Statistica Sinica **22** (2012), 1427-1456
doi:http://dx.doi.org/10.5705/ss.2010.167

A DIRECT SEMIPARAMETRIC RECEIVER OPERATING CHARACTERISTIC CURVE REGRESSION WITH UNKNOWN LINK AND BASELINE FUNCTIONS

Huazhen Lin[1], Xiao-Hua Zhou[2,3,4] and Gang Li[5]

[1] *Southwestern University of Finance and Economics,*
[2] *HSR&D Center of Excellence,* [3] *University of Washington,*
[4] *Peking University and* [5] *University of California at Los Angeles*

Abstract: In this article, we study a direct receiver operating characteristic (ROC) curve regression model with completely unknown link and baseline functions. A semiparametric procedure is proposed to estimate both the parametric and nonparametric components of the model. The resulting parameter estimates and ROC curve estimates are shown to be consistent and asymptotically normal with a $n^{-1/2}$ convergence rate. With arbitrary link and baseline functions, our model is more robust than existing direct ROC regression models that require either complete or partially complete specification of the link and baseline functions. Moreover, the robustness of our new method is gained at little cost to efficiency, as evidenced by the parametric convergence rate of our estimators and by the simulation study. An illustrative example is given using a hearing test data set.

Key words and phrases: Diagnostic tests, kernel smoothing, nonparametric, ROC regression, transformation models.

1. Introduction

The diagnostic accuracy of a medical test is often assessed using a receiver operating characteristic (ROC) curve (Zhou, Obuchowski, and McClish (2002)). ROC regression methodology offers a useful means of investigating how patient characteristics influence test accuracy. Several approaches to ROC regression have been developed during the past decade. Tosteson and Begg (1988) proposed regression models for ROC curves of ordinal-scale tests using ordinal regression models. Thompson and Zucchini (1989) proposed regression models for a summary measure, such as the area under the ROC curve. More recently, Pepe (1997, 2003) proposed to directly model covariate effects on the ROC curve using the model

$$ROC(t, X) = F\{\theta' X + H(t)\}, \tag{1.1}$$

where t is a false positive rate varying from 0 to 1, X is a p_0-dimensional covariate vector, F is a link function, representing the way covariates affect the ROC curve, H is a baseline function that satisfies $H(0) = -\infty$ and $H(1) = +\infty$, essentially defining the location and shape of the ROC curve, and θ is a vector of unknown regression coefficients, quantifying covariate effects. Model (1.1) includes the classic binormal model when $F(\cdot) = \Phi(\cdot)$ and $H(\cdot) = \alpha_0 + \alpha_1\Phi^{-1}(\cdot)$. Assuming that F and H are known, Alonzo and Pepe (2002) and Pepe (2003) proposed to estimate the regression parameters using estimating equations. Cai and Pepe (2002) and Cai (2004) studied a more flexible direct ROC model by assuming a parametric form for the link function together with a nonparametric baseline function H. The direct ROC regression approach has some appealing features. First of all, it directly models the effects of covariates on the test accuracy, and hence the results are easy to interpret. Secondly, the direct ROC regression model only makes assumptions on a functional form of the ROC curve and hence enjoys a certain degree of robustness (Hanley (1998), Metz, Herman, and Shen (1998)). Thirdly, the direct regression model preserves the property of invariance to monotone data transformations, a fundamental property of an ROC curve. Finally, direct modeling of ROC curves allows one to make inference on ROC curves in a restricted range of false-positive rates, incorporate interactions between covariates and false-positive rates, and compare ROC curves of tests with different numerical scales.

Since parametric or single semiparametric methods with a misspecified link or baseline function can lead to biased ROC curve estimates, the purpose of this paper is to study a more robust direct ROC regression model by allowing arbitrary link and baseline functions in the model (1.1). We propose a semiparametric method to estimate H, θ, and F, based on the observation that our model is equivalent to a transformation model with unknown transformation and error distribution functions. We show that the proposed estimators for the ROC curve and the regression parameters are consistent and asymptotically normal with the parametric convergence rate of $n^{-1/2}$. This result suggests that robustness of our method, realized by allowing nonparametric link and baseline functions, is obtained at little cost to efficiency. This asymptotic result is supported empirically by the simulation studies presented in Section 4.

The article is organized as follows. In Section 2, we describe our procedure for estimating $\theta, F(\cdot)$, $H(\cdot)$, and ROC curves. Section 3 gives the asymptotic distribution theory for the proposed estimators. In Section 4, we report a simulation study to evaluate the robustness and efficiency of the ROC curve and regression parameter estimates. An illustration is given in Section 5 using a data set from a hearing study.

2. Estimation Methods

2.1. Notation and model

Throughout the paper, subscripts $\bar{d}$ and d denote terms related to the non-diseased and diseased subjects, respectively. Let $Y_{\bar{d}}$ and Y_d denote test results, Z be the covariates that are relevant to both diseased and non-diseased subjects, and Z_d be the covariates that are specific to diseased subjects, such as disease severity. Take $X = (Z', Z_d')'$. Let $ROC(\cdot, X)$ be the ROC curve of the test in the subpopulation of diseased subjects with covariates Z and Z_d and the subpopulation of non-diseased subjects with the same covariates Z. We then model the effects of covariates X on the ROC curves by (1.1), where H is an unknown monotone increasing function satisfying $H(0) = -\infty$ and $H(1) = +\infty$, and F is an unknown cumulative distribution function.

Our data consist of n_d diseased subjects with multiple observations per subject, $\{Y_{d,ik}, Z_{ik}, Z_{d,ik}\}, k = 1, \ldots, m_i, i = 1, \ldots, n_d$, and $n_{\bar{d}}$ non-diseased subjects $\{Y_{\bar{d},ik}, Z_{ik}\}$, where $k = 1, \ldots, m_i, i = n_d + 1, \ldots, n_d + n_{\bar{d}}$. Here m_i is the number of repeated observations for the ith subject. In our data example, each patient has two observations, one from each of his/her two ears. Let $N_d = \sum_{i=1}^{n_d} m_i$, $N_{\bar{d}} = \sum_{i=n_d+1}^{n} m_i$, and $n = n_{\bar{d}} + n_d$. Observations are correlated if they are from the same subject and are independent otherwise. Let (t_0, t_1) be a region of false positive rates of interest. In the paper, we focus on the estimation of the ROC curve on (t_0, t_1), $0 < t_0 < t_1 < 1$.

To estimate the parameters and nonparametric functions, we first need to estimate $S_{\bar{d},Z}(y) = Pr(Y_{\bar{d}} \geq y|Z)$, the survival function of test results of non-diseased subjects given covariates Z. Following Cai and Pepe (2002), we use the semiparametric location model for $S_{\bar{d},Z}(y)$: $S_{\bar{d},Z}(y) = S_{\bar{d}}(y - \gamma' Z)$, where $S_{\bar{d}}$ is the unknown survival function. A more general choice for modeling $S_{\bar{d},Z}(y)$ is provided in Section 6, with similar results obtained. Denote the estimator of $S_{\bar{d},Z}(y)$ by $\widehat{S}_{\bar{d},Z}(y)$.

We consider the estimation of $\theta, H(\cdot)$, and $F(\cdot)$. It is easy to show that

$$E\left\{I\{Y_d \geq S_{\bar{d},Z}^{-1}(t)\}|X\right\} = F\{\theta' X + H(t)\} \tag{2.1}$$

is equivalent to the transformation model

$$H(T) = -\theta' X + \varepsilon, \tag{2.2}$$

where $T = S_{\bar{d},Z}(Y_d)$, and ε is a random error with the distribution function F. When F is specified up to a finite-dimensional vector of parameters, Cai (2004) extended the procedure proposed by Cheng, Wei, and Ying (1995) to estimate the ROC curve based on Model (2.2). When the observations of T are independent, Horowitz (1996) and Zhou, Lin, and Johnson (2009) proposed a semiparametric method to estimate H and F. However, there are several problems with a direct application of their methods to estimate the ROC curve. First, subjects may have multiple data records, so the original data are correlated. Second, since the distribution of $Y_{\bar{d}}$ is usually unknown, $S_{\bar{d}}$ needs to be estimated in order to apply these methods to observations of $T = S_{\bar{d},Z}(Y_d)$ and, as a result, the independence assumption on observations of T, required by these methods, no longer holds. As a result, the issue of correlation is more complicated than that caused by the multiple records of subjects. In this paper, we extend the Zhou, Lin, and Johnson (2009) method to estimate the ROC curve. However, because of the complicated correlation between $\widehat{T} = \widehat{S}_{\bar{d},Z}(Y_d)$, it is more involved when establishing asymptotic properties.

2.2. Estimation methods

Due to identification concerns, we set $E[\varepsilon] = 0$, $Var[\varepsilon] = 1$, $H(t_0) = 0$ and X includes an intercept term. Under the restrictions of $E[\varepsilon] = 0$ and $Var[\varepsilon] = 1$, F is identifiable, so the effect of X (excluding the intercept) is. Hence the estimated θ can be regarded as covariate effects under the standardized link function, defined as a distribution function of standardized random variable. As a result, θ in our model has the same interpretation as in the models proposed by Alonzo and Pepe (2002), Pepe (2003), Cai and Pepe (2002) and Cai (2004).

For the estimation of H, we observe that in the transformation model (2.2), T depends on X only through the index $W = \theta' X$. Let $G(\cdot|w)$ be the cumulative distribution function (CDF) of T conditional on $W = w$. Assume that H, F, and G are differentiable with respect to all their arguments. Let $h(t) = dH(t)/dt$, $f(t) = dF(t)/dt$, $p(t|w) = dG(t|w)/dt$, and $\mathrm{g}(t|w) = dG(t|w)/dw$. Model (2.2) implies that $G(t|w) = F(H(t) + w)$. Therefore, $p(t|w) = f(H(t) + w)h(t)$ and $\mathrm{g}(t|w) = f(H(t) + w)$. So, $\mathrm{g}(t|w)h(t) = p(t|w)$. Let $\mathrm{g}(t, w) = \mathrm{g}(t|w)p(w)$ and $p(t, w) = p(t|w)p(w)$, where $p(\cdot)$ is the density function of W. We then have

$$\mathrm{g}(t, w)h(t) = p(t, w). \tag{2.3}$$

Write $X_{ik} = (Z'_{ik}, Z'_{d,ik})'$. Replacing w in (2.3) by $W_{ik} = X'_{ik}\theta$ and summing over all observations, we get that $\sum_{i=1}^{n_d} \sum_{k=1}^{m_i} \mathrm{g}(t, W_{ik})h(t) = \sum_{i=1}^{n_d} \sum_{k=1}^{m_i} p(t, W_{ik})$. Therefore

$$h(t) = \frac{\sum_{i=1}^{n_d} \sum_{k=1}^{m_i} p(t, W_{ik})}{\sum_{i=1}^{n_d} \sum_{k=1}^{m_i} \mathrm{g}(t, W_{ik})}. \tag{2.4}$$

Integrating both sides of (2.4) gives us

$$H(t) = \int_{t_0}^{t} \frac{\sum_{i=1}^{n_d} \sum_{k=1}^{m_i} p(u, W_{ik})}{\sum_{i=1}^{n_d} \sum_{k=1}^{m_i} \mathrm{g}(u, W_{ik})} du. \tag{2.5}$$

Hence, by using (2.5), to construct an estimator H_n of H, we need to estimate $p(t)$, $G(t|w)$ and its derivatives. Let K_0 and K_1 be one-dimensional density functions, and b_0 and b_1 be bandwidths. We estimate $p(\cdot)$ by

$$p_n(w) = \frac{1}{N_d b_0} \sum_{i=1}^{n_d} \sum_{k=1}^{m_i} K_0 \left(\frac{W_{ik} - w}{b_0} \right). \tag{2.6}$$

We then estimate $G(t|w)$ by $G_n(t|w) = (N_d b_0 p_n(w))^{-1} \sum_{i=1}^{n_d} \sum_{k=1}^{m_i} I(\widehat{T}_{ik} \leq t)$ $K_0((W_{ik} - w)/b_0)$. We obtain an estimate of $\mathrm{g}(t|w)$ by differentiating $G_n(t|w)$ with respect to w,

$$\mathrm{g}_n(t|w) = \partial G_n(t|w)/\partial w. \tag{2.7}$$

Although $p(t|w) = \partial G(t|w)/\partial t$, we cannot use $\partial G_n(t|w)/\partial t$ to estimate $p(t|w)$, because $G_n(t|w)$ is a step function of t. Instead, we estimate $p(t|w)$ by

$$p_n(t|w) = \frac{1}{N_d b_0 b_1 p_n(w)} \sum_{i=1}^{n_d} \sum_{k=1}^{m_i} K_1 \left(\frac{\widehat{T}_{ik} - t}{b_1} \right) K_0 \left(\frac{W_{ik} - w}{b_0} \right). \tag{2.8}$$

The estimator H_n of H is obtained by substituting (2.6), (2.7), and (2.8) into (2.5).

Without imposing a parametric structure on F, it is natural to estimate θ by a solution to the estimating equation $\sum_{i=1}^{n_d} \sum_{k=1}^{m_i} \left(H(\widehat{T}_{ik}) + X'_{ik}\theta \right) X_{ik} = 0$. When given H, the estimator of θ is

$$\theta_n = - \left(\sum_{i=1}^{n_d} \sum_{k=1}^{m_i} X_{ik} X'_{ik} \right)^{-1} \sum_{i=1}^{n_d} \sum_{k=1}^{m_i} X_{ik} H(\widehat{T}_{ik}). \tag{2.9}$$

2.3. Implementation

We outline the algorithm for estimating θ, $H(\cdot)$, and $F(\cdot)$ as follows

Step 1: Specify an initial value of θ.

Step 2: Repeat (a) and (b) below until successive values of θ do not differ significantly.

(a) Given θ, estimate H by (2.5) with $p(u, W_{ik})$ and $\mathrm{g}(u, W_{ik})$ replaced by $p_n(u|W_{ik})p_n(W_{ik})$ and $\mathrm{g}_n(u|W_{ik})p_n(W_{ik})$, respectively.

(b) Given H, estimate θ by (2.9) as θ_n. Given H and θ_n, estimate $\sigma^2 = Var(\varepsilon)$ by the sample variance $\widehat{\sigma}^2$ and update θ_n by $\theta_n/\widehat{\sigma}$.

Let $\widehat{H}$ and $\widehat{\theta}$ be the results from the algorithm.

Step 3: Estimate F by the empirical distribution function of $\widehat{U} = \widehat{H}(\widehat{T}) + \widehat{\theta}' X$, denoted by $\widehat{F}$.

Step 4: The ROC curve for a test with covariate values x is estimated by $\widehat{ROC}(t, x) = \widehat{F}\{\widehat{\theta}' x + \widehat{H}(t)\}$.

Simulation suggests that our method is not sensitive to the initial value of θ. We use the estimation of θ from the classic binormal model as the initial value, because the classic binormal model is robust in some degree.

2.4. The selection of the bandwidths

Our estimation procedure involves the selections of the bandwidth b_0 and b_1. This can be achieved by using K-fold cross-validation (Tian, Zucker, and Wei (2005), Fan, Lin, and Zhou (2006)), to minimize the prediction error of $I(Y_{d,ik} \geq \widehat{S}^{-1}_{\bar{d},Z_{ik}}(t))$ based on (2.1) with $t = u_1, \ldots, u_D$, uniformly distributed on $[t_0, t_1]$. Concretely, denote the full dataset by T, and denote training and test sets by $T - T^v$ and T^v, respectively, for $v = 1, \ldots, K$. For each pair of bandwidths (b_0, b_1) and v, we find the estimator $\widehat{ROC}^v(t, x) = \widehat{F}\left(\widehat{\theta}' x + \widehat{H}(t)\right)$ of $F(\theta' x + H(t))$ using the training set $T - T^v$, and form the cross-validation criterion as

$$CV(b_0, b_1) = \sum_{v=1}^{K} \sum_{i \in T^v} \sum_{k=1}^{n_i} \sum_{\ell=1}^{D} \{I(Y_{d,ik} \geq \widehat{S}^{-1}_{\bar{d},Z_{ik}}(u_\ell)) - \widehat{ROC}^v(u_\ell, x)\}^2.$$

We then find the bandwidths (b_0, b_1) that minimize the criterion $CV(b_0, b_1)$. The number K is usually chosen to be $K = 5$ or $K = 10$. In the data analysis presented later, $K = 10$ is used.

3. Large Sample Properties

In this section, we establish the asymptotic properties for $\widehat{\theta}$, $\widehat{H}(\cdot)$, $\widehat{F}(\cdot)$, and $\widehat{ROC}(t, x)$. Generally, it is difficult to establish the asymptotic properties for the infinite-dimensional parameters $\widehat{H}(\cdot)$ and $\widehat{F}(\cdot)$ simultaneously. Fortunately, the expressions of $\widehat{\theta}$ and $\widehat{H}(\cdot)$ are not related to $\widehat{F}(\cdot)$. Hence, the derivation of asymptotic results can be done separately using existing techniques. Particularly, our proof on the asymptotic properties for all the estimators relies on five steps, with the second step being crucial. The first step consists of expansions for $\widehat{S}_{\bar{d},Z}(\cdot)$, which have already been obtained by Cai and Pepe (2002). The second step consists of an expansion for $\widehat{H}(t)$, where the key is to establish asymptotic forms of the $\widehat{p}(t, x'\widehat{\theta})$ and $\widehat{\mathrm{g}}(t, x'\widehat{\theta})$, the estimators of $p(t, x'\theta)$ and $\mathrm{g}(t, x'\theta)$, respectively; these asymptotic forms can be obtained by combining the technique in Zhou, Lin, and Johnson (2009) and the expansions of $\widehat{S}_{\bar{d},Z}(y)$. The third to fifth steps consist of proofs of the asymptotic normality for $\widehat{\theta}$, $\widehat{F}$ and $\widehat{ROC}(t, x)$, respectively.

We assume that as $n_{\bar{d}}$ and n_d converge to ∞, $n_d/n_{\bar{d}} \to r_0$, $n_{\bar{d}}/N_{\bar{d}} \to a_{\bar{d}}$, and $n_d/N_d \to a_d$. Lemma A in Appendix A.3 consists of an expansion of $\widehat{H}(t)$ that is a key to establishing the asymptotic properties of all the estimators in Theorems 1 to 4. All the proofs can be found in the Appendix.

Theorem 1. *Under conditions given in Appendix A.*1*, as* $n_{\bar{d}} \to \infty$ *and* $n_d \to \infty$,

$$\sqrt{n_d}\left(\widehat{\theta}-\theta\right) \to N\left(0, E\left[(a_{\bar{d}} A_1^{-1}\Delta_{d,\theta,i})^{\otimes 2}\right] + r_0 E\left[(a_d A_1^{-1}\Delta_{\bar{d},\theta,i})^{\otimes 2}\right]\right), \qquad (3.1)$$

where A_1, $\Delta_{d,\theta,i}$ *and* $\Delta_{\bar{d},\theta,i}$ *are defined in Appendix A.*2.

The asymptotic variance is a summation of two terms; the first term reflects the variation from the diseased subjects, and the second reflects the variation from the non-diseased subjects. Substituting the expression for $\widehat{\theta}-\theta$ into that for $\widehat{H}$ in (A.3) in Appendix A.3, we obtain $\widehat{H}(t)-H(t)$ as a sum of independent random variables.

Theorem 2. *Under conditions given in Appendix A.*1*, as* $n_{\bar{d}} \to \infty$ *and* $n_d \to \infty$,

$$\sqrt{n_d}\left(\widehat{H}(t)-H(t)\right) \to N(0, a_{\bar{d}}^2 E\Delta_{d,H,i}^2(t) + a_d^2 r_0 E\Delta_{\bar{d},H,i}^2(t)), \qquad (3.2)$$

where $\Delta_{d,H,i}(t)$ *and* $\Delta_{\bar{d},H,i}(t)$ *are defined in Appendix A.*2.

Thus, $\widehat{H}(t)$ is a $\sqrt{n_d}$ consistent, asymptotically normal estimator of $H(t)$, and we estimate the function $H(\cdot)$ at the parametric convergent rate. A similar conclusion on the parametric convergence rate of the estimated transformation function was reached in Horowitz (1996), Ye and Duan (1997), and Zhou, Lin, and Johnson (2009). The conclusion also assures that the resulting estimators for F, and then for the ROC curve, converge to their true values with the parametric convergence rate.

Theorem 3. *Under conditions given in Appendix A.1, as* $n_{\bar{d}} \to \infty$ *and* $n_d \to \infty$,

$$\sqrt{n_d}\left(\widehat{F}(z)-F(z)\right) \to N(0, a_d^2 E\Delta_{d,F,i}^2(t) + r_0 a_{\bar{d}}^2 f^2(z) E\Delta_{\bar{d},F,i}^2(z)), \qquad (3.3)$$

where $\Delta_{d,F,i}(z)$ *and* $\Delta_{\bar{d},F,i}(z)$ *are defined in Appendix A.*2.

Theorem 4. *Under conditions given in Theorem* 1,

$$\sqrt{n_d}\left(\widehat{ROC}(t,x)-ROC(t,x)\right) \\ \to N\left(0, a_d^2 E\Delta_{d,R,i}^2(t,x) + r_0 a_{\bar{d}}^2 f^2\left(\theta' x + H(t)\right) E\Delta_{\bar{d},R,i}^2(t,x)\right), \qquad (3.4)$$

where $\Delta_{d,R,i}(t,x)$ *and* $\Delta_{\bar{d},R,i}(t,x)$ *are defined in Appendix A.*2.

The variance estimation of $\widehat{H}, \widehat{\theta}, \widehat{F}$ and $\widehat{ROC}$ involves estimating derivatives of unknown functions. Because it may be difficult to get a good estimate of a derivative, we use the bootstrap method to approximate the variance or covariance matrix in our simulation studies and in the data example. The number of bootstrap samples is chosen to be 100.

From Theorems 1−4, we have

$$\begin{aligned}&\sqrt{n_d}\left(\widehat{ROC}(t,x)-ROC(t,x)\right)\\&\quad=\sqrt{n_d}f\left(\theta' x+H(t)\right)\frac{1}{N_{\bar{d}}}\sum_{i=n_d+1}^{n}\Delta_{\bar{d},R,i}(t,x)+\frac{\sqrt{n_d}}{N_d}\sum_{i=1}^{n_d}\Delta_{d,R,i}(t,x)+o_p(1).\end{aligned}$$

Following the resampling technique (Cai and Pepe (2002), Parzen, Wei, and Ying (1994)), we can construct a $(1-\alpha)100\%$ simultaneous confidence band for $\{ROC(t,x), t_0 \le t \le t_1\}$. However, the resampling technique requires good estimates for $\Delta_{\bar{d},R,i}(t,x)$ and $\Delta_{d,R,i}(t,x)$, which again involve estimating derivatives of unknown functions. A new method is needed to construct a simultaneous confidence band for $\{ROC(t,x), t_0 \le t \le t_1\}$. We will consider the problem in future work.

4. Simulation Study

4.1. Robustness

Since our procedure does not require a specification of parametric forms for the link and baseline functions, we would expect that the proposed procedure would be more robust than existing fully parametric and single semiparametric procedures. To investigate this issue, we examined the performance of the proposed method in comparison with the single-semiparametric method of Cai (2004) and the parametric method of Pepe and Cai (2004). We choose these methods because they are the most efficient among existing single-semiparametric and parametric direct regression methods, respectively. The performance of an estimator of the ROC curve, $\widehat{ROC}(\cdot,\cdot)$, is assessed via bias and the integrated mean square errors(IMSE),

$$\mathrm{IMSE}(X)=\sum_{k=1}^{n_{grid}}E[\widehat{ROC}(u_k,X)-ROC(u_k,X)]^2, \tag{4.1}$$

where $\{u_k, k=1,\cdots,n_{grid}\}$ are the grid points at which the functions $ROC(\cdot)$ are estimated. In the simulation, we chose $n_{grid}=100$ and u_k is distributed uniformly on $(0,1)$. We took K_1 to be a second-order kernel, K_0 to be a sixth-order kernel. The conditions in Appendix A.1 can be satisfied under these choices. We used the high-order kernels given in Muller (1984). Our simulations suggest that the proposed method is not sensitive to the selection of the kernel, provided the order of the kernel is satisfied.

For each of $n_d = 400$ diseased subjects, we first generated Z as uniform on $(0, 10)$ and $\varepsilon_d = \delta X_1 + (1-\delta)X_2$, with δ a binomial variable with $p = 0.5$, X_1 and X_2 were normal with standard deviation 0.01 and means -2 and 2, respectively. Here δ, X_1 and X_2 were independent. Hence, the distribution function of ε_d was $F(z) = (1/2)\Phi((z+2)/0.01) + (1/2)\Phi((z-2)/0.01)$, where $\Phi(z)$ is the standard normal distribution function. We then constructed the response Y_d of a diseased patient as $Y_d = \upsilon(\theta Z + \varepsilon_d) + Z$, where $\upsilon(z) = d_0 + (1/d_1)\sinh((z - d_2)/d_3)$, $d_0 = 8, d_1 = 2, d_2 = 2.5, d_3 = 3$, and $\theta = 0.5$. For each of $n_{\bar{d}} = 400$ non-diseased subjects, we generated Z and $\varepsilon_{\bar{d}}$ as uniform on $(0, 10)$ and standard normal, respectively, and then constructed the response $Y_{\bar{d}}$ of a non-diseased patient as $Y_{\bar{d}} = 7 + Z + \varepsilon_{\bar{d}}$. The induced ROC curve with $Z = z$ was then given by $ROC(t, z) = F(-\upsilon^{-1}(7 - \Phi^{-1}(t)) + \theta z)$.

For each simulated data set, we obtained estimates of θ and the ROC curve at $z = 3, 5, 7, 9$ using the proposed approach with the bandwidths $b_0 = 1.5$ and $b_1 = 0.05$, the single semiparametric method with the misspecified link function $F = \Phi$, and the parametric approach with the misspecified link $F = \Phi$ and misspecified baseline $H = \alpha_0 + \alpha_1\Phi^{-1}$. Here, we misspecified the link and baseline function as $\Phi(\cdot)$ and $\alpha_0 + \alpha_1\Phi^{-1}(\cdot)$, respectively, to induce the misspecified binormal model, since the binormal model is the most commonly used one in the ROC curve literature.

The averaged ROC curves and the distributions of the IMSE at $z = 3, 5, 7, 9$ over the 200 replications are displayed in Figures 1 and 2, respectively. Those show that the misspecification of both the link and baseline functions (indicated by "parametric") and the misspecification of the link function (indicated by "Single") lead to biased ROC curve estimates and, as a result, large IMSE. The single semiparametric method with the misspecified link function can lead to substantially biased estimates and its bias can be larger than the parametric estimate when both the link and baseline function are misspecified. These results suggest that the single semiparametric estimators of the regression parameters of Cai (2004) may not be robust, and the lack of robustness of the single semiparametric method in estimating the regression parameters can in part be attributed to the strong dependence of the parameter estimates on the specification of the link function (see (3.1) and the expression of $\xi(x)$ in Cai (2004)). The proposed estimate was much closer to the true ROC curve and had less IMSE than those produced by the parametric and single semiparametric estimates, suggesting that the proposed method is robust and accurate.

Numerical studies were also conducted with a smaller sample size under the same simulation scheme as in Figure 1 and Figure 2. Figure 3 and Figure 4 show the averaged ROC curves and the distributions of the IMSE at $z = 3, 5, 7, 9$ over the 500 replications with $n_0 = n_1 = 200$. For the smaller sample size, the

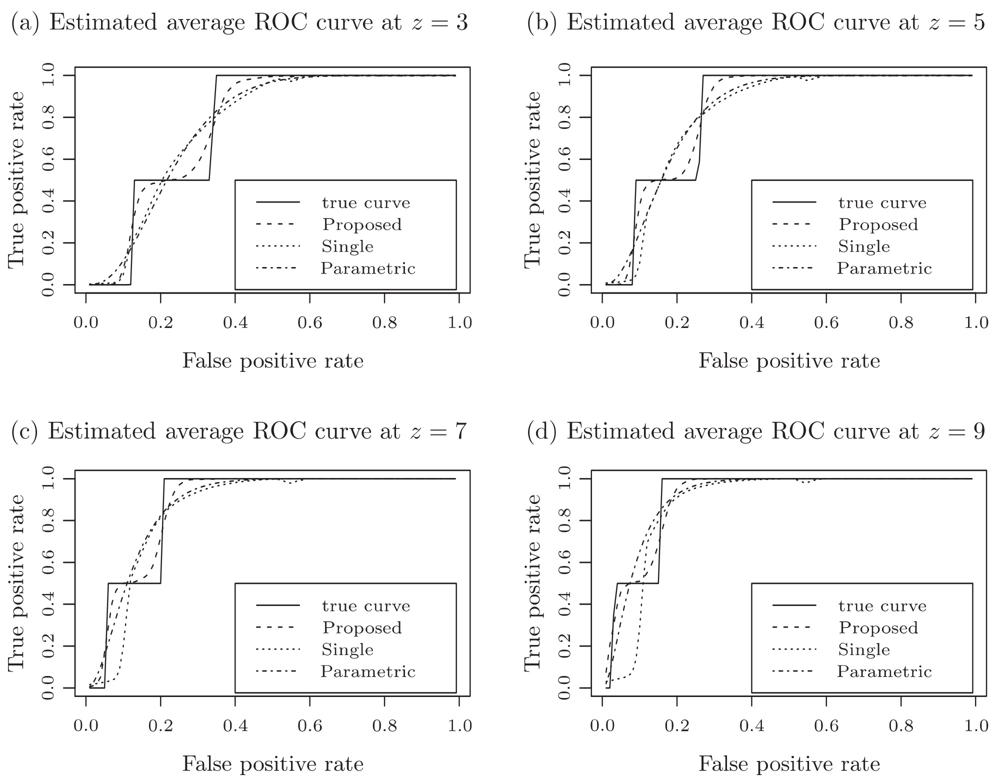

Figure 1. Average of the estimated ROC curves at $z = 3, 5, 7, 9$ over the 200 simulated data sets with $n_0 = n_1 = 400$.

performance of all estimators was worse; however, the relative performance of the three methods was similar to those in Figure 1 and Figure 2.

4.2. Relative efficiency

We next investigated whether the added robustness in our approach was gained at the expense of reduced statistical efficiency. We generated data from the binormal model and included both a covariate, Z, common to both the diseased and non-diseased subjects, and a covariate, Z_d, relevant only to diseased subjects. We generated independent observations for $n_d = 400$ diseased and $n_{\bar{d}} = 400$ non-diseased subjects as $Y_d = \alpha_1^{-1}\{\alpha_0 + \theta_1 Z_d + (\theta_2 + 0.5)Z + \varepsilon_d\}$ and $Y_{\bar{d}} = 0.5Z + \varepsilon_{\bar{d}}$, respectively. Here, $\alpha_0 = \alpha_1 = 1$, $\theta = (\theta_1, \theta_2)' = (0.5, 0.7)'$, $\varepsilon_{\bar{d}}$ and ε_d are the standard normal random variables, and Z_d and Z are Bernoulli ($p = 0.5$) and uniform $(0, 1)$ random variables, respectively. Given Z_d and Z, the induced ROC curve is $ROC_{Z_d,Z}(t) = \Phi(\alpha_0 + \alpha_1\Phi^{-1}(t) + \theta_1 Z_d + \theta_2 Z)$.

For each simulated data set, we obtained estimates of θ_1, θ_2, and the ROC curve at $X = (Z_d, Z) = (0, 0.5), (0, 0.75), (1, 0.25), (1, 0.5)$ using the proposed

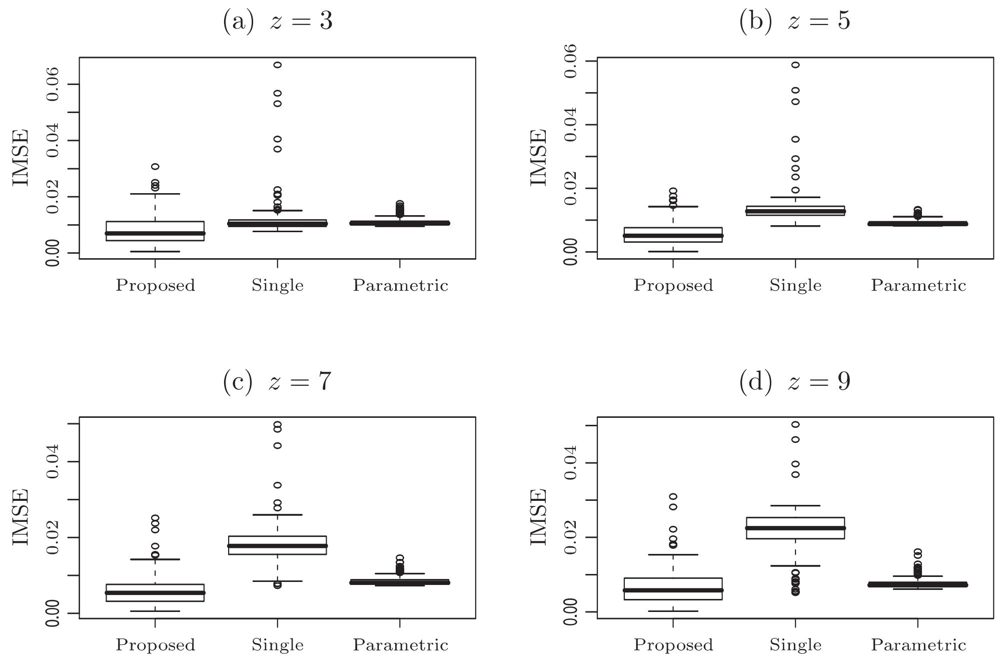

Figure 2. The boxplots for the distribution of the IMSE over 200 replications, using the proposed method, single semiparametric method, and the parametric method at $z = 3, 5, 7, 9$.

approach with the bandwidths $b_0 = 0.01, b_1 = 9$, the single semiparametric approach with the correctly specified link function, and the parametric approach with correct specification of both the link and baseline functions. The averaged ROC curves at $X = (0, 0.5), (0, 0.75), (1, 0.25), (1, 0.5)$ over the 200 simulated data sets, using the three methods, were almost the same and basically unbiased; hence, we do not report the results here. The distributions of the IMSE over the 200 replications are displayed in Figure 5. The IMSE of the proposed estimators for the ROC curve was larger than those of the single semiparametric and the parametric methods when X was close to its boundary, and they were comparable when X was in the interior of its support. This is not surprising, because the parametric estimation is carried out under the true link and baseline functions, the single semiparametric estimation is carried out under the true link function, whereas the proposed method assumes these two functions are unknown. In addition, we also note that the IMSE of the single semiparametric method was close to that of the parametric methods in all cases, suggesting that the single semiparametric method is efficient.

The distributions of the IMSE over the 500 replications with $n_0 = n_1 = 200$ are displayed in Figure 6. Similar conclusions with those from Figure 5 are obtained.

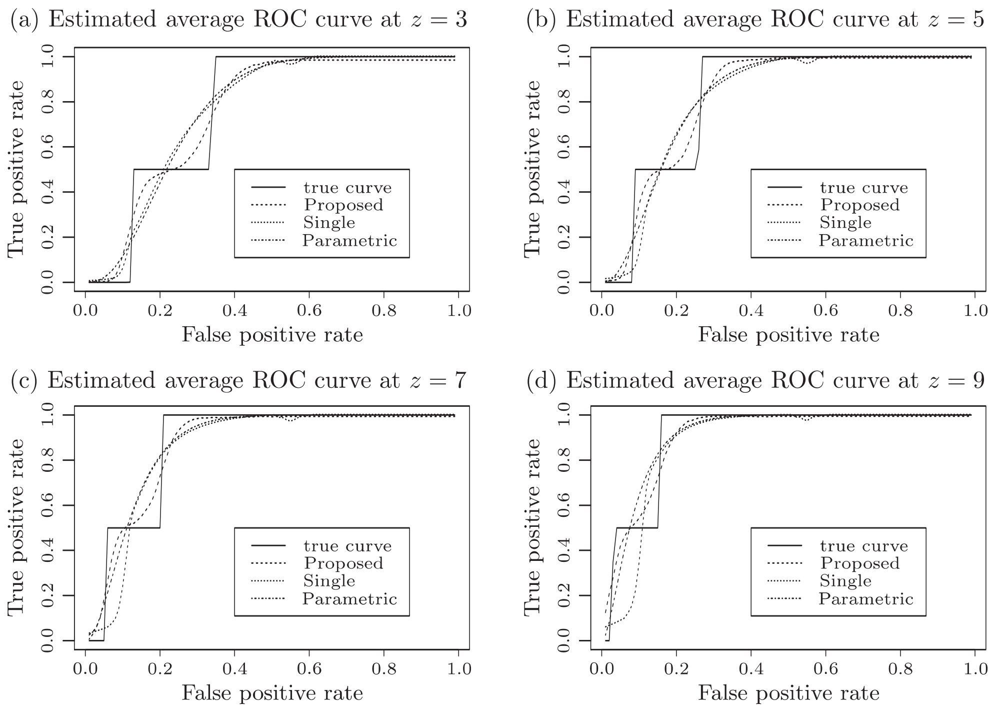

Figure 3. Average of the estimated ROC curves at $z = 3, 5, 7, 9$ over the 500 simulated data sets with $n_0 = n_1 = 200$.

4.3. The issues of initial value, bandwidth and kernel

Our method requires initial values to start the estimation. Based on the classical binormal model, we suggested an initial value in Section 2.3. It may be not close to the true value. Hence, it is necessary to test if the proposed algorithm is sensitive to the initial value of θ. To investigate the issue, we first took a series of initial values of $\theta = (\theta_1, \theta_2)$, varied in a 7×7 design with $\theta_1 = (-15, -10, -5, 0.1, 5, 10, 15)$ and $\theta_2 = (-15, -10, -5, 0.1, 5, 10, 15)$, then we selected a typical sample from the simulation in Section 4.2 with $n_0 = n_1 = 400$. The estimated MSE-value of the typical sample for θ is the median of the 200 MSE-values. All the selected initial values had the algorithm converge and gave the estimator $\widehat{\theta} = (0.3885, 0.5183)$. We also investigated the sensitivity of the initial value using a typical sample from the simulation in Section 4.1, and a similar conclusion is obtained.

In the derivation of the large sample properties, higher order kernel functions are needed to ensure the $\sqrt{n}$ convergent rate. With higher order kernel function, however, $p_n(t)$, $g_n(t|w)$, and $p_n(t|w)$ are not always positive and, hence, the resultant $ROC(t; x)$ may not be monotone increasing in t as desired. As posed

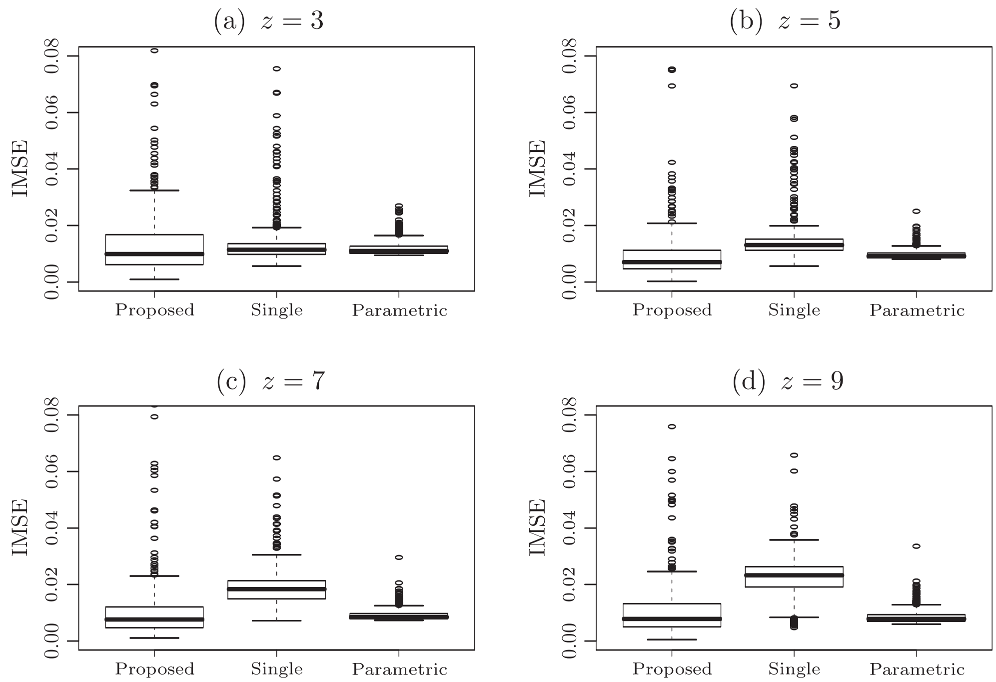

Figure 4. The boxplots for the distribution of the IMSE over 500 replications, using the proposed method, single semiparametric method, and the parametric method at $z = 3, 5, 7, 9$ for the simulation data with $n_0 = n_1 = 200$.

by a referee, from a practical point of view is the proposed method sensitive to the choice of the kernel function? By changing the parameters k, μ and ν in the kernels function provided by Muller (1984), we obtained the rectangular kernel ($k = 2, \mu = 0, \nu = 0$), the Epanechnikov kernel ($k = 2, \mu = 1, \nu = 0$), the Legendre kernel of order j ($k = 2j+2, \mu = 0, \nu = 0$, Deheuvels (1977)), the Eddy kernel (any $k, \mu = 1, \nu = 0$, Eddy (1980)), Ramlau-Hansen kernel (any k, any μ, $\nu = 0$, Ramlau-Hansen (1983)) and the Gasser kernel (any $k, \mu = 0, 1$, any ν, Gasser and Muller (1979)). Because the theoretical conclusion require sixth-order kernel, we take $k = 6$. To investigate the sensitivity of the kernel function, we varied μ and ν in a 2×3 design with $\mu = 0$ or 2 and $\nu = 0, 2$ or 4. For the typical sample with $n_0 = n_1 = 400$ from the simulation in Section 4.2, the estimated ROC curve at $X = (0, 0.5), (0, 0.75), (1, 0.25), (1, 0.5)$ with each pair of (μ, ν) is displayed in Figure 7. Figure 7 shows that all the estimated ROC curves with different kernel functions fully overlap, suggesting the proposed method is not sensitive to the selection of the kernel functions. In addition, from Figure 7, we can see that the estimated ROC curve is very close to a monotone increasing function, although is not always monotone increasing.

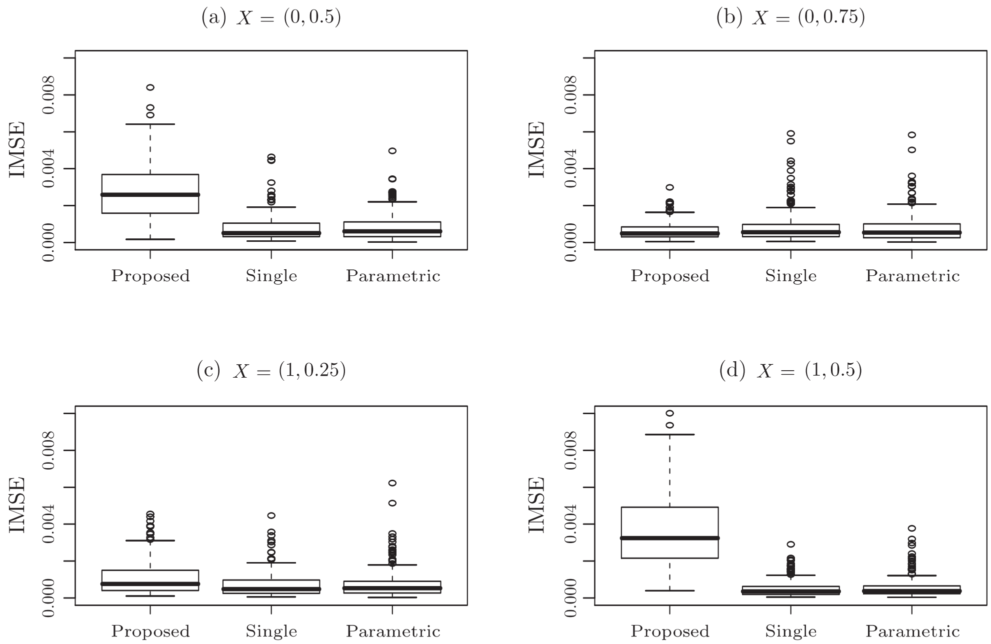

Figure 5. The boxplots for the distribution of the IMSE using data with $n_0 = n_1 = 200$ over the 200 replications, using the proposed method, single semiparametric method and the parametric method at $X = (0, 0.5), (0, 0.75), (1, 0.25), (1, 0.5)$.

Finally, we examined the performance of the proposed cross-validation method for selecting the bandwidths b_0 and b_1. We ran the sample with $n_{\bar{d}} = 400$ and $n_d = 400$ according to the simulation in Section 4.1 and the sample with $n_{\bar{d}} = 400$ and $n_d = 400$ from the simulation in Section 4.2. For the first case, we varied the bandwidth b_0 in $\{1.5, 1.8, 2.2, 2.5, 2.8\}$ and the bandwidth b_1 in $\{0.01, 0.05, 0.1, 0.15, 0.2\}$. For the second, we varied the bandwidth b_0 in $\{6, 7, 8, 9, 10, 11, 12\}$ and b_1 in $\{0.01, 0.03, 0.05, 0.07, 0.09, 0.12, 0.15\}$. We obtained the $CV(b_0, b_1)$ and $IMSE(b_0, b_1)$. Figure 8 is the plot of CV vs IMSE for the two cases. Figure 8 shows that CV increases as $\|\widehat{\theta} - \theta\|$ increases, suggesting that the proposed cross-validation method provides a reasonable estimator of bandwidths.

5. Example: Hearing Test Study

We illustrate the application of our method in the hearing test study of Stover, et al. (1996), which has also been reported on by Pepe (2003). For each ear, the distortion product otoacoustic emission (DPOAE) test is applied under nine different settings for the input stimulus. Each setting is defined

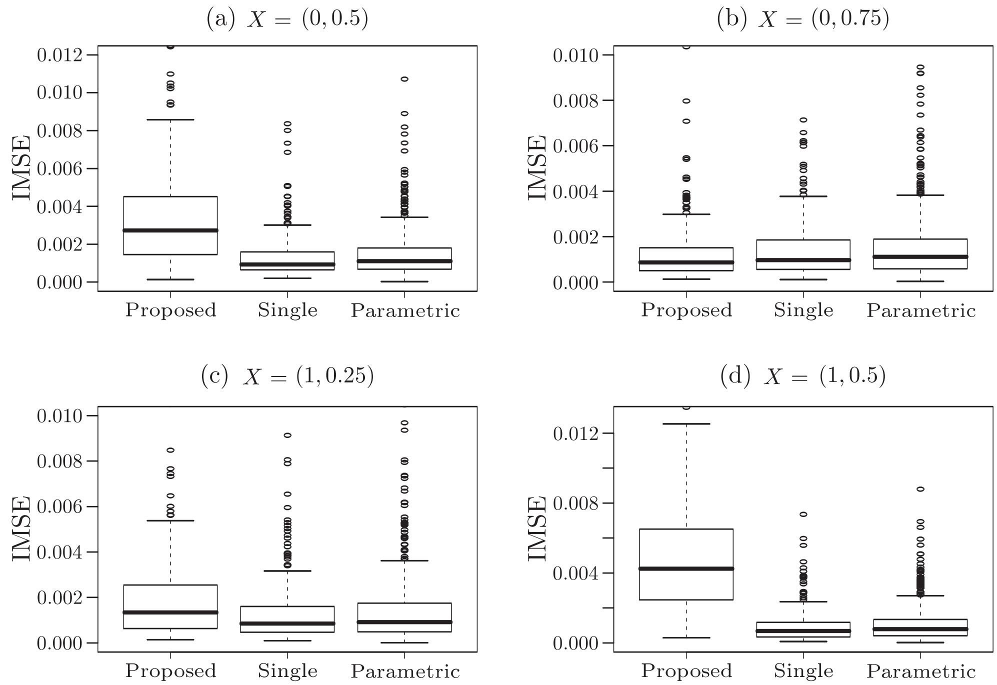

Figure 6. The boxplots for the distribution of the IMSE using data with $n_0 = n_1 = 200$ over the 500 replications, using the proposed method, single semiparametric method and the parametric method at $X = (0, 0.5), (0, 0.75), (1, 0.25), (1, 0.5)$.

by a particular frequency (f) and intensity (L) of the auditory stimulus. The response of the ear to the stimulus can be affected by the stimulus parameters, as well as by the hearing status of the ear. Among hearing-impaired ears, the severity of hearing impairment, as measured by the true hearing threshold, is expected to affect the results of the DPOAE test. The test result, Y, is the negative signal-to-noise ratio response, which coincides with our convention that higher values are associated with hearing impairment. The disease variable, D, is hearing impairment of 20 decibels (dB) or more. Covariates are the frequency and intensity levels of the stimulus. With the same set-up as in Pepe (2003), we have two covariates, X_f = frequency/100, measured in Hertz, and X_L = intensity/10, measured in dB. We also use a disease-specific covariate, Z_d = (hearing threshold − 20)/10, measured in decibels, taking values greater than 0 dB for hearing-impaired ears, and is undefined for normally-hearing ears. We fit the model $ROC_{X_L,X_f,Z_d}(u) = F(H(u) + \theta_1 X_L + \theta_2 X_f + \theta_3 Z_d)$ and analyzed the data using the proposed method with bandwidths $b_0 = 1$ and $b_1 = 0.01$, the single semiparametric method with the link function $F = \Phi$ and unknown

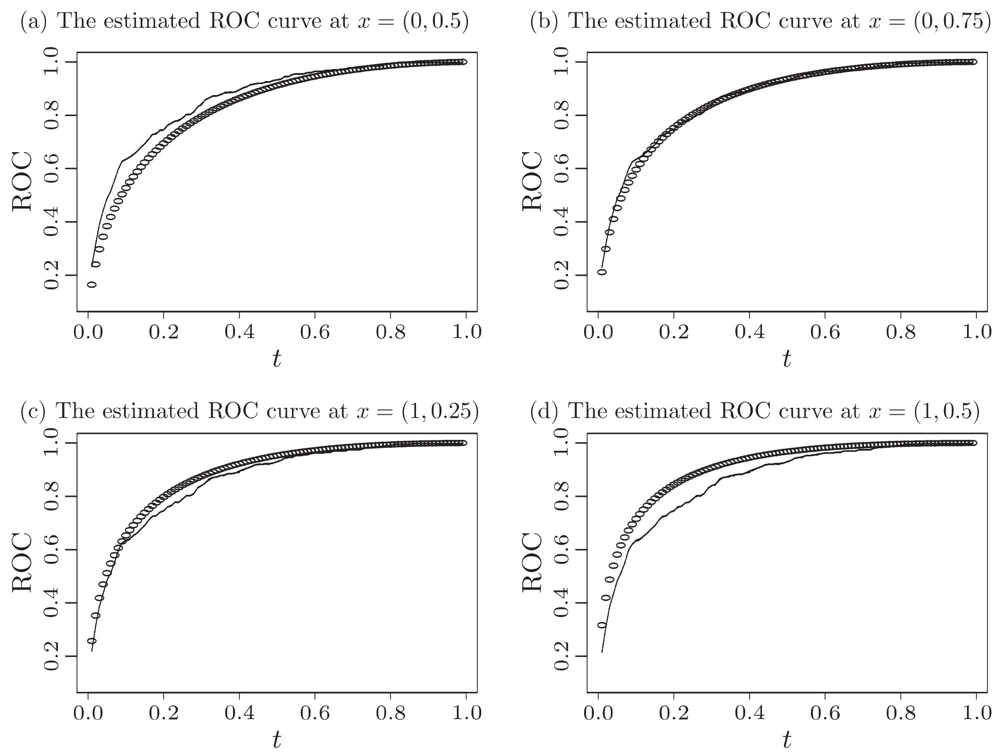

Figure 7. The estimated ROC curve (solid lines) at (a) $X = (0, 0.5)$, (b) $X = (0, 0.75)$, (c) $X = (1, 0.25)$ and (d) $X = (1, 0.5)$ with varying kernel functions based on the typical sample from the simulation in Section 4.2. Pointed line is the true ROC curve.

H, and the parametric method with $F = \Phi$ and $H = \alpha_1 + \alpha_2\Phi^{-1}$. Table 1 presents the coefficient estimates and their bootstrap standard errors using the three methods. Here, to compare our model with the binormal model, we took $t_0 = \min_{i,k} \widehat{T}_{ik}$ and $H(t_0) = \Phi^{-1}(t_0)$. The hearing test appears to perform better when the stimulus used has a low intensity and a higher frequency. The estimates of θ_3 clearly indicate that it is easier to detect hearing-impairment among severely hearing impaired patients than mildly impaired patients.

The estimated ROC curves at $X_L = 6.0$, $X_f = 14.16$, and $Z_d = 0.5$ using the three methods are plotted in Figure 9. Table 1 shows that the absolute value of the parameter estimates using our method are smaller than those using the existing single semiparametric and parametric methods. However, the estimated ROC curve of our method is higher than those from the other two methods. Our models make some assumptions, such as the location model for test results of the non-diseased, and a linear relationship between covariates and the transformation of test results of the diseased. We should investigate the validity of these

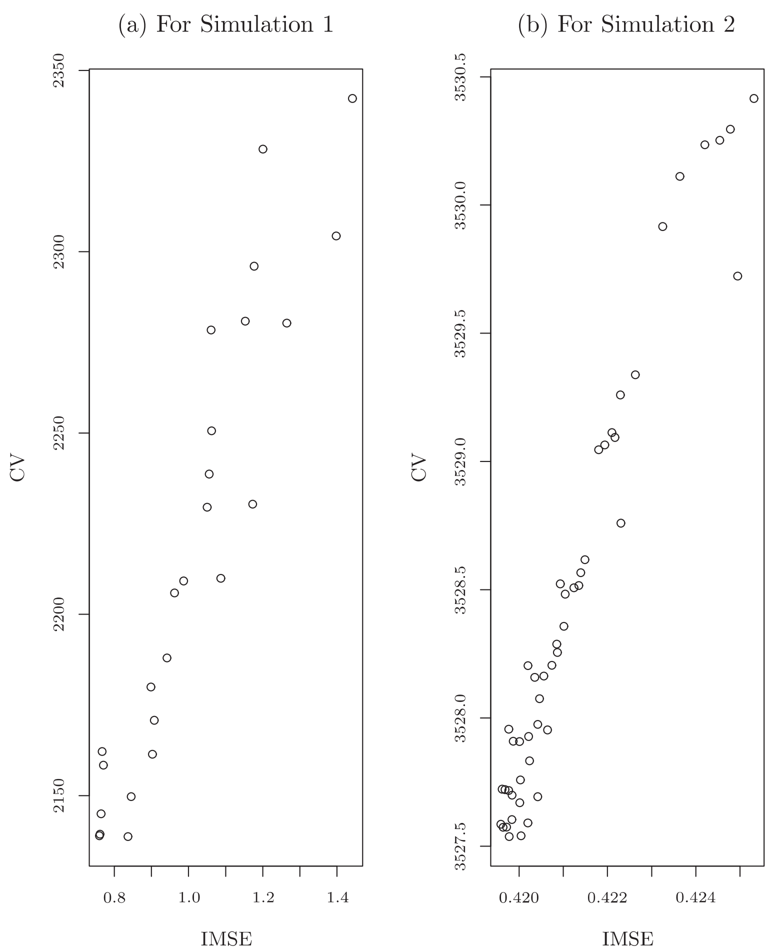

Figure 8. The scatter plot of CV vs IMSE for the typical samples (a) with $n_0 = n_1 = 400$ from the simulation in Section 4.1; and (b) with $n_0 = n_1 = 400$ from the simulation in Section 4.2.

assumptions. In Figure 9, we also displayed the empirical ROC curve, based on 147 non-diseased patients, with $X_L = 6.0, X_f = 14.16$, and 20 diseased patients, with $X_L = 6.0, X_f = 14.16$, and $Z_d = 0.5$. We see that on average, our estimate is closer to the empirical ROC curve than the two other estimated curves when the false positive rate is between 0.15 and 1.0. When the false positive rate is between 0 and 0.15, the estimated empirical ROC curve is not reliable due to the small number of observations available in this region.

6. Discussion

In this paper, we have developed double semi-parametric regression mod-

Table 1. Hearing test: estimated parameters and standard errors.

	Proposed	Single	Parametric
	estimated(SD)	estimated(SD)	estimated(SD)
θ_1	-0.4951(0.1137)	-0.5525(0.1136)	-0.5435(0.0955)
θ_2	0.0401(0.0157)	0.0394(0.0173)	0.0370(0.0171)
θ_3	0.3740(0.0462)	0.4241(0.0483)	0.3819(0.0484)

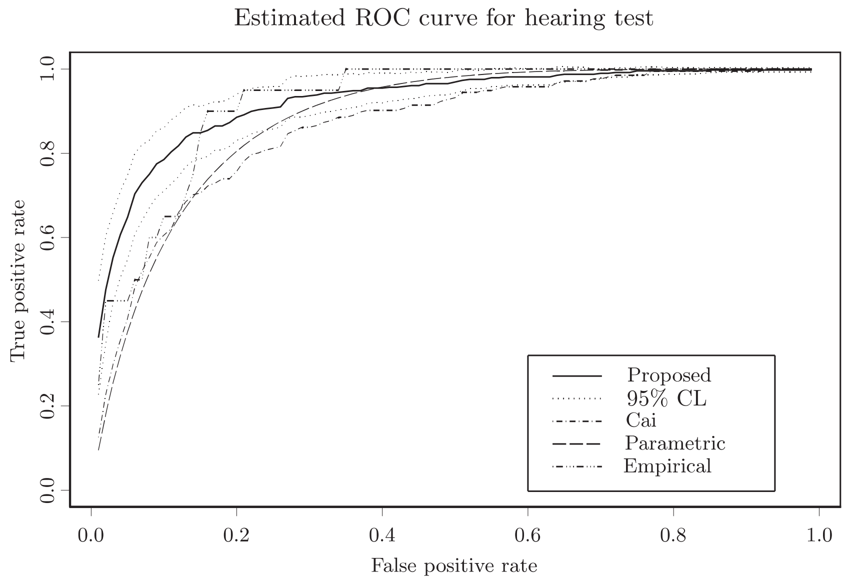

Figure 9. The estimated ROC curves of the hearing test at $X_L = 6.0, X_f = 14.16, Z_d = 0.5$.

els for the ROC curves, that allow both the baseline and link functions to be non-parametric. Our method has the following two apparent advantages over existing methods: non-parametric specifications of the link and baseline functions increase robustness; the new estimators for the regression parameters and the ROC curve converge to their true values at the parametric rate $n_d^{-1/2}$, suggesting that extra flexibility is gained at little cost to efficiency; this is also confirmed by our simulation studies.

We propose using the semi-parametric location model for $S_{\bar{d},Z}(y)$, the survival distribution of test results of the non-diseased patients with covariates Z; however, we can extend the method to a more general model, such as the semi-parametric linear transformation model (1.1) for $S_{\bar{d},Z}(y)$. Under the framework of the semi-parametric linear transformation model, we can estimate $S_{\bar{d},Z}(y)$ at rate $n^{-1/2}$ by the method of Zhou, Lin, and Johnson (2009), and hence the requirement on $\widehat{S}_{\bar{d},Z}(y)$ by the method proposed in the paper is satisfied. Therefore, related large sample properties can be obtained by replacing the asymptotic expansion of $\widehat{S}_{\bar{d},Z}(y)$ in the paper with that under the framework of the semiparametric linear transformation model.

We make linear assumptions, $\theta' X$ and $\gamma' Z$, at (1.1) and for the model for test results of the non-diseased subjects, respectively. The linear assumption is commonly used for dimension reduction so that we can make reliable inferences based on limited data. If we add additional higher-order and interaction terms to the models, the linear assumption can be approximately satisfied. However, in doing so, we may end up with a model with too many covariates. Further research is needed for efficiently selecting a subset of significant variables from the model (1.1) and the model for test results of the non-diseased subjects. Variable selection for (1.1) is challenging, because we need to consider model selection in two parts of the model: the nonparametric component of the outcome transformation and the parametric component of significant variables.

Acknowledgement

Lin's research was supported by the National Natural Science Funds for Distinguished Young Scholar (No. 11125104), the National Natural Science Foundation of China (No. 11071197) and Program for New Century Excellent Talents in University.

Appendix

We outline the proofs for the asymptotic results given in Section 3. More details of the proofs can be found in Zhou, Lin, and Johnson (2009). We make some assumptions that are needed to prove asymptotic normality of estimators for $\theta, H(\cdot)$, $F(\cdot)$ and the ROC curve. Hereafter, we let $f^{(k_1,k_2,\ldots)}(x_1, x_2, \ldots) = d^{(k_1+k_2+\cdots)} f(x_1, x_2, \ldots)/dx_1^{k_1} dx_2^{k_2} \cdots$ be the $(k_1+k_2+\cdots)$th order partial derivative of f.

A.1. Conditions

1. Let K_0 and K_1 be one-dimensional bounded and symmetric functions with compact supports. Without loss of generality, these supports are $[-1, 1]$. K_1 is of bounded variation, while K_0 has a bounded, continuous second derivative, for which $|K_0^{(2)}(x_1) - K_0^{(2)}(x_2)| \leq M|x_1 - x_2|$ for some $M < \infty$.
2. As $n_d \to \infty$, $n_d b_0^8 \to \infty$, $\log n_d/\sqrt{n_d b_0^6 b_1} \to 0$, $n_d b_0^{2s_0} \to 0$, and $n_d b_1^{2s_1} \to 0$, where s_0 and s_1 are the orders of K_0 and K_1, respectively.
3. Z and Z_d have a bounded support. The true values of θ belong to the interior of a known compact set, and the support of $W = \theta' X$ is taken to be Θ.
4. H is strictly increasing, and its derivatives, $H^{(k)}(t)(k = 1, \ldots, s_1 + 1)$, exist and are uniformly bounded over $t \in [t_0, t_1]$. The derivatives, $p^{(k)}(w)$ and

$p^{(k_1,k)}(t,w), k_1 = 1,\ldots,s_1+1, k = 1,\ldots,s_0+1$, exist and are uniformly bounded over $t \in [t_0,t_1]$ and $w \in \Theta$.

5.
$$\frac{1}{N_d}\sum_{i=1}^{n_d}\sum_{k=1}^{m_i} I\left(T_{ik} \notin [t_0,t_1]\right) = o_p(n_d^{-1/2}). \tag{A.1}$$

6. $\max_{1\le i\le n_d+n_{\bar d}} m_i < \infty$.

7. There is a sequence $\widehat{\theta} = \widehat{\theta}_n$ such that $\|\widehat{\theta} - \theta\| = o_p(1)$.

The assumption on b_0 and b_1 can be satisfied, for example, if K_1 is a second-order kernel, K_0 is a sixth-order kernel, $b_1 \propto n_d^{-1/3}$, and $b_0 \propto n_d^{-1/10}$. Since g_n is a function of the derivatives of K_0 and derivative functionals converge relatively slowly, the higher-order kernel for K_0 is needed to insure a sufficiently rapid convergence; We use one given in Muller (1984). With a higher order kernel function, however, $p_n(t)$, $\mathrm{g}_n(t|w)$, and $p_n(t|w)$ may not be positive. Condition (A.1) is used to avoid the tail problem. Condition 7 is commonly assumed in the semiparametric literature; see Carroll et al. (1997) and Horowitz (1996), and this condition requires that the initial value for θ be close to its true value.

A.2. Notation

To express Theorems 1 to 4, we write $B_1 = E[Z]$, $B_2 = E[ZZ']$

$$\bar{\mathrm{g}}(t) = E\mathrm{g}(t,W),\ \ q_0(w) = E[X|W=w],\ \ q_1(w) = E[Z|W=w],$$
$$\eta(t,w) = \frac{2h(t)p^{(1)}(w)}{\bar{\mathrm{g}}(t)},\ \ \pi(t) = \int_{t_0}^{t} h(u)\frac{E\left[\mathrm{g}(u,W)q_0^{(1)}(W)\right]}{\bar{\mathrm{g}}(u)}du,$$
$$Q(u,w) = (B_1 - q_1(w))S_{\bar d}^{(1)}(S_{\bar d}^{-1}(u))p(u|w),$$
$$\lambda(u) = \frac{E\left[\left\{p^{(10)}(u|W) - h(u)p^{(01)}(u|W)\right\}p(W)\right]}{\bar{\mathrm{g}}(u)},\ \ \text{and}$$
$$\Gamma(t) = \int_{t_0}^{t}\frac{E\left[\left\{Q^{(10)}(u,W) - h(u)Q^{(01)}(u,W)\right\}p(W)\right]}{\bar{\mathrm{g}}(u)}du.$$

Then we define

$$A_1 = EX\{X+\pi(T)\}',\ \ A_2 = EX\left[h(T)S_{\bar d}^{(1)}(S_{\bar d}^{-1}(T))(Z-B_1)+\Gamma(T)\right]',$$
$$A_3(t) = EX + \int_{-\infty}^{+\infty}\pi(y)p(t-H(y))dH(y),$$
$$A_4(t) = \int_{-\infty}^{+\infty}\left\{h(y)S_{\bar d}^{(1)}(S_{\bar d}^{-1}(y))[B_1+q_1(t-H(y))]-\Gamma(y)\right\}p(t-H(y))h(y)dy,$$
$$A_5(t,x) = (x+\pi(t))'A_1^{-1}A_2 - \Gamma'(t) - A_4'(\theta'x+H(t)) - A_3'(\theta'x+H(t))A_1^{-1}A_2,$$

$$A_6(t,x) = \left(A_3\left(\theta' x + H(t)\right) - x - \pi(t)\right)' A_1^{-1}.$$

Let E_{XT} and E_W be the expectation with respect to (X,T) and W, respectively. Related to the components of the asymptotical expansion, we take

$$\begin{aligned}
\mu_{\bar{d},ik} &= Y_{\bar{d},ik} - \gamma' Z_{ik}, \;\; e_{\bar{d},ik}(u) = I(\mu_{\bar{d},ik} \geq S_{\bar{d}}^{-1}(u)) - u,\\
\psi_{\bar{d},ik}(t) &= I(\mu_{\bar{d},ik} \geq S_{\bar{d}}^{-1}(t)) h(S_{\bar{d}}(\mu_{\bar{d},ik})) - H(t) + \int_{t_0}^{t} \lambda(u) e_{\bar{d},ik}(u) du,\\
\tau_{\bar{d},ik}(t) &= h(t) e_{\bar{d},ik}(t) - \psi_{\bar{d},ik}(t), \;\; \tilde{\tau}_{\bar{d},ik} = E_{XT}\left[X \tau_{\bar{d},ik}(T)\right],\\
\delta_{\bar{d},ik}(t) &= \psi_{\bar{d},ik}(t) + \Gamma'(t) B_2^{-1} Z_{ik} \mu_{\bar{d},ik},\\
\delta_{d,ik}(t) &= \int_{t_0}^{t} \eta(u, W_{ik}) \left\{I(T_{ik} \leq u) - G(u|W_{ik})\right\} du\\
&\quad + \left\{\frac{I(t_0 \leq T_{ik} \leq t)}{\overline{g}(T_{ik})} - \int_{t_0}^{t} \frac{p(u|W_{ik})}{\overline{g}(u)}\right\} p(W_{ik}),\\
\tilde{\delta}_{d,ik} &= E_{XT}\left[X \delta_{d,ik}(T)\right], \;\; \breve{\tau}_{\bar{d},ik}(t) = E_W\left[\tau_{\bar{d},ik}\left(H^{-1}(t-W)\right)\right],\\
\breve{\delta}_{d,ik}(t) &= E_W\left[\delta_{d,ik}(H^{-1}(t-W))\right].
\end{aligned}$$

The following are the components of the asymptotical expansion for $\widehat{\theta}$, $\widehat{H}(t)$, $\widehat{F}(t)$ and $\widehat{ROC}(t,x)$,

$$\begin{aligned}
\Delta_{\bar{d},\theta,i} &= \sum_{k=1}^{m_i} \left\{A_2 B_2^{-1} Z_{ik} \mu_{\bar{d},ik} - \tilde{\tau}_{\bar{d},ik}\right\},\\
\Delta_{d,\theta,i} &= \sum_{k=1}^{m_i} \left\{\tilde{\delta}_{d,ik} + \varepsilon_{ik} X_{ik}\right\},\\
\Delta_{d,H,i}(t) &= \sum_{k=1}^{m_i} \left\{\delta_{d,ik}(t) - \pi'(t) A_1^{-1}\left[\tilde{\delta}_{d,ik} + \varepsilon_{ik} X_{ik}\right]\right\},\\
\Delta_{\bar{d},H,i}(t) &= \sum_{k=1}^{m_i} \left\{\pi'(t) A_1^{-1}\left\{A_2 B_2^{-1} Z_{ik} \mu_{\bar{d},ik} - \tilde{\tau}_{\bar{d},ik}\right\} - \delta_{\bar{d},ik}(t)\right\},\\
\Delta_{d,F,i}(t) &= \sum_{k=1}^{m_i} \Big\{f(t) \breve{\delta}_{d,ik}(t) - A_3'(t) A_1^{-1}\left(f(t) \tilde{\delta}_{d,ik} + \varepsilon_{ik} X_{ik}\right)\\
&\quad - (I(\varepsilon_{ik} \geq t) - F(t))\Big\},\\
\Delta_{\bar{d},F,i}(t) &= \sum_{k=1}^{m_i} \left\{\breve{\tau}_{\bar{d},ik}(t) - A_3'(t) A_1^{-1} \tilde{\tau}_{\bar{d},ik} + \left\{A_4'(t) + A_3'(t) A_1^{-1} A_2\right\} B_2^{-1} Z_{ik} \mu_{\bar{d},ik}\right\},\\
\Delta_{d,R,i}(t,x) &= \sum_{k=1}^{m_i} \Big\{f(v) A_6(t,x)\left[\tilde{\delta}_{d,ik} + \varepsilon_{ik} X_{ik}\right] + f(v) \delta_{d,ik}(t)
\end{aligned}$$

$$-f(v)E_X\left[\delta_{d,ik}(H^{-1}(v-X'\theta))\right]+\left[I(\varepsilon_{ik}\geq v)-F(v)\right]\Big\},$$

$$\Delta_{\bar{d},R,i}(t,x)=\sum_{k=1}^{m_i}\left\{A_5(t,x)B_2^{-1}Z_{ik}\mu_{\bar{d},ik}+A_6(t,x)\tilde{\tau}_{\bar{d},ik}-\psi_{\bar{d},ik}(t)-\breve{\tau}_{\bar{d},ik}(t)\right\},$$

where $v=\theta'x+H(t)$.

A.3. Lemmas

Lemma A.1. Let

$$\Upsilon_n(y)=\frac{1}{N_d}\sum_{i=1}^{n_d}\sum_{k=1}^{m_i}\int_{y_0}^{y}\frac{S(u)}{b}K(\frac{\zeta_{ik}-u}{b})du,$$

and

$$\vartheta_n(y)=\frac{1}{N_d}\sum_{i=1}^{n_d}\sum_{k=1}^{m_i}I(y_0\leq\zeta_{ik}\leq y)S(\zeta_{ik}),$$

where K is bounded and symmetric density function with support $[-1,1]$. Here the derivative, $S^{(r)}$, exists and is uniformly bounded over $y\in[y_0,y_1]$, and r is the order of K. If $n_d b^{2r}\to 0$,then $\Upsilon_n(y)-\vartheta_n(y)=o_p(n_d^{-1/2})$ uniformly in $y\in[y_0,y_1]$.

Lemma A.2. Under $A.1$, we have

$$(-1)^{k_0}\frac{1}{b_0^{k_0+1}}ES_1(W)K_0^{(k_0)}(\frac{W-w}{b_0})=(S_1(w)p(w))^{(k)}+O(b_0^{s_0}),$$

$$\frac{(-1)^{k_0+k_1}}{b_0^{k_0+1}b_1^{k_1+1}}ES_2(W,T)K_1^{(k_1)}(\frac{T-t}{b_1})K_0^{(k_0)}(\frac{W-w}{b_0})$$
$$=\{S_2(w,t)p(w,t)\}^{(k_1,k_0)}+O(b_0^{s_0}+b_1^{s_1}),$$

for $k_0,k_1=0,1$, where the derivatives, $S_1^{(s_0+1)}(w)$ and $S_2^{(s_0+1,s_1+1)}(w,t)$, exist and are uniformly bounded over $w\in\Theta$ and $t\in[t_0,t_1]$.

The proofs of Lemma A.1 and Lemma A.2. See Horowitz (1996) and Zhou, Lin, and Johnson (2009).

Lemma A.3. Under conditions given in Cai and Pepe (2002), we have

$$\widehat{\gamma}-\gamma=\frac{B_2^{-1}}{N_{\bar{d}}}\sum_{i=n_d+1}^{n}\sum_{k=1}^{m_i}Z_{ik}\mu_{\bar{d},ik}+o_p(n_{\bar{d}}^{-1/2}),\ \text{and}$$

$$\widehat{S}_{\bar{d}}(c)-S_{\bar{d}}(c)=\frac{1}{N_{\bar{d}}}\sum_{i=n_d+1}^{n}\sum_{k=1}^{m_i}\Big\{e_{\bar{d},ik}(S_{\bar{d}}(c))+S_{\bar{d}}^{(1)}(c)(B_1'B_2^{-1}Z_{ik})\mu_{\bar{d},ik}\Big\}$$
$$+o_p(n_{\bar{d}}^{-1/2}).$$

The Proof of Lemma A.3. See the proof of Lemma A.1 in Cai and Pepe (2002).

Lemma A.4 establishes asymptotic forms of $\widehat{p}(t, x'\widehat{\theta})$ and $\widehat{\mathrm{g}}(t, x'\widehat{\theta})$, respectively, the estimates of $p(t, x'\theta)$ and $\mathrm{g}(t, x'\theta)$. These are used in proving Lemma A.5 and Lemma A.

Lemma A.4. With $\widehat{W} = X'\widehat{\theta}$ and $\widehat{W}_{ik} = X'_{ik}\widehat{\theta}$, let

$$
\begin{aligned}
\widehat{p}(t,w) &= \frac{1}{N_d b_0 b_1}\sum_{i=1}^{n_d}\sum_{k=1}^{m_i} K_1\Big(\frac{\widehat{T}_{ik}-t}{b_1}\Big)K_0\Big(\frac{X'_{ik}\widehat{\theta}-w}{b_0}\Big),\\
p_n(t,w) &= \frac{1}{N_d b_0 b_1}\sum_{i=1}^{n_d}\sum_{k=1}^{m_i} K_1\Big(\frac{T_{ik}-t}{b_1}\Big)K_0\Big(\frac{X'_{ik}\theta-w}{b_0}\Big),\\
\widehat{p}(w) &= \frac{1}{N_d b_0}\sum_{i=1}^{n_d}\sum_{k=1}^{m_i} K_0\Big(\frac{X'_{ik}\widehat{\theta}-w}{b_0}\Big),\\
p_n(w) &= \frac{1}{N_d b_0}\sum_{i=1}^{n_d}\sum_{k=1}^{m_i} K_0\Big(\frac{X'_{ik}\theta-w}{b_0}\Big),\\
\widehat{G}(t,w) &= \frac{1}{N_d b_0}\sum_{i=1}^{n_d}\sum_{k=1}^{m_i} I(\widehat{T}_{ik}\le t)K_0\Big(\frac{X'_{ik}\widehat{\theta}-w}{b_0}\Big),\\
G_n(t,w) &= \frac{1}{N_d b_0}\sum_{i=1}^{n_d}\sum_{k=1}^{m_i} I(T_{ik}\le t)K_0\Big(\frac{X'_{ik}\theta-w}{b_0}\Big),\\
\Lambda_1(t,w,x) &= -p(t,w)\left[q_0(w)-x\right],\\
\Lambda_2(w,x) &= -p(w)[q_0(w)-x],\\
\Lambda_3(t,w,x) &= -G(t|w)p(w)[q_0(w)-x],\\
P_1(t,w) &= -\sum_{i=n_d+1}^{n}\sum_{k=1}^{m_i}\Big\{e_{\bar{d},ik}(t)+(B_1-q_1(w))'S^{(1)}_{\bar{d}}(S^{-1}_{\bar{d}}(t))B_2^{-1}Z_{ik}\mu_{\bar{d},ik}\Big\}\,p(t|w),\\
P_2(t,w) &= -\sum_{i=n_d+1}^{n}\sum_{k=1}^{m_i}\Big\{\frac{1}{b_4}\int_{S^{-1}_{\bar{d}}(t)}^{c_0} K_4\Big(\frac{\mu_{\bar{d},ik}-u}{b_4}\Big)du - t\\
&\qquad +(B_1-q_1(w))'S^{(1)}_{\bar{d}}(S^{-1}_{\bar{d}}(t))B_2^{-1}Z_{ik}\mu_{\bar{d},ik}\Big\}p(t|w),
\end{aligned}
$$

where $c_0 = S^{-1}_{\bar{d}}(t_0)$, K_4 is a kernel function defined on $[-1,1]$, b_4 is a bandwidth, satisfying $n_{\bar{d}}b_4^{2s_4} \to 0$, and s_4 is the order of K_4. Under A.1,

$$
\begin{aligned}
p(t, x'\widehat{\theta}) &= p_n(t, x'\theta) + (\widehat{\theta}-\theta)'\Lambda_1^{(010)}(t, x'\theta, x) + \frac{1}{N_{\bar{d}}}P_2^{(10)}(t, x'\theta)p(x'\theta)\\
&\qquad + o_p(n^{-1/2}),
\end{aligned}
$$

$$\widehat{p}^{(r)}(x'\widehat{\theta}) = p_n^{(r)}(x'\theta) + (\widehat{\theta}-\theta)'\Lambda_2^{(r+1,0)}(x'\theta, x) + o_p(n^{-1/2}), \qquad \text{and}$$

$$\widehat{G}^{(0r)}(t, x'\widehat{\theta}) = G_n^{(0r)}(t, x'\theta) + (\widehat{\theta}-\theta)'\Lambda_3^{(0,r+10)}(t, x'\theta, x) + \frac{1}{N_{\bar{d}}}\frac{\partial^r \left\{p(x'\theta)P_1(t, x'\theta)\right\}}{\partial(x'\theta)^r} + o_p(n^{-1/2}),$$

uniformly over $t \in [t_0, t_1]$ and x in the bounded support of X.

Proof of Lemma A.4. Noting that

$$\widehat{T}_{ik} - T_{ik} = -S_0^{(1)}(\mu_{d,ik})Z'_{ik}(\widehat{\gamma}-\gamma) + \left\{\widehat{S}_{\bar{d}}(\mu_{d,ik}) - S_{\bar{d}}(\mu_{d,ik})\right\}, \qquad \text{(A.2)}$$

where $\mu_{d,ik} = Y_{d,ik} - \gamma' Z_{ik}$, using Lemma A.3 and following the arguments of Lemma 3 in Zhou, Lin, and Johnson (2009), we can complete the proof of Lemma A.4.

Lemma A.5. Let $p(t, w)$ denote the joint density function of (T, W),

$$\Psi_{n1}(t) = \frac{1}{N_d^2 b_0 \bar{\mathrm{g}}(t)} \sum_{j=1}^{n_d}\sum_{\ell=1}^{m_j}\sum_{i=1}^{n_d}\sum_{k=1}^{m_i} \Bigg\{ b_1^{-1}K_1\Big(\frac{T_{ik}-t}{b_1}\Big) + \frac{h(t)p^{(1)}(W_{\bar{d},j\ell})}{p(W_{\bar{d},j\ell})}\left(I(T_{ik} \le t) - G(t|W_{\bar{d},j\ell})\right)\Bigg\} K_0\Big(\frac{W_{ik}-W_{\bar{d},j\ell}}{b_0}\Big),$$

$$\Psi_{n2}(t) = \frac{h(t)}{N_d^2 b_0^2 \bar{\mathrm{g}}(t)} \sum_{j=1}^{n_d}\sum_{\ell=1}^{m_j}\sum_{i=1}^{n_d}\sum_{k=1}^{m_i} (I(T_{ik} \le t) - G(t|W_{\bar{d},j\ell}))K_1^{(1)}\Big(\frac{W_{ik}-W_{\bar{d},j\ell}}{b_0}\Big),$$

$$\Psi_{n3}(t) = \frac{1}{N_{\bar{d}}\bar{\mathrm{g}}(t)} E_W\left\{\left[P_2^{(10)}(t, W) - h(t)P_1^{(01)}(t, W)\right]p(W)\right\},$$

and $\Sigma(t) = h(t)E[\mathrm{g}(t, W)q_0^{(1)}(W)]/\bar{\mathrm{g}}(t)$. Under A.1,

$$\frac{\sum_{i=1}^{n_d}\sum_{k=1}^{m_i}\widehat{p}(t, \widehat{W}_{ik})}{\sum_{i=1}^{n_d}\sum_{k=1}^{m_i}\widehat{\mathrm{g}}(t, \widehat{W}_{ik})} - \frac{\sum_{i=1}^{n_d}\sum_{k=1}^{m_i}p(t, W_{ik})}{\sum_{i=1}^{n_d}\sum_{k=1}^{m_i}\mathrm{g}(t, W_{ik})} = \Psi_{n1}(t) + \Psi_{n2}(t) + \Psi_{n3}(t) + \Sigma(t)'(\widehat{\theta}-\theta) + o_p(n^{-1/2}),$$

uniformly in $t \in [t_0, t_1]$.

Proof of Lemma A.5. Take $G(t, w) = G(t|w)p(w)$, $\widehat{\mathrm{g}}(t, w) = \frac{\partial \widehat{G}(t|w)}{\partial w}\widehat{p}(w)$. Since

$$\widehat{\mathrm{g}}(t, x'\widehat{\theta}) = \widehat{G}^{(01)}(t, x'\widehat{\theta}) - \frac{\widehat{G}(t, x'\widehat{\theta})\widehat{p}^{(1)}(x'\widehat{\theta})}{\widehat{p}(x'\widehat{\theta})},$$

$$\mathrm{g}(t, x'\theta) = G^{(01)}(t, x'\theta) - \frac{G(t, x'\theta)p^{(1)}(w)}{p(w)},$$

and $p(t, w) = h(t)\mathrm{g}(t, w)$, Lemma A.5 follows by combining Lemma A.4, the conditions on b_0 and b_1, and some tedious computation.

Lemma A. *Under the conditions given in Appendix A.1, as $n_{\bar{d}} \to \infty$ and $n_d \to \infty$,*

$$\widehat{H}(t) - H(t) = \frac{1}{N_d}\sum_{i=1}^{n_d}\sum_{k=1}^{m_i}\delta_{d,ik}(t) - \frac{1}{N_{\bar{d}}}\sum_{i=n_d+1}^{n}\sum_{k=1}^{m_i}\delta_{\bar{d},ik}(t) + (\widehat{\theta}-\theta)'\pi(t) + o_p(|\widehat{\theta}-\theta|), \tag{A.3}$$

uniformly over $t \in [t_0, t_1]$.

Proof of Lemma A. Since $h(u) = p(u, w)/\mathrm{g}(u, w)$, by (2.5) and Lemma A.5, we have

$$\widehat{H}(t) - H(t) = \int_{t_0}^{t}\Psi_{n1}(u)du + \int_{t_0}^{t}\Psi_{n2}(u)du + \int_{t_0}^{t}\Psi_{n3}(u)du + (\widehat{\theta}-\theta)'\int_{t_0}^{t}\Sigma(u)du + o_p(n^{-1/2}). \tag{A.4}$$

Exchanging the summations in $\Psi_{n1}(t)$ and $\Psi_{n2}(t)$, using Lemma A.2 and the conditions on b_0, we can show that

$$\begin{aligned}
&\Psi_{n1}(t) + \Psi_{n2}(t)\\
&= \frac{1}{N_d\bar{\mathrm{g}}(t)}\sum_{i=1}^{n_d}\sum_{k=1}^{m_i}p(W_{ik})\Big\{b_1^{-1}K_1\Big(\frac{T_{ik}-t}{b_1}\Big) + \frac{h(t)p^{(1)}(W_{ik})}{p(W_{ik})}\Big(I(T_{ik}\le t) - G(t|W_{ik})\Big)\Big\}\\
&\quad + \frac{h(t)}{N_d\bar{\mathrm{g}}(t)}\sum_{i=1}^{n_d}\sum_{k=1}^{m_i}\Big\{p^{(1)}(W_{ik})(I(T_{ik}\le t) - G(t|W_{ik})) - \mathrm{g}(t, W_{ik})\Big\} + o_p(n^{-1/2}),
\end{aligned} \tag{A.5}$$

uniformly over $t \in [t_0, t_1]$. Consider $\Psi_{n3}(u)$. Let

$$\tilde{p}_{1ik}(t, w) = -e_{\bar{d},ik}(t) - (B_1 - q_1(w))'S_{\bar{d}}^{(1)}(S_{\bar{d}}^{-1}(t))B_2^{-1}Z_{ik}\mu_{\bar{d},ik},$$

and

$$\begin{aligned}
\tilde{p}_{2ik}(t, w) = &-\Big\{\frac{1}{b_4}\int_{S_{\bar{d}}^{-1}(t)}^{c_0}K_4\left(\frac{\mu_{\bar{d},ik}-u}{b_4}\right)du - t\Big\}\\
&-(B_1 - q_1(w))'S_{\bar{d}}^{(1)}(S_{\bar{d}}^{-1}(t))B_2^{-1}Z_{ik}\mu_{\bar{d},ik}.
\end{aligned}$$

Then

$$\Psi_{n3}(t) = \frac{1}{N_{\bar{d}}\bar{\mathrm{g}}(t)}\sum_{i=n_d+1}^{n}\sum_{k=1}^{m_i}E_W\left[\tilde{p}_{2ik}^{(10)}(t, W)p(t, W) + \tilde{p}_{2ik}(t, W)p^{(10)}(t, W)\right.$$

$$-h(t)\tilde{p}_{1ik}^{(01)}(t,W)p(t,W)-h(t)\tilde{p}_{1ik}(t,W)p^{(01)}(t|W)p(W)\Big]$$
$$\equiv\frac{1}{\overline{\mathrm{g}}(t)}\left\{\Psi_{n31}(t)+\Psi_{n32}(t)-\Psi_{n33}(t)-\Psi_{n34}(t)\right\}. \tag{A.6}$$

It can be shown that

$$\begin{aligned}&\Psi_{n33}(t)+\Psi_{n34}(t)\\&=\frac{1}{N_{\bar{d}}}\sum_{i=n_d+1}^{n}\sum_{k=1}^{m_i}h(t)E_W\Big[(q_z^{(1)}(W))'S_{\bar{d}}^{(1)}(S_{\bar{d}}^{-1}(t))B_2^{-1}Z_{ik}\mu_{\bar{d},ik}p(t,W)\\&\quad-\Big\{e_{\bar{d},ik}(t)+(B_1-q_1(W))'S_{\bar{d}}^{(1)}(S_{\bar{d}}^{-1}(t))B_2^{-1}Z_{ik}\mu_{\bar{d},ik}\Big\}p^{(01)}(t|W)p(W)\Big].\end{aligned} \tag{A.7}$$

By Lemma A.1 and assumption that $n_{\bar{d}}b_4^{2s_4}\to 0$, we have

$$\begin{aligned}\Psi_{n32}(t)&=-\frac{1}{N_{\bar{d}}}\sum_{i=n_d+1}^{n}\sum_{k=1}^{m_i}E_W\Big[\Big\{e_{\bar{d},ik}(t)+(B_1-q_1(W))'S_{\bar{d}}^{(1)}(S_{\bar{d}}^{-1}(t))B_2^{-1}Z_{ik}\mu_{\bar{d},ik}\Big\}\\&\qquad\times p^{(10)}(t,W)\Big]+o_p(n^{-1/2}).\end{aligned} \tag{A.8}$$

In addition,

$$\begin{aligned}&\int_{t_0}^{t}\frac{\Psi_{n31}(u)}{\overline{\mathrm{g}}(u)}du\\&=-\frac{1}{N_{\bar{d}}}\sum_{i=n_d+1}^{n}\sum_{k=1}^{m_i}E_W\left\{-\int_{t_0}^{t}\frac{p(u,W)K_4((\mu_{\bar{d},ik}-S_{\bar{d}}^{-1}(u))/b_4)}{b_4\overline{\mathrm{g}}(u)S_{\bar{d}}^{(1)}(S_{\bar{d}}^{-1}(u))}du-\int_{t_0}^{t}\frac{p(u,W)}{\overline{\mathrm{g}}(u)}du\right\}\\&\quad-\frac{1}{N_{\bar{d}}}\sum_{i=n_d+1}^{n}\sum_{k=1}^{m_i}E_W\left\{\int_{t_0}^{t}\frac{p(u,W)S_{\bar{d}}^{(2)}(S_{\bar{d}}^{-1}(u))}{\overline{\mathrm{g}}(u)S_{\bar{d}}^{(1)}(S_{\bar{d}}^{-1}(u))}du(B_1-q_1(W))'B_2^{-1}Z_{ik}\mu_{\bar{d},ik}\right\}.\end{aligned} \tag{A.9}$$

Note that $S_{\bar{d}}(\mu_{\bar{d},ik})$ is the uniform distribution on $[t_0,t_1]$. By Lemma A.1, we can show that

$$-\frac{1}{N_{\bar{d}}}\sum_{i=n_d+1}^{n}\sum_{k=1}^{m_i}E_W\left\{\int_{t_0}^{t}\frac{p(u,W)K_4((\mu_{\bar{d},ik}-S_{\bar{d}}^{-1}(u))/b_4)}{b_4\overline{\mathrm{g}}(u)S_{\bar{d}}^{(1)}(S_{\bar{d}}^{-1}(u))}du\right\}$$

in (A.9) can be replaced by

$$\frac{1}{N_{\bar{d}}}\sum_{i=n_d+1}^{n}\sum_{k=1}^{m_i}E_W\left\{\frac{I(t_0\le S_{\bar{d}}(\mu_{\bar{d},ik})\le t)p(S_{\bar{d}}(\mu_{\bar{d},ik}),W)}{\overline{\mathrm{g}}(S_{\bar{d}}(\mu_{\bar{d},ik}))}\right\}.$$

Then, by (A.6), (A.7), (A.8), (A.9), and some algebraic calculations, we can show that

$$\int_{t_0}^{t} \Psi_{n3}(u)du = -\frac{1}{N_{\bar{d}}} \sum_{i=n_d+1}^{n} \sum_{k=1}^{m_i} \Big\{ I(t_0 \le S_{\bar{d}}(\mu_{\bar{d},ik}) \le t) h(S_{\bar{d}}(\mu_{\bar{d},ik})) - H(t)$$
$$+\Gamma(t) B_2^{-1} Z_{ik} \mu_{\bar{d},ik} + \int_{t_0}^{t} \lambda(u) e_{\bar{d},ik}(u) du \Big\}. \tag{A.10}$$

Furthermore, noting that the term

$$\frac{1}{N_d} \sum_{i=1}^{n_d} \sum_{k=1}^{m_i} \int_{t_0}^{t} \frac{1}{b_1 \bar{g}(u)} K_1 \left(\frac{T_{ik} - u}{b_1} \right) p(W_{ik}) du$$

in $\int \{\Psi_{n1}(u) + \Psi_{n2}(u)\} du$ can be replaced by $(1/N_d) \sum_{i=1}^{n_d} \sum_{k=1}^{m_i} I(t_0 \le T_{ik} \le t) p(W_{ik}) / \bar{g}(T_{ik})$ by Lemma A.1 and the condition on b_1, we conclude Lemma A from (A.4), (A.5), and (A.10).

A.4. Proof of Theorem 1.

We consider the asymptotic expression form of $\widehat{\theta} - \theta$. By (2.9) and a Taylor expansion, we can write

$$\begin{aligned} \widehat{\theta} - \theta &= -D^{-1} \frac{1}{N_d} \sum_{i=1}^{n_d} \sum_{k=1}^{m_i} h(T_{ik}) (\widehat{T}_{ik} - T_{ik}) X_{ik} \\ &\quad -D^{-1} \frac{1}{N_d} \sum_{i=1}^{n_d} \sum_{k=1}^{m_i} \left(\widehat{H}(T_{ik}) - H(T_{ik}) \right) X_{ik} \\ &\quad -D^{-1} \frac{1}{N_d} \sum_{i=1}^{n_d} \sum_{k=1}^{m_i} \varepsilon_{ik} X_{ik} + o_p(n_d^{-1/2}) \\ &\equiv C_{n1} + C_{n2} + C_{n3} + o_p(n_d^{-1/2}), \end{aligned} \tag{A.11}$$

where $D = E[XX']$. By (A.2), substituting Lemma A.3 into C_{n1} and exchanging the summations, we get

$$C_{n1} \approx \frac{D^{-1}}{N_{\bar{d}}} \sum_{i=n_d+1}^{n} \sum_{k=1}^{m_i} E_{XT} h(T) X \Big\{ S_0^{(1)}(S_{\bar{d}}^{-1}(T))(Z - B_1)' B_2^{-1} Z_{ik} \mu_{\bar{d},ik} - e_{\bar{d},ik}(T) \Big\}.$$

Substituting (A.3) into C_{n2}, exchanging the summations and using Condition 4, we get

$$C_{n2} \approx -D^{-1} \frac{1}{N_d} \sum_{j=1}^{n_d} \sum_{\ell=1}^{m_j} E_{XT} \left[X \delta_{\bar{d},ik}(T) \right] - D^{-1} E \left[X \pi'(T) \right] (\widehat{\theta} - \theta)$$

$$+D^{-1}\frac{1}{N_{\bar{d}}}\sum_{i=n_d+1}^{n}\sum_{k=1}^{m_i}E_{XT}\left[\delta_{\bar{d},ik}(T)X\right].$$

Theorem 1 follows by substituting the expressions of C_{n1}, C_{n2} and C_{n3} into (A.11).

A.5. Proof of Theorem 3

Denote the compact support of $H(T)+X'\theta$ for $T\in[t_0,t_1]$ by $[e_0,e_1]$. For any $t\in[e_0,e_1]$, take

$$\widehat{F}(t)=\frac{1}{N_d}\sum_{i=1}^{n_d}\sum_{k=1}^{m_i}I(\widehat{H}(\widehat{T}_{ik})+X_{ik}'\widehat{\theta}\le t),\ \text{ and }\ F_n(t)=\frac{1}{N_d}\sum_{i=1}^{n_d}\sum_{k=1}^{m_i}I(\varepsilon_{ik}\le t).$$

Using the argument as in Lemma 9 of Horowitz (1996), we get that

$$N_d^{1/2}\sup_{t\in[e_0,e_1]}|\left(\widehat{F}(t)-\tilde{E}I(\widehat{H}(\widehat{T}_{ik})+X_{ik}'\widehat{\theta}\le t)\right)-(F_n(t)-F(t))|\to 0,$$

where $\tilde{E}h(X,Y_d)=\int h(x,y)dP(x,y)$ for any function h, and P is the cumulative distribution function of (X,Y_d). Hence,

$$N_d^{1/2}\left(\widehat{F}(t)-F(t)\right)=D_{n1}(t)+D_{n2}(t)+o_p(1),\tag{A.12}$$

where $D_{n1}(t)=N_d^{1/2}(F_n(t)-F(t))$, and $D_{n2}(t)=N_d^{1/2}(\tilde{E}I(\widehat{H}(\widehat{T}_{ik})+X_{ik}'\widehat{\theta}\le t)-F(t))$. Using the techniques of Horowitz (1996), we get that

$$\begin{aligned}D_{n2}(t)=&-N_d^{1/2}f(t)\int\Big\{h(y)\left(\widehat{S}_{\bar{d}}(S_{\bar{d}}^{-1}(y))-y\right)+\left(\widehat{H}(y)-H(y)\right)\\&+h(y)S_0^{(1)}(S_{\bar{d}}^{-1}(y))(\widehat{\gamma}-\gamma)'E[Z|X'\theta=t-H(y)]\Big\}p(t-H(y))h(y)dy\\&-N_d^{1/2}f(t)E[X'](\widehat{\theta}-\theta).\end{aligned}$$

Substituting Lemma A.3, the expressions of $\widehat{\theta}-\theta$ and $\widehat{H}(w)-H(w)$ into $D_{n2}(t)$, we can show that

$$\begin{aligned}D_{n2}(t)=&-\frac{N_d^{1/2}f(t)}{N_{\bar{d}}}\sum_{i=n_d+1}^{n}\sum_{k=1}^{m_i}\Big\{E_{XT}\left[\tau_{\bar{d},ik}\left(H^{-1}(t-W)\right)-A_3'(t)A_1^{-1}\tau_{\bar{d},ik}(T)X\right]\\&+\left(A_4'(t)+A_3'(t)A_1^{-1}A_2\right)B_2^{-1}Z_{ik}\mu_{\bar{d},ik}\Big\}\\&-\frac{f(t)}{N_d^{1/2}}\sum_{i=1}^{n_d}\sum_{k=1}^{m_i}\Big\{E_{XT}\left[\delta_{d,ik}(H^{-1}(t-W))-A_3'(t)A_1^{-1}\delta_{d,ik}(T)X\right]\\&-A_3'(t)A_1^{-1}\varepsilon_{ik}X_{ik}\Big\}.\end{aligned}\tag{A.13}$$

Theorem 3 then follows from (A.12) and (A.13).

A.6. Proof of Theorem 4

By the definition of $\widehat{ROC}(t,x)$, we have that

$$\begin{aligned}
&\widehat{ROC}(t,x) - ROC(t,x) \\
&\quad = \widehat{F}\left(\widehat{\theta}'x + \widehat{H}(t)\right) - F\left(\widehat{\theta}'x + \widehat{H}(t)\right) + F\left(\widehat{\theta}'x + \widehat{H}(t)\right) - F\left(\theta'x + H(t)\right).
\end{aligned}$$

Then, by Taylor expansions and the smoothing approximation in Lemma A.1, we have

$$\begin{aligned}
&\widehat{ROC}(t,x) - ROC(t,x) \\
&\quad = \left\{\widehat{F}\left(\theta'x + H(t)\right) - F\left(\theta'x + H(t)\right)\right\} \\
&\qquad + f\left(\theta'x + H(t)\right)\left((\widehat{\theta} - \theta)'x + (\widehat{H}(t) - H(t))\right) + o_p(n^{-1/2}). \qquad \text{(A.14)}
\end{aligned}$$

Substituting the expressions of $\widehat{\theta} - \theta$, $\widehat{H}(w) - H(w)$, and $\widehat{F}(t) - F(t)$ into (A.14), we can prove Theorem 4.

References

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

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

Cai, T. and Pepe, M. S. (2002). Semi-parametric ROC Analysis to Evaluate Biomarker for Disease. *J. Amer. Statist. Assoc.* **9**, 1099-107.

Carroll, R. J., Fan, J., Gijbels, I. and Wand, M. P. (1997). Generalized partially linear single-index models. *J. Amer. Statist. Assoc.* **92**, 477-489.

Cheng, S. C., Wei, L. J. and Ying, Z. (1995). Analysis of transformation models with censored data. *Biometrika* **82**, 835-845.

Deheuvels, P. (1977). Estimation non parametrique de la densitepar histogrammes generalises (II), *Pub. Inst. Stat. Univ. Paris XXII*, 1-23.

Eddy, W. F. (1980). Optimum kernel estimators of the mode, *Ann. Statist.* **8**, 870-882.

Fan, J., Lin, H. Z. and Zhou, Y. (2006). Local partial-likelihood estimation for lifetime data. *Ann. Statist.* **34**, 290-325.

Gasser, T. and Muller, H. G. (1979). Kernel estimation of regression functions. In *Smoothing Techniques for Curve Estimation* (Edited by T. Gasser and M. Rosenblatt), 23-68. Springer, Heidelberg.

Hanley J. A. (1998). The robustness of the “binormal” assumptions used in fitting ROC curves. *Medical Decision Making* **8**, 197-203.

Henrik Ramlau-Hansen (1983). Smoothing counting process intensities by means of kernel functions. *Ann. Statist.* **11**, 453-466.

Horowitz, J. L. (1996). Semiparametric estimation of a regression model with an unknown transformation of the dependent variable. *Econometrica* **64**, 103-137.

Metz, C. E., Herman, B. A. and Shen, J. H. (1998) Maximum likelihood estimation of receiver operating characteristic (ROC) curves from continuously-distributed data. *Statist. Medicine* **17**, 1033-1053.

Muller, H. G. (1984). Smooth optimum kernel estimators of densities, regression curves and modes, *Ann. Statist.* **12**, 766-774.

Parzen, M. I., Wei, L. J. and Ying, Z. (1994). A resampling method based on pivotal estimating functions. *Biometrika*, **81**, 341-350.

Pepe, M. S. (1997). A regression modelling framework for ROC curves in medical diagnostic testing. *Biometrika* **84**,595-608.

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

Pepe, M. S. and Cai, T. (2004). The analysis of placement values for evaluating discriminatory measures. *Biometrics* **60**: 528-535.

Thompson, M. L. and Zucchini, W. (1989). On the statistical analysis of ROC curves. *Statist. Medicine* **8**, 1277-1290.

Tian, L. Zucker, D. and Wei, L. J. (2005). On the Cox Model With Time-Varying Regression Coefficients. *J. Amer. Statist. Assoc.* **100**, 172-183.

Tosteson, A. N. A. and Begg, C. B. (1988). A general regression methodology for ROC curve estimation. *Medical Decision Making* **3**, 204-215.

Ye, J. and Duan, N. (1997). Nonparametric $n^{-1/2}$-consistent estimation for the general transformation models. *Ann. Statist.* **25**, 2682-2717.

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

Zhou, X. H., Lin, H. Z. and Johnson, E. (2009). Nonparametric heteroscedastic transformation regression models for skewed data with an application to health care costs. *J. Roy. Statist. Soc. Ser. B* **70**, 1029-1047.

School of Statistics, Southwestern University of Finance and Economics, Chengdu, 611130, China.

E-mail: huazhenlin@hotmail.com

HSR&D Center of Excellence, VA Puget Sound Health Care System, Seattle, WA 98108, USA.

Department of Biostatistics, University of Washington, Seattle, WA, USA.

Beijing International Center for Mathematical Research, Peking University, Beijing, China.

E-mail: azhou@u.washington.edu

Department of Biostatistics, University of California at Los Angeles, Los Angeles, CA 90095, USA.

E-mail: vli@ucla.edu

(Received July 2010; accepted September 2011)