v(0) - \sum_{j=1}^{k} \int_{a_{j-1}}^{a_j} u'(x)v(x)dx \\
&= \quad u(L)v(L) - u(0)v(0) - \sum_{j=1}^{k} b_j \big(u(a_j) - u(a_{j-1})\big) \\
&= \quad u(L)v(L) - u(0)v(0) + b_1 u(a_0) - b_k u(a_k) + \sum_{j=1}^{k-1} (b_{j+1} - b_j) u(a_j)
\end{aligned}
$$

Since $a_0 = 0$, $a_k = L$, $v(a_0) = b_1$ and $v(a_k) = b_k$, we have

$$
\Big[u(x)v(x)\Big]_{x=0}^{x=L} - \int_0^L u'(x)v(x)dx \quad = \quad \sum_{j=1}^{k-1} (b_{j+1} - b_j) u(a_j) \tag{7}
$$

Now, using (6) and (7), the lemma is proved.

□

Proof of Theorem 2.

Denoting $\widetilde{F}^B_{h,g}(t|x) = 1 - \widetilde{S}^B_{h,g}(t|x)$ and $\widehat{F}^B_h(dt|x) = 1 - \widehat{S}^B_h(dt|x)$, standard algebra gives

$$
\widetilde{F}^B_{h,g}(t|x) - F(t|x) \quad = \quad \int \mathbb{K}\left(\frac{t-u}{g}\right) \widehat{F}^B_h(du|x) - F(t|x) = A_1 + A_2, \tag{8}
$$

where

$$
A_1 = \int \mathbb{K}\left(\frac{t-u}{g}\right) \big(\widehat{F}^B_h(du|x) - F(du|x)\big)
$$

and

$$
A_2 = \int \mathbb{K}\left(\frac{t-u}{g}\right) F(du|x) - F(t|x).
$$

Using Lemma 2 and Theorem 1, it is obtained

$$
\begin{aligned}
A_1 \quad &= \quad \int K(u) \big(\widehat{F}^B_h(t - gu|x) - F(t - gu|x)\big) du \\
&= \quad \int K(u) \Big(-\big(1 - F(t - gu|x)\big) \sum_{i=1}^{n} w_{n,i}(x) \xi(Z_i, \delta_i, t - gu, x) \\
&\qquad + R_n(t - gu|x) \Big) du \\
&= \quad A_{11} + A_{12},
\end{aligned} \tag{9}
$$

where $A_{11} = -\int K(u)\big(1 - F(t-gu|x)\big)\sum_{i=1}^{n} w_{n,i}(x)\xi(Z_i, \delta_i, t-gu, x)du$ and $A_{12} = \int K(u)R_n(t-gu|x)du$.

First considering A_{11} in (9),

$$\begin{aligned} A_{11} &= -\int K(u)\big(1 - F(t-gu|x)\big)\sum_{i=1}^{n} w_{n,i}(x)\xi(Z_i, \delta_i, t-gu, x)du \\ &= -\sum_{i=1}^{n} w_{n,i}(x)\int K(u)\big(1 - F(t-gu|x)\big)\xi(Z_i, \delta_i, t-gu, x)du, \end{aligned}$$

and denoting $\eta(Z,\delta,t,x) := \int K(u)\big(1 - F(t-gu|x)\big)\xi(Z,\delta,t-gu,x)du$, it is obtained

$$A_{11} = -\sum_{i=1}^{n} w_{n,i}(x)\eta(Z_i, \delta_i, t, x). \tag{10}$$

Considering A_{12} in (9) and using A.8, it follows that

$$|A_{12}| \le \int_{-1}^{1} K(u)\Big|R_n(t-gu|x)\Big|du \le \sup_{z\in[t-g,t+g]}\Big|R_n(z|x)\Big|.$$

Fix $\varepsilon > 0$ and define $a' = a + \varepsilon$, $b' = b - \varepsilon$. Then,

$$\sup_{(t,x)\in[a',b']\times I} |A_{12}| \le \sup_{(t,x)\in[a',b']\times I}\left\{\sup_{z\in[t-g,t+g]}\Big|R_n(z|x)\Big|\right\} \tag{11}$$

On the one hand, there exists $n_0 \in \mathbb{N}$ such that $g = g_n \le \varepsilon$ for all $n \ge n_0$. So, $z \in [t-g, t+g]$ implies that $|z-t| \le g \le \varepsilon$ and equivalently, $t-\varepsilon \le z \le t+\varepsilon$.

On the other hand, $t \in [a', b']$ implies that $a + \varepsilon = a' \le t \le b' = b - \varepsilon$.

Therefore,

$$z \le t + \varepsilon \le (b-\varepsilon) + \varepsilon = b \Rightarrow z \le b$$

and also,

$$z \ge t - \varepsilon \ge (a+\varepsilon) - \varepsilon = a \Rightarrow z \ge a.$$

Hence, $z \in [a, b]$ and $x \in I$. So, for $t \in [a', b']$,

$$\sup_{z\in[t-g,t+g]}\Big|R_n(z|x)\Big| \le \sup_{(t',x')\in[a,b]\times I}\Big|R_n(t'|x')\Big|. \tag{12}$$

Recalling the inequality obtained in (11) and applying the inequality in (12), one has

$$\sup_{(t,x)\in[a',b']\times I} |A_{12}| \le \sup_{(t,x)\in[a',b']\times I}\left\{\sup_{(t',x')\in[a,b]\times I}\Big|R_n(t'|x')\Big|\right\}$$

and from Theorem 1,

$$\sup_{(t',x')\in[a,b]\times I}\left|R_n(t'|x')\right| = O\left(\frac{\ln n}{nh}\right)^{3/4} \quad \text{a. s.}$$

Finally, defining $R_n^1(t|x) = A_{12}$, the following is obtained

$$\sup_{(t,x)\in[a',b']\times I}\left|R_n^1(t|x)\right| = \sup_{(t,x)\in[a',b']\times I}|A_{12}| = O\left(\frac{\ln n}{nh}\right)^{3/4} \quad \text{a. s.} \tag{13}$$

Now, considering A_2 in (8) and using Lemma 2, it follows that

$$\begin{aligned} A_2 &= \int_{-\infty}^{+\infty}\frac{1}{g}K\left(\frac{t-y}{g}\right)F(y|x)dy - F(t|x) \\ &= \int_{-\infty}^{+\infty}K(u)F(t-gu|x)du - F(t|x). \end{aligned}$$

Assuming that $g = g_n$ tends to zero when n tends to infinity, Taylor's formula for $F(t-gu|x)$ gives:

$$F(t-gu|x) = F(t|x) - guF'(t|x) + \frac{1}{2}(gu)^2F''(t|x) + o(g^2)$$

Then, using assumption A.8,

$$A_2 = \frac{1}{2}d_KF''(t|x)g^2 + R_n^2(t|x) \tag{14}$$

with $R_n^2(t|x) = o(g^2)$.

Finally, Equations (8) - (14) give

$$\widetilde{F}^B_{h,g}(t|x) - F(t|x) = -\sum_{i=1}^n w_{n,i}(x)\eta(Z_i,\delta_i,t,x) + \frac{1}{2}d_KF''(t|x)g^2 + R_n^1(t|x) + R_n^2(t|x)$$

where

$$\sup_{(t,x)\in[a',b']\times I}\left|R_n^1(t|x)\right| = O\left(\frac{\ln n}{nh}\right)^{3/4} \quad \text{a. s.}$$

which proves Theorem 2 since $\widetilde{S}^B_{h,g}(t|x) - S(t|x) = F(t|x) - \widetilde{F}^B_{h,g}(t|x)$.

□

Proof of Lemma 1.

Theorem 2 gives

$$\widetilde{S}^B_{h,g}(t|x) - S(t|x) = \sum_{i=1}^n w_{n,i}(x)\eta(Z_i,\delta_i,t,x) - \frac{1}{2}d_KF''(t|x)g^2 + R_n^1(t|x) + R_n^2(t|x) \quad a.s.,$$

where

$$w_{n,i}(x) = \frac{K\big((x - X_i)/h\big)}{\sum_{j=1}^n K\big((x - X_j)/h\big)}.$$

Note that

$$\widehat{m}_h(x) := \frac{1}{nh}\sum_{j=1}^n K\big((x - X_j)/h\big)$$

is the Parzen-Rosenblatt estimator of the density function of X, $m(x)$. Then,

$$\begin{aligned}
\sum_{i=1}^n w_{n,i}(x)\eta(Z_i, \delta_i, t, x) &= \sum_{i=1}^n \frac{1}{nh}\frac{K\big((x - X_i)/h\big)}{\widehat{m}_h(x)}\eta(Z_i, \delta_i, t, x) \\
&= \sum_{i=1}^n \frac{1}{nh}\frac{K\big((x - X_i)/h\big)}{m(x)}\eta(Z_i, \delta_i, t, x) \\
&\quad + \frac{m(x) - \widehat{m}_h(x)}{\widehat{m}_h(x)}\sum_{i=1}^n \frac{1}{nh}\frac{K\big((x - X_i)/h\big)}{m(x)}\eta(Z_i, \delta_i, t, x) \\
&= \sum_{i=1}^n w_{n,i}^A(x)\eta(Z_i, \delta_i, t, x) + R_n^3(t|x)
\end{aligned}$$

where

$$w_{n,i}^A(x) = \frac{1}{nh}\frac{K\big((x - X_i)/h\big)}{m(x)}$$

for all $i = 1, ..., n$ and

$$R_n^3(t|x) = \frac{m(x) - \widehat{m}_h(x)}{\widehat{m}_h(x)}\sum_{i=1}^n w_{n,i}^A(x)\eta(Z_i, \delta_i, t, x)$$

Since $\widehat{m}_h(x)$ is a consistent estimator of $m(x)$ and its bias and variance convergence rates are $O(h^2)$ and $O(1/nh)$, respectively (see [18]),

$$R_n^3(t|x) = O_p\left(h^2 + \frac{1}{\sqrt{nh}}\right)\sum_{i=1}^n w_{n,i}^A(x)\eta(Z_i, \delta_i, t, x).$$

□

Proof of Theorem 3.

Lemma 1 gives

$$\begin{aligned}
\widetilde{S}_{h,g}^{AB}(t|x) - S(t|x) &= \sum_{i=1}^n w_{n,i}^A(x)\eta(Z_i, \delta_i, t, x) - \frac{1}{2}d_K F''(t|x)g^2 \\
&= \frac{1}{m(x)n}\sum_{i=1}^n \varphi_{n,i}(t, x) - \frac{1}{2}d_K F''(t|x)g^2
\end{aligned}$$

where $\varphi_{n,i}(t,x) = \frac{1}{h}K\left(\frac{x-X_i}{h}\right)\eta(Z_i,\delta_i,t,x)$ are independent and identically distributed random variables for all $i = 1,...,n$. Consequently,

$$\text{Bias}\big(\widetilde{S}_{h,g}^{AB}(t|x)\big) \;=\; \frac{1}{m(x)}E\big(\varphi_{n,1}(t,x)\big) - \frac{1}{2}d_K F''(t|x)g^2 \qquad (15)$$

and

$$\text{Cov}\big(\widetilde{S}_{h,g}^{AB}(t_1|x), \widetilde{S}_{h,g}^{AB}(t_2|x)\big) = \frac{1}{m^2(x)n}\text{Cov}\big(\varphi_{n,1}(t_1,x), \varphi_{n,1}(t_2,x)\big). \qquad (16)$$

First, an expression for $E\big(\varphi_{n,1}(t,x)\big)$ is found:

$$\begin{aligned}
E\big(\varphi_{n,1}(t,x)\big) &= \frac{1}{h}E\left[K\left(\frac{x-X_1}{h}\right)E\big[\eta(Z_1,\delta_1,t,x)|X_1\big]\right] \\
&= \int K(v)\Phi_\eta(x-hv,t,x)m(x-hv)dv \\
&= \int K(v)\bigg(\Phi_\eta(x,t,x)m(x) - hv\frac{\partial\Phi_\eta(u,t,x)m(u)}{\partial u}\Big|_{u=x} \\
&\quad + \frac{h^2v^2}{2}\frac{\partial^2\Phi_\eta(u,t,x)m(u)}{\partial u^2}\Big|_{u=x}\bigg)dv + o(h^2) \\
&= \Phi_\eta(x,t,x)m(x) + \frac{h^2}{2}d_K\Big(\Phi_\eta(x,t,x)m''(x) \\
&\quad + 2\Phi'_\eta(x,t,x)m'(x) + \Phi''_\eta(x,t,x)m(x)\Big) + o(h^2)
\end{aligned}$$

Next, an explicit expression for Φ_η is obtained,

$$\Phi_\eta(u,t,x) = \int K(v)\big(1-F(t-gv|x)\big)\Phi_\xi(u,t-gv,x)dv$$

where $\Phi_\xi(u,t,x) = E\big[\xi(Z_1,\delta_1,t,x)|X_1 = u\big]$ can be written as follows:

$$\Phi_\xi(u,t,x) = \int_0^t \frac{dH_1(z|u)}{1-H(z|x)} - \int_0^t \frac{1-H(v|u)}{\big(1-H(v|x)\big)^2}dH_1(v|x).$$

Then, $\Phi_\xi(x,t,x) = 0$ for any x and t and, consequently, $\Phi_\eta(x,t,x) = 0$ for any x and t. Hence,

$$E\big(\varphi_{n,1}(t,x)\big) = \frac{d_K}{2}\Big(2\Phi'_\eta(x,t,x)m'(x) + \Phi''_\eta(x,t,x)m(x)\Big)h^2 + o(h^2),$$

and replacing it in (15) the bias of $\widetilde{S}_{h,g}^{AB}(t|x)$ is available.

In order to achieve the covariance of the estimator, an asymptotic expression for $\text{Cov}\big(\varphi_{n,1}(t_1,x), \varphi_{n,1}(t_2,x)\big)$ is obtained

$$\text{Cov}\big[\varphi_{n,1}(t_1,x), \varphi_{n,1}(t_2,x)\big] \;=\; C_{11} - C_{12} + C_2, \qquad (17)$$

where

$$C_{11} = E\left[\frac{1}{h^2}K^2\left(\frac{x-X_1}{h}\right)\Phi_\eta(X_1,t_1,x)\Phi_\eta(X_1,t_2,x)\right],$$

$$C_{12} = E\left[\frac{1}{h}K\left(\frac{x-X_1}{h}\right)\Phi_\eta(X_1,t_1,x)\right]E\left[\frac{1}{h}K\left(\frac{x-X_1}{h}\right)\Phi_\eta(X_1,t_2,x)\right]$$

and

$$C_2 = E\Big[\mathrm{Cov}\big[\varphi_{n,1}(t_1,x),\varphi_{n,1}(t_2,x)|X_1\big]\Big].$$

Asymptotic expressions for the terms involved in (17) are found. The first becomes

$$\begin{aligned} C_{11} &= \int \frac{1}{h^2}K^2\left(\frac{x-u}{h}\right)\Phi_\eta(u,t_1,x)\Phi_\eta(u,t_2,x)m(u)du \\ &= \int\int\int \frac{1}{h}K^2(w)K(v_1)K(v_2)\big(1-F(t_1-gv_1|x)\big)\big(1-F(t_2-gv_2|x)\big) \\ &\quad B(x-hw,t_1-gv_1,t_2-gv_2,x)dv_1dv_2dw \end{aligned}$$

where $B(u,z_1,z_2,x) := \Phi_\xi(u,z_1,x)\Phi_\xi(u,z_2,x)m(u)$.

By means of a Taylor expansion of $B(u,t_1-gv_1,t_2-gv_2,x)$ when $u = x-hw$ around $u = x$,

$$B(x-hw,t_1-gv_1,t_2-gv_2,x) = h^2w^2\Phi'_\xi(x,t_1-gv_1,x)\Phi'_\xi(x,t_2-gv_2,x)m(x)+O(h^3).$$

Thus,

$$\begin{aligned} C_{11} &= \int\int\int hw^2K^2(w)K(v_1)K(v_2)\big(1-F(t_1-gv_1|x)\big)\big(1-F(t_2-gv_2|x)\big) \\ &\quad \cdot\Phi'_\xi(x,t_1-gv_1,x)\Phi'_\xi(x,t_2-gv_2,x)m(x)dv_1dv_2dw+O(h^2) \end{aligned}$$

Now, using Taylor expansions of the functions involved when $z_1 = t_2 - gv_1$ and $z_2 = t_2 - gv_2$ around $z_1 = t_1$ and $z_2 = t_2$, respectively, leads to

$$\begin{aligned} C_{11} &= hd_{K^2}m(x)\big(1-F(t_1|x)\big)\big(1-F(t_2|x)\big)\Phi'_\xi(x,t_1,x)\Phi'_\xi(x,t_2,x) \\ &\quad +O(h^2)+O(hg^2). \end{aligned} \tag{18}$$

Let us denote $C(u,z,x) = \Phi_\xi(u,z,x)m(u)$. Then,

$$\begin{aligned} &E\left[\frac{1}{h}K\left(\frac{x-X_1}{h}\right)\Phi_\eta(X_1,t,x)\right] \\ &\quad = \frac{h^2}{2}d_K\int K(v)\big(1-F(t-gv|x)\big)\Big(\Phi''_\xi(x,t-gv,x)m(x) \\ &\qquad +2\Phi'_\xi(x,t-gv,x)m'(x)\Big)dv+O(h^3) = O(h^2) \end{aligned}$$

Hence,

$$C_{12} \quad = \quad O(h^4). \tag{19}$$

Now,

$$\begin{aligned} C_2 &= \int \frac{1}{h^2} K^2\left(\frac{x-z}{h}\right) \mathrm{Cov}\big[\eta(Z_1,\delta_1,t_1,x),\eta(Z_1,\delta_1,t_2,x)|X_1=z\big] m(z)dz \\ &= \frac{1}{h} c_K D(x,t_1,t_2,x) + h d_{K^2} D''(x,t_1,t_2,x) + O(h^2) \end{aligned} \tag{20}$$

where $D(u,t_1,t_2,x)$ is defined in the statement of Theorem 3. An expression for $D(x,t_1,t_2,x)$ is calculated. Since

$$E\big[\eta(Z_1,\delta_1,t,x)|X_1=x\big] = \Phi_\xi(x,t,x) = 0$$

and

$$\begin{aligned} & E\big[\eta(Z_1,\delta_1,t_1,x)\eta(Z_1,\delta_1,t_2,x)|X_1=x\big] \\ &\quad = \int\int K(v_1)K(v_2)\big(1-F(t_1-gv_1|x)\big)\big(1-F(t_2-gv_2|x)\big) \\ &\qquad E\big[\xi(Z_1,\delta_1,t_1-gv_1,x)\xi(Z_1,\delta_1,t_2-gv_2,x)|X_1=x\big]dv_1dv_2, \end{aligned}$$

it follows that

$$\begin{aligned} & D(x,t_1,t_2,x) \\ &\quad = E\big[\eta(Z_1,\delta_1,t_1,x)\eta(Z_1,\delta_1,t_2,x)|X_1=x\big]m(x) \\ &\quad = m(x)\int\int K(v_1)K(v_2)\big(1-F(t_1-gv_1|x)\big)\big(1-F(t_2-gv_2|x)\big) \\ &\qquad E\big[\xi(Z_1,\delta_1,t_1-gv_1,x)\xi(Z_1,\delta_1,t_2-gv_2,x)|X_1=x\big]dv_1dv_2. \end{aligned}$$

Long calculations lead to the following expression for $D(x,t_1,t_2,x)$:

$$\begin{aligned} D(x,t_1,t_2,x) &= m(x)V_1(t_1,t_2,x) + m(x)V_2(t_1,t_2,x)g \\ &\quad + m(x)V_3(t_1,t_2,x)g^2 + O(g^3) \end{aligned} \tag{21}$$

where $V_1(t_1,t_2,x)$, $V_2(t_1,t_2,x)$ and $V_3(t_1,t_2,x)$ are defined in Theorem 3.

By means of similar but more tedious calculations, omitted here, a general expression for $D(u,t_1,t_2,x)$ could be obtained. Thus, using expression (21) in

(20), gives:

$$\begin{aligned} C_2 &= c_K m(x) V_1(t_1,t_2,x)\frac{1}{h} + c_K m(x) V_2(t_1,t_2,x)\frac{g}{h} + \\ &\quad c_K m(x) V_3(t_1,t_2,x)\frac{g^2}{h} + \frac{d_{K^2}}{2} D''(x,t_1,t_2,x)h \\ &\quad + O\left(\frac{g^3}{h}\right) + O(h^2). \end{aligned} \tag{22}$$

Now, plugging (18), (19) and (22) in (17) gives

$$\begin{aligned} \mathrm{Cov}\big(\varphi_{n,1}(t_1,x),\varphi_{n,1}(t_2,x)\big) &= c_K m(x) V_1(t_1,t_2,x)\frac{1}{h} + c_K m(x) V_2(t_1,t_2,x)\frac{g}{h} \\ &\quad + c_K m(x) V_3(t_1,t_2,x)\frac{g^2}{h} + d_{K^2} V_4(t_1,t_2,x)h \\ &\quad + O(h^2) + O\left(\frac{g^3}{h}\right) + O(hg^2). \end{aligned}$$

where

$$\begin{aligned} V_4(t_1,t_2,x) &= m(x)\big(1-F(t_1|x)\big)\big(1-F(t_2|x)\big)\Phi'_\xi(x,t_1,x)\Phi'_\xi(x,t_2,x) \\ &\quad + \frac{1}{2}D''(x,t_1,t_2,x) \end{aligned}$$

Replacing the expression of $\mathrm{Cov}\big(\varphi_{n,1}(t_1,x),\varphi_{n,1}(t_2,x)\big)$ in (16) leads to

$$\begin{aligned} \mathrm{Cov}\big(\widetilde{S}^{AB}_{h,g}(t_1|x),\widetilde{S}^{AB}_{h,g}(t_2|x)\big) &= \frac{c_K}{m(x)} V_1(t_1,t_2,x)\frac{1}{nh} + \frac{c_K}{m(x)} V_2(t_1,t_2,x)\frac{g}{nh} \\ &\quad + \frac{c_K}{m(x)} V_3(t_1,t_2,x)\frac{g^2}{nh} + \frac{d_{K^2}}{m^2(x)} V_4(t_1,t_2,x)\frac{h}{n} \\ &\quad + O\left(\frac{h^2}{n}\right) + O\left(\frac{g^3}{nh}\right) + O\left(\frac{hg^2}{n}\right). \end{aligned}$$

and the covariance part of the theorem is proved.

□

Proof of Theorem 4.

From Theorem 3 an expression of the asymptotic variance of $\widetilde{S}^{AB}_{h,g}(t|x)$ is obtained:

$$\begin{aligned} \mathrm{Var}\big(\widetilde{S}^{AB}_{h,g}(t|x)\big) = &\ \frac{c_K}{m(x)} V_1(t,t,x)\frac{1}{nh} + \frac{c_K}{m(x)} V_2(t,t,x)\frac{g}{nh} + \frac{c_K}{m(x)} V_3(t,t,x)\frac{g^2}{nh} \\ & + \frac{d_{K^2}}{m^2(x)} V_4(t,t,x)\frac{h}{n} + o\left(\frac{g^2}{nh} + \frac{h}{n}\right). \end{aligned}$$

where

$$\begin{aligned}
V_1(t,t,x) &= 2\big(1-F(t|x)\big)^2 L(t|x)\mathbb{K}*K(0),\\
V_2(t,t,x) &= 2\big(1-F(t|x)\big)^2 L'(t|x)\mathbb{K}*K_1(0),\\
V_3(t,t,x) &= \big(-f'(t|x)L(t|x)\big(1-F(t|x)\big)\\
&\quad -2f(t|x)L'(t|x)\big(1-F(t|x)\big)\big)\mathbb{K}*K_2(0)\\
&\quad -\big(1-F(t|x)\big)L(t|x)f'(t|x)\big(d_K-\mathbb{K}*K_2(0)\big)\\
&\quad \big(-2f^2(t|x)L(t|x)+\big(1-F(t|x)\big)L'(t|x)f(t|x)\big)\mathbb{K}_1*K_1(0)\\
V_4(t,t,x) &= m(x)\big(1-F(t|x)\big)^2\big(\Phi'_\xi(x,t,x)\big)^2+\frac{1}{2}D''(x,t,t,x),
\end{aligned}$$

Definitions of $K_l(u)$ and $\mathbb{K}_l(u)$ in (3) and assumption A.8 give: $\mathbb{K}*K(0)=\dfrac{1}{2}$, $\mathbb{K}*K_1(0)=-\dfrac{1}{2}+\dfrac{1}{2}c_{\mathbb{K}}$, $\mathbb{K}*K_2(0)=\dfrac{1}{2}-\mu_1(\mathbb{K}^2)$, $\mathbb{K}_1*K_1(0)=0$. Therefore,

$$\begin{aligned}
V_1(t,t,x) &= \big(1-F(t|x)\big)^2 L(t|x),\\
V_2(t,t,x) &= (c_{\mathbb{K}}-1)\big(1-F(t|x)\big)^2 L'(t|x),\\
V_3(t,t,x) &= -d_K\big(1-F(t|x)\big)L(t|x)f'(t|x)\\
&\quad +\left(\frac{1}{2}-\mu_1(\mathbb{K}^2)\right)\big(1-F(t|x)\big)^2 L''(t|x)\\
&\quad +\big(2\mu_1(\mathbb{K}^2)-1\big)\big(1-F(t|x)\big)L'(t|x)f(t|x),
\end{aligned}$$

from which the expression of $\mathrm{Var}\big(\widetilde{S}^{AB}_{h,g}(t|x)\big)$ derives.

Finally, the asymptotic normality of $\tilde{S}^{B}_{h,g}(t|x)$ is proved. In the proof of Theorem 3 the estimator $\widetilde{S}^{B}_{h,g}(t|x)$ is splited up into the following terms:

$$\begin{aligned}
\widetilde{S}^{B}_{h,g}(t|x)-S(t|x) &= \frac{1}{m(x)n}\sum_{i=1}^{n}\varphi_{n,i}(t,x)-\frac{1}{2}d_K F''(t|x)g^2\\
&\quad +R^1_n(t|x)+R^2_n(t|x)+R^3_n(t|x) \quad a.s.,
\end{aligned}$$

where $\varphi_{n,i}(t,x)=\dfrac{1}{h}K\left(\dfrac{x-X_i}{h}\right)\eta(Z_i,\delta_i,t,x)$ are independent and identically distributed random variables for all $i=1,...,n$.

Since the supports of the functions K and m are compact and $F(t|x)$ is bounded, it is guaranteed that $\mathrm{Var}\big(\varphi_{n,i}(t,x)\big)<\infty$. On the other hand, it is clear that

$$\sigma_n^2=\frac{1}{m^2(x)n^2}\sum_{i=1}^{n}\mathrm{Var}\big(\varphi_{n,i}(t,x)\big)>0.$$

Therefore, if the Lindeberg's condition is satisfied, then

$$\frac{\sum_{i=1}^{n}\Big(\varphi_{n,i}(t,x)-E\big[\varphi_{n,i}(t,x)\big]\Big)}{\sigma_n}\xrightarrow{d} N(0,1).$$

The Lindeberg's condition requires that

$$\frac{1}{\sigma_n^2}\sum_{i=1}^{n}\int_{|\varphi_{n,i}(t,x)|>\varepsilon\sigma_n}\big(\varphi_{n,i}(t,x)-E\big[\varphi_{n,i}(t,x)\big]\big)^2 dP\longrightarrow 0\quad\forall\varepsilon>0.$$

Denoting $\mathbb{1}(A)$ as the indicator function of A the following indicator is defined

$$\mathbb{1}_{n,i}=\mathbb{1}\big(|\varphi_{n,i}(t,x)-E[\varphi_{n,i}(t,x)]|>\varepsilon\sigma_n\big)$$

$$=\mathbb{1}\left(\left(\frac{1}{h}K\left(\frac{x-X_i}{h}\right)\eta(Z_i,\delta_i,t,x)-E\left[\frac{1}{h}K\left(\frac{x-X_i}{h}\right)\eta(Z_i,\delta_i,t,x)\right]\right)^2>\varepsilon^2\sigma_n^2\right)$$

the Lindeberg's condition can be expressed as follows

$$\begin{aligned}&\frac{1}{\sigma_n^2}\sum_{i=1}^{n}\int_{\{|\varphi_{n,i}(t,x)-E[\varphi_{n,i}(t,x)]|>\varepsilon\sigma_n\}}\Big(\varphi_{n,i}(t,x)-E\big[\varphi_{n,i}(t,x)\big]\Big)^2 dP\\ &\quad=\quad\frac{1}{\sigma_n^2}E\left[\sum_{i=1}^{n}\Big(\varphi_{n,i}(t,x)-E\big[\varphi_{n,i}(t,x)\big]\Big)^2\mathbb{1}_{n,i}\right]=\frac{1}{\sigma_n^2}E\big(\varphi_n\big)\end{aligned}$$

By applying assumption A.3d, it is easy to prove that $\xi(Z,\delta,t,x)$, defined in Theorem 1, is bounded:

$$\begin{aligned}|\xi(Z,\delta,t,x)| &= \left|\frac{1_{\{Z\le t,\delta=1\}}}{1-H(Z|x)}-\int_0^t\frac{dH_1(u|x)}{\big(1-H(u|x)\big)^2}\right|\\ &\le \frac{1_{\{Z\le t,\delta=1\}}}{1-H(Z|x)}+\int_0^t\frac{dH_1(u|x)}{\big(1-H(u|x)\big)^2}\le\frac{1}{\theta}+\int_0^t\frac{dH_1(u|x)}{\theta^2}\\ &\le \frac{1}{\theta}+\frac{H(t|x)}{\theta^2}\le\frac{1}{\theta}+\frac{1}{\theta^2}\end{aligned}$$

and, consequently, η is also bounded:

$$\begin{aligned}|\eta(Z,\delta,t,x)| &\le \int K(u)\big(1-F(t-gu|x)\big)\left(\frac{1}{\theta}+\frac{1}{\theta^2}\right)du\\ &= \left(\frac{1}{\theta}+\frac{1}{\theta^2}\right)\left(\big(1-F(t|x)\big)+\frac{g^2}{2}d_K\big(1-F''(t|x)\big)\right)+O(g^2)\end{aligned}$$

Due to the fact that $(nh)^{-1}\longrightarrow 0$ and K and η are bounded,

$$\exists n_0\in\mathbb{N}/n\ge n_0\Rightarrow\mathbb{1}_{n,i}(w)=0,\ \forall w\text{ and }\forall i\in\{1,2,\dots,n\}$$

$$\Leftrightarrow\exists n_0\in\mathbb{N}/n\ge n_0\Rightarrow\varphi_n(w)=0,\forall w$$

$$\Leftrightarrow \exists n_0 \in \mathbb{N}/n \geq n_0 \Rightarrow E\big(\varphi_n\big) = 0$$

Hence, $\lim\limits_{n\to\infty} \dfrac{1}{\sigma_n^2} E\big(\varphi_n\big) = 0$ and the Lindeberg's condition is proved. As a consequence,

$$\frac{\sum_{i=1}^{n} \Big(\varphi_{n,i}(t,x) - E\big[\varphi_{n,i}(t,x)\big]\Big)}{\sigma_n} \xrightarrow{d} N(0,1)$$

where $\sigma_n^2 = \mathrm{Var}\big(\widetilde{S}_{h,g}^{AB}(t|x)\big)$. The asymptotic normality of the estimator then holds.

$\square$

Acknowledgements

This research has been supported by MINECO Grant MTM2017-82724-R, and by the Xunta de Galicia (Grupos de Referencia Competitiva ED431C-2016-015, ED431C-2020-14 and Centro Singular de Investigación de Galicia ED431G 2019/01), all of them through the ERDF.

References

[1] R. Beran, Nonparametric regression with randomly censored survival data, Technical report, University of California (1981).

[2] E. L. Kaplan, P. Meier, Nonparametric estimation from incomplete observations, Journal of American Statistical Association 53 (1958) 457–481.

[3] D. M. Dabrowska, Uniform consistency of the kernel conditional Kaplan-Meier estimate, The Annals of Statistics 17 (3) (1989) 1157–1167.

[4] W. Gonzalez-Manteiga, C. Cadarso-Suárez, Asymptotic properties of a generalized Kaplan-Meier estimator with applications, Journal of Nonparametric Statistics 4 (1994) 65–78.

[5] I. Van Keilegom, N. Veraverbeke, Uniform strong convergence results for the conditional Kaplan-Meier estimator and its quantiles, Communications in Statistics, Theory Methods 25 (2) (1996) 2251–2265.

[6] M. C. Iglesias-Pérez, W. González-Manteiga, Strong representation of a generalized product-limit estimator for truncated and censored data with some applications, Journal of Nonparametric Statistics 10 (3) (1999) 213–244.

[7] I. Van Keilegom, M. Akritas, Transfer of tail information in censored regression models, The Annals of Statistics 27 (5) (1999) 1745–1784.

[8] I. Van Keilegom, M. Akritas, N. Veraverbeke, Estimation of the conditional distribution in regression with censored data: a comparative study., Computational Statistics & Data Analysis 53 (2001) 457–481.

[9] A. Gannoun, J. Saracco, A. Yuan, G. Bonney, Nonparametric quantile regression with censored data, Scandinavian Journal of Statistics 32 (2005) 527–550.

[10] A. Gannoun, J. Saracco, K. Yu, Comparison of kernel estimator of conditional distribution function and quantile regression under censoring, Statistical Modelling 7 (4) (2007) 329–344.

[11] Z. Cai, Weighted local linear approach to censored nonparametric regression, in: M. G. Akritas, D. N. Politis (Eds.), Recent Advances and Trends in Nonparametric Statistics, 2003, p. 217–231.

[12] L. N. Allen, L. C. Rose, Financial survival analysis of defaulted debtors, Journal of the Operational Research Society 57 (6) (2006) 630–636.

[13] R. Cao, J. M. Vilar, A. Devia, Modelling consumer credit risk via survival analysis (with discussion), Statistics and Operations Research Transactions 33 (1) (2009) 3–30.

[14] L. Dirick, G. Claeskens, B. Baesens, Time to default in credit scoring using survival analysis: a benchmark study, Journal of the Operational Research Society 68 (2017) 652–665.

[15] R. Peláez, R. Cao, J. M. Vilar, Probability of default estimation in credit risk using a nonparametric approach, TEST (to appear).
URL https://doi.org/10.1007/s11749-020-00723-1

[16] A. Földes, L. Rejtø, B. B. Winter, Strong consistency properties of nonparametric estimators for randomly censored data, ii: estimation of density and failure rate, Periodica Mathematica Hungarica 12 (1981) 15–29.

[17] E. Leconte, S. Poiraud-Casanova, C. Thomas-Agnan, Smooth conditional distribution function and quantiles under random censorship, Lifetime Data Analysis 8 (2002) 229–246.

[18] B. W. Silverman, Density estimation for statistics & data analysis, Monographs on Statistics and Applied Probability Chapman and Hall (1986).