Smoothed time-dependent ROC curve for right censored survival data

Kassu Mehari Beyene* | Anouar El Ghouch

Institute of Statistics, Biostatistics and Actuarial Sciences, Catholic University of Louvain, Louvain la Neuve, Belgium

Correspondence
*Kassu Mehari Beyene
m2kassu@gmail.com

Summary

The prediction reliability is of primary concern in many clinical studies when the objective is to develop new predictive models or improve existing risk scores. In fact, before using a model in any clinical decision making, it is very important to check its ability to discriminate between subjects who are at risk of, for example, developing certain disease in a near future from those who will not. To that end, the time-dependent receiver operating characteristic curve (ROC) is the most commonly used method in practice. Several approaches have been proposed in the literature to estimate the ROC non-parametrically in the context of survival data. But, except one recent approach, all the existing methods provide a non-smooth ROC estimator whereas, by definition, the ROC curve is smooth. In this article we propose and study a new nonparametric smooth ROC estimator based on a weighted kernel smoother. More precisely, our approach relies on a well-known kernel method used to estimate cumulative distribution functions of random variables with bounded supports. We derived some asymptotic properties for the proposed estimator. As bandwidth is the main parameter to be set, we present and study different methods to appropriately select one. A simulation study is conducted, under different scenarios, to prove the consistency of the proposed method and to compare its finite sample performance with a competitor. The results show that the proposed method performs better and appear to be quite robust to bandwidth choice. As for inference purposes, our results also reveal the good performances of a proposed nonparametric bootstrap procedure. Furthermore, we illustrate the method using a real data example.

KEYWORDS:
AUC, kernel estimation, sensitivity and specificity, bandwidth selection, weighted distribution

1 | INTRODUCTION

In clinical medicine, identifying individuals with high risk of developing a certain event of interest is crucial as it helps to plan early prevention and possibly treatment. The event can be death due to certain disease, diagnosis of a disease or the recurrences of a disease. In particular, the use of prediction models to predict the risk of developing a disease (e.g., diabetes, breast cancer, asthma, dementia, ...), given the individual's characteristics and diagnostic test(s) measures, is very popular. Nowadays, prognostic risk scores (or markers) derived from statistical models are widely used to identify high risk individuals. In public health, for example, the Framingham risk score is a well-known tool used to predict the probability of developing cardiovascular disease within 10-year period. This probability is estimated from a risk score obtained by combining risk factors such as age, blood pressure, smoking status, diabetes, high density lipoproteins and total cholesterol levels.

However, before a prognostic risk score is used in routine clinical practice, its predictive quality has to be proven. One common way to do that is to measure its ability to correctly discriminate subjects with and without the outcome of interest. Different accuracy measures exist in the literature. For binary marker, the most common known measures are the sensitivity and specificity. Sensitivity is defined as the probability that a test result will be positive when the disease is present (true positive rate) and specificity is the probability that a test result will be negative when the disease is not present (true negative rate). For continuous diagnostic test, the receiver operating characteristic (ROC) curve and area under the ROC curve (AUC) are the two commonly used measures. The ROC curve is the plot of sensitivity against one minus specificity for all possible cutoff values of the marker. In the literature, both ROC curve and AUC have been receiving increased attention over the past few decades.[1,2]

In classical analysis, disease status is assumed to be known. However, in prognostic studies, as for example in survival analysis, the disease status of the subject is time-dependent. In such a case, a time-dependent event is dichotomized as positive or "diseased" and negative or "non-diseased" by considering a particular fixed time of interest, t. This leads to defining time-dependent sensitivity, time-dependent specificity and the resulting time-dependent ROC and AUC. Heagerty et al[3], Heagerty and Zheng[4], Etzioni et al[5] and Slate and Turnbull[6], among others, proposed several definitions and extension to expand the classical methodology to the case of time to event data. The cumulative sensitivity and dynamic specificity are, however, the most widely used definitions in clinical applications as they are more relevant; see[7,8]. Therefore, in this article we will focus on these definitions for measuring the discrimination ability of a given risk score.

Let T denote the time to occurrence of a certain event of interest, and M be a continuous marker measured at baseline. The event status of a subject at time horizon t is defined as $D_t = I(T \leq t)$, with 1 indicating that the subject has the event before t (positive) and 0 otherwise (negative). Without loss of generality, hereafter, we assume that a higher value of M is indicative of a shorter survival time. Hence, for a given cutoff value c, a subject i is classified as positive if its marker value is greater than c ($M_i > c$). Accordingly, the time-dependent true positive rate (Tp_t) and false positive rate (Fp_t) are defined as

$$\mathrm{Tp}_t(c) = P(M > c | D_t = 1),$$
$$\mathrm{Fp}_t(c) = P(M > c | D_t = 0),$$

where $c \in (-\infty, \infty)$ is a fixed cutoff value. The corresponding time-dependent ROC function (ROC_t) is

$$ROC_t(u) = \mathrm{Tp}_t\{\mathrm{Fp}_t^{-1}(u)\}, \tag{1}$$

where $u \in (0, 1)$ and $c = \mathrm{Fp}_t^{-1}(u)$ is the threshold value such that $u = \mathrm{Fp}_t(c)$.

ROC_t can be quantitatively summarized by the time-dependent area under the ROC curve (AUC_t) defined as

$$AUC_t = \int_0^1 ROC_t(u)du. \tag{2}$$

The main challenge in estimating the above quantities is that, because of censoring, the disease status is not known for all subject under study. In fact, in survival analysis, some individuals are typically lost during the follow-up period, due to various causes, or did not develop the event of interest by the end of the study. There is abundant literature about the estimation of ROC curves taking censoring into account using either parametric, nonparametric or semiparametric techniques. Heagerty et al[3] proposed two non-parametric methods. The first one is based on the Kaplan-Meier estimator of the survival function whereas the second is based on the bivariate nearest neighbor estimator of Akritas.[9] Their work can be considered as the cornerstone for the area of time-dependent accuracy measures estimation since it opened the door for many proposals. For instance, Chambless and Diao[10] proposed a non-parametric Kaplan-Meier like method and a semi-parametric model based approaches. Martínez-Camblor et al[11] and Li et al[12] independently proposed methods for ROC curve and AUC estimation based on imputation idea where the missing event status for censored subjects are replaced by their "expected" values obtained from the Cox model or the (conditional) Kaplan-Meier estimator. There are plenty of other proposals; see for example[13–15] and the references given in these papers. For a comprehensive summary the reader is also referred to Blanche et al[8] and Kamarudin et al[16].

Like a cumulative distribution function, the ROC is commonly estimated non-parametrically using empirical method. This simple and very popular approach leads to a valid and reliable estimator. However, unlike the population ROC which is continuous and smooth, the empirical ROC is a step function. This is a serious drawback in many real applications where the ROC curves are used not only to evaluate and compare the discriminatory ability of markers but also to find an optimal cut-off point. To achieve this, the most commonly used methods are the point on the ROC curve closest to $(0, 1)$ and the point on the ROC curve farthest from the non-informative marker (diagonal line), also called the Youden index; see, for example [17,18], for more details. Finding a cut-point from a smooth ROC curve is clearly more convenient as the empirical ROC will typically lead to a non-unique solution. Fluss et al [19] compare empirical method with a kernel based method and come to the conclusion that the former "has the worst performance and is not recommended unless sample size is very large" and advocate the use of kernel method in most situations. We shall not discuss further the problem of cut-point in the present work and refer interested readers to [20–22] for more information. In their refined analysis, Lloyd and Yong [23] compare, theoretically and via simulations, the empirical ROC estimator with a kernel estimator, in the case of uncensored data, and prove that the latter dominate the former for moderate to large samples. Similar conclusions hold when comparing the classical empirical and the kernel smooth estimator of a distribution function, see [24–26]. These and other related works show that kernel based estimators are, in general, asymptotically more efficient than corresponding classical empirical estimators. Another argument in favor of smoothing is that the discontinuity of the empirical ROC makes its graphical representations not very nice from a visual point of view and not intuitive. All these justify our interest on developing a smooth nonparametric estimator for the time-dependent ROC for censored data. Note that, in contrast to cross-sectional studies, where there is an abundance literature about smoothing ROC curves, see for example the recent paper of Pulit [27] and the references given therein, this problem is not well studied in the context of survival analysis. In fact, to the best of our knowledge, the only related work is the recent paper of Martínez-Camblor and Pardo-Fernández [28]. These authors proposed a smooth time-dependent ROC curve estimators for both Cumulative sensitivity/Dynamic specificity and Incident sensitivity/Dynamic specificity. Their estimators are based on bivariate kernel density estimation and they showed that their Cumulative/Dynamic smoothed ROC estimator out performs over the nearest neighbor estimator of Heagerty et al [3].

To construct our kernel based smooth estimator we use the fact that the time-dependent ROC can be expressed as a weighted cumulative distribution of a random variable supported on the unit interval. The weights are nothing but a random variable used to impute the "missing" values of $I(T_i \leq t)$ for censored observations. In this proposal there are some important challenges that must be addressed. The first is the bounded support of ROC function. This is a problem because standard kernel estimates suffer from boundary effects when the support of the variable of interest is bounded. As this may affect the performances (convergence rate and bias) of the estimator near the boundaries, a "boundary correction" is needed. For that, one of the most used method, that we adopt in this work, is to use a transformation mapping the unit interval to the real line. After transforming the observed data, standard kernel method may be satisfactorily applied without further consideration. The second difficulty is that the weight variable is unknown and hence needs to be estimated from the observed data. For this we use a plug-in method based on Beran's nonparametric estimator. [29] As an additional problem, in addition to the smoothing parameter needed to compute Beran's estimator, we need to choose a bandwidth parameter for our kernel estimator of the ROC curve itself. For that, we adapt some well-known existing bandwidth selection methods, used in kernel estimation of a classical (unweighted) cumulative distribution, to our situation.

The rest of this article is organized as follows. Next section gives the main statements and definitions of the paper and describes in details our estimation approach. In this section, we also present some asymptotic results and introduce a bootstrap procedure for estimating the standard deviation and constructing confidence intervals. Section 3 is devoted to the problem of the bandwidth selection. In Section 4, the finite sample performances of our estimator is evaluated by simulations. A real data illustration is provided in Section 5. Finally, discussions and conclusions are presented in Section 6.

2 | ESTIMATION METHOD

We assume that the observed data set consists of $\{(Y_i, \Delta_i, M_i), i = 1, 2..., n\}$ where $Y_i = \min(T_i, C_i)$ is the observed survival time, T_i is the time to the event of interest, C_i is the censoring time, $\Delta_i = I(T_i \leq C_i)$ is the censoring indicator and M_i is the marker observed at the ith subject. In the following, the event status of interest will be denoted by $D(t) = I(T \leq t)$, with 1 (0) indicating that the event happened before (after) a given time horizon, t.

2.1 | Imputation

As mentioned in the Introduction, our objective is to measure the predictive ability of M using a smooth ROC curve. The main difficulty is that event status $D(t)$ is unknown for some subjects as a result of censoring. Hence, following an idea suggested independently by Martínez-Camblor et al[11] and Li et al[12], we impute $D(t)$ by its expected value given the observed data. Thus, we impute $D(t)$ by the random variable

$$W(t) = E\{D(t)|Y, \Delta, M\} = \left[1 - (1-\Delta)\frac{S(t|M)}{S(Y|M)}\right] I(Y \le t),$$

where $S(t|m) = P(T > t|M = m)$ denotes the conditional survival function of T given the marker M. The second equality in the above equation is true under the usual assumption, in survival analysis, that T and C are independent given M. Unlike other existing methods in the literature, here all the available information is used in the imputation. Also, it is interesting to observe that $W(t) = D(t)$ except in the case when $Y = y < t$ and $\Delta = 0$. In this case, $W(t)$ is nothing but the conditional probability $P(T \le t|T > y, M)$. In the above expression of $W(t)$, $S(t|m)$ is unknown and needs to be estimated form the observed data. This can be done via, for example, the kernel estimator of Beran that can be written in the following form:

$$\hat{S}(t|m) = \prod_{i:\, Y_i \le t} \left(1 - \frac{v_i(m)}{\sum_{j=1}^n v_j(m) I(Y_j \ge Y_i)}\right)^{\Delta_i},$$

where, for a given bandwidth $b \equiv b_n$ and a given kernel density k, $v_i(m) = \frac{k(\frac{M_i - m}{b})}{\sum_j k(\frac{M_j - m}{b})}$. The consistency and asymptotic normality of this estimator were established by several researchers (see for example[29–32]). In the following, we will use the notation $\hat{W}_i(t)$ to denote $\left[1 - (1-\Delta_i)\frac{\hat{S}(t|M_i)}{\hat{S}(Y_i|M_i)}\right] I(Y_i \le t)$.

2.2 | Estimation

In this section, we first introduce an empirical (non-smooth) estimation of the ROC and from this we propose a smooth version of it. As given by (1), the ROC can be written as

$$\mathrm{ROC}_t(u) = P(Fp_t(M) \le u|D_t = 1) = \frac{E\{I(Z(t) \le u)W(t)\}}{E\{W(t)\}}, \tag{3}$$

where $Z(t)$ denote the random variable $Fp_t(M)$. The function Fp_t can also be written on terms of $W(t)$ as

$$Fp_t(m) = \frac{E\{I(M \ge m)(1 - W(t))\}}{E\{(1 - W(t))\}}.$$

An empirical estimator, say $\hat{F}p_t(m)$, of this quantity can be obtained by replacing expectations with simple averages. Plugging-in this on (3), leads to the following empirical estimator of the ROC,

$$\widehat{\mathrm{ROC}}_{emp}(u) = \frac{\sum\limits_{i=1}^n \hat{W}_i I(\hat{Z}_i \le u)}{\sum\limits_{i=1}^n \hat{W}_i}, \tag{4}$$

where $\hat{Z}_i \equiv \hat{Z}_i(t) = \hat{F}p_t(M_i)$ and $\hat{W}_i \equiv \hat{W}_i(t)$. Here and in the following, we suppress the dependence on time t for notational simplicity.

The ROC curve estimator given in equation (4) is nothing but a weighted empirical distribution function with weights given by $\hat{\mathcal{W}}_i = \hat{W}_i / n^{-1}\sum_{i=1}^n \hat{W}_i$, i.e., $\widehat{ROC}_{emp} = n^{-1}\sum_i \hat{\mathcal{W}}_i I(\hat{Z}_i \le u)$. To the best of our knowledge, this simple and promising estimator has never been proposed or studied in the literature. However, it has a serious drawback of being a step function. For this reason, and given the arguments invoked in the Introduction, here we are interested in getting a smooth ROC curve estimator. As done to smooth out the classical empirical distribution, see[33], this goal can be easily achieved by replacing the indicator function $I(.)$ in $\widehat{\mathrm{ROC}}_{emp}$ by a kernel distribution function. More precisely, for a bandwidth parameter $h \equiv h_n$, i.e., a sequence of strictly positive real numbers converging to zero with n, a kernel smooth estimator of $ROC(u)$ is given by

$$\begin{aligned}\widehat{\mathrm{ROC}}(u) &= \int_{-\infty}^{\infty} k\left(\frac{u - x}{h}\right) d\widehat{\mathrm{ROC}}_{emp}(x) \\ &= n^{-1}\sum_{i=1}^n \hat{\mathcal{W}}_i K\left(\frac{u - \hat{Z}_i}{h}\right),\end{aligned} \tag{5}$$

where $K(x) = \int_{-\infty}^{x} k(s)ds$ is a (kernel) distribution function with a density k. Clearly $\widehat{\text{ROC}}$ inherits its smoothness from K. Also, observe that when K is chosen to be the indicator function $I(x \geq 0)$ then $\widehat{\text{ROC}}$ reduces to $\widehat{\text{ROC}}_{emp}$. So that this latter can be seen as a particular case of equation (5).

2.3 | Some asymptotic results

In this section we establish some asymptotic properties of the proposed estimator including its consistency and asymptotic mean squared error. For that let us assume, for simplicity, that the kernel k is a bounded symmetric (around zero) density with a bounded support, say $[-1,1]$. Put $\mu_2(k) = \int y^2 k(y)dy$ and $\rho(k) = 2\int yk(y)K(y)dy$. The bandwidth h is such that $h \to 0$ and $nh \to \infty$. We will also need the following assumption.

Assumption A: For the fixed t, (i) $\sup_m |\hat{S}(t|m) - S(t|m)| = O(r_n)$ a.s., for some $r_n = o(1)$, and (ii) $\inf_m S(t|m) > s$, for some $s > 0$.
Under some suitable conditions, see[34], Assumption $A(i)$ holds with $r_n = (\log(b^{-1})/(nb))^{1/2} = o(b^{1+\delta})$, for some $0 < \delta < 1$. This Assumption guarantees that $\max_i |\hat{W}_i - W_i| = O(r_n)$ a.s.. Which, by some simple algebra, implies that $\max_i |\hat{Z}_i - Z_i| = O(r_n)$ a.s.. Using these two equalities, it can be shown that, uniformly in u,

$$\widehat{\text{ROC}}(u) = ROC_n(u) + O(r_n) \text{ a.s.,} \tag{6}$$

where $ROC_n(u) = n^{-1}\sum_{i=1}^{n} \mathcal{W}_i K\left(\frac{u - Z_i}{h}\right)$ and $\mathcal{W}_i = \frac{W_i}{E(W_i)} = \frac{W_i}{P(T \leq t)}$.
Let f and F denote the density and cumulative distribution of Z. Define $\psi_p(z) = E(\mathcal{W}^p | Z = z)$, $p = 1, 2$, and

$$\xi_p(u) = \int_{-\infty}^{u} \psi_p(z)dF(z) = E\left\{\mathcal{W}^p I(Z \leq u)\right\}.$$

Observe that $\xi_1(u) = ROC(u)$ and

$$E\left\{\mathcal{W}^p K^p\left(\frac{u - Z}{h}\right)\right\} = E\left\{\psi_p(Z)K^p\left(\frac{u - Z}{h}\right)\right\} = \int_{-\infty}^{\infty} K^p\left(\frac{u - z}{h}\right)d\xi_p(z).$$

Standard tools from kernel estimation literature, see e.g. Chapter 1 of Li and Racine[35], lead to the following result.

Theorem 1. Assume that f', ψ_1' and ψ_2^2 are continuous at u, then

(i) $E\left\{ROC_n(u)\right\} = ROC(u) + \frac{h^2}{2}\mu_2(k)ROC''(u) + o(h^2)$.

(ii) $nVar\left\{ROC_n(u)\right\} = Var\left\{\mathcal{W}I(Z \leq u)\right\} - h\rho(k)\psi_2(u)f(u) + o(h)$.

By combining this Theorem and equation (6), we obtain the following result on the MSE of $\widehat{\text{ROC}}(u)$:

$$MSE\left\{\widehat{\text{ROC}}(u)\right\} = AMSE(u) + o(h^4 + \frac{h}{n} + b^{2+2\delta} + h^2 b^{1+\delta}),$$

where

$$AMSE(u) = \frac{h^4}{4}\mu_2^2(k)(ROC''(u))^2 + n^{-1}Var\left\{\mathcal{W}I(Z \leq u)\right\} - \frac{h}{n}\rho(k)\psi_2(u)f(u). \tag{7}$$

The mean square error of $\widehat{ROC}(u)$ is of order $o(h^4 + \frac{h}{n} + b^{2+2\delta} + h^2 b^{1+\delta})$ which is different from the order of MSE of the classical (unweighted) kernel smoothed distribution estimator that is $o(h^4 + \frac{h}{n})$ (see, e.g., Li and Racine[35], p.22). The extra term $b^{2+2\delta} + h^2 b^{1+\delta}$ comes from the estimation of an unknown weight W. The order of the variance term here is $o(hn^{-1})$, which is the same as the classical kernel smoothed distribution estimator. However, the bias term has an extra term $b^{1+\delta}$ which comes from the estimation of the weight. These asymptotic results reduce to the classical case if the weights are known constant.

2.4 | Boundary correction

As any standard kernel based nonparametric estimator $\widehat{\mathrm{ROC}}(u)$ is subject to boundary bias because $\hat{Z}$ (the pseudo data) is supported in the unit interval $[0,1]$. This bias comes from the fact that the kernel, which has no knowledge of the boundary, assigns positive weight outside the compact support of the variable. To be more clear, the calculation done in the previous section assumes, implicitly, that $u \in [h, 1-h]$. When, for example, $u \in [0,h)$, i.e. $u \equiv u_n = \alpha h$, for some $0 \leq \alpha < 1$, it can be shown that

$$E\left\{ROC_n(u)\right\} = ROC(u)\int_{-1}^{\alpha} k(y)dy + O(h).$$

This implies that our estimator $\widehat{\mathrm{ROC}}$ is inconsistent in the boundary regions. Such a problem is very well documented in the literature, especially in the case of kernel density estimation, and several modified estimators have been proposed. The most used methods are the reflection method, the boundary kernel method, the local linear approach and the transformation method; see [36,37] and the references therein for more about this subject. In the case of completely observed data, Koláček and Karunamuni [38] has recently suggested and studied a boundary correction method for the classical kernel estimator of the ROC curve. Here, the procedure we shall use to handle this problem is to apply the method of transformation that aims at "sending away" the boundaries to $\pm\infty$; see [39,40]. Specifically, the idea consist on transforming the variable with bounded support (Z in our case) to that with an unbounded support. The distribution of the transformed variable can then be estimated using the standard kernel estimator. Let $Q : [0,1] \to \mathbb{R}$ be any continuous and increasing function such that $\lim_{x\to 0} Q(x) = -\infty$ and $\lim_{x\to 1} Q(x) = +\infty$, since $\mathrm{ROC}_t(u) = P(Q(Z_t) \leq Q(u)|D_t = 1)$, we suggest the following modified kernel ROC estimator

$$\widehat{\mathrm{ROC}}_{TR}(u) = n^{-1}\sum_{i=1}^{n} \hat{\mathcal{W}}_i K\left(\frac{Q(u) - Q(\hat{Z}_i)}{h}\right). \tag{8}$$

The most commonly used transformations are the quantile functions of random variables with unbounded support like for example the Probit transform corresponding to the standard normal quantile function denoted hereafter by Φ^{-1}. In the case of kernel density estimation, with completely observed data, this approach was recently investigated by Geenens [41] and Wen and Wu [42]. They show some drawbacks of this simple method and suggest some refinements. Here we stick to the "naive" approach, as given by equation (8), but we believe that it could be interesting to investigate this problem in more details.

Note that when Q is chosen to be the identity function, $\widehat{\mathrm{ROC}}_{TR}$ reduces to the estimator given in equation (5).

Remark 1. As for the AUC, since

$$AUC_t = \int_0^1 ROC_t(u)du = \frac{\int_0^1 E(I(Z_t \leq u)W_t)du}{E(W_t)} = 1 - E(\mathcal{W}_t Z_t),$$

an estimate of it can be obtained easily using either the empirical version of this last expression, i.e., $1 - n^{-1}\sum_{i=1}^{n} \hat{\mathcal{W}}_i \hat{Z}_i$, or by a numerical integration method, i.e., $\int_0^1 \widehat{ROC}_{TR}(u)du$. The former has the advantage of being simple and independent of the bandwidth parameter needed for the smooth ROC estimator.

2.5 | Variance estimation and inference

To make inference about the ROC or the AUC, one need to estimate the (asymptotic) distribution of the proposed estimators. In the following we will focus on the AUC but the same procedure can be applied to the ROC. Given the difficulties on getting an asymptotic distribution or at least a close and easy to estimate formula for the asymptotic variance, here we propose to use the well-known "naive" nonparametric bootstrap method of Efron [43]. This consist of first drawing, with replacement, a large number, say B, of random samples of size n from the original observed data $\{(M_i, Y_i, \Delta_i), i = 1, 2, ..., n\}$. Next, from each bootstrap sample b, $b = 1, \ldots, B$, we calculate $\widehat{AUC}_b$, the estimated bootstrap AUC statistic. The empirical variance of this bootstrap statistic, i.e.,

$$S_B^2 = B^{-1}\sum_b \left(\widehat{AUC}_b - B^{-1}\sum_b \widehat{AUC}_b\right)^2$$

can be then used as an estimator of the variance of $\widehat{AUC}$, the original estimate of the AUC. $\{\widehat{AUC}_b, b = 1, 2, ..., B\}$ can also be used to construct asymptotic confidence intervals. For that, several approaches exist in the literature; see for example [44]. Among theme there is (i) the standard normal method: $\widehat{AUC} \pm z_{1-\alpha/2} S_B$, where z_α is the $\alpha-$quantile of a $N(0,1)$ and (ii) the percentile method: $[\widehat{AUC}_B(\alpha/2), \widehat{AUC}_B(1-\alpha/2)]$, where $\widehat{AUC}_B(\alpha)$ is the $\alpha-$quantile of the B bootstrap sample estimates. The validity and precision of these bootstrap procedures will be investigated via simulations in Section 4. Note that the use of bootstrap in the context of time-dependent ROC and AUC estimations is not new but appears in several papers including[3,12,13].

3 | BANDWIDTH SELECTION

To compute the proposed estimator we have to choose the kernel and the two bandwidths: b, needed by Beran's estimator in the imputation stage, and h needed to smooth the indicator function. It is well-known form kernel smoothing literature, see e.g., Jones[45], that the kernel function has little effect on the performance of the resulting estimates but this is not the case, in general, for the bandwidth(s) that has to be selected carefully. Now, the effect of b should be less important then the effect of h as the latter is the parameter that really controls the degree of smoothness of our estimator. Also, observe that b intervenes only in the calculation of W_i for subjects with $Y_i < t$ and $\Delta_i = 0$. More precisely, for these subjects, we need b to get $\hat{S}(t|M_i)/\hat{S}(Y_i|M_i)$. Dabrowska[46] showed that, uniformly in t, the bias of $\hat{S}(t|M_i)$ is asymptotically proportional to b^2. Hence, as noted by Li et al[12], for inappropriately larger or smaller b, the biases in $\hat{S}(t|M_i)$ and $\hat{S}(Y_i|M_i)$ are in the same direction and may cancel out to some extent, particularly when Y_i and t are close to each other. This is confirmed by our simulation study reported in the next section where we investigate the robustness of our estimator to the selection of b. So in this section we will focus on the bandwidth h that may have a significant influence on the ROC estimate.

Let start with some theoretical considerations. The asymptotic integrated mean square error ($AMISE$) is found by integrating the $AMSE(u)$, see (7). The bandwidth which minimizes this $AMISE$ can be easily obtained by differentiation and it can be expressed as

$$h = \left(\frac{E(\mathcal{W}^2)\rho(k)}{n\mu_2^2(k)\kappa(ROC)} \right)^{1/3},$$

where $\kappa(ROC) = \int_{-\infty}^{\infty} \left\{ROC''(u)\right\}^2 du = \int \left\{(\psi_1(u)f(u))'\right\}^2 du$, where $f(.)$ is the density of Z and $\psi_1(z) = E(\mathcal{W}|Z=z)$. It is interesting to note that when W is a constant (non-stochastic) then $ROC(u) = P(Z \le u)$ and h reduces to $h_F = \left(\frac{\rho(k)}{n\mu_2^2(k)\kappa(F)} \right)^{1/3}$, which is the optimal bandwidth that minimizes the $AMISE$ of the classical kernel distribution function, see e.g., Polansky and Baker[47]. The difference between h and h_F comes from $E(\mathcal{W}^2) = Var\{\mathcal{W}\} + 1$. As a consequence, when the variability of $\mathcal{W}$ increases a larger bandwidth will be needed, which sounds intuitive. $E(\mathcal{W}^2) = E\left\{W^2\right\}/(E\{W\})^2$ can be easily estimated empirically using the estimated weights $\hat{W}_i$ given at the end of Section 2.1.

To estimate h, the real problem comes from the estimation of $\kappa(ROC)$. In the case of the (unweighted) distribution function, the simplest and most used approach, known as Rule-of-Thumb, is to estimate $\kappa(F)$ using a parametric model. For example, if we assume that f is a normal density with mean μ and variance σ^2, then it is very simple to show that $\kappa(F) = (4\sigma^3\pi^{0.5})^{-1}$; see e.g., Polansky and Baker[47]. An estimate for $\kappa(F)$ is given by $\widehat{\kappa(F)} = (4\hat{\sigma}^3\pi^{0.5})^{-1}$, where $\hat{\sigma}$ is an estimator for the population standard deviation σ. Silverman[48] suggest to use the minimum of sample standard deviation and sample interquartile range divided by 1.349 as a robust estimate of σ.

Since the ROC can be seen as a weighted distribution function, one may adapt the above normal reference rule and estimate $\kappa(ROC)$ by

$$\widehat{\kappa(ROC)}_{NR} = (4\hat{\sigma}_w^3\pi^{0.5})^{-1}, \tag{9}$$

where $\hat{\sigma}_w$ is the smaller of weighted sample standard deviation and weighted sample interquartile range divided by 1.349. The resulting bandwidth is given by

$$\hat{h}_{NR} = \left[\frac{\hat{E}\{\hat{\mathcal{W}}^2\}\rho(k)}{n\mu_2^2(k)\widehat{\kappa(ROC)}_{NR}} \right]^{1/3}, \tag{10}$$

and will be called *normal reference* bandwidth.

Another option is to estimate the quantity $\kappa(ROC)$ non-parametrically as it was done by Polansky and Baker[47] for $\kappa(F)$. Following these authors, using integration by parts, we can easily show that

$$\kappa(ROC) = -E\left\{\mathcal{W}ROC^{(3)}(Z)\right\},$$

where $ROC^{(3)}(.)$ is the third derivative of the ROC function. Which suggests estimating $\kappa(ROC)$ by

$$\kappa\widehat{(ROC)}_{PI} = -\hat{E}\big\{\hat{\mathcal{W}}\widehat{ROC^{(3)}}(Z)\big\} = -n^{-2}g^{-3}\sum_{i=1}^{n}\sum_{j=1}^{n}\hat{\mathcal{W}}_i\hat{\mathcal{W}}_j K^{(3)}\Big(\frac{\hat{Z}_i-\hat{Z}_j}{g}\Big), \tag{11}$$

where $K^{(3)}(.)$ is the third derivative of the kernel distribution function K and g is a smoothing parameter that needs to be chosen. We will refer to the resulting bandwidth, obtained by replacing $\kappa\widehat{(ROC)}_{NR}$ by $\kappa\widehat{(ROC)}_{PI}$ in (10), as the *plug-in* bandwidth and will be denoted by h_{PI}. The main difficulty of this approach is that one still needs to choose the pilot bandwidth g in order to calculate $h_{PI} \equiv h_{PI}(g)$. For the classical (unweighted) kernel distribution estimation, Polansky and Baker[47] proposed an iterative method where g is selected based on the asymptotic mean squared error of $\widehat{\kappa(F)}$. Due to the complexity of deriving the asymptotic mean squared error for $\kappa\widehat{(ROC)}_{PI}$, in this study we use the normal reference method given in (10) to calculate the bandwidth g.

An alternative approach is to use Cross-Validation (CV). In the case of the distribution function, different approaches exist in the literature; see for example[49–51]. In the latter the authors proposed a modified cross-validation estimator that is asymptotically optimal and performs well in simulation studies and real data analysis. The suggested bandwidth is obtained by minimizing the function

$$CV_F(h) = \sum_{i=1}^{n}\int_{-\infty}^{\infty}\left[I(x \geq x_i) - \hat{F}^{-i}(x_i)\right]^2 dx,$$

where $\hat{F}^{-i}(x_i)$ denotes the kernel distribution estimator calculated with the bandwidth h and without the i^{th} observation. To adapt this method to our case, we suggest incorporating the weight variable into the CV function and select the bandwidth that minimizes

$$CV_{ROC}(h) = \sum_{i=1}^{n}\int_{0}^{1}\left[\hat{\mathcal{W}}_i I(u \geq \hat{Z}_i) - \widehat{\mathrm{ROC}}^{-i}(u)\right]^2 du, \tag{12}$$

where $\widehat{\mathrm{ROC}}^{-i}(u) = (n-1)^{-1}\sum_{j\neq i}\hat{\mathcal{W}}_j K(\frac{u-\hat{Z}_j}{h})$ is the ROC curve estimate without the i^{th} observation.

The performances of the three methods discussed above (normal reference (NR), plug-in (PI) and Cross-Validation (CV)) will be investigated in the next section.

4 | SIMULATION STUDY

In this section, we present a Monte Carlo simulation study to investigate the finite sample performance of the proposed ROC curve estimators: the empirical (non-smoothed), see equation (4), the smoothed estimator without boundary correction, see equation (5), and the smoothed estimator with boundary correction, see equation (8). We are also interested in comparing the performance of the latter with the smoothed time-dependent ROC curve estimator proposed by Martínez-Camblor and Pardo-Fernández[28], hereafter denoted by (MP). As mentioned above, the ROC curve estimator with and without the boundary correction requires selecting two bandwidth parameters. Given the difficulty to select these parameters in an optimal way, we also investigate here the robustness of these ROC estimators to the choice of the bandwidth parameters and assess the performance of the proposed modified bandwidth selection methods. Furthermore, the validity of the proposed bootstrap approach is also examined. The behavior of the estimator depends on various aspects, including sample size and censoring rates. Therefore, we generate data considering different scenarios, the details of which are given in the upcoming subsection.

4.1 Data generation

To generate the data, we considered two scenarios

- Scenario I: The survival time T is assumed to follow a log-normal distribution with parameters $\mu = 0$ and $\sigma = 2$. The censoring time C is independently generated from a log-normal distribution with parameters μ and $\sigma = 2$. The value of μ is chosen to produce the required censoring proportions. The marker is given by $M = \sqrt{p}T + \sqrt{1-p}R$, where R is a log-normal distributed random variable with parameters $\mu = 0$ and $\sigma = 2$.
- Scenario II: The survival time is assumed to follow an $Exponential(1)$ distribution. The censoring time is taken to be independent of T and also follows an exponential distribution with parameter λ, where the value of λ is chosen to produce the required censoring proportions. The marker is given by $M = \sqrt{p}T + \sqrt{1-p}R$, where R is an $Exponential(1)$ distributed random variable.

These scenarios are very similar to those of Martínez-Camblor and Pardo-Fernández[28]. The values of p are set to be 0.25 and 0.50. The later corresponds to a stronger association between the survival time and the marker. The parameter of the censoring time distributions are chosen to achieve 20% and 50% censoring rates. Furthermore, for the ROC curves calculation, we consider the time horizon t corresponding to the quartile values (Q_1, Q_2 and Q_3) of the survival time T. For both scenarios, we generate $N = 1000$ replicates of samples of sizes $n = 100$, 200, 400. The true time-dependent ROC curves for both scenarios are presented in Figure 1.

The performance measures for the proposed ROC estimator with and without boundary correction are calculated using the three proposed data-driven bandwidth selection methods for h: the normal reference bandwidth (NR), the plug-in (PI) and the cross-validation (CV). To select b, the bandwidth parameter needed for Beran's weight calculation, we simply use the plug-in bandwidth selection method of the classical density estimation that was proposed by Sheather and Jones[52]. This method is implemented in the `bandwidth` function of the `stats` package of R.[53]

The MP estimator of[28] is based on the bivariate kernel density estimation and hence it uses a 2×2 matrix of smoothing parameters. In their simulations, the authors considered six data-driven bandwidths selection methods: plug-in (PI), least-squared cross-validation (LSCV) and unbiased cross-validation (UCV), based either on diagonal or non-diagonal smoothing matrices. They found that the non-diagonal version has better performance and hence in our comparison we only consider this latter. The MP method is implemented in the `R` package `smoothROCtime`[54], which was used in our simulations.

To compare the performance of the ROC estimators we chose two criteria: the integrated mean bias given by

$$\mathrm{MIB}(\widehat{\mathrm{ROC}}) = 1000^{-1} \sum_{l=1}^{1000} \int (\widehat{\mathrm{ROC}}_l(u) - \mathrm{ROC}(u))du$$

where $\widehat{\mathrm{ROC}}_l(.)$ is the ROC estimate from the l^{th} sampled data and the integrated mean square error (MISE) given by

$$\mathrm{MISE}(\widehat{\mathrm{ROC}}) = 1000^{-1} \sum_{l=1}^{1000} \mathrm{ISE}(\widehat{\mathrm{ROC}}_l),$$

where $\mathrm{ISE}(\widehat{\mathrm{ROC}}_l) = \int (\widehat{\mathrm{ROC}}_l(u) - \mathrm{ROC}(u))^2 du$. In the above formulas, the true ROC(.) was calculated from a simulated sample (T_i, M_i), $i = 1, ..., 4 \times 10^6$, using equation (4) with $\hat{W}_i = I(T_i \leq t)$.

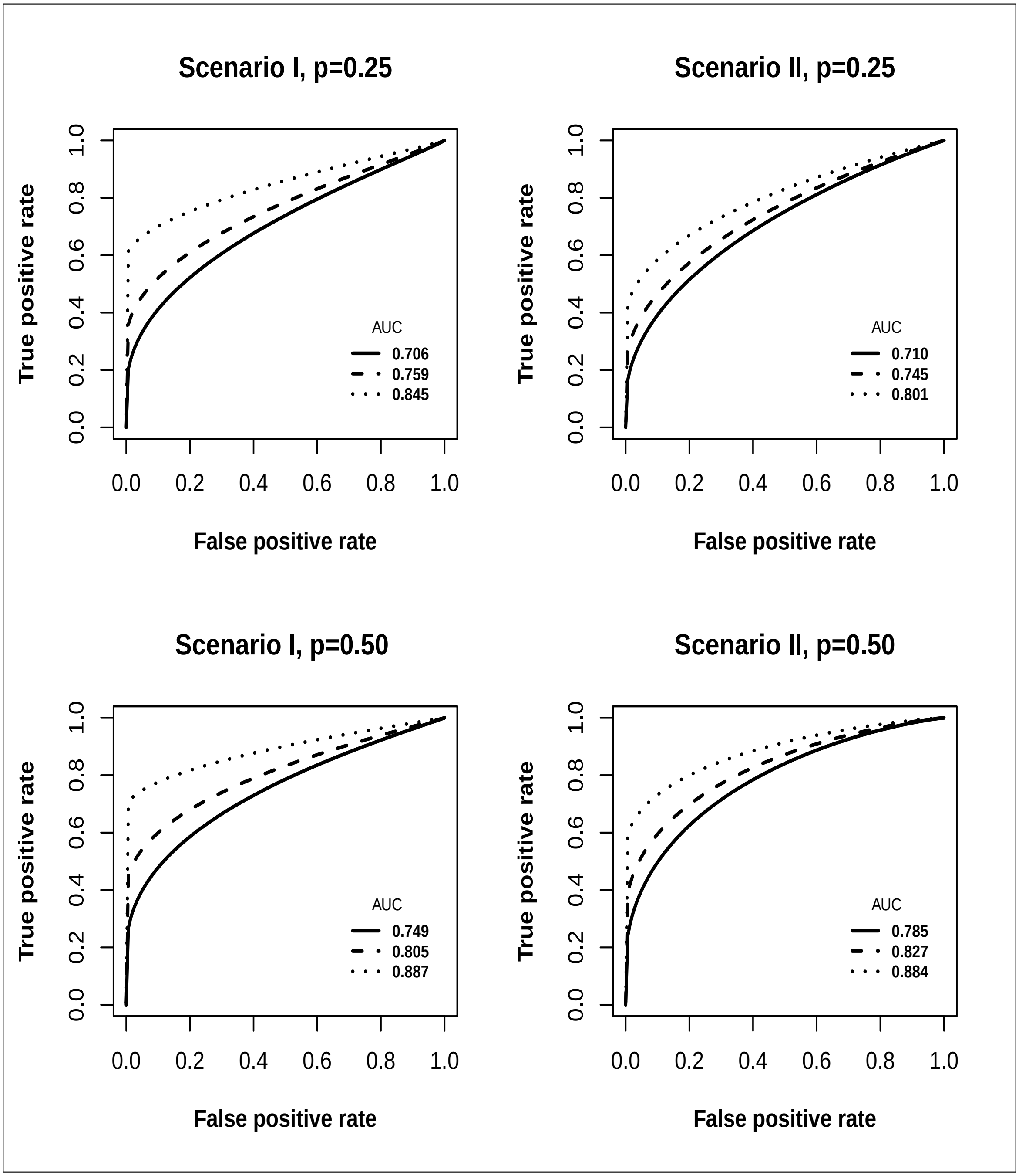

FIGURE 1 The true ROC curves and the corresponding AUC values of the first scenario (left column) and the second scenario (right column) for the prediction time Q1 (solid line), Q2 (dashed line) and Q3 (doted line).

4.2 | Simulation results

The simulation results below are organized into two parts. First, we compare the overall performance of the smooth ROC curve estimator with and without boundary correction and also compare the performance of the proposed bandwidth selection methods. Furthermore, we also explore the robustness of these estimators to the choice of the bandwidth parameters. Secondly, we compare the finite sample performances of the boundary corrected smooth ROC estimator with the empirical non-smoothed ROC estimator and the MP method of Martínez-Camblor and Pardo-Fernández[28].

4.2.1 Bandwidth selection and boundary correction

We first present the results of the simulation study conducted to investigate the finite sample behavior of the ROC curve estimator with and without boundary correction given in equation (5) and equation (8), respectively. Table 1 shows the MIB and the MISE for both estimators for the first scenario obtained with sample sizes $n = 100, 200, 400$, censoring rates 20% and 50%, $p = 0.25$, 0.50 and $t = Q_1, Q_2, Q_3$. These results clearly show that the ROC curve estimator with boundary correction has smaller MIB and MISE than the ROC curve estimator without boundary correction. This is true for all censoring rates, sample sizes, p and t values. Both the MIB and the MISE decrease with the sample size and increases with the censoring percentages. Results for the second scenario presented in Table 2 also show similar findings.

Regarding the bandwidth selection methods, the results presented in Table 1 and Table 2 indicated that all the three bandwidth selection methods produce very similar results. However, the NR method has slightly better performance for the ROC curve estimator with boundary correction whereas for the ROC estimator without boundary correction the PI method seems to be better. Globally, the PI method also shows smaller bias and, typically, it performs better for a relatively large time horizon ($t = Q_3$).

TABLE 1 MIB($\times 10^{-2}$) and MISE($\times 10^{-3}$) of the ROC estimators with and without boundary correction for different sample sizes (n), censoring rates (% cen), p and t values obtained with the **first scenario** using the three bandwidth selection methods: NR, PI and CV.

$n = 100$			ROC with boundary correction						ROC without boundary correction					
			NR		**PI**		**CV**		**NR**		**PI**		**CV**	
p	**% cen**	***t***	**MIB**	**MISE**	**MIB**	**MISE**	**MIB**	**MISE**	**MIB**	**MISE**	**MIB**	**MISE**	**MIB**	**MISE**
0.25	20	0.3	1.294	5.271	0.941	5.565	2.617	5.745	2.094	5.883	1.332	6.156	2.913	7.700
0.25	20	1.0	1.080	3.752	0.492	4.029	1.072	4.734	2.068	5.361	1.010	4.945	0.618	5.359
0.25	20	3.9	0.389	3.461	-0.063	3.193	-0.134	3.222	1.083	5.802	0.228	4.136	-0.114	3.356
0.25	50	0.3	1.825	6.384	1.513	6.695	2.682	6.681	2.438	7.300	1.775	7.565	3.074	8.747
0.25	50	1.0	2.086	6.379	1.619	6.636	1.954	6.952	2.684	8.191	1.846	7.791	1.736	8.085
0.25	50	3.9	0.509	7.486	0.203	7.192	0.072	7.357	1.018	9.838	0.390	8.237	0.034	7.491
0.50	20	0.3	1.425	5.028	0.949	5.343	3.121	5.963	2.321	6.270	1.394	6.245	2.608	8.154
0.50	20	1.0	1.064	3.513	0.421	3.609	0.567	4.231	2.020	5.766	0.926	4.736	0.254	4.172
0.50	20	3.9	0.156	2.691	-0.064	2.390	-0.106	2.387	0.443	4.050	0.051	2.900	-0.082	2.533
0.50	50	0.3	2.126	6.304	1.711	6.584	3.200	6.968	2.793	7.859	1.996	7.786	3.027	9.290
0.50	50	1.0	2.136	6.158	1.620	6.194	1.867	6.550	2.731	8.528	1.848	7.501	1.463	6.987
0.50	50	3.0	0.259	5.359	0.105	4.973	-0.012	5.042	0.546	7.113	0.212	5.701	-0.009	5.191
$n = 200$														
0.25	20	0.3	0.730	2.860	0.431	3.036	1.333	3.036	1.521	3.390	0.740	3.486	1.305	4.172
0.25	20	1.0	0.728	1.976	0.318	2.116	0.323	2.359	1.730	3.319	0.726	2.739	0.215	2.426
0.25	20	3.9	0.354	1.694	0.074	1.572	0.065	1.570	1.038	3.626	0.263	2.156	0.084	1.735
0.25	50	0.3	1.323	3.268	1.059	3.448	1.659	3.417	1.952	4.064	1.261	4.097	1.758	4.617
0.25	50	1.0	1.599	3.431	1.269	3.553	1.376	3.617	2.262	4.974	1.440	4.375	1.101	3.988
0.25	50	3.9	0.169	3.235	-0.081	3.138	-0.109	3.177	0.727	5.276	0.060	3.831	-0.099	3.356
0.50	20	0.3	0.835	2.692	0.418	2.886	1.596	3.074	1.723	3.657	0.787	3.495	0.827	4.146
0.50	20	1.0	0.724	1.840	0.278	1.891	0.205	2.036	1.665	3.688	0.652	2.640	0.200	2.091
0.50	20	3.9	0.115	1.366	-0.008	1.258	-0.009	1.245	0.337	2.279	0.034	1.480	0.017	1.413
0.50	50	0.3	1.466	3.168	1.108	3.338	1.854	3.428	2.162	4.433	1.337	4.180	1.435	4.622
0.50	50	1.0	1.526	3.324	1.154	3.324	1.208	3.391	2.159	5.343	1.315	4.239	0.988	3.573
0.50	50	3.9	-0.004	2.384	-0.125	2.266	-0.143	2.281	0.220	3.582	-0.078	2.601	-0.123	2.464
$n = 400$														
0.25	20	0.3	0.468	1.392	0.217	1.471	0.642	1.498	1.138	1.838	0.436	1.760	0.367	1.983
0.25	20	1.0	0.384	0.982	0.114	1.042	0.072	1.114	1.281	2.058	0.384	1.445	0.073	1.175
0.25	20	3.9	0.079	0.894	-0.063	0.866	-0.061	0.850	0.672	2.348	0.025	1.165	-0.042	1.012
0.25	50	0.3	0.922	1.677	0.712	1.768	0.968	1.759	1.471	2.311	0.844	2.192	0.802	2.335
0.25	50	1.0	1.051	1.737	0.828	1.785	0.842	1.819	1.707	2.995	0.947	2.328	0.718	1.964
0.25	50	3.9	-0.014	1.594	-0.155	1.609	-0.155	1.608	0.516	3.211	-0.081	1.978	-0.141	1.789
0.50	20	0.3	0.517	1.315	0.177	1.398	0.582	1.521	1.299	2.056	0.446	1.764	0.070	1.733
0.50	20	1.0	0.363	0.907	0.083	0.925	0.053	0.958	1.213	2.352	0.316	1.393	0.063	1.056
0.50	20	3.9	-0.037	0.796	-0.094	0.758	-0.093	0.747	0.117	1.355	-0.092	0.853	-0.065	0.905
0.50	50	0.3	1.004	1.628	0.724	1.707	1.007	1.748	1.636	2.606	0.878	2.246	0.574	2.102
0.50	50	1.0	0.934	1.624	0.681	1.610	0.677	1.641	1.559	3.262	0.791	2.218	0.602	1.778
0.50	50	3.9	-0.110	1.233	-0.176	1.220	-0.176	1.212	0.051	1.977	-0.170	1.344	-0.150	1.383

TABLE 2 MIB($\times 10^{-2}$) and MISE($\times 10^{-3}$) of the ROC estimators with and without boundary correction for different sample sizes (n), censoring rates (% cen), p and t values obtained with the **second scenario** using the three bandwidth selection methods: NR, PI and CV.

$n = 100$			ROC with boundary correction						ROC without boundary correction					
			NR		**PI**		**CV**		**NR**		**PI**		**CV**	
p	**% cen**	t	**MIB**	**MISE**	**MIB**	**MISE**	**MIB**	**MISE**	**MIB**	**MISE**	**MIB**	**MISE**	**MIB**	**MISE**
0.25	20	0.3	1.392	5.548	1.104	5.788	2.413	6.006	2.021	6.022	1.359	6.283	2.779	7.534
0.25	20	0.7	1.321	3.941	0.860	4.225	1.544	4.635	2.086	4.875	1.218	4.840	1.309	5.613
0.25	20	1.4	1.515	4.288	0.957	4.404	0.850	4.716	2.324	6.170	1.332	5.280	0.812	4.799
0.25	50	0.3	1.619	6.393	1.348	6.656	2.272	6.668	2.157	7.032	1.545	7.297	2.638	8.249
0.25	50	0.7	2.081	5.511	1.686	5.770	2.096	5.926	2.673	6.594	1.911	6.533	1.984	7.061
0.25	50	1.4	2.342	10.44	1.874	10.53	1.912	10.84	2.882	12.21	2.079	11.42	1.737	10.96
0.50	20	0.3	1.550	4.813	1.128	5.037	2.892	5.671	2.147	5.963	1.339	5.841	2.173	7.044
0.50	20	0.7	1.314	3.278	0.742	3.408	1.011	3.968	1.965	4.941	1.068	4.251	0.592	4.096
0.50	20	1.4	1.077	3.114	0.674	2.890	0.609	2.922	1.460	4.862	0.847	3.590	0.624	3.040
0.50	50	0.3	1.859	5.597	1.476	5.810	2.695	6.119	2.364	6.896	1.622	6.749	2.244	7.748
0.50	50	0.7	2.113	4.681	1.629	4.730	1.897	4.955	2.639	6.521	1.833	5.728	1.468	5.409
0.50	50	1.4	1.990	7.238	1.652	6.873	1.599	6.951	2.328	9.330	1.781	7.775	1.546	7.014
$n = 200$														
0.25	20	0.3	0.810	2.721	0.553	2.837	1.230	2.893	1.409	3.079	0.735	3.186	1.380	3.727
0.25	20	0.7	0.576	1.961	0.214	2.116	0.374	2.307	1.306	2.629	0.492	2.495	0.141	2.558
0.25	20	1.4	0.668	1.998	0.301	2.089	0.247	2.192	1.472	3.460	0.567	2.635	0.257	2.281
0.25	50	0.3	1.172	3.096	0.927	3.225	1.417	3.211	1.707	3.565	1.072	3.662	1.608	4.162
0.25	50	0.7	1.272	2.760	0.967	2.907	1.125	2.960	1.853	3.616	1.134	3.427	0.866	3.383
0.25	50	1.4	1.391	4.503	1.050	4.603	1.014	4.709	1.996	6.094	1.219	5.262	0.972	4.812
0.50	20	0.3	0.946	2.345	0.555	2.458	1.582	2.656	1.513	3.135	0.710	2.946	0.751	3.343
0.50	20	0.7	0.620	1.669	0.186	1.760	0.123	1.911	1.222	2.856	0.424	2.259	0.079	1.972
0.50	20	1.4	0.518	1.523	0.255	1.452	0.241	1.455	0.838	2.762	0.343	1.816	0.258	1.600
0.50	50	0.3	1.356	2.724	0.998	2.830	1.696	2.941	1.856	3.652	1.106	3.420	1.155	3.784
0.50	50	0.7	1.335	2.438	0.965	2.476	1.032	2.565	1.829	3.868	1.109	3.131	0.815	2.695
0.50	50	1.4	1.121	3.052	0.861	2.934	0.833	2.936	1.417	4.655	0.937	3.450	0.846	3.096
$n = 400$														
0.25	20	0.3	0.719	1.627	0.503	1.691	0.814	1.714	1.246	1.950	0.632	1.936	0.734	2.161
0.25	20	0.7	0.485	1.053	0.228	1.129	0.191	1.216	1.130	1.643	0.420	1.395	0.147	1.275
0.25	20	1.4	0.373	1.050	0.150	1.106	0.132	1.133	1.091	2.261	0.310	1.462	0.141	1.234
0.25	50	0.3	0.968	1.808	0.770	1.883	1.004	1.886	1.449	2.212	0.870	2.186	0.939	2.378
0.25	50	0.7	0.995	1.567	0.784	1.639	0.837	1.666	1.536	2.288	0.898	1.992	0.671	1.806
0.25	50	1.4	0.867	2.276	0.646	2.356	0.614	2.396	1.453	3.636	0.759	2.788	0.619	2.501
0.50	20	0.3	0.784	1.401	0.452	1.461	0.859	1.567	1.304	2.008	0.577	1.769	0.316	1.803
0.50	20	0.7	0.479	0.887	0.181	0.928	0.137	0.966	1.026	1.845	0.331	1.252	0.137	1.047
0.50	20	1.4	0.256	0.812	0.091	0.790	0.087	0.778	0.520	1.752	0.117	0.987	0.103	0.939
0.50	50	0.3	1.068	1.554	0.772	1.614	1.071	1.680	1.541	2.272	0.856	1.992	0.624	1.984
0.50	50	0.7	1.019	1.328	0.758	1.347	0.746	1.388	1.488	2.479	0.851	1.772	0.684	1.485
0.50	50	1.4	0.663	1.559	0.493	1.535	0.484	1.527	0.905	2.751	0.519	1.793	0.498	1.698

Now we turn to the question of robustness with respect to bandwidth selection. To investigate this problem, we reran the simulations but instead of using a data-driven bandwidth we choose the Beran's bandwidth to be $b = c_1 \times n^{-1/5}$, with $c_1 = 0.001, 0.1, 0.5, 1, 1.5, 2, 3$, and the ROC's bandwidth to be $h = c_2 \times n^{-1/3}$, with $c_2 = 1, 2, 3, 4, 5, 6$. Table 3 (Table 4) shows the MISEs obtained with the first scenario for $n = 100$ ($n = 400$). For the sake of brevity, the results for the other scenarios are not reported as they lead to the same general conclusion. In these tables, the columns show the effect of b for a fixed h and the rows show the effect of h for a fixed b. From these results, one can clearly see that both estimators are almost insensitive to the choice of b as the change in MISE is almost negligible. For example, at 20% censoring rate and $h = 100^{-1/3}$, as the b value increase by 5 folds the MISE increased by less than 0.17% for the ROC curve with boundary correction and increased by less than 2.32% for the ROC curve without boundary correction. The results also reveal that the ROC estimator without boundary correction is more sensitive to the choice of h than the ROC estimator with boundary correction. For example, at 20% censoring rate and

$b = 100^{-1/5}$, doubling the value of h decreased the MISE by less than 10% for the ROC curve with boundary correction and increased the MISE by almost 75% for the ROC curve without boundary correction. In these tables, the minimum obtained value of the MISE using the grid search is indicated in bold and the minimum data-driven MISE is indicated in italic. From these results one can observe that the data-driven bandwidth selection methods of both the bandwidth b and h perform reasonably well, especially for a large sample size and a small censoring rate. For example, for the boundary corrected smooth ROC curve estimator with $n = 100$ and 20% censoring rate, the minimum MISE value using the grid search (for both b and h) is 3.247 versus 3.725 for the data-driven bandwidths. When $n = 400$, the corresponding values are 0.964 and 0.982, respectively. This is also true for 50% censoring rate but with relatively larger differences. This suggest that, although Sheather and Jones[52] method does not take censoring into account, it results in a smooth ROC estimate close to the "optimal" one.

TABLE 3 The MISE($\times 10^{-3}$) of ROC estimators with and without boundary correction computed with different censoring rates (% cen), $\boldsymbol{n} = \mathbf{100}$, $p = 0.25$ and $t = Q2 = 1$ for the first scenario and with $b = c_1 \times n^{-1/5}$ and $h = c_2 \times n^{-1/3}$. The last two columns correspond to the NR and the PI selection methods for h and the last row ($\hat{b}$) corresponds to Sheather and Jones's selection methods for b.

% cen=20	ROC estimator **with** boundary correction							
c_1/c_2	1	2	3	4	5	6	**NR**	**PI**
0.001	4.212	3.754	3.491	3.491	3.875	4.730	3.762	4.058
0.100	4.164	3.664	3.337	**3.247**	3.515	4.238	3.631	4.011
0.500	4.171	3.701	3.400	3.338	3.638	4.398	3.687	4.032
1.000	4.165	3.718	3.436	3.398	3.730	4.527	3.723	4.037
1.500	4.157	3.723	3.454	3.432	3.786	4.609	3.739	4.033
2.000	4.147	3.720	3.461	3.453	3.823	4.665	3.743	4.025
3.000	4.133	3.717	3.475	3.490	3.887	4.758	3.750	4.014
$\hat{b}$	4.148	3.725	3.472	3.470	3.848	4.698	*3.752*	4.029
% cen=50								
0.001	6.267	5.617	5.271	5.321	5.881	6.991	5.741	6.010
0.100	6.045	5.134	4.419	3.951	**3.858**	4.255	5.067	5.866
0.500	6.110	5.362	4.781	4.463	4.545	5.130	5.374	5.969
1.000	6.228	5.578	5.078	4.862	5.086	5.841	5.674	6.111
1.500	6.405	5.801	5.356	5.222	5.559	6.446	5.949	6.294
2.000	6.597	6.023	5.623	5.555	5.980	6.965	6.203	6.490
3.000	6.923	6.394	6.071	6.111	6.668	7.790	6.606	6.820
$\hat{b}$	6.738	6.183	5.810	5.778	6.245	7.274	*6.379*	6.636
% cen=20	ROC estimator **without** boundary correction							
0.001	5.657	9.841	17.716	26.530	34.479	41.175	5.330	4.931
0.100	**5.303**	9.332	17.147	25.964	33.945	40.680	5.096	4.785
0.500	5.426	9.518	17.349	26.160	34.126	40.846	5.176	4.836
1.000	5.526	9.647	17.489	26.298	34.256	40.967	5.253	4.886
1.500	5.597	9.732	17.582	26.390	34.344	41.048	5.306	4.916
2.000	5.649	9.794	17.650	26.459	34.409	41.109	5.340	4.932
3.000	5.741	9.911	17.778	26.586	34.530	41.222	5.390	4.952
$\hat{b}$	5.679	9.828	17.686	26.494	34.443	41.140	5.361	*4.945*
% cen=50								
0.001	7.535	11.53	19.382	28.125	35.965	42.544	7.473	7.114
0.100	**5.712**	8.708	16.152	24.893	32.901	39.697	6.622	6.643
0.500	6.470	9.853	17.384	26.076	33.994	40.697	7.026	6.874
1.000	6.969	10.468	18.050	26.733	34.615	41.272	7.374	7.133
1.500	7.424	11.002	18.627	27.303	35.153	41.771	7.706	7.390
2.000	7.827	11.458	19.113	27.781	35.603	42.188	8.001	7.627
3.000	8.502	12.229	19.929	28.579	36.352	42.880	8.471	8.011
$\hat{b}$	8.072	11.724	19.385	28.042	35.846	42.412	8.191	*7.791*

TABLE 4 The MISE($\times 10^{-3}$) of ROC estimators with and without boundary correction computed with different censoring rates (% cen), $\boldsymbol{n} = \mathbf{400}$, $p = 0.25$ and $t = Q2 = 1$ for the first scenario and with $b = c_1 \times n^{-1/5}$ and $h = c_2 \times n^{-1/3}$. The last two columns correspond to the NR and the PI selection methods for h and the last row ($\hat{b}$) corresponds to Sheather and Jones's selection methods for b.

% cen=20	ROC estimator **with** boundary correction							
c_1/c_2	1	2	3	4	5	6	**NR**	**PI**
0.001	1.094	1.015	0.970	0.968	1.016	1.125	0.987	1.078
0.100	1.111	1.029	0.980	0.977	1.027	1.139	0.996	1.101
0.500	1.086	1.014	0.973	0.977	1.034	1.152	0.988	1.077
1.000	1.062	0.998	0.965	0.975	1.037	1.162	0.979	1.052
1.500	1.053	0.992	**0.964**	0.978	1.044	1.174	0.978	1.042
2.000	1.051	0.991	0.966	0.983	1.052	1.186	0.980	1.039
3.000	1.054	0.996	0.973	0.994	1.068	1.210	0.987	1.041
$\hat{\boldsymbol{b}}$	1.053	0.994	0.968	0.984	1.052	1.185	*0.982*	1.042
% cen=50								
0.001	1.819	1.657	1.519	1.406	1.334	1.333	1.590	1.762
0.100	1.598	1.449	1.328	1.251	1.227	1.265	1.367	1.590
0.500	1.465	1.360	1.284	**1.249**	1.265	1.345	1.318	1.452
1.000	1.491	1.410	1.361	1.350	1.391	1.504	1.392	1.469
1.500	1.660	1.587	1.548	1.550	1.609	1.751	1.580	1.634
2.000	1.877	1.804	1.768	1.777	1.851	2.017	1.801	1.847
3.000	2.317	2.243	2.209	2.228	2.324	2.528	2.244	2.283
$\hat{\boldsymbol{b}}$	1.813	1.741	1.705	1.713	1.782	1.941	*1.737*	1.785
% cen=20	ROC estimator **without** boundary correction							
0.001	**2.542**	4.545	7.6831	12.370	17.960	23.681	1.950	1.406
0.100	2.570	4.642	7.814	12.509	18.098	23.812	1.945	1.393
0.500	2.601	4.702	7.895	12.602	18.193	23.907	1.962	1.395
1.000	2.645	4.759	7.961	12.675	18.270	23.984	1.996	1.413
1.500	2.693	4.815	8.023	12.742	18.339	24.053	2.034	1.432
2.000	2.736	4.866	8.080	12.803	18.402	24.116	2.068	1.448
3.000	2.811	4.954	8.176	12.905	18.508	24.221	2.122	1.473
$\hat{\boldsymbol{b}}$	2.723	4.850	8.062	12.784	18.382	24.096	2.058	*1.445*
% cen=50								
0.001	2.793	4.141	6.8972	11.434	16.995	22.739	2.518	2.158
0.100	**2.760**	4.484	7.397	11.961	17.498	23.204	2.280	1.866
0.500	2.890	4.825	7.898	12.539	18.101	23.802	2.316	1.822
1.000	3.151	5.156	8.284	12.963	18.541	24.242	2.526	1.955
1.500	3.469	5.510	8.672	13.380	18.971	24.671	2.801	2.166
2.000	3.786	5.845	9.027	13.756	19.357	25.054	3.080	2.396
3.000	4.378	6.456	9.660	14.417	20.028	25.717	3.599	2.846
$\hat{\boldsymbol{b}}$	3.686	5.738	8.911	13.632	19.229	24.927	2.995	*2.328*

4.2.2 | Comparing the proposed method with the competitors

In this section, we present the results of a simulation study conducted to compare the finite sample performance of the proposed boundary corrected ROC curve estimator with the proposed empirical (non-smoothed) estimator and the MP method of Martínez-Camblor and Pardo-Fernández[28]. Table 5 (Table 6) shows the MIB and the MISE of the first (second) scenario obtained for different sample sizes, censoring rates, p values and time horizons. As can be seen from these tables, the MP method has a notably larger MIB and MISE than both the empirical (non-smoothed) and the boundary corrected smooth estimators proposed in this work; see equation (4) and equation (8), respectively. In general this is true for all censoring rates, sample sizes and time horizons. In addition, compared to the MP method, our smooth estimator has relatively faster rate of convergence as its MISE decays faster. For example, at 20% censoring rate, $p = 0.25$ and $t = 0.3$, as the sample size increases from $n = 100$ to $n = 200$ the MISE decreases by around 50% for the smooth ROC estimator with boundary correction and decreases only by around 20%

for the MP method. Furthermore, we also observed that the variance of the integrated squared errors of the MP method is larger than both the proposed smooth and the empirical (non-smoothed) estimators (the results are not reported here) which is consistent with the results presented in Table 5 and Table 6. Finally, these tables show that the proposed boundary corrected ROC estimator has smaller MISE than the non-smoothed (empirical) one except for a few cases where the prediction time is relatively large (i.e., $t = Q_3$). In such a situation, the plug-in bandwidth method (PI) seems to be more adapted as it performs better than the NR method (see Tables 1 and 2) and gives results that are very close to the empirical estimator. This is consistent with the finding of a simulation study by Lloyd and Yong[23] that was conducted to compare the empirical ROC estimator with a kernel based estimator in the case of uncensored data.

In conclusion, we can say that the proposed boundary corrected smoothed time-dependent ROC estimator has better performance than both the proposed empirical (non-smoothed) estimator and the MP method.

TABLE 5 The MIB($\times 10^{-2}$) and MISE($\times 10^{-3}$) of the empirical, the smooth (boundary corrected), and the MP estimators computed for different sample sizes (n), censoring rates (% cen), p and t values obtained with the **first scenario**. The bandwidth(s) are selected using the NR method for our estimator and the LSCV and the PI methods for the MP estimator.

			Empirical unsmoothed		Proposed		MP			
$n = 100$					NR		LSCV		PI	
p	**% cen**	***t***	**MIB**	**MISE**	**MIB**	**MISE**	**MIB**	**MISE**	**MIB**	**MISE**
0.25	20	0.3	0.197	7.454	1.294	5.271	3.337	13.288	4.656	10.855
0.25	20	1.0	0.184	4.674	1.080	3.752	3.963	13.073	8.872	21.811
0.25	20	3.9	-0.076	3.141	0.389	3.461	4.464	13.437	11.02	35.811
0.25	50	0.3	0.817	8.654	1.825	6.384	2.995	9.5497	7.773	15.740
0.25	50	1.0	1.238	7.395	2.086	6.379	5.074	12.281	11.14	28.206
0.25	50	3.9	0.062	7.271	0.509	7.486	9.049	22.848	15.12	51.266
0.50	20	0.3	0.202	6.863	1.425	5.028	3.627	14.964	4.722	12.394
0.50	20	1.0	0.198	3.925	1.064	3.513	3.914	13.506	9.203	25.066
0.50	20	3.9	-0.025	2.196	0.156	2.691	4.330	11.962	10.48	35.452
0.50	50	0.3	0.989	8.154	2.126	6.304	3.958	11.793	9.025	20.339
0.50	50	1.0	1.295	6.591	2.136	6.158	6.111	14.837	12.48	35.521
0.50	50	3.9	0.041	4.857	0.259	5.359	9.904	25.819	15.58	55.915
$n = 200$										
0.25	20	0.3	0.026	3.837	0.730	2.860	3.860	10.555	5.098	9.8133
0.25	20	1.0	0.214	2.332	0.728	1.976	4.408	11.576	9.052	21.026
0.25	20	3.9	0.123	1.481	0.354	1.694	4.901	12.906	10.45	33.018
0.25	50	0.3	0.640	4.277	1.323	3.268	3.243	8.0379	7.255	13.896
0.25	50	1.0	1.058	3.874	1.599	3.431	5.112	11.524	10.11	25.000
0.25	50	3.9	-0.068	3.116	0.169	3.235	8.447	21.377	13.32	44.034
0.50	20	0.3	0.050	3.525	0.835	2.692	4.522	12.717	5.324	11.686
0.50	20	1.0	0.224	1.975	0.724	1.840	4.898	13.554	9.502	24.762
0.50	20	3.9	0.075	1.040	0.115	1.366	5.001	13.317	9.973	32.953
0.50	50	0.3	0.693	4.004	1.466	3.168	4.163	10.069	8.588	18.643
0.50	50	1.0	0.983	3.489	1.526	3.324	6.158	14.420	11.50	32.253
0.50	50	3.9	-0.070	2.108	-0.004	2.384	9.286	24.232	13.85	48.584
$n = 400$										
0.25	20	0.3	0.008	1.766	0.468	1.392	4.678	10.024	5.377	9.5197
0.25	20	1.0	0.091	1.103	0.384	0.982	5.655	13.474	9.018	20.559
0.25	20	3.9	0.003	0.729	0.079	0.894	5.834	16.270	9.859	30.817
0.25	50	0.3	0.466	2.104	0.922	1.677	3.925	8.3241	6.873	12.733
0.25	50	1.0	0.716	1.911	1.051	1.737	5.875	13.357	9.418	22.864
0.25	50	3.9	-0.102	1.523	-0.014	1.594	8.649	23.373	12.08	38.910
0.50	20	0.3	0.007	1.612	0.517	1.315	5.456	14.057	5.768	11.603
0.50	20	1.0	0.087	0.927	0.363	0.907	6.309	17.167	9.596	24.661
0.50	20	3.9	-0.003	0.506	-0.037	0.796	5.898	16.969	9.442	30.803
0.50	50	0.3	0.488	1.970	1.004	1.628	5.098	11.376	8.244	17.489
0.50	50	1.0	0.601	1.674	0.934	1.624	7.062	17.313	10.82	29.878
0.50	50	3.9	-0.093	1.001	-0.110	1.233	9.438	26.390	12.63	43.121

TABLE 6 The MIB($\times 10^{-2}$) and MISE($\times 10^{-3}$) of the empirical, the smooth (boundary corrected), and the MP estimators computed for different sample sizes (n), censoring rates (% cen), p and t values obtained with the **second scenario**. The bandwidth(s) are selected using the NR method for our estimator and the LSCV and the PI methods for the MP estimator.

			Empirical unsmoothed		Proposed		MP			
$n = 100$					NR		LSCV		PI	
p	% cen	t	MIB	MISE	MIB	MISE	MIB	MISE	MIB	MISE
0.25	20	0.3	0.281	7.788	1.392	5.548	0.915	4.969	1.141	3.533
0.25	20	0.7	0.470	5.064	1.321	3.941	2.503	4.541	3.000	4.593
0.25	20	1.4	0.815	4.723	1.515	4.288	4.470	7.180	5.014	7.953
0.25	50	0.3	0.588	8.632	1.619	6.393	1.902	5.926	2.067	4.977
0.25	50	0.7	1.292	6.566	2.081	5.511	4.247	7.316	4.404	7.134
0.25	50	1.4	1.699	10.76	2.342	10.44	7.931	16.13	8.097	16.25
0.50	20	0.3	0.292	6.458	1.550	4.813	0.846	4.019	1.057	2.833
0.50	20	0.7	0.467	3.802	1.314	3.278	2.780	4.092	3.324	4.468
0.50	20	1.4	0.660	2.859	1.077	3.114	4.907	7.956	5.400	8.874
0.50	50	0.3	0.678	7.169	1.859	5.597	2.730	5.841	3.024	5.094
0.50	50	0.7	1.311	5.073	2.113	4.681	5.686	9.075	5.910	9.159
0.50	50	1.4	1.567	6.778	1.990	7.238	9.580	20.25	9.786	20.72
$n = 200$										
0.25	20	0.3	0.068	3.641	0.810	2.721	0.623	2.506	0.837	2.054
0.25	20	0.7	0.041	2.439	0.576	1.961	1.788	2.438	2.294	2.776
0.25	20	1.4	0.276	2.199	0.668	1.998	3.602	4.450	4.125	5.341
0.25	50	0.3	0.462	4.025	1.172	3.096	1.159	3.139	1.407	2.896
0.25	50	0.7	0.753	3.243	1.272	2.760	3.301	4.343	3.533	4.546
0.25	50	1.4	0.984	4.721	1.391	4.503	6.484	10.49	6.720	10.99
0.50	20	0.3	0.086	3.033	0.946	2.345	0.655	2.044	0.857	1.693
0.50	20	0.7	0.093	1.899	0.620	1.669	2.189	2.322	2.704	2.863
0.50	20	1.4	0.295	1.376	0.518	1.523	4.225	5.683	4.680	6.619
0.50	50	0.3	0.529	3.388	1.356	2.724	1.892	3.130	2.333	3.174
0.50	50	0.7	0.814	2.623	1.335	2.438	4.680	5.825	5.040	6.476
0.50	50	1.4	0.876	2.868	1.121	3.052	8.067	14.31	8.376	15.29
$n = 400$										
0.25	20	0.3	0.230	2.027	0.719	1.627	0.756	1.480	0.916	1.333
0.25	20	0.7	0.160	1.247	0.485	1.053	1.676	1.542	2.136	1.944
0.25	20	1.4	0.161	1.122	0.373	1.050	3.258	3.327	3.723	4.127
0.25	50	0.3	0.500	2.230	0.968	1.808	0.717	1.924	0.980	1.870
0.25	50	0.7	0.672	1.775	0.995	1.567	2.657	2.912	2.921	3.211
0.25	50	1.4	0.635	2.389	0.867	2.276	5.257	7.301	5.523	7.842
0.50	20	0.3	0.209	1.689	0.784	1.401	0.839	1.211	0.972	1.110
0.50	20	0.7	0.159	0.961	0.479	0.887	2.120	1.605	2.557	2.124
0.50	20	1.4	0.142	0.690	0.256	0.812	3.987	4.897	4.357	5.657
0.50	50	0.3	0.514	1.850	1.068	1.554	1.465	1.924	1.878	2.117
0.50	50	0.7	0.690	1.403	1.019	1.328	4.094	4.310	4.448	4.998
0.50	50	1.4	0.531	1.455	0.663	1.559	6.986	11.13	7.283	12.03

4.2.3 | Validity of the bootstrap

In this section, we investigate the validity of the bootstrap procedures described in Section 2.5 to estimate the variance and construct confidence intervals for the time-dependent AUC. For that, form each of the 1000 simulated data, we draw $B = 2000$ bootstrap samples and then calculate the bootstrap variance S_B^2 and the two confidence intervals (standard normal and percentile) using the formulas provided in Section 2.5. The consistency of S_B will be assessed by comparing its average (here after ASD_B) with the simulation-based empirical standard-deviation $ESD = \sqrt{1000^{-1}\sum_l \left(\widehat{AUC}_l - 1000^{-1}\sum_l \widehat{AUC}_l\right)^2}$, where $\widehat{AUC}_l$ is the AUC estimate from the lth simulated data. On the other hand, the validity and precision of the proposed confidence intervals will be assessed by calculating their coverage probabilities (CP) and their average widths (AW), respectively. The former refers to the proportion of times that the confidence interval encloses the true AUC value and the later is the average difference between the lower and upper limits of the confidence intervals. Table 7 shows the obtained results for data generated according to the first scenario with sample sizes $n = 100, 400$, censoring rates 20% and 50%, $p = 0.25, 0.5$ and $t = 0.3, 1.0, 3.9$. From this table, we can first notice that ASD_B increases with the censoring rate and decrease with the sample size. Also, ASD_B remains close to ESD and the difference between these two quantities decreases as sample size increases. Next, regarding the confidence intervals, we notice that, globally, for all the cases, the coverage probabilities are close to the nominal level of 95% and improve (deteriorate) as n (% of censoring) increases. On the other hand, the average widths decrease (increase) with sample size (censoring). The percentile method lead to the best performances, both in terms of coverage probability and average width. The classical normal method gives some unsatisfactory results when the sample size is small and/or the percentage of censoring is high. These results clearly demonstrate that the proposed bootstrap provides a good approximation of the sampling distribution of $\widehat{AUC}$.

TABLE 7 The empirical standard deviation (ESD), the average bootstrap standard deviation (ASD_B), the coverage probability (CP) and the average width (AW) of 95%−bootstrap confidence intervals obtained with the smooth estimator using the standard normal method (Normal) and the percentile method computed for different sample sizes (n), censoring rates (% cen), p and t values. Data simulated according to the **first scenario.**

$n = 100$					**Normal**		**Percentile**	
p	**% cen**	**t**	**ESD**	**ASD_B**	**CP**	**AW**	**CP**	**AW**
0.25	20	0.3	0.061	0.062	0.943	0.243	0.945	0.243
0.25	20	1.0	0.050	0.050	0.942	0.194	0.945	0.193
0.25	20	3.9	0.044	0.043	0.937	0.168	0.946	0.166
0.25	50	0.3	0.067	0.066	0.936	0.261	0.943	0.260
0.25	50	1.0	0.063	0.059	0.940	0.233	0.944	0.232
0.25	50	3.9	0.066	0.061	0.890	0.240	0.930	0.236
0.50	20	0.3	0.059	0.059	0.944	0.233	0.945	0.232
0.50	20	1.0	0.047	0.046	0.934	0.180	0.935	0.179
0.50	20	3.9	0.037	0.036	0.925	0.140	0.939	0.139
0.50	50	0.3	0.065	0.064	0.933	0.249	0.932	0.248
0.50	50	1.0	0.058	0.055	0.928	0.216	0.937	0.214
0.50	50	3.9	0.054	0.050	0.890	0.197	0.917	0.194
$n = 400$								
0.25	20	0.3	0.030	0.031	0.956	0.122	0.959	0.122
0.25	20	1.0	0.024	0.025	0.956	0.096	0.952	0.096
0.25	20	3.9	0.020	0.020	0.940	0.079	0.948	0.079
0.25	50	0.3	0.033	0.034	0.956	0.131	0.955	0.131
0.25	50	1.0	0.029	0.029	0.943	0.115	0.954	0.115
0.25	50	3.9	0.029	0.028	0.928	0.109	0.944	0.108
0.50	20	0.3	0.029	0.030	0.954	0.116	0.958	0.116
0.50	20	1.0	0.022	0.023	0.957	0.089	0.955	0.088
0.50	20	3.9	0.017	0.017	0.939	0.067	0.947	0.066
0.50	50	0.3	0.031	0.032	0.954	0.125	0.953	0.125
0.50	50	1.0	0.027	0.027	0.942	0.105	0.948	0.105
0.50	50	3.9	0.024	0.023	0.922	0.090	0.945	0.089

5 | REAL DATA ANALYSIS

To illustrate our method, we apply it to a real data set from a randomized placebo-controlled trial of the drug D-penicillamine for treating primary biliary cirrhosis (PBC) of the liver conducted at the Mayo Clinic between 1974 and 1984. A total of 312 subjects were randomized to the treatments and among these 125 died by the end of the follow-up and the remaining are censored subjects. Heagerty and Zheng[4] used this data to illustrate their proposed time-dependent ROC curve estimator. They fitted a Cox model and developed two prognostic scores. This data is available in the R package `survivalROC`.[55] For the purpose of demonstration, we use here the prognostic score derived by Heagerty and Zheng using the following five variables: log(bilirubin), albumin, log(prothrombin time), edema, and age. Figure 2 displays the time-dependent ROC curves estimated using the proposed boundary corrected smooth ROC curve estimator, the MP method, and the proposed empirical (non-smoothed) method evaluated at the prediction times $t = 3, 6$ years. The boundary corrected smooth ROC estimator was calculated using the NR bandwidth selection method. As for the MP method, we used the non-diagonal least squares cross-validation method. The estimated ROC curves for the boundary corrected smooth and the empirical (non-smoothed) methods are very close to each other, whereas the MP method results in a smaller ROC, especially at 6 years prediction time. Table 8 shows the estimated time-dependent AUC with the corresponding 95% bootstrap confidence intervals for the proposed smoothed and empirical methods. The reported confidence intervals were calculated based on 2000 bootstrap samples using either the standard normal approach or the percentile method. Regarding the importance of the marker's discrimination ability, the obtained confidence intervals reveals similar decision about the significance of the AUC values, since the "null" value (i.e. 0.5) is not contained in the intervals.

TABLE 8 The estimated AUC value with 95% bootstrap confidence intervals using the percentile method and the standard normal method (Normal) from both the proposed smoothed and empirical (Non-smoothed) methods for $t = 3, 6$ years.

Proposed smoothed		**95% bootstrap confidence interval**	
t	**Estimate**	**Percentile**	**Normal**
3	0.892	[0.782, 0.897]	[0.781, 0.897]
6	0.873	[0.719, 0.845]	[0.719, 0.843]
Empirical (Non-smoothed)			
3	0.897	[0.786, 0.901]	[0.787, 0.903]
6	0.877	[0.722, 0.848]	[0.723, 0.848]

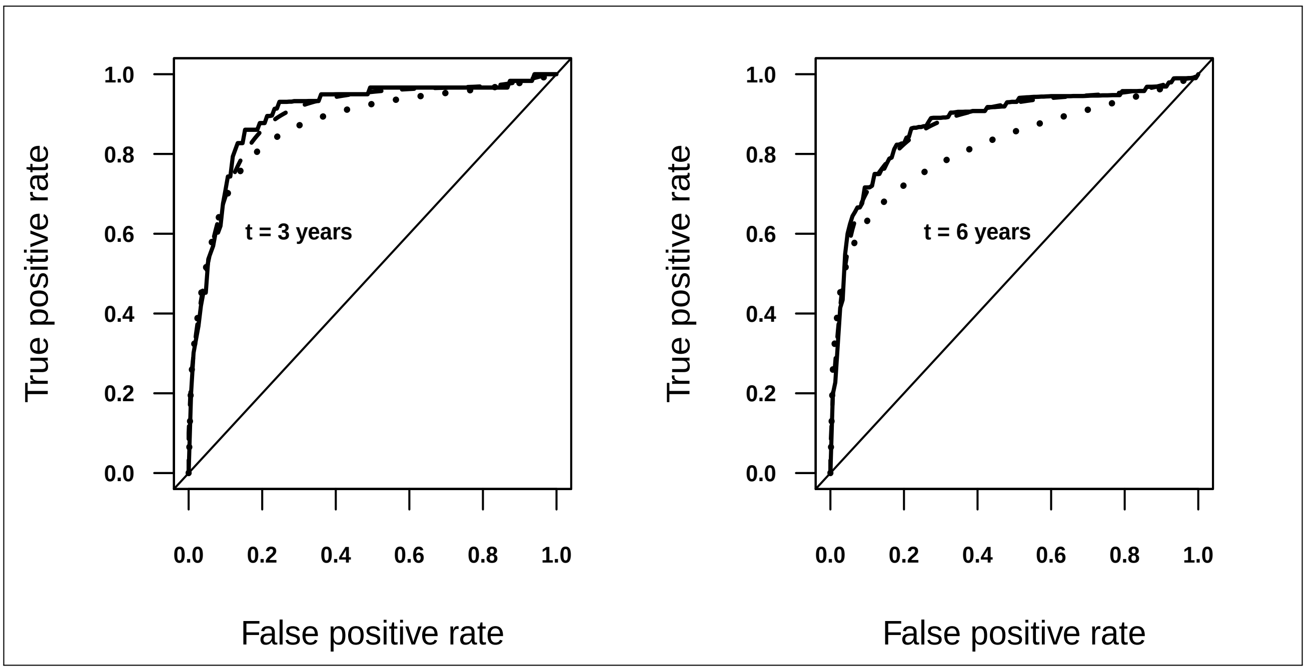

FIGURE 2 Estimated time-dependent ROC curve from the empirical (solid line), boundary corrected smooth (dashed line) and the MP method (doted line) for 3 years (left) and 6 years (right) prediction time.

6 | DISCUSSION

In this article, we proposed and investigated a nonparametric smoothed time-dependent ROC curve estimators based on a weighted kernel smoothing technique for right censored time-to-event data. Our method relies on well-known kernel techniques used to estimate cumulative distribution functions of random variables with bounded support. We derived some asymptotic properties of the proposed estimators and suggested a boundary correction based on a quantile transformation method. An important problem in conducting kernel based estimation is the choice of bandwidth. The proposed smooth estimators require selecting two bandwidth parameters: one for Beran's weight calculation and one for smoothing the ROC curve itself. To select the latter, we proposed three data-driven methods: the normal reference (NR), the plug-in (PI) and the cross-validation (CV). In order to estimate the variability and construct confidence intervals for the time-dependent ROC curve and its corresponding AUC, we proposed a bootstrap method.

Our simulation results show better performance, in terms of the MIB and MISE, for the boundary corrected smooth ROC estimator compared to the one without the boundary correction. Results for comparing the robustness of these estimators suggest that the latter is relatively highly sensitive to the choice of the bandwidth parameter needed for smoothing the ROC curve. However, both these estimators are almost insensitive to the choice the bandwidth needed for Beran's weight calculation. We also compared the performance of the three proposed bandwidth selection methods. From the results, we found that these methods produce very similar results with slightly better performances of the NR (PI) method for the ROC estimator with (without) boundary correction. The simulation results also show that the proposed smooth estimators lead, in general, to better results compared to the method of Martínez-Camblor and Pardo-Fernández[28], which is, to the best of our knowledge, the only available competitor in the literature. Regarding the performance of the proposed bootstrap confidence intervals estimation approach, we found that, globally, for all the cases, the coverage probabilities are close to the nominal level of 95%. The percentile method lead to the best performances, both in terms of coverage probability and average width, compared to the classical normal method which gives some unsatisfactory results when the sample size is small and/or the percentage of censoring is high. The results clearly demonstrate that the proposed bootstrap provides a good approximation of the sampling distribution of $\widehat{AUC}$. The methods also illustrated with a real data analysis from the PBC dataset. Finally, it is important to note that, in practice, a marker is typically built from the observed data using a given statistical model (Cox, AFT, ...). In such a situation, before the model/marker can be really used, the discrimination analysis, based on the ROC or the AUC, should always be coupled with a good calibration algorithm.[56,57]

In this study we only considered a quantile transformation technique to correct the boundary problem. However, considering other boundary correcting approaches, such as the reflection method, and comparing its performance with the one used in this paper is an interesting open problem. Furthermore, selecting the pilot bandwidth g of the plug-in method, see equation (11), based on the asymptotic mean square error is also another interesting research topic. Finally, extending and investigating the proposed smoothing method to other metrics, such as standardized net benefit and net reclassification improvement at the event rate, would be an interesting future research topic.

Finally, the proposed method is implemented in the open-source `R`-package `cenROC`, which is publicly available from the GitHub repository (https://github.com//elghouch/cenROC).

ACKNOWLEDGMENTS

Kassu Mehari Beyene would like to thank the National Fund for Scientific Research (FRNS) of Belgium grant PDR(FRNS)-T0080.16 for supporting financially.

CONFLICT OF INTEREST

The authors have declared no conflict of interests.

REFERENCES

1. Pepe MS. *The statistical evaluation of medical tests for classification and prediction*. Oxford Statistical Sciences Series . 2003.
2. Krzanowski WJ, Hand DJ. *ROC csurves for continuous data*. Chapman and Hall. 1st ed. 2009.
3. Heagerty PJ, Lumley T, Pepe MS. Time-dependent ROC curves for censored survival data and a diagnostic marker. *Biometrics* 2000; 56(2): 337–344.
4. Heagerty PJ, Zheng Y. Survival model predictive accuracy and ROC curves. *Biometrics* 2005; 61(1): 92–105.
5. Etzioni R, Pepe M, Longton G, Hu C, Goodman G. Incorporating the time dimension in receiver operating characteristic curves: A case study of prostate cancer. *Medical Decision Making* 1999; 19(3): 242 – 251.
6. Slate EH, Turnbull BW. Statistical models for longitudinal biomarkers of disease onset. *Statistics in Medicine* 2000; 19(4): 617 – 637.
7. Lambert J, Chevret S. Summary measure of discrimination in survival models based on cumulative/dynamic time-dependent ROC curves. *Statistical methods in medical research* 2016; 25(5): 2088–2102.
8. Blanche P, Dartigues JF, Jacqmin-Gadda H. Review and comparison of ROC curve estimators for a time-dependent outcome with marker-dependent censoring. *Biometrical Journal* 2013; 55(5): 687–704.
9. Akritas MG. Nearest neighbor estimation of a bivariate distribution under random censoring. *The Annals of Statistics* 1994; 22(3): 1299–1327.
10. Chambless LE, Diao G. Estimation of time-dependent area under the ROC curve for long-term risk prediction. *Statistics in medicine* 2006; 25(20): 3474–3486.
11. Martínez-Camblor P, Bayón GF, Pérez-Fernández S. Cumulative/dynamic ROC curve estimation. *Journal of Statistical Computation and Simulation* 2016; 86(17): 3582 – 3594.
12. Li L, Greene T, Hu B. A simple method to estimate the time-dependent receiver operating characteristic curve and the area under the curve with right censored data. *Statistical Methods in Medical Research* 2018; 27(8): 2264-2278.
13. Uno H, Cai T, Tian L, Wei L. Evaluating prediction rules for t-year survivors with censored regression models. *Journal of the American Statistical Association* 2007; 102(478): 527–537.
14. Hung H, Chiang CT. Estimation methods for time-dependent AUC models with survival data. *Canadian Journal of Statistics* 2010; 38(1): 8–26.
15. Wolf P, Schmidt G, Ulm K. The use of ROC for defining the validity of the prognostic index in censored data. *Statistics & Probability Letters* 2011; 81(7): 783 – 791.
16. Kamarudin AN, Cox T, Kolamunnage-Dona R. Time-dependent ROC curve analysis in medical research: current methods and applications. *BMC medical research methodology* 2017; 17(1): 53.
17. Sima CS, Gönen M. Optimal cutpoint estimation with censored data. *Journal of Statistical Theory and Practice* 2013; 7(2): 345-359.
18. López-Ratón M, Rodríguez-Álvarez M, Cadarso-Suárez C, Gude-Sampedro F. OptimalCutpoints: An R package for selecting optimal cutpoints in diagnostic tests. *Journal of Statistical Software, Articles* 2014; 61(8): 1–36.
19. Fluss R, Faraggi D, Reiser B. Estimation of the Youden Index and its associated cutoff point. *Biometrical Journal* 2005; 47(4): 458–472.
20. Hsieh F, Turnbull BW. Nonparametric methods for evaluating diagnostic tests. *Statistica Sinica* 1996: 47–62.

21. Perkins J, Schisterman F. The inconsistency of “optimal” cut-points using two ROC based criteria. *Am J Epidemiol* 2006; 163: 670–675.

22. Hwang YT, Hung YH, Wang CC, Terng HJ. Finding the optimal threshold of a parametric ROC curve under a continuous diagnostic measurement. *REVSTAT–Statistical Journal* 2018; 16(1): 23–43.

23. Lloyd CJ, Yong Z. Kernel estimators of the ROC curve are better than empirical. *Statistics & Probability Letters* 1999; 44(3): 221 – 228.

24. Reiss RD. Nonparametric estimation of smooth distribution functions. *Scandinavian Journal of Statistics* 1981; 8(2): 116–119.

25. Falk M. Relative efficiency and deficiency of kernel type estimators of smooth distribution functions. *Statistica Neerlandica* 1983; 37(2): 73–83.

26. Swanepoel JW. Mean intergrated squared error properties and optimal kernels when estimating a diatribution function. *Communications in Statistics - Theory and Methods* 1988; 17(11): 3785–3799.

27. Pulit M. A new method of kernel-smoothing estimation of the ROC curve. *Metrika* 2016; 79(5): 603–634.

28. Martínez-Camblor P, Pardo-Fernández JC. Smooth time-dependent receiver operating characteristic curve estimators. *Statistical Methods in Medical Research* 2018; 27(3): 651 – 674.

29. Beran R. Nonparametric regression with randomly censored survival data. tech. rep., Technical Report, Univ. California, Berkeley; 1981.

30. Dabrowska DM. Uniform consistency of the kernel conditional Kaplan-Meier estimate. *The Annals of Statistics* 1989; 17(3): 1157-1167.

31. Gonzalez-Manteiga W, Cadarso-Suarez C. Asymptotic properties of a generalized kaplan-meier estimator with some applications. *Journal of Nonparametric Statistics* 1994; 4(1): 65–78.

32. Van Keilegom I, Veraverbeke N. Estimation and bootstrap with censored data in fixed design nonparametric regression. *Annals of the Institute of Statistical Mathematics* 1997; 49(3): 467–491.

33. Nadaraya EA. Some new estimates for distribution functions. *Theory of Probability & Its Applications* 1964; 9(3): 497–500.

34. Van Keilegom I, Akritas MG. Transfer of tail information in censored regression models. *Annals of Statistics* 1999: 1745–1784.

35. Li Q, Racine JS. *Nonparametric econometrics: theory and practice*. Princeton University Press . 2007.

36. Karunamuni RJ, Alberts T. On boundary correction in kernel density estimation. *Statistical Methodology* 2005; 2(3): 191–212.

37. Tenreiro C. Boundary kernels for distribution function estimation. *REVSTAT Statistical Journal* 2013; 11: 169–190.

38. Koláček J, Karunamuni RJ. On boundary correction in kernel estimation of ROC curves. *Austrian Journal of Statistics* 2016; 38(1): 17–32.

39. Wand MP, Marron JS, Ruppert D. Transformations in density estimation. *Journal of the American Statistical Association* 1991; 86(414): 343–353.

40. Marron JS, Ruppert D. Transformations to reduce boundary bias in kernel density estimation. *Journal of the Royal Statistical Society. Series B (Methodological)* 1994: 653–671.

41. Geenens G. Probit transformation for kernel density estimation on the unit interval. *Journal of the American Statistical Association* 2014; 109(505): 346–358.

42. Wen K, Wu X. An improved transformation-based kernel estimator of densities on the unit interval. *Journal of the American Statistical Association* 2015; 110(510): 773–783.

43. Efron B. Bootstrap methods: another look at the Jackknife. *The Annals of Statistics* 1979; 7(1): 1–26.

44. Efron B. Nonparametric standard errors and confidence intervals. *Canadian Journal of Statistics* 1981; 9(2): 139–158.

45. Jones MC. The performance of kernel density functions in kernel distribution function estimation. *Statistics & Probability Letters* 1990; 9(2): 129–132.

46. Dabrowska DM. Non-parametric regression with censored survival time data. *Scandinavian Journal of Statistics* 1987; 14(3): 181–197.

47. Polansky AM, Baker ER. Multistage plug—in bandwidth selection for kernel distribution function estimates. *Journal of Statistical Computation and Simulation* 2000; 65(1-4): 63–80.

48. Silverman BW. *Density estimation for statistics and data analysis*. London: Chapman and Hall . 1986.

49. Sarda P. Smoothing parameter selection for smooth distribution functions. *Journal of Statistical Planning and Inference* 1993; 35(1): 65 – 75.

50. Altman N, Léger C. Bandwidth selection for kernel distribution function estimation. *Journal of Statistical Planning and Inference* 1995; 46(2): 195 – 214.

51. Bowman A, Hall P, Trvan T. Bandwidth selection for the smoothing of distribution functions. *Biometrika* 1998; 85: 799 – 808.

52. Sheather SJ, Jones MC. A reliable data-based bandwidth selection method for kernel density estimation. *Journal of the Royal Statistical Society. Series B (Methodological)* 1991; 53(3): 683–690.

53. R Core Team . *R: A language and environment for statistical computing*. R Foundation for Statistical Computing; Vienna, Austria: 2020.

54. Diaz-Coto S. *smoothROCtime: Smooth time-dependent ROC curve estimation*. 2018. R package version 0.1.0.

55. Heagerty PJ, Saha-Chaudhuri P. *survivalROC: Time-dependent ROC curve estimation from censored survival data*. 2013. R package version 1.0.3.

56. Cook NR. Statistical evaluation of prognostic versus diagnostic models: beyond the ROC curve. *Clinical chemistry* 2008; 54(1): 17–23.

57. Van Calster B, McLernon DJ, Van Smeden M, Wynants L, Steyerberg EW. Calibration: the Achilles heel of predictive analytics. *BMC medicine* 2019; 17(1): 1–7.

□