Econometrics and Statistics xxx (xxxx) xxx

Contents lists available at ScienceDirect

Econometrics and Statistics

journal homepage: www.elsevier.com/locate/ecosta

Addressing robust estimation in covariate–specific ROC curves

Ana M. Bianco[a,*], Graciela Boente[b]

[a] Instituto de Cálculo, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires and CONICET, Argentina
[b] Departamento de Matemáticas and Instituto de Cálculo, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires and CONICET, Argentina

ARTICLE INFO

Article history:
Received 7 September 2022
Revised 11 February 2023
Accepted 6 April 2023
Available online xxx

MSC:
62F35
62G35

Keywords:
Functional Covariates
Regression
Robustness
ROC curves

ABSTRACT

Proposals given in the field of ROC curves focusing on their robust aspects and contributions are considered. The motivation is the extended belief that ROC curves are robust. Without being exhaustive, some recent advances in the area are mentioned. The attention is placed on those situations where the presence of covariates related to the diagnostic marker may increase the discriminating power of the ROC curve. Recent robust procedures given in the framework of the induced methodology are extended to the situation where functional covariates are also present. Consistency results for this proposal are derived under mild conditions. The reported numerical study illustrates that the robust estimators of the covariate specific ROC curve improve the performance of the classical ones for contaminated samples.

1. Introduction

The Receiver Operating Characteristic curve (ROC curve) is a graphical tool that assesses the accuracy of a classification method. Nowadays it is a well–accepted technique for this purpose. In this sense, given a binary classifier, the ROC curve reflects how well this classifier discriminates between two different groups or classes. ROC curves were developed during World War II in the field of radar signal detection to analyse the capability of a radar receiver operator, who led to its name, to detect an enemy object and to distinguish it from simple noise. From the early 60’s ROC curves gained popularity in biomedicine and pharmaceutical environments and since then, they became a key tool to evaluate diagnostic tests and they grew into a very active area of research.

It is worth noticing that its use is widely extended to evaluate, for example, the performance of an instrumental system or a machine learning process, see Chapter 10 in Krzanowski and Hand (2009) for more applications. However, just to fix ideas, we will focus on medical diagnosis. In fact, in any decision making medical scenario the correct diagnosis of a patient is crucial and it is present in the everyday clinical practice. Let us suppose that the diagnostic test depends on a continuous variable, henceforth the marker Y and that, for certain value c, we classify as diseased all the subjects with Y greater or equal than the threshold c and as healthy, otherwise. Since subjects may be misclassified, two types of errors are involved in the assignation process, namely false positives and false negatives. The rates of these two misclassification errors do matter. Thus, two notions emerge as relevant: the *sensitivity*, that is related to the skill of correctly detecting the condition of interest, and the *specificity*, related to the ability of correctly assigning a subject to the non–diseased group. For instance,

* Corresponding author.
E-mail addresses: abianco@dm.uba.ar (A.M. Bianco), gboente@dm.uba.ar (G. Boente).

https://doi.org/10.1016/j.ecosta.2023.04.001

in a perfect diagnostic test one would choose the value of the threshold that leads to sensitivity and specificity equal to 1. However, this is not usually possible and it seems intuitive that the relationship between of sensitivity and specificity varies with the threshold c. The ROC curve is a statistical plot that represents for each value of c the sensitivity against the complementary of its specificity, that is, versus $1-$specificity.

In most situations it is usual to have additional information related to some other features of the patient, so, along with the marker, some covariates are often registered. As Pepe (2003) showed, in some cases the covariates impact on the performance of the ROC curve, so it is advisable to incorporate this additional information into the study. There are different ways to adapt the ROC curve methodology so as to introduce the covariate effects. On the one hand, some methods directly estimate the conditional distribution functions. In this direction, we can mention the work of López-de Ullibarri et al. (2008) who considered, in the conditional setup, the problem of fitting the ROC curve through a smooth estimator of the marker conditional distribution function in both populations. On the other hand, most of the procedures studied in the literature account for the covariate effect through regression models, by means either of the direct or the indirect method. In the direct methodology, the ROC curve is directly fitted through a generalized linear model using the covariates and suitable observations. In the indirect model, for each population, the markers distribution is modelled separately in terms of the covariates and just after, the induced ROC curve is computed.

As mentioned in Gonçalves et al. (2014), different uses of the word robustness have been considered in the literature. This paper tackles the concept of robustness in the sense of protection against anomalous data in the sample. Aware of the impact that outlying values may have on the diagnostic test accuracy, we center our attention on the robust aspects of the estimation procedures of the conditional ROC curve. Moreover, since regression models are involved in both the direct and induced approaches, atypical data among the responses or the covariates may severely affect the estimation methods. This robustness issue is even more complex when dealing with functional data, since, in such a situation, different types of atypical data may arise. As mentioned in Bali and Boente (2011) and Hubert et al. (2015), in a functional framework, the outliers need not to be "extreme" data points, but might consist of curves that behave differently from the others, or that display a persistent behaviour either in shift, amplitude and/or shape. More precisely, they may correspond to atypical trajectories entirely outlying, that is, with extreme values for the L^2 norm, to isolated points within otherwise typical trajectories (corresponding to a single extreme measurement) or they can be related to an extreme on some principal components, being the latter the more difficult to detect. This justifies the need of developing robust procedures to estimate the ROC curve when functional covariates are present.

In the first part of this paper, we will overview some of the contributions done in this field mainly focusing on the description of existing robust developments. Both, the selection of topics and given references, are far from being exhaustive. According to the data characteristics as well as to the model structure, we will highlight the connection between ROC curves and robust regression proposals.

Secondly, with this motivation, another contribution of this work rests extending the robust procedure given in Bianco et al. (2022) to regression models with increasing complexity. In fact, in some situations, together with the marker one may collect an additional infinite-dimensional covariate. In this case, the indirect methodology could be used, but it would involve a functional covariate. Inácio et al. (2012) extended the indirect procedure so as to include a functional explanatory variable, but they used classical procedures that could be unstable when facing outliers. Indeed, we generalize the proposal given in Bianco et al. (2022) to a very general scenario in which the markers are modelled in terms of a functional partially linear model. The considered situations include the functional linear regression model and also the nonparametric or additive regression ones with real valued covariates. In this way, the given approach enables to face with this robust perspective a vast class of cases. More precisely, we model the effect of the covariates over the markers through a general location–scale regression model, which may be nonparametric, additive, semiparametric and may include a functional covariate as well. We obtain consistency results for the proposed conditional ROC curve estimators in this very wide framework.

The remainder of the paper is organized as follows. Section 2.1 recalls basic definitions, while in Section 2.2 we discuss the different interpretations of robustness within the ROC curve literature. Some robust contributions for adjusted ROC curves are given in Section 2.3. For the case where functional covariates are present, we introduce our robust proposal in Section 3, where we also derive consistency results under mild conditions. A numerical study carried out to evaluate the finite sample performance of the proposals is described and analysed in Section 4. Some final remarks are given in Section 5, while proofs are relegated to the Appendix.

2. The ROC curve

2.1. Definition and basic notions

Assume that we are involved in a decision making scenario and we have two groups or classes. Just to fix ideas, we will identify them as diseased (D) and healthy (H). The classification rule or test depends on a continuous marker, Y, and a threshold value c, since it assigns an individual to the diseased group if $Y \geq c$ and to the healthy one if $Y < c$.

Let F_D be the distribution of the marker on the diseased population and F_H the distribution of Y in the healthy one. Henceforth, 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 clear that in this context, the *sensitivity* (true positive rate) and the complementary of the *specificity* (false positive rate) are functions of the threshold c and depend on these two distributions. In fact, the ability of the test to detect the condition of interest is measured through $1-F_D(c)$, while the inability of the test to identify healthy people is evaluated through $1-F_H(c)$. So, the practitioner's interest focusses on the study of the pairs $\{(1-F_H(c), 1-F_D(c))\}$ with $c \in \mathbb{R}$. The analysis of their evolution as c varies gives a complete picture of the performance of the classifier over all the possible threshold values. Indeed, these pairs describe a geometrical object called ROC curve. Now, taking into account that we can rewrite it in terms of $p = 1-F_H(c)$, we obtain the usual expression $\mathrm{ROC}(p) = 1-F_D(F_H^{-1}(1-p)))$, $p \in (0,1)$. Two properties distinguishes the ROC curve as a measure of accuracy of a given marker or test. First, the ROC curve is invariant under any strictly increasing transformation of the marker Y and secondly, it does no depend on the units in which Y is measured.

To facilitate the approach we will follow below when covariates are present, consider the case in which the distributions F_D and F_H belong to a location–scale family, i.e., for $j = D, H$, we have that

$$F_j(t) = G_j\left(\frac{t-\mu_j}{\sigma_j}\right), \tag{1}$$

where μ_j and σ_j denote the parameters of location and scale, respectively, and the distribution G_j has mean 0 and scale 1. Then, the ROC curve can be written as

$$\mathrm{ROC}(p) = G_D\left(\frac{\mu_H-\mu_D}{\sigma_D} + \frac{\sigma_H}{\sigma_D} G_H^{-1}(p)\right). \tag{2}$$

An extensively used model is the bi-normal one, that assumes that in both populations the marker is normally distributed, i.e., $G_j = \Phi$, for $j = D, H$. In this case, the distributions are characterized by the means μ_j and the standard deviations σ_j, $j = D, H$, leading to $\mathrm{ROC}(p) = \Phi\big((\mu_H-\mu_D)/\sigma_D + \sigma_H\Phi^{-1}(p)/\sigma_D\big)$. Since we are dealing with a curve, it is useful to summarize its whole information in a single number. Certainly the most widely used summary index is the area under the ROC curve (AUC) that measures the accuracy of the test and is formally defined as $\mathrm{AUC} = \int_0^1 \mathrm{ROC}(p)dp$. This summary index takes values between 0 and 1 corresponding to low and high discriminatory capability, respectively. After a change of variable and a bit of algebra, it is easy to see that $\mathrm{AUC} = \mathbb{P}(Y_D > Y_H)$, where Y_D and Y_H represents the markers of two independent individuals randomly chosen from the diseased and healthy populations, respectively. For that reason, values of AUC close to 1 suggest a high diagnostic accuracy of the marker. Another popular summary statistic is the Youden index, YI, which measures the difference between the ROC curve and the identity function, and is defined as $YI = \max_{0<p<1}\{\mathrm{ROC}(p) - p\}$. It is worth mentioning that when Y_D is stochastically greater that Y_H, then $\mathrm{ROC}(p) \geq p$, so the area under the curve is greater or equal than 0.5 and the Youden index is non–negative.

2.2. Estimation and robustness

Suppose that we have a random sample $y_{D,1}, \ldots, y_{D,n_D}$ from the diseased population where $y_{D,1} \sim F_D$ and another one from the healthy population $y_{H,1}, \ldots, y_{H,n_H}$ verifying that $y_{H,1} \sim F_H$. Assume that the samples are independent from each other.

Natural estimators of the ROC curve are based on the plug–in method. They can be obtained just by replacing the distribution functions by their empirical versions or by smooth estimators based on the convolution with a kernel. There exists a wide literature concerning nonparametric proposals for the ROC estimation, Pepe (2003) and Zhou et al. (2011) account for a very complete reference of the first contributions based on these approaches. More recently, Pulit (2016) also considered kernel smoothing estimators of the ROC curve based on the idea that this curve is a distribution function.

If we suppose that the distributions F_D and F_H belong to a location–scale family as in (1), an estimation procedure of the ROC curve given in (2) can be obtained following the next steps: i) compute estimators of the location–scale parameters $\widehat{\mu}_j$ and $\widehat{\sigma}_j$ for $j = D, H$, ii) compute the corresponding residuals and replace G_j by empirical distribution estimators, $\widehat{G}_j$ and iii) plug–in all these estimators in (2). This procedure yields the plug–in estimator given by $\widehat{\mathrm{ROC}}(p) = 1-\widehat{G}_D\big((\widehat{\mu}_H-\widehat{\mu}_D)/\widehat{\sigma}_D + \widehat{\sigma}_H\,\widehat{G}_H^{-1}(1-p)/\widehat{\sigma}_D\big)$. In the particular case, where the errors distribution is assumed to be known, as in the bi–normal model, $\widehat{G}_D = G_D$ and $\widehat{G}_H = G_H$ in the previous expression.

After the seminal work of Huber (1964), robust estimation and inference are topics that have gained a place among the different areas of Statistics. Rousseeuw and Leroy (1987), Dutter et al. (2003), Hampel et al. (2005), Huber and Ronchetti (2009), Maronna et al. (2019) and, more recently, Ronchetti (2021) gave a very complete overview of the proposed robust developments and results obtained in the last decades. However, robustness has received less attention in the topic of ROC curves and related indexes. Gonçalves et al. (2014) presented a discussion concerning the bi–normal model and its supposed robustness. However, they warned about the possible variations in the interpretation of the word robustness given by different authors. As Gonçalves et al. (2014) mentioned, the choice of the bi–normal model is sometimes justified by theoretical arguments, but also by customary reasons due to its simplicity. One other justification to fit a ROC curve using this model is its supposed robustness. The notion of robustness related to this paper focusses on studying how ROC curve estimators produce valid inferences in circumstances where atypical observations are present.

On the one hand, Walsh (1997) carried out a numerical experiment where it is shown that the bi–normal estimator is affected by model misspecifications. On the other hand, Farcomeni and Ventura (2012) revisited robust methods in medical research making focus on resistance with respect to outliers and illustrated the use of these procedures through applications to real data sets. In the context of stress-strength reliability, Greco and Ventura (2011) concerned about robust estimation and inference of the reliability parameter which is directly related to the area under the curve, namely AUC, and they proposed a methodology based on $M-$estimation. Recently, Ruli et al. (2022) uses the Tsallis scoring rule to provide confidence intervals for the AUC. An alternative option is to parametrically model the ROC curve and different estimators can be implemented using this approach. Focusing on these ideas, Devlin et al. (2013) explored the robustness properties of parametric modelling the ROC curve considering also misspecification of the ROC shape. In the last decades, the interest on the effects of model misspecification and outliers has increased in different fields as practitioners become more aware that robust procedures provide more stable tools for inferential purposes.

2.3. ROC curve with covariates

In the previous section, we have introduced ROC curves related to binary classifiers constructed from a primary marker Y. In practice, the discriminatory ability of the marker may be improved by several factors. In a medical scenario these may be age, gender, body mass index, systolic pressure or any other covariate of interest. Pepe (2003) exemplified through different situations how the discriminatory capability of a test is improved by the presence of covariates. When this is the case, it seems wise to assess the possible covariate effects in the ROC analysis so as to avoid oversimplification. In brief, we may say that the information registered all along the covariates may be incorporated in different ways. Pardo-Fernández et al. (2014) and Inácio et al. (2021b) gave a review on this topic.

Henceforth, we assume that, together with the markers Y_D and Y_H, covariates $\mathbf{X}_D$ and $\mathbf{X}_H$ are also registered for the diseased and healthy populations, respectively. Throughout this paper, we assume that the same covariates are measured in both populations, even when their distribution may vary. The notions related to the ROC curve and the summary indexes are naturally extended to this setting where the interest is focused on each $\mathbf{x}$ in the common support of $\mathbf{X}_D$ and $\mathbf{X}_H$, namely $\mathcal{S}$. Therefore, the conditional ROC curve is defined as

$$\mathrm{ROC}_{\mathbf{x}}(p) = 1 - F_D(F_H^{-1}(1-p|\mathbf{x})|\mathbf{x}), \tag{3}$$

where $F_j(\cdot|\mathbf{x})$ stands for the conditional distribution of $Y_j|\mathbf{X}_j = \mathbf{x}$, i.e., $F_j(y|\mathbf{x}) = \mathbb{P}(Y_j \le y|\mathbf{X}_j = \mathbf{x})$, for $j = D, H$. Analogously to the case considered in Section 2.1, in order to measure the performance of the discriminatory ability of the marker, the definition of the conditional area under the curve and the Youden index are extended as $\mathrm{AUC}_{\mathbf{x}} = \int_0^1 \mathrm{ROC}_{\mathbf{x}}(p)\, dp$ and $YI_{\mathbf{x}} = \max_{0<p<1}\{\mathrm{ROC}_{\mathbf{x}}(p) - p\}$.

Expression (3) suggests intuitive estimators based on an empirical approach just by simply plugging-in estimators of the conditional distributions functions. That is, given a value $\mathbf{x}$ and suitable estimators of $\widehat{F}_j(\cdot|\mathbf{x})$ obtained from independent samples, we could estimate $\mathrm{ROC}_{\mathbf{x}}(p)$ as $\widehat{\mathrm{ROC}}_{\mathbf{x}}(p) = 1 - \widehat{F}_D(\widehat{F}_H^{-1}(1-p|\mathbf{x})|\mathbf{x})$. To overcome the drawback of discontinuities that empirical estimators present, López-de Ullibarri et al. (2008) extended the nonparametric estimator of Peng and Zhou (2004) to the conditional setup by modelling the ROC curve through a smooth local linear estimate of the conditional marker distribution for diseased and non-diseased subjects. In the case of multidimensional covariates, an alternative nonparametric proposal that combines $B-$splines with a Bayesian approach was given in Inácio de Carvalho et al. (2013), avoiding in this way burden computations associated to the involved dependent Dirichlet processes.

Alternatively, in the direct and the induced methodologies regression tools are used to model the covariates effect on the classifier accuracy. On the one hand, in the direct modelling, the effect of the covariates is directly evaluated on the ROC curve by means of a generalized linear model, leading to the so-called ROC–GLM model. A crucial point to understand the direct approach is that the conditional ROC curve may be interpreted either as the conditional distribution function of the random variable $1 - F_H(Y_D|\mathbf{x})$, known as *placement values* (see Pepe, 2000 and Pepe and Cai, 2004 for further interpretation), or as the conditional expected value of the binary indicator $\mathbb{I}_{\{1-F_H(Y_D|\mathbf{x})<p\}}$ due to the following equalities

$$\mathrm{ROC}_{\mathbf{x}}(p) = \mathbb{P}(1 - F_H(Y_D|\mathbf{x}) < p|\mathbf{X}_D = \mathbf{x}) = \mathbb{E}\Big[\mathbb{I}_{\{1-F_H(Y_D|\mathbf{x})<p\}}|\mathbf{X}_D = \mathbf{x}\Big].$$

Hence, a regression model for the binary variables emerges as a natural possibility and it seems reasonable to postulate the following model to describe the functional relationship between the covariates and the ROC curve

$$\mathrm{ROC}_{\mathbf{x}}(p) = g(\mu(\mathbf{x}) + h(p)) . \tag{4}$$

The expression (4) has three components: the (inverse) link function $g: \mathbb{R} \to [0,1]$, the regression function μ that captures the effect of the covariates and the baseline function h, which models the shape of the ROC curve. Note that the baseline function represents the effect of the specificity on the sensibility. Usually, the functions h and g are chosen to ensure that the right hand side in (4) is effectively a monotone increasing function in p and satisfies $g(\mu(\mathbf{x}) + h(0)) = 0$ and $g(\mu(\mathbf{x}) + h(1)) = 1$, for all $\mathbf{x}$. The different assumptions made on these three components induce a vast statistical literature on the ROC–GLM model. For instance, it may be assumed that the link function g is known and that μ has either a parametric form, such as $\mu(\mathbf{x}) = \mathbf{x}^t\boldsymbol{\beta}$, or we may postulate that μ follows a nonparametric model, such as an additive one. Similarly a parametric or nonparametric model can be assumed for the baseline function h. Alonzo and Pepe (2002), Pepe (2000), Cai and Pepe (2002) and Cai (2004) contributed with the earliest developments in the direct methodology. Pepe (2000) and Alonzo and Pepe (2002) assumed a parametric model for the baseline function h and at the same time that the link function is known and μ is linear, while in Cai (2004) and Cai and Pepe (2002) the function h is unspecified. It is worth mentioning the generalized additive model assumed for the regression function μ proposed in Rodríguez-Álvarez et al. (2011b). This methodology leads to a ROC–GAM model for the ROC curve which may be a flexible tool to capture nonlinear trends in the covariates effect. Suppose that independent samples $(y_{j,i}, \mathbf{x}_{j,i})$, $1 \le i \le n_j$, $j = D, H$, are available. In many of these proposals the algorithm of estimation is very similar and based on a stepwise procedure that can be described as follows: i) from the sample $(y_{H,i}, \mathbf{x}_{H,i})$ compute an estimator $\widehat{F}_H(\cdot|\mathbf{x})$ of the conditional distribution $F_H(\cdot|\mathbf{x})$, ii) given a grid of points of p_ℓ of length n_p, compute the estimated *placement values* and with them the indicator variables $\widehat{U}_{\ell i} = \mathbb{I}_{\{1-\widehat{F}_H(y_{D,i}|\mathbf{x})<p_\ell\}}$ for $i = 1, \dots, n_D$, iii) fit the regression model postulated in (4) with the pseudo-observations $(\widehat{U}_{\ell i}, \mathbf{x}_{D,i})$. Charaf (2022) verified through a simulation study the sensitivity to outliers of this procedure when it is based on classical estimators. This author also numerically explored the stability of a robust stepwise version that combines robust parametric estimators of the regression components in the ROC–GLM model with an adaptive weighted empirical distribution based on the residuals from a robust fit in the healthy population. It is worth mentioning that robust methods in biostatistics including robust estimation in regression and generalized linear models are described in Heritier et al. (2009).

On the other hand, in the indirect or induced methodology, in both populations, the marker distribution is modelled separately in terms of the covariates and just after, the induced ROC curve is computed. A broad formulation to incorporate covariates in the ROC curve is through a general location–scale regression model. For the sake of simplicity, let us assume an homoscedastic regression model for both populations, that is,

$$Y_D = \mu_{0,D}(\mathbf{X}_D) + \sigma_{0,D}\,\epsilon_D \quad \text{and} \quad Y_H = \mu_{0,H}(\mathbf{X}_H) + \sigma_{0,H}\,\epsilon_H \,,$$

where, for $j = D, H$, $\mu_{0,j}$ is the true regression function and $\sigma_{0,j}$ corresponds to the scale parameter. For $j = D, H$, the errors $\epsilon_j \sim G_j$ are independent of $\mathbf{X}_j$ and have scale 1 to properly identify $\sigma_{0,j}$, while to identify the regression function it is assumed that $\mathbb{E}\epsilon_j = 0$. It is worth noticing that since the errors and the covariates are independent, for a given $\mathbf{x} \in \mathcal{S}$, we have that, for $j = D, H$,

$$F_j(y|\mathbf{x}) = \mathbb{P}(Y_j \le y|\mathbf{X}_j = \mathbf{x}) = G_j\left(\frac{y - \mu_{0,j}(\mathbf{x})}{\sigma_{0,j}}\right).$$

Therefore, the quantiles of the healthy conditional distribution are related to those of the errors through $F_H^{-1}(p|\mathbf{x}) = \sigma_{0,H}\, G_H^{-1}(p) + \mu_{0,H}(\mathbf{x})$, with $G_H^{-1}(\cdot)$ the quantile function of the errors ϵ_H. As a consequence, we obtain the following induced expression for the conditional ROC curve

$$\mathrm{ROC}_{\mathbf{x}}(p) = 1 - G_D\left(\frac{\mu_{0,H}(\mathbf{x}) - \mu_{0,D}(\mathbf{x})}{\sigma_{0,D}} + \frac{\sigma_{0,H}}{\sigma_{0,D}}\, G_H^{-1}(1-p)\right). \tag{5}$$

A clear advantage of this approach is that we can model the between the marker and the covariates and the user's preferences, the regression functions may be modelled parametrically, nonparametrically or partially parametrically. The works of Tosteson and Begg (1988), Pepe (1998), Faraggi (2003), Zheng and Heagerty (2004), González-Manteiga et al. (2011) and Rodríguez-Álvarez et al. (2011a) go in this direction. In particular, Faraggi (2003) worked in a semiparametric context assuming normal errors and an additive parametric model for the conditional means, while in a closed line of research Pepe (1998) did not assume a specific errors distribution and Dodd and Pepe (2003) use a semiparametric approach and regression models to estimate the area under the curve. For the case of univariate covariates, Yao et al. (2010), González-Manteiga et al. (2011) and Rodríguez-Álvarez et al. (2011a) proposed nonparametric estimators of the covariate-specific ROC curve using kernel-type regression estimators. Besides, in the multivariate covariates setting, Rodríguez and Martínez (2014) developed a Bayesian semiparametric regression model using Gaussian process priors to model the conditional mean of the marker for each of the populations under study. Some of these methods are implemented through a stepwise procedure based on the following steps: i) for $j = D, H$ and from the sample $(y_{j,1}, \mathbf{x}_{j,1}), \dots, (y_{j,n_j}, \mathbf{x}_{j,n_j})$ estimate the regression functions $\mu_{0,j}(\mathbf{x})$ and scale parameters $\sigma_{0,j}$, ii) compute the corresponding standardized residuals and calculate estimators of the diseased errors distributions and healthy quantile functions using their empirical versions and iii) plug–in all these estimators in (5).

Rodríguez-Álvarez et al. (2011c) gave a detailed review on the direct and induced methodologies. They also numerically compared the performance of these ROC regression techniques and among other characteristics, they considered the resistance to departures from the central model. It is worth mentioning that, since most of the existing proposals are based on classical least squares procedures or local averages, they may be badly influenced by outliers or by small deviations from the model assumptions. Motivated by this fact, Inácio et al. (2021a) and Bianco et al. (2022) looked for proposals to robustly estimate the covariate–specific ROC curve assuming a location–scale regression model for the marker in the diseased and healthy populations. Both papers considered the induced methodology, and exemplified the lack of robustness of the classical ROC curve estimator.

In particular, Inácio et al. (2021a) assumed an additive regression model for $\mu_{0,j}$ and combined additive $B-$splines with $M-$estimators, while for the estimation of the errors distributions they computed weighted empirical distribution functions of the standardized residuals. The implemented weighted empirical estimator is based on hard–rejection weights that use fixed cut–off constants γ_j, $j = D, H$, not depending on the sample size. Hence, for each sample, standardized residuals with absolute value larger than the corresponding γ_j are eliminated. The purpose of using a weighted empirical distribution function of the standardized residuals is clearly to control the influence of large residuals. In fact, even when the regression functions and scale parameters were robustly estimated, large standardized residuals corresponding to atypical data would appear and therefore, affect the classical empirical estimators of the distribution function and quantiles of the errors, see

Bianco et al. (2022) for an example. Despite that Inácio et al. (2021a) validated their proposal through a simulation study, their method would not be consistent, as proved in Bianco et al. (2022), since $\lim_{n_j\to\infty}\gamma_j\neq\infty$. Instead, the latter authors proposed an alternative robust stepwise procedure that indeed yields a consistent method. More precisely, their focus is on a semiparametric approach where first, a location–scale regression model is robustly fitted to each diagnostic variable and then, robust adaptive weighted empirical estimators based on the regression residuals are considered. Under mild assumptions the uniform consistency of the proposal is derived. At this stage, the choice of the cut–off constants γ_j does matter since consistency is obtained only when the sequence increases to infinity. In Section 3 we provide more details on this proposal.

3. Functional covariates

Similarly to what happens in the finite dimensional case, classical regression methods can be seriously affected when atypical observations arise among the functional covariates. Under a functional linear model, Maronna and Yohai (2013) and Kalogridis and Van Aelst (2019) illustrate through numerical experiments that the least squares approach is not resistant to atypical functional explanatory data, even when combined with splines or functional principal components to reduce the dimensionality. As when dealing with euclidean covariates, it is expected that this lack of robustness would be inherited by the conditional ROC curve.

As mentioned in the Introduction, when considering functional data, different types of atypical data may arise. For that reason their detection is generally a challenging problem. In the last years, several detection criteria have been given in the literature based on different notions of depths, dimension reduction and/or visualization tools. Among others, we can mention the procedures described in Sun and Genton (2011), Arribas-Gil and Romo (2014), Hubert et al. (2015), Rousseeuw et al. (2018), Dai and Genton (2019) and the references therein.

As it is well known, when considering linear regression models with covariates in $\mathbb{R}^p$, atypical data in the explanatory variables can be classified as good or bad high–leverage observations, since the first ones help in the fitting process. For that reason, when considering robust regression methods, outliers in the covariates are not automatically eliminated using some diagnostic method, but instead $MM-$estimators with bounded loss functions are used. The same behaviour may arise when dealing with functional covariates so, even when different outlier detection rules exist, it is better to adapt the best practices of robust estimation to this setting. In this section, we discuss how the proposal given in Bianco et al. (2022) can be extended to functional covariates specific ROC curves.

Estimation of the conditional ROC curve when the covariate is functional within the induced context has been considered in Inácio et al. (2012). These authors considered the situation in which each population follows a fully functional nonparametric model and also the case of an homoscedastic functional linear model. In the latter, the estimation is performed using a projection of the covariates over the linear space spanned by the first p estimated principal directions. In the former, a kernel–based estimator is considered using one of the semi–metrics described in Ferraty and Vieu (2006). Inácio de Carvalho et al. (2016) faced the problem of estimating the adjusted partial area under the specificity-ROC curve which is a useful measure and corresponds to the area for sensitivities in a specified interval. They also discussed the situation where functional and scalar covariates are used to adjust the ROC curve by fitting the regression through a semi-functional partial linear regression, that is, a model linear on the scalar covariates and nonparametric on the functional ones, see Aneiros-Pérez and Vieu (2006) for a description on how to estimate the regression function in such models.

In this paper, we also consider the situation where functional and scalar covariates are useful to adjust the average sensitivity for all specificity values. However, instead of a semi-functional partial linear regression, we deal with a general model for each population, that includes the functional partial linear model (FPLM), and we provide a detailed study of the consistency properties for the ROC curve, the area under the curve and the Youden index.

More precisely, we assume that for $j=D,H$

$$Y_j=\langle V_j,\beta_{0,j}\rangle+\eta_{0,j}(\mathbf{Z}_j)+\sigma_{0,j}\,\epsilon_j\,, \tag{6}$$

where $V_j\in L^2(\mathcal{I})$, $\mathbf{Z}_j\in\mathcal{Z}_j\subset\mathbb{R}^{q_j}$ is a bounded interval, $\sigma_{0,j}>0$ is the unknown error scale parameter and $\epsilon_j\sim G_j$ is independent of $\mathbf{X}_j$. The nonparametric component $\eta_{0,j}:\mathcal{Z}_j\to\mathbb{R}$ is an unknown smooth function and $\beta_{0,j}\in L^2(\mathcal{I})$. As usual $\langle\cdot,\cdot\rangle$ denotes the $L^2(\mathcal{I})$ inner product and $\mathcal{I}$ is also a compact interval which we assume to be [0,1]. To simplify the notation, and without loss of generality, we assume that $q_j=q$ and $\mathcal{Z}_j=\mathcal{Z}=[0,1]^q$, for $j=D,H$. Then, if $\mathbf{X}_j=(V_j,\mathbf{Z}_j)$, the support of both regression functions $\mu_{0,j}(\mathbf{X}_j)=\langle V_j,\beta_{0,j}\rangle+\eta_{0,j}(\mathbf{Z}_j)$ is $\mathcal{S}=L^2(0,1)\times[0,1]^q$. Note that by allowing multivariate covariates $\mathbf{Z}_j$, model (6) enables to consider a wide variety of models. Among others, we include the usual functional partial linear one (when $q=1$), the fully nonparametric or additive model (when $\beta_{0,j}=0$) or the functional partial linear additive one, when $\eta_{0,j}(\mathbf{Z}_j)=\alpha+\sum_{s=1}^q\eta_{0,j,s}(Z_{j,s})$ where $\int_0^1\eta_{0,j,s}(z)dz=0$ for identification purposes and $\mathbf{Z}_j=(Z_{j,1},\dots,Z_{j,q})^t$.

In the particular setting of model (6), the conditional ROC curve given in (5) may be expressed as

$$\mathrm{ROC}_{(V,\mathbf{z})}(p)=1-G_D\left(\frac{\langle V,\beta_{0,H}-\beta_{0,D}\rangle+\eta_{0,H}(\mathbf{z})-\eta_{0,D}(\mathbf{z})}{\sigma_{0,D}}+\frac{\sigma_{0,H}}{\sigma_{0,D}}\,G_H^{-1}(1-p)\right).$$

Hence, as mentioned in Section 2.3, estimates of the ROC curve can be obtained by plugging–in appropriate estimators of the unknown quantities.

The purpose of this section is to provide a framework where the ROC curve estimators defined in Bianco et al. (2022) remain be consistent when functional and scalar covariates give additional information and a semi-functional partial linear regression is assumed. Assume that we have independent samples from the diseased and healthy populations, that is, $(y_{j,i}, V_{j,i}, \mathbf{z}_{j,i})$, $1 \le i \le n_j$, are i.i.d. with the same distribution as $(Y_j, V_j, \mathbf{Z}_j)$, for $j = D, H$. To robustly estimate the ROC curve under model (6), we then extend their proposal through following stepwise procedure.

Step 1. Estimate $\beta_{0,j}$, $\eta_{0,j}$ and $\sigma_{0,j}$, for $j = D, H$ in a robust fashion from the samples $\{(y_{j,i}, \mathbf{x}_{j,i})\}_{1\le i\le n_j}$, where $\mathbf{x}_{j,i} = (V_{j,i}, \mathbf{z}_{j,i})$. Denote the resulting estimators by $\widehat{\beta}_j$, $\widehat{\eta}_j$ and $\widehat{\sigma}_j$.

Step 2. Compute for each sample the standardized regression residuals

$$\widehat{\epsilon}_{j,i} = \frac{y_{j,i} - \langle V_{j,i}, \widehat{\beta}_j \rangle + \widehat{\eta}_j(\mathbf{z}_{j,i})}{\widehat{\sigma}_j}\,.$$

From these residuals, estimate the errors distribution function using the adaptive weighted empirical distribution $\widehat{G}_j$ given by

$$\widehat{G}_j(s) = \frac{1}{\sum_{\ell=1}^{n_j} w_{j,\ell}} \sum_{i=1}^{n_j} w_{j,i} \mathbb{I}_{\{\widehat{\epsilon}_{j,i} \le s\}} \quad \text{with} \quad w_{j,i} = w\left(\frac{\widehat{\epsilon}_{j,i}}{\gamma_{n_j}}\right), \tag{7}$$

where γ_{n_j} is an increasing sequence of cut–off values and $w : \mathbb{R} \to [0, 1]$ 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 $u \ge 1$. Using (7), compute the robust adaptive estimators of the diseased distribution and healthy quantile functions, denoted, $\widehat{G}_D$ and $\widehat{G}_H^{-1}$, respectively.

Step 3. Plug–in the robust estimators computed in the first two steps into the expression of the conditional ROC curve to obtain

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

where $\mathbf{x} = (V, \mathbf{z})$ and $\widehat{\mu}_j(\mathbf{x}) = \langle V, \widehat{\beta}_j \rangle + \widehat{\eta}_j(\mathbf{z})$.

Estimates of the area under the curve may be obtained as $\widehat{\mathrm{AUC}}_{\mathbf{x}} = \int_0^1 \widehat{\mathrm{ROC}}_{\mathbf{x}}(p)dp$.

The following assumptions on the errors distributions and on the regression estimates are needed to derive consistency of the estimators of the conditional ROC curve and its related summary measures.

A1 $G_H : \mathbb{R} \to (0, 1)$ has an associated density g_H such that $g_H(u) > 0$, for all $u \in \mathbb{R}$.
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 $j = D, H$, the estimators $\widehat{\beta}_j$, $\widehat{\eta}_j$ and $\widehat{\sigma}_j$ are such that $\widehat{\sigma}_j \xrightarrow{a.s.} \sigma_{0,j}$ and for each fixed $(V, \mathbf{z}) \in L^2(0, 1) \times [0, 1]^q$, $\left|\langle V, \widehat{\beta}_j - \beta_{0,j} \rangle + \widehat{\eta}_j(\mathbf{z}) - \eta_{0,j}(\mathbf{z})\right| \xrightarrow{a.s.} 0$.

Conditions ensuring that assumptions A3 and A4 are fulfilled for some classes of robust estimators are presented in Section 3.2.

Theorem 1 below provides conditions ensuring that the resulting estimates of the ROC curve are consistent.

Theorem 1. *Let $\{(y_{j,i}, V_{j,i}, \mathbf{z}_{j,i})\}_{1\le i\le n_j}$, $j = D, H$, be independent observations satisfying (6) and let $\widehat{\beta}_j$, $\widehat{\eta}_j$ and $\widehat{\sigma}_j$ be estimators of $\beta_{0,j}$, $\eta_{0,j}$ and $\sigma_{0,j}$, respectively. Assume that A1 to A4 hold. Then, for each fixed $\mathbf{x} \in \mathcal{S}$, we have that*

(a) $\sup_{0<p<1} |\widehat{\mathrm{ROC}}_{\mathbf{x}}(p) - \mathrm{ROC}_{\mathbf{x}}(p)| \xrightarrow{a.s.} 0$,
(b) $\widehat{\mathrm{AUC}}_{\mathbf{x}} \xrightarrow{a.s.} \mathrm{AUC}_{\mathbf{x}}$,
(c) $\widehat{YI}_{\mathbf{x}} \xrightarrow{a.s.} YI_{\mathbf{x}}$.

The proof of Theorem 1 is omitted, since (a) follows using analogous arguments as those considered in the proof of Theorem 1 in Bianco et al. (2022). Moreover, with respect to (b) and (c), consistency of the estimators of the area under the curve and the Youden index is obtained straightforwardly from the inequalities $\left|\int_0^1 \widehat{\mathrm{ROC}}_{\mathbf{x}}(p)dp - \int_0^1 \mathrm{ROC}_{\mathbf{x}}(p)dp\right| \le \int_0^1 |\widehat{\mathrm{ROC}}_{\mathbf{x}}(p) - \mathrm{ROC}_{\mathbf{x}}(p)|dp$ and $\left|\widehat{YI}_{\mathbf{x}} - YI_{\mathbf{x}}\right| \le \sup_{0<p<1} |\widehat{\mathrm{ROC}}_{\mathbf{x}}(p) - \mathrm{ROC}_{\mathbf{x}}(p)|$.

3.1. Robust estimators of the regression function under a FPLM

It is worth mentioning that model (6) includes the functional linear regression one when $\eta_{0,j} \equiv \mu$ is a constant, and the nonparametric regression model, when $\beta_{0,j} \equiv 0$. This nonparametric regression model has been considered in González-Manteiga et al. (2011) to fit the marker and the asymptotic properties of the classical induced ROC estimators were studied therein. Robust nonparametric regression estimators have been considered by several authors who combined M–estimators with a preliminary scale estimator either with kernel weights or with B–splines or penalized splines. Regarding robust proposals based on kernels, we can mention the pioneering paper by Cleveland (1979), the local $M-$estimators proposed in Härdle and Tsybakov (1988) or Boente and Fraiman (1989) and the local regression quantiles studied in Welsh (1996). Penalized $M-$estimators were considered in Cox (1983), Cantoni and Ronchetti (2001) and Oh et al. (2007), while $S-$ and $M-$type smoothing splines including a penalty term were studied in Tharmaratnam et al. (2010) and Kalogridis and Van Aelst (2020, 2021) and Kalogridis (2021), respectively. Robust estimators for partial linear models when both covariates are euclidean were proposed in He and Shi (1996) and He et al. (2002), who studied $M-$estimators based on splines and in Bianco and Boente (2004) who considered robust estimators based on local $M-$estimators. Additive models are also included in model (6), for which robust kernel–based estimators were considered in Bianco and Boente (1998), Boente et al. (2017) and Boente and Martínez (2017). Moreover, robust $B-$spline estimators for additive models can be obtained as a particular case of those introduced in Boente and Martínez (2022) for partial linear additive models.

When considering functional linear regression models, the distorting effect of atypical responses and/or covariates on the procedures based on least squares, principal component and partial least squares approaches have already been described in the literature. Quantile regression estimators were studied in Cardot et al. (2005) and Kato (2012). However, these estimators are not resistant to bad high–leverage points. To overcome this drawback, robust estimators using $MM-$regression estimators combined with a penalization term were proposed in Maronna and Yohai (2013), while $B-$splines estimators were obtained as a particular case of those considered in Boente et al. (2020) who derived their consistency and convergence rates. Finally, a different approach was studied in Kalogridis and Van Aelst (2019) who proposed robust estimators based on a dimensionality reduction principle using robust functional principal components. They also considered a penalized version of these estimators and derived their consistency.

Following this last line of research, for functional partial linear models that is when $Z_j \in \mathbb{R}$, Qingguo (2015) proposed $M-$estimators using functional principal components for the slope. However, this author estimated the principal directions using the eigenfunctions of the sample covariance operator leading to a final procedure which is not resistant to outliers in the functional covariates. Besides, the proposed $M-$estimators used a monotone score function and no scale estimator was considered to standardize residuals. In contrast, Huang et al. (2015) used $B-$splines and allowed for bounded loss functions, however, their M-estimators are not scale equivariant. To overcome these issues, Boente et al. (2020) first computed a robust preliminary scale estimator and then, considered $M-$regression estimators with a bounded loss function for the slope and the nonparametric function. To reduce the dimensionality, they approximated both the functional regression parameter and the nonparametric component by means of B-splines.

For the sake of completeness, as an illustrative example of an approach based on a fixed basis approximating method, we recall the definition of the proposal given in Boente et al. (2020), considering only the diseased sample $(y_{D,i}, V_{D,i}, z_{D,i})$, $1 \le i \le n_D$. To define the B-splines estimators, we fix its order ℓ and the number of knots $m_{n_D}^{(1)}$ and $m_{n_D}^{(2)}$ to be used to approximate $\beta_{0,D}$ and $\eta_{0,D}$, respectively. Then, the corresponding (normalized) B-splines bases have dimensions $k_{n_D,\beta} = m_{n_D}^{(1)} + \ell$ and $k_{n_D,\eta} = m_{n_D}^{(2)} + \ell$, respectively (see Corollary 4.10 of Schumaker, 1981). Denote these bases by $\{B_s^{(1)} : 1 \le s \le k_{n_D,\beta}\}$ and $\{B_s^{(2)} : 1 \le s \le k_{n_D,\eta}\}$ and to simplify the notation, denote their sizes with $p_1 = p_{1,D} = k_{n_D,\beta}$ and $p_2 = p_{2,D} = k_{n_D,\eta}$, respectively.

Boente et al. (2020) defined robust $MM-$estimators using bounded ρ-functions, as defined in Maronna et al. (2019).

To define the estimators, for any vectors $\mathbf{b} \in \mathbb{R}^{p_1}$ and $\mathbf{a} \in \mathbb{R}^{p_2}$, let $r_i(\beta_\mathbf{b}, \eta_\mathbf{a})$, $1 \le i \le n_D$, be the residuals with respect to the spline approximations of β_0 and η_0 given by $\beta_\mathbf{b}(t) = \sum_{s=1}^{p_1} b_s B_s^{(1)}(t)$ and $\eta_\mathbf{a}(z) = \sum_{s=1}^{p_2} a_s B_s^{(2)}(z)$, respectively. More precisely, $r_i(\beta_\mathbf{b}, \eta_\mathbf{a}) = y_{D,i} - \mathbf{b}^t \mathbf{v}_{D,i} - \mathbf{a}^t \mathbf{B}_{D,i}$, with $\mathbf{v}_{D,i} = (v_{i1}, \ldots, v_{ip_1})^t$, $v_{is} = \langle V_{D,i}, B_s^{(1)} \rangle$ and $\mathbf{B}_{D,i} = (B_1^{(2)}(z_{D,i}), \ldots, B_{p_2}^{(2)}(z_{D,i}))^t$. The case $\beta_{0,D} \equiv 0$ corresponds to the nonparametric regression model, so in the above expression we understand that $\mathbf{b} = \mathbf{0}$ and the final estimator equals 0. Similarly, when the model is a functional linear one, that is, when no additional covariates $z_{D,i}$ are available, we understand that $p_2 = 1$ and $\mathbf{B}_{D,i} = 1$, while $\eta_\mathbf{a} = a \in \mathbb{R}$ are the candidates to estimate the model intercept.

The $MM-$procedure for the FPLM can be summarized as follows. First, we recall that the preliminary scale estimator is defined using a bounded $\rho-$function ρ_0. For that purpose, let $s_n(\beta_\mathbf{b}, \eta_\mathbf{a})$ be the M-scale estimator of the residuals solution to the following equation

$$\frac{1}{n_D - (p_1 + p_2)} \sum_{i=1}^{n_D} \rho_0\left(\frac{r_i(\beta_\mathbf{b}, \eta_\mathbf{a})}{s_{n_D}(\beta_\mathbf{b}, \eta_\mathbf{a})}\right) = b, \tag{8}$$

where $b = E(\rho_0(\epsilon_D))$. The residual scale estimator is given by

$$\widehat{\sigma}_D = \min_{\mathbf{b},\mathbf{a}} s_{n_D}(\beta_\mathbf{b}, \eta_\mathbf{a}). \tag{9}$$

Secondly, to define the final estimators, we consider a $\rho-$function ρ_1, such that $\rho_1 \le \rho_0$ and $\sup_t \rho_1(t) = \sup_t \rho_0(t)$ and compute an M-estimator using the residual scale estimator $\widehat{\sigma}_D$ and the loss function ρ_1

$$(\widehat{\mathbf{b}}_D, \widehat{\mathbf{a}}_D) = \text{argmin}_{\mathbf{b},\mathbf{a}} \sum_{i=1}^{n_D} \rho_1\left(\frac{r_i(\beta_\mathbf{b}, \eta_\mathbf{a})}{\widehat{\sigma}_D}\right). \tag{10}$$

The resulting estimators of the regression function $\beta_{0,D}$ and the nonparametric component $\eta_{0,D}$ are given by

$$\widehat{\beta}_D(t) = \sum_{s=1}^{p_1} \widehat{b}_{D,s} B_s^{(1)}(t) \quad \text{and} \quad \widehat{\eta}_D(z) = \sum_{s=1}^{p_2} \widehat{a}_{D,s} B_s^{(2)}(z), \tag{11}$$

where $\widehat{\mathbf{b}}_D = (\widehat{b}_{D,1}, \ldots, \widehat{b}_{D,p_1})^t$ and $\widehat{\mathbf{a}}_D = (\widehat{a}_{D,1}, \ldots, \widehat{a}_{D,p_2})^t$.

The choice of p_1 and p_2 is a task that deserves attention. Boente et al. (2020) described a robust criterion to select the bases dimensions and gave conditions to guarantee that $\widehat{\sigma}_D \xrightarrow{a.s.} \sigma_{0,D}$ and $\|\widehat{\beta}_D - \beta_{0\,D}\|_\infty + \|\widehat{\eta}_D - \eta_{0\,D}\|_\infty \xrightarrow{a.s.} 0$.

3.2. Consistency of the ROC curve when considering basis estimators

In this section, we discuss conditions ensuring that $\sup_{0<p<1} |\widehat{\mathrm{ROC}}_{\mathbf{x}}(p) - \mathrm{ROC}_{\mathbf{x}}(p)| \xrightarrow{a.s.} 0$ for some families of estimators under model (6). For the sake of simplicity, henceforth we will assume that $q = 1$, that is, the usual functional partial linear model. When $q > 1$, the description given below can be adapted using appropriate bases. From Theorem 1, it will be enough to prove that $\|\widehat{G}_j - G_j\|_\infty \xrightarrow{a.s.} 0$, $j = D, H$ and that assumption A4 holds.

The family of robust $MM-$estimators defined in Section 3.1 is based on $B-$splines which provide a flexible basis, often accurate and computational convenient. However, other bases such as that of natural splines of order ℓ used in Maronna and Yohai (2013) or that obtained from Bernstein polynomials may be used to construct finite dimensional candidates in (9) and (11). For that reason, in this section, we consider a more general family of estimators. More precisely, we assume that, for $j = D, H$, the estimators of the functional slope and nonparametric component belong to a finite–dimensional space spanned by fixed bases denoted $\{B_{j,1}^{(1)}, \ldots, B_{j,p_{1,j}}^{(1)}\}$ and $\{B_{j,1}^{(2)}, \ldots, B_{j,p_{2,j}}^{(2)}\}$, whose dimensions increase with the sample size. Hence, the estimators can be written as

$$\widehat{\beta}_j(t) = \sum_{s=1}^{p_{1,j}} \widehat{b}_{j,s} B_{j,s}^{(1)}(t) \quad \text{and} \quad \widehat{\eta}_j(z) = \sum_{s=1}^{p_{2,j}} \widehat{a}_{j,s} B_{j,s}^{(2)}(z),$$

where $\widehat{\mathbf{b}}_j = (\widehat{b}_{j,1}, \ldots, \widehat{b}_{j,p_{1,j}})^t$ and $\widehat{\mathbf{a}}_j = (\widehat{a}_{j,1}, \ldots, \widehat{a}_{j,p_{2,j}})^t$ are the estimators coefficients, so that the inner product $\langle V, \widehat{\beta}_j\rangle$ equals $\sum_{s=1}^{p_{1,j}} \widehat{b}_{j,s} \langle V, B_{j,s}^{(1)}\rangle = \widehat{\mathbf{b}}_j^{\,t} \mathbf{v}_j$ with $\mathbf{v}_j = (v_{j,1}, \ldots, v_{j,p_{1,j}})^t$ and $v_{j,s} = \langle V, B_{j,s}^{(1)}\rangle$.

For brevity, in assumptions D2 and D3, it will be understood that $j = D, H$.

- D1 The weight function $w : \mathbb{R} \to [0, 1]$ is even, non-increasing on $[0, +\infty)$, continuous, $w(0) = 1$, $w(u) > 0$ for $0 < u < 1$.
- D2 G_j is a continuous distribution function.
- D3 The bases dimensions $p_{1,j}$ and $p_{2,j}$ are such that $p_{1,j} + p_{2,j} = o(n_j)$.

Note that D2 corresponds to assumption A2 when $j = D$, while A1 implies D2 for $j = H$.

The following result is an extension of Proposition 1 in Bianco et al. (2022) to the functional partial linear model. Its proof may be found in the Appendix.

Proposition 1. *For $j = D$ or H, let $\left(y_{j,i}, V_{j,i}, z_{j,i}\right) \in \mathbb{R} \times L^2(0, 1) \times [0, 1]$, $1 \le i \le n_j$, be observations that satisfy the functional partial linear regression model (6). Assume that $w(t) = \mathbb{I}_{[-1,1]}(t)$ or w satisfies assumption D1. Let $\widehat{G}_j$ be defined as in (7) with $\gamma_{n_j} \xrightarrow{a.s.} \infty$. Then, under A4, D2 and D3, we have that $\|\widehat{G}_j - G_j\|_\infty \xrightarrow{a.s.} 0$.*

Note that in the statement of the Proposition 1, the sequence γ_{n_j} may be random, meaning that the cut–off values may be adaptive and hence, data dependent.

It is worth mentioning that assumption A4 is a consistency requirement for the estimated regression function and it holds whenever $\|\widehat{\beta}_j - \beta_{0\,j}\| + \|\widehat{\eta}_j - \eta_{0\,j}\|_\infty \xrightarrow{a.s.} 0$, where $\|\cdot\|$ stands for the norm in $L^2(0, 1)$. The results in Boente et al. (2020) ensure that the robust $MM-$estimators based on $B-$splines defined in Section 3.1 through (9) and (11) satisfy A4. In the case of partial additive linear models with real–valued covariates, Boente and Martínez (2022) highlighted that, when using a general basis, one key point to derive uniform consistency results for the estimators of the nonparametric components is that there exists an element in the finite–dimensional linear space spanned by the chosen basis providing a uniform approximation to the true function at a given rate. A similar argument is valid in our framework by taking the spaces spanned by $\{B_{j,s}^{(2)}, 1 \le j \le p_{2,j}\}$ or $\{B_{j,s}^{(1)}, 1 \le j \le p_{1,j}\}$ when estimating $\eta_{0,j}$ or $\beta_{0,j}$, respectively. When considering Bernstein polynomials, Theorem 3.2 in Powell (1981) ensures this condition. Note that this uniform approximation condition is a requirement in Newey (1997) who derived weak consistency results for series function estimators under a nonparametric regression model.

We describe below some particular cases of the functional partial linear model which have major interest. The first one is the nonparametric regression model which corresponds to $\beta_{0,j} \equiv 0$. For this case, our results show that the robust ROC curve estimators considered in this paper are consistent, when either robust $B-$ or $P-$spline estimators or Bernstein polynomials are used to estimate $\eta_{0,j}$.

The second example corresponds to a functional linear model, that is, when $\eta_{0,j} \equiv 0$. Under such model, Proposition 1 and Theorem 1 show that the robust alternative to the ROC procedure described in Inácio et al. (2012) is

Table 1
Summary measures for the classical (CL) and robust (ROB) estimates of the conditional AUC and ROC curves and for the Youden index, labelled YI.

	AUC		ROC		YI	
	KS	*MSE*	*KS*	*MSE*	*KS*	*MSE*
	C_0					
CL	0.044	0.001	0.108	0.001	0.063	0.001
ROB	0.047	0.001	0.111	0.002	0.066	0.002
	$C_{1,H}$					
CL	0.099	0.005	0.707	0.016	0.129	0.008
ROB	0.047	0.001	0.114	0.002	0.066	0.002
	$C_{2,H}$					
CL	0.087	0.004	0.374	0.018	0.209	0.023
ROB	0.047	0.001	0.114	0.002	0.066	0.002

consistent if instead of reducing the dimensionality by means of functional principal components, we use $B-$splines. It is worth noticing that the classical approach using empirical distribution functions is included in our results taking $w \equiv 1$, so by means of $B-$splines our proposal also provides consistent estimators in the classical framework. As mentioned above, assumption A4 allows to include slope estimators defined not only using $B-$splines for which the results in Boente et al. (2020) guarantee that $\|\widehat{\beta}_j - \beta_{0\,j}\| \xrightarrow{a.s.} 0$, but also those obtained taking other finite–dimensional fixed bases combined eventually with a penalization term. For instance, penalized estimators in functional linear regression have been defined in Crambes et al. (2009) or Cardot et al. (2005) who introduced a penalty on the second derivative of the candidates combined with least squares or quantile estimators, respectively.

It is worth mentioning that our results can easily be extended to the situation of a partial linear additive model or a purely additive one, using similar arguments to those considered in the proof of Proposition 1 combined with those in Lemmas A.2 and A.3 from Boente and Martínez (2022).

4. Monte Carlo study for functional covariates

In this section, we summarize the results of a Monte Carlo study, where we consider a functional linear model, which is a particular case of the FPLM considered in Section 3, and $B-$spline based estimators for the slope coefficient. Through this numerical study, we illustrate the lack of resistance of the classical procedure when atypical observations arise in the samples. At the same time, the simulation reveals the stability of our proposal that combines robust adaptive weighted empirical distribution functions with robust $MM-$estimators. More precisely, the classical estimator combines a least squares approach to estimate the intercept, the regression slope and the scale parameter and the residuals empirical distribution to obtain the final ROC curve. In contrast, the robust procedure uses the estimators described in Section 3.1 to estimate the parameters of both the healthy and diseased populations and the adaptive weighted empirical distribution functions defined in (7). For the MM-estimators, we chose as $\rho-$function the very well known Tukey bisquare function with tuning constants $c_0 = 1.54764$ $(b = 1/2)$ in (8) and $c_1 = 3.444$ in (10). To compute the weights in (7), we used the hard–rejection function $w(t) = \mathbb{I}_{[-1,1]}(t)$, while the cut–off values were chosen as described in Bianco et al. (2022). We selected the bases dimension for the robust $MM-$estimators by means of the robust Bayesian information criterion defined in Boente et al. (2020) using as loss function ρ_1. Instead, for the classical procedure the squared loss function was considered. In all cases, cubic splines and equally spaced knots were used. All calculations were performed in R and the simulation codes are available at https://github.com/gboente/FDA-ROC

Our simulation framework is the following. For $j = D, H$, we considered i.i.d. samples $(y_{j,i}, V_{j,i})$, $1 \le i \le n_j$, generated with the same distribution as (Y_j, V_j) which satisfy the following functional linear model $Y_j = \alpha_j + \langle V_j, \beta_{0,j}\rangle + \sigma_{0,j}\,\epsilon_j$, where $\sigma_{0,D} = 2$, $\sigma_{0,H} = 1$, $\alpha_D = 2$, $\alpha_H = 1.5$. Let $\phi_\ell(t) = \sqrt{2}\sin((2\ell-1)\pi t/2)$, $\ell \ge 1$, be the principal directions of a Wiener process. The slope parameter for the healthy population $\beta_{0,H}$ equals the function $\beta_0 = \phi_1 + \phi_2$, while for the diseased population, $\beta_{0,D} = 2\beta_0$.

For clean samples, denoted C_0 from now on, the errors $\epsilon_j \sim N(0,1)$, $j = D, H$. The covariates V_j were generated independently between them according to the model $V_j(t) = \sum_{\ell=1}^{5} \xi_{j,\ell}\phi_\ell(t)$, with $\xi_{j,\ell} \sim N(0, \lambda_\ell)$, where $\lambda_1 = 4$, $\lambda_2 = 1$, $\lambda_3 = 0.5$, $\lambda_4 = 0.1$ and $\lambda_5 = 0.05$. This model is similar to the one used in Inácio et al. (2012).

To evaluate the stability of the conditional ROC curve when atypical data arise, two contaminations have been considered

- Scenario $C_{1,H}$: here only the regression errors $\epsilon_{H,i}$ are contaminated in order to produce "vertical outliers". They are generated as $\epsilon_{H,i} \sim 0.9\,N(0,1) + 0.1\,N(\mu, 0.25)$. We chose $\mu = 8$.
- Scenario $C_{2,H}$: in this setting we introduced outliers in the healthy population with high-leverage by contaminating the functional covariates V_H and the errors simultaneously. Outliers in the $V_{H,i}$'s are generated by perturbing the distribution of the second score in the Karhunen-Loève representation of the process. Specifically, the contaminated data $(y^{(c)}_{H,i}, V^{(c)}_{H,i})$ are obtained as $y^{(c)}_{H,i} = \langle \beta_0, V^{(c)}_{H,i}\rangle + \epsilon^{(c)}_{H,i}$, where to generate the outliers, we sampled $d_{H,i} \sim Bi(1, 0.10)$ and then we took

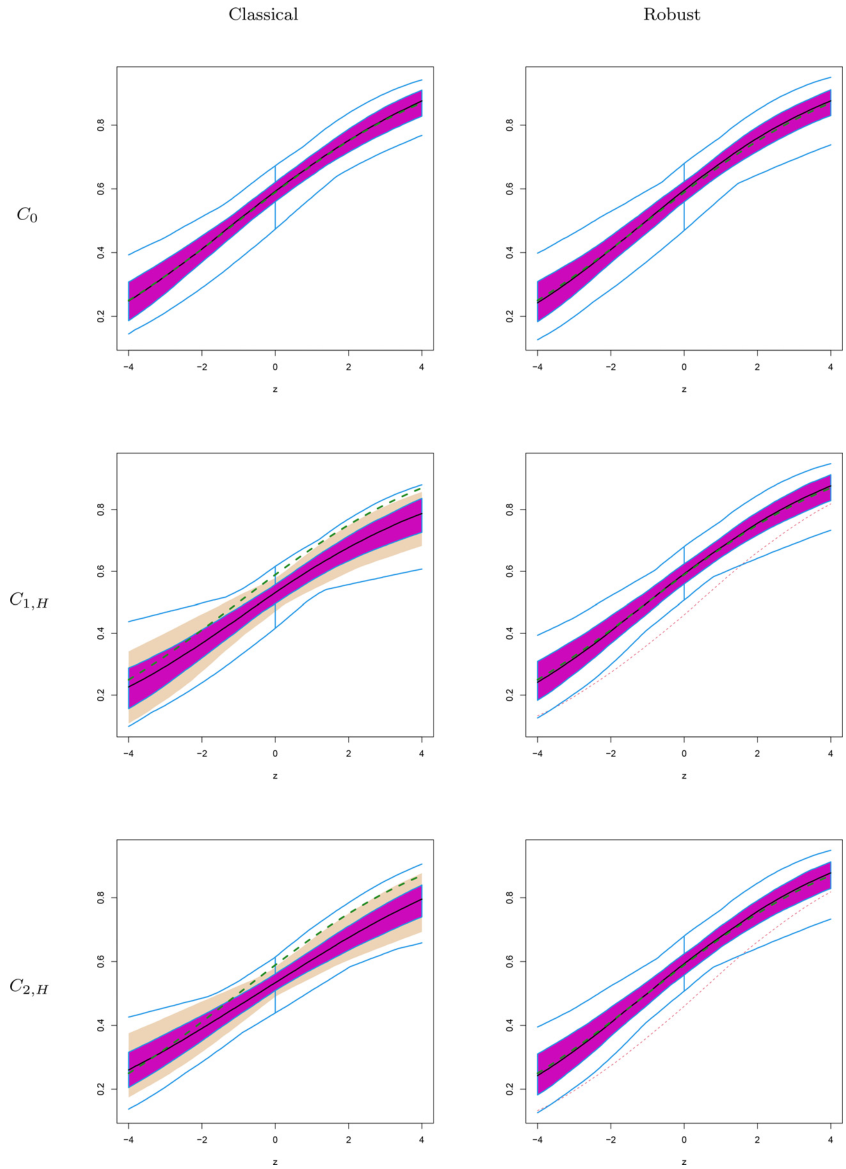

Fig. 1. Functional boxplots of the estimators of the area under the curve. The true function is shown with a green dashed line, while the black solid one is the central curve of the $Nrep = 1000$ estimates $\widehat{\mathrm{AUC}}_z$. Columns correspond to estimation methods, while rows to clean and contaminated samples.

- $\epsilon_{H,i}^{(c)} = \epsilon_{H,i}$ and $V_{H,i}^{(c)} = V_{H,i}$, if $d_{H,i} = 0$
- $\epsilon_{H,i}^{(c)} \sim N(\mu, 0.25)$ and $V_{H,i}^{(c)} = \sum_{s=1}^{5} \xi_{H,i,s}^{(c)} \phi_s(t)$, with $\xi_{H,i,s}^{(c)} \sim N(0, \lambda_s)$ for $s \neq 2$ and $\xi_{H,i,2}^{(c)} \sim N(\mu/2, 0.25)$, if $d_{H,i} = 1$. We select $\mu = 12$.

The values of μ chosen in $C_{1,H}$ and $C_{2,H}$ are illustrative cases of the performance of the classical and robust estimators, since similar results were obtained when μ varies from 8 to 24. Under these scenarios, we generated $Nrep = 1000$ datasets of size $n_D = n_H = 300$ and the functional covariates were evaluated at an equidistant grid of points of size 100.

As in Inácio et al. (2012), the conditional ROC curve and its estimates were evaluated at covariates of the form $V_z(t) = z\phi_1(t)$, in this way, they can be plotted as a function of z. For simplicity, we denote as $\mathrm{ROC}_z(p)$ the conditional ROC curve and analogous notation is used for its estimates and for the area under the curve and Youden index. The values of z were taken on an equidistant grid of points between $-2\sqrt{\lambda_1}$ and $2\sqrt{\lambda_1}$, with step of 0.05. We denote $N_z = 161$ the grid length. Besides, the possible values for p were chosen between 0.01 and 0.99 with a step of 0.01, leading to a grid of size $N_p = 99$. The performance of the ROC curve estimators was measured through the mean over replications of

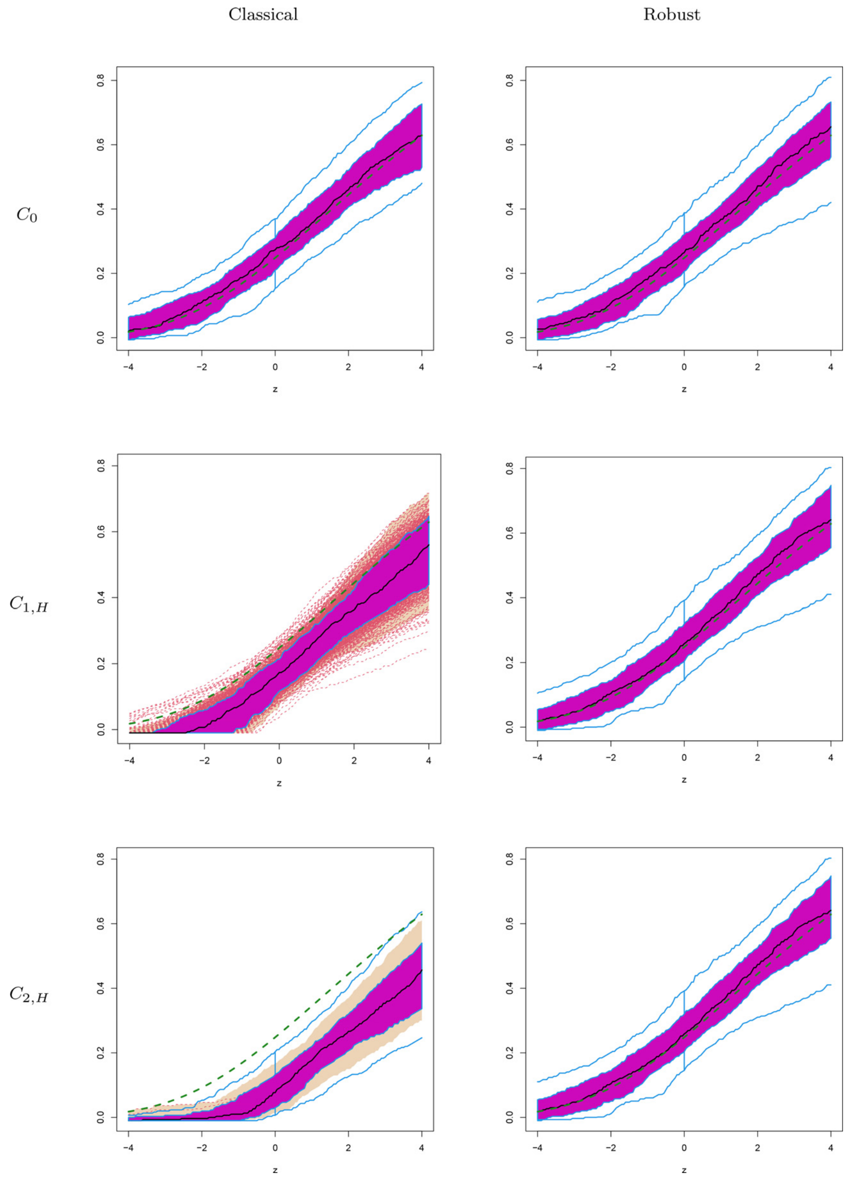

Fig. 2. Functional boxplots of the estimators of the Youden index. The true function is shown with a green dashed line, while the black solid one is the central curve of the $Nrep = 1000$ estimates $\widehat{\mathrm{YI}}_z$. Columns correspond to estimation methods, while rows to clean and contaminated samples.

- the Mean Squared Error (MSE) given by

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

- the Kolmogorov–Smirnov distance (KS) calculated as

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

that give global summaries of the mismatch between the estimated ROC curves and the true ones. Similar measures were considered for the area under the curve AUC_z and the Youden index YI_z.

Table 1 reports the obtained results and illustrates the stability of the robust proposal. In contrast, when considering the classical estimators of the ROC curve, the mean over replications of the KS measure is enlarged almost seven times, under $C_{1,H}$ and three times under $C_{2,H}$, while the MSE increases more than sixteen times. When considering the classical estimator

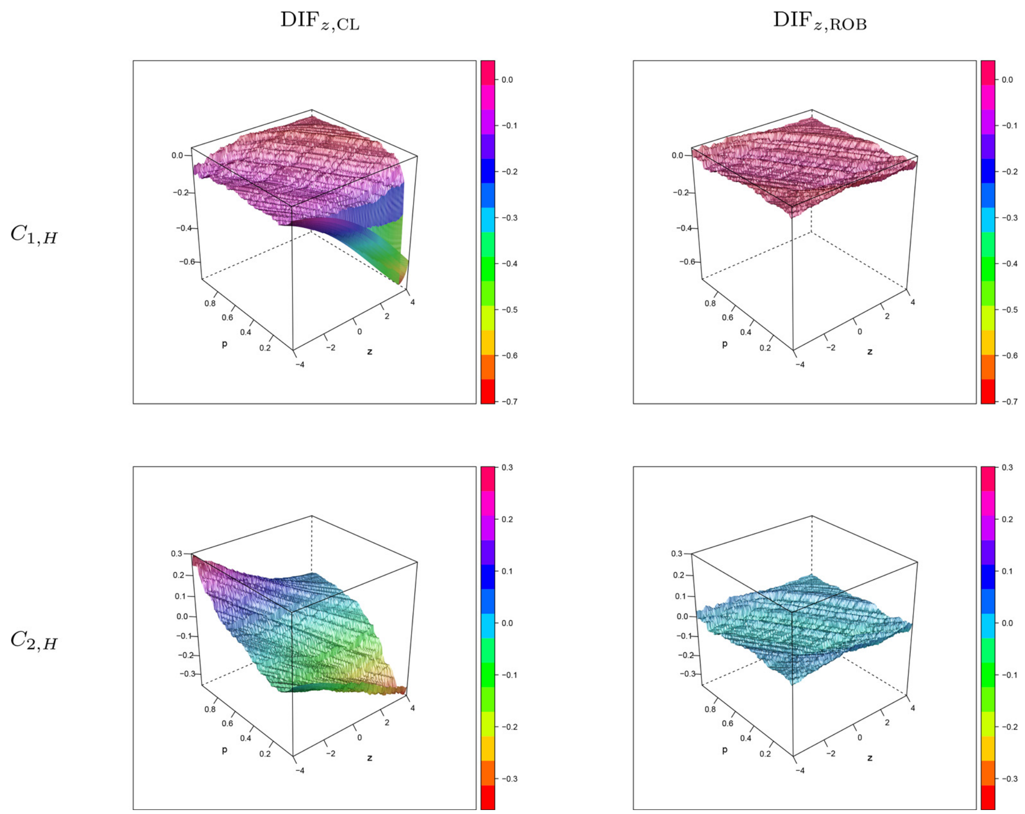

Fig. 3. Differences between the estimates of the conditional ROC curve and the true function for iteration 50 denoted $\mathrm{DIF}_{z,cl} = \widehat{\mathrm{ROC}}_{z,cl} - \mathrm{ROC}_z$ and $\mathrm{DIF}_{z,rob} = \widehat{\mathrm{ROC}}_{z,rob} - \mathrm{ROC}_z$, for the classical and robust estimators, respectively. The rows correspond to the two contaminations settings.

of the conditional area under the curve, the mean square error is increased more than five times and the mean uniform distance, that is, the mean over replications of $\sup_{1\le i\le N_z} \left|\widehat{\mathrm{AUC}}_{z_i} - \mathrm{AUC}_{z_i}\right|$ doubles its value with respect to the one obtained for clean samples. A similar behaviour is observed for the classical estimates of the Youden index.

To illustrate the effect of these contaminations, Figures 1 and 2 display the functionals boxplots of the classical and robust estimators of the area under the curve AUC_z and the Youden index YI_z, respectively. Functional boxplots, introduced by Sun and Genton (2011), are a useful tool 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_z or YI_z curves. For the classical estimators, under contamination, we also present in beige the regions containing the 90% of the curves. Once again, it becomes evident that the classical estimator suffers from the introduced contamination, in particular, when considering the Youden index, since for both contaminations the true curve goes beyond the limit of the central area and some times beyond the limits of the 90% central curves. In contrast the robust estimators remain very stable.

To have a deeper comprehension of the behaviour of the ROC curve estimators under contamination, Figure 3 presents the differences between the estimates of the ROC surface obtained in replication number 50 and the true surface, that is, $\mathrm{DIF}_{z,cl} = \widehat{\mathrm{ROC}}_{z,cl} - \mathrm{ROC}_z$ and $\mathrm{DIF}_{z,rob} = \widehat{\mathrm{ROC}}_{z,rob} - \mathrm{ROC}_z$, where $\widehat{\mathrm{ROC}}_{z,cl}$ and $\widehat{\mathrm{ROC}}_{z,rob}$ stand for the classical and robust estimators, respectively. Note that in order to facilitate the comparisons, the left and right panels are in the same scale. These plots show how the outliers highly influence the classical estimators, in particular, for values of p smaller than 0.2 under $C_{1,H}$ and within the whole range under $C_{2,H}$. In contrast the robust estimators do not exhibit this behaviour.

To visualize the effect that contaminations have on the ROC surface estimators across replications, Figure 4 presents the surface boxplots of $\widehat{\mathrm{ROC}}_{z,rob}$ and $\widehat{\mathrm{ROC}}_{z,cl}$, as defined in Genton et al. (2014), under C_0 and the two contamination settings. For these plots, the volume depth is used to order the surfaces. The median is represented in dark violet, the central region containing the 50% deepest surfaces is represented in blue, while the surfaces in pink indicate the whiskers, beyond whose limits a surface is declared as outlier. The true function ROC_z is plotted in green. The effect of outliers on the classical estimators is clearly reflected in these plots, where the true surface lies above the pink surface whiskers making evident that most classical surfaces underestimate the true conditional ROC. In contrast the robust procedure is very stable for the considered contaminations.

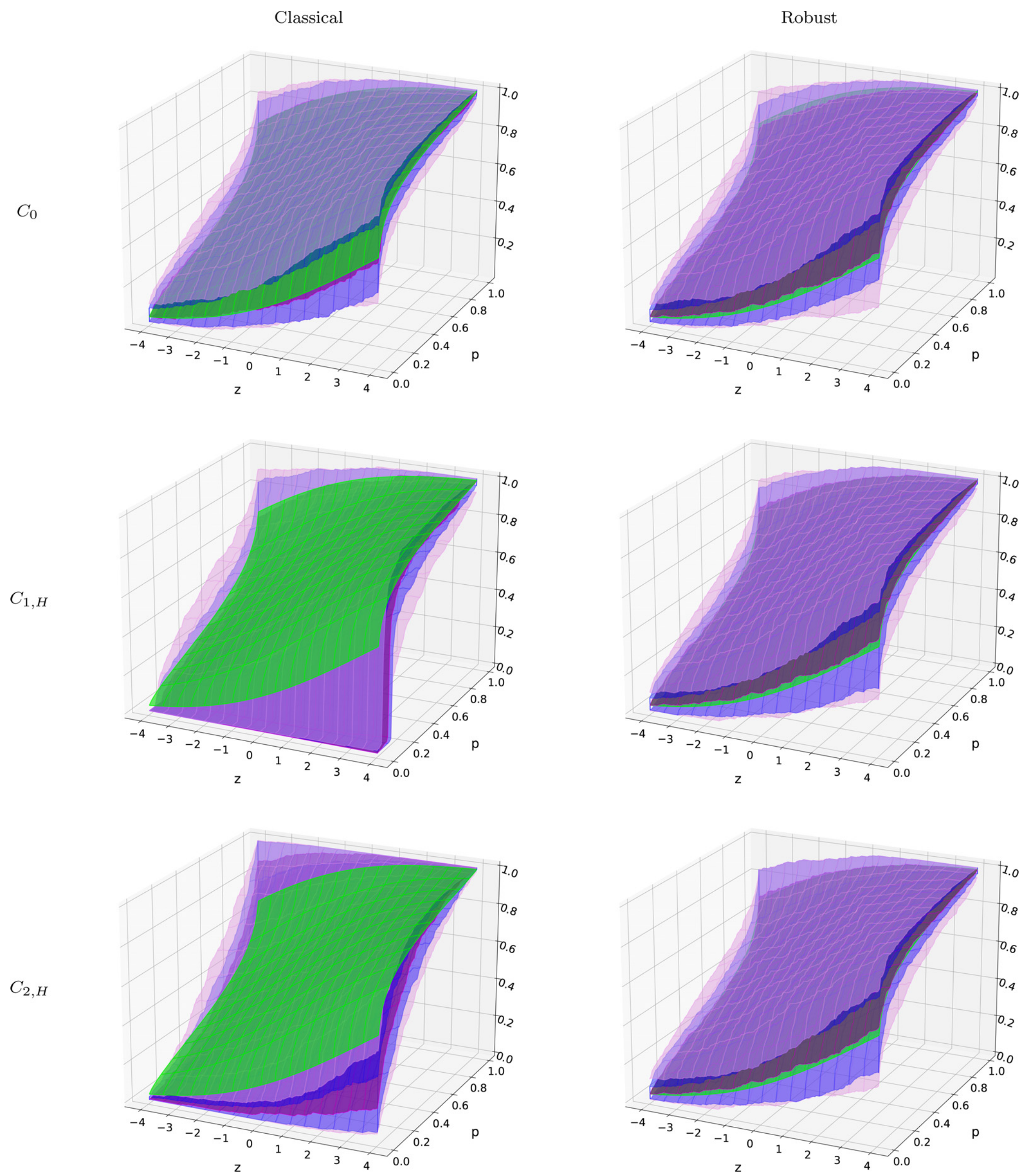

Fig. 4. Surface boxplots of the estimators for $\widehat{\mathrm{ROC}}_{z,cl}$ and $\widehat{\mathrm{ROC}}_{z,rob}$. Columns correspond to the estimation method, while rows to clean and contaminated samples.

5. Final comments

The ROC curve is an extended graphical tool useful to size up the accuracy of a diagnostic test based on a biomarker. In many situations, the presence of covariates related to the marker may increase its discriminating power. In such cases, it is suitable to use the conditional ROC curve. Direct and indirect methods are two different ways to incorporate the additional information contained in the covariates. As it is well known, conditional ROC curves may be easily estimated using a plug–in procedure. The usual techniques to estimate the conditional ROC curve based on classical regression methods and empirical distribution and quantile functions are not robust in the sense that they are unstable when atypical observations are present.

In this paper, besides providing a revision on the existing literature on this robust topic, we address the problem of robustly estimate the conditional ROC curve for some complex models, in particular, when some of the covariates are functional. More precisely, we generalize the proposal given in Bianco et al. (2022) to a very general scenario in which the markers are modelled in terms of a functional partially linear model. The considered situations include the functional linear regression model and also the nonparametric or additive regression ones with real valued covariates. In this way, the given approach enables us to cover a wide range of cases using a robust perspective. We obtain consistency results under standard regularity conditions. Our simulation study illustrates the stability of our proposal under different contamination scenarios.

Some possible future lines of research may include the robust estimation of the volume under the ROC surface in presence of covariates, see To et al. (2022), the derivation of appropriate robust confidence regions for the AUC and ROC, as well as the extension of the proposal given by Bianco et al. (2022) to deal with high–dimensional covariates. These areas have their own interest even when they are beyond the scope of the actual paper.

Acknowledgements

The authors wish to thank the Associate Editor and two anonymous referees for their valuable comments which led to an improved version of the original paper. This research was supported by Grants PICT 2018-00740 from ANPCYT and 20020170100022BA from the Universidad de Buenos Aires, Argentina and also by the Spanish Project PID2020-116587GB-I00 from the Ministry of Science and Innovation (MCIN/ AEI/FEDER, UE), Spain.

A. Proofs

The following Lemma is needed to derive Proposition 1. It corresponds to the functional counterpart of Lemma 1 in Bianco et al. (2022).

Lemma A.1. *Assume that either $w(t) = \mathbb{I}_{[-1,1]}(t)$ or w satisfies D1. For a fixed $s \in \mathbb{R}$ and $j = D, H$, define the class of functions $\mathcal{F}^{(j)} = \{f_{\mathbf{b},\mathbf{a},\kappa,\nu} : \mathbb{R} \times L^2(0,1) \times (0,1) \to \mathbb{R}$ with index $(\mathbf{b}, \mathbf{a}, \kappa, \nu) \in \mathbb{R}^{p_{1,j}} \times \mathbb{R}^{p_{2,j}} \times \mathbb{R}_{\geq 0} \times \mathbb{R}_{\geq 0}\}$, where $f_{\mathbf{b},\mathbf{a},\kappa,\nu}(y, v, z) = w\Big(\nu(y - \mathbf{b}^t\mathbf{v}(v) - \mathbf{a}^t\mathbf{B}(z))\Big)\,\mathbb{I}_{\{y-\mathbf{b}^t\mathbf{v}(v)-\mathbf{a}^t\mathbf{B}(z)\leq\kappa\, s\}}$, with $\mathbf{v}(v) = (\langle v, B^{(1)}_{j,1}\rangle, \ldots, \langle v, B^{(1)}_{j,p_{1,j}}\rangle)^t$ and $\mathbf{B}(z) = (B^{(2)}_{j,1}(z), \ldots, B^{(2)}_{j,p_{2,j}}(z))^t$.*

Then, under D3, we have that $\sup_{f\in\mathcal{F}^{(j)}} |P_{n_j} f - Pf| \xrightarrow{a.s.} 0$, where $P_{n_j} f = (1/n_j)\sum_{i=1}^{n_j} f(y_{j,i}, V_{j,i}, z_{j,i})$ and $Pf = \mathbb{E} f(Y_j, V_j, Z_j)$, for $j = D, H$.

Proof of Lemma A1. Fix $j = D$ or H and assume that w satisfies D1. Note that $\mathcal{F}^{(j)} \subset \mathcal{F}_1 \cdot \mathcal{F}_2$ where

$$\mathcal{F}_1 = \{f_{\mathbf{b},\mathbf{a},\nu}(y, v, z) = w\Big(\nu(y - \mathbf{b}^t\mathbf{v}(v) - \mathbf{a}^t\mathbf{B}(z))\Big) \text{ for } (\mathbf{b}, \mathbf{a}, \nu) \in \mathbb{R}^{p_{1,j}} \times \mathbb{R}^{p_{2,j}} \times \mathbb{R}_{\geq 0}\}$$
$$\mathcal{F}_2 = \{f_{\mathbf{b},\mathbf{a},\kappa}(y, v, z) = \mathbb{I}_{\{y-\mathbf{b}^t\mathbf{v}(v)-\mathbf{a}^t\mathbf{B}(z)\leq\kappa\, s\}} \text{ for } (\mathbf{b}, \mathbf{a}, \kappa) \in \mathbb{R}^{p_{1,j}} \times \mathbb{R}^{p_{2,j}} \times \mathbb{R}_{\geq 0}\}\,.$$

The classes $\mathcal{F}^{(j)}$, $\mathcal{F}_1$ and $\mathcal{F}_2$ have envelope 1, hence we have easily that, for any measure Q,

$$N(2\,\epsilon, \mathcal{F}^{(j)}, L_1(Q)) \leq N(\epsilon, \mathcal{F}_1, L_1(Q))N(\epsilon, \mathcal{F}_2, L_1(Q)),$$

so that to show $\sup_{f\in\mathcal{F}^{(j)}} |P_{n_j} f - Pf| \xrightarrow{a.s.} 0$, it will be enough to prove that, for $s = 1, 2$,

$$\frac{1}{n_j}\log N(\epsilon, \mathcal{F}_s, L_1(P_{n_j})) \xrightarrow{p} 0\,. \tag{P.1}$$

As in the proof of Lemma S.2 in Bianco et al. (2022), it is straightforward to see that $\mathcal{F}_2$ is a VC-class with index $V_2 = q_{n_j} + 3$, where $q_{n_j} = p_{1,j} + p_{2,j}$, so

$$N(\epsilon, \mathcal{F}_2, L_1(Q)) \leq K\, V_2(16e)^{V_2}\left(\frac{1}{\epsilon}\right)^{V_2-1}. \tag{P.2}$$

Using that $\log(q_{n_j} + 3)/(q_{n_j} + 3) < 1$ and assuming without loss of generality that $K > 1$, from (A.2), we easily get that $\log\left(N(\epsilon, \mathcal{F}_2, L_1(\mathbb{Q}))\right) \leq C(p_{1,j} + p_{2,j})\log(1/\epsilon)$, for $\epsilon < \min((16e)^{-1}, e^{-K})$ and some constant C. Thus, we obtain (A.1), when $s = 2$, from $(1/n_j)\log N(\epsilon, \mathcal{F}_{n_j}, L_1(P_{n_j})) \leq C\{(p_{1,j} + p_{2,j})/n_j\}\log(1/\epsilon) \to 0$, since $p_{1,j} + p_{2,j} = o(n_j)$ from D3.

On the other hand, the family $\mathcal{R} = \Big\{\nu(y - \mathbf{b}^t\mathbf{v}(v) - \mathbf{a}^t\mathbf{B}(z)) : (\mathbf{b}, \mathbf{a}, \nu) \in \mathbb{R}^{p_{1,j}} \times \mathbb{R}^{p_{2,j}} \times \mathbb{R}_{\geq 0}\Big\}$ is a subset of the vector space of all linear functions in $q_{n_j} + 1$ variables. It follows from Lemma 2.6.15 of van der Vaart and Wellner (1996) that $\mathcal{R}$ has VC-index at most $q_{n_j} + 3$. Note that w is an even function, non-increasing on $[0, +\infty)$, hence it can be written as $w = w^{(1)} + w^{(2)}$, where $w^{(1)}(t) = w(t)\mathbb{I}_{[0,+\infty)}(t)$ is non–increasing and $w^{(2)}(t) = w(t)\mathbb{I}_{(-\infty,0)}(t)$ is non–decreasing. Using the permanence property for VC-classes, see Lemma 9.9(viii) in Kosorok (2008), we obtain that the classes of functions $\mathcal{R}_{w^{(1)}} = w^{(1)} \circ \mathcal{R}$ and $\mathcal{R}_{w^{(2)}} = w^{(2)} \circ \mathcal{R}$ are VC–classes with VC–index at most $q_n + 3$. Furthermore, the classes $\mathcal{R}_{w^{(\ell)}}$, $\ell = 1, 2$, have envelope 1. Then, Theorem 2.6.7 of van der Vaart and Wellner (1996) implies that there exists a universal constant K such that, for any probability measure $\mathbb{Q}$ on $\mathbb{R}^{q_{n_j}+1}$ and any $0 < \epsilon < 1$, we have that

$$N(\epsilon, \mathcal{R}_{w^{(\ell)}}, L_1(\mathbb{Q})) \leq K(q_{n_j} + 3)\ (16e)^{(q_{n_j}+3)}\left(\frac{1}{\epsilon}\right)^{q_{n_j}+2}.$$

Note that $\mathcal{R}_{w^{(1)}} + \mathcal{R}_{w^{(2)}}$ has also constant envelope equal to 1. Therefore, using that $\mathcal{F}_1$ has constant envelope equal to 1, $\mathcal{F}_1 \subset \mathcal{R}_{w^{(1)}} + \mathcal{R}_{w^{(2)}}$ and $N(2\epsilon, \mathcal{R}_{w^{(1)}} + \mathcal{R}_{w^{(2)}}, L_1(Q)) \leq N(\epsilon, \mathcal{R}_{w^{(1)}}, L_1(Q)) \times N(\epsilon, \mathcal{R}_{w^{(2)}}, L_1(Q))$, we get that

$$N(2\epsilon, \mathcal{F}_1, L_1(P_{n_j})) \leq \left[K(q_{n_j} + 3)\ (16e)^{(q_{n_j}+3)}\left(\frac{1}{\epsilon}\right)^{q_{n_j}+2}\right]^2,$$

which arguing as above, allows to conclude that (A.1) holds, when $s=1$, since $q_{n_j}=o(n_j)$.

When $w(t)=\mathbb{I}_{[-1,1]}(t)$ the result is straightforward using that

$$\mathcal{F}_1=\{f_{\mathbf{b},\mathbf{a},\,\nu}(y,\mathbf{x})=\mathbb{I}_{\{\nu(y-\mathbf{b}^t\mathbf{v}(\nu)-\mathbf{a}^t\mathbf{B}(z))\le 1\}}\mathbb{I}_{\{-\nu(y-\mathbf{b}^t\mathbf{v}(\nu)-\mathbf{a}^t\mathbf{B}(z))\le 1\}}\text{ for }(\mathbf{b},\mathbf{a},\nu)\in\mathbb{R}^{p_{1,j}}\times\mathbb{R}^{p_{2,j}}\times\mathbb{R}_{\ge 0}\}$$

$p_{1,j}=o(n_j)$ and $p_{2,j}=o(n_j)$ and similar arguments to those considered above. □

Proof of Proposition 1. 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 of Bianco et al. (2022), it will be enough to show that for each $s\in\mathbb{R}$, $\widehat{G}_j(s)\xrightarrow{a.s.}G_j(s)$.

To avoid burden notation, we define u_i as $u_i=\sigma_{0,j}\epsilon_{j,i}$ and $U=\sigma_{0,j}\epsilon_j$ and we will omit the subscript j. Hence, for instance, V, Z, $\sigma_{0,j}$, $\widehat{\beta}$, $\widehat{\eta}$, $\widehat{\sigma}$, γ_n and n will stand for V_j, Z_j, σ_0, $\widehat{\beta}_j$, $\widehat{\eta}_j$, $\widehat{\sigma}_j$, γ_{n_j} and n_j, respectively.

From now on, let $Pf=\mathbb{E}f(Y,V,Z)$ with (Y,V,Z) a random vector with the same distribution as $(y_{j,i},V_{j,i},z_{j,i})$, $1\le i\le n_j$. Furthermore, denote as $L(\mathbf{b},\mathbf{a},\,\kappa,\,\nu)=\mathbb{E}f_{\mathbf{b},\mathbf{a},\,\kappa,\nu}(Y,V,Z)$, where $f_{\mathbf{b},\mathbf{a},\,\kappa,\nu}$ is defined in Lemma A.1, that is, $f_{\mathbf{b},\mathbf{a},\,\kappa,\nu}(y,v,z)=w(\nu(y-\langle V,\beta_{\mathbf{b}}\rangle-\eta_{\mathbf{a}}(z)))\,\mathbb{I}_{\{y-\langle V,\beta_{\mathbf{b}}\rangle-\eta_{\mathbf{a}}(z)\le\kappa\, s\}}$ with $\beta_{\mathbf{b}}(t)=\sum_{s=1}^{p_1}b_sB_s^{(1)}(t)$ and $\eta_{\mathbf{a}}(z)=\sum_{s=1}^{p_2}a_sB_s^{(2)}(z)$. Define also $L^\star(\mathbf{b},\mathbf{a},\,\kappa)=\mathbb{E}g_{\mathbf{b},\mathbf{a},\,\kappa}(Y,V,Z)$, where $g_{\mathbf{b},\mathbf{a},\,\kappa}(y,v,z)=\mathbb{I}_{\{y-\langle V,\beta_{\mathbf{b}}\rangle-\eta_{\mathbf{a}}(z)\le\kappa\, s\}}$. Then, we have that $L(\widehat{\mathbf{b}},\widehat{\mathbf{a}},\,\widehat{\sigma},\,\widehat{\nu})=Pf_{\widehat{\mathbf{b}},\widehat{\mathbf{a}},\,\widehat{\sigma},\,\widehat{\nu}}$ and $L^\star(\widehat{\mathbf{b}},\widehat{\mathbf{a}},\,\widehat{\sigma})=\mathbb{E}\,\mathbb{I}_{\{Y-\langle V,\widehat{\beta}\rangle-\widehat{\eta}(Z)\le\widehat{\sigma}\, s\}}$, where the expectations are taking with respect to (Y,V,Z), conditional to the sample.

Denote as $\widehat{\nu}=1/(\gamma_n\widehat{\sigma})$, $\nu_0=0$. Then $\widehat{\nu}\xrightarrow{a.s.}\nu_0$, since $\gamma_n\xrightarrow{a.s.}\infty$ and $\widehat{\sigma}\xrightarrow{a.s.}\sigma_0$. We will begin by showing that for each fixed real number s, we have that

$$\frac{1}{n}\sum_{i=1}^{n}w_i\,\mathbb{I}_{\{\widehat{\epsilon}_i\le s\}}\xrightarrow{a.s.}G_j(s),\tag{P.3}$$

where $w_i=w(\widehat{\epsilon}_i/\gamma_n)$ and $\widehat{\epsilon}_i$ is defined in (). For that purpose and noting that $\widehat{\epsilon}_i=(y_i-\widehat{\mathbf{b}}^t\mathbf{v}_i-\widehat{\mathbf{a}}^t\mathbf{B}_i)/\widehat{\sigma}$, we consider the class of functions $\mathcal{F}$ given in Lemma A.1 which entails that $\sup_{f\in\mathcal{F}}|P_nf-Pf|\xrightarrow{a.s.}0$. Then, using that $P_nf_{\widehat{\mathbf{b}},\widehat{\mathbf{a}},\,\widehat{\sigma},\,\widehat{\nu}}=(1/n)\sum_{i=1}^n w_i\mathbb{I}_{\{\widehat{\epsilon}_i\le s\}}$, we obtain that $\sum_{i=1}^n w_i\mathbb{I}_{\{\widehat{\epsilon}_i\le s\}}/n-L(\widehat{\mathbf{b}},\widehat{\mathbf{a}},\,\widehat{\sigma},\,\widehat{\nu})\xrightarrow{a.s.}0$.

It remains to show that $L(\widehat{\mathbf{b}},\widehat{\mathbf{a}},\,\widehat{\sigma},\,\widehat{\nu})\xrightarrow{a.s.}G_j(s)$. Note that $L(\widehat{\mathbf{b}},\widehat{\mathbf{a}},\,\widehat{\sigma},\,\widehat{\nu})-G_j(s)=A_n(\widehat{\mathbf{b}},\widehat{\mathbf{a}},\,\widehat{\sigma},\,\widehat{\nu})+B_n(\widehat{\mathbf{b}},\widehat{\mathbf{a}},\,\widehat{\sigma})$, where $A_n(\mathbf{b},\mathbf{a},\,\sigma,\,\nu)=L(\mathbf{b},\mathbf{a},\,\sigma,\,\nu)-L^\star(\mathbf{b},\mathbf{a},\,\sigma)$ and $B_n(\mathbf{b},\mathbf{a},\,\sigma)=L^\star(\mathbf{b},\mathbf{a},\,\sigma)-G_j(s)$, where we recall that $L^\star(\mathbf{b},\mathbf{a},\,\sigma)=\mathbb{E}\,\mathbb{I}_{\{Y-\langle V,\beta_{\mathbf{b}}\rangle-\eta_{\mathbf{a}}(Z)\le\sigma\, s\}}$. Hence, it will be enough to show that $A_n(\widehat{\mathbf{b}},\widehat{\mathbf{a}},\,\widehat{\sigma},\,\widehat{\nu})\xrightarrow{a.s.}0$ and $B_n(\widehat{\mathbf{b}},\widehat{\mathbf{a}},\,\widehat{\sigma})\xrightarrow{a.s.}0$.

We begin by considering the situation where w satisfies D1. Noting that

$$|A_n(\mathbf{b},\mathbf{a},\,\kappa,\,\nu)|\le\mathbb{E}|w(\nu(U-\langle V,\beta_{\mathbf{b}}-\beta_0\rangle-[\eta_{\mathbf{a}}(Z)-\eta_0(Z)]))-1|,$$

and using the Dominated Convergence Theorem, the continuity of w, the fact that $\widehat{\nu}\xrightarrow{a.s.}\nu_0=0$ and A4, we obtain that $A_n(\widehat{\mathbf{b}},\widehat{\mathbf{a}},\,\widehat{\sigma},\,\widehat{\nu})\xrightarrow{a.s.}0$.

When $w=\mathbb{I}_{[-1,1]}$, denote $\widehat{r}=Y-\langle V,\widehat{\beta}\rangle-\widehat{\eta}(Z)=U-\langle V,\widehat{\beta}-\beta_0\rangle-[\widehat{\eta}(Z)-\eta_0(Z)]$ and $r_{\mathbf{b},\mathbf{a}}=Y-\langle V,\beta_{\mathbf{b}}\rangle-\eta_{\mathbf{a}}(Z)=U-\langle V,\beta_{\mathbf{b}}-\beta_0\rangle-[\eta_{\mathbf{a}}(Z)-\eta_0(Z)]$. Note that $\widehat{r}=r_{\widehat{\mathbf{b}},\widehat{\mathbf{a}}}$.

Using that $w\big(\nu\, r_{\mathbf{b},\mathbf{a}}\big)-1=\mathbb{I}_{\{\nu\, r_{\mathbf{b},\mathbf{a}}\le 1\}}\left\{\mathbb{I}_{\{-1\le\nu\, r_{\mathbf{b},\mathbf{a}}\}}-1\right\}-\left\{1-\mathbb{I}_{\{\nu\, r_{\mathbf{b},\mathbf{a}}\le 1\}}\right\}$, we get that, for any $\nu>0$,

$$\begin{aligned}|A_n(\mathbf{b},\mathbf{a},\,\kappa,\,\nu)|&\le\mathbb{E}\Big|\mathbb{I}_{\{-1\le\nu\, r_{\mathbf{b},\mathbf{a}}\}}-1\Big|+\mathbb{E}\Big|1-\mathbb{I}_{\{\nu\, r_{\mathbf{b},\mathbf{a}}\le 1\}}\Big|=\mathbb{E}\Big|\mathbb{I}_{\{r_{\mathbf{b},\mathbf{a}}<-\nu^{-1}\}}\Big|+\mathbb{E}\Big|1-\mathbb{I}_{\{r_{\mathbf{b},\mathbf{a}}\le\nu^{-1}\}}\Big|\\&\le\mathbb{E}\Big|\mathbb{I}_{\{U<-\nu^{-1}+\langle V,\beta_{\mathbf{b}}-\beta_0\rangle+[\eta_{\mathbf{a}}(Z)-\eta_0(Z)]\}}\Big|+\mathbb{E}\Big|1-\mathbb{I}_{\{U\le\nu^{-1}+\langle V,\beta_{\mathbf{b}}-\beta_0\rangle+[\eta\mathbf{a}(Z)-\eta_0(Z)]\}}\Big|\\&\le\mathbb{E}\left|G_j\left(\frac{1}{\sigma_0}\left[-\frac{1}{\nu}+\langle V,\beta_{\mathbf{b}}-\beta_0\rangle+[\eta\mathbf{a}(Z)-\eta_0(Z)]\right]\right)\right|\\&\quad+\mathbb{E}\left|G_j\left(\frac{1}{\sigma_0}\left[\frac{1}{\nu}+\langle V,\beta_{\mathbf{b}}-\beta_0\rangle+[\eta\mathbf{a}(Z)-\eta_0(Z)]\right]\right)-1\right|,\end{aligned}$$

Now the proof follows from the fact that $\lim_{u\to-\infty}G_j(u)=0$ while $\lim_{u\to+\infty}G_j(u)=1$ using A4 and $1/\widehat{\nu}=\gamma_n\widehat{\sigma}\xrightarrow{a.s.}+\infty$.

To derive that $B_n(\widehat{\mathbf{b}},\widehat{\mathbf{a}},\,\widehat{\sigma})\xrightarrow{a.s.}0$, note that for any $(\mathbf{b},\mathbf{a},\,\sigma)$, we have

$$B_n(\mathbf{b},\mathbf{a},\,\sigma)=\mathbb{E}\,\mathbb{I}_{\{U-\langle V,\beta_{\mathbf{b}}-\beta_0\rangle-[\eta_{\mathbf{a}}(Z)-\eta_0(Z)]\le\sigma\, s\}}-\mathbb{E}\,\mathbb{I}_{\{U\le\sigma_0\, s\}}\,.$$

We will distinguish two possibilities. In the first one, let us assume that $\langle V,\beta_{\mathbf{b}}-\beta_0\rangle+[\eta_{\mathbf{a}}(Z)-\eta_0(Z)]+\sigma\, s\le\sigma_0\, s$, then $\mathbb{I}_{\{U\le\langle V,\beta_{\mathbf{b}}-\beta_0\rangle+[\eta_{\mathbf{a}}(Z)-\eta_0(Z)]+\sigma\, s\}}=1$ implies that $\mathbb{I}_{\{U\le\sigma_0\, s\}}=1$, so

$$\Delta_{\mathbf{b},\mathbf{a},\sigma}(U)=\Big|\mathbb{I}_{\{U\le\langle V,\beta_{\mathbf{b}}-\beta_0\rangle+[\eta_{\mathbf{a}}(Z)-\eta_0(Z)]+\sigma\, s\}}-\mathbb{I}_{\{U\le\sigma_0\, s\}}\Big|=0\,.$$

Similarly, when $\mathbb{I}_{\{U\le\sigma_0\, s\}}=0$, we have that $\mathbb{I}_{\{U\le\langle V,\beta_{\mathbf{b}}-\beta_0\rangle+[\eta_{\mathbf{a}}(Z)-\eta_0(Z)]+\sigma\, s\}}=0$ and $\Delta_{\mathbf{b},\mathbf{a},\sigma}(U)=0$. Therefore, when $\langle V,\beta_{\mathbf{b}}-\beta_0\rangle+[\eta_{\mathbf{a}}(Z)-\eta_0(Z)]+\sigma\, s\le\sigma_0\, s$, $\Delta_{\mathbf{b},\mathbf{a},\sigma}(U)=1$ if and only if $\langle V,\beta_{\mathbf{b}}-\beta_0\rangle+[\eta_{\mathbf{a}}(Z)-\eta_0(Z)]+\sigma\, s<U\le\sigma_0\, s$. Secondly, when $\langle V,\beta_{\mathbf{b}}-\beta_0\rangle+[\eta_{\mathbf{a}}(Z)-\eta_0(Z)]+\sigma\, s\ge\sigma_0\, s$, we have that $\Delta_{\mathbf{b},\mathbf{a},\sigma}(U)=1$ if and only if $\sigma_0\, s<U\le\langle V,\beta_{\mathbf{b}}-\beta_0\rangle+[\eta_{\mathbf{a}}(Z)-\eta_0(Z)]+\sigma\, s$.

Let $\mathcal{C}_{\mathbf{b},\mathbf{a},\sigma} = \{(v,z) \in L^2(0,1)\times(0,1) : \langle v, \beta_{\mathbf{b}} - \beta_0\rangle + [\eta_{\mathbf{a}}(Z) - \eta_0(Z)] + \sigma\, s \le \sigma_0\, s\}$ and $\overline{\mathcal{C}}_{\mathbf{b},\mathbf{a},\sigma}$ its complement, then using that $U/\sigma_0 \sim G_j$, we get the bound

$$\begin{aligned}|B_n(\mathbf{b},\mathbf{a},\ \sigma)| &\le \mathbb{E}\,\mathbb{I}_{\mathcal{C}_{\mathbf{b},\mathbf{a},\sigma}}\ \mathbb{I}_{\{\langle V,\beta_{\mathbf{b}}-\beta_0\rangle+[\eta_{\mathbf{a}}(Z)-\eta_0(Z)]+\sigma\, s<U\le\sigma_0\, s\}} + \mathbb{E}\,\mathbb{I}_{\overline{\mathcal{C}}_{\mathbf{b},\mathbf{a},\sigma}}\ \mathbb{I}_{\{\sigma_0\, s<U\le\langle V,\beta_{\mathbf{b}}-\beta_0\rangle+[\eta_{\mathbf{a}}(Z)-\eta_0(Z)]+\sigma\, s\}}\\ &\le \mathbb{E}\,\mathbb{I}_{\mathcal{C}_{\mathbf{b},\mathbf{a},\sigma}}\left\{G_j(s) - G_j\left(\frac{\langle V,\beta_{\mathbf{b}}-\beta_0\rangle+[\eta_{\mathbf{a}}(Z)-\eta_0(Z)]+\sigma\, s}{\sigma_0}\right)\right\}\\ &\quad+\mathbb{E}\,\mathbb{I}_{\overline{\mathcal{C}}_{\mathbf{b},\mathbf{a},\sigma}}\left\{G_j\left(\frac{\langle V,\beta_{\mathbf{b}}-\beta_0\rangle+[\eta_{\mathbf{a}}(Z)-\eta_0(Z)]+\sigma\, s}{\sigma_0}\right) - G_j(s)\right\}\\ &\le \mathbb{E}\left|G_j\left(\frac{\langle V,\beta_{\mathbf{b}}-\beta_0\rangle+[\eta_{\mathbf{a}}(Z)-\eta_0(Z)]+\sigma\, s}{\sigma_0}\right) - G_j(s)\right|,\end{aligned}$$

so $|B_n(\widehat{\mathbf{b}},\widehat{\mathbf{a}},\ \widehat{\sigma})| \le M(\widehat{\mathbf{b}},\widehat{\mathbf{a}},\ \widehat{\sigma})$, where $M(\mathbf{b},\mathbf{a},\ \sigma) = \mathbb{E}\Delta^{\star}_{\mathbf{b},\mathbf{a},\ \sigma}(v,z)$ and

$$\Delta^{\star}_{\mathbf{b},\mathbf{a},\ \sigma}(v,z) = \left|G_j\left(\frac{\langle v,\beta_{\mathbf{b}}-\beta_0\rangle+[\eta_{\mathbf{a}}(z)-\eta_0(z)]+\sigma\, s}{\sigma_0}\right) - G_j(s)\right|.$$

Note that A4 and the fact that $\widehat{\sigma} \xrightarrow{a.s.} \sigma_0$ entail that $\langle v, \widehat{\beta} - \beta_0\rangle + [\widehat{\eta}(z) - \eta_0(z)] + \widehat{\sigma}\, s \xrightarrow{a.s.} \sigma_0\, s$, for each $(v,z) \in L^2(0,1)\times(0,1)$. Using the continuity of G_j and the fact that $\widehat{\beta} = \beta_{\widehat{\mathbf{b}}}$ and $\widehat{\eta} = \eta_{\widehat{\mathbf{a}}}$, we get that $\Delta^{\star}_{\widehat{\mathbf{b}},\widehat{\mathbf{a}},\ \widehat{\sigma}}(v,z) \xrightarrow{a.s.} 0$, for each $(v,z) \in L^2(0,1)\times(0,1)$ which together with the Dominated Convergence Theorem implies that $B_n(\widehat{\mathbf{b}},\widehat{\mathbf{a}},\ \widehat{\sigma}) \xrightarrow{a.s.} 0$, concluding the proof of (A.3).

Similar arguments allow to show that $\sum_{i=1}^n w_i/n \xrightarrow{a.s.} 1$ which concludes the proof. □

References

Alonzo, T.A., Pepe, M.S., 2002. Distribution-free ROC analysis using binary regression techniques. Biostatistics 3, 421–432.
Aneiros-Pérez, G., Vieu, P., 2006. Semi-functional partial linear regression. Statistics and Probability Letters 76, 1102–1110.
Arribas-Gil, A., Romo, J., 2014. Shape outlier detection and visualization for functional data: the outliergram. Biostatistics 15, 603–619.
Bali, J.L., Boente, G., 2011. Robust functional principal component analysis. In: Pacheco, A., Santos, R., Oliveira, R., Paulino, C. (Eds.), New advances in statistical modeling and applications, pp. 41–53.
Bianco, A., Boente, G., 1998. Robust kernel estimators for additive models with dependent observations. The Canadian Journal of Statistics 6, 239–255.
Bianco, A., Boente, G., 2004. Robust estimators in semiparametric partly linear regression models. Journal of Statistical Planning and Inference 122, 229–252.
Bianco, A.M., Boente, G., González-Manteiga, W., 2022. Robust consistent estimators for ROC curves with covariates. Electronic Journal of Statistics 16, 4133–4161.
Boente, G., Fraiman, R., 1989. Robust nonparametric regression estimation for dependent observations. Annals of Statistics 17, 1242–1256.
Boente, G., Martínez, A., 2017. Marginal integration $m-$estimators for additive models. TEST 26, 231–260.
Boente, G., Martínez, A., 2022. A robust spline approach in partially linear additive models. In: Computational statistics and data analysis.
Boente, G., Martínez, A., Salibián-Barrera, M., 2017. Robust estimators for additive models using backfitting. Journal of Nonparametric Statistics 29, 744–767.
Boente, G., Salibián-Barrera, M., Vena, P., 2020. Robust estimation for semi-functional linear regression models. Computational Statistics and Data Analysis 152, 107041.
Cai, T., 2004. Semiparametric ROC regression analysis with placement values. Biostatistics 5, 45–60.
Cai, T., Pepe, M.S., 2002. Semiparametric receiver operating characteristic analysis to evaluate biomarkers for disease. Journal of the American Statistical Association 97, 1099–1107.
Cantoni, E., Ronchetti, E., 2001. Resistant selection of the smoothing parameter for smoothing splines. Statistics and Computing 11, 141–146.
Cardot, H., Crambes, C., Sarda, P., 2005. Quantile regression when the covariates are functions. Journal of Nonparametric Statistics 17, 841–856.
Charaf, J. (2022). Sensibilidad y estimación robusta en modelos de regresión directa para curvas ROC. Unpublished manuscript. Available at http://cms.dm.uba.ar/academico/carreras/licenciatura/tesis/2022/.
Cleveland, W.S., 1979. Robust locally-weighted regression and smoothing scatterplots. Journal of the American Statistical Association 74, 829–836.
Cox, D.D., 1983. Asymptotics for m-type smoothing splines. Annals of Statistics 11, 530–551.
Crambes, C., Kneip, A., Sarda, P., 2009. Smoothing splines estimators for functional linear regression. Annals of Statistics 37, 35–72.
Dai, W., Genton, M., 2019. Directional outlyingness for multivariate functional data. Computational Statistics and Data Analysis 131, 50–65.
Devlin, S.M., Thomas, E.G., Emerson, S.S., 2013. Robustness of approaches to ROC curve modelling under misspecification of the underlying probability model. Communications in Statistics - Theory and Methods 42, 3655–3664.
Dodd, L., Pepe, M., 2003. Semiparametric regression for the area under the receiver operating characteristic curve. Journal of the American Statistical Association 98, 409–417.
Dutter, R., Filzmoser, P., Gather, U., Rousseeuw, P., 2003. Developments in Robust Statistics. Physica-Verlag, Heidelberg, Germany.
Faraggi, D., 2003. Adjusting receiver operating characteristic curves and related indices for covariates. Journal of the Royal Statistical Society, Series D 52, 1152–1174.
Farcomeni, A., Ventura, L., 2012. An overview of robust methods in medical research. Statistical Methods in Medical Research 21, 111–133.
Ferraty, F., Vieu, P., 2006. Nonparametric Functional Data Analysis. Springer, New York.
Genton, M.G., Johnson, C., Potter, K., Stenchikov, G., Sun, Y., 2014. Surface boxplots. Stat 3, 1–11.
Gonçalves, L., Subtil, A., Oliveira, M.R., Bermudez, P., 2014. ROC curve estimation: An overview. REVSTAT-Statistical Journal 12, 1–20.
González-Manteiga, W., Pardo-Fernández, J.C., Van Keilegom, I., 2011. ROC curves in non-parametric location-scale regression models. Scandinavian Journal of Statistics 38, 169–184.
Greco, L., Ventura, L., 2011. Robust inference for the stress-strength reliability. Statistical Papers 52, 773–788.
Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J., Stahel, W.A., 2005. Robust statistics: the approach based on influence functions. Wiley, New York.
Härdle, W., Tsybakov, B., 1988. Robust nonparametric regression with simultaneous scale curve estimation. Annals of Statistics 16, 120–135.
He, X., Shi, P., 1996. Bivariate tensor–product b–spline in a partly linear model. Journal of Multivariate Analysis 58, 162–181.
He, X., Zhue, Z.Y., Fung, W.K., 2002. Estimation in a semiparametric model for longitudinal data with unspecified dependence structure. Biometrika 89, 579–590.
Heritier, S., Cantoni, E., Copt, S., Victoria-Feser, M.P., 2009. Robust Methods in Biostatistics. Wiley, New York.

Huang, L., Wang, H., Cui, H., Wang, S., 2015. Sieve m-estimator for a semi-functional linear model. Science China, Mathematics 58, 2421–2434.
Huber, P., 1964. Robust estimation of a location parameter. Annals of Mathematical Statistics 35, 73–101.
Huber, P., Ronchetti, E., 2009. Robust Statistics, 2nd Wiley, New York.
Hubert, M., Rousseeuw, P., Segaert, P., 2015. Multivariate functional outlier detection. Statistical Methods and Applications 24, 177–202.
Inácio, V., González-Manteiga, W., Febrero-Bande, M., Gude, F., Alonzo, T., Cadarso-Suárez, C., 2012. Extending induced ROC methodology to the functional context. Biostatistics 13, 594–608.
Inácio, V., Lourenço, V., de Carvalho, M., Parker, R.A., Gnanapragasam, V., 2021a. Robust and flexible inference for the covariate-specific receiver operating characteristic curve. Statistics in Medicine 40, 5779–5795.
Inácio, V., Rodríguez-Álvarez, M.X., Gayoso-Diz, P., 2021b. Statistical evaluation of medical tests. Annual Review of Statistics and Its Application 8, 41–67.
Inácio de Carvalho, V., de Carvalho, M., Alonzo, T., González-Manteiga, W., 2016. Functional covariate-adjusted partial area under the specificity-ROC curve with an application to metabolic syndrome diagnosis. Annals of Applied Statistics 10, 1472–1495.
Inácio de Carvalho, V., Jara, A., Hanson, T.E., de Carvalho, M., 2013. Bayesian nonparametric ROC regression modeling. Bayesian Analysis 3, 623–646.
Kalogridis, I., 2021. Asymptotics for m-type smoothing splines with non-smooth objective functions. Test 1–17.
Kalogridis, I., Van Aelst, S., 2019. Robust functional regression based on principal components. Journal of Multivariate Analysis 173, 393–415.
Kalogridis, I., Van Aelst, S., 2020. M-type penalized splines with auxiliary scale estimation. Journal of Statistical Planning and Inference 212, 97–113.
Kalogridis, I., Van Aelst, S., 2021. Robust penalized spline estimation with difference penalties. Econometrics and Statistics 1–30.
Kato, K., 2012. Estimation in functional linear quantile regression. Annals of Statistics 40, 3108–3136.
Kosorok, M., 2008. Introduction to Empirical Processes and Semiparametric Inference. Springer–Verlag, New York.
Krzanowski, W.J., Hand, D.J., 2009. ROC curves for continuous data. Chapman and Hall/CRC, Boca Raton.
Maronna, M., Yohai, V., 2013. Robust functional linear regression based on splines. Computational Statistics and Data Analysis 65, 46–55.
Maronna, R.A., Martin, R.D., Yohai, V.J., Salibián-Barrera, M., 2019. Robust Statistics: Theory and Methods (with R), 2nd Wiley, New York.
Newey, W., 1997. Convergence rates and asymptotic normality for series estimators. Journal of Econometrics 79, 147–168.
Oh, H.-S., Nychka, D.W., Lee, T.C.M., 2007. The role of pseudo data for robust smoothing with applications to wavelet regression. Biometrika 94, 893–904.
Pardo-Fernández, J.C., Rodríguez-Alvarez, M.X., Van Keilegom, I., 2014. A review on ROC curves in the presence of covariates. REVSTAT Statistical Journal 12, 21–41.
Peng, L., Zhou, X., 2004. Local linear smoothing of receiver operating characteristic (ROC) curves. Journal of Statistical Planning and Inference 18, 129–143.
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., 2000. An interpretation for the ROC curve and inference using GLM procedures. Biometrics 56, 352–359.
Pepe, M.S., 2003. The Statistical Evaluation of Medical Tests for Classification and Prediction. Oxford University Press, New York.
Pepe, M.S., Cai, T., 2004. The analysis of placement values for evaluating discriminatory measures. Biometrics 60, 528–535.
Powell, M.J.D., 1981. Approximation theory and methods. Cambridge University Press, Cambridge.
Pulit, M., 2016. A new method of kernel–smoothing estimation of the ROC curve. Metrika 79, 603–634.
Qingguo, T., 2015. Estimation for semi-functional linear regression. Statistics 49, 1262–1278.
Rodríguez, A., Martínez, J.C., 2014. Bayesian semiparametric estimation of covariate-dependent ROC curves. Biostatistics 15, 353–369.
Rodríguez-Álvarez, M.X., Roca-Pardiñas, J., Cadarso-Suárez, C., 2011a. ROC curve and covariates: extending the induced methodology to the non-parametric framework. Statistics and Computing 21, 483–495.
Rodríguez-Álvarez, M.X., Roca-Pardiñas, J., Cadarso-Suárez, C., 2011b. A new flexible direct ROC regression model: Application to the detection of cardiovascular risk factors by anthropometric measures. Computational Statistics and Data Analysis 55, 3257–3270.
Rodríguez-Álvarez, M.X., Tahoces, P.C., Cadarso-Suárez, C., Lado, M.J., 2011c. 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.
Ronchetti, E., 2021. The main contributions of robust statistics to statistical science and a new challenge. Metron 79, 127–1359.
Rousseeuw, P., Leroy, A.M., 1987. Robust Regression and Outlier Detection. Series in Probability and Mathematical Statistics. Wiley-Interscience, New York.
Rousseeuw, P., Raymaekers, J., Hubert, M., 2018. A measure of directional outlyingness with applications to image data and video. Journal of Computational and Graphical Statistics 27, 345–359.
Ruli, E., Ventura, L., Musio, M., 2022. Robust confidence distributions from proper scoring rules. Statistics 56, 455–478.
Schumaker, L., 1981. Spline Functions: Basic Theory. Wiley, New York.
Sun, Y., Genton, M.G., 2011. Functional boxplots. Journal of Computational and Graphical Statistics 20, 316–334.
Tharmaratnam, K., Claeskens, G., Croux, C., Salibian-Barrera, M., 2010. $s-$estimation for penalized regression splines. Journal of Computational and Graphical Statistics 5, 609–625.
To, D.K., Adimari, G., Chiogna, M., 2022. Estimation of the volume under a ROC surface in presence of covariates. Computational Statistics and Data Analysis 174, 107434.
Tosteson, A.N., Begg, C.B., 1988. A general regression methodology for ROC curve estimation. Medical Decision Making 8, 204–215.
López-de Ullibarri, I., Cao, R., Cadarso-Suárez, C., Lado, M.J., 2008. Nonparametric estimation of conditional ROC curves: Application to discrimination tasks in computerized detection of early breast cancer. Computational Statistics and Data Analysis 52, 2623–2631.
van der Vaart, A., Wellner, J., 1996. Weak Convergence and Empirical Processes with Applications to Statistics. Springer–Verlag, New York.
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.
Welsh, A.H., 1996. Robust estimation of smooth regression and spread functions and their derivatives. Statistica Sinica 6, 347–366.
Yao, F., Craiu, R.V., Reiser, B., 2010. Nonparametric covariate adjustment for receiver operating characteristic curves. The Canadian Journal of Statistics 38, 27–46.
Zheng, Y., Heagerty, P.J., 2004. Semiparametric estimation of time-dependent ROC curves for longitudinal marker data. Biostatistics 4, 615–632.
Zhou, X.H., McClish, D.K., Obuchowski, N.A., 2011. Statistical Methods in Diagnostic Medicine. John Wiley & Sons, New York.