The Canadian Journal of Statistics
Vol. 38, No. 1, 2010, Pages 27–46
La revue canadienne de statistique

Nonparametric covariate adjustment for receiver operating characteristic curves

Fang YAO[1], Radu V. CRAIU[1]* and Benjamin REISER[2]

[1]*Department of Statistics, University of Toronto, Toronto, Ontario, Canada M5S 3G3*
[2]*Department of Statistics, University of Haifa, Haifa 31905, Israel*

Key words and phrases: Area under curve; asymptotics; covariate adjustment; Mann–Whitney; nonparametric; smoothing; uniform convergence.

MSC 2000: Primary 62G08; secondary 62N05.

Abstract: The accuracy of a diagnostic test is typically characterized using the receiver operating characteristic (ROC) curve. Summarizing indexes such as the area under the ROC curve (AUC) are used to compare different tests as well as to measure the difference between two populations. Often additional information is available on some of the covariates which are known to influence the accuracy of such measures. The authors propose nonparametric methods for covariate adjustment of the AUC. Models with normal errors and possibly non-normal errors are discussed and analyzed separately. Nonparametric regression is used for estimating mean and variance functions in both scenarios. In the model that relaxes the assumption of normality, the authors propose a covariate-adjusted Mann–Whitney estimator for AUC estimation which effectively uses available data to construct *working samples* at any covariate value of interest and is computationally efficient for implementation. This provides a generalization of the Mann–Whitney approach for comparing two populations by taking covariate effects into account. The authors derive asymptotic properties for the AUC estimators in both settings, including asymptotic normality, optimal strong uniform convergence rates and mean squared error (MSE) consistency. The MSE of the AUC estimators was also assessed in smaller samples by simulation. Data from an agricultural study were used to illustrate the methods of analysis. *The Canadian Journal of Statistics* 38: 27–46; 2010 © 2010 Statistical Society of Canada

Résumé: La précision d'un test diagnostique est habituellement établie en utilisant les courbes caractéristiques de fonctionnement du récepteur (« ROC »). Des statistiques telles que l'aire sous la courbe ROC (« AUC ») sont utilisées afin de comparer différents tests et pour mesurer la différence entre deux populations. Souvent de l'information supplémentaire est disponible sur quelques covariables dont l'influence sur de telles statistiques est connue. Les auteurs suggèrent des méthodes non paramétriques afin d'ajuster la statistique AUC pour prendre en compte les covariables. Des modèles avec des erreurs gaussiennes et même non gaussiennes sont présentés et analysés séparément. Une régression non paramétrique est utilisée afin d'estimer les fonctions moyenne et variance dans les deux scénarios. Pour le modèle sans l'hypothèse de normalité, les auteurs proposent un estimateur de Mann-Whithney tenant compte des covariables pour l'AUC qui utilise l'information disponible dans les données afin de construire des échantillons d'analyse pour n'importe quelle valeur des covariables. Cet estimateur est implanté, car il est calculable de façon efficace. Il généralise l'approche de Mann-Whitney pour comparer deux populations en considérant l'effet des covariables. Les auteurs obtiennent les propriétés asymptotiques des estimateurs AUC pour les deux scénarios incluant la normalité asymptotique, les vitesses optimales de convergence uniforme forte et la convergence en erreur quadratique moyenne (« MSE »). Le MSE de l'estimateur de l'AUC est aussi étudié pour les petits échantillons à l'aide de simulations. Des données provenant d'une étude dans le domaine agricole sont utilisées afin d'illustrer les méthodes d'analyse. *La revue canadienne de statistique* 38: 27–46; 2010 © 2010 Société statistique du Canada

* *Author to whom correspondence may be addressed.*
E-mail: craiu@utstat.toronto.edu

1. INTRODUCTION

The receiver operating characteristic (ROC) curve is a commonly used tool for summarizing the accuracy of a test with binary results. The sensitivity, or true positive rate, of a binary test is the probability that a truly diseased subject is diagnosed as diseased. The specificity, which is also equal to one minus false positive rate, is defined as the probability that a healthy subject produces a negative test. Suppose that the result of a test is a random variable Y; depending on whether $Y < c$ or $Y \geq c$ the test result is considered negative or positive, respectively. If the distribution of Y is continuous, each value of the threshold c will correspond to different sensitivity and specificity values. The ROC curve is obtained when plotting sensitivity against $1-$specificity as c is increasing over its range of possible values (or, alternatively, as $1-$specificity increases from zero to one). In general the ROC curve summarizes how well two populations can be separated by a specified variable. Frequently a number of tests (a.k.a. markers or classifiers) are performed on each individual subject. A global univariate summary of the corresponding ROC curve is used to determine which classifier is more accurate. A number of such summaries are available but the most commonly used one is the *area under the ROC curve* (*AUC*). The AUC can be interpreted as the probability that a randomly chosen diseased subject will have a marker value greater than that of a randomly chosen nondiseased subject and can be used as an alternative measure of difference between two populations (e.g., Zhou, Obuchowski & McClish, 2002). Its range of application extends from medical applications to reliability theory (Reiser & Guttman, 1986).

The presence of ROC curves has become ubiquitous in medical studies (Metz, 1986; Hsiao, Bartko & Potter, 1989; Aoki et al., 1997; Otto et al., 1998; Stover, Gorga & Neely, 1996; Zhou, Obuchowski & McClish, 2002), its usage being spurred by the highly cited paper of Hanley & McNeil (1982) and the classic text of Swets & Pickett (1982). Parametric and nonparametric methods for estimating individual ROC curves are available as well as methods that do not assume independent observations (Begg, 1991; Delong, Delong & Clarke-Pearson, 1988; Molodianovitch, Faraggi & Reiser, 2006; Pepe, 2003).

In a large number of situations, additional information is available in the form of covariates which are known to influence the accuracy of the test. Only recently, statistical methods have been devised to incorporate such information in the ROC-based analysis. Some of the earlier methods have been produced by Thompson & Zucchini (1989), Obuchowski (1995), Tosteson & Begg (1988), and Toledano & Gatsonis (1995). Pepe (1997) formulated a general regression framework to model the dependence of the ROC curve directly on the covariates. Pepe (2000) and Dodd & Pepe (2003) propose semiparametric approaches to model the ROC and AUC directly using generalized linear models. Cai & Pepe (2002) extend the parametric ROC regression model by allowing an arbitrary nonparametric baseline function. Cai (2004) finds a more efficient estimator in the semiparametric setting. Brumback, Pepe & Alonzo (2006) used an alternative procedure by applying a generalized regression framework directly to the AUC in order to adjust the Mann–Whitney test for covariates. However, this approach loses the connection with the threshold value, does not allow the prediction of the sensitivity and specificity at a given threshold conditional on covariates nor does it model covariate effects on the individual marker values. Consequently we prefer to directly model the covariate effects on the marker values and through this modelling process obtain the analyses of interest.

The methods proposed in this paper fall within the first category of methods described in Pepe (1998). We propose a nonparametric approach to adjust for covariates the computation of AUC and other ROC-related quantities of interest. The main motivation for our method is the robustness to model mis-specification which may beset a parametric adjustment. We thus generalize in two ways the approaches of Faraggi (2003) and Schisterman, Faraggi & Reiser (2004) who use normal regression models to adjust the index AUC for covariates. We describe the regression model, distinguishing between the model with normal random error and the model

that relaxes the normality assumption. In a first extension of previous work, we estimate the mean and variance functions using nonparametric regression techniques, more specifically, local polynomial regression instead of parametric linear models. Our main contribution leading to the second extension is to construct a covariate-adjusted Mann–Whitney estimator (CAMWE) in the model that relaxes the normality assumption, which relies on *working samples* created at any possible covariate value $Z = z$ of interest for the estimation of AUC. Such working samples have, for any $Z = z$, the same size as the original sample and can be used to estimate a number of covariate-adjusted characteristics of the ROC curve. In practice the computation is kept minimal by utilizing the estimated mean and variance functions for all $Z = z$ of interest. We recommend bootstrapping in order to obtain confidence intervals for the covariate-adjusted AUC. Although we focus on covariate-adjusted AUC estimation, the proposed methods can be readily extended to other measures related to ROC curves, for example, the covariate-adjusted specificity, sensitivity and Youden Index (Youden, 1950).

A theoretical investigation provides asymptotic results for models which assume normal random errors and models for which we relax the normal error assumption. The asymptotic normality and optimal strong uniform convergence rates for the covariate-adjusted AUC estimators for the normal error model are established. For the model without the normality assumption we first derive asymptotic normality of the "hypothetical" CAMWE and then characterize the asymptotic behaviour of the mean squared error (MSE) of the CAWME. We performed simulations under a number of scenarios to demonstrate effectiveness and robustness of the proposed estimators as well as the validity of the bootstrap scheme for confidence band construction.

2. MODEL AND ESTIMATION

2.1. Regression Model

To motivate our proposal, we first note that parametric methods are used mainly for simple interpretation but may mis-specify the correct model forms, while nonparametric models provide an alternative solution and are more robust and data-adaptive. We attempt to achieve the robustness from two perspectives. First, we do not assume any parametric forms for the mean and variance functions of the test response variables, X for nondiseased individuals and Y for diseased individuals. Although we refer to "diseased" and "nondiseased" groups, the above framework applies to any two populations of interest. We utilize nonparametric regression models

$$X|Z = \mu_1(Z) + \sqrt{v_1(Z)}\epsilon_1, \tag{1}$$

$$Y|Z = \mu_2(Z) + \sqrt{v_2(Z)}\epsilon_2, \tag{2}$$

where Z denotes the covariate, the standardized errors ϵ_1 and ϵ_2 are independent of each other with zero mean and unit variance, and the variance functions $0 < v_1(z) < \infty$ and $0 < v_2(z) < \infty$ for all $z \in \Re$. Note that the errors here can depend heteroscedastically on the covariate Z through v_1 and v_2. Second, we do not assume specific distributions for the noises in order to guard against mis-specification of error distributions. Denote the conditional cumulative distribution functions (c.d.f.) of X and Y given Z by $F(\cdot|Z)$ and $G(\cdot|Z)$, and c.d.f.s of ϵ_1 and ϵ_2 by $F^*(\cdot)$ and $G^*(\cdot)$. Here we assume F^* and G^* do not depend on Z, that is, the dependence of X and Y on Z are expressed only through μ_1, μ_2, v_1, and v_2, which is equivalent to a location-scale model. It is worth mentioning that, if the response variable is appropriately chosen at $Z = z$, then marker values of the diseased sample should be greater than that of the nondiseased sample on average. This is equivalent to $P(Y > X|Z = z) > 0.5$, an assumption implicitly made for the remaining of the paper. If the baseline distributions F^* and G^* are symmetric about 0, it implies the assumption $\mu_2(z) > \mu_1(z)$. In practice, we can simply constrain all the AUC estimators to be greater than

0.5. This would not affect any subsequent development due to the consistency of the unrestricted estimators as presented in Section 3. For notational convenience, we use the unrestricted forms throughout the paper.

This extends the first type of models discussed by Pepe (1998), where linear forms were assumed for μ_1 and μ_2 with variances not depending on the covariate Z, that is, $\mu_2(z) = \alpha_0 + \alpha_1 + (\alpha_2 + \alpha_3)z$, $\mu_1(z) = \alpha_0 + \alpha_2 z$, $v_1(z) = v_1$ and $v_2(z) = v_2$. It is also noticed that we do not require the same baseline distributions of the standardized error ϵ_1 and ϵ_2 in contrast to Pepe (1998). Moreover, when the noise is not normally distributed, we shall propose a new estimator for the area under the ROC curve that extends the Mann–Whitney estimator for covariate-adjustment by using standardized residuals via the so-called *working samples*.

2.2. Estimation for Model With Normal Error Assumption

Let $A(z)$ be the area under the ROC curve with the covariate adjustment $Z = z$. From models (1) and (2), when the errors ϵ_1 and ϵ_2 are normally distributed, that is, $F^* = G^* = \Phi$, where $\Phi(\cdot)$ is the c.d.f. of the standard normal, it is straightforward to derive the following explicit expression:

$$A_{\mathrm{N}}(z) = P(Y > X|Z = z) = \Phi\left\{\frac{\mu_2(z) - \mu_1(z)}{\sqrt{v_1(z) + v_2(z)}}\right\}, \tag{3}$$

where the subscript N stands for the normal assumption. One can also obtain closed forms of the sensitivity $q_{\mathrm{N}}(z)$ and specificity $p_{\mathrm{N}}(z)$ for $Z = z$,

$$q_{\mathrm{N}}(z) = \Phi\left\{\frac{\mu_2(z) - c}{\sqrt{v_2(z)}}\right\}, \qquad p_{\mathrm{N}}(z) = \Phi\left\{\frac{c - \mu_1(z)}{\sqrt{v_1(z)}}\right\}, \tag{4}$$

for a given threshold c. The ROC curve for the covariate $Z = z$ is the plot of $q(z)$ versus $1 - p(z)$ for all possible values of c, and this can be explicitly written as

$$q_{\mathrm{N}}(z) = \Phi\left[\frac{\mu_2(z) - \mu_1(z) + \sqrt{v_1(z)}\Phi^{-1}\{1 - p_{\mathrm{N}}(z)\}}{\sqrt{v_2(z)}}\right]. \tag{5}$$

The unknown functions f, g, v_1, v_2, are estimated by using nonparametric regression methods as addressed in Section 3.1, providing a "nonparametric adjustment" as discussed in Section 1.

2.3. Estimation for Model Without Normal Error Assumption

The assumption of normal error above simplifies the calculations of the AUC via (3) but is not always supported by the data. In addition, the normality assumption hampers the full generality one expects from a nonparametric model. We propose here a fully nonparametric yet simple estimator of the AUC with covariate adjustment, $A(z) = P(Y > X|Z = z)$, for the model with the normal error assumption.

The proposed estimator is motivated by the classical Mann–Whitney statistic, which is formulated for two samples $\{x_1, \ldots, x_m\}$ and $\{y_1, \ldots, y_n\}$ as

$$M_{m,n} = \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n} 1_{[0,\infty)}(y_j - x_i), \tag{6}$$

where $1_{[0,\infty)}(x) = 1$ if $x \geq 0$ and $1_{[0,\infty)}(x) = 0$ otherwise. The data obtained from nondiseased and diseased samples consist of $\{(z_{i,x}, x_i) : i = 1, \ldots, m\}$ and $\{(z_{j,y}, y_j) : j = 1, \ldots, n\}$, where $z_{i,x}$ is the observed covariate value in the nondiseased sample and $z_{j,y}$ in the diseased sample. It should be noticed that the markers X and Y are evaluated at possibly different values of the

covariate Z, and we are often interested in estimating $A(z)$ even for z-values which were not measured in either group or both. To estimate $A(z)$ at $Z = z$, one possibility is to include the marker values x_i and y_j that fall into neighbourhoods of z with appropriate weight functions. This consideration naturally leads to a bivariate kernel estimator that is fully nonparametric,

$$\widehat{A}_{\mathrm{K}}(z) = \frac{\sum_{i=1}^{m}\sum_{j=1}^{n} 1_{[0,\infty)}(y_j - x_i)K_{h_x}(z_{i,x} - z)K_{h_y}(z_{j,y} - z)}{\sum_{i=1}^{m}\sum_{j=1}^{n} K_{h_x}(z_{i,x} - z)K_{h_y}(z_{j,y} - z)}, \tag{7}$$

where h_x and h_y are bandwidths, $K_h(\cdot) = (1/h)K(\cdot/h)$ when $K(\cdot)$ is a symmetric kernel density. However, $\widehat{A}_{\mathrm{K}}$, does not efficiently use the available data due to the restriction on the local windows, nor do the regression models (1) and (2) play any role here. Note that $\widehat{A}_{\mathrm{K}}$ is obtained by smoothing the binary variables $1_{[0,\infty)}(y_j - x_i)$ corresponding to covariate observations $(z_{i,x}, z_{j,y}) \in [z - h_x, z + h_x] \times [z - h_y, z + h_y]$ that are not necessarily located on the diagonal (in fact, $\{z_{i,x}\}$ and $\{z_{j,y}\}$ may have no overlap). It is unclear how to choose the bandwidths h_x and h_y which are critical to the kernel regression estimation, as the standard cross-validation procedure does not apply due to the absence of the observed $(z_{i,x}, z_{j,y}, 1_{[0,\infty)}(y_j - x_i))$ on the diagonal of the bivariate covariate surface. More discussion and comparisons concerning $\widehat{A}_{\mathrm{K}}(z)$ will be presented in simulations in Section 4.

Based on the above considerations, we propose a different nonparametric estimator of $A(z)$ which utilizes the entire collection of data available and the regression models (1) and (2). First, suppose that we can observe all the standardized residuals, $i = 1, \ldots, m$, $j = 1, \ldots, n$,

$$\epsilon_{i,x} = \frac{x_i - \mu_1(z_{i,x})}{\sqrt{v_1(z_{i,x})}}, \qquad \epsilon_{j,y} = \frac{y_j - \mu_2(z_{j,y})}{\sqrt{v_2(z_{j,y})}}. \tag{8}$$

Recall that the distributions of ϵ_1 and ϵ_2 do not depend on Z, implying that $\epsilon_{1,i}$ are independently and identically distributed (i.i.d.) with the c.d.f. F^* for $i = 1, \ldots, m$, and $\epsilon_{2,j}$ are i.i.d. with the c.d.f. G^* for $j = 1, \ldots, n$. In Pepe (1998) these standardized residuals can be used to obtain the empirical distributions of ϵ_1 and ϵ_2. In a similar sprit, we propose a different way to utilize these residuals to construct *working samples* $\{x_{i,z}, \ldots, x_{m,z}\}$ and $\{y_{1,z}, \ldots, y_{n,z}\}$ as if they were all observed at $Z = z$,

$$x_{i,z} = \mu_1(z) + \sqrt{v_1(z)}\epsilon_{i,x}, \qquad y_{j,z} = \mu_2(z) + \sqrt{v_2(z)}\epsilon_{j,y}. \tag{9}$$

Then it is intuitive to use the proposed Covariate-Adjusted Mann–Whitney Estimator (CAMWE) for $A(z)$,

$$A_M(z) = \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n} 1_{[0,\infty)}(y_{j,z} - x_{i,z}). \tag{10}$$

This is a natural extension of the Mann–Whitney estimator since in the case of no covariate effect μ_1, μ_2, v_1, v_2 are constant in z and (10) becomes the traditional Mann–Whitney statistic. For practical implementation, after obtaining nonparametric estimates of μ_1, μ_2, v_1, and v_2, we do not have to choose other tuning parameters for each covariate value $Z = z$, while (7) requires retuning. Analogously we can calculate the sensitivity and specificity from the working samples for $Z = z$,

$$q_{\mathrm{M}}(z) = \frac{1}{n}\sum_{j=1}^{n} 1_{[0,\infty)}(y_{j,z} \geq c), \qquad p_{\mathrm{M}}(z) = \frac{1}{m}\sum_{i=1}^{m} 1_{[0,\infty)}(x_{i,z} \leq c), \tag{11}$$

for a given threshold c. The ROC curves for $Z = z$ can be obtained by plotting $q_{\mathrm{M}}(z)$ versus $1 - p_{\mathrm{M}}(z)$ for all possible values of c.

Remark 1. Note that the central idea is to construct the *working sample* $\{x_{i,z}, \ldots, x_{m,z}\}$ and $\{y_{1,z}, \ldots, y_{n,z}\}$ for each $Z = z$. The entire conditional ROC curve, given the covariate value $Z = z$, can be obtained from (11). One can estimate any index of interest at $Z = z$ using this working sample. For instance, the Youden Index (YI) (Youden, 1950) can be calculated by $\mathrm{YI}_{\mathrm{M}}(z) = p_{\mathrm{M}}(z) + q_{\mathrm{M}}(z) - 1$, where $p_{\mathrm{M}}(z)$ and $q_{\mathrm{M}}(z)$ are defined by (11), and its optimal threshold given $Z = z$ can be found via a numerical search.

Remark 2. In principle, the proposed approach can be extended to the case of multiple covariates using different strategies. A natural consideration is to use multivariate nonparametric smoothing techniques that require extensive computation. An alternative is to use additive frameworks for mean and variance structures, respectively, then construct the working sample in a similar spirit for each set of covariate values of interest.

2.4. Implementation Via Nonparametric Regression

We exploit the local polynomial regression models for estimating the functions μ_1 and μ_2. Let $K(\cdot)$ be a compactly supported symmetric kernel density function with a finite variance, $h_1 = h_1(m)$ a sequence of bandwidths used to estimate μ_1, and $h_2 = h_2(n)$ a sequence of bandwidths for μ_2. Let p be the order of local polynomial fit, for example, $p = 0$ and $p = 1$ correspond to local constant and local linear fits, respectively. An odd order fit is often suggested (Fan & Gijbels, 1996) for both theoretical and practical considerations. In particular, for estimating the regression function itself, a common choice is the local linear fit with $p = 1$. Denote the resulting pth order local polynomial estimators of $\mu_1(z)$ and $\mu_2(z)$ by $\hat{\mu}_1(z)$ and $\hat{\mu}_2(z)$. Next, the variance functions $v_1(z)$ and $v_2(z)$ for heteroscedastic errors are estimated by fitting local polynomial regression to the squared residuals, $v_{i,x}$ and $v_{j,y}$, $i = 1, \ldots, m$, $j = 1, \ldots, n$,

$$v_{i,x} = \{x_i - \hat{\mu}_1(z_{i,x})\}^2, \qquad v_{j,y} = \{y_j - \hat{\mu}_2(z_{j,y})\}^2, \tag{12}$$

with bandwidths $b_1 = b_1(m)$ and $b_2 = b_2(n)$. The detailed formulas of the aforementioned local polynomial estimators are given in the Appendix. In the case of homoscedastic errors, $v_1(z) \equiv v_1$ and $v_2(z) \equiv v_2$, it is easy to obtain root-n consistent estimators (Hall & Marron, 1990; Hall, Kay & Titterington, 1990). The theoretical properties in Section 4 are still valid with slight modifications. In practice, the bandwidths h_1, h_2, b_1, and b_2 are chosen by the standard technique of leave-one-out cross-validation for estimating the mean and variance functions, while other existing techniques can certainly be applied. Such bandwidths usually fulfill the assumptions needed for theoretical developments in Section 3 for sufficiently large sample sizes. Substituting the local polynomial estimators $\hat{\mu}_1(z)$, $\hat{\mu}_2(z)$, $\hat{v}_1(z)$, and $\hat{v}_2(z)$ for these unknown quantities in formulae (3)–(5), (10), and (11) provides the point estimators $\widehat{A}_{\mathrm{N}}(z)$, $\hat{p}_{\mathrm{N}}(z)$, $\hat{q}_{\mathrm{N}}(z)$, $\widehat{A}_{\mathrm{M}}(z)$ $\hat{p}_{\mathrm{M}}(z)$, and $\hat{q}_{\mathrm{M}}(z)$ for covariate $Z = z$.

To evaluate confidence limits and variances for AUC under normality assumption, the existing formulation (Guttman et al., 1988; Faraggi, 2000, among others) is no longer valid due to nonparametric regression. In principle we can derive the approximate variance for AUC under the normality assumption, based on the asymptotic normality of the local polynomial estimators (Fan & Gijbels, 1996) using the Cramér–Wold device. However, due to the complicated asymptotic expressions with unknown functionals and their derivatives, the evaluation of such asymptotic quantities will require extensive pilot smoothing and further approximations. This might deteriorate the accuracy and not be worth further pursuing. Thus we choose to obtain confidence limits and variance estimates for AUC via "bootstrapping the original data" as proposed by Efron &

Tibshirani (1993). We do not repeat the procedure here for conciseness. While this approach can be justified in normal error model due to the limiting distributions in Theorem 1, it may not be the case under the model without normal error assumption for which the asymptotic normality of the CAMWE $\widehat{A}_{\mathrm{M}}(z)$ is unknown at this moment. Nevertheless, the simulation performed in Section 4.1 offers empirical support to this bootstrap procedure for the non-normal error models.

Remark. Jointly choosing four bandwidths for aiming at the AUC estimator is prohibitively expensive, even impossible with available computing resources. Even if the computation load were not an issue, we would have no suitable criterion to perform the joint optimization for two reasons. First, if one bases the criterion on asymptotic bias and variance, these quantities involve unknown functionals and their derivatives and are too complicated for practical use. It should also be noticed that such asymptotic expressions are established only for the normal error model. Second, if one attempts cross-validation directly for $A(z)$, there are no observed values of AUC at $Z = z$ available, which is a similar issue as the one discussed for $\widehat{A}_{\mathrm{K}}$ in Section 2.3.

3. ASYMPTOTIC PROPERTIES OF AUC ESTIMATORS

In this section we present the asymptotic theory developed for the nonparametric estimators of the AUC with covariate adjustment for $Z = z$ under both normal and non-normal error assumptions. We defer the regularity conditions and technical proofs to the Appendix, and two auxiliary lemmas are provided in the Supplement file available at http://fisher.utstat.toronto.edu/fyao/publication.html due to space limitations. One can easily extend these arguments to obtain the corresponding asymptotic theory for the sensitivity $q(z)$ and specificity $p(z)$ with a given threshold value c. These are not presented here for conciseness.

3.1. Asymptotic Properties for the Model With Normal Error Assumption

We begin with the asymptotic normality of the estimated AUC under the normal error assumption, where the target $A(z)$ is exactly $A_{\mathrm{N}}(z)$, that is, $A(z) \equiv A_{\mathrm{N}}(z)$. Let $\theta(z)$ be the density function of the covariate Z that is treated as a random variable. Put $\eta_1(z) = E(\epsilon_1^3|Z = z)$, $\eta_2(z) = E(\epsilon_2^3|Z = z)$, $\kappa_1(z) = \mathrm{Var}(\epsilon_1^2|Z = z)$, and $\kappa_2(z) = \mathrm{Var}(\epsilon_2^2|Z = z)$. Recall that $h_1 = h_1(m)$, $b_1 = b_1(m)$, $h_2 = h_2(n)$, and $b_2 = b_2(n)$ are the sequences of bandwidths for estimating $\mu_1(z)$, $v_1(z)$, $\mu_2(z)$, and $v_2(z)$. One can see that, if the bandwidths h_1 and b_1 are chosen optimally for estimating $\mu_1(z)$ and $v_1(z)$, then h_1 and b_1 will be of the same order in terms of the sample size m. Here we consider the odd order p of local polynomial estimators for μ_1, v_1, μ_2, and v_2 as argued in Section 2.4. The same order p is used mainly for notational convenience, while we certainly can choose different orders in practice. With slight modifications, the results can be easily adapted to possibly different orders as well as the case of even p.

For the symmetric kernel density $K(\cdot)$, we assume that the j-th moment $m_j(K) = \int u^j K(u)\,\mathrm{d}u$ exists for all integers $j \geq 0$, moreover, $R(K) = \int K^2(u) < \infty$, $\mu_2(K) > 0$. For convenience, we introduce the notion of the order of a kernel function. We say $\tilde{K}$ is an ℓth order kernel function, provided that $m_0(\tilde{K}) = 1$, $m_j(\tilde{K}) = 0$ for $j = 1, \ldots, \ell - 1$ and $m_\ell(\tilde{K}) \neq 0$. It is obvious that $K(\cdot)$ is a 2nd order kernel. Let the $(p + 1) \times (p + 1)$ matrix $S_p = \{m_{j+l}(K)\}_{0 \leq j,l \leq p}$, $\boldsymbol{e}_k$ be the $(p + 1) \times 1$ vector with the kth element equal to 1 and 0 elsewhere, and

$$K^*(u) = \boldsymbol{e}_1^{\mathrm{T}} S_p^{-1}(1, u, \ldots, u^p)^{\mathrm{T}} K(u), \tag{13}$$

which is often referred to as the equivalent kernel. One can verify that $K^*(\cdot)$ is a $(p + 1)$th order kernel when p is odd. Also denote $R(K^*, \rho) = \int K^*(u)K^*(u/\rho)\,\mathrm{d}u$ for any $0 < \rho < \infty$.

Theorem 1. *Under the assumptions* (A1)–(A5) *for a given z,*

- *if* $n/m \to \infty$, $\sqrt{mh_1}\{\widehat{A}_{\mathrm{N}}(z) - A_{\mathrm{N}}(z)\} \xrightarrow{D} N\{B_1(z), V_1(z)\}$, *where* $\phi(u) = (2\pi)^{-1/2}\,\mathrm{e}^{-u^2/2}$, $\delta(z) = \{\mu_2(z) - \mu_1(z)\}/\sqrt{v_1(z) + v_2(z)}$,

$$B_1(z) = -\frac{\phi\{\delta(z)\} m_{p+1}(K^*) d_1}{(p+1)!\sqrt{v_1(z) + v_2(z)}} \left[\mu_1^{(p+1)}(z) + \frac{\{\mu_2(z) - \mu_1(z)\} v_1^{(p+1)}(z) \rho_1^{p+1}}{2\{v_1(z) + v_2(z)\}}\right],$$

$$V_1(z) = \frac{\phi^2\{\delta(z)\}}{\theta(z)\{v_1(z) + v_2(z)\}} \left[R(K^*) v_1(z) + \frac{\{\mu_2(z) - \mu_1(z)\} R(K^*, \rho_1) \eta_1(z)}{\{v_1(z) + v_2(z)\}\rho_1} + \frac{\{\mu_2(z) - \mu_1(z)\}^2 R(K^*) \kappa_1(z)}{4\{v_1(z) + v_2(z)\}^2 \rho_1}\right], \tag{14}$$

- *if* $n/m \to 0$, $\sqrt{nh_2}\{\widehat{A}_{\mathrm{N}}(z) - A_{\mathrm{N}}(z)\} \xrightarrow{D} N\{B_2(z), V_2(z)\}$, *where*

$$B_2(z) = \frac{\phi\{\delta(z)\} m_{p+1}(K^*) d_2}{(p+1)!\sqrt{v_1(z) + v_2(z)}} \left[\mu_2^{(p+1)}(z) - \frac{\{\mu_2(z) - \mu_1(z)\} v_2^{(p+1)}(z) \rho_2^{p+1}}{2\{v_1(z) + v_2(z)\}}\right],$$

$$V_2(z) = \frac{\phi^2\{\delta(z)\}}{\theta(z)\{v_1(z) + v_2(z)\}} \left[R(K^*) v_2(z) - \frac{\{\mu_2(z) - \mu_1(z)\} R(K^*, \rho_2) \eta_2(z)}{\{v_1(z) + v_2(z)\}\rho_2} + \frac{\{\mu_2(z) - \mu_1(z)\}^2 R(K^*) \kappa_2(z)}{4\{v_1(z) + v_2(z)\}^2 \rho_2}\right], \tag{15}$$

- *if* $n/m \to \lambda$ *for some* $0 < \lambda < \infty$, $\sqrt{mh_1}\{\widehat{A}_{\mathrm{N}}(z) - A_{\mathrm{N}}(z)\} \xrightarrow{D} N\{B_3(z), V_3(z)\}$, *where*

$$B_3(z) = B_1(z) + \lambda^{-(p+1)/(2p+3)} B_2(z), \qquad V_3(z) = V_1(z) + \lambda^{-(2p+2)/(2p+3)} V_2(z) \tag{16}$$

Besides the pointwise limiting distributions, we also establish the optimal rates for strong uniform convergence of $\widehat{A}_{\mathrm{N}}$ in Theorem 2, where a.s. is the abbreviation of "almost surely".

Theorem 2. *Under the assumptions* $(A1^\dagger)$–$(A5^\dagger)$ *and* $(A6)$–$(A8)$, *let* $\tau_m = h_1^{p+1} + \sqrt{\log(1/h_1)/(mh_1)}$ *and* $\omega_n = h_2^{p+1} + \sqrt{\log(1/h_2)/(nh_2)}$, *then*

$$\sup_{z \in \mathcal{Z}} |\widehat{A}_{\mathrm{N}}(z) - A_{\mathrm{N}}(z)| = O(\tau_m + \omega_n) \quad \text{a.s.} \tag{17}$$

3.2. Asymptotic Properties for the Model Without Normal Error Assumption

Now we turn to the asymptotic properties of the CAMWE $\widehat{A}_{\mathrm{M}}(z)$ of $A(z)$ for the model that relaxes the normal error assumption. We first state the asymptotic normality of the "hypothetical" estimator $A_{\mathrm{M}}(z)$ (10) that contains true values of the unknown mean and variance functions, while our target is $A(z) = P(Y > X | Z = z)$. Recall that F^* and G^* are the c.d.f.s of standardized errors ϵ_1 and ϵ_2, and do not depend on the covariate Z. Define

$$h_{1,0}(\epsilon_1; z) = G^*\left\{\sqrt{\frac{v_1(z)}{v_2(z)}}\epsilon_1 + \frac{\mu_1(z) - \mu_2(z)}{\sqrt{v_2(z)}}\right\},$$

$$h_{0,1}(\epsilon_2; z) = F^*\left\{\sqrt{\frac{v_2(z)}{v_1(z)}}\epsilon_2 + \frac{\mu_2(z) - \mu_1(z)}{\sqrt{v_1(z)}}\right\}.$$

Set $\xi^2_{1,0}(z) = \text{var}\{h_{1,0}(\epsilon_1; z)\}$ and $\xi^2_{0,1}(z) = \text{var}(h_{0,1}\{\epsilon_2; z)\}$.

Theorem 3. *For the regression models* (1) *and* (2) *and a given* z*,*

$$E\{A_{\text{M}}(z)\} = A(z), \quad \text{var}\{A_{\text{M}}(z)\} = O\left(\frac{1}{m+n}\right). \tag{18}$$

If $n/m \to \lambda$ *for some* $0 < \lambda < \infty$*,* $\xi^2_{1,0}(z) > 0$ *and* $\xi^2_{0,1}(z) > 0$*, then*

$$\sqrt{m+n}\{A_{\text{M}}(z) - A(z)\} \xrightarrow{D} N\left\{0, \frac{\xi^2_{1,0}(z)}{\lambda^*} + \frac{\xi^2_{0,1}(z)}{1-\lambda^*}\right\}, \tag{19}$$

where $\lambda^* = 1/(1+\lambda)$*.*

In the next theorem we establish the MSE consistency of the CAMWE $\widehat{A}_{\text{M}}(z)$ for the "hypothetical" estimator $A_{\text{M}}(z)$ for a given covariate $Z = z$, based on uniform consistency of the estimated mean and variance functions. It is noticed in the proof that we actually do not need the optimal strong uniform convergence rates stated in Lemma 2, as these rates cannot be passed to $\widehat{A}(z)$, while uniform consistency in probability is sufficient. Thus the assumptions (A3†) and (A5†) can be relaxed as (A3*) and (A5*), see Appendix for details.

Theorem 4. *Under* (*A*9) *and the assumptions for Theorem* 2 *with* (*A*3†) *and* (*A*5†) *replaced by* (*A*3*) *and* (*A*5*)*, for a given* z*,*

$$E[\{\widehat{A}_{\text{M}}(z) - A_{\text{M}}(z)\}^2] \longrightarrow 0. \tag{20}$$

We conclude this section with the following corollary that is a direct consequence of Theorems 3 and 4. Note that the MSE discrepancy between estimated and true AUC at $Z = z$ is dominated by the nonparametric rate in (20) that is usually slower the parametric rate $(m+n)^{-1/2}$, although its order of magnitude is not obtainable, at least to our knowledge.

Corollary 1. *Under* (*A*9) *and the assumptions for Theorem* 2 *with* (*A*3†) *and* (*A*5†) *replaced by* (*A*3*) *and* (*A*5*)*, for a given* z*,*

$$E[\{\widehat{A}_{\text{M}}(z) - A(z)\}^2] \longrightarrow 0. \tag{21}$$

4. SIMULATION AND DATA EXAMPLE

4.1. Simulations

The purpose of the simulations is to assess the performance of the methods for estimating AUC in nonparametric regression settings. We have not compared our method with parametric models since the two approaches address different situations. If a parametric model is correctly specified, its performance will be superior to a nonparametric procedure; however, if there is no known parametric model suitable for the data considered, one will have no choice but to use the nonparametric tools available.

We consider three situations for illustration. In the first situation the underlying models are, for nondiseased and diseased individuals, respectively,

$$\begin{aligned} x_i &= 6 + 1.5z_{i,x} + 1.5\sin(z_{i,x}) + \sqrt{v_1(z_{i,x})}\epsilon_{i,x} \\ y_j &= 6 + 1.5z_{j,y} + 1.5\sin(z_{j,y}) + \sqrt{z_{j,y} - 0.5} + \sqrt{v_1(z_{i,x})}\epsilon_{j,y}, \end{aligned} \tag{22}$$

where the errors $\epsilon_{i,x}$ and $\epsilon_{j,y}$ are standard normal, the conditional variance functions are $v_1(z) = 0.3 + \Phi(2z - 6)$ and $v_2(z) = 1.5 + \Phi(2z - 6)$, $i = 1, \ldots, m$, $j = 1, \ldots, n$. The covariates $z_{i,x}$ and $z_{j,y}$ are independently generated from $U[1, 5]$, and moderate sample sizes $n = m = 40$ are used. The identical setting is used in the second situation, except that the errors $\epsilon_{i,x}$ and $\epsilon_{j,y}$ are generated from a Student-t distribution with 3 degrees of freedom and rescaled to have zero mean and unit variance.

The third situation, in which the log-transformed responses have normal errors $\epsilon^*_{i,x}$ and $\epsilon^*_{j,y}$, that is, the responses are generated from log-normal models, is designed to demonstrate the robustness of the proposed CAMWE $\widehat{A}_{\mathrm{M}}(z)$. Since a log-transform often stabilizes the variability, we assume a constant variance σ^2 on log-scale for both groups. Let $\mu^*_1(\cdot)$ and $\mu^*_2(\cdot)$ be the mean functions on log-scale, while $\mu_1(\cdot)$ and $\mu_2(\cdot)$ correspond to the original scale. From the properties of the log-normal distribution, one has

$$\log\{\mu_1(z)\} = \mu^*_1(z) + \frac{\sigma^2}{2}, \qquad v_1(z) = (e^{\sigma^2} - 1)\mu^2_1(z)$$

$$\log\{\mu_2(z)\} = \mu^*_2(z) + \frac{\sigma^2}{2}, \qquad v_2(z) = (e^{\sigma^2} - 1)\mu^2_2(z).$$

We choose $\mu_1(z) = 1 - 0.5z - 0.25\sin(\pi z)$ and $\mu_2(z) = 1 - 0.5z - 0.25\sin(\pi z) + 1.5\sqrt{z + 0.5}$, $z \in [0, 1]$, and $\sigma^2 = 1/3$. Then the models are completely determined and can be written as

$$x_i = \exp\{\mu^*_1(z_{i,x}) + \sigma\epsilon^*_{i,x}\}, \qquad y_j = \exp\{\mu^*_2(z_{j,y}) + \sigma\epsilon^*_{j,y}\}, \tag{23}$$

where the covariates $z_{i,x}$ and $z_{j,y}$ are independently generated from $U[0, 1]$, $\epsilon^*_{i,x}$ and $\epsilon^*_{j,y}$ are standard normal errors, $i = 1, \ldots, m$, $j = 1, \ldots, n$.

With the generated data we compared three estimators, $\widehat{A}_{\mathrm{N}}(z)$ with normal error assumption, CAMWE $\widehat{A}_{\mathrm{M}}(z)$ with non-normal error assumption as well as the kernel estimator $\widehat{A}_{\mathrm{K}}(z)$. For bandwidth choices, recall that joint selection aiming for $\widehat{A}_{\mathrm{N}}(z)$ and $\widehat{A}_{\mathrm{M}}(z)$ is not feasible and that cross-validation fails for $\widehat{A}_{\mathrm{K}}(z)$. To make the comparisons possible, for $\widehat{A}_{\mathrm{N}}(z)$ and $\widehat{A}_{\mathrm{M}}(z)$ we minimized the true integrated squared errors, respectively, say $\int\{\hat{\mu}_1(z; h_1) - \mu_1(z)\}^2\,\mathrm{d}z$ to select h_1, and similarly for h_2, b_1, and b_2, while $\int\{\widehat{A}_{\mathrm{K}}(z; h_x, h_y) - A(z)\}^2\,\mathrm{d}z$ was minimized for choosing h_x and h_y in $\widehat{A}_{\mathrm{K}}(z)$. One can see that, if one targets at $A(z)$, the bandwidths chosen for $\widehat{A}_{\mathrm{N}}(z)$ and $\widehat{A}_{\mathrm{M}}(z)$ may not be as "optimal" as those for $\widehat{A}_{\mathrm{K}}(z)$. However, it is demonstrated below that even in such a disadvantageous situation, the proposed estimators, especially $\widehat{A}_{\mathrm{M}}(z)$, are still preferable. We used the sample sizes of $n = m = 40$ and $n = m = 100$, while all the estimates were improved with increased sample sizes as expected. All three AUC estimates are obtained by applying the estimation procedures to the simulated data $\{(z_{i,x}, x_i)\}_{i=1,\ldots,m}$ and $\{(z_{j,y}, y_j)\}_{j=1,\ldots,n}$ (on original scale throughout) in the aforementioned three situations. Monte Carlo averages (calculated from 500 runs in each case) of Mean Squared Errors at different values of z are presented in Figure 1. We can see that, for the normal error model the CAMWE $\widehat{A}_{\mathrm{M}}(z)$ and normal estimator $\widehat{A}_{\mathrm{N}}(z)$ are comparable and both outperform the kernel estimator $\widehat{A}_{\mathrm{K}}(z)$. Although $\widehat{A}_{\mathrm{K}}(z)$ improves upon $\widehat{A}_{\mathrm{N}}(z)$ under the heavy-tailed Student-t error model, the CAMWE $\widehat{A}_{\mathrm{M}}(z)$ is still the most effective. For the log-normal model, when we apply these three estimation procedures to the original responses, the CAMWE and kernel estimators yield comparable results (CAMWE seems slightly better), and both significantly improved upon the normal estimator (Figures 2 and 3).

Now we examine the empirical performance of the pointwise confidence bands and variance estimates obtained by "bootstrapping the data" in the possibly non-normal error models, that is when the CAMWE $\widehat{A}_{\mathrm{M}}(z)$ is used for estimation, we carried out an additional study. We

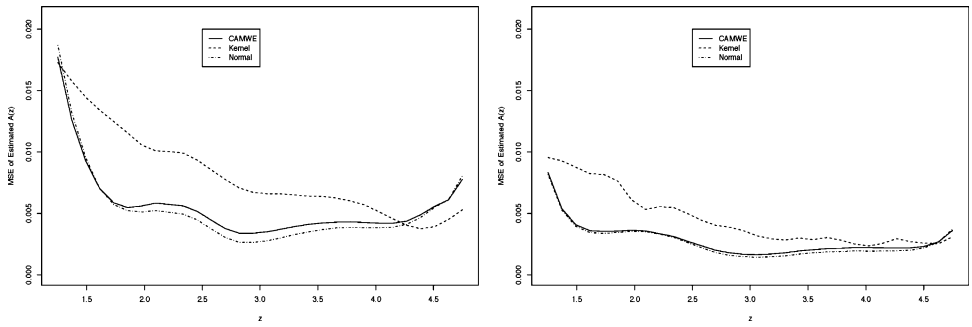

FIGURE 1: Simulation results for the normal model, where the sample sizes are $m = n = 40$ (left) and $m = n = 100$ (right), respectively. Shown are Monte Carlo averages of Mean Squared Errors (MSE) of three estimators, $\widehat{A}_{\mathrm{M}}$ (CAMWE, solid), $\widehat{A}_{\mathrm{N}}$ (Normal, dash-dotted), and $\widehat{A}_{\mathrm{K}}$ (Kernel, dashed) at different values of z.

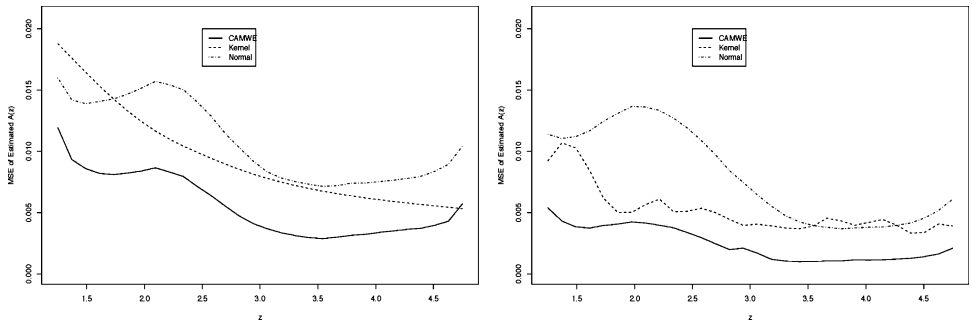

FIGURE 2: Simulation results with the same setting as Figure 1, except that the model has a Student-t error with 3 degrees of freedom.

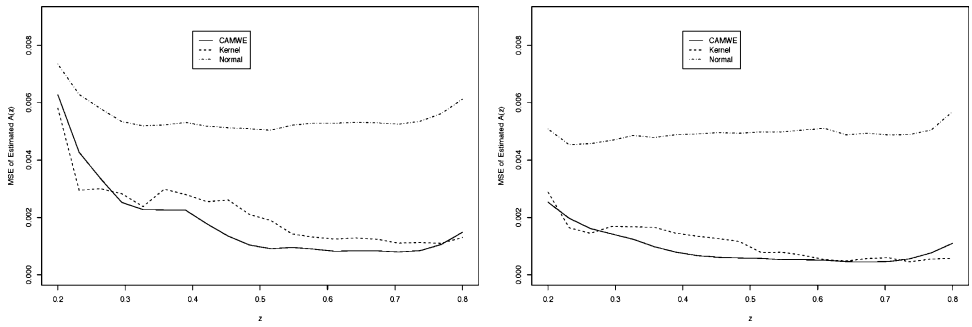

FIGURE 3: Simulation results with the same setting as Figure 1, except that the model has a lognormal error.

used the same settings for the three models with normal, Student with 3 degrees of freedom and log-normal errors, respectively. The benchmark used for comparison is the 95% pointwise confidence bands and variance estimates averaged from 500 Monte Carlo runs. In each Monte Carlo run, $\widehat{A}_{\mathrm{M}}(z)$ was obtained and we bootstraped the data 1,000 times to calculate 95% bootstrap bands (defined between the 2.5th and the 97.5th percentiles) and bootstrap sample variance. All the bandwidths involved in the estimation are selected, respectively, by leave-one-out cross-validation in smoothing steps. In the left panels of Figures 4–6 we reported, for all three data-generating

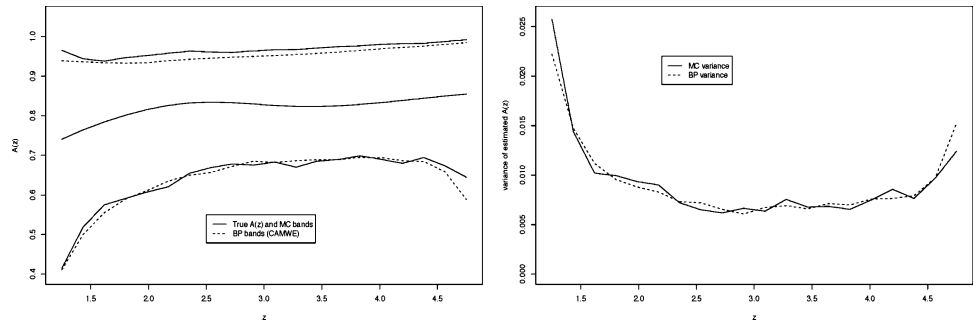

FIGURE 4: Simulation results of 95% pointwise bootstrap confidence bands (left) and variance comparisons (right) for the model with normal error with the same settings as in Figure 1 and sample sizes $n = m = 40$. Left: True AUC and 95% pointwise Monte Carlo (MC) bands (solid) obtained from 500 runs, and the MC averages of 95% pointwise bootstrap bands (dashed). Right: MC variance estimates (solid) versus MC averages of bootstrap variance estimates (dashed), as described in Section 4.1.

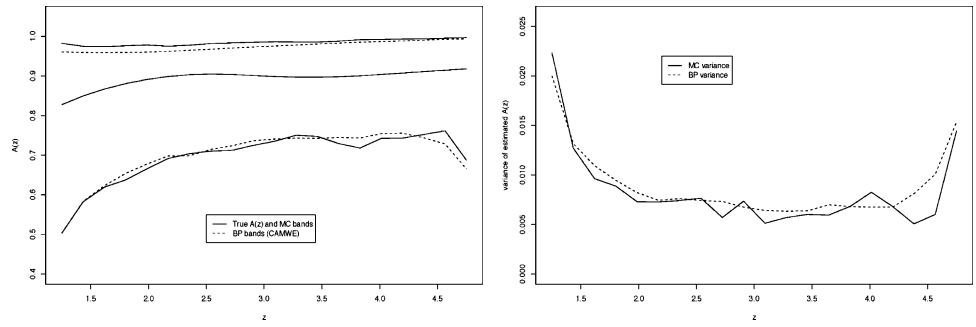

FIGURE 5: Simulation results with the same setting as Figure 4, except that the model has a Student-t error with 3 degrees of freedom.

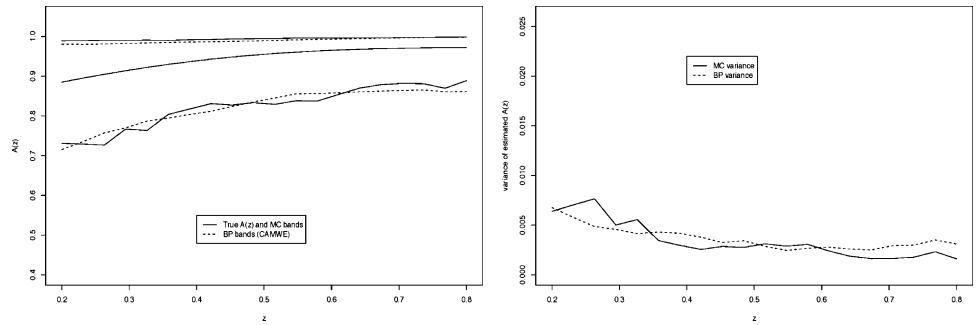

FIGURE 6: Simulation results with the same setting as Figure 4, except that the model has a lognormal error.

models with moderate sample sizes $n = m = 40$, the comparisons between the Monte Carlo averages of the bootstrap bands and the Monte Carlo bands. In the right panels, similar comparisons were shown for the averaged bootstrap variance estimates of $\widehat{A}_{\mathrm{M}}(z)$ against the Monte Carlo variances. From Figures 4–6, for the CAMWE $\widehat{A}_{\mathrm{M}}(z)$, the averages of confidence bands obtained

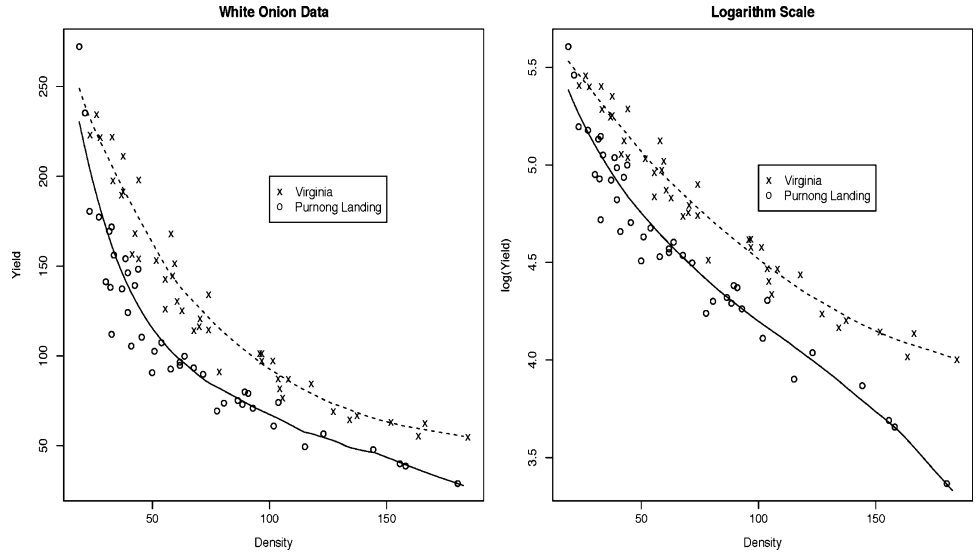

FIGURE 7: Spanish Onion Data with response on the original scale (left) and the logarithmic scale (right), with the smooth estimates of the mean functions for two populations, Pumong Landing (solid) and Virginia (dashed).

by "bootstrapping the data" approximate well the 95% pointwise Monte Carlo bands, and the same can be said about the averages of bootstrap variance estimates. This provides some empirical evidence for using the bootstrap confidence bands and variance estimates for the CAMWE $\widehat{A}_{\mathrm{M}}(z)$ in models without normality assumption. For the normal error model, we have done similar comparisons and the results are almost identical to those obtained for $\widehat{A}_{\mathrm{M}}(z)$ (thus not reported for brevity).

4.2. Example: Spanish Onion Growth—Yield and Density

We consider the white onions data originally reported by Ratkowski (1983) on the density-yield relationship of varieties of white Spanish Onion grown in various regions of Australia. These data have been the subject of a nonparametric analysis of covariance in Young & Bowman (1995). One can see from Figure 7 that the relationship between the density and yield is nonlinear for the two regions considered here: Virginia and Purnong Landing. A question of interest is whether the two regions of origin for the onions can be separated simply by looking at the yield. Figure 7 shows that the difference between yields depends on the density which will be the covariate under consideration in our study.

If we apply directly the method of Faraggi (2003) to the data on the original scale we observe a large discrepancy between the parametric and nonparametric analyses, as illustrated by the left panel in Figure 8. We also notice that bootstrapping the data produces wider 95% confidence bands for large values of the density due to the sparseness and high variability. But even such confidence bands do not cover the parametric estimator of the AUC. We should note that due to the sparseness of observations with densities larger than 150 we focus on the covariate range (0, 150). On the logarithmic scale, the relationship between yield and density is more linear as can be seen from the right panel in Figure 7. In addition, the transformation seems to stabilize the variance so it is not unexpected that he difference between the nonparametric approach and the parametric one diminishes. We can also notice that, on both original and logarithmic scales, the estimates obtained under the normal assumption are more conservative indicat-

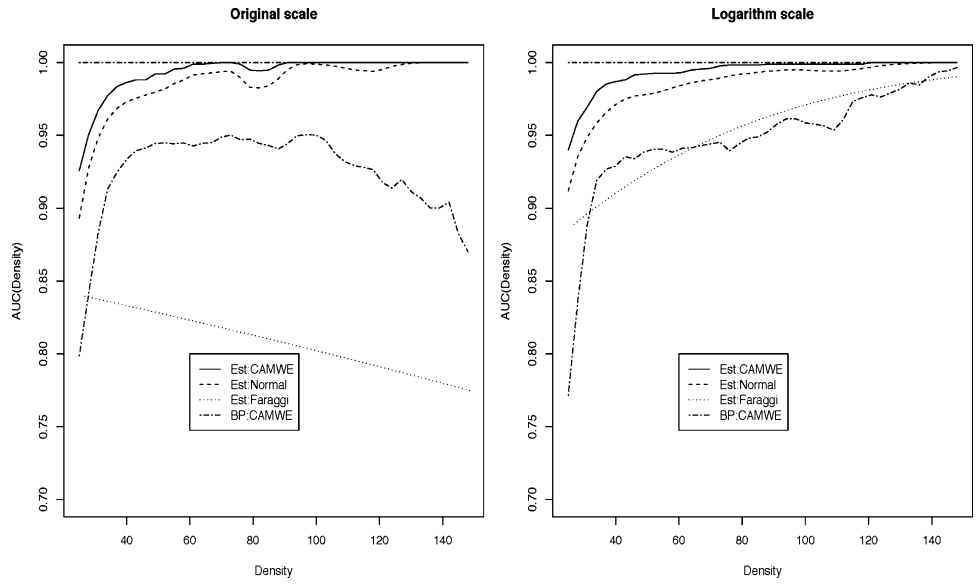

FIGURE 8: Left panel: Comparison of estimated functional relationship between AUC and density obtained using the nonparametric approach with and without normal error assumption, denoted by Normal and CAMWE, respectively, with the parametric estimate following Faraggi (2003). Also shown are the 95% pointwise confidence bands obtained from nonparametric Bootstrap method. Right panel: Same comparison as in the top panel with response on the logarithmic scale.

ing a smaller AUC for small densities. This indicates that the normal assumptions may not be valid for this dataset and that the nonparametric approach is more suitable due to its robustness.

5. CONCLUSIONS

We introduce nonparametric adjustment for covariate information in the context of ROC analysis, more specifically for the AUC index. The essential idea in our proposal is that the conditional ROC curve and all the indexes associated with it (e.g., Youden Index (YI) and its optimal cutoff value) can be computed using the statistical model and, subsequently, the reconstructed *working sample*. The theoretical properties of the index estimators deserve further investigation. The approach bears some similarity to the work on nonparametric adjustment for covariates when estimating a treatment effect as in Young & Bowman (1995) and Cantoni & de Luna (2006) and advances in that field are likely to yield newer results for the ROC covariate adjustment. In contrast to their work we focus on a generalized Mann–Whitney approach. Our simulations demonstrate effectiveness and robustness of the proposed method. While the discussion is limited to the case of only one covariate, the proposed approach can be extended to multiple covariates in various ways (e.g.; additive models). It is expected that the computational load will significantly increase with each additional covariate added to the model. In principle one may consider reasonable parametric approximations suggested by nonparametric approaches that lead to simpler interpretations. For instance, one possibility is to use parametric models for the mean and variance functions following the nonparametrically estimated forms. Similar strategy applies to approximating the empirical c.d.f. of the noise by parametric functions.

APPENDIX

Recall that $\{(z_{i,x}, x_i)\}_{1\le i\le m}$ and $\{(z_{j,y}, y_j)\}_{1\le j\le n}$ are nondiseased and diseased samples. The local polynomial regression estimator of $\mu_1(z)$ is obtained by minimizing

$$\sum_{i=1}^{m}\left\{x_i-\sum_{k=0}^{p}\beta_k(z_{i,x}-z)^k\right\}^2 K_{h_1}(z_{i,x}-z), \tag{24}$$

where $h_1 = h_1(m)$ is the bandwidth controlling the amount of smoothing, and $K_{h_1}(\cdot) = K(\cdot/h_1)/h_1$. It is more convenient to work with matrix notation. Denote the design matrix of (24) by Z_x,

$$Z_x = \begin{pmatrix} 1 & (z_{1,x}-z) & \cdots & (z_{1,x}-z)^p \\ \vdots & \vdots & & \vdots \\ 1 & (z_{m,x}-z) & \cdots & (z_{m,x}-z)^p \end{pmatrix},$$

and put $W_{x,h_1} = \mathrm{diag}\{K_{h_1}(z_{i,x}-z) : i = 1,\ldots,m\}$ and $\boldsymbol{x} = (x_1,\ldots,x_m)^{\mathrm{T}}$. The local polynomial estimator is then given by

$$\hat{\mu}_1(z) = \boldsymbol{e}_1^{\mathrm{T}}(Z_x^{\mathrm{T}}W_{x,h_1}Z_x)^{-1}Z_xW_{x,h_1}\boldsymbol{x}. \tag{25}$$

Analogously for the diseased sample $(z_{j,y}, y_j)$, $j = 1,\ldots,n$, the design matrix Z_y and weight matrix W_{y,h_2} are similarly defined, letting $\boldsymbol{y} = (y_1,\ldots,y)^{\mathrm{T}}$, then the local polynomial estimator for μ_2 is $\hat{\mu}_2(z) = \boldsymbol{e}_1^{\mathrm{T}}(Z_y^{\mathrm{T}}W_{y,h_2}Z_y)^{-1}Z_yW_{y,h_2}\boldsymbol{y}$.

We next estimate the variance functions $v_1(z)$ and $v_2(z)$ for heteroscedastic errors according to models (1) and (2). The nonparametric estimators $\hat{v}_1(z)$ and $\hat{v}_2(z)$ are obtained by fitting local polynomial regression to the squared residuals, that is, the variance observations $v_{i,x}$ and $v_{j,y}$ as in (12). Let $b_1 = b_1(m)$ and $b_2 = b_2(n)$ be the sequences of bandwidths for $\hat{v}_1(z)$ and $\hat{v}_2(z)$. Denote $\boldsymbol{v}_x = (v_{1,x},\ldots,v_{m,x})^{\mathrm{T}}$ and $\boldsymbol{v}_y = (v_{1,y},\ldots,v_{n,y})^{\mathrm{T}}$, we have

$$\hat{v}_1(z) = \boldsymbol{e}_1^{\mathrm{T}}(Z_x^{\mathrm{T}}W_{x,b_1}Z_x)^{-1}Z_xW_{x,b_1}\boldsymbol{v}_x, \qquad \hat{v}_2(z) = \boldsymbol{e}_1^{\mathrm{T}}(Z_y^{\mathrm{T}}W_{y,b_2}Z_y)^{-1}Z_yW_{y,b_2}\boldsymbol{v}_y,$$

where Z_x and Z_y are defined as the above, $W_{x,b_1} = \mathrm{diag}\{K_{b_1}(z_{i,x}-z) : i = 1,\ldots,m\}$ and $W_{y,b_2} = \mathrm{diag}\{K_{b_2}(z_{j,y}-z) : j = 1,\ldots,n\}$.

We list below the regularity conditions that are standard in nonparametric smoothing. Recall the notations defined in Section 3.1. Denote by $N(z)$ a neighbourhood of z. Assume that, for a given value z of Z,

(A1) $\theta(z) > 0$ and $\theta(\cdot)$ is continuous in $N(z)$.

(A2) $v_1(z) > 0$, $\mu_1^{(p+1)}(\cdot)$, $v_1^{(p+1)}(\cdot)$, $\eta_1(\cdot)$ and $\kappa_1(\cdot)$ are continuous in $N(z)$.

(A3) $h_1 \to 0, mh_1 \to \infty, mh_1^{2p+3} \to d_1^2$ for some $d_1 > 0$, $b_1/h_1 \to \rho_1$ for some $0 < \rho_1 < \infty$, as $m \to \infty$.

(A4) $v_2(z) > 0$, $\mu_2^{(p+1)}(\cdot)$, $v_2^{(p+1)}(\cdot)$, $\eta_2(\cdot)$ and $\kappa_2(\cdot)$ are continuous in $N(z)$;

(A5) $h_2 \to 0, nh_2 \to \infty, nh_2^{2p+3} \to d_2^2$ for some $d_2 > 0$, and $b_2/h_2 \to \rho_2$ for some $0 < \rho_2 < \infty$.

Denote by $\mathcal{Z}$ the set of possible values of Z (usually an interval on the real line). Additional assumptions below are needed for the uniform convergence results,

(A6) K^* is uniform continuous, absolutely integrable with respect to Lebesgue measure on $\Re$ and of bounded variation, $K^*(u) \to 0$ as $|u| \to \infty$, $\int \{|u \log(|u|)|\}^{1/2} |\, \mathrm{d}K^*(u)| < \infty$.

(A7) $E(|X|^s) < \infty$, $\sup_{z\in\mathcal{Z}} \int |x|^s p_{(Z,X)}(z, x)\, \mathrm{d}x < \infty$ for some $s \geq 2$, where $p_{(Z,X)}$ is the joint density of (Z, X).

(A8) $E(|Y|^s) < \infty$, $\sup_{z\in\mathcal{Z}} \int |y|^s p_{(Z,Y)}(z, y)\, \mathrm{d}y < \infty$ for some $s \geq 2$, where $p_{(Z,Y)}$ is the joint density of (Z, Y).

For the proof of Theorem 2 we need to modify (A1)–(A5) as follows. For convenience we impose conditions on the equivalent kernel K^* (13) instead of the original kernel K.

(A1†) $\theta(\cdot) > 0$, and $\theta^{(p+1)(\cdot)}$ is bounded and continuous on $\mathcal{Z}$.

(A2†) On the domain $\mathcal{Z}$, $v_1(\cdot) > \delta_1$ for some $\delta_1 > 0$ and is bounded, $\mu_1(\cdot)$ is bounded, $\mu_1^{(p+1)}(\cdot)$, $v_1^{(p+1)}(\cdot)$, $\eta_1(\cdot)$, and $\kappa_1(\cdot)$ are bounded and continuous.

(A3†) $\sum_m h_1^{\Delta_1} < \infty$ for some $\Delta_1 > 0$, $m^{2\rho_1-1}h_1 \to \infty$ for some $\rho_1 < 1 - s^{-1}$, where $s > 2$ satisfies (A7.1).

(A4†) On the domain $\mathcal{Z}$, $v_2(\cdot) > \delta_2$ for some $\delta_2 > 0$ and is bounded, $\mu_2(\cdot)$ is bounded, $\mu_2^{(p+1)}(\cdot)$, $v_2^{(p+1)}(\cdot)$, $\eta_2(\cdot)$, and $\kappa_2(\cdot)$ are bounded and continuous.

(A5†) $\sum_n h_2^{\Delta_2} < \infty$ for some $\Delta_2 > 0$, $n^{2\rho_2-1}h_2 \to \infty$ for some $\rho_2 < 1 - s^{-1}$, where $s > 2$ satisfies (A7.2).

For establishing Theorem 4, as mentioned in Section 3.1, the regularity conditions (A3†) and (A5†) can be relaxed as follows, with an additional assumption (A9).

(A3*) $h_1 \to 0$, $m^{\rho_1}h_1 \to \infty$ for some $\rho_1 < 1 - s^{-1}$, where s satisfies (A7.1).

(A5*) $h_2 \to 0$, $n^{\rho_2}h_2 \to \infty$ for some $\rho_2 < 1 - s^{-1}$, where s satisfies (A7.2).

(A9) $F^*(\cdot)$ and $G^*(\cdot)$ are continuous on their domains.

Due to limited space we provide two lemmas with proofs in Supplementary file available at http://fisher.utstat.toronto.edu/fyao/publication.html, where Lemma 1 states the asymptotic normality of $\sqrt{mh_1}\{\hat{\mu}_1(z) - \mu_1(z), \hat{v}_1(z) - v_1(z)\}$ and $\sqrt{mh_1}\{\hat{\mu}_1(z) - \mu_1(z), \hat{v}_1 - v_1(z)\}$ and $\sqrt{mh_2}\{\hat{\mu}_2(z) - \mu_2(z), \hat{v}_2(z) - v_2(z)\}$ and Lemma 2 states the uniform consistency of $\hat{\mu}_1$, $\hat{v}_1$, $\hat{\mu}_2$ and $\hat{v}_2$ (Mack and Silverman, 1982; Horng, 2006). Then Theorem 1 and 2 follows immediate by applying the Cramér–Wold device and Slustsky's theorem to $\widehat{A}_{\mathrm{N}}(z)$, respectively, where the proofs are straightforward and thus not shown.

Proof of Theorem 3. For a given $Z = z$, one can see that "hypothetical" estimator $A_{\mathrm{M}}(z)$ is in fact a two-sample U-statistic. The argument used in the theory of U-statistics can be applied here. The unbiasedness of $A_{\mathrm{M}}(z)$ is obvious. For the asymptotic variance at a given z, put $h(X, Y; z) = 1_{[0,\infty)}(Y - X|Z = z) - A(z)$, $h^*_{0,0} = E\{h(X, Y; z)\} \equiv 0$. $h^*_{1,0}(X; z) = E\{h(X, Y; z)|X\}$, $h^*_{0,1}(Y; z) = E\{h(X, Y; z)|Y\}$. Note that

$$
\begin{aligned}
h^*_{0,1}(Y; z) &= P(Y \geq X|Y, Z = z) \\
&= P\left(\mu_1(z) + \epsilon_1\sqrt{v_1(z)} \leq \mu_2(z) + \epsilon_2\sqrt{v_2(z)} \,\middle|\, \epsilon_2\right) \\
&= P\left(\epsilon_1 \leq \sqrt{\frac{v_2(z)}{v_1(z)}}\epsilon_2 + \frac{\mu_2(z) - \mu_1(z)}{\sqrt{v_1(z)}} \,\middle|\, \epsilon_2\right) \equiv h_{1,0}(\epsilon_2; z),
\end{aligned}
$$

and similarly $h^*_{1,0}(X;z) \equiv h_{1,0}(\epsilon_1;z)$, that is, $\xi^2_{1,0} \equiv \mathrm{var}\{h^*_{1,0}(Y;z)\}$, $\xi^2_{0,1} \equiv \mathrm{var}\{h^*_{0,1}(Y;z)\}$ as specified in Section 3.2. The unbiasedness of $A_{\mathrm{M}}(z)$ is obvious from $h^*_{0,0} \equiv 0$. For the variance calculation, after some counting techniques, one has,

$$\mathrm{var}\{A_{\mathrm{M}}(z)\} = \frac{1}{mn}\sum_{c=0,1}\sum_{d=0,1} C^1_c C^{m-1}_{1-c} C^1_d C^{n-1}_{1-d}\xi_{c,d} = \frac{\xi^2_{1,0}(z)}{m} + \frac{\xi^2_{0,1}(z)}{n} + o\left(\frac{1}{m+n}\right), \quad (26)$$

where C^n_k is the combination of choosing k from n. This proves (18).

To show the asymptotic normality (19), define

$$T_{m,n}(z) = \sqrt{m+n}\left\{\frac{1}{m}\sum_{i=1}^{m} h^*_{1,0}(x_{i,z}) + \frac{1}{n}\sum_{j=1}^{n} h^*_{0,j}(y_{j,z})\right\},$$

which is in fact the projection of $\sqrt{m+n}\{A_{\mathrm{M}}(z) - A(z)\}$ on the space formed by random variables of the form of $\{\sum_{i=1}^m \psi(x_{i,z}) + \sum_{j=1}^n \psi^*(y_{i,z})\}$, where ψ and ψ^* are arbitrary measurable functions. From Hájek's Projection Theorem and (26), we have, as $m, n \to \infty$,

$$\mathrm{var}\{\sqrt{m+n}A_{\mathrm{M}}(z) - T_{m,n}(z)\} = \mathrm{var}\{\sqrt{m+n}A_{\mathrm{M}}(z)\} - \mathrm{var}\{T_{m,n}(z)\} \longrightarrow 0,$$

which, together with unbiasedness, implies that $\sqrt{m+n}\{A_{\mathrm{M}}(z) - A(z)\}$ is asymptotically equivalent to $T_{m,n}(z)$. Then following central limit theorem, when $n/(m+n) \to \lambda^*$ and $\min\{\xi^2_{1,0}(z), \xi^2_{0,1}(z)\} > 0$, $T_{m,n}(z)$ has the limiting distribution as specified in (19). So does $\sqrt{m+n}\{A_{\mathrm{M}}(z) - A(z)\}$. ■

Proof of Theorem 4. Define $w_{ij} = y_{i,z} - x_{i,z}$ and $\hat{w}_{ij} = \hat{y}_{i,z} - \hat{x}_{i,z}$, and the dependences of w_{ij} and $\hat{w}_{ij}$ on $x_{i,z}, y_{j,z}, z_{i,x}, z_{j,y}$ and z are suppressed for simplicity. Let $a_1(z) = \mu_2(z) - \mu_1(z)$, $a_2(z_{j,y}, z) = \sqrt{v_2(z)/v_2(z_{j,y}, z)}$, $a_3(z_{i,x}, z) = -\sqrt{v_1(z)/v_1(z_{i,x})}$, $a_4(z_{j,y}, z) = -\mu_2(z_{j,y})a_2(z_{j,y}, z)$, $a_5(z_{i,x}, z) = -\mu_1(z_{i,x})a_3(z_{i,x}, z)$, and then

$$w_{ij} = a_1(z) + a_2(z_{j,y}, z)y_j + a_3(z_{i,x}, z)x_i + a_4(z_{j,y}, z) + a_5(z_{i,x}, z),$$
$$\hat{w}_{ij} = \hat{a}_1(z) + \hat{a}_2(z_{j,y}, z)y_j + \hat{a}_3(z_{i,x}, z)x_i + \hat{a}_4(z_{j,y}, z) + \hat{a}_5(z_{i,x}, z),$$

where " $\hat{}$ " is the generic notation for estimated quantities. By analogy to the proof of Lemma 2 with the assumptions (A3†) and (A5†) replaced by (A3*) and (A5*), we obtain weak (in probability) uniform consistency of $\hat{\mu}_1$, $\hat{\mu}_2$, $\hat{v}_1$, and $\hat{v}_2$. This is sufficient for our purpose, the reason of which will be singled out below. Again by analogy to the proof of Theorem 2 with Slutsky's Theorem (in probability instead of almost sure), we have, for a given z, $\hat{a}_1(z) \xrightarrow{p} a_1(z)$, $\sup_{z_{j,y}} |\hat{a}_k(z_{j,y}, z) - a_k(z_{j,y}, z)| = o_p(1)$, $\sup_{z_{i,x}} |\hat{a}_l(z_{i,x}, z) - a_l(z_{i,x}, z)| = o_p(1)$, for $k = 2, 4$ and $l = 3, 5$. Since $\epsilon_{1,i} \overset{\text{i.i.d.}}{\sim} F^*$, one has $\epsilon_{1,i} = O_p(1)$ and, analogously, $\epsilon_{2,j} = O_p(1)$, regardless of i and j. Also note that μ_1, μ_2, v_1 and v_2 are bounded on $\mathcal{Z}$, then we obtain $\sup_{i,j,z_{i,x},z_{j,y}} |\hat{w}_{ij} - w_{ij}| = o_p(1)$ that only depends on the given z. We observe that $E[\{\widehat{A}_{\mathrm{M}}(z) - A_{\mathrm{M}}(z)\}^2] = E_{0,0} + E_{1,0} + E_{0,1} + E_{1,1}$, where

$$E_{0,0} = \frac{1}{m^2n^2}\sum_{i\neq i', j\neq j'} E\left[\{1_{[0,\infty)}(\hat{w}_{ij}) - 1_{[0,\infty)}(w_{ij})\}\{1_{[0,\infty)}(\hat{w}_{i'j'}) - 1_{[0,\infty)}(w_{i'j'})\}\right],$$

while $E_{1,0}$, $E_{0,1}$, and $E_{1,1}$ are defined in the same way, with $E_{1,0}$ corresponds to $\sum_{i=i',j\neq j'}$, $E_{0,1}$ to $\sum_{i\neq i',j=j'}$ and $E_{1,1}$ to $\sum_{i=i',j=j'}$. We first focus on $E_{0,0}$,

$$\begin{aligned} E_{0,0} = \frac{1}{m^2n^2} \sum_{i\neq i',j\neq j'} \big\{ & P(\hat{w}_{ij} \geq 0, \hat{w}_{i'j'} \geq 0) + P(w_{ij} \geq 0, w_{i'j'} \geq 0) \\ & - P(\hat{w}_{ij} \geq 0, w_{i'j'} \geq 0) - P(w_{ij} \geq 0, \hat{w}_{i'j'} \geq 0) \big\} \\ \leq \sup_{i,i',j,j'} \big| & P(\hat{w}_{ij} \geq 0, \hat{w}_{i'j'} \geq 0) + P(w_{ij} \geq 0, w_{i'j'} \geq 0) \\ & - P(\hat{w}_{ij} \geq 0, w_{i'j'} \geq 0) - P(w_{ij} \geq 0, \hat{w}_{i'j'} \geq 0) \big| . \end{aligned} \tag{27}$$

For any given z, from Slutsky's Theorem, we have $(\hat{w}_{ij}, \hat{w}_{i'j'})^{\mathrm{T}}$, $(\hat{w}_{ij}, w_{i'j'})^{\mathrm{T}}$ and $(w_{ij}, \hat{w}_{i'j'})^{\mathrm{T}}$ converge in probability to $(w_{ij}, w_{i'j'})^{\mathrm{T}}$ uniformly in all arguments except z, which implies uniform convergence in distribution. Therefore the four sequences of probabilities in (27) all uniformly converge to $P(w_{ij} \geq 0, w_{i'j'} \geq 0)$ as $m, n \to \infty$, which leads to $E_{0,0} \to 0$. From the above argument, one can see that the weak uniform consistency is sufficient, also that the convergence rates cannot be preserved for evaluating upper bounds for those probability differences. Using similar arguments, it is easy to show that $E_{1,0} = O(E_{0,0}/m)$, $E_{0,1} = O(E_{0,0}/n)$ and $E_{1,1} = O\{E_{0,0}/(mn)\}$. ■

ACKNOWLEDGEMENTS

We would like to thank Professor Hans-Georg Müller for constructive remarks. We also thank the Editor and an Associate Editor for helpful suggestions that have improved the paper. Research of RVC and FY is supported by grants from the Natural Sciences and Engineering Research Council of Canada. RVC also acknowledges grant support from the Canadian Institutes of Health Research.

BIBLIOGRAPHY

K. Aoki, J. Misumi, T. Kimura, W. Zhao & T. Xie (1997). Evaluation of cutoff levels for screening of gastric cancer using serum pepsinogens and distributions of levels of serum pepsinogens I, II and of PG I/PG II ratios in a gastric cancer case-control study. *Journal of Epidemiology*, 7, 143–151.

C. B. Begg (1991). Advances in statistical methodology for diagnostic medicine in the 1980's. *Statistics in Medicine*, 10, 1887–1895.

L. C. Brumback, M. S. Pepe & T. A. Alonzo (2006). Using the ROC curve for gauging treatment effect in clinical trials. *Statistics in Medicine*, 25, 575–590.

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

T. Cai & M. S. Pepe (2002). Semiparametric receiver operating characteristic analysis to evaluate biomarkers for disease. *Journal of the American Statistical Association*, 97, 1099–1107.

E. Cantoni & X. de Luna (2006). Non-parametric adjustment for covariates when estimating a treatment effect. *Nonparametric Statistics*, 18, 227–244.

E. R. Delong, D. M. Delong & D. L. Clarke-Pearson (1988). Comparing the area under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*, 44, 837–844.

L. Dodd & M. S. Pepe (2003). Semi-parametric regression for the area under the Receiver Operating Characteristic Curve. *Journal of the American Statistical Association*, 98, 409–417.

B. Efron & R. J. Tibshirani (1993). "*An Introduction to the Bootstrap*", Chapman & Hall, New York.

J. Fan & I. Gijbels (1996). "*Local Polynomial Modelling and Its Applications*", Chapman & Hall, London.

D. Faraggi (2000). The effect of random measurement error on receiver operating characteristic (ROC) curves. *Statistics in Medicine*, 19, 61–70.

D. Faraggi (2003). Adjusting receiver operating curves and related indices for covariates. *The Statistician*, 52, 179–192.

I. Guttman, R. A. Johnson, G. K. Bhattacharyya & B. Reiser (1988). Confidence limits for stress-strength models with explanatory variables. *Technometrics*, 30, 161–168.

P. Hall & J. S. Marron (1990). On variance estimation in nonparametric regression. *Biometrika*, 77, 415–419.

P. Hall, J. W. Kay & D. M. Titterington (1990). Asymptotically optimal difference-based estimation of variance in nonparametric regression. *Biometrika*, 77, 521–528.

J. A. Hanley & B. J. McNeil (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, 143, 29–36.

W. J. Horng (2006). Uniform consistency and convergence rate in the local polynomial estimation of regression. *Journal of Science and Engineering Technology*, 2, 61–66.

J. K. Hsiao, J. J. Bartko & W. Z. Potter (1989). Diagnosing diagnoses: receiver operating characteristics methods and psychiatry. *Archives of General Psychiatry*, 46, 664–667.

Y. P. Mack & B. W. Silverman (1982). Weak and strong uniform consistency of kernel regression estimates. *Z. Wahrscheinlichkeitstheorie and Verwandte Gebiete*, 61, 405–415.

C. E. Metz (1986). ROC methodology in radiologic imaging.*Investigative Radiology*, 21, 720–733.

K. Molodianovitch, D. Faraggi & B. Reiser (2006). Comparing the areas under two correlated ROC curves: Parametric and non-parametric approaches. *Biometrical Journal*, 48, 745–757.

N. A. Obuchowski (1995). Multireader, multimodality receiver operating characteristic curve studies: Hypothesis testing and sample size estimation using an analysis of variance approach with dependent observations. *Academic Radiology*, 2, 522–529.

M. Otto, J. Wiltfang, E. Schütz, A. Otto, A. Pfahlberg, O. Gefeller, M. Uhr, A. Giese, T. Weber, H. A. Kretzschmar & S. Poser (1998). Diagnosis of Creutzfeldt-Jacob disease by measurement of s100 protein in serum: prospective case-control study. *British Medical Journal*, 316, 577–586.

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

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

M. S. Pepe (2000). An interpretation for the ROC curve and inference using GLM procedures. *Biometrics*, Oxford University Press, New York, 56, 352–359.

M. S. Pepe (2003). *The Statistical Evaluation of Medical Tests for Classification and Prediction*. Oxford Statistical Sciences Series.

D. A. Ratkowsky (1983). "*Nonlinear Regression Modelling (Statistics: Textbooks and Monographs, Vol. 48)*", Marcel Dekker, New York.

B. Reiser & I. Guttman (1986). Statistical-inference for pr(y-less-than-x)—the normal case. *Technometrics*, 28, 253–257.

E. Schisterman, D. Faraggi & B. Reiser (2004). Adjusting the generalized ROC curve for covariates. *Statistics in Medicine*, 23, 3319–3331.

L. Stover, M. Gorga & S. Neely (1996). Toward optimizing the clinical utility of distortion product otoacoustic emission measurements. *Journal of the Acoustical Society of America*, 100, 956–967.

J. A. Swets & R. M. Pickett (1982). "*Evaluation of Diagnostic Systems: Methods from Signal Detection Theory*", Academic Press, New York.

M. L. Thompson & W. Zucchini (1989). On the statistical analysis of ROC curves. *Statistics in Medicine*, 8, 1277–1290.

A. Toledano & C. Gatsonis (1995). Regression analysis of correlated receiver operating characteristic data. *Academic Radiology*, 2, 530–536.

A. N. A. Tosteson & C. B. Begg (1988). A general regression methodology for ROC curve estimation. *Medical Decision Making*, 8, 204–215.

W. J. Youden (1950). Index for rating diagnostic tests. *Cancer*, 3, 32–35.

S. G. Young & A. W. Bowman (1995). Non-parametric analysis of covariance. *Biometrics*, 51, 920–931.

X. H. Zhou, N. A. Obuchowski & D. K. McClish (2002). "*Statistical Methods in Diagnostic Medicine*", Wiley, New York.

Received 21 October 2008
Accepted 10 August 2009