Electronic Journal of Statistics
Vol. 16 (2022) 4133–4161
ISSN: 1935-7524
https://doi.org/10.1214/22-EJS2042

Robust consistent estimators for ROC curves with covariates*

Ana M. Bianco, Graciela Boente

Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires and CONICET, Argentina
e-mail: abianco@dm.uba.ar; gboente@dm.uba.ar
and

Wenceslao González–Manteiga

Facultad de Matemática, Universidad de Santiago de Compostela, Spain
e-mail: wenceslao.gonzalez@usc.es

Abstract: The Receiver Operating Characteristic (ROC) curve is a useful tool to measure the classification capability of a continuous variable to assess the accuracy of a medical test that distinguishes between two conditions. Sometimes, covariates related to the diagnostic variable may increase the discriminating power of the ROC curve. Due to the lack of stability of classical ROC curves estimators to outliers, we introduce a procedure to obtain robust estimators in presence of covariates. The considered proposal focusses on a semiparametric approach which robustly fits a location-scale regression model to the diagnostic variable and considers robust adaptive empirical estimators of the regression residuals. The uniform consistency of the proposal is derived under mild assumptions. A Monte Carlo study is carried out to compare the performance of the robust proposed estimators with the classical ones both, in clean and contaminated samples. A real data set is also analysed.

MSC2020 subject classifications: Primary 62G35; secondary 62G05.
Keywords and phrases: Covariates, robustness, ROC curves, parametric regression.

Received November 2021.

Contents

1 Introduction 4134
2 Proposal 4137
2.1 Preliminaries 4137
2.2 The general procedure 4139
3 Consistency results 4141
4 Monte Carlo study 4144
5 Analysis of real data set 4148

*This research was partially supported by Grants PICT 2018-00740 from ANPCYT and 20020170100022BA from the Universidad de Buenos Aires, Argentina and also by the Spanish Projects MTM2016-76969P and PID2020-116587GB-I00 from the from the Ministry of Science and Innovation (MCIN/ AEI/FEDER, UE), Spain.

6 Final remarks 4151
A Appendix 4152
A.1 Proof of Theorem 1 4152
A.2 Proof of Propositions 1 and 2 4155
Acknowledgments 4160
Supplementary Material 4160
References 4160

1. Introduction

The Receiver Operating Characteristic (ROC) curve is a useful tool to size up the capability of a continuous variable or the accuracy of a pharmaceutical or medical test to distinguish between two conditions. ROC curves are a very well known technique in medical studies where a continuous variable or marker (biomarker) is used to diagnose a disease or to evaluate the progression of a disease. The use of ROC curves has become more and more popular in medicine from the early 60's due to their applications in classification or discrimination (see Gonçalves *et al.*, 2014, for a historical note and Krzanowski and Hand, 2009 for further details). Thinking of this scenario, we assume that we deal with two populations, henceforth, identified as diseased (D) and healthy (H) and that a continuous score usually called *biomarker* or *diagnostic variable*, Y, is considered for the assignment purpose and whose rule is based on a cut–off value c. Thus, according to this assignment rule, an individual is classified as diseased if $Y \geq c$ and as healthy when $Y < c$. Let F_D be the distribution of the marker on the diseased population and F_H the distribution of Y in the healthy one. From now on, for practical reasons, we denote as $Y_D \sim F_D$ the marker in the diseased population and $Y_H \sim F_H$ the score in the healthy one. It is worth mentioning that the decision of classifying an individual as diseased when the biomarker is larger or equal than c, is a natural one when Y_D is stochastically greater than Y_H, that is, when $\mathbb{P}(Y_D \leq c) \leq \mathbb{P}(Y_H \leq c)$ for all c. It is clear that the classification errors depend on the threshold c. Therefore, it becomes of interest to study the pairs $\{(1-F_H(c), 1-F_D(c)),\ c \in \mathbb{R}\}$, which describe a geometrical object called ROC curve that reflects the discriminatory capability of the biomarker. This suggests a different parametrization of this curve in terms of the false positive rate, $1 - F_H(c)$, leading to $\{(p, 1 - F_D(F_H^{-1}(1-p))),\ p \in (0,1)\}$ and therefore, to $\mathrm{ROC}(p) = 1 - F_D(F_H^{-1}(1-p))), \quad p \in (0,1)$. In this manner, the ROC curve is a complete picture of the performance of the assignment procedure over all possible threshold values. As it is well known, when Y_D is stochastically greater than Y_H, then $\mathrm{ROC}(p) \geq p$, for any $0 < p < 1$ and the discriminatory capability of the biomarker may be reflected on how far above from the identity function is the ROC curve.

In practical situations, the discriminatory effectiveness of the biomarker may be improved by several factors. Thus, when for each individual there is additional information contained in measured covariates, it is sensible to include them in the ROC analysis. Pepe (2003) illustrates how the discriminatory capability of

a test is improved by the presence of covariates, through several examples. For an overview on this topic, we refer to Pardo-Fernández *et al.* (2014). In brief, we may say that the information registered all along the covariates may impact the discrimination capability of the marker. In this situation, in order to have a deeper comprehension of the effect of the covariates, it would be advisable to incorporate to the ROC curve the additional information they provide instead of considering a *joint* ROC curve, that may lead to oversimplification. This issue can be accomplished in different ways. In the direct methodology, the ROC curve is directly regressed onto the covariates by means of a generalized linear model. Among others, Alonzo and Pepe (2002), Pepe (2003) and Cai (2004) follow this approach. In contrast, in the induced methodology, the markers distribution in each population is modelled separately in terms of the covariates and just after, the induced ROC curve is computed. The papers by Pepe (1998), Faraggi (2003), González-Manteiga *et al.* (2011) and Rodríguez-Álvarez *et al.* (2011a) go in this direction, while Rodríguez-Álvarez *et al.* (2011b) perform a comparative study of the direct and induced methodologies. From now on, we denote as $\mathbf{X}_D$ and $\mathbf{X}_H$ the covariates for the disease and healthy populations and we assume that they have the same dimension. In such case, for any $\mathbf{x}$ in the common support of $\mathbf{X}_D$ and $\mathbf{X}_H$, the conditional ROC curve is defined as

$$\mathrm{ROC}_{\mathbf{x}}(p) = 1 - F_D(F_H^{-1}(1-p|\mathbf{x})|\mathbf{x})\,, \tag{1}$$

where $F_j(\cdot|\mathbf{x})$ stands for conditional distribution of $Y_j|\mathbf{X}_j = \mathbf{x}$, $j = H, D$. In this paper, we focus on the induced approach through a general regression model.

The general methodology to estimate the conditional ROC curve consists in a plug–in procedure, where estimators of the regression and of the variance functions together with empirical distribution and quantile function estimators based on the residuals are plugged into the general expression of the conditional ROC curve. Pepe (1997, 1998, 2003), Faraggi (2003), González-Manteiga *et al.* (2011) propose estimators that implement these ideas. Since most of these estimators are based on classical least squares procedures or local averages, they may be influenced by anomalous data or small deviations from the model assumptions. The bi–normal model, in which both populations are assumed to be normal, is a very popular choice to fit a ROC curve and one justification for its broad use is its robustness, meaning that the estimator based on a bi–normal model produces valid inferences in circumstances where the observations are not normal. In fact, Gonçalves *et al.* (2014) discuss the scope of the so–called robustness in the ROC curve scenario. Walsh (1997) performs a simulation study that shows that the bi–normal estimator is affected by model misspecifications and by the location of the decision thresholds.

In this paper, we focus on robustness, in the sense, that we seek for procedures resistant to deviations from the underlying model and highly efficient when this central model holds. During the last decades, robust statistics has pursued the aim of developing procedures that enable reliable inference results, even if small deviations from the model assumptions occur or in the presence of a moderate percentage of outliers. Even when these efforts have been sustained over time across different statistical areas, up to our knowledge, ROC

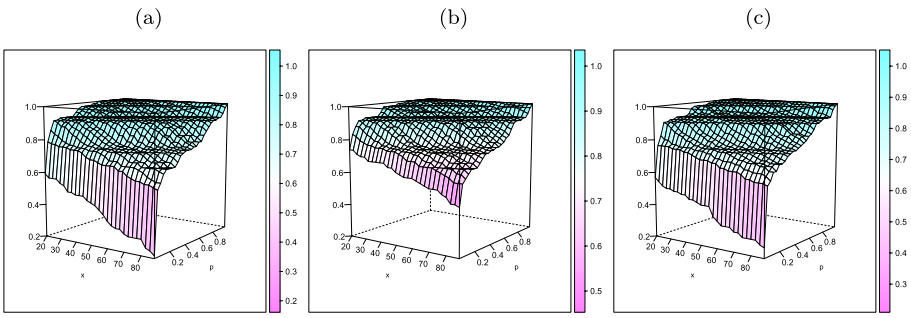

FIG 1. *Estimated ROC surfaces for the Diabetes Data using Age (X) as covariate: (a) Classical estimator, (b) Classical estimator without the detected outliers and (c) Naive estimator.*

curves have received little attention from this robustness point of view. When no covariates are available, robust estimators of the area under the ROC curve were given in Greco and Ventura (2011) assuming that the distribution functions are known up to a finite–dimensional parameter (see also Farcomeni and Ventura, 2012). In this sense, when covariates are recorded to improve the discrimination power of the biomarker, the main contribution of our paper is to bridge the gap between ROC curves and robustness. We achieve this goal by fitting a location-scale regression model to the diagnostic variable and considering weighted empirical estimators of the regression residuals distributions. In this respect, our proposal is semiparametric since the errors distribution is not assumed to be known, for example, as in the bi–normal model and may be easily implemented in `R` through the functions available in the library `robustbase`.

Our motivating example consists of the real dataset of a marker for diabetes previously analysed in Faraggi (2003) and Pardo-Fernández *et al.* (2014), in which we add to their analysis a robust perspective focussing on the potential effect of influential data. The observations, that come from a population-based pilot survey of diabetes mellitus in Cairo, Egypt, consist of postprandial blood glucose measurements (Y) from a fingerstick in 286 subjects who were divided into healthy (198) and diseased (88) groups according to gold standard criteria of the World Health Organization (1985). It is believed that the aging process may be associated with resistance or relative insulin deficiency among healthy people, therefore postprandial fingerstick glucose levels would be expected to be higher for older persons who do not have diabetes. According to this belief, Smith and Thompson (1996) adjust the ROC curve analysis for covariate information using age (X). The obtained ROC curve of the transformed biomarker is given in Figure 1 together with the ROC curve obtained after removing the 6 outliers detected in the healthy sample through a robust regression fit. Figure 1 also displays the ROC curve built using the naive approach of using robust regression estimators combined with the usual empirical distribution and quantile function estimators based on the residuals. These plots illustrate that the use of robust regression and variance estimators are not enough to protect the es-

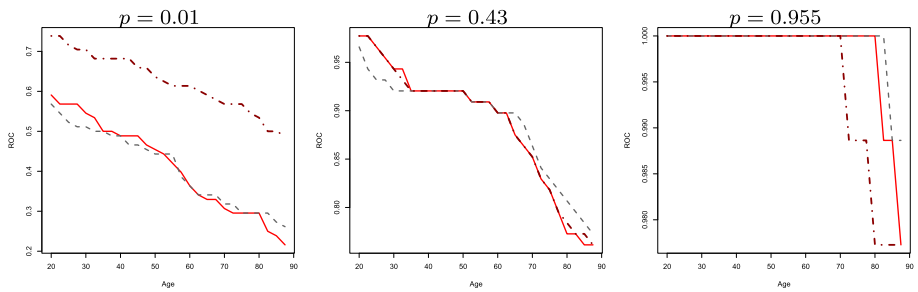

FIG 2. *Estimated ROC for the Diabetes Data for three values of p using Age (X) as covariate. The solid red line corresponds to the classical estimator, the dark red dotted-dashed one to the classical estimator without the detected outliers and the dashed grey to the naive one.*

timation of the ROC curve from the influence of atypical data. This effect may be explained by the fact that large residuals are still present when empirical distribution estimators are computed affecting, in particular, extreme quantiles and producing a distorted ROC surface for values of p close to 0 and 1, which is better appreciated in Figure 2. This motivates the need of defining appropriate robust estimators of the ROC curve.

In the rest of the paper we will introduce a robust proposal and we will study some of its properties. The paper is organized as follows. Section 2 reviews some general concepts regarding the conditional ROC curve and introduces the robust proposal to estimate the ROC curve. Section 3 presents some consistency results. In Section 4, a numerical study is conducted to examine the small sample properties of the proposed procedures under a non–linear regression model, while the results under a linear model are given in the supplementary file available online. The advantages of the proposed methodology are illustrated in Section 5 on a real data set, while some final comments are presented in Section 6. Assumptions and proofs of the main results are relegated to the Appendix, while the convergence of a proposal to choose the tuning constants involved is studied in the supplementary file (Bianco *et al.*, 2022).

2. Proposal

2.1. Preliminaries

In this section, we recall the approach considered to model the induced ROC curve when covariates are measured. For that purpose, denote as Y_D and $\mathbf{X}_D$ the biomarker and the covariates measured in the diseased population and as Y_H and $\mathbf{X}_H$ the corresponding ones in the healthy individuals. Even though in many situations the same covariates are measured in both populations, in this paper, we will only assume that $\mathbf{X}_D$ and $\mathbf{X}_H$ have the same dimension and we will call $\mathcal{S}$ their common support.

A general way to include covariates is through a general location–scale regression model which, for simplicity of presentation, we assume homoscedastic in both populations, that is,

$$Y_D = \mu_{0,D}(\mathbf{X}_D) + \sigma_{0,D}\,\epsilon_D\,, \tag{2}$$
$$Y_H = \mu_{0,H}(\mathbf{X}_H) + \sigma_{0,H}\,\epsilon_H\,, \tag{3}$$

where, for $j = D, H$, $\mu_{0,j}$ is the true regression function and $\sigma_{0,j}$ corresponds to the model dispersion, respectively. It is also assumed that the errors $\epsilon_j \sim G_j$ are independent of $\mathbf{X}_j$, for $j = D, H$ and have scale 1 to properly identify $\sigma_{0,j}$. Furthermore, to identify the regression function in the classical framework it is assumed that $\mathbb{E}\epsilon_j = 0$, for $j = D, H$. Instead, in the robust setting it is usual to avoid the existence of moments. For that reason, to ensure consistency of the robust estimators to the target regression function $\mu_{0,j}$, it is standard to assume that G_j has a symmetric distribution. Otherwise, Fisher–consistency of the related regression functionals should be required. It is worth noticing that since the errors and the covariates are independent, for a given $\mathbf{x} \in \mathcal{S}$, we have that

$$\begin{aligned} F_D(y|\mathbf{x}) &= F_{Y_D|X_D}(y|\mathbf{x}) = \mathbb{P}(Y_D \le y|\mathbf{X}_D = \mathbf{x}) \\ &= \mathbb{P}(\mu_{0,D}(\mathbf{X}_D) + \sigma_{0,D}\,\epsilon_D \le y|\mathbf{X}_D = \mathbf{x}) = G_D\left(\frac{y - \mu_{0,D}(\mathbf{x})}{\sigma_{0,D}}\right)\,. \end{aligned}$$

Analogously, we get that the conditional distribution in the healthy distribution satisfies

$$F_H(y|\mathbf{x}) = F_{Y_H|X_H}(y|\mathbf{x}) = G_H\left(\frac{y - \mu_{0,H}(\mathbf{x})}{\sigma_{0,H}}\right)\,.$$

As a consequence, the quantiles of the conditional distributions are related to those of the errors through $F_j^{-1}(p|\mathbf{x}) = \sigma_{0,j}\,G_j^{-1}(p) + \mu_{0,j}(\mathbf{x})$, for $j = D, H$, where $G_j^{-1}(\cdot)$ denotes the quantile function of the errors ϵ_j. Thus, the conditional ROC curve given $\mathbf{x} \in \mathcal{S}$ defined in (1) can be computed as

$$\mathrm{ROC}_{\mathbf{x}}(p) = 1 - G_D\left(\frac{\mu_{0,H}(\mathbf{x}) - \mu_{0,D}(\mathbf{x})}{\sigma_{0,D}} + \frac{\sigma_{0,H}}{\sigma_{0,D}}\,G_H^{-1}(1-p)\right)\,. \tag{4}$$

One advantage of this approach is that it enables a very general modelling of the regression functions $\mu_{0,j}$, for $j = D, H$, since this task can be accomplished from different perspectives. This means that according to the information about the relationship between the biomarker and the covariates and the user's preferences, the regression functions may be modelled parametrically, nonparametrically or partly parametrically.

Different summary measures of the ROC curve are useful to sum up particular features of the curve. One of the most popular indices is the conditional *area under the curve* ($\mathrm{AUC}_{\mathbf{x}}$), which is computed as $\mathrm{AUC}_{\mathbf{x}} = \int_0^1 \mathrm{ROC}_{\mathbf{x}}(p)dp$.

Suppose that we have a sample from the diseased population, $(y_{D,i}, \mathbf{x}_{D,i})$, $1 \le i \le n_D$, that verifies model (2) and one from the healthy population, $(y_{H,i}, \mathbf{x}_{H,i})$,

$1 \leq i \leq n_H$, verifying model (3). Furthermore, assume that the samples are independent from each other. As mentioned in the Introduction, expression (4) of the conditional ROC curve suggests a natural estimation procedure which may be described through the following steps

Step 1. For $j = D, H$ and from the sample $(y_{j,1}, \mathbf{x}_{j,1}), \ldots, (y_{j,n_j}, \mathbf{x}_{j,n_j})$ estimate $\mu_{0,j}(\mathbf{x})$ and $\sigma_{0,j}$. Denote the resulting estimators by $\widetilde{\mu}_j(\mathbf{x})$ and $\widetilde{\sigma}_j$, for $j = D, H$.
Step 2. Compute for each sample the standardized regression residuals

$$r_{H,i} = \frac{y_{H,i} - \widetilde{\mu}_H(\mathbf{x}_{H,i})}{\widetilde{\sigma}_H} \quad \text{and} \quad r_{D,i} = \frac{y_{D,i} - \widetilde{\mu}_D(\mathbf{x}_{D,i})}{\widetilde{\sigma}_D}\,.$$

On the basis of these residuals, the empirical distribution function estimator can be computed as $\widehat{G}_{j,\mathrm{EMP}}(r) = (1/n_j)\sum_{i=1}^{n_j} \mathbb{I}_{r_{j,i} \leq r}$, for $j = D, H$, leading in the healthy population, to the quantile function estimator $\widehat{G}^{-1}_{H,\mathrm{EMP}}$.
Step 3. Plug–in the estimators computed in the first two steps into equation (4) to obtain the estimated conditional ROC function

$$\widetilde{\mathrm{ROC}}_{\mathbf{x}}(p) = 1 - \widehat{G}_{D,\mathrm{EMP}}\left(\frac{\widetilde{\mu}_H(\mathbf{x}) - \widetilde{\mu}_D(\mathbf{x})}{\widetilde{\sigma}_D} + \frac{\widetilde{\sigma}_H}{\widetilde{\sigma}_D}\,\widehat{G}^{-1}_{H,\mathrm{EMP}}(1-p)\right).$$

2.2. The general procedure

From now on, we consider the regression models given in (2) and (3), however the arguments given below may be easily extended to the heteroscedastic case.

Our goal is to introduce a procedure to get reliable and stable $\mathrm{ROC}_{\mathbf{x}}$ estimators, even when a moderate percentage of outliers arise in one sample or in both of them. As it is well known, the regression and scale estimators computed in **Step 1** may be affected by outliers if classical methods, such as least squares estimators, are considered. A naive approach could be to compute robust estimators of $\mu_{0,j}(\mathbf{x})$ and $\sigma_{0,j}$, for $j = D, H$, instead of classical ones. However, as shown with the diabetes data set considered in the Introduction, this is not enough to provide final robust estimators of the ROC curve, since the presence of large residuals may still affect **Steps 2** and **3**. In fact, even when robust estimators are considered for the estimation of the regression and variance functions, large residuals would influence the classical empirical distribution and quantile function estimators wasting the efforts made in the first step to get robustness.

Note that large values of $|r_{j,i}|$ suggest that the corresponding pairs $(y_{j,i}, \mathbf{x}_{j,i})$ may be outliers. In that case, under a normal error model, it seems wise to consider as atypical those points whose residuals are larger than a certain cut–off value c, that is, such that $|r_{j,i}| > c$. Typically, c is chosen as 2.5 by taking the standard normal distribution as a benchmark. For that reason, for $j = D, H$, once the residuals $r_{j,i}$ from a robust fit have been computed, we have to downweight large residual values, as it is done when computing robust regression

estimators. Unlike the empirical distribution estimators, where all the observations have the same weight, downweighting in the second step atypical points, i.e., those values that lie far away from the bulk of the data, may result in a more resistant procedure. More precisely, in **Step 2**, for $j = D, H$, we need to consider a weighted empirical distribution defined as:

$$\widehat{G}_j(r) = \frac{1}{\sum_{\ell=1}^{n_j} w_{j,\ell}} \sum_{i=1}^{n_j} w_{j,i} \mathbb{I}_{r_{j,i} \leq r} \quad \text{where} \quad w_{j,i} = w\left(\frac{r_{j,i}}{c}\right), \tag{5}$$

where $w : \mathbb{R} \to [0, 1]$ is an even weight function, non-increasing in $[0, +\infty)$, such that $w(0) = 1$, $w(u) > 0$ for $0 < u < 1$ and $w(u) = 0$ for $u \geq 1$. Further assumptions on the weight function will be required in Propositions 1 and 2 to ensure consistency of the resulting weighted empirical distribution. The fact that $w(u) = 0$ for $u \geq 1$ ensures that $w_{j,i} = 0$ when $|r_{j,i}|$ is larger than the selected cut–off value c, so, observations with large residuals will be completely eliminated in the weighted estimators ensuring robustness of the final ROC curve. It is worth noting that beneath this criterion to downweight large residuals lays the idea that the errors distribution G_j is symmetric. If the practitioner suspects that a skewed distribution underlies, another kind of weights, such as asymmetric ones, may be preferable.

At this stage, the choice of the constant c does matter and is a keypoint. Recall that when considering MM-estimators in linear or non–linear regression models, the tuning constant is selected to attain a given efficiency ensuring at the same time robustness. However, in our setting, it is quite intuitive that for a fixed cut–off value $\widehat{G}_j(r)$ will not converge to G_j but to

$$\frac{\mathbb{E}\left\{w\left(\frac{\epsilon_j}{c}\right)\mathbb{I}_{\epsilon_j \leq r}\right\}}{\mathbb{E}w\left(\frac{\epsilon_j}{c}\right)},$$

leading to unconsistent estimators of the distribution functions G_j, $j = D, H$, and of the ROC function (see Propositions 1 and 2 below). Therefore, cut–off values c_{n_j}, increasing to infinite but at the same time, ensuring good robustness properties for small samples are needed. We refer to Section S.1 in the supplementary file for a description of a possible procedure to select the cut–off values.

Taking these ideas into account, we propose the following stepwise procedure:

Step 1R. Estimate $\mu_{0,H}(\mathbf{x})$, $\sigma_{0,H}$, $\mu_{0,D}(\mathbf{x})$, $\sigma_{0,D}$ in a robust fashion from the samples $(y_{H,1}, \mathbf{x}_{H,1}), \ldots, (y_{H,n_H}, \mathbf{x}_{H,n_H})$ and $(y_{D,1}, \mathbf{x}_{D,1}), \ldots, (y_{D,n_D}, \mathbf{x}_{D,n_D})$, respectively. Denote the resulting estimators by $\widehat{\mu}_H(\mathbf{x})$, $\widehat{\sigma}_H$, $\widehat{\mu}_D(\mathbf{x})$ and $\widehat{\sigma}_D$.

Step 2R. Compute for each sample the standardized regression residuals

$$r_{H,i} = \frac{y_{H,i} - \widehat{\mu}_H(\mathbf{x}_{H,i})}{\widehat{\sigma}_H} \quad \text{and} \quad r_{D,i} = \frac{y_{D,i} - \widehat{\mu}_D(\mathbf{x}_{D,i})}{\widehat{\sigma}_D}.$$

From these residuals, evaluate robust adaptive estimators of the distribution and quantile functions, denoted, $\widehat{G}_D$ and $\widehat{G}_H^{-1}$, respectively as in (5), using an unbounded increasing sequence of cut–off values c_{n_j}, $j = D, H$.
Step 3R. Plug–in the robust estimators computed in the first two steps into equation (4) to obtain

$$\widehat{\mathrm{ROC}}_{\mathbf{x}}(p) = 1 - \widehat{G}_D\left(\frac{\widehat{\mu}_H(\mathbf{x}) - \widehat{\mu}_D(\mathbf{x})}{\widehat{\sigma}_D} + \frac{\widehat{\sigma}_H}{\widehat{\sigma}_D}\,\widehat{G}_H^{-1}(1-p)\right).$$

Regarding **Step 1R**, the considered regression models (2) and (3) may be either parametric, nonparametric or semiparametric. In each case, suitable robust estimators must be used. In particular, in the parametric case, linear or non–linear models may be adequate. To illustrate, assume that, for $j = D, H$, the observations satisfy the non–linear regression model

$$y_{j,i} = f_j(\mathbf{x}_{j,i}, \beta_{0,j}) + \sigma_{0,j}\epsilon_{j,i}\,, \qquad i = 1 \ldots n_j\,, \tag{6}$$

with $\beta_{0,j} \in \mathbb{R}^q$, $\sigma_{0,j}$ the scale parameter and f_j a known function. Note that the dimension of the regression parameter $\beta_{0,j}$ may be equal or not to that of the covariates. The errors $\epsilon_{j,i}$ are independent and identically distributed (i.i.d.) with unknown distribution G_j and independent of the covariates $\mathbf{x}_{j,i}$. In such setting, the MM-estimators introduced in Yohai (1987) are a recommended option for the linear model, while, under a general non–linear regression model, the weighted MM-estimators presented in Bianco and Spano (2019) may be used.

The challenge in **Step 2R** is to implement a version of the empirical distribution with adaptive weights ensuring at the same time consistency under the central model and robustness.

As mentioned above, our proposal may be easily implemented when one or both regression models are heteroscedastic. In this case, if the scale function is modelled in terms of the covariates, estimators of the scale function need to be computed in **Step 1R**. Then, in **Step 2R** the residuals $r_{j,i}$, for $j = D$ and/or H, are obtained using $\widehat{\sigma}_j(\mathbf{x}_{j,i})$ instead of $\widehat{\sigma}_j$. At this point different approaches may be followed. In the nonparametric setting, the local median of the absolute deviations from the median (local MAD) considered in Boente and Fraiman (1989) provides a robust consistent scale estimator. Instead, when the scale function is parametrically modelled, the robust estimators that have been studied in Carroll and Ruppert (1982) and Bianco *et al.* (2000), under an heteroscedatic linear regression model, may be implemented. The results in Theorem 1 can be extended to the heteroscedastic setting by requiring to the true scale function $\sigma_j(\mathbf{x})$ to be bounded and bounded from below on any compact set $\mathcal{K} \subset \mathcal{S}$ and to the estimators $\widehat{\sigma}_j(\mathbf{x})$ consistency assumptions analogous to those stated for $\widehat{\mu}_j(\mathbf{x})$ in **A4** and **A5**.

3. Consistency results

In this section, we derive uniform consistency results for the ROC curve estimators. Theorem 1 and Corollary 1 recall on the uniform consistency of the

distribution function estimators. For the weighted distribution function estimators defined in (5) uniform consistency results are derived in Propositions 1 and 2 under a linear and non–linear regression model, respectively. Extensions to other settings, such as nonparametric regression models, may be also obtained but are beyond the scope of this manuscript.

Henceforth, $(y_{j,i}, \mathbf{x}_{j,i})$, $1 \leq i \leq n_j$, for $j = D, H$, stand for independent random samples from the diseased and healthy populations with the same distribution as $(Y_D, \mathbf{X}_D) \in \mathbb{R}^{p+1}$ and $(Y_H, \mathbf{X}_H) \in \mathbb{R}^{p+1}$, respectively, where $(Y_D, \mathbf{X}_D)$ satisfy (2) and $(Y_H, \mathbf{X}_H)$ fulfils (3). The errors $\epsilon_j \sim G_j$ are independent of $\mathbf{X}_j$, for $j = D, H$. In this situation, we have that (4) holds. To provide a general framework, for $j = D, H$, we will denote as $\widehat{G}_j$ an estimator of the errors distribution function G_j obtained from the sample $(y_{j,i}, \mathbf{x}_{j,i})$, $1 \leq i \leq n_j$ using robust consistent estimators $\widehat{\mu}_j$ and $\widehat{\sigma}_j$ of $\mu_{0,j}$ and $\sigma^2_{0,j}$, respectively. Then, the estimator of the ROC curve whose uniform consistency we will study is given by

$$\widehat{\mathrm{ROC}}_{\mathbf{x}}(p) = 1 - \widehat{G}_D \left(\frac{\widehat{\mu}_H(\mathbf{x}) - \widehat{\mu}_D(\mathbf{x})}{\widehat{\sigma}_D} + \frac{\widehat{\sigma}_H}{\widehat{\sigma}_D} \widehat{G}_H^{-1}(1-p) \right) .$$

Theorem 1. *Let $(y_{j,i}, \mathbf{x}_{j,i})$, $1 \leq i \leq n_j$, $j = D, H$, be independent observations satisfying (2) and (3), respectively and let $\widehat{\mu}_j$ and $\widehat{\sigma}_j$ be estimators of $\mu_{0,j}$ and $\sigma_{0,j}$, respectively. Assume that $\widehat{\sigma}_j \xrightarrow{a.s.} \sigma_{0,j}$, for $j = D, H$, and that **A1** to **A4** in the Appendix hold. Then,*

(i) $\sup_{0<p<1} |\widehat{ROC}_{\mathbf{x}}(p) - ROC_{\mathbf{x}}(p)| \xrightarrow{a.s.} 0$, *for each fixed* $\mathbf{x} \in \mathcal{S}$.

*(ii) If, in addition, assumptions **A5** and **A6** in the Appendix hold and G_D has a bounded density g_D, we have that, for any $\delta > 0$*

$$\sup_{\delta<p<1-\delta} \sup_{\mathbf{x}\in\mathcal{K}} |\widehat{ROC}_{\mathbf{x}}(p) - ROC_{\mathbf{x}}(p)| \xrightarrow{a.s.} 0 .$$

*(iii) Furthermore, when assumptions **A5** and **A6** in the Appendix hold, G_D has a bounded density g_D and the conditional ROC function is such that, for any $\epsilon > 0$, there exists $0 < \eta < 1$ such that, for any $\mathbf{x} \in \mathcal{K}$, $ROC_{\mathbf{x}}(\eta) < \epsilon$ and $1 - ROC_{\mathbf{x}}(1-\eta) < \epsilon$, then* $\sup_{0<p<1} \sup_{\mathbf{x}\in\mathcal{K}} |\widehat{ROC}_{\mathbf{x}}(p) - ROC_{\mathbf{x}}(p)| \xrightarrow{a.s.} 0$.

As a consequence of Theorem 1, we immediately get the following result for parametric regression.

Corollary 1. *For $j = D, H$, let $(y_{j,i}, \mathbf{x}_{j,i})$, $1 \leq i \leq n_j$, be independent observations satisfying (6), where the errors $\epsilon_{j,i} \sim G_j$ are independent of the covariates $\mathbf{x}_{j,i}$ and assume that $\widehat{\beta}_j$ and $\widehat{\sigma}_j$ are strongly consistent estimators of $\beta_{0,j}$ and $\sigma_{0,j}$, respectively. Then, under assumptions **A1** to **A3** and **A7** in the Appendix,*

(i) $\sup_{0<p<1} |\widehat{ROC}_{\mathbf{x}}(p) - ROC_{\mathbf{x}}(p)| \xrightarrow{a.s.} 0$, *for each fixed* $\mathbf{x} \in \mathcal{S}$.

*(ii) If, in addition, G_D has a bounded density g_D and the regression functions f_j satisfy, for $j = D, H$, assumption **A8** in the Appendix, then*

$\sup_{\mathbf{x}\in\mathcal{K}} |\widehat{ROC}_{\mathbf{x}}(p) - ROC_{\mathbf{x}}(p)| \xrightarrow{a.s.} 0$. *Moreover, for any* $\delta > 0$

$$\sup_{\delta<p<1-\delta} \sup_{\mathbf{x}\in\mathcal{K}} |\widehat{ROC}_{\mathbf{x}}(p) - ROC_{\mathbf{x}}(p)| \xrightarrow{a.s.} 0\,.$$

(iii) Furthermore, assume that G_D *has a bounded density* g_D*, the regression functions* f_j *satisfy assumption* ***A8*** *in the Appendix, for* $j = D, H$*, and the conditional ROC function is such that, for any* $\epsilon > 0$*, there exists* $0 < \eta < 1$ *such that, for any* $\mathbf{x} \in \mathcal{K}$*,* $ROC_{\mathbf{x}}(\eta) < \epsilon$ *and* $1-ROC_{\mathbf{x}}(1-\eta) < \epsilon$*, then* $\sup_{0<p<1} \sup_{\mathbf{x}\in\mathcal{K}} |\widehat{ROC}_{\mathbf{x}}(p) - ROC_{\mathbf{x}}(p)| \xrightarrow{a.s.} 0$.

It is worth noticing that the requirement $\sup_{\mathbf{x}\in\mathcal{K}} \mathrm{ROC}_{\mathbf{x}}(\eta) < \epsilon$ and $\sup_{\mathbf{x}\in\mathcal{K}} \{1 - \mathrm{ROC}_{\mathbf{x}}(1-\eta)\} < \epsilon$ in (iii) is satisfied when **A8** holds and G_H has support on the whole line as stated in **A1**.

Propositions 1 and 2 below show that, under mild assumptions, when $c_{n_j} \xrightarrow{a.s.} \infty$ assumption **A3** is fulfilled for the weighted empirical distribution function estimators defined in (5), that is, we have that $\|\widehat{G}_j - G_j\|_\infty \xrightarrow{a.s.} 0$, $j = D, H$. Note that these results allow for random cut–off values to include the adaptive procedure described in Section S.1 and whose convergence is analysed in the supplementary file. In particular, from Propositions 1 and 2, we also obtain the uniform consistency of the empirical residuals distribution $\widehat{G}_{j,\mathrm{EMP}}$ which correspond to the choice $w(t) \equiv 1$.

We will consider linear and non–linear models separately, since weaker assumptions are required for the former one. In particular, for the linear regression model we also obtain uniform consistency of the weighted empirical distribution function estimators for the hard rejection weight function $w(t) = \mathbb{I}_{[-1,1]}(t)$. The approach allows to fit different models for the healthy and diseased populations.

From now on, we fix $j = D$ or H and we assume that the observations $(y_{j,i}, \mathbf{x}_{j,i})$, $1 \le i \le n_j$, satisfy the regression model (6) and that $\widehat{\beta}_j$ and $\widehat{\sigma}_j$ are estimators of $\beta_{0,j}$ and $\sigma_{0,j}$, respectively.

Consider the functions

$$h_{j,\infty}(c) = \mathbb{E}_{G_j} w\left(\frac{\epsilon_j}{c}\right) \tag{7}$$

$$h_{j,0}(c,s) = \mathbb{E}_{G_j} w\left(\frac{\epsilon_j}{c}\right) \mathbb{I}_{\epsilon_j \le s}\,. \tag{8}$$

Proposition 1. *Let* $j = D$ *or* H *and assume that the observations* $(y_{j,i}, \mathbf{x}_{j,i})$*,* $1 \le i \le n_j$*, satisfy a linear regression model, that is, (6) holds with* $f_j(\mathbf{x}, \beta) = \mathbf{x}^{\mathrm{T}}\beta$*, and that* $w(t) = \mathbb{I}_{[-1,1]}(t)$ *or* w *satisfies assumption* ***C1*** *in the Appendix. Let*

$$\widehat{G}_j(r) = \frac{1}{\sum_{\ell=1}^n w_{j,\ell}} \sum_{i=1}^n w_{j,i} \mathbb{I}_{r_{j,i} \le r} \quad \text{with} \quad w_{j,i} = w\left(\frac{r_{j,i}}{c_{n_j}}\right), \tag{9}$$

where $c_{n_j} \xrightarrow{a.s.} c_{j,0}$*. If assumptions* ***C2*** *to* ***C3*** *in the Appendix hold, we have that*

a) if $c_{j,0} < \infty$*,*

$$\sup_{s\in\mathbb{R}} \left| \widehat{G}_j(s) - \frac{h_{j,0}(c_{j,0}, s)}{h_{j,\infty}(c_{j,0})} \right| \xrightarrow{a.s.} 0\,,$$

with $h_{j,\infty}(c_{j,0})$ *and* $h_{j,0}(c_{j,0}, s)$ *defined in* (7) *and* (8)*, respectively.*

b) if $c_{j,0} = \infty$, $\|\widehat{G}_j - G_j\|_\infty \xrightarrow{a.s.} 0$.

Proposition 2. *Let* $j = D$ *or* H *and assume that the observations* $(y_{j,i}, \mathbf{x}_{j,i})$, $1 \leq i \leq n_j$, *satisfy the non–linear regression model* (6) *and let* $\widehat{G}_j$ *be as in* (9) *with* $c_{n_j} \xrightarrow{a.s.} c_{j,0}$. *If assumptions* ***C1*** *and* ***C3*** *to* ***C5*** *in the Appendix hold, we have that*

a) if $c_{j,0} < \infty$,

$$\sup_{s \in \mathbb{R}} \left| \widehat{G}_j(s) - \frac{h_{j,0}(c_{j,0}, s)}{h_{j,\infty}(c_{j,0})} \right| \xrightarrow{a.s.} 0\,,$$

with $h_{j,\infty}(c_{j,0})$ *and* $h_{j,0}(c_{j,0}, s)$ *defined in* (7) *and* (8)*, respectively.*

b) if $c_{j,0} = \infty$, $\|\widehat{G}_j - G_j\|_\infty \xrightarrow{a.s.} 0$.

Note that the results in Propositions 1 and 2 entail that, unlike the tuning constant in robust regression estimators, fixed cut–off values for the empirical weighted distribution functions given in (5) are not a sensible choice, since they do not lead to consistent estimators of the residuals distribution function. Section S.1 in the supplementary file discusses a possible adaptive procedure for the selection of the cut–off values that will ensure consistency under the assumed errors model $G_{j,0}$.

4. Monte Carlo study

In this section, we summarize the results of a simulation study conducted to have an insight on the finite sample performance of the proposal given in Section 2. The goal of this numerical experiment is two–fold. On the one hand, we want to illustrate the lack of stability of the classical methods when deviations from the central model arise. On the other hand, we want to evaluate the performance of our robust proposal under different contamination schemes and to compare it with the classical one. For that purpose, we considered different scenarios and contaminations schemes. In all cases, we generate $Nrep = 1000$ datasets of size $n_D = n_H = n = 100$ and $n_D = n_H = n = 200$. We considered two regression models, a linear and a non–linear one. The results for the linear model, relegated to the supplementary file available online, also include the performance of the naive approach which uses robust regression estimators combined with the usual empirical distribution and quantile function estimators based on the residuals. Through this numerical experiment, we illustrate the necessity of using robust estimators in all the stages of the stepwise procedure. In both cases, different contaminating schemes are analysed either contaminating just one population or both of them.

To summarize the discrepancy between the estimator and the true ROC surface, we consider two grids of points: $\mathcal{G}_p = \{p_j\}_{j=1}^{N_p}$ corresponding to equidistant values between 0.01 and 0.99 with step 0.01 and $\mathcal{G}_x = \{x_i\}_{i=1}^{N_x}$, where the net has step 0.05 within the interval $[a, b]$ with $a = -0.5$ and $b = 0.5$ for the non–linear

model considered. The estimators performance is then evaluated using the mean over replications of

- the Mean Squared Error (MSE) given by

$$MSE = \frac{1}{N_x N_p} \sum_{i=1}^{N_x} \sum_{j=1}^{N_p} \left(\widehat{\mathrm{ROC}}_{x_i}(p_j) - \mathrm{ROC}_{x_i}(p_j) \right)^2 ,$$

- a measure inspired on the Kolmogorov–Smirnov distance (KS) calculated as

$$KS = \sup_{1 \le i \le N_x} \sup_{1 \le j \le N_p} \left| \widehat{\mathrm{ROC}}_{x_i}(p_j) - \mathrm{ROC}_{x_i}(p_j) \right| ,$$

that give global summaries of the mismatch between the estimated ROC curve and the true one.

We consider an exponential model, that is, we assume that the observations follow the non–linear regression models

$$y_{D,i} = \beta_{D,1} \exp(\beta_{D,2}\, x_{D,i}) + \epsilon_{D,i} , \tag{10}$$

$$y_{H,i} = \beta_{H,1} \exp(\beta_{H,2}\, x_{H,i}) + \epsilon_{H,i} , \tag{11}$$

with $(\beta_{D,1}, \beta_{D,2})^{\mathrm{T}} = (5, 2)$, $(\beta_{H,1}, \beta_{H,2})^{\mathrm{T}} = (3, 1)$ for all $i = 1, \ldots, n$ $\epsilon_{j,i} \sim N(0, 1)$ are independent and independent from $x_{j,i} \sim U(-0.5, 0.5)$, for $j = D, H$. Besides, the sample from one population was generated independently from that of the other one.

We evaluate the performance of the classical procedure which uses in **Step 1** the usual least squares estimators for non–linear regression models and the empirical distribution function with the robust method proposed in Section 2.2. More precisely, in **Step 1R** the robust regression estimators correspond to the weighted MM-estimators defined in Bianco and Spano (2019), while in **Step 2R** we used the weighted empirical distribution defined in (5). The weight function was taken as $w(t) = \mathbb{I}_{[-1,1]}(t)$ and the cut–off values as described in Section S.1 of the supplementary file with $\eta = 2.5$ and $G_{j,0} = \Phi$ the normal distribution.

To assess the impact of anomalous data on the estimation of the conditional ROC curve, we introduce *shift outliers* in both populations and to explore the influence of the shift size, we vary its magnitude on a grid of fixed values. To this end, the first m observations of each sample were replaced by observations following the models

$$y_{D,i} = \beta_{D,1} \exp(\beta_{D,2}\, x_{D,i}) + S + 0.01\epsilon_{D,i} , \tag{12}$$

$$y_{H,i} = \beta_{H,1} \exp(\beta_{H,2}\, x_{H,i}) + S + 0.01\epsilon_{H,i} , \tag{13}$$

where $x_{j,i} \sim U(0.49, 0.5)$ and $\epsilon_{j,i}$ are as above, for $j = D, H$. The level shift S is a fixed number taking values equal to $2.5, 5, 7.5, 10, 12.5$ or 15. The aim of this contamination scheme is to introduce bad leverage points, that is, high leverage covariates with vertical outliers. As it is well known, including several repeated points (x, y) may lead to numerical instability of the algorithms used

to compute both estimators. To avoid this problem, we randomly choose high leverage values of the covariates uniformly distributed in $(0.49, 0.5)$, combined with a level shift S. The Gaussian errors with standard deviation 0.01 ensure that the responses vary close enough to the level shift.

We consider a 5% or a 10% proportion of anomalous points, that is, we replace $m = n\delta$ points, $\delta = 0.05$ and 0.10 by observations generated as in (12) and (13). We denote this contamination $C_{\delta,S}$, while C_0 stands for the situation of clean samples.

For the sake of brevity, Table 1 summarizes the discrepancy between the true and estimated ROC curves in terms of the mean over replications of the MSE. In the supplementary file, when considering a linear model, the results regarding the Kolmogorov–Smirnov distance are also given. The damage of shift outliers on the conditional ROC curve is striking, since the MSE increases more than 10 times for $n = 100$ and more than 20 times if $n = 200$ when S takes the largest values.

TABLE 1

Influence of the shift size S on the MSE, when $n = 100$ and 200, under the non–linear model (10) and (11), when considering the contamination scheme $C_{\delta,S}$.

				S					
δ	n		C_0	2.5	5	7.5	10	12.5	15
0.05	100	Robust	0.0023	0.0028	0.0023	0.0023	0.0024	0.0024	0.0024
		Classical	0.0019	0.0023	0.0066	0.0127	0.0183	0.0231	0.0269
0.10	100	Robust	0.0023	0.0032	0.0031	0.0024	0.0024	0.0024	0.0024
		Classical	0.0019	0.0029	0.0092	0.0180	0.0240	0.0248	0.0268
0.05	200	Robust	0.0011	0.0018	0.0011	0.0011	0.0011	0.0011	0.0011
		Classical	0.0010	0.0015	0.0062	0.0124	0.0181	0.0229	0.0267
0.10	200	Robust	0.0011	0.0022	0.0014	0.0011	0.0011	0.0011	0.0011
		Classical	0.0010	0.0021	0.0088	0.0178	0.0239	0.0242	0.0249

TABLE 2

Mean of MSE and KS over replications for clean and contaminated samples, under the non–linear model (10) and (11).

n		C_0		$C_{0.05,10}$		$C_{0.10,10}$	
		Classical	Robust	Classical	Robust	Classical	Robust
100	MSE	0.0019	0.0023	0.0183	0.0024	0.0240	0.0024
	KS	0.1881	0.1944	0.9334	0.2001	0.6893	0.2048
200	MSE	0.0010	0.0011	0.0181	0.0011	0.0239	0.0011
	KS	0.1352	0.1367	0.9364	0.1395	0.7102	0.1445

Henceforth, we focus on the particular case of outliers with shift value $S = 10$, a mild value among those considered, so as to have a deeper comprehension of the effect of the introduced anomalous points. Table 2 summarizes the results through the mean of the measures MSE and KS. Note that the mean of the summary measures are distorted for the classical procedure. In particular, when considering the measure KS based on the Kolmogorov–Smirnov distance, the

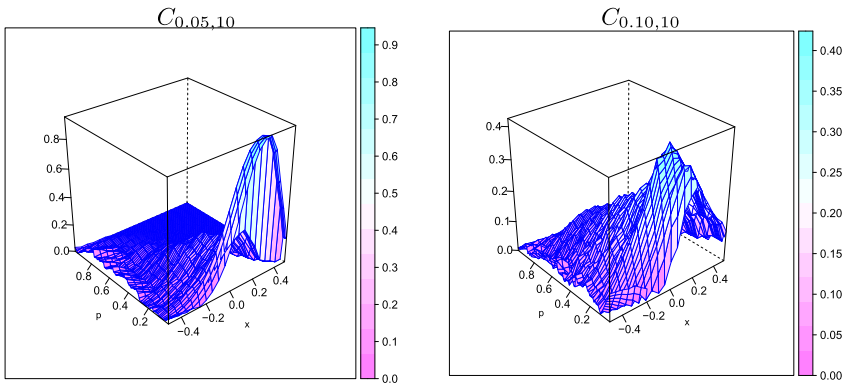

FIG 3. *Surface plot of* $\Delta(\mathbf{x}, p) = \left|\widehat{ROC}_x(p) - ROC_x(p)\right|$ *for the samples generated in one of the considered replications, under two contamination schemes.*

mean is enlarged almost 7 times, under $C_{0.05,S}$ when $n = 200$. It is worth mentioning that, when considering the classical procedure, the Kolmogorov–Smirnov distance decreases when moving from the contaminated model $C_{0.05,10}$ to $C_{0.10,10}$, although the former scenario contains a higher proportion of contaminated data. Figure 3 where we plot the surface $\Delta(\mathbf{x}, p) = \left|\widehat{\mathrm{ROC}}_x(p) - \mathrm{ROC}_x(p)\right|$ for the samples generated in one of the considered replications, explains this unexpected effect. The left panel corresponds to $C_{0.05,10}$ and the right one to $C_{0.10,10}$. The decrease in KS is mainly due to the fact that under $C_{0.10,10}$ the whole ROC surface is distorted leading to a large area under the surface $\Delta(\mathbf{x}, p)$, but with a lower maximum than under $C_{0.05,10}$. This area is represented by MSE which approximates $\int_0^1 \int_{-0.5}^{0.5} \Delta^2(\mathbf{x}, p) d\mathbf{x}\, dp$. In contrast, under $C_{0.05,10}$, there is a high peak for values of p close to 0, but in a small region, so this distortion is not reflected on the MSE but it is well registered by the KS. For that reason, the two measures are important giving complementary information, since they capture different effects.

Figures 4 and 5 show the functionals boxplots of the classical and robust AUC_x obtained for $n = 100$ and $n = 200$, respectively. Functional boxplots, introduced by Sun and Genton (2011), are useful to visualize a collection of curves. The area in purple represents the 50% inner band of curves, the dotted red lines correspond to outlying curves, the black line indicates the central (deepest) function, while the green line in the plot corresponds to the true AUC_x curve. Notice that in these boxplots, the estimators of conditional area under the curve were plotted in the range $(-0.5, 0.2)$, since for this simulation scheme the AUC_x is almost 1 when the covariate takes values from 0.2 to 0.5. Once again, it becomes evident that the classical estimator suffers from the introduced contamination and that the classical estimator of AUC_x is completely deviated from the true conditional area under the curve, which is plotted in green, while the robust AUC_x estimator remains very stable.

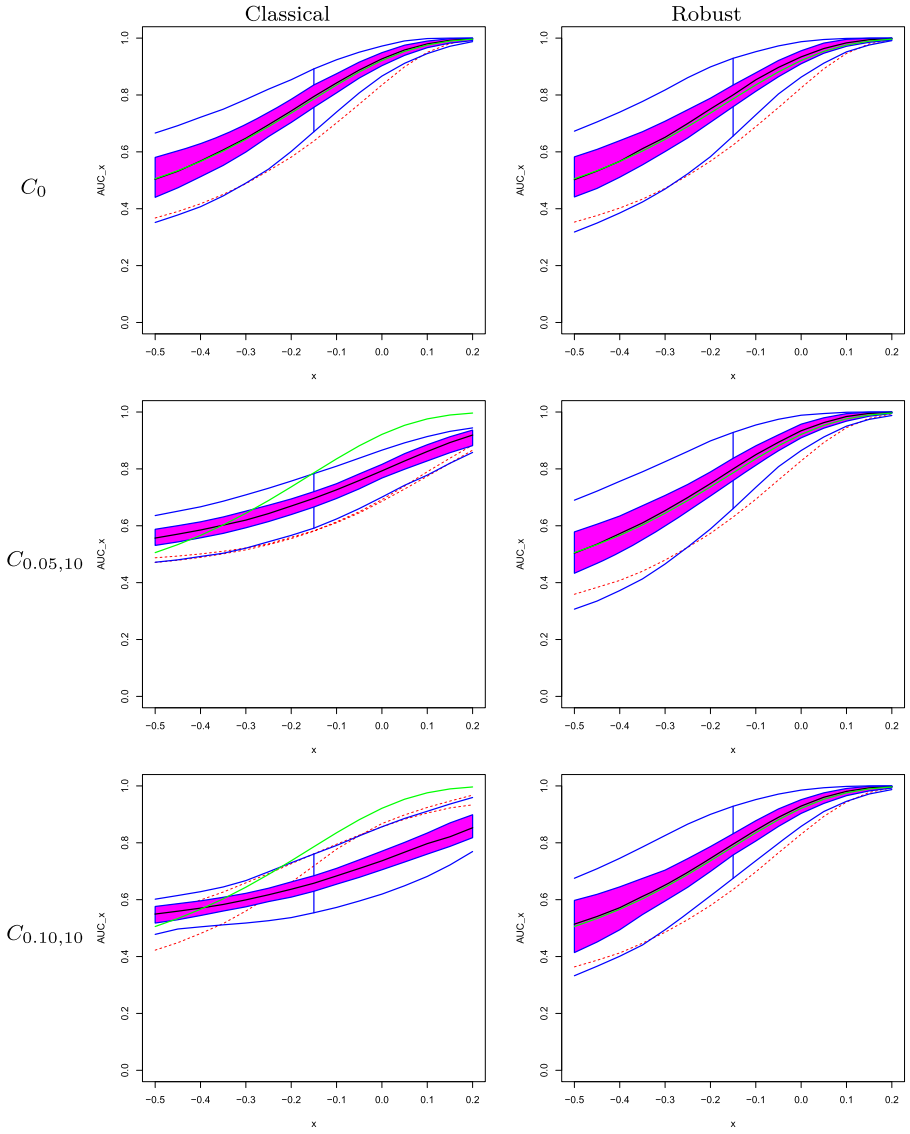

FIG 4. *Functional boxplots of* $\widehat{AUC}_x$ *obtained with the classical and robust estimators for* $n = 100$ *with clean samples and under 5% and 10% of contamination with level shift* $S = 10$, *under the non–linear model* (10) *and* (11). *The green line corresponds to the true* AUC_x *and the dotted red lines to the outlying curves detected by the functional boxplot.*

5. Analysis of real data set

In this section, we illustrate the benefits of the robust proposed methodology by means of the diabetes real dataset described in the Introduction.

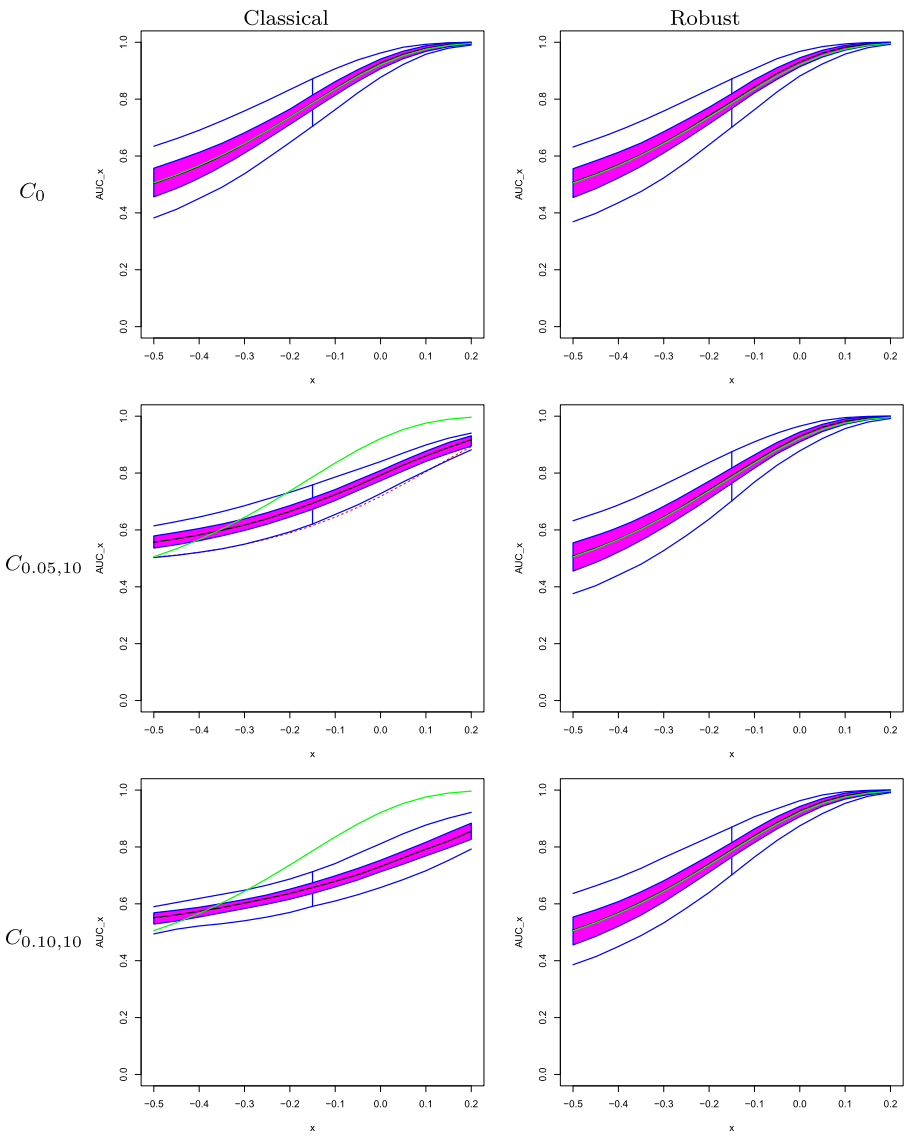

FIG 5. *Functional boxplots of* $\widehat{AUC}_x$ *obtained with the classical and robust estimators for* $n = 200$ *with clean samples and under 5% and 10% of contamination with level shift* $S = 10$*, under the non–linear model* (10) *and* (11)*. The green line corresponds to the true* AUC_x *and the dotted red lines to the outlying curves detected by the functional boxplot.*

Following the analysis given in Faraggi (2003), we transform the marker from both populations using power function $f(t) = -t^{-1/2}$. After this, we assume a linear regression model in each population for the transformed marker y, i.e.,

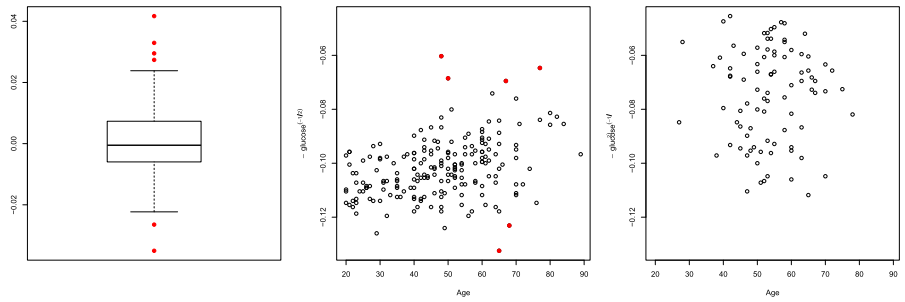

FIG 6. *The left panel corresponds to the boxplots of the residuals obtained after a robust fit for healthy sample, while the central and right panels to the scatter plots for the healthy and diseased samples. The red points correspond to the detected outliers.*

$$\begin{aligned} y_{D,i} &= \beta_{D,1} + \beta_{D,2}\, x_{D,i} + \epsilon_{D,i}\,, \qquad 1 \le i \le 88\,, \\ y_{H,i} &= \beta_{H,1} + \beta_{H,2}\, x_{H,i} + \epsilon_{H,i}\,, \qquad 1 \le i \le 198 \end{aligned}$$

and we compute the classical and robust estimators of the conditional ROC curves, denoted $\widehat{\mathrm{ROC}}_{\mathbf{x},\mathrm{CL}}$ and $\widehat{\mathrm{ROC}}_{\mathbf{x}}$, respectively. For the robust procedure, we compute MM-regression estimators using the bi-square loss function and an initial S-estimator with 50% breakdown point. The tuning constant in the M-step was set equal to 3.444 to ensure an 85% of efficiency with respect to the least squares estimator. In **Step 2R** we choose $w(t) = \mathbb{I}_{[-1,1]}(t)$, $\eta = 2.5$ and $G_{j,0} = \Phi$, as in the simulation study.

Based on the residuals boxplots of a robust fit, 6 outliers were detected in the healthy sample, labelled as 37, 78, 125, 137, 141 and 150, see the left panel of Figure 6. Note that Figures 1 and 2 illustrate the damaging effect that these outliers have on both the classical and naive estimates of the ROC surface.

The filled red points on the central panel of Figure 6 represent the atypical observations encountered in the healthy sample which correspond to vertical outliers. After removing them, the classical estimator of the conditional ROC curves is recomputed with the remaining points, namely $\widehat{\mathrm{ROC}}^{(-6)}_{\mathbf{x},\mathrm{CL}}$. The upper panel of Figure 7 displays the estimated surfaces with these three procedures using equidistant grids of points of size 29 and 28 in p and x, respectively, $p \in [0.01, 0.99]$ and $x \in [20, 87.5]$. In order to facilitate the comparison between the estimated surfaces, the lower panel in Figure 7 shows the differences between these estimators. This Figure makes evident that the robust estimator and the classical one when it is computed without the outliers are very similar all along the studied range, while the classical estimator computed from the whole sample shows a different pattern, especially for large values of age (see the left panel of Figure 7).

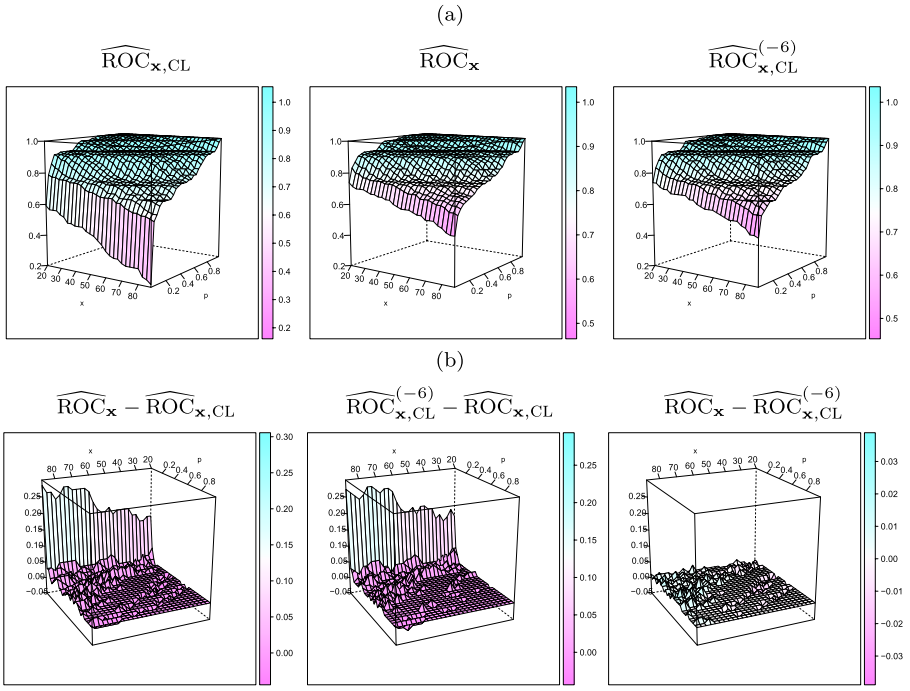

FIG 7. *Diabetes Data: (a) Estimated ROC surfaces and (b) Difference between the estimated ROC surfaces.*

6. Final remarks

The ROC curve is a useful graphical tool that measures the discriminating power of a biomarker to distinguish between two conditions or classes. When the practitioner is able to measure covariates related to the diagnostic variable which can increase the discriminating power, it is sensible to incorporate them in the analysis. To have a deeper comprehension of the effect of the covariates, it would be advisable to incorporate the covariates information to the ROC analysis instead of considering the marginal ROC curve. Conditional ROC curves may be easily estimated using a plug–in procedure. However, the use of classical regression estimators and empirical distribution and quantile functions may lead to estimates which may breakdown in the presence of a small amount of atypical data, as it is illustrated in our numerical experiment.

In this paper, we introduce a procedure to robustly estimate the conditional ROC curve. The methodology combines robust regression estimators with a weighted empirical distribution function which downweights the effect of large residuals. It is worth mentioning that our proposal may be easily implemented in `R` using the function `weighted.fractile` to compute the quantiles of the robust adaptive distribution function estimators and any function allowing to compute

robust estimators of the regression functions and scale parameters. When a linear model is assumed, robust MM-estimators are obtained through the function `lmrob` of the library `robustbase` which, by default, uses a bi-square loss function and an initial S-estimator with 50% breakdown point. Furthermore, when fitting a nonlinear model robust MM-estimators may be implemented through the function `nlrob`. When considering more complex models, such as partial linear or partial additive models, the function `lmrob` also allows an easy implementation of the estimators when B-splines are used to approximate the nonparametric components, see for instance, the proposals studied in Boente and Martínez (2021) or in Boente *et al.* (2020) who also considered the case of asymmetric log–Gamma errors.

We prove that the estimators are uniformly strongly consistent under standard regularity conditions when the sequence of cut-off parameters increases to infinity. A simulation study shows that our proposed estimators have good robustness and finite-sample statistical properties. Even though our numerical studies focus on a parametric regression approach, it should be mentioned that our proposal could also be implemented when considering nonparametric or partly parametric regression models simply combining a robust fit with the weighted empirical procedure. Using similar arguments to those given in Boente *et al.* (2020) or Boente and Martínez (2021), the consistency results stated in Propositions 1 and 2 may be extended to the case in which the biomarker is fitted by means of a partial linear or partial additive regression model and B-splines combined with MM-estimators are used in the first step. This interesting topic and as well as the extension to functional regressors will be object of future research.

Appendix A: Appendix

A.1. Proof of Theorem 1

To derive consistency results for the ROC curve, we will need the following assumptions on the errors distributions and on their estimates:

- **A1** $G_H : \mathbb{R} \to (0,1)$ has an associated density g_H such that $g_H(y) > 0$, for all $y \in \mathbb{R}$.
- **A2** $G_D : \mathbb{R} \to (0,1)$ is continuous.
- **A3** $\|\widehat{G}_j - G_j\|_\infty \xrightarrow{a.s.} 0$, $j = D, H$.
- **A4** For each fixed $\mathbf{x}$, $|\widehat{\mu}_j(\mathbf{x}) - \mu_{0,j}(\mathbf{x})| \xrightarrow{a.s.} 0$, $j = D, H$.
- **A5** For any compact set $\mathcal{K} \subset \mathcal{S}$, $\sup_{\mathbf{x}\in\mathcal{K}} |\widehat{\mu}_j(\mathbf{x}) - \mu_{0,j}(\mathbf{x})| \xrightarrow{a.s.} 0$, $j = D, H$.
- **A6** The regression functions $\mu_{0,j}$ are such that, for any compact set $\mathcal{K}$, $\sup_{\mathbf{x}\in\mathcal{K}} |\mu_{0,j}(\mathbf{x})| = A_j < \infty$.

Remark 1. If we are dealing with parametric regression models for $j = D, H$, i.e., when $\mu_{0,j}(\mathbf{x}) = f_j(\mathbf{x}, \beta_{0,j})$ and $\widehat{\beta}_j$ and $\widehat{\sigma}_j$ stand for robust consistent estimators

of $\beta_{0,j}$ and $\sigma^2_{0,j}$, respectively, the estimator of the ROC curve equals

$$\widehat{\mathrm{ROC}}_{\mathbf{x}}(p) = 1 - \widehat{G}_D\left(\frac{f_H(\mathbf{x},\widehat{\beta}_H) - f_D(\mathbf{x},\widehat{\beta}_D)}{\widehat{\sigma}_D} + \frac{\widehat{\sigma}_H}{\widehat{\sigma}_D}\,\widehat{G}_H^{-1}(1-p)\right).$$

In this framework, conditions under which **A3** holds for the linear model $f_j(\mathbf{x},\beta_{0,j}) = \mathbf{x}^{\mathrm{T}}\beta_{0,j}$ or more generally, for a non–linear model are given in Propositions 1 and 2.

On the other hand, **A4** to **A6** hold if the non–linear regression functions satisfy

A7 For each fixed $\mathbf{x}$, the regression functions $f_j(\mathbf{x},\mathbf{b})$ are continuous in $\mathbf{b}$.

A8 The functions f_j are such that $\sup_{\mathbf{x}\in\mathcal{K}}|f_j(\mathbf{x},\beta_n) - f_j(\mathbf{x},\beta_{0,j})| \to 0$ and $\sup_{\mathbf{x}\in\mathcal{K}}|f_j(\mathbf{x},\beta_j)| = A_j < \infty$, for any compact set $\mathcal{K}$ and any sequence $\beta_n \to \beta_{0,j}$.

In particular, these assumptions hold if the regression model is a linear one. ■

Proof of Theorem 1. We begin by proving (i). Using assumption **A3** for $j = H$ and the continuity of the quantile functionals when **A1** holds, we get that, for the healthy subjects, $\widehat{G}_H^{-1}(p) \xrightarrow{a.s.} G_H^{-1}(p)$, for each $0 < p < 1$. To avoid burden notation denote as

$$\widehat{\Delta}(\mathbf{x},p) = \frac{\widehat{\mu}_H(\mathbf{x}) - \widehat{\mu}_D(\mathbf{x})}{\widehat{\sigma}_D} + \frac{\widehat{\sigma}_H}{\widehat{\sigma}_D}\,\widehat{G}_H^{-1}(1-p)\,,$$
$$\Delta(\mathbf{x},p) = \frac{\mu_{0,H}(\mathbf{x}) - \mu_{0,D}(\mathbf{x})}{\sigma_{0,D}} + \frac{\sigma_{0,H}}{\sigma_{0,D}}\,G_H^{-1}(1-p)\,.$$

Note that the consistency of $\widehat{\sigma}_j$ and **A4** together with the fact that $\widehat{G}_H^{-1}(p) \xrightarrow{a.s.} G_H^{-1}(p)$, entail that for each fixed p and $\mathbf{x}$, $\widehat{\Delta}(\mathbf{x},p) \xrightarrow{a.s.} \Delta(\mathbf{x},p)$. Therefore, we have that,

$$\begin{aligned}
|\widehat{\mathrm{ROC}}_{\mathbf{x}}(p) - \mathrm{ROC}_{\mathbf{x}}(p)| &= \left|\widehat{G}_D\left(\widehat{\Delta}(\mathbf{x},p)\right) - G_D\left(\Delta(\mathbf{x},p)\right)\right| \\
&\le \left|\widehat{G}_D\left(\widehat{\Delta}(\mathbf{x},p)\right) - G_D\left(\widehat{\Delta}(\mathbf{x},p)\right)\right| + \left|G_D\left(\widehat{\Delta}(\mathbf{x},p)\right) - G_D\left(\Delta(\mathbf{x},p)\right)\right| \\
&\le \left\|\widehat{G}_D - G_D\right\|_\infty + \left|G_D\left(\widehat{\Delta}(\mathbf{x},p)\right) - G_D\left(\Delta(\mathbf{x},p)\right)\right|
\end{aligned}$$

which together with the continuity of G_D lead to $\widehat{\mathrm{ROC}}_{\mathbf{x}}(p) \xrightarrow{a.s.} \mathrm{ROC}_{\mathbf{x}}(p)$, for each fixed $\mathbf{x}$ and $0 < p < 1$. Note that for each fixed $\mathbf{x}$, $\mathrm{ROC}_{\mathbf{x}}(p)$ satisfies the conditions in Lemma S.1 in the supplementary file, so $\sup_{0<p<1}|\widehat{\mathrm{ROC}}_{\mathbf{x}}(p) - \mathrm{ROC}_{\mathbf{x}}(p)| \xrightarrow{a.s.} 0$.

(ii) Using that

$$\begin{aligned}
\left|\widehat{\Delta}(\mathbf{x},p) - \Delta(\mathbf{x},p)\right| &\le \frac{1}{\widehat{\sigma}_D}\left\{|\widehat{\mu}_H(\mathbf{x}) - \mu_{0,H}(\mathbf{x})| + |\widehat{\mu}_D(\mathbf{x}) - \mu_{0,D}(\mathbf{x})|\right\} \\
&\quad + \left|\frac{1}{\widehat{\sigma}_D} - \frac{1}{\sigma_{0,D}}\right|\,|\mu_{0,H}(\mathbf{x}) - \mu_{0,D}(\mathbf{x})| + \frac{\widehat{\sigma}_H}{\widehat{\sigma}_D}\left|\widehat{G}_H^{-1}(1-p) - G_H^{-1}(1-p)\right|
\end{aligned}$$

$$+ |G_H^{-1}(1-p)| \left| \frac{\widehat{\sigma}_H}{\widehat{\sigma}_D} - \frac{\sigma_{0,H}}{\sigma_{0,D}} \right| ,$$

assumption **A6**, the consistency of $\widehat{\sigma}_j$ and the uniform consistency of $\widehat{\mu}_j$, we get easily that $\sup_{\mathbf{x}\in\mathcal{K}} \left|\widehat{\Delta}(\mathbf{x},p) - \Delta(\mathbf{x},p)\right| \xrightarrow{a.s.} 0$. Hence, the inequality

$$\sup_{\mathbf{x}\in\mathcal{K}} |\widehat{\mathrm{ROC}}_{\mathbf{x}}(p) - \mathrm{ROC}_{\mathbf{x}}(p)| \quad \leq \quad \left\|\widehat{G}_D - G_D\right\|_\infty + \|g_D\|_\infty \sup_{\mathbf{x}\in\mathcal{K}} \left|\widehat{\Delta}(\mathbf{x},p) - \Delta(\mathbf{x},p)\right|$$

leads to $\sup_{\mathbf{x}\in\mathcal{K}} |\widehat{\mathrm{ROC}}_{\mathbf{x}}(p) - \mathrm{ROC}_{\mathbf{x}}(p)| \xrightarrow{a.s.} 0$.

Denote $\widehat{B} = \sup_{\delta<p<1-\delta} \sup_{\mathbf{x}\in\mathcal{K}} |\widehat{\mathrm{ROC}}_{\mathbf{x}}(p) - \mathrm{ROC}_{\mathbf{x}}(p)|$. Then, after some algebra we get that $\widehat{B} \leq \sum_{\ell=1}^{5} \widehat{B}_\ell$ where $\widehat{B}_1 = \|\widehat{G}_D - G_D\|_\infty$,

$$\begin{aligned}
\widehat{B}_2 &= \|g_D\|_\infty \left| \frac{1}{\widehat{\sigma}_D} - \frac{1}{\sigma_{0,D}} \right| (A_H + A_D) , \\
\widehat{B}_3 &= \|g_D\|_\infty \frac{1}{\widehat{\sigma}_D} \sup_{\mathbf{x}\in\mathcal{K}} \{|\widehat{\mu}_H(\mathbf{x}) - \mu_{0,H}(\mathbf{x})| + |\widehat{\mu}_D(\mathbf{x}) - \mu_{0,D}(\mathbf{x})|\} , \\
\widehat{B}_4 &= \frac{\widehat{\sigma}_H}{\widehat{\sigma}_D} \sup_{\delta<p<1-\delta} \left|\widehat{G}_H^{-1}(1-p) - G_H^{-1}(1-p)\right| , \\
\widehat{B}_5 &= \sup_{\delta<p<1-\delta} |G_H^{-1}(1-p)| \left| \frac{\widehat{\sigma}_H}{\widehat{\sigma}_D} - \frac{\sigma_{0,H}}{\sigma_{0,D}} \right| .
\end{aligned}$$

Assumptions **A1** to **A3** together with **A5** and the consistency of $\widehat{\sigma}_j$ entail that $\widehat{B}_\ell \xrightarrow{a.s.} 0$, for $\ell = 1, 2, 3, 5$. Note that $\widehat{G}_H^{-1}$ is a non–decreasing function. Besides, using that **A1** entails that G_H is continuous and strictly increasing, we immediately obtain that the quantile function G_H^{-1}, which in this case equals the inverse of G_H, is also strictly increasing and uniformly continuous over the compact interval $[\delta, 1-\delta]$. Hence, taking into account that $\widehat{G}_H^{-1}(p) \xrightarrow{a.s.} G_H^{-1}(p)$, for each $0 < p < 1$, analogous arguments to those considered in the proof of Lemma S.1 in the supplementary file when bounding $\sup_{a\leq t\leq b} |F_n(t) - F(t)|$ allow to derive that $\sup_{\delta\leq p\leq 1-\delta} \left|\widehat{G}_H^{-1}(1-p) - G_H^{-1}(1-p)\right| \xrightarrow{a.s.} 0$, concluding the proof of (ii).

We now proceed to derive (iii). Let $\epsilon > 0$ be fixed and choose $0 < \eta < 1$ such that, $\sup_{\mathbf{x}\in\mathcal{K}} \mathrm{ROC}_{\mathbf{x}}(\eta) < \epsilon/6$ and $\sup_{\mathbf{x}\in\mathcal{K}} (1 - \mathrm{ROC}_{\mathbf{x}}(1-\eta)) < \epsilon/6$. Denote as

$$\begin{aligned}
\widehat{R}(\eta) &= \sup_{\eta<p<1-\eta} \sup_{\mathbf{x}\in\mathcal{K}} |\widehat{\mathrm{ROC}}_{\mathbf{x}}(p) - \mathrm{ROC}_{\mathbf{x}}(p)| , \\
\widehat{R}_1(\eta) &= \sup_{p\leq\eta} \sup_{\mathbf{x}\in\mathcal{K}} |\widehat{\mathrm{ROC}}_{\mathbf{x}}(p) - \mathrm{ROC}_{\mathbf{x}}(p)| , \\
\widehat{R}_2(\eta) &= \sup_{1-\eta\leq p} \sup_{\mathbf{x}\in\mathcal{K}} |\widehat{\mathrm{ROC}}_{\mathbf{x}}(p) - \mathrm{ROC}_{\mathbf{x}}(p)| .
\end{aligned}$$

Hence, $\sup_{0<p<1} \sup_{\mathbf{x}\in\mathcal{K}} |\widehat{\mathrm{ROC}}_{\mathbf{x}}(p) - \mathrm{ROC}_{\mathbf{x}}(p)| \leq \widehat{R}(\eta) + \widehat{R}_1(\eta) + \widehat{R}_2(\eta)$, where $\widehat{R}(\eta) \xrightarrow{a.s.} 0$ from (ii). Besides, using that $\mathrm{ROC}_{\mathbf{x}}(p)$ is a distribution function and

$\widehat{\mathrm{ROC}}_{\mathbf{x}}(p)$ is non-decreasing in p, we get that for any $p \leq \eta$, $\mathbf{x} \in \mathcal{K}$, $|\widehat{\mathrm{ROC}}_{\mathbf{x}}(p) - \mathrm{ROC}_{\mathbf{x}}(p)| \leq \max\left\{\widehat{\mathrm{ROC}}_{\mathbf{x}}(\eta), \mathrm{ROC}_{\mathbf{x}}(\eta)\right\}$, so

$$\widehat{R}_1(\eta) \leq \sup_{\mathbf{x}\in\mathcal{K}} \max\left\{\widehat{\mathrm{ROC}}_{\mathbf{x}}(\eta), \mathrm{ROC}_{\mathbf{x}}(\eta)\right\} = \widehat{C}_1(\eta)\,.$$

Similarly, we obtain that

$$\widehat{R}_2(\eta) \leq \sup_{\mathbf{x}\in\mathcal{K}} \max\left\{1 - \widehat{\mathrm{ROC}}_{\mathbf{x}}(1-\eta), 1 - \mathrm{ROC}_{\mathbf{x}}(1-\eta)\right\} = \widehat{C}_2(\eta)\,.$$

Taking into account that $\sup_{\mathbf{x}\in\mathcal{K}} \mathrm{ROC}_{\mathbf{x}}(\eta) < \epsilon$ and $\sup_{\mathbf{x}\in\mathcal{K}}(1 - \mathrm{ROC}_{\mathbf{x}}(1-\eta)) < \epsilon$ and that for any fixed $0 < p < 1$, $\sup_{\mathbf{x}\in\mathcal{K}} |\widehat{\mathrm{ROC}}_{\mathbf{x}}(p) - \mathrm{ROC}_{\mathbf{x}}(p)| \xrightarrow{a.s.} 0$, we conclude that there exists $\mathcal{N}$ such that $\mathbb{P}(\mathcal{N}) = 0$ and for $\omega \notin \mathcal{N}$, $\widehat{R}(\eta) \to 0$, $\widehat{C}_1(\eta) \to \sup_{\mathbf{x}\in\mathcal{K}} \mathrm{ROC}_{\mathbf{x}}(\eta) < \epsilon/6$ and $\widehat{C}_2(\eta) \to \sup_{\mathbf{x}\in\mathcal{K}} 1 - \mathrm{ROC}_{\mathbf{x}}(1-\eta) < \epsilon/6$. Hence, for n_H and n_D large enough, we obtain that $\widehat{R}(\eta) < \epsilon/3$, $\widehat{C}_\ell(\eta) < \epsilon/3$, for $\ell = 1, 2$ which leads to $\sup_{0<p<1} \sup_{\mathbf{x}\in\mathcal{K}} |\widehat{\mathrm{ROC}}_{\mathbf{x}}(p) - \mathrm{ROC}_{\mathbf{x}}(p)| \leq \epsilon$, concluding the proof. ■

A.2. Proof of Propositions 1 and 2

The following assumptions are needed to show that $\|\widehat{G}_j - G_j\|_\infty \xrightarrow{a.s.} 0$ under a linear model

- **C1** The weight function $w : \mathbb{R} \to [0, 1]$ is even, non-increasing on $[0, +\infty)$, continuous, $w(0) = 1$, $w(u) > 0$ for $0 < u < 1$.
- **C2** G_j is a continuous distribution function.
- **C3** The estimators $\widehat{\beta}_j$ and $\widehat{\sigma}_j$ are such that $\widehat{\beta}_j \xrightarrow{a.s.} \beta_{0,j}$ and $\widehat{\sigma}_j \xrightarrow{a.s.} \sigma_{0,j}$.

The following Lemmas are needed to derive Propositions 1 and 2, their proof can be found in the supplementary file.

Lemma 1. *Assume that either $w(t) = \mathbb{I}_{[-1,1]}(t)$ or w satisfies **C1**. Let s be a fixed real number and define the class of functions*

$$\mathcal{F} = \left\{f_{\theta,\,\kappa,\nu}(u, \mathbf{x}) = w\left(\nu(u - \mathbf{x}^{\mathrm{T}}\theta)\right)\, \mathbb{I}_{u-\mathbf{x}^{\mathrm{T}}\theta\leq\kappa\, s} \text{ for } (\theta, \kappa, \nu) \in \mathbb{R}^p \times \mathbb{R}_{\geq 0} \times \mathbb{R}_{\geq 0}\right\}.$$

Then, we have that $\sup_{f\in\mathcal{F}} |P_{n_j} f - Pf| \xrightarrow{a.s.} 0$, where we use the standard notation in empirical processes, i.e., $P_{n_j} f = (1/n_j)\sum_{i=1}^{n_j} f(u_{j,i}, \mathbf{x}_{j,i})$ and $Pf = \mathbb{E}f(U_j, \mathbf{X}_j)$, with $u_{j,i} = \sigma_{0,j}\epsilon_{j,i}$ and $U_j = \sigma_{0,j}\epsilon_j$, for $j = D, H$.

Denote $\nu_0 = 1/(c_0\,\sigma_0)$, where we understand that if $c_0 = \infty$, $\nu_0 = 0$. Furthermore, let $\mathcal{J}_0$ be a compact interval with non–empty interior, such that $\nu_0 \in \mathcal{J}_0$.

To provide a general framework to deal with non–linear models, we consider the assumptions

- **C4** G_j has a bounded density g_j.

C5 $f_j(\mathbf{x},\beta)$ is a continuous function of β for each $\mathbf{x}$. Furthermore, for each $\mathbf{x}$, $\sup_{\|\beta-\beta_{0,j}\|\leq 1}|f_j(\mathbf{x},\beta)| \in L^1(P_j)$.

It is clear that **C4** entails **C2**. Lemma 2 is the non–linear counterpart of Lemma 1. Note that a bounded density is needed when a general non–linear model is considered, as well as a continuous weight function.

Lemma 2. *Assume that the observations* $(y_{j,i},\mathbf{x}_{j,i})$, $1\leq i\leq n_j$, *satisfy the non–linear regression model* (6). *Let* s *be a fixed real number and define the class of functions*

$$\mathcal{G}=\{g_{\beta,\,\sigma,\,\nu}(y,\mathbf{x})=w\left(\nu(y-f(\mathbf{x},\beta))\right)\mathbb{I}_{y-f(\mathbf{x},\beta)\leq\sigma\, s} \text{ for } (\beta,\sigma,\nu)\in\mathcal{V}_0\times\mathcal{I}_0\times\mathcal{J}_0\}\,.$$

Under ***C1***, ***C4*** *and* ***C5***, *we have that* $\sup_{g\in\mathcal{G}}|P_{n_j}g-Pg|\xrightarrow{a.s.}0$, *where* $Pg=\mathbb{E}g(Y_j,\mathbf{X}_j)$ *and* $P_{n_j}g=(1/n_j)\sum_{i=1}^{n_j}g(y_{j,i},\mathbf{x}_{j,i})$.

Proof of Proposition 1. When $c_{j,0}=\infty$, using that G_j is a bounded, monotone and continuous function and that $\widehat{G}_j$ is monotone, according to Lemma S.1 in the supplementary file, it will be enough to show that for each $s\in\mathbb{R}$, $\widehat{G}_j(s)\xrightarrow{a.s.}G_j(s)$. On the other hand, when $c_{j,0}<\infty$, standard arguments imply that $F(s)=h_{j,0}(c_{j,0},s)/h_{j,\infty}(c_{j,0})$ is a bounded, monotone and continuous function of s and the uniform convergence also follows from the pointwise one.

To avoid burden notation, we define u_i as $u_i=\sigma_{0,j}\epsilon_{j,i}$ and $U=\sigma_{0,j}\epsilon_j$. When there is no confusion, we will omit the subscript j. Hence, for instance, $\mathbf{X}$, $\sigma_{0,j}$, $\widehat{\beta}$, $\widehat{\sigma}$, c_n, n, h_∞ and h_0 will stand for $\mathbf{X}_j$, σ_0, $\widehat{\beta}_j$, $\widehat{\sigma}_j$, c_{n_j}, n_j, $h_{j,\infty}$ and $h_{j,0}$, respectively.

Denote as $\widehat{\nu}=1/(c_n\,\widehat{\sigma})$, $\nu_0=1/(c_0\,\sigma_0)$, where we understand that if $c_0=+\infty$, $\nu_0=0$. Then $\widehat{\nu}\xrightarrow{a.s.}\nu_0$.

We will begin by showing that for each fixed real number s, we have that

$$\frac{1}{n}\sum_{i=1}^{n}w_i\,\mathbb{I}(r_i\leq s)\xrightarrow{a.s.}h_0(c_0,s)=\begin{cases}\mathbb{E}_{G_j}w\left(\dfrac{\epsilon_j}{c_0}\right)\mathbb{I}_{\epsilon_j\leq s} & \text{if } c_0<\infty\\ \mathbb{E}_{G_j}\mathbb{I}_{\epsilon_j\leq s}=G_j(s) & \text{if } c_0=\infty\,.\end{cases}\tag{A.1}$$

For that purpose and noting that $r_i=(u_i-\mathbf{x}_i^{\mathrm{T}}\widehat{\theta})/\widehat{\sigma}$ with $\widehat{\theta}=\widehat{\beta}-\beta_0$, define the class of functions

$$\mathcal{F}=\{f_{\theta,\,\kappa,\nu}(u,\mathbf{x})=w\left(\nu(u-\mathbf{x}^{\mathrm{T}}\theta)\right)\mathbb{I}_{u-\mathbf{x}^{\mathrm{T}}\theta\leq\kappa\, s} \text{ for } (\theta,\kappa,\nu)\in\mathbb{R}^p\times\mathbb{R}_{\geq 0}\times\mathbb{R}_{\geq 0}\}\,.$$

Lemma 1 entails that $\sup_{f\in\mathcal{F}}|P_nf-Pf|\xrightarrow{a.s.}0$. Then, using that $P_nf_{\widehat{\theta},\,\widehat{\sigma},\,\widehat{\nu}}=(1/n)\sum_{i=1}^{n}w_i\mathbb{I}(r_i\leq s)$, we obtain that

$$\frac{1}{n}\sum_{i=1}^{n}w_i\mathbb{I}(r_i\leq s)-Pf_{\widehat{\theta},\,\widehat{\sigma},\,\widehat{\nu}}\xrightarrow{a.s.}0\,.$$

It remains to show that $Pf_{\widehat{\theta},\,\widehat{\sigma},\,\widehat{\nu}}\xrightarrow{a.s.}Pf_{0,\,\sigma_0,\,\nu_0}=h_0(c_0,s)$, which will follow if we derive that

$$A_n \;=\; Pf_{\widehat{\theta},\,\widehat{\sigma},\,\widehat{\nu}}-\mathbb{E}w\left(\nu_0U\right)\mathbb{I}_{U-\mathbf{X}^{\mathrm{T}}\widehat{\theta}\leq\widehat{\sigma}\, s}\xrightarrow{a.s.}0\tag{A.2}$$

$$B_n \quad = \quad \mathbb{E}w\left(\nu_0 U\right) \mathbb{I}_{U-\mathbf{X}^{\mathrm{T}}\widehat{\theta}\leq\widehat{\sigma}\, s} - h_0(c_0, s) \xrightarrow{a.s.} 0\,. \tag{A.3}$$

We begin by considering the situation where w satisfies **C1**. Noting that

$$\begin{aligned} |A_n| &= \left|\mathbb{E}\left\{w\left(\widehat{\nu}(U-\mathbf{X}^{\mathrm{T}}\widehat{\theta})\right) - w\left(\nu_0\, U\right)\right\}\mathbb{I}_{U-\mathbf{X}^{\mathrm{T}}\widehat{\theta}\leq\widehat{\sigma}\, s}\right| \\ &\leq \mathbb{E}\left|w\left(\widehat{\nu}(U-\mathbf{X}^{\mathrm{T}}\widehat{\theta})\right) - w\left(\nu_0\, U\right)\right|\,, \end{aligned}$$

using the Dominated Convergence Theorem, the continuity of w and the fact that $\widehat{\nu} \xrightarrow{a.s.} \nu_0$ and $\widehat{\theta} \xrightarrow{a.s.} 0$, we obtain that $A_n \xrightarrow{a.s.} 0$, concluding the proof of (A.2).

When $w = \mathbb{I}_{[-1,1]}$, we have that

$$\begin{aligned} w\left(\widehat{\nu}(u-\mathbf{x}^{\mathrm{T}}\widehat{\theta})\right) - w\left(\nu_0\, u\right) &= \mathbb{I}_{\widehat{\nu}(u-\mathbf{x}^{\mathrm{T}}\widehat{\theta})\leq 1}\, \mathbb{I}_{-1\leq\widehat{\nu}(u-\mathbf{x}^{\mathrm{T}}\widehat{\theta})} - \mathbb{I}_{\nu_0\, u\leq 1}\, \mathbb{I}_{-1\leq\nu_0\, u} \\ &= \mathbb{I}_{\widehat{\nu}(u-\mathbf{x}^{\mathrm{T}}\widehat{\theta})\leq 1}\, \left\{\mathbb{I}_{-1\leq\widehat{\nu}(u-\mathbf{x}^{\mathrm{T}}\widehat{\theta})} - \mathbb{I}_{-1\leq\nu_0\, u}\right\} \\ &\quad - \left\{\mathbb{I}_{\nu_0\, u\leq 1} - \mathbb{I}_{\widehat{\nu}(u-\mathbf{x}^{\mathrm{T}}\widehat{\theta})\leq 1}\right\}\, \mathbb{I}_{-1\leq\nu_0\, u}\,, \end{aligned}$$

so

$$\begin{aligned} |A_n| &\leq \mathbb{E}\left|w\left(\widehat{\nu}(U-\mathbf{X}^{\mathrm{T}}\widehat{\theta})\right) - w\left(\nu_0\, U\right)\right| \\ &\leq \mathbb{E}\left|\mathbb{I}_{-1\leq\widehat{\nu}(U-\mathbf{X}^{\mathrm{T}}\widehat{\theta})} - \mathbb{I}_{-1\leq\nu_0\, U}\right| + \mathbb{E}\left|\mathbb{I}_{\nu_0\, U\leq 1} - \mathbb{I}_{\widehat{\nu}(U-\mathbf{X}^{\mathrm{T}}\widehat{\theta})\leq 1}\right| \\ &\leq \mathbb{E}\left|\mathbb{I}_{-\widehat{\nu}^{-1}+\mathbf{X}^{\mathrm{T}}\widehat{\theta}\leq U} - \mathbb{I}_{-{\nu_0}^{-1}\leq U}\right| + \mathbb{E}\left|\mathbb{I}_{U\leq{\nu_0}^{-1}} - \mathbb{I}_{U\leq\widehat{\nu}^{-1}+\mathbf{X}^{\mathrm{T}}\widehat{\theta}}\right| \\ &\leq \mathbb{E}\left|\mathbb{I}_{U<-\widehat{\nu}^{-1}+\mathbf{X}^{\mathrm{T}}\widehat{\theta}} - \mathbb{I}_{U<-{\nu_0}^{-1}}\right| + \mathbb{E}\left|\mathbb{I}_{U\leq{\nu_0}^{-1}} - \mathbb{I}_{U\leq\widehat{\nu}^{-1}+\mathbf{X}^{\mathrm{T}}\widehat{\theta}}\right| \\ &\leq \mathbb{E}\left|G_j\left(\frac{1}{\sigma_0}\left[-\frac{1}{\widehat{\nu}}+\mathbf{X}^{\mathrm{T}}\widehat{\theta}\right]\right) - G_j\left(-\frac{1}{\sigma_0\ \nu_0}\right)\right| \\ &\quad + \mathbb{E}\left|G_j\left(\frac{1}{\sigma_0}\left[\frac{1}{\widehat{\nu}}+\mathbf{X}^{\mathrm{T}}\widehat{\theta}\right]\right) - G_j\left(\frac{1}{\sigma_0\ \nu_0}\right)\right|\,, \end{aligned}$$

where we understand that $\mathbb{I}_{u<-1/\nu_0} = 0$, $G_j\left(-1/\sigma_0\ \nu_0\right) = 0$, $\mathbb{I}_{u<1/\nu_0} = 1$ and $G_j\left(1/\sigma_0\ \nu_0\right) = 1$ if $\nu_0 = 0$. Now the proof follows from the continuity of G_j if $c_0 < \infty$ and from the fact that $\lim_{u\to-\infty} G_j(u) = 0$ while $\lim_{u\to+\infty} G_j(u) = 1$.

To derive (A.3), note that

$$|B_n| = \left|\mathbb{E}w\left(\nu_0 U\right)\, \mathbb{I}_{U-\mathbf{X}^{\mathrm{T}}\widehat{\theta}\leq\widehat{\sigma}\, s} - \mathbb{E}w\left(\nu_0\, U\right)\mathbb{I}_{U\leq\sigma_0\, s}\right| \leq \mathbb{E}\left|\mathbb{I}_{U\leq\mathbf{X}^{\mathrm{T}}\widehat{\theta}+\widehat{\sigma}\, s} - \mathbb{I}_{U\leq\sigma_0\, s}\right|\,.$$

If $\mathbf{x}^{\mathrm{T}}\widehat{\theta} + \widehat{\sigma}\, s \leq \sigma_0\, s$, then $\mathbb{I}_{u\leq\mathbf{x}^{\mathrm{T}}\widehat{\theta}+\widehat{\sigma}\, s} = 1$ implies that $\mathbb{I}_{u\leq\sigma_0\, s} = 1$, so that $\Delta(u) = \mathbb{I}_{u\leq\mathbf{x}^{\mathrm{T}}\widehat{\theta}+\widehat{\sigma}\, s} - \mathbb{I}_{u\leq\sigma_0\, s} = 0$. Similarly, if $\mathbb{I}_{u\leq\sigma_0\, s} = 0$, then $\mathbb{I}_{u\leq\mathbf{x}^{\mathrm{T}}\widehat{\theta}+\widehat{\sigma}\, s} = 0$ and $\Delta(u) = 0$. Therefore, when $\mathbf{x}^{\mathrm{T}}\widehat{\theta} + \widehat{\sigma}\, s \leq \sigma_0\, s$, $\Delta(u) = 1$ if and only if $\mathbf{x}^{\mathrm{T}}\widehat{\theta} + \widehat{\sigma}\, s < u \leq \sigma_0\, s$.

On the other hand, if $\mathbf{x}^{\mathrm{T}}\widehat{\theta} + \widehat{\sigma}\, s \geq \sigma_0\, s$, then $\Delta(u) = 1$ if and only if $\sigma_0\, s < u \leq \mathbf{x}^{\mathrm{T}}\widehat{\theta} + \widehat{\sigma}\, s$.

Note that the fact that $\widehat{\theta} \xrightarrow{a.s.} 0$ and $\widehat{\sigma} \xrightarrow{a.s.} \sigma_0$ entails that $\mathbf{x}^{\mathrm{T}}\widehat{\theta} + \widehat{\sigma}\, s \xrightarrow{a.s.} \sigma_0\, s$, for each $\mathbf{x}$. Let $\mathcal{C} = \{\mathbf{x} : \mathbf{x}^{\mathrm{T}}\widehat{\theta} + \widehat{\sigma}\, s \leq \sigma_0\, s\}$ and $\overline{\mathcal{C}}$ its complement, then

$$
\begin{aligned}
|B_n| &\leq \mathbb{E}\, \mathbb{I}_{\mathcal{C}}\, \mathbb{I}_{\mathbf{X}^{\mathrm{T}}\widehat{\theta}+\widehat{\sigma}\, s<U\leq\sigma_0\, s} + \mathbb{E}\, \mathbb{I}_{\overline{\mathcal{C}}}\, \mathbb{I}_{\sigma_0\, s<U\leq\mathbf{X}^{\mathrm{T}}\widehat{\theta}+\widehat{\sigma}\, s} \\
&\leq \mathbb{E}\, \mathbb{I}_{\mathcal{C}} \left\{ G_j(\, s) - G_j\left(\frac{\mathbf{X}^{\mathrm{T}}\widehat{\theta} + \widehat{\sigma}\, s}{\sigma_0} \right) \right\} + \mathbb{E}\, \mathbb{I}_{\overline{\mathcal{C}}} \left\{ G_j\left(\frac{\mathbf{X}^{\mathrm{T}}\widehat{\theta} + \widehat{\sigma}\, s}{\sigma_0} \right) - G_j(\, s) \right\} \\
&\leq \mathbb{E} \left| G_j\left(\frac{\mathbf{X}^{\mathrm{T}}\widehat{\theta} + \widehat{\sigma}\, s}{\sigma_0} \right) - G_j(\, s) \right|
\end{aligned}
$$

and (A.3) follows immediately from the continuity of G_j and **C3**, concluding the proof of (A.1).

Similar arguments allow to show that

$$
\frac{1}{n}\sum_{i=1}^{n} w_i \xrightarrow{a.s.} h_\infty(c_0) = \begin{cases} \mathbb{E}_{G_j} w\left(\dfrac{\epsilon_j}{c_0}\right) & \text{if } c_0 < \infty \\ 1 & \text{if } c_0 = \infty \end{cases} \tag{A.4}
$$

and the desired result follows now easily combining (A.1) and (A.4). ■

Proof of Proposition 2. As in the proof of Proposition 1, when $c_{0,j} = \infty$, using that **C4** implies that G_j is a bounded, monotone and continuous function and that $\widehat{G}_j$ is monotone, from Lemma S.1, it will be enough to show that for each $s \in \mathbb{R}$, $\widehat{G}_j(s) \xrightarrow{a.s.} G_j(s)$. On the other hand, when $c_{0,j} < \infty$, standard arguments allow to show that $F(s) = h_{j,0}(c_{j,0}, s)/h_{j,\infty}(c_{j,0})$ is a bounded, monotone and continuous function of s and the uniform convergence also follows from the pointwise one.

The proof follows the same steps as that of Proposition 1 and we will omit the subscript j to avoid burden notation, when there is no confusion.

Denote $\widehat{\nu} = 1/(c_n\, \widehat{\sigma})$, $\nu_0 = 1/(c_0\, \sigma_0)$, where we understand that if $c_0 = \infty$, $\nu_0 = 0$. Furthermore, let $\mathcal{J}_0$ be a compact interval with non–empty interior, such that $\nu_0 \in \mathcal{J}_0$. Taking into account that $\widehat{\nu} \xrightarrow{a.s.} \nu_0$, we have that with probability 1, for n large enough $\widehat{\nu} \in \mathcal{J}_0$.

As in the proof of Proposition 1, we will begin by showing that for each fixed real number s we have that

$$
\frac{1}{n}\sum_{i=1}^{n} w_i \mathbb{I}(r_i \leq s) \xrightarrow{a.s.} h_0(c_0, s) = \begin{cases} \mathbb{E}_{G_j} w\left(\dfrac{\epsilon_j}{c_0}\right) \mathbb{I}_{\epsilon_j \leq s} & \text{if } c_0 < \infty\,, \\ \mathbb{E}_{G_j} \mathbb{I}_{\epsilon_j \leq s} = G_j(s) & \text{if } c_0 = \infty\,. \end{cases} \tag{A.5}
$$

For that purpose and noting that $r_i = (y_i - f(\mathbf{x}, \widehat{\beta}))/\widehat{\sigma}$, define the class of functions

$$
\mathcal{G} = \{ g_{\beta,\, \sigma,\, \nu}(y, \mathbf{x}) = w\left(\nu(y - f(\mathbf{x}, \beta))\right) \mathbb{I}_{y - f(\mathbf{x},\beta) \leq \sigma\, s} \text{ for } (\beta, \sigma, \nu) \in \mathcal{V}_0 \times \mathcal{I}_0 \times \mathcal{J}_0 \}\,.
$$

Lemma 2 entails that $\sup_{g\in\mathcal{G}} |P_n g - P g| \xrightarrow{a.s.} 0$, then, using that

$$
\frac{1}{n}\sum_{i=1}^{n} w_i \mathbb{I}(r_i \leq s) = P_n g_{\widehat{\beta},\, \widehat{\sigma},\, \widehat{\nu}}\,,
$$

we obtain that

$$\frac{1}{n}\sum_{i=1}^{n} w_i \mathbb{I}(r_i \le s) - Pg_{\widehat{\beta},\,\widehat{\sigma},\,\widehat{\nu}} \xrightarrow{a.s.} 0\,.$$

It remains to show that $Pg_{\widehat{\beta},\,\widehat{\sigma},\,\widehat{\nu}} \xrightarrow{a.s.} Pg_{\beta_0,\,\sigma_0,\,\nu_0} = h_0(c_0, s)$, which will follow if we derive that

$$A_n = Pg_{\widehat{\beta},\,\widehat{\sigma},\,\widehat{\nu}} - \mathbb{E}w\,(\nu_0 U)\;\mathbb{I}_{Y - f(\mathbf{X},\widehat{\beta}) \le \widehat{\sigma}\, s} \xrightarrow{a.s.} 0 \tag{A.6}$$

$$B_n = \mathbb{E}w\,(\nu_0 U)\;\mathbb{I}_{Y - f(\mathbf{X},\widehat{\beta}) \le \widehat{\sigma}\, s} - h_0(c_0, s) \xrightarrow{a.s.} 0\,. \tag{A.7}$$

Noting that

$$\begin{aligned} |A_n| &= \left|\mathbb{E}\left\{w\left(\widehat{\nu}\left[Y - f(\mathbf{X},\widehat{\beta})\right]\right) - w\,(\nu_0\, U)\right\}\mathbb{I}_{Y - f(\mathbf{X},\widehat{\beta}) \le \widehat{\sigma}\, s}\right| \\ &\le \mathbb{E}\left|w\left(\widehat{\nu}\left[Y - f(\mathbf{X},\widehat{\beta})\right]\right) - w\,(\nu_0\, U)\right|\,, \end{aligned}$$

using the Dominated Convergence Theorem, the continuity of w and the fact that $\widehat{\nu} \xrightarrow{a.s.} \nu_0$ and $\widehat{\theta} \xrightarrow{a.s.} 0$, we obtain that $A_n \xrightarrow{a.s.} 0$, concluding the proof of (A.6).

To derive (A.7), using that $0 \le w(t) \le 1$, we get that

$$\begin{aligned} |B_n| &= \left|\mathbb{E}w\,(\nu_0 U)\;\mathbb{I}_{Y - f(\mathbf{X},\widehat{\beta}) \le \widehat{\sigma}\, s} - \mathbb{E}w\,(\nu_0\, U)\,\mathbb{I}_{U \le \sigma_0\, s}\right| \\ &\le \mathbb{E}\left|\mathbb{I}_{U \le f(\mathbf{X},\widehat{\beta}) - f(\mathbf{X},\beta_0) + \widehat{\sigma}\, s} - \mathbb{I}_{U \le \sigma_0\, s}\right|\,. \end{aligned}$$

As in the proof of Proposition 1, we have that, if $f(\mathbf{x},\widehat{\beta}) - f(\mathbf{x},\beta_0) + \widehat{\sigma}\, s \le \sigma_0\, s$, then $\Delta(u) = \mathbb{I}_{u \le \mathbf{x}^{\mathrm{T}}\widehat{\theta} + \widehat{\sigma}\, s} - \mathbb{I}_{u \le \sigma_0\, s} = 1$ if and only if $f(\mathbf{x},\widehat{\beta}) - f(\mathbf{x},\beta_0) + \widehat{\sigma}\, s < u \le \sigma_0\, s$.

On the other hand, if $f(\mathbf{x},\widehat{\beta}) - f(\mathbf{x},\beta_0) + \widehat{\sigma}\, s \ge \sigma_0\, s$, then $\Delta(u) = 1$ if and only if $\sigma_0\, s < u \le f(\mathbf{x},\widehat{\beta}) - f(\mathbf{x},\beta_0) + \widehat{\sigma}\, s$.

Note that the fact that $\widehat{\beta} \xrightarrow{a.s.} 0$ and $\widehat{\sigma} \xrightarrow{a.s.} \sigma_0$ together with the continuity of $f(\mathbf{x},\beta)$ entails that $f(\mathbf{x},\widehat{\beta}) - f(\mathbf{x},\beta_0) + \widehat{\sigma}\, s \xrightarrow{a.s.} \sigma_0\, s$, for each $\mathbf{x}$. Let $\mathcal{C} = \{\mathbf{x} : f(\mathbf{x},\widehat{\beta}) - f(\mathbf{x},\beta_0) + \widehat{\sigma}\, s \le \sigma_0\, s\}$ and $\overline{\mathcal{C}}$ its complement, then

$$\begin{aligned} |B_n| &\le \mathbb{E}\,\mathbb{I}_{\mathcal{C}}\,\mathbb{I}_{f(\mathbf{X},\widehat{\beta}) - f(\mathbf{X},\beta_0) + \widehat{\sigma}\, s < U \le \sigma_0\, s} + \mathbb{E}\,\mathbb{I}_{\overline{\mathcal{C}}}\,\mathbb{I}_{\sigma_0\, s < U \le f(\mathbf{X},\widehat{\beta}) - f(\mathbf{X},\beta_0) + \widehat{\sigma}\, s} \\ &\le \mathbb{E}\left|G_j\left(\frac{f(\mathbf{X},\widehat{\beta}) - f(\mathbf{X},\beta_0) + \widehat{\sigma}\, s}{\sigma_0}\right) - G_j(\,s)\right| \end{aligned}$$

and (A.7) follows immediately from the continuity of G_j and **C3**, concluding the proof of (A.5).

Similar arguments allow to show that

$$\frac{1}{n}\sum_{i=1}^{n} w_i \xrightarrow{a.s.} h_\infty(c_0) = \begin{cases} \mathbb{E}_{G_j} w\left(\dfrac{\epsilon_j}{c_0}\right) & \text{if } c_0 < \infty \\ 1 & \text{if } c_0 = \infty \end{cases} \tag{A.8}$$

and the desired result follows now easily combining (A.5) and (A.8). ■

Acknowledgments

The authors wish to thank the anonymous referees for their valuable comments which led to an improved version of the original paper.

Supplementary Material

Supplementary material to "Robust consistent estimators for ROC curves with covariates"
(doi: 10.1214/22-EJS2042SUPP; .pdf). The supplementary material contains some results regarding the choice of the cut-off values as well as the proofs of Lemmas 1 and 2. We also include the results of a numerical study performed under a linear regression model.

References

Alonzo, T. A. and Pepe, M. S. (2002). Distribution-free ROC analysis using binary regression techniques. *Biostatistics*, **3**, 421–432.

Bianco, A. M. and Spano, P. (2019). Robust inference for nonlinear regression models. *Test*, **28**, 369–398. MR3962059

Bianco, A. M., Boente, G. and González-Manteiga, W. Supplementary material to "Robust consistent estimators for ROC curves with covariates". 2022.

Bianco, A. M., Boente, G. and di Rienzo, J. (2000). Some results for robust GM-based estimators in heteroscedatic regression models. *Journal of Statistical Inference and Planning*,**89**, 215–242. MR1794423

Boente, G. and Fraiman, R. (1989). Robust nonparametric regression estimation for dependent observations. *Annals of Statistics*,**17**, 1242–1256. MR1015148

Boente, G. and Martínez, A. (2021). A robust spline approach in partially linear additive models. Preprint available at https://arxiv.org/abs/2107.12987

Boente, G., Rodríguez, D. and Vena, P. (2020). Robust estimators in a generalized partly linear regression model under monotony constraints. *Test*, **29**, 50–89. MR4063383

Cai, T. (2004). Semiparametric ROC regression analysis with placement values. *Biostatistics*, **5**, 45–60.

Carroll, R.J. and Ruppert, D. (1982). Robust estimation in heteroscedastic linear models. *Annals of Statistics*,**10**, 429–441. MR0653518

Faraggi, D. (2003). Adjusting receiver operating characteristic curves and related indices for covariates. *Journal of the Royal Statistical Society*, Ser. D, **52**, 1152–1174. MR1977259

Farcomeni, A. and Ventura, L. (2012). An overview of robust methods in medical research. *Statistical Methods in Medical Research*, **21**, 111–133. MR2893816

Gonçalves, L., Subtil, A., Oliveira, M. R. and Bermudez, P. (2014). ROC curve estimation: An overview. *REVSTAT-Statistical Journal*, **12**, 1–20. MR3195207

González-Manteiga, W., Pardo-Fernández, J. C. and Van Keilegom, I. (2011). ROC curves in non-parametric location-scale regression models.*Scandinavian Journal of Statistics*, **38**, 169–184. MR2760145

Greco, L. and Ventura, L. (2011). Robust inference for the stress-strength reliability. *Statistical Papers*, **52**, 773–788. MR2846684

Krzanowski, W. J. and Hand, D. J. (2009). *ROC curves for continuous data*. Chapman and Hall/CRC, Boca Raton. MR2522628

Pardo-Fernández, J. C., Rodríguez-Alvarez, M. X. and Van Keilegom, I. (2014). A review on ROC curves in the presence of covariates. *REVSTAT Statistical Journal*, **12**, 21–41. MR3195208

Pepe, M. S. (1997). A regression modelling framework for receiver operating characteristic curves in medical diagnostic testing. *Biometrika*, **84**, 595–608. MR1603993

Pepe, M. S. (1998). Three approaches to regression analysis of receiver operating characteristic curves for continuous test results. *Biometrics*, **54**, 124–135.

Pepe, M. S. (2003). *The Statistical Evaluation of Medical Tests for Classification and Prediction*, Oxford University Press, New York. MR2260483

Rodríguez-Álvarez, M. X., Roca-Pardiñas, J. and Cadarso-Suárez, C. (2011a). ROC curve and covariates: extending the induced methodology to the non-parametric framework. *Statistics and Computing*, **21**, 483–495. MR2826687

Rodríguez-Álvarez, M. X., Tahoces, P. C., Cadarso-Suárez, C. and Lado, M. J. (2011b). Comparative study of ROC regression techniques: Applications for the computer-aided diagnostic system in breast cancer detection. *Computational Statistics and Data Analysis*, **55**, 888–902. MR2736605

Smith, P. J. and Thompson, T. J. (1996). Correcting for confounding in analyzing receiver operating characteristic curves. *Biometrical Journal*, **38**, 857–863.

Sun, Y. and Genton, M. G. (2011). Functional boxplots. *Journal of Computational and Graphical Statistics*, **20**, 316–334. MR2847798

Walsh, S. J. (1997). Limitations to the robustness of binormal ROC curves: Effects of model misspecification and location of decision thresholds on bias, precision, size and power, *Statistics in Medicine*, **16**, 669–679.

Yohai, V. J. (1987). High breakdown-point and high efficiency robust estimates for regression. *Annals of Statistics*, **15**, 642–656. MR0888431