Nonparametric Covariate Adjustment for Receiver Operating Characteristic Curves

by

Fang Yao
Department of Statistics
University of Toronto

and

Radu V. Craiu
Department of Statistics
University of Toronto

and

Benjamin Reiser
Department of Statistics
University of Haifa


TECHNICAL REPORT SERIES

University of Toronto
Department of Statistics
Nonparametric Covariate Adjustment for Receiver Operating Characteristic Curves

FANG YAO
Department of Statistics
University of Toronto

RADU V. CRAIU
Department of Statistics
University of Toronto

BENJAMIN REISER
Department of Statistics
University of Haifa

July, 2007
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. We propose nonparametric methods for covariate adjustment of the AUC. Models with normal errors and non-normal errors are discussed and analyzed separately. Nonparametric regression is used for estimating mean and variance functions in both scenarios. In the general noise distribution case we 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. We derive asymptotic properties for the AUC estimators in both settings, including asymptotic normality, optimal strong uniform convergence rates and $L^2$ consistency. The usefulness of the proposed methods is demonstrated through simulated and real data examples.

KEY WORD: Area Under Curve, Asymptotics, Covariate Adjustment, Mann-Whitney, Nonparametric, Smoothing, Uniform Convergence,
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, respectively positive. If the distribution of \( Y \) is continuous, each value of the threshold \( c \) will correspond to different sensitivity and specificity values. Hanley (1989), among many others, argued that instead of summarizing the performance of the test for a particular \( c \), it is preferable to sum up the properties of the test using the ROC curve, that is the plot of sensitivity against one minus specificity when \( c \) is varied over the whole range. 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). Bamber (1975) has shown that 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. Wolfe and Hogg (1971) have proposed this probability as a measure of the difference between two populations and argued that this is often more meaningful than examining mean differences. Hauck et al. (2000) have emphasized its use as a measure of treatment effects in clinical trials. In addition, this probability naturally arises in reliability theory (Reiser and Cuttman, 1986).

The presence of ROC curves has become ubiquitous in medical studies (Metz, 1989; Hsiao et al., 1989; Aoki et al., 1997; Otto et al., 1998; Stover et al, 1996; Zou et al., 2002), its usage being spurred by the now classic text of Swets and 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 et al., 1988; Molodanovitch et al., 2006; Pepe, 2003).

In a large number of situations, there is available additional information in the form of covariates which are known to influence the accuracy of the test. Only recently, statistical methods have been devised for incorporating such information in the ROC-based analysis. Thompson and Zucchi (1989) and Obuchowski (1995) proposed calculating the ROC curve and its summary statistic (e.g. AUC) for a number of distinct combinations of covariates and then applied a general regression model. Tosteson and Begg (1988) and Toledano and Gatsonis (1995) proposed modeling the distribution of the test result $Y$ in the diseased and nondiseased populations using a latent variable ordinal regression model. Pepe (1997) formulated a general regression framework to model the dependence of the ROC curve directly on the covariates. Pepe (2000) and Dodd and Pepe (2003) propose semiparametric approaches to model the ROC and AUC directly using generalized linear models. Brumback et al. (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. 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 modeling process obtain the analyses of interest.

The proposed approaches fall within the first category of methods described in Pepe (1998). In the present paper we generalize in two directions the approaches of Faraggi (2003) and Schisterman et al. (2006) who use normal regression models to adjust the index AUC for covariates. Their approach is based on Guttman et al. (1988) who discussed covariate adjustment in a reliability context. In Section 2 we describe the regression model, distinguishing between the normal noise assumption and the general noise assumption. In both scenarios, we estimate the mean and variance functions using nonparametric regression techniques, more specifically, local polynomial regression (discussed in Section 3.1 and
procedures provided in Appendix A.1) instead of parametric linear models. This is the first direction of our generalization and provides flexibility for possibly nonlinear patterns and heteroscedastic errors. In the latter scenario we propose a covariate-adjusted Mann-Whitney estimator (CAMWE) for a general noise distribution in Section 2.3. This is the second direction of our generalization and a critical methodological contribution of this paper. The proposed CAMWE constructs working samples at any possible value \( Z = z \) of interest for the estimation of AUC. Such working samples have, for any \( Z = z \), the same sizes as the original samples, so that we can most efficiently use the available data. In practice the computation is kept minimal by utilizing the estimated mean and variance functions for all \( Z = z \) of interest. Confidence bands for the functional dependence between the covariate(s) and the AUC are obtained via bootstrap based on the proposed methods as described in Section 3.2.

A theoretical investigation provides asymptotic results for both the normal noise and general noise models. The asymptotic normality and optimal strong uniform convergence rates for the covariate-adjusted AUC estimators for normal noise are presented in Section 4.1. For the general noise distribution we first derive asymptotic normality of the "hypothetical" CAMWE and then characterize the asymptotic behavior of the Mean Squared Error (MSE) of the CAWME. Although we focus on covariate-adjusted AUC estimation, the proposed methods and corresponding theory can be readily extended to other measures of ROC curves, e.g., the covariate-adjusted specificity and sensitivity. We illustrate the performance of the proposed methods with simulated examples in Section 5.1 and a real data application in Section 5.2. Conclusions are offered in Section 6. Fitting procedures, auxiliary results and technical proofs are deferred to the Appendix.
2. MODEL AND ESTIMATION

2.1 Regression Model

Consider the test response variable for nondiseased individuals $X$ and for diseased individuals $Y$. We assume nonparametric regression models

\begin{align}
X|Z &= f(Z) + \sqrt{v_1(Z)} \epsilon_1, \\
Y|Z &= g(Z) + \sqrt{v_1(Z)} \epsilon_2,
\end{align}

where $Z$ denotes the covariate, the standardized errors $\epsilon_1$ and $\epsilon_2$ 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 $z \in \mathbb{R}$. Note that the errors here can depend heteroscedastically on the covariate $Z$ through $v_1$ and $v_2$, leading to more general and realistic model specifications. For the diseased and nondiseased populations, the covariate $Z$ is related to the marker values $Y$ and $X$ through the functional relationship $f(Z)$ and $g(Z)$. In general, 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 $\epsilon_1$ and $\epsilon_2$ by $F^*(\cdot)$ and $G^*(\cdot)$. Here we assume $F^*$ and $G^*$ do not depend on $Z$, i.e., the dependence of $X$ and $Y$ on $Z$ are only through $f$, $g$, $v_1$ and $v_2$. Although we refer to “diseased” and “nondiseased” groups, the above framework applies to any two populations of interest.

2.2 Estimation under Normal Noise 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 $\epsilon_1$ and $\epsilon_2$ are normally distributed, i.e., $F^* = G^* = \Phi$, where $\Phi(\cdot)$ is the cumulative distribution function (c.d.f.) of standard normal, it is straightforward to derive the explicit expression as follows,

\begin{align}
A_N(z) = P(Y > X|Z = z) = \Phi \left\{ \frac{g(z) - f(z)}{\sqrt{v_1(z) + v_2(z)}} \right\},
\end{align}

(3)
where the subscript "N" stands for normal assumption. One can also obtain closed forms of the sensitivity \(q_N(z)\) and specificity \(p_N(z)\) for \(Z = z\),

\[
q_N(z) = \Phi \left( \frac{g(z) - c}{\sqrt{v_2(z)}} \right), \quad p_N(z) = \Phi \left( \frac{c - f(z)}{\sqrt{v_1(z)}} \right),
\]

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

\[
q_N(z) = \Phi \left[ \frac{g(z) - f(z) + \sqrt{v_1(z)} \Phi^{-1}(1 - p(z))}{\sqrt{v_2(z)}} \right],
\]

The unknown functions in the foregoing expressions, \(f, g, v_1, v_2\), are estimated by using nonparametric smoothing as addressed in Section 3.1, providing a "nonparametric adjustment" in the first direction discussed in Section 1.

### 2.3 Estimation under General Noise Assumption

The assumption of normal noise above simplifies the calculations of the AUC via (3) as well as other quantities, 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 a general noise distribution, which constitutes the other aspect of the "nonparametric adjustment".

The proposed estimator is motivated by the traditional Mann-Whitney statistic, which is generally formulated for two samples \(\{x_1, \ldots, x_m\}\) and \(\{y_1, \ldots, y_n\}\) as follows,

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

where \(1_{[0,\infty)}(x) = 1\) if \(x \geq 0\) and \(1_{[0,\infty)}(x) = 0\) otherwise, i.e., the indicator function on \([0, \infty)\). 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 the diseased or nondiseased groups. To estimate $A(z)$ at $Z = z$, one possibility is to only include the marker values $x_i$ and $y_j$ that fall into a neighborhood or local window of $z$, say $N(z)$,

$$A_L(z) = \sum_{z_{i,x} \in N(z)} \sum_{z_{j,y} \in N(z)} 1_{[0,\infty)}(y_j - x_i) / \left\{ \sum_{i=1}^{m} 1_{N(z)}(z_{i,x}) \sum_{j=1}^{n} 1_{N(z)}(z_{j,y}) \right\}$$

(7)

However, this estimator does not efficiently use the available data due to the restriction on the neighborhood $N(z)$, nor do the regression models (1) and (2) play any role here. It is also computationally intensive, as it requires choosing an appropriate $N(z)$ for each $z$ of interest.

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

$$\epsilon_{i,x} = \frac{x_i - f(z_{i,x})}{\sqrt{v_1(z_{i,x})}}, \quad \epsilon_{j,y} = \frac{y_j - f(z_{j,y})}{\sqrt{v_2(z_{j,y})}}.$$  

(8)

Recall that the distributions of $\epsilon_1$ and $\epsilon_2$ 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 $j = 1, \ldots, n$. Therefore we can construct working samples $\{x_{i,x}, \ldots, x_{m,x}\}$ and $\{y_{1,z}, \ldots, y_{n,z}\}$ for $Z = z$, as if they were all observed at $Z = z$,

$$x_{i,z} = f(z) + \sqrt{v_1(z)}\epsilon_{i,x}, \quad y_{j,z} = g(z) + \sqrt{v_2(z)}\epsilon_{j,y}.$$  

(9)

Now we have the working samples for any $Z = z$ that have the identical distributions and as much information as the original samples. Then it is intuitive to construct the proposed Covariate-Adjusted Mann-Whitney Estimator (CAMWE) for $A(z)$,

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

(10)

This can be seen when we consider the situation of no covariate effect in which case $f$, $g$, $v_1$, $v_2$ are constant in $z$ and (10) becomes the traditional Mann-Whitney statistic. For practical implementation, after obtaining nonparametric estimates of the unknown functions $f$, $g$, $v_1$.
and \(v_2\), we do not have to choose other tuning parameters for each covariate value \(Z = z\), while (7) requires this. Analogously we can calculate the sensitivity and specificity from the working samples for \(Z = z\),

\[
q_M(z) = \frac{1}{n} \sum_{j=1}^{n} 1_{[0, \infty)}(y_j, z \geq c), \quad p_M(z) = \frac{1}{m} \sum_{i=1}^{m} 1_{[0, \infty)}(x_{i,z} \leq c),
\]

for a given threshold \(c\). The ROC curves for \(Z = z\) can be obtained by plotting \(q_M(z)\) versus \(1 - p_M(z)\) for all possible values of \(c\).

3. NONPARAMETRIC SMOOTHING AND BOOTSTRAP CONFIDENCE LIMITS

3.1 Nonparametric Smoothing Procedures

Although standard parametric models will find features in the data which have been already incorporated \textit{a priori}, these models may not be adequate if the "true" patterns of \(f\), \(g\), \(v_1\) and \(v_2\) are not well defined and do not fall into a preconceived class of functions. In such situations an analysis through nonparametric regression is advisable, since it is a more flexible and data-adaptive way to characterize the functional relationship.

We exploit the local polynomial regression models for estimating the functions \(f\) and \(g\). 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 \(f\), and \(h_2 = h_2(n)\) a sequence of bandwidths for \(g\). Let \(p\) be the order of local polynomial fit, e.g., \(p = 0\) and \(p = 1\) correspond to local constant and local linear fits, respectively. An odd order fit is often suggested (Fan and Cuijbers, 1996) from 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 \(p\)th order local polynomial estimators of \(f(z)\) and \(g(z)\) by \(\hat{f}(z)\) and \(\hat{g}(z)\). 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 \(v_{j,y}\), \(i = 1, \ldots, m, j = 1, \ldots, n\),

\[
v_{i,x} = (x_i - \hat{f}(z_{i,x}))^2, \quad v_{j,y} = (y_j - \hat{g}(z_{j,y}))^2,
\]

\(12\)
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 Appendix A.1. 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 and Marron, 1990; Hall et al., 1990; among others). The theoretical properties developed in Section 4 can be easily carried over 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, while other existing techniques can certainly be applied. For theoretical developments with the sample sizes tending to infinity, the magnitude of the bandwidths in terms of the sample sizes that yield optimal asymptotic properties are appropriately assumed in Section 4. The bandwidths chosen by cross-validation or other established approaches usually fulfill such assumptions for sufficiently large sample sizes.

Substituting the local polynomial estimators \( \hat{f}(z), \hat{g}(z), \hat{\theta}_1(z) \) and \( \hat{\theta}_2(z) \) for these unknown quantities in formulae (3)-(5), (10) and (11) provides the point estimators \( \hat{A}_N(z), \hat{p}_N(z), \hat{q}_N(z), \hat{A}_M(z), \hat{p}_M(z) \) and \( \hat{q}_M(z) \) for covariate \( Z = z \).

3.2 Bootstrap Confidence Limits Based on Proposed Method

For the construction of the confidence limits for AUC under normal noise, the formulae proposed by Guttman et al. (1988) are no longer valid due to the use of local polynomial regression. In principle we can derive the confidence limits for AUC based on the normal assumption, and the asymptotic normality of the local polynomial estimators \( \hat{f}(z), \hat{g}(z), \hat{\theta}_1(z) \) and \( \hat{\theta}_2(z) \) (Fan and Gijbels, 1996) using the Cramér-Wold device. However, the resulting asymptotic bias and variance formulae are prohibitively complicated and will involve unknown functions and their derivatives as well. The evaluation of such asymptotic quantities will require extensive pilot smoothing which might deteriorate the estimation of the confidence limits and not be worth further pursuing. Therefore we prefer to obtain confidence limits for AUC via standard parametric bootstrap (Efron and Tibishirani, 1993) for the normal noise case.

For the general noise situation, in order to construct confidence bands for the estimated
AUC, we would like to present two bootstrap approaches based on our proposed CAMWB (10), while other bootstrap-based alternatives are also possible. The first approach makes use of the estimated mean and variance functions from the original data and only resamples the estimated standardized residuals as described below.

1. Sample with replacement from the estimated standardized residuals \( \{ \hat{e}_{i,x} : i = 1, \ldots, m \} \)
   and \( \{ \hat{e}_{j,y} : j = 1, \ldots, n \} \) to form bootstrap sets \( \{ \hat{e}_{i,x}^{(b)} : i = 1, \ldots, m \} \) and \( \{ \hat{e}_{j,y}^{(b)} : j = 1, \ldots, n \} \), respectively, where \( \hat{e}_{i,x} \) and \( \hat{e}_{j,y} \) are obtained by plugging in the estimated quantities of (8).

2. Using the estimated mean and variance functions from the observed data, construct the bootstrapped working samples at covariate value \( Z = z \),
   
   \[
   \hat{x}_{i,x}^{(b)} = \hat{f}(z) + \hat{e}_{i,x}^{(b)} \sqrt{\hat{v}_1(z)}, \quad \hat{y}_{j,y}^{(b)} = \hat{g}(z) + \hat{e}_{j,y}^{(b)} \sqrt{\hat{v}_2(z)}, \quad i = 1, \ldots, m, \quad j = 1, \ldots, n.
   \]

3. Estimate \( A^{(b)}(z) \) using (10), i.e.,
   
   \[
   \hat{A}^{(b)}_{M}(z) = \frac{1}{mn} \sum_{i=1}^{m} \sum_{j=1}^{n} T_{0,\infty}(\hat{y}_{j,y}^{(b)} - \hat{x}_{i,x}^{(b)}).
   \]

   Then the set \( \{ \hat{A}^{(b)}_{M}(z) : b = 1, \ldots, B \} \) is used to obtain confidence limits for \( \hat{A}(z) \).

The second approach is in the same vein as "bootstrapping the original data" as presented by Efron and Tibshirani (1993), and estimates all the unknown functions and \( A(z) \) using the bootstrapped data. We do not repeat the procedure here for conciseness. Our experience shows that while the second approach is computationally more demanding, the reliability of the first approach can be compromised if the mean and variance functions are poorly estimated. Our simulation studies and data example show that the first approach tends to create confidence bands narrower than the second approach in many instances.

4. THEORETICAL PROPERTIES

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 general noise
assumptions. One can easily extend exactly similar 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.

4.1 Asymptotic Properties for Estimation under Normal Noise

We begin with the asymptotic normality of the estimated AUC under the normal noise assumption, where the target \( A(z) \) is exactly \( A_N(z) \), i.e., \( A(z) \equiv A_N(z) \). Let \( \theta(z) \) be the density function of the covariate \( Z \) that is treated as a random variable. Denote by \( N(z) \) a neighborhood of \( z \). Assume that, for a given value \( z \) of \( Z \),

\[
(A1) \quad \theta(z) > 0 \text{ and } \theta(\cdot) \text{ is continuous in } N(z).
\]

Put \( \eta_1(z) = E(e_1^2|Z = z) \), \( \eta_2(z) = E(e_2^2|Z = z) \), \( \kappa_1(z) = \text{Var}(e_1^2|Z = z) \) and \( \kappa_2(z) = \text{Var}(e_2^2|Z = z) \). Assume that, for a given \( z \),

\[
(A2) \quad v_1(z) > 0, f^{(p+1)}(\cdot), v_1^{(p+1)}(\cdot), \eta_1(\cdot) \text{ and } \kappa_1(\cdot) \text{ are continuous in } N(z).
\]

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 \( f(z) \), \( v_1(z) \), \( g(z) \) and \( v_2(z) \). One can see that, if the bandwidths \( h_1 \) and \( b_1 \) are chosen optimally for estimating \( f(z) \) and \( v_1(z) \), then \( h_1 \) and \( b_1 \) will be of the same order in terms of the sample size \( m \). Thus we assume the following, as \( m \to \infty \),

\[
(A3) \quad h_1 \to 0, nh_1 \to \infty, nh_1^{2p+3} \to d_1^2 \text{ for some } d_1 > 0, b_1/h_1 \to \rho_1 \text{ for some } 0 < \rho_1 < \infty.
\]

Analogously, for the estimation of \( g(z) \) and \( v_2(z) \), we assume that, for a given \( z \),

\[
(A4) \quad v_2(z) > 0, g^{(p+1)}(\cdot), v_2^{(p+1)}(\cdot), \eta_2(\cdot) \text{ and } \kappa_2(\cdot) \text{ are continuous in } N(z);
\]

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

Here we consider the odd order \( p \) of local polynomial estimators for \( f \), \( v_1 \), \( g \) 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 \( p \).
We also need the following assumption and notation for the kernel function \( K(\cdot) \) which is a symmetric density function. Denote \( \mu_j(K) = \int u^j K(u) du \) which exists for all integer \( j \).

\[ R(K) = \int K^2(u) < \infty, \mu_2(K) > 0. \]

For convenience, we introduce the notion of the order of a kernel function. We say \( K_0 \) is an \( \ell \)th order kernel function, provided that \( \mu_j(K_0) = 1, \mu_j(K_0) = 0 \) for \( j = 1, \ldots, \ell - 1 \) and \( \mu_\ell(K_0) \neq 0 \). It is obvious that \( K(\cdot) \) is a 2nd order kernel. Let the \((p + 1) \times (p + 1)\) matrix \( S_p = \{\mu_{j+i}(K)\}_{0 \leq j, i \leq p}, e_k \) be the \((p + 1) \times 1\) vector with the \( k \)th element equal to 1 and 0 elsewhere, and

\[ K^*(u) = e_1^T S_p^{-1}(1, u, \ldots, u^p)^T K(u), \tag{13} \]

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 \( R(K^*, \rho) = \int K^*(u)K^*(u/\rho) du \) for any \( 0 < \rho < \infty \).

Lemma 1 in Appendix A.2 provides the joint asymptotic distributions of the local polynomial estimators of \( \{\hat{f}(z), v_1(z)\}^T \) and \( \{\hat{g}(z), v_2(z)\}^T \), which is the basis for deriving the asymptotic distributions of \( \hat{A}_N(z) \). The difficulty in the proof of Lemma 1 is to deal with the dependence between the mean and variance estimators, though \( \{\hat{f}(z), v_1(z)\}^T \) and \( \{\hat{g}(z), v_2(z)\}^T \) are independent, see Appendix A.2 for details. Based on the joint limiting distributions in Lemma 1, we exploit the Cramér-Wold device to obtain the asymptotic distribution of \( \hat{A}_N(z) \) as follows.

**Theorem 1** Under the assumptions (A1)-(A6) for a given \( z \),

- if \( n/m \to \infty, \sqrt{m} h(z) \{\hat{A}_N(z) - A_N(z)\} \overset{D}{\to} N\{B_1(z), V_1(z)\} \), where \( \phi(u) = (2\pi)^{-1/2}e^{-u^2/2} \),

\[ \delta(z) = \{g(z) - f(z)\}/\sqrt{v_1(z) + v_2(z)}, \]

\[ B_1(z) = -\frac{\phi(\delta(z))\mu_{p+1}(K^*)d_1}{(p + 1)\sqrt{v_1(z) + v_2(z)}} \left[ f(p+1)(z) + \frac{\{g(z) - f(z)\}v_1^{(p+1)}(z)\rho_1^{p+1}}{2\{v_1(z) + v_2(z)\}} \right], \]

\[ V_1(z) = \frac{\phi^2(\delta(z))}{\theta(z)\{v_1(z) + v_2(z)\}} \left[ R(K^*)v_1(z) + \frac{\{g(z) - f(z)\}R(K^*, \rho_1)\eta_1(z)}{\{v_1(z) + v_2(z)\}\rho_1} + \frac{\{g(z) - f(z)\}^2R(K^*)\kappa_1(z)}{4\{v_1(z) + v_2(z)\}^2\rho_1} \right], \tag{14} \]
• if \( n/m \to 0 \), \( \sqrt{mh_2} \{ \bar{A}_N(z) - A_N(z) \} \xrightarrow{D} N\{B_2(z), V_2(z)\} \), where

\[
B_2(z) = \frac{\phi \{ \delta(z) \} \mu_{p+1}(K^*) d_2}{(p + 1)! \sqrt{v_1(z) + v_2(z)}} \left[ g^{(p+1)}(z) - \frac{g(z) - f(z)}{2v_1(z) + v_2(z)} v_2^{(p+1)}(z) \right] \left[ \frac{R(K^*) v_2(z)}{v_1(z) + v_2(z)} - \frac{\{g(z) - f(z)\}^2 R(K^*) \kappa_2(z)}{4\{v_1(z) + v_2(z)\}^2 \rho_2} \right],
\]

\[\tag{15}\]

• if \( n/m \to \lambda \) for some \( 0 < \lambda < \infty \), \( \sqrt{mh_1} \{ \bar{A}_N(z) - A_N(z) \} \xrightarrow{D} N\{B_3(z), V_3(z)\} \), where

\[
B_3(z) = B_1(z) + \lambda^{-\frac{p+1}{2}} B_2(z), \quad V_3(z) = V_1(z) + \lambda^{-\frac{p+1}{2}} V_2(z)
\]

Besides the pointwise results on the limiting distributions, we also establish the optimal rates for strong uniform convergence of \( \bar{A}_N p_N \) and \( q_N \), in Theorem 2. An additional assumption is needed for the uniform results, given by

\[\tag{A7.1} E(|X|^s) < \infty, \sup_{x \in \mathbb{Z}} \int |x|^s p_{(Z,X)}(z, x) dx < \infty \text{ for some } s \geq 2, \text{ where } p_{(Z,X)} \text{ is the joint density of } (Z, X).\]

\[\tag{A7.2} E(|Y|^s) < \infty, \sup_{y \in \mathbb{Z}} \int |y|^s p_{(Z,Y)}(z, y) dy < \infty \text{ for some } s \geq 2, \text{ where } p_{(Z,Y)} \text{ is the joint density of } (Z, Y).\]

For the functions and quantities involved in (A1)-(A6), we need to modify those assumptions for our purpose. Denote the set of all possible values of \( Z \) by \( \mathcal{Z} \) that is usually an interval on the real line.

(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, \( f(\cdot) \) is bounded, \( f^{(p+1)}(\cdot), v_1^{(p+1)}(\cdot), \eta_1(\cdot) \) and \( \kappa_1(\cdot) \) are bounded and continuous.

(A3') \( \sum h_1^{\Delta_1} < \infty \) 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 $Z$, $v_2(\cdot) > \delta_2$ for some $\delta_2 > 0$ and is bounded, $g(\cdot)$ is bounded,
$g^{(p+1)}(\cdot), v_2^{(p+1)}(\cdot)$, $\eta_2(\cdot)$ 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).

We impose conditions on the equivalent kernel $K^*$ (13) instead of the original kernel $K$.

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

Lemma 2 in Appendix A.2 states the strong uniform convergence rates of the local polynomial estimators of the mean and variance functions. Slutsky’s Theorem yields strong
uniform convergence rates of $\hat{A}_N$ as presented below, where a.s. is the abbreviation of “almost
surely”.

**Theorem 2** Under the assumptions (A1')-(A6'), (A7.1) and (A7.2), let $\tau_m = h_1^{p+1} + \sqrt{\log(1/h_1)/(m\overline{h_1})}$ and $\omega_n = h_2^{p+1} + \sqrt{\log(1/h_2)/(n\overline{h_2})}$, then

$$\sup_{z \in \mathbb{Z}} |\hat{A}_N(z) - A_N(z)| = O(\tau_m + \omega_n) \quad \text{a.s.} \quad (17)$$

4.2 Asymptotic Properties for Estimation under General Noise

Now we turn to the asymptotic properties of the covariate-adjusted Mann-Whitney estimator
of $A(z)$ under the general noise assumption. We first state the asymptotic normality of the
“hypothetical” estimator $A_M(z)$ (given in (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 $\epsilon_1$ and $\epsilon_2$, and do not depend on the covariate
$Z$. Define

$$h_{1,0}(\epsilon_1, z) = G^* \left( \sqrt{\frac{v_1(z)}{v_2(z)}} \epsilon_1 + \frac{f(z) - g(z)}{\sqrt{v_2(z)}} \right), \quad h_{0,1}(\epsilon_2, z) = F^* \left( \frac{v_2(z)}{v_1(z)} \epsilon_2 + \frac{g(z) - f(z)}{\sqrt{v_1(z)}} \right).$$

Denote $\xi_1^2(z) = \text{var}(h_{1,0}(\epsilon_1; z))$ and $\xi_0^2(z) = \text{var}(h_{0,1}(\epsilon_2; z))$. 

14
Theorem 3 For the regression models (1) and (2) and a given $z$,

$$E[A_M(z)] = A(z), \quad \text{var}[A_M(z)] = O\left(\frac{1}{m+n}\right).$$  \hspace{1cm} (18)

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

$$\frac{\sqrt{m+n}}{m+n}[A_M(z) - A(z)] \xrightarrow{D} N\left\{0, \frac{\xi_{l,0}^2(z)}{\lambda^*} + \frac{\xi_{0,1}^2(z)}{1-\lambda^*}\right\},$$  \hspace{1cm} (19)

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

In the next theorem we establish the $L^2$ consistency of the covariate-adjusted Mann-Whitney estimator $\widehat{A}_M(z)$ for the “hypothetical” estimator $A_M(z)$ 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 $\widehat{A}(z)$. We only need the mean and variance estimators to be uniformly consistent in probability in Theorem 4. Thus the regularity conditions (A3') and (A5') can be relaxed to the following.

(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).

We also need the following additional assumptions,

(A8) $F^*(\cdot)$ and $G^*(\cdot)$ are continuous on their domains.

Theorem 4 Under (A8) and the assumptions for Theorem 2 with (A3') and (A5') replaced by (A3*) and (A5*), for a given $z$,

$$E[(\widehat{A}_M(z) - A_M(z))^2] \to 0.$$  \hspace{1cm} (20)

We conclude this section with the following corollary that is a direct consequence of Theorem 4 and Theorem 5. Note that the $L^2$ distance between estimated and true AUC at $Z = z$ is dominated by the nonparametric rate in (20) which is usually slower than the parametric rate $(m+n)^{-1/2}$, though its order of magnitude is not obtainable, at least to our knowledge.
Corollary 1. Under (A8) and the assumptions for Theorem 2 with (A3\dag) and (A5\dag) replaced by (A3^*) and (A5^*), for a given $z$,

$$E[(\hat{A}_M(z) - A(z))^2] \to 0.$$  \hspace{1cm} (21)

5. SIMULATIONS AND DATA EXAMPLE

5.1 Simulations

The purpose of the simulations is to assess the accuracy of the method in estimating AUC in nonparametric settings. We have not compared our method with parametric models since the two approaches address different situations. If a parametric model is appropriate, its performance will be superior to a nonparametric procedure which essentially ignores information about the existing structure in the data; however, if there is no known parametric model suitable for the data considered, one will have no choice but to use the nonparametric inferential tools available.

There are infinitely many choices for the mean and variance functions (obviously) so our study is certain to not cover all the scenarios of interest. However, we tried to incorporate in our functions patterns that are not uncommon, at least as far as our experience goes.

We consider for illustration a model in which for non-diseased individuals:

$$X_i = \alpha_0 + \alpha_1 Z_i + \alpha_2 \sin(Z_i) + \epsilon_i$$

where the normal deviate $\epsilon$ has conditional variance $\var(\epsilon_i|Z_i) = \xi_0 + \xi_1 \Phi(\delta_0 + \delta_1 Z_i)$. For diseased individuals we consider the model

$$Y_i = \beta_0 + \beta_1 Z_i + \beta_2 \sin(Z_i) + \beta_3 \sqrt{Z_i - 1} + \eta_i,$$

with $\eta$ Gaussian with conditional variance $\var(\eta_i|Z_i) = \var(\epsilon_i|Z_i) + \gamma$.

In Table 1 we show some representative simulation results. It can be noticed that even with small samples the AUC can be reconstructed quite accurately. However, the "distance" between the mean functions for the non-diseased and the diseased is central to the accuracy of the method. In the normal noise scenario AUC estimators $\hat{A}_N(z)$ and $\hat{A}_M(z)$ are comparable.
and get closer to the true AUC when sample size increases. The bootstrap confidence bands produced by resampling residuals tend to be narrower than those obtained by resampling the data.

An alternative scenario is one in which the noise is not normal. We modify the above scenarios so that the distribution of the noise is Student with 3 degrees of freedom. In Table 2 we can see that in this case the nonparametric approach with general noise (solid curve) is performing better than the model in which the noise is assumed to follow a Gaussian distribution, regardless of the sample size.

5.2 Real Data Example

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. The data has been the subject of a nonparametric analysis of covariance in Young and Bowman (1995). One can see from Figure 1 that the relationship between the density and yield is non-linear 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 1 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 can be observed from Figure 2. We also notice that bootstrapping the data produces wider confidence bands due to the sparseness and high variability of the observations for large values of the density. But even the larger confidence band does not cover the parametric estimation of the AUC which is the decreasing black line in Figure 2. 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 Figure 3. While the difference between the nonparametric approach and the
parametric adjustment diminishes, we can notice in Figure 4 that the parametric estimate is more conservative indicating a smaller AUC for small densities. This indicates that the normal assumptions may not be entirely suitable for this dataset and the nonparametric approach is more fitting due to its robustness.

6. CONCLUSIONS

We introduce nonparametric adjustment for covariate information in the context of ROC analysis, more specifically for the AUC index. The approach proposed can be extended to calculations related to the Youden Index (YI) (Youden, 1950) and the optimal threshold for YI. In addition, the ROC conditional on a covariate value and all the indexes associated to this conditional curve can be computed in similar fashion to the one described in the paper. The approach bears some similarity to the work on nonparametric adjustment for covariates when estimating a treatment effect as in Young and Bowman (1995) and Cantori and 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 Mann-Whitney type approach. Our simulations show that even with moderate sample sizes the shape of the relationship between AUC and covariate values can be accurately recovered. In the paper the discussion is limited to the case of only one covariate but the approach proposed can certainly be extended to a larger number of quantitative covariates although the computational load increases significantly with each newly added covariate.

APPENDIX

A.1 Local Polynomial Estimators

We focus on the randomized sample \( (z_i,x_i), i = 1, \ldots, m \), which is assumed to consist of i.i.d. realizations from a random vector \((Z,X)\), and the results can be carried over to the diseased i.i.d. sample \((z_j,x_j,y_j), j = 1, \ldots, n\).
The local polynomial regression estimator of $f(z)$ is obtained by minimizing

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

where $h_1$ is a 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 (22) by $Z_x$,

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

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

$$
\hat{f}(z) = e_1^T (Z_x^T W_{x,h_1} Z_x)^{-1} Z_x W_{x,h_1} x.
$$

(23)

Analogously for the diseased sample $(z_{j,y}, y_j), j = 1, \ldots, n$, we define the quantities $Z_y = \begin{pmatrix}
1 & (z_{1,y} - z) & \cdots & (z_{1,y} - z)^p \\
\vdots & \vdots & \ddots & \vdots \\
1 & (z_{n,y} - z) & \cdots & (z_{n,y} - z)^p
\end{pmatrix}$, $W_{y,h_2} = \text{diag}\{K_{h_2}(z_{j,y} - z) : j = 1, \ldots, n\}$ and $y = (y_1, \ldots, y_n)^T$, where and obtain the local polynomial fit for $g(z)$,

$$
\hat{g}(z) = e_1^T (Z_y^T W_{y,h_2} Z_y)^{-1} Z_y W_{y,h_2} y.
$$

(24)

We next estimate the variance function $v_1(z)$ and $v_2(z)$ for heteroscedastic errors $\epsilon_1$ and $\epsilon_2$ 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, i.e., the variance observations, $v_{i,x}$ and $v_{j,y}, i = 1, \ldots, m, j = 1, \ldots, n$, defined by

$$
v_{i,x} = \{x_i - \hat{f}(z_{i,x})\}^2, \quad v_{j,y} = \{y_j - \hat{g}(z_{j,y})\}^2,
$$

(25)

where $\hat{f}(\cdot)$ and $\hat{g}(\cdot)$ are obtained by (23) and (24). Let $b_1 = b_1(n)$ and $b_2 = b_2(m)$ be the sequences of bandwidths that control the amount of smoothing for $\hat{v}_1(z)$ and $\hat{v}_2(z)$. Denote
\[ v_x = (v_{1,x}, \ldots, v_{m,x})^T \] and \[ v_y = (v_{1,y}, \ldots, v_{m,y})^T. \] Applying the local polynomial regression to the data \((z_{i,x}, v_{i,x})\), and \((z_{j,y}, v_{j,y})\), \(i = 1, \ldots, m, j = 1, \ldots, n\), we obtain

\[ \hat{\nu}_1(z) = e_1^T (Z_x^T W_{x,b_1} Z_x)^{-1} Z_x W_{x,b_1} v_x, \quad \hat{\nu}_2(z) = e_1^T (Z_y^T W_{y,b_2} Z_y)^{-1} Z_y W_{y,b_2} v_y, \]

(26)

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

**A.2 Auxiliary Results and Proofs**

**Lemma 1** If the assumptions (A1)-(A3), (A6) hold, and \(m \to \infty\), for a given \(z\),

\[ \sqrt{m} h_i \{ \hat{f}(z) - f(z), \hat{\nu}_1(z) - \nu_1(z) \}^T \xrightarrow{D} N(b_1(z), \Sigma_1(z)), \]

(27)

where \(b_1(z) = (b_{11}(z), b_{12}(z))^T\) and \(\Sigma_1(z) = \{\sigma_{x,ij}(z)\}_{1 \leq i,j \leq 2}\) with

\[ b_{11}(z) = \frac{\mu_{p+1}(K^*)}{(p+1)!} d_1 f^{(p+1)}(z), \quad b_{12}(z) = \frac{\mu_{p+1}(K^*)}{(p+1)!} d_1 \rho_1^{p+1} v_1^{(p+1)}(z), \]

\[ \sigma_{x,11}(z) = \frac{R(K^*) \nu_1(z)}{\theta(z)}, \quad \sigma_{x,22}(z) = \frac{R(K^*) \kappa_1(z)}{\theta(z) \rho_1}, \quad \sigma_{x,12}(z) = \frac{R(K^*, \rho_1) \eta_1(z)}{\theta(z) \rho_1}. \]

Analogously, if the assumptions (A1), (A4)-(A6) hold, and \(n \to \infty\), for a given \(z\),

\[ \sqrt{n} h_i \{ \hat{g}(z) - g(z), \hat{\nu}_2(z) - \nu_2(z) \}^T \xrightarrow{D} N(b_2(z), \Sigma_2(z)), \]

(28)

where \(b_2(z) = (b_{21}(z), b_{22}(z))^T\) and \(\Sigma_2(z) = \{\sigma_{y,ij}(z)\}_{1 \leq i,j \leq 2}\) with

\[ b_{21}(z) = \frac{\mu_{p+1}(K^*)}{(p+1)!} d_2 g^{(p+1)}(z), \quad b_{22}(z) = \frac{\mu_{p+1}(K^*)}{(p+1)!} d_2 \rho_2^{p+1} v_2^{(p+1)}(z), \]

\[ \sigma_{y,11}(z) = \frac{R(K^*) \nu_2(z)}{\theta(z)}, \quad \sigma_{y,22}(z) = \frac{R(K^*) \kappa_2(z)}{\theta(z) \rho_2}, \quad \sigma_{y,12}(z) = \frac{R(K^*, \rho_2) \eta_2(z)}{\theta(z) \rho_2}. \]

**Proof of Lemma 1.** The asymptotic normality of \(\hat{f}(z)\) with the bias \(b_{11}\) and the variance \(\sigma_{x,11}\) is standard in local polynomial regression. Let \(v_{i,x}^* = \{x_i - f(z_{i,x})\}^2\), note that the input data \(v_{i,x} = \{x_i - \hat{f}(z_{i,x})\}^2 = v_{i,x}^* + 2\{x_i - f(z_{i,x})\}\{\hat{f}(z_{i,x}) - f(z_{i,x})\} + \{\hat{f}(z_{i,x}) - f(z_{i,x})\}^2\).
Applying a local polynomial fit to \((z_{i,x}, v_{i,x}), i = 1, \ldots, m\), one can see that the second term will result in a quantity of the order \(O_p(\theta^{p+1} + 1/\sqrt{mb_1})\) and the third term will yield \(O_p(h_1^{2(p+1)} + 1/(mh_1))\). It is obvious that both quantities are ignorable, compared to the local polynomial estimator \(v^*_i(z)\) obtained by fitting \((z_{i,x}, v^*_{i,x})\). Therefore the estimators \(\hat{v}_1(z)\) and \(v^*_1(z)\) are asymptotically equivalent with the same limit distribution. Again we apply the standard argument of local polynomial regression to obtain the asymptotic normality of \(\hat{v}_1(z)\) with the bias \(b_{12}\) and variance \(\sigma_{x,22}\). To derive the covariance of the limit distribution between \(\hat{f}(z)\) and \(\hat{v}_1(z)\), it is equivalent to work with \(\hat{f}(z)\) and \(v^*(z)\). Moreover, using the equivalent kernel notation \(K^*\), the limiting covariance is identical to \(\text{cov}\{\hat{f}(z) - f(z), \hat{v}_1(z)\}\), where

\[
\hat{f}(z) = \frac{1}{m h_1 \theta(z)} \sum_{i=1}^{m} K^*(\frac{z_{i,x} - z}{h_1}) x_i, \quad \hat{v}(z) = \frac{1}{m b_1 \theta(z)} \sum_{i=1}^{m} K^*(\frac{z_{i,x} - z}{b_1}) v^*_i.
\]

Note that

\[
\text{cov}\{\hat{f}(z) - f(z), \hat{v}_1(z)\} = \frac{1}{m h_1 \theta^2(z)} \left( E[K^*(\frac{Z - z}{h_1}) K^*(\frac{Z - z}{b_1}) \{X - f(z)\} \{X - f(Z)\}^2] - E[K^*(\frac{Z - z}{h_1}) \{X - f(z)\} E[K^*(\frac{Z - z}{b_1}) \{X - f(Z)\}^2] \right).
\]

Employ Taylor expansion around \(z\) for the above first term, and the asymptotic bias \(b_{11}, \hat{v}_{12}\) are used to compute the order of the second term. Then one obtains

\[
\text{cov}\{\hat{f}(z) - f(z), \hat{v}_1(z)\} = \frac{1}{m h_1 \rho_1 \theta(z)} \left\{ \int K^*(u) K^*(u/\rho_1) du \eta(z) + O(h) \right\},
\]

which leads to the covariance of the limit distribution, \(\sigma_{x,12}\). The same arguments can be applied to obtain the joint asymptotic distribution in (28).

**Proof of Theorem 1.** The Cramér-Wold device is exploited to derive the asymptotic distributions of \(\widehat{A}_N(z)\) for three possible cases. It is obvious that if \(n/m \to \infty\), the estimation error of \(\hat{g}(z)\) and \(\hat{v}_2(z)\) are of smaller order than that of \(\hat{f}(z)\) and \(\hat{v}_1(z)\). Thus

\[
\Phi \left\{ \frac{\hat{g}(z) - \hat{f}(z)}{\sqrt{\hat{v}_1(z) + \hat{v}_2(z)}} \right\} \sim \Phi \left\{ \frac{g(z) - \hat{f}(z)}{\sqrt{\hat{v}_1(z) + v_2(z)}} \right\},
\]

\[21\]
where \( \sim \) denotes the asymptotic equivalence. Put \( H_1(f, v_1) = (g - f)/\sqrt{v_1 + v_2} \). Then the gradient vector of \( H_1 \) with respect to \((f, v_1)^T\),

\[
DH_1(f, v_1) = -\left(\frac{1}{\sqrt{v_1 + v_2}}, \frac{g - f}{2(v_1 + v_2)^{3/2}}\right)^T.
\]

Applying the Cramér-Wold device and the joint asymptotic distribution of \( \{\hat{f}(z), \hat{v}_1(z)\}\),

\[
B_1(z) = \phi\{\delta(z)\} DH_1\{f(z), v_1(z)\}^T b_1(z),
\]

\[
V_1(z) = \phi^2\{\delta(z)\} DH_1\{f(z), v_1(z)\}^T \Sigma_1(z) DH_1\{f(z), v_1(z)\},
\]

which leads to (17). The same argument is applied again to obtain (15) when \( n/m \to 0 \).

When \( n/m \to \lambda \) for \( 0 < \lambda < \infty \), we have the joint asymptotic distribution

\[
\sqrt{m} b_1\{\hat{f}(z) - f(z), \hat{v}_1(z) - v_1(z), \hat{g}(z) - g(z), \hat{v}_2(z) - v_2(z)\}^T \\
\xrightarrow{D} N\left\{ \begin{pmatrix} b_1(z) \\ \lambda^{-\frac{p+1}{2p+3}} b_2(z) \end{pmatrix}, \begin{pmatrix} \Sigma_1(z) & 0 \\ 0 & \lambda^{-\frac{2p+2}{2p+3}} \Sigma_2(z) \end{pmatrix} \right\}
\]

Put \( H_2(f, v_1, g, v_2) = (g - f)/\sqrt{v_1 + v_2} \), and the gradient vector of \( H_2 \) with respect to \((f, v_1, g, v_2)^T\) is

\[
DH_2(f, v_1, g, v_2) = \left(\frac{1}{\sqrt{v_1 + v_2}}, \frac{g - f}{2(v_1 + v_2)^{3/2}}, \frac{1}{\sqrt{v_1 + v_2}}, \frac{g - f}{2(v_1 + v_2)^{3/2}}\right)^T.
\]

Applying Cramér-Wold device and the joint asymptotic distribution (29) leads to (16).

**Lemma 2** If the assumptions (A1\(^*\))-(A3\(^*\)), (A6\(^*\)) and (A7.1) hold, and \( m \to \infty \),

\[
\sup_{z \in \mathcal{Z}} |\hat{f}(z) - f(z)| = O(\tau_m), \quad \sup_{z \in \mathcal{Z}} |\hat{v}_1(z) - v_1(z)| = O(\tau_m), \quad w.p.1., \tag{29}
\]

and If the assumptions (A1\(^*\)), (A4\(^*\))-(A6\(^*\)) and (A7.2) hold, and \( n \to \infty \),

\[
\sup_{z \in \mathcal{Z}} |\hat{g}(z) - g(z)| = O(\omega_n), \quad \sup_{z \in \mathcal{Z}} |\hat{v}_1(z) - v_1(z)| = O(\omega_n), \quad w.p.1., \tag{30}
\]

where \( \tau_m = h_1^{p+1} + \sqrt{\log(1/h_1)/(mh_1)} \) and \( \omega_n = h_2^{p+1} + \sqrt{\log(1/h_2)/(nh_2)} \) as defined in Theorem 2.
Proof of Lemma 2. It sufficient to show (29). The strong uniform convergence rate \( \tau_m \) for \( \hat{f} \) was obtained by Horng (2006), which is based on the arguments in Silverman (1978) and Mack and Silverman (1982) and the equivalent kernel \( \hat{v}_1 \), we follow the similar argument used in the proof of Lemma 1. Recall that \( v_{i,x}^* = \{x_i - f(z_i,x)\}^2 \), and \( v_{i,x} = \{x_i - \hat{f}(z_i,x)\}^2 \). Applying a local polynomial fit to \( (z_{i,x}, v_{i,x}) \), \( i = 1, \ldots, m \), the second and third terms of the resulting estimator tend to 0 with probability 1, and the leading term has the strong uniform convergence rate \( \tau_m \) by using the same argument for \( \hat{f} \).

Proof of Theorem 2. The proof follows Lemma 2 and the uniform version of Slutsky’s Theorem. It is only needed to note that, if (A2) and (A4) hold, \( A_N = \Phi(f, g, v_1, v_2) \) has bounded partial derivative in each argument, and thus satisfies Lipschitz continuity.

Proof of Theorem 3. For a given \( Z = z \), one can see that “hypothetical” estimator \( A_M(z) \) is essentially a two-sample U-statistic. The argument used in U-statistic can be applied here. The unbiasedness of \( A_M(z) \) is obvious. For the asymptotic variance at a given \( z \), put \( h(X, Y; z) = 1_{(0, \infty)}(Y - X|Z = z) - A(z) \), \( h_{0,0}^* = E\{h(X, Y; z)\} = 0 \). \( h_{1,0}^*(X; z) = E\{h(X, Y; z)|X\} \), \( h_{0,1}^*(Y; z) = E\{h(X, Y; z)|Y\} \). Note that

\[
h_{0,1}^*(Y; z) = P(Y \geq X|Y, Z = z) = P\left( f(z) + \epsilon_1 \sqrt{v_1(z)} \leq g(z) + \epsilon_2 \sqrt{v_2(z)} \right) = P\left( \epsilon_1 \leq \frac{g(z) - f(z)}{\sqrt{v_1(z)}} \right) \equiv h_{1,0}(\epsilon_2; z),
\]

and similarly \( h_{1,0}^*(X; z) \equiv h_{1,0}(\epsilon_1; z) \), i.e., \( \xi_{1,0}^2 \equiv \var\{h_{1,0}^*(Y; z)\} \), \( \xi_{0,1}^2 \equiv \var\{h_{0,1}^*(Y; z)\} \) as specified in Theorem 3. The unbiasedness of \( A_M(z) \) is obvious from \( h_{0,0}^* = 0 \). For the variance calculation, after some counting techniques, one has,

\[
\begin{align*}
\var\{A_M(z)\} &= \frac{1}{nm} \sum_{c=0}^{m-1} \sum_{d=0}^{n-1} C_m^c C_n^d \xi_{c,d}^2 = \frac{2}{m} + \frac{2}{n} + o\left( \frac{1}{m + n} \right), \tag{31}
\end{align*}
\]

where \( C_k^a \) is the combination of choosing \( k \) from \( n \). This proves (18).

To show the asymptotic normality (19), define

\[
T_{n,n}(z) = \sqrt{m + n} \left\{ \frac{1}{m} \sum_{i=1}^m h_{1,0}^*(x_i z) + \frac{1}{n} \sum_{j=1}^n h_{0,1}^*(y_j z) \right\},
\]

23
which is in fact the projection of $\sqrt{m + n} \{A_M(z) - A(z)\}$ 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 $\psi$ and $\psi^*$ are arbitrary measurable functions. From Hájek’s Projection Theorem and (31), we have, as $m,n \to \infty$,

$$\text{var}\{\sqrt{m + n}A_M(z) - T_{m,n}(z)\} = \text{var}\{\sqrt{m + n}A_M(z)\} - \text{var}\{T_{m,n}(z)\} \to 0,$$

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

**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 note that the dependences of $w_{ij}$ and $\hat{w}_{ij}$ on $x_{i,z}, y_{i,z}, z_{i,x}, z_{j,y}$ and $z$ are suppressed for simplicity. Let $a_1(z) = g(z) - f(z)$, $a_2(z_{j,y}, z) = \sqrt{\nu_2(z)/\nu_2(z_{j,y}, z)}$, $a_3(z_{i,x}, z) = -\sqrt{\nu_1(z)/\nu_1(z_{i,x})}$, $a_4(z_{j,y}, z) = -g(z_{j,y})a_2(z_{j,y}, z)$, $a_5(z_{i,x}, z) = -f(z_{i,x})a_3(z_{i,x}, z)$, 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 "-" 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 $\hat{f}$, $\hat{g}$, $\hat{v}_1$ 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 the uniform version of Slutsky’s Theorem (in probability instead of almost sure), we have, for a given $z$, $\hat{a}_1(z) \overset{P}{\to} a_1(z)$, $\sup_{z_{j,y}} |\hat{a}_k(z_{j,y}, z) - a_k(z_{j,y}, z)| = o_p(1)$, $\sup_{z_{i,x}} |\hat{a}_l(z_{i,x}, z) - a_l(z_{i,x}, z)| = o_p(1)$, for $k = 2, 4$ and $l = 3, 5$. Since $\epsilon_{i,j} \overset{i.i.d.}{\sim} F^*$, one has $\epsilon_{i,j} = O_P(1)$and, analogously, $\epsilon_{j,i} = O_P(1)$, regardless of $i$ and $j$. Also note that $f$, $g$, $v_1$ and $v_2$ are bounded on $Z$, then we obtain $\sup_{z_{i,x}, z_{j,y}} |\hat{w}_{ij} - w_{ij}| = o_p(1)$ that only depends on the given $z$.

To show (20), we observe that $E[(\hat{A}_M(z) - A_M(z))^2] = E_{0,0} + E_{1,0} + E_{0,1} + E_{1,1}$, where

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

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

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

$$E_{1,1} = \frac{1}{m^2n^2} \sum_{i \neq i', j \neq j'} E\left[\left(1_{[0,\infty)}(\hat{w}_{ij}) - 1_{[0,\infty)}(w_{ij})\right)\left(1_{[0,\infty)}(\hat{w}_{ij}) - 1_{[0,\infty)}(w_{ij})\right)\right].$$
while $E_{1,0}$, $E_{0,1}$ and $E_{1,1}$ are defined in the same way, with $E_{1,0}$ corresponds to $\sum_{i \neq i', j \neq j'}$, $E_{0,1}$ to $\sum_{i \neq i', j = j'}$ and $E_{1,1}$ to $\sum_{i = i', j = j'}$. We first focus on $E_{0,0}$,

$$E_{0,0} = \frac{1}{m^2n^2} \sum_{i \neq i', j \neq j'} \left\{ P(\hat{w}_{ij} \geq 0, \hat{w}_{ij'} \geq 0) + P(w_{ij} \geq 0, w_{ij'} \geq 0) \right. \left. -P(\hat{w}_{ij} \geq 0, w_{ij'} \geq 0) - P(w_{ij} \geq 0, \hat{w}_{ij'} \geq 0) \right\}$$

$$\leq \sup_{i \neq i', j \neq j'} \left| P(\hat{w}_{ij} \geq 0, \hat{w}_{ij'} \geq 0) + P(w_{ij} \geq 0, w_{ij'} \geq 0) \right. \left. -P(\hat{w}_{ij} \geq 0, w_{ij'} \geq 0) - P(w_{ij} \geq 0, \hat{w}_{ij'} \geq 0) \right|$$

(32)

For any given $z$, from Slutsky's Theorem, we have $(\hat{w}_{ij}, \hat{w}_{ij'})^T$, $(\hat{w}_{ij}, w_{ij'})^T$ and $(w_{ij}, \hat{w}_{ij'})^T$ converge in probability to $(w_{ij}, w_{ij'})^T$ uniformly in all arguments except $z$, which implies uniform convergence in distribution. Therefore the four sequences of probabilities in (32) all uniformly converge to $P(w_{ij} \geq 0, w_{ij'} \geq 0)$ 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 $E_{1,1} = O(E_{0,0}/mn)$. This completes the proof of Theorem 4.

REFERENCES


Table 1: Simulation results for two scenarios with normal noise: (a) mean functions non-diseased (lower curve) and diseased (upper curve); (b) variance functions for non-diseased (lower curve) and diseased (upper curve); (c) True AUC curve along with nonparametric estimates with and without the normal error assumption and the confidence bands obtained using the two bootstrap methods.

(a) ![Graph A](image1)

(b) ![Graph B](image2)

(c) ![Graph C](image3)

Scenario 1: \( n = 40, \beta_0 = \alpha_0 = 0, \alpha_1 = \alpha_2 = \beta_2 = \beta_1 = 3, \beta_3 = 1 \)

\[ \xi_0 = 0.3, \xi = 3, \delta_1 = 2, \delta_0 = -6, \gamma = 1.2 \]

Scenario 2: \( n = 100, \beta_0 = \alpha_0 = 0, \alpha_1 = \alpha_2 = \beta_2 = \beta_1 = 1.5, \beta_3 = 1 \)

\[ \xi_0 = 0.3, \xi = 1, \delta_1 = 2, \delta_0 = -6, \gamma = 1.2 \]
Table 2: Simulation results for two scenarios with Student(3) distribution of noise: (a) mean functions non-diseased (lower curve) and diseased (upper curve); (b) variance functions for non-diseased (lower curve) and diseased (upper curve); (c) True AUC curve along with nonparametric estimates with and without the normal error assumption and the confidence bands obtained using the two bootstrap methods.

(a)  

(b)  

(c)  

Scenario 1: $n = 50$, $\beta_0 = \alpha_0 = 0$, $\alpha_1 = \alpha_2 = \beta_2 = \beta_1 = 3$, $\beta_3 = 1$  
$\xi_0 = 0.3$, $\xi = 3$, $\delta_1 = 2$, $\delta_0 = -6$, $\gamma = 1.2$  

Scenario 2: $n = 100$, $\beta_0 = \alpha_0 = 0$, $\alpha_1 = \alpha_2 = \beta_2 = \beta_1 = 1.5$, $\beta_3 = 2.5$  
$\xi_0 = 0.3$, $\xi = 1$, $\delta_1 = 2$, $\delta_0 = -6$, $\gamma = 1.2$  

31
Figure 1: Spanish Onion Data with response on: the original scale (left) the logarithmic scale (right).
Figure 2: Comparison of estimated dependency between AUC and density obtained using the nonparametric approach with and without normal noise with the parametric estimation of the same dependency following Faraggi (2003).
Figure 3: Same comparison as in Figure 6 with response on the logarithmic scale.