

## A GOODNESS OF FIT TEST FOR MULTIPLICATIVE-INTERCEPT RISK MODELS BASED ON CASE-CONTROL DATA

Biao Zhang

*The University of Toledo*

*Abstract:* Qin and Zhang (1997) considered a goodness-of-fit test for the logistic regression model under a case-control sampling plan on the basis of a Kolmogorov-Smirnov-type statistic. There, however, does not exist a goodness-of-fit test for the multiplicative-intercept risk model or the odds-linear model described in the literature. By extending the work of Qin and Zhang (1997), and by indicating the equivalence of the multiplicative-intercept risk model and a two-sample semi-parametric selection bias model, we propose a Kolmogorov-Smirnov-type statistic to test the validity of the multiplicative-intercept risk model based on case-control data. We also propose a bootstrap procedure for finding the P-values of the proposed test. In addition, we establish some asymptotic results associated with the proposed test statistic and justify the proposed bootstrap procedure. As an application of the proposed test procedure, we consider simulation results and the analysis of two real data sets.

*Key words and phrases:* Biased sampling problem, bootstrap, case-control data, confidence band, Gaussian process, Kolmogorov-Smirnov two-sample statistic, logistic regression, mixture sampling, odds-linear model, prospective analysis, semi-parametric selection bias model, strong consistency, weak convergence.

### 1. Introduction

Let  $Y$  be a binary response variable and  $X$  be the associated  $1 \times p$  covariate vector. Hsieh, Manski and McFadden (1985) introduced the following multiplicative-intercept risk model:

$$\frac{P(Y = 1|X = x)}{1 - P(Y = 1|X = x)} = \theta^* r(x; \beta), \quad \text{or} \quad P(Y = 1|X = x) = \frac{\theta^* r(x; \beta)}{1 + \theta^* r(x; \beta)}, \quad (1.1)$$

where  $P(Y = 1|X = x)$  is the probability of disease given  $1 \times p$  covariate vector  $x = (x_1, \dots, x_p)$ ,  $r(x; \beta)$  is, for fixed  $x$ , a known function from  $\mathbf{R}^p$  to  $\mathbf{R}^+$ ,  $\theta^*$  is an unknown positive scalar, and  $\beta = (\beta_1, \dots, \beta_p)^\tau$  is a  $p$ -vector of unknown parameters. Note that linearity is not assumed in  $r(x; \beta)$ . Note also that model (1.1) reduces to the standard logistic regression model if  $\theta^* = \exp(\alpha^*)$  and

$r(x; \beta) = \exp(x\beta)$ , and reduces to the odds-linear model (Weinberg and Sandler (1991) and Wacholder and Weinberg (1994)) when  $r(x; \beta) = 1 + x\beta$ . By generalizing earlier works of Anderson (1972, 1979), Farewell (1979), and Prentice and Pyke (1979) in the context of logistic regression models, Weinberg and Wacholder (1993) showed that under model (1.1), case-control data can be analyzed by maximum likelihood as if they had arisen prospectively, up to an unidentifiable multiplicative constant which depends on the relative sampling fractions. Moreover, they showed that the prospective analysis leads not only to valid point estimates for  $\beta$ , but to valid estimates of standard errors and likelihood ratio testing under model (1.1). Since there does not exist a goodness-of-fit test for the multiplicative-intercept risk model or the odds-linear model described in the literature, we consider in this paper testing the validity of model (1.1) based on case-control data, as specified below.

Let  $X_1, \dots, X_{n_0}$  be a random sample from  $P(x|Y = 0)$  and, independent of the  $X_i$ , let  $Z_1, \dots, Z_{n_1}$  be a random sample from  $P(x|Y = 1)$ . Write  $\pi = P(Y = 1) = 1 - P(Y = 0)$ . Let  $g(x) = f(x|Y = 0)$  and  $h(x) = f(x|Y = 1)$  be, respectively, the conditional density or frequency functions of  $X$  given  $Y = 0$  and  $Y = 1$ . Then an application of Bayes' rule yields the following two-sample semiparametric model:

$$X_1, \dots, X_{n_0} \stackrel{\text{i.i.d.}}{\sim} g(x), \quad Z_1, \dots, Z_{n_1} \stackrel{\text{i.i.d.}}{\sim} h(x) = \exp[\theta + s(x; \beta)]g(x), \quad (1.2)$$

where  $\theta = \log \theta^* + \log \frac{1-\pi}{\pi}$  and  $s(x; \beta) = \log r(x; \beta)$ . Throughout this paper, let  $G(x)$  and  $H(x)$  be, respectively, the corresponding cumulative distribution functions of  $g(x)$  and  $h(x)$ . Note that model (1.2) includes the two-sample length-biased sampling model (Vardi (1982)) with  $\theta = -\log \mu$  and  $s(x; \beta) \equiv \log x$ . Note also that model (1.2), equivalent to model (1.1), is a two-sample semiparametric selection-bias model with weight functions  $w_1(x, \theta, \beta) = \exp[\theta + s(x; \beta)]$  and  $w_2(x, \theta, \beta) = 1$  depending on the unknown parameters  $\theta$  and  $\beta$ . The  $s$ -sample semiparametric selection-bias model was proposed by Vardi (1985) and was further developed by Gilbert, Lele, and Vardi (1999). Vardi (1982, 1985), Gill, Vardi and Wellner (1988), and Qin (1993) have discussed estimation problems in biased sampling models with known weight functions. Qin and Zhang (1997) considered testing the validity of model (1.2) when  $s(x; \beta) = x\beta$ . Our focus of attention in this paper is to test the validity of model (1.2) for some smooth function  $s(x; \beta)$ .

This paper is structured as follows. In Section 2, we propose our test statistic. In Section 3, we present some asymptotic results including the weak convergence of the proposed test statistic to a Gaussian process. In Section 4, we propose and justify a bootstrap procedure which allows us to find P-values of the proposed test. Also in Section 4, we report some results on analysis of two real data problems. Section 5 concerns constructing confidence bands for  $G$  under model (1.2).

A simulation study is presented in Section 6 to demonstrate the performance of the maximum semiparametric likelihood estimation of  $(\theta, \beta)$ . Finally, proofs of the main theoretical results appear in Section 7.

**2. Methodology**

We first discuss the problem of identifiability of  $(\theta, \beta)$  and  $G$ . According to Theorem 2 of Gilbert, Lele and Vardi (1999), it can be shown that model (1.2) is identifiable if and only if  $s(x_0; \beta') \equiv s(x_0; \beta'')$  for some  $x_0 \in \mathbf{R}^p$  and all  $\beta', \beta'' \in \mathbf{R}^p$  with  $\beta' \neq \beta''$ . In particular, if  $s(x_0; \beta) \equiv 0$  for some  $x_0$  and all  $\beta$ , then model (1.2) is identifiable. Clearly, the logistic regression model with  $s(x; \beta) = x\beta$  and the odds-linear model with  $s(x; \beta) = \log(1 + x\beta)$  are identifiable because  $s(0; \beta) \equiv 0$  for all  $\beta$  in either case.

We next consider semiparametric maximum likelihood estimation of  $(\theta, \beta)$  and  $G$ . To this end, let  $\{T_1, \dots, T_n\}$  denote the pooled sample  $\{X_1, \dots, X_{n_0}; Z_1, \dots, Z_{n_1}\}$  with  $n = n_0 + n_1$ . Throughout this paper, we assume that the partial derivatives  $\frac{\partial r(t; \beta)}{\partial \beta}$  and  $\frac{\partial^2 r(t; \beta)}{\partial \beta \partial \beta^T}$  exist for each  $t$  and  $\beta$ . Based on the observed data in (1.2), we can write the full likelihood function as

$$\mathcal{L}(\theta, \beta, G) = \prod_{i=1}^{n_0} dG(X_i) \prod_{j=1}^{n_1} w_1(Z_j, \theta, \beta) dG(Z_j) = \left\{ \prod_{i=1}^n p_i \right\} \left\{ \prod_{j=1}^{n_1} w_1(Z_j, \theta, \beta) \right\},$$

where  $w_1(x, \theta, \beta) = \exp[\theta + s(x; \beta)]$  and  $p_i = dG(T_i)$ , for  $i = 1, \dots, n$ , are (non-negative) jumps with total mass being unity. By employing the two-step profile maximization approach described in Owen (1988, 1990) and Qin and Lawless (1994), we are led to the following estimates of  $G$  and  $(\theta, \beta)$  :

$$\tilde{G}(t) = \sum_{i=1}^n \tilde{p}_i I_{[T_i \leq t]} = \frac{1}{n_0} \sum_{i=1}^n \frac{I_{[T_i \leq t]}}{1 + \rho \exp[\tilde{\theta} + s(T_i; \tilde{\beta})]} \tag{2.1}$$

and  $(\tilde{\theta}, \tilde{\beta})$  maximizes the profile semiparametric log-likelihood function of  $(\theta, \beta)$  given by

$$\ell(\theta, \beta) = -n \log n_0 - \sum_{i=1}^n \log\{1 + \rho \exp[\theta + s(T_i; \beta)]\} + n_1 \theta + \sum_{j=1}^{n_1} s(Z_j; \beta), \tag{2.2}$$

where  $\rho = n_1/n_0$ . Since the weight functions  $w_1(x, \theta, \beta) = \exp[\theta + s(x; \beta)]$  and  $w_2(x, \theta, \beta) = 1$  are strictly positive and  $\ell(\theta, \beta)$  is identical to the logarithm of the partial likelihood (4.3) of Gilbert, Lele and Vardi (1999) plus  $n \log n_0$ , Theorem 4 of Gilbert, Lele and Vardi (1999) implies that if  $s(x_0; \beta') \equiv s(x_0; \beta'')$  for some  $x_0 \in \mathbf{R}^p$  and all  $\beta', \beta'' \in \mathbf{R}^p$  with  $\beta' \neq \beta''$ , then  $(\tilde{\theta}, \tilde{\beta}, \tilde{G})$  maximizes the full likelihood function  $\mathcal{L}(\theta, \beta, G)$ . Consequently,  $\mathcal{L}(\theta, \beta, G)$  will have a unique maximum

if  $\ell(\theta, \beta)$  has a unique maximum. Let  $s_1(t; \beta) = \frac{\partial s(t; \beta')}{\partial \beta'}|_{\beta'=\beta}$ . If  $\frac{s_1(t; \beta)}{s(t; \beta)}$  is not degenerate at the unit vector,  $\min(n_0, n_1) > 0$ , and  $\frac{\partial^2 s(t; \beta)}{\partial \beta \partial \beta^T} = 0$ , then Theorem 5 of Gilbert, Lele and Vardi (1999) implies that the profile semiparametric log-likelihood function  $\ell(\theta, \beta)$  defined in (2.2) is strictly concave over the parameter space in which  $\beta$  lives. Furthermore, under these three conditions, if  $(\tilde{\theta}, \tilde{\beta})$  is a solution to the following system of score equations:

$$\begin{aligned} \frac{\partial \ell(\theta, \beta)}{\partial \theta} &= n_1 - \sum_{i=1}^n \frac{\rho \exp[\theta + s(T_i; \beta)]}{1 + \rho \exp[\theta + s(T_i; \beta)]} = 0, \\ \frac{\partial \ell(\theta, \beta)}{\partial \beta} &= \sum_{j=1}^{n_1} s_1(Z_j; \beta) - \sum_{i=1}^n \frac{\rho \exp[\theta + s(T_i; \beta)]}{1 + \rho \exp[\theta + s(T_i; \beta)]} s_1(T_i; \beta) = 0, \end{aligned} \quad (2.3)$$

then  $(\tilde{\theta}, \tilde{\beta})$  uniquely maximizes  $\ell(\theta, \beta)$ . In this case, we call  $(\tilde{\theta}, \tilde{\beta}, \tilde{G})$  the maximum semiparametric likelihood estimator of  $(\theta, \beta, G)$  under model (1.2). It is easy to see that the logistic regression model with  $s(x; \beta) = x\beta$  satisfies the above three conditions. However, the odds-linear model with  $s(x; \beta) = \log(1 + x\beta)$  does not satisfy  $\frac{\partial^2 s(t; \beta)}{\partial \beta \partial \beta^T} = 0$  and, in this case, a solution to (2.3) may be a local maximum. Although asymptotically a local maximum is not a problem (since only local statistical properties are considered when the sample size is sufficiently large), it can be considerably biased with a larger variance when the sample size is small.

**Remark 2.1.** The two-step profile maximization procedure, by which we derive the maximum semiparametric likelihood estimator  $(\tilde{\theta}, \tilde{\beta}, \tilde{G})$ , is similar to that of Murphy, Rossini, and van der Vaart (1997) in which they considered maximum likelihood estimation of the parameters in the proportional odds model with right-censored data. Both procedures rely on first maximizing the nonparametric part in the full likelihood function with the parametric part fixed, and then maximizing the profile log-likelihood function with respect to the parametric part.

**Remark 2.2.** We can also derive the estimator  $(\tilde{\theta}, \tilde{\beta}, \tilde{G})$  by employing the “plug-in” method of Dabrowska and Doksum (1988) in the context of a two-sample generalized odds-rate model. As indicated below by this approach, the maximum semiparametric likelihood estimator  $(\tilde{\theta}, \tilde{\beta})$  is identical to a method of moments estimator of  $(\theta, \beta)$ . Motivated by the work of Gill, Vardi, and Wellner (1988), let  $F = \frac{n_0}{n}G + \frac{n_1}{n}H$  be the “average distribution function”. Then by (1.2) we have

$$G(t) = n \int \frac{1}{n_0} \frac{1}{1 + \rho \exp[\theta + s(y; \beta)]} I_{[y \leq t]} dF(y).$$

Let  $F_n(t) = \frac{1}{n} \sum_{i=1}^n I_{[T_i \leq t]}$  be the empirical distribution function of the pooled sample  $\{T_1, \dots, T_n\}$ . Then we can estimate  $G$  by

$$G_n(t) = n \int \frac{1}{n_0} \frac{1}{1 + \rho \exp[\theta + s(y; \beta)]} I_{[y \leq t]} dF_n(y) = \frac{1}{n_0} \sum_{i=1}^n \frac{I_{[T_i \leq t]}}{1 + \rho \exp[\theta + s(T_i; \beta)]}$$

with the constraint  $G_n(\infty) = 1$ . Let  $\hat{G}(t) = \frac{1}{n_0} \sum_{i=1}^{n_0} I_{[X_i \leq t]}$  be the empirical distribution function based on the control data  $X_1, \dots, X_{n_0}$ , and let  $\psi(t, \theta, \beta) = (\psi_0(t, \theta, \beta), \psi_1(t, \theta, \beta), \dots, \psi_p(t, \theta, \beta))^T$  be a vector function from  $\mathbf{R}^{2p+1}$  to  $\mathbf{R}^{p+1}$ . Then for a particular choice of  $\psi(t, \theta, \beta)$ , we can estimate  $(\theta, \beta)$  by matching the expectation of  $\psi(t, \theta, \beta)$  under  $G_n$  with that under  $\hat{G}$ :  $\int \psi(t, \theta, \beta) dG_n(t) = \int \psi(t, \theta, \beta) d\hat{G}(t)$ . In other words, we can estimate  $(\theta, \beta)$  by seeking a root to the following system of equations:

$$\frac{1}{n_0} \sum_{i=1}^n \frac{1}{1 + \rho \exp[\theta + s(T_i; \beta)]} \psi(T_i, \theta, \beta) = \frac{1}{n_0} \sum_{i=1}^{n_0} \psi(X_i, \theta, \beta).$$

It is easy to see that the above system of equations reduces to the system of score equations in (2.3) if we take  $\psi(t, \theta, \beta) = (1, s_1^T(t; \beta))^T$ .

The discrepancy between the two estimators  $\tilde{G}(t)$  and  $\hat{G}(t)$  allows us to assess the validity of model (1.2). Thus, the difference

$$\Delta_n(t) = \sqrt{n} |\tilde{G}(t) - \hat{G}(t)|, \quad \Delta_n = \sup_{-\infty \leq t \leq \infty} \Delta_n(t) \tag{2.4}$$

can be employed to measure the departure from the assumption of the multiplicative-intercept risk model (1.1). Note that the test statistic proposed in (2.4) reduces to that of Qin and Zhang (1997) when  $\theta^* = \exp(\alpha^*)$  and  $r(x; \beta) = \exp(x\beta)$  in model (1.1).

**Remark 2.3.** One can estimate  $H(t)$  by  $\tilde{H}(t) = \sum_{i=1}^n \tilde{p}_i \exp[\tilde{\theta} + s(T_i; \tilde{\beta})] I_{[T_i \leq t]}$  based on the case-control data  $T_1, \dots, T_n$  under model (1.2).

**Remark 2.4.** The test statistic  $\Delta_n$  can be applied to mixture sampling data. See Remark 3 of Qin and Zhang (1997).

**Remark 2.5.** In light of Anderson (1972, 1979), we may treat the case-control data as the prospective data to compute the maximum likelihood estimate of  $(\theta^*, \beta)$  under model (1.1). Let  $(\tilde{\theta}, \tilde{\beta})$  denote the solution to the likelihood equations in (2.3) and  $(\hat{\theta}^*, \hat{\beta})$  denote the maximum likelihood estimate of  $(\theta^*, \beta)$  under model (1.1). Then  $\tilde{\theta} = \log \hat{\theta}^* + \log \frac{n_0}{n_1}$  and  $\tilde{\beta} = \hat{\beta}$ . Thus, the maximum likelihood estimates of  $\beta$  are identical under the retrospective sampling scheme and the prospective sampling scheme. In addition, estimated asymptotic variance-covariance matrices for  $\hat{\beta}$  and  $\tilde{\beta}$  based on the observed information coincide.

Weinberg and Wacholder (1993) obtained these results under the assumption that  $X$  in model (1.1) is discrete. Their approach is based on the EM algorithm. Moreover, in the context of the standard logistic regression model with  $\theta^* = \exp(\alpha^*)$  and  $r(x; \beta) = \exp(x\beta)$ , we have  $\tilde{\theta} = \hat{\alpha}^* + \log \frac{n_0}{n_1}$ .

### 3. Asymptotic Results

In this section, we study the asymptotic properties of the proposed estimator  $\tilde{G}(t)$  in (2.1), and the proposed test statistic  $\Delta_n$  in (2.4). To this end, let  $(\theta_0, \beta_0)$  be the true value of  $(\theta, \beta)$  under model (1.2). Throughout this paper, we assume  $\rho = n_1/n_0$  is positive and finite, and remains fixed as  $n = n_0 + n_1 \rightarrow \infty$ . Write

$$\begin{aligned} s_{1k}(t; \beta) &= \frac{\partial s(t; \beta)}{\partial \beta_k} = \frac{1}{r(t; \beta)} \frac{\partial r(t; \beta)}{\partial \beta_k}, \quad k = 1, \dots, p, \quad s_{10}(t; \beta) \equiv 1, \\ s_1(t; \beta) &= \frac{\partial s(t; \beta)}{\partial \beta} = \frac{1}{r(t; \beta)} \frac{\partial r(t; \beta)}{\partial \beta} = (s_{11}(t; \beta), \dots, s_{1p}(t; \beta))^T, \\ s_{2kl}(t; \beta) &= \frac{\partial^2 s(t; \beta)}{\partial \beta_k \partial \beta_l} = \frac{\partial s_{1k}(t; \beta)}{\partial \beta_l} \\ &= \frac{1}{r(t; \beta)} \frac{\partial^2 r(t; \beta)}{\partial \beta_k \partial \beta_l} - \frac{1}{r^2(t; \beta)} \left( \frac{\partial r(t; \beta)}{\partial \beta_k} \right) \left( \frac{\partial r(t; \beta)}{\partial \beta_l} \right), \quad k, l = 1, \dots, p, \\ s_2(t; \beta) &= \frac{\partial^2 s(t; \beta)}{\partial \beta \partial \beta^\tau} = \frac{\partial s_1(t; \beta)}{\partial \beta^\tau} \\ &= \frac{1}{r(t; \beta)} \frac{\partial^2 r(t; \beta)}{\partial \beta \partial \beta^\tau} - \frac{1}{r^2(t; \beta)} \left( \frac{\partial r(t; \beta)}{\partial \beta} \right) \left( \frac{\partial r(t; \beta)}{\partial \beta^\tau} \right) = (s_{2kl})_{k, l = 1, \dots, p}, \\ s_{3klm}(t; \beta) &= \frac{\partial^3 s(t; \beta)}{\partial \beta_k \partial \beta_l \partial \beta_m} = \frac{\partial s_{2kl}(t; \beta)}{\partial \beta_m}, \quad k, l, m = 1, \dots, p, \\ A_{11}(t) &= \int \frac{\exp[\theta_0 + s(y; \beta_0)]}{1 + \rho \exp[\theta_0 + s(y; \beta_0)]} I_{[y \leq t]} dG(y), \quad A_{11} = A_{11}(\infty), \quad S_{11} = \frac{\rho}{1 + \rho} A_{11}, \\ A_{21}(t) &= \int \frac{\exp[\theta_0 + s(y; \beta_0)]}{1 + \rho \exp[\theta_0 + s(y; \beta_0)]} s_1(y; \beta_0) I_{[y \leq t]} dG(y), \quad A_{21} = A_{21}(\infty), \\ S_{21} &= \frac{\rho}{1 + \rho} A_{21}, \\ A_{22} &= \int \frac{\exp[\theta_0 + s(t; \beta_0)]}{1 + \rho \exp[\theta_0 + s(t; \beta_0)]} s_1(t; \beta_0) s_1^\tau(t; \beta_0) dG(t), \quad S_{22} = \frac{\rho}{1 + \rho} A_{22}, \\ A &= \begin{pmatrix} A_{11} & A_{21}^\tau \\ A_{21} & A_{22} \end{pmatrix}, \quad S = \begin{pmatrix} S_{11} & S_{21}^\tau \\ S_{21} & S_{22} \end{pmatrix} = \frac{\rho}{1 + \rho} A, \quad \Sigma = \frac{1 + \rho}{\rho} \left[ A^{-1} - \begin{pmatrix} 1 + \rho & 0 \\ 0 & 0 \end{pmatrix} \right]. \end{aligned}$$

In order to formulate our results, we state some assumptions.

- (A1) There exists a neighborhood  $\Theta_0$  of the true parameter point  $\beta_0$  such that for all  $t$ , the function  $r(t; \beta)$  admits all third derivatives  $\frac{\partial^3 r(t; \beta)}{\partial \beta_k \partial \beta_l \partial \beta_m}$  for all  $\beta \in \Theta_0$ .

- (A2) There exists a function  $Q_1$  such that  $|s_{1k}(t; \beta)| \leq Q_1(t)$  for all  $\beta \in \Theta_0$  and  $k = 1, \dots, p$ , where  $q_{1j} = \int Q_1^j(y) \{1 + \rho \exp[\theta_0 + s(y; \beta_0)]\} dG(y) < \infty$ ,  $j = 1, 2, 3$ .
- (A3) There exists a function  $Q_2$  such that  $|s_{2kl}(t; \beta)| \leq Q_2(t)$  for all  $\beta \in \Theta_0$  and  $k, l = 1, \dots, p$ , where  $q_{2j} = \int Q_2^j(y) \{1 + \rho \exp[\theta_0 + s(y; \beta_0)]\} dG(y) < \infty$ ,  $j = 1, 2$ .
- (A4) There exists a function  $Q_3$  such that  $|s_{3klm}(t; \beta)| \leq Q_3(t)$  for all  $\beta \in \Theta_0$  and  $k, l, m = 1, \dots, p$ , where  $q_3 = \int Q_3(y) \{1 + \rho \exp[\theta_0 + s(y; \beta_0)]\} dG(y) < \infty$ .

The first theorem concerns the strong consistency and the asymptotic distribution of  $(\tilde{\theta}, \tilde{\beta})$ .

**Theorem 3.1.** *Suppose that model (1.2) and Assumptions (A1)–(A4) hold, and that  $A$  is positive definite.*

- (a) *As  $n \rightarrow \infty$ , with probability 1 there exists a sequence  $(\tilde{\theta}, \tilde{\beta})$  of roots of the system of score equations (2.3) such that  $(\tilde{\theta}, \tilde{\beta})$  is strongly consistent for estimating  $(\theta_0, \beta_0)$ .*
- (b) *As  $n \rightarrow \infty$ , one can write*

$$\begin{pmatrix} \tilde{\theta} - \theta_0 \\ \tilde{\beta} - \beta_0 \end{pmatrix} = \frac{1}{n} S^{-1} \begin{pmatrix} \frac{\partial \ell(\theta_0, \beta_0)}{\partial \theta} \\ \frac{\partial \ell(\theta_0, \beta_0)}{\partial \beta} \end{pmatrix} + o_p(n^{-1/2}), \tag{3.1}$$

where  $\frac{\partial \ell(\theta_0, \beta_0)}{\partial \theta} = \frac{\partial \ell(\theta, \beta)}{\partial \theta} |_{(\theta, \beta) = (\theta_0, \beta_0)}$  and  $\frac{\partial \ell(\theta_0, \beta_0)}{\partial \beta} = \frac{\partial \ell(\theta, \beta)}{\partial \beta} |_{(\theta, \beta) = (\theta_0, \beta_0)}$ . As a result,

$$\sqrt{n} \begin{pmatrix} \tilde{\theta} - \theta_0 \\ \tilde{\beta} - \beta_0 \end{pmatrix} \xrightarrow{d} N_{p+1}(\mathbf{0}, \Sigma). \tag{3.2}$$

We now establish the weak convergence of  $\sqrt{n}(\tilde{G} - \hat{G})$  to a Gaussian process by representing  $\tilde{G} - \hat{G}$  as the mean of a sequence of independent and identically distributed stochastic processes with a remainder term of order  $o_p(n^{-1/2})$ .

**Theorem 3.2.** *Under the conditions of Theorem 3.1, one can write  $\tilde{G}(t) - \hat{G}(t) = H_1(t) - \tilde{G}(t) - H_2(t) + R_n(t)$ , where*

$$\begin{aligned} H_1(t) &= \frac{1}{n_0} \sum_{i=1}^n \frac{I_{[T_i \leq t]}}{1 + \rho \exp[\theta_0 + s(T_i; \beta_0)]}, \\ H_2(t) &= \frac{\rho}{n} (A_{11}(t), A_{21}^T(t)) S^{-1} \begin{pmatrix} \frac{\partial \ell(\theta_0, \beta_0)}{\partial \theta} \\ \frac{\partial \ell(\theta_0, \beta_0)}{\partial \beta} \end{pmatrix}, \end{aligned} \tag{3.3}$$

and the remainder term  $R_n(t)$  satisfies  $\sup_{-\infty \leq t \leq \infty} |R_n(t)| = o_p(n^{-1/2})$ . Then  $\sqrt{n}(\tilde{G} - \hat{G}) \xrightarrow{\mathcal{D}} W$  in  $D[-\infty, \infty]^p$ , where  $W$  is a Gaussian process with continuous

sample paths and satisfies, for  $-\infty \leq s, t \leq \infty$ ,

$$EW(t) = 0, \quad EW(s)W(t) = \rho(1+\rho) \left[ A_{11}(s \wedge t) - (A_{11}(s), A_{21}^T(s)) A^{-1} \begin{pmatrix} A_{11}(t) \\ A_{21}(t) \end{pmatrix} \right]. \tag{3.4}$$

Theorem 3.2 forms the basis for testing the validity of model (1.2) on the basis of the test statistic  $\Delta_n$  in (2.4). Let  $w_\alpha$  denote the  $\alpha$ -quantile of the distribution of  $\sup_{-\infty \leq t \leq \infty} |W(t)|$ . According to Theorem 3.2 and the Continuous Mapping Theorem (Billingsley (1968, p.30)), we have

$$\begin{aligned} \lim_{n \rightarrow \infty} P(\Delta_n \geq w_{1-\alpha}) &= \lim_{n \rightarrow \infty} P\left( \sup_{-\infty \leq t \leq \infty} \sqrt{n} |\tilde{G}(t) - G(t)| \geq w_{1-\alpha} \right) \\ &= P\left( \sup_{-\infty \leq t \leq \infty} |W(t)| \geq w_{1-\alpha} \right) = \alpha. \end{aligned}$$

Thus, our proposed goodness-of-fit test procedure has the following decision rule: reject model (1.2) at level  $\alpha$  if  $\Delta_n > w_{1-\alpha}$ . In order for this proposed test procedure to be useful in practice, we need to find the distribution of  $\sup_{-\infty \leq t \leq \infty} |W(t)|$  and be able to calculate the  $(1 - \alpha)$ -quantile  $w_{1-\alpha}$ . Unfortunately, no analytic expressions appear to be available for the distribution function of  $\sup_{-\infty \leq t \leq \infty} |W(t)|$  and the quantile function thereof. Alternatively, one may employ a bootstrap procedure as described in the next section.

**4. A Bootstrap Procedure**

In this section we present a bootstrap procedure which can be employed to approximate the quantile  $w_{1-\alpha}$  defined at the end of the last section. If model (1.2) is valid, we can generate bootstrap data, respectively, from  $\tilde{G}$  and  $\tilde{H}$ , where  $\tilde{G}$  is given by (2.1) and  $\tilde{H}$  is defined in Remark 2.3. Specifically, let  $X_1^*, \dots, X_{n_0}^*$  be a random sample from  $\tilde{G}$  and, independent of the  $X_i^*$ , let  $Z_1^*, \dots, Z_{n_1}^*$  be a random sample from  $\tilde{H}$ . Note that some of the  $X_i^*$  could come from  $Z_1, \dots, Z_{n_1}$  and some of the  $Z_j^*$  could be from  $X_1, \dots, X_{n_0}$ . Let  $\{T_1^*, \dots, T_n^*\}$  denote the combined bootstrap sample  $\{X_1^*, \dots, X_{n_0}^*; Z_1^*, \dots, Z_{n_1}^*\}$  and  $(\tilde{\theta}^*, \tilde{\beta}^*)$  be the solution to the system of score equations in (2.3) with the  $T_i^*$  in place of the  $T_i$ . Moreover, similar to (2.1)–(2.4), let  $\hat{G}^*(t) = \frac{1}{n_0} \sum_{i=1}^{n_0} I_{[X_i^* \leq t]}$  and

$$\tilde{G}^*(t) = \sum_{i=1}^n \tilde{p}_i^* I_{[T_i^* \leq t]} = \frac{1}{n_0} \sum_{i=1}^n \frac{I_{[T_i^* \leq t]}}{1 + \rho \exp[\tilde{\theta}^* + s(T_i^*; \tilde{\beta}^*)]}.$$

Then the corresponding bootstrap version of the test statistic  $\Delta_n$  in (2.4) is given by  $\Delta_n^*(t) = \sqrt{n} |\tilde{G}^*(t) - \hat{G}^*(t)|$  and  $\Delta_n^* = \sup_{-\infty \leq t \leq \infty} \Delta_n^*(t)$ .

We now study the asymptotic behavior of  $(\tilde{\theta}^*, \tilde{\beta}^*)$ ,  $\tilde{G}^*(t)$ , and  $\Delta_n^*$  by deriving the bootstrap versions of the representations and weak convergence as given in

Theorems 3.1 and 3.2. Let  $A_{n11}(t)$ ,  $A_{n21}(t)$ ,  $A_{n11}$ ,  $A_{n21}$ ,  $A_{n22}$ ,  $S_{n11}$ ,  $S_{n21}$ ,  $S_{n22}$ ,  $A_n$ ,  $S_{nn}$ , and  $\Sigma_n$  be the bootstrap versions of previously defined expressions, with  $(\theta_0, \beta_0, G)$  replaced by  $(\tilde{\theta}, \tilde{\beta}, \tilde{G})$ . Furthermore, let

$$\begin{aligned} \frac{\partial \ell^*(\tilde{\theta}, \tilde{\beta})}{\partial \theta} &= n_1 - \sum_{i=1}^n \frac{\rho \exp[\tilde{\theta} + s(T_i^*; \tilde{\beta})]}{1 + \rho \exp[\tilde{\theta} + s(T_i^*; \tilde{\beta})]}, \\ \frac{\partial \ell^*(\tilde{\theta}, \tilde{\beta})}{\partial \beta} &= \sum_{j=1}^{n_1} s_1(Z_j^*; \tilde{\beta}) - \sum_{i=1}^n \frac{\rho \exp[\tilde{\theta} + s(T_i^*; \tilde{\beta})]}{1 + \rho \exp[\tilde{\theta} + s(T_i^*; \tilde{\beta})]} s_1(T_i^*; \tilde{\beta}). \end{aligned}$$

**Theorem 4.1.** *Suppose that model (1.2) and Assumptions (A1)–(A4) hold,  $A$  is positive definite, and  $\int Q_1^2(y)Q_2(y)\{1 + \rho \exp[\theta_0 + s(y; \beta_0)]\}dG(y) < \infty$ .*

(a) *With probability one as  $n \rightarrow \infty$ ,*

$$\begin{pmatrix} \tilde{\theta}^* - \tilde{\theta} \\ \tilde{\beta}^* - \tilde{\beta} \end{pmatrix} = \frac{1}{n} S_{nn}^{-1} \begin{pmatrix} \frac{\partial \ell^*(\tilde{\theta}, \tilde{\beta})}{\partial \theta} \\ \frac{\partial \ell^*(\tilde{\theta}, \tilde{\beta})}{\partial \beta} \end{pmatrix} + o_p^*(n^{-1/2}), \tag{4.1}$$

where  $o_p^*$  stands for  $o_p$  in bootstrap probability under  $\tilde{G}$  or  $\tilde{H}$ , and

$$\sqrt{n} \begin{pmatrix} \tilde{\theta}^* - \tilde{\theta} \\ \tilde{\beta}^* - \tilde{\beta} \end{pmatrix} \xrightarrow{d} N_{p+1}(\mathbf{0}, \Sigma). \tag{4.2}$$

(b) *With probability one as  $n \rightarrow \infty$ ,  $\tilde{G}^*(t) - \hat{G}^*(t) = H_1^*(t) - \hat{G}^*(t) - H_2^*(t) + R_n^*(t)$ , where*

$$\begin{aligned} H_1^*(t) &= \frac{1}{n_0} \sum_{i=1}^n \frac{I_{[T_i^* \leq t]}}{1 + \rho \exp[\tilde{\theta} + s(T_i^*; \tilde{\beta})]}, \\ H_2^*(t) &= \frac{\rho}{n} (A_{n11}(t), A_{n21}^T(t)) S_{nn}^{-1} \begin{pmatrix} \frac{\partial \ell^*(\tilde{\theta}, \tilde{\beta})}{\partial \theta} \\ \frac{\partial \ell^*(\tilde{\theta}, \tilde{\beta})}{\partial \beta} \end{pmatrix}, \end{aligned}$$

and the remainder term  $R_n^*(t)$  satisfies  $\sup_{-\infty \leq t \leq \infty} |R_n^*(t)| = o_p^*(n^{-1/2})$ . With probability one as  $n \rightarrow \infty$ ,  $\sqrt{n}(\tilde{G}^* - \hat{G}^*) \xrightarrow{D} W$  in  $D[-\infty, \infty]^p$ , where  $W$  is the Gaussian process defined in Theorem 3.2.

Theorem 3.2 and part (b) of Theorem 4.1 indicate that the limit process of  $\sqrt{n}(\tilde{G}^* - \hat{G}^*)$  agrees with that of  $\sqrt{n}(\tilde{G} - \hat{G})$ . It follows from the Continuous Mapping Theorem that  $\Delta_n^* = \sup_{-\infty \leq t \leq \infty} \sqrt{n}|\tilde{G}^*(t) - \hat{G}^*(t)|$  has the same limiting behavior as does  $\Delta_n = \sup_{-\infty \leq t \leq \infty} \sqrt{n}|\tilde{G}(t) - G(t)|$ . Thus, we can approximate the quantiles of the distribution of  $\Delta_n$  by those of  $\Delta_n^*$ . This heuristic argument is justified in Theorem 4.2 with Assumption (A5).

(A5) The distribution of  $\sup_{-\infty \leq t \leq \infty} |W(t)|$  is continuous.

By Theorem 1 of Tsirel'son (1975), when  $p = 1$ , this distribution is continuous except perhaps at the lower endpoint of its support. (See also Assumption 4 of Bickel and Krieger (1989)). In case  $p > 1$ , according to Theorem 6.9.2 of Adler (1981), the function  $S(w) = P(\sup_{-\infty \leq t \leq \infty} |W(t)| \geq w)$  is continuous for sufficiently large  $w$ .

**Theorem 4.2.** *Suppose that model (1.2) and Assumptions (A1)–(A5) hold, and  $A$  is positive definite. Fix  $\alpha$  with  $0 < \alpha < 1$ , let  $w_{1-\alpha}^n = \inf\{t; P^*(\Delta_n^* \leq t) \geq 1 - \alpha\}$ , where  $P^*$  stands for the bootstrap probability under  $\tilde{G}$  or  $\tilde{H}$ . Then as  $n \rightarrow \infty$ ,*

$$\lim_{n \rightarrow \infty} P(\Delta_n \geq w_{1-\alpha}^n) = \lim_{n \rightarrow \infty} P\left(\sup_{-\infty \leq t \leq \infty} \sqrt{n}|\tilde{G}(t) - \hat{G}(t)| \geq w_{1-\alpha}^n\right) = \alpha.$$

Theorem 4.2 immediately implies the following bootstrap decision rule: reject model (1.2) at level  $\alpha$  if  $\Delta_n > w_{1-\alpha}^n$ . We now apply the proposed goodness-of-fit test procedure to two real data sets.

**Example 4.1.** The data set in Hosmer and Lemeshow (1989, p.3) gives age and the status of coronary heart disease for 100 subjects who participated in a study. Hosmer and Lemeshow analyzed the relationship between age and the status of coronary heart disease by employing the standard logistic regression model. Qin and Zhang (1997) also analyzed this data set by testing the validity of the logistic regression model. Their analysis supports the goodness-of-fit for the logistic regression model. Here we analyze this data set from a different prospective on the basis of the following odds-linear model:

$$\frac{P(Y = 1|X = x)}{1 - P(Y = 1|X = x)} = \alpha_0 + \alpha_1 x_1 + \cdots + \alpha_p x_p. \quad (4.3)$$

Let  $\theta^* = \alpha_0$  and  $\beta_k = \alpha_k/\alpha_0$  for  $k = 1, \dots, p$ . Then it is seen that model (4.3) is a special case of model (1.1) with  $r(x; \beta) = 1 + x\beta$  and is equivalent to model (1.2) with  $\theta = \log \theta^* + \log \frac{1-\pi}{\pi} = \log \alpha_0 + \log \frac{1-\pi}{\pi}$  and  $s(x; \beta) = \log r(x; \beta) = \log(1 + x\beta)$ . Here we consider the case  $p = 1$ . Let  $X$  denote age and  $Y = 1$  or 0 represent the presence or absence of coronary heart disease. Since the sample data  $(X_i, Y_i)$ ,  $i = 1, \dots, 100$ , can be thought of as being drawn independently from the joint distribution of  $(X, Y)$ , Remark 2.4 implies that we can use the statistic  $\Delta_n$  in (2.4) to test the validity of model (4.3). The system of score equations in (2.3) yields  $(\tilde{\theta}, \tilde{\beta}) = (-7.62, 47.47)$ . A plot (not shown here) of the estimated profile semiparametric log-likelihood function  $\tilde{\ell}(\beta) = \ell(\tilde{\theta}, \beta)$  of  $\beta$  for  $\beta \in [10, 100]$  indicates that  $\tilde{\beta}$  indeed maximizes  $\tilde{\ell}(\beta)$  with  $\tilde{\ell}(\tilde{\beta}) + n \log(n_0) = -50.26353$ . Furthermore, the proposed test statistic  $\Delta_n$  in (2.4) is identical to  $\Delta_n = 1.55$ . Moreover, the observed P-value is found to be 0.005 based on 1000

bootstrap replications of  $\Delta_n^*$ . Consequently, we can conclude that the odds-linear model (4.3) is not appropriate for studying the relationship between age and the status of coronary heart disease for this data set.

Figure 1 shows the curves of  $\tilde{G}$  and  $\hat{G}$  along with the curves  $\tilde{H}$  and  $\hat{H}$  based on this data set. The plot clearly supports our conclusion.

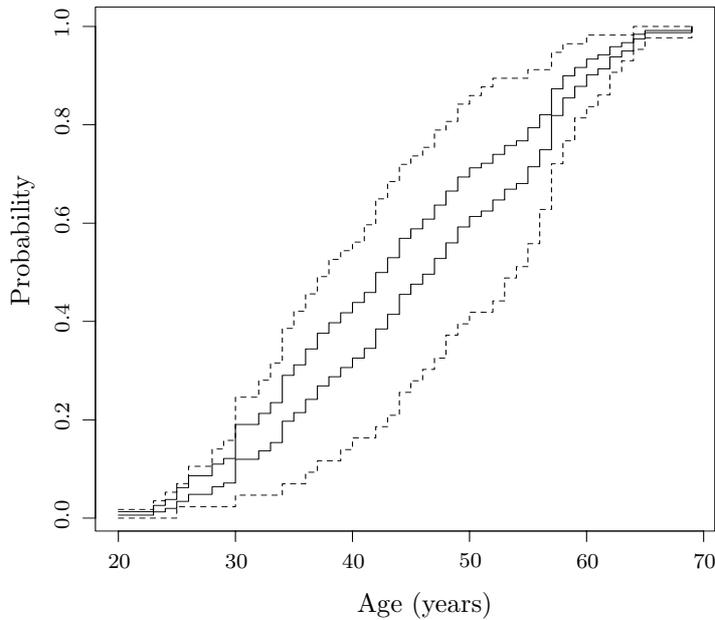


Figure 1. Example 4.1: Estimated cumulative distribution functions  $\tilde{G}, \hat{G}, \tilde{H}$  and  $\hat{H}$ . The solid curve and dashed curve on the upper left represent  $\tilde{G}$  and  $\hat{G}$ , respectively, whereas the solid curve and dashed curve on the lower right stand for  $\tilde{H}$  and  $\hat{H}$ , respectively.

**Example 4.2.** Bliss (1935) reported the number of beetles killed after five hours' exposure to gaseous carbon disulphide at various concentrations. The data are also listed in Table 4.7 of Agresti (1990, p.106). Agresti (1990) analyzed the relationship between beetle mortality and the log dosage by employing the model with complementary log-log link:

$$P(Y = 1|X = x) = 1 - \exp[-\exp(\beta_1 + \beta_2x)]. \tag{4.4}$$

He reported a good fit with a P-value of 0.744. Let  $X$  denote log dosage and let  $Y = 1$  if the beetle dies and  $Y = 0$  if the beetle survives. Then model (1.2) holds with  $\theta = \log \frac{1-\pi}{\pi}$  and  $s(x; \beta) = \log\{1 - \exp[-\exp(\beta_1 + \beta_2x)]\} + \exp(\beta_1 + \beta_2x)$ . Since the sample data  $(Y_i, X_i), i = 1, \dots, 481$ , can be thought of as being drawn independently from the joint distribution of  $(Y, X)$ , Remark 2.4 implies that we

can use the statistic  $\Delta_n$  in (2.4) to test the validity of model (4.4). Under model (1.2), we found  $(\tilde{\theta}, \tilde{\beta}_1, \tilde{\beta}_2) = (-3.21090, -18.47785, 10.96803)$  and  $\Delta_n = 0.24856$  with the observed P-value 0.667 based on 1000 bootstrap replications of  $\Delta_n^*$ , closely agreeing with Agresti's (1990) conclusion.

Figure 4 shows the curves of  $\tilde{G}$  and  $\hat{G}$  along with the curves  $\tilde{H}$  and  $\hat{H}$  based on this data set. Figure 4 demonstrates that the curve of  $\tilde{G}$  ( $\tilde{H}$ ) bears a strong resemblance to that of  $\hat{G}$  ( $\hat{H}$ ), thus further indicating a good fit of model (4.4) or (1.2) to these data.

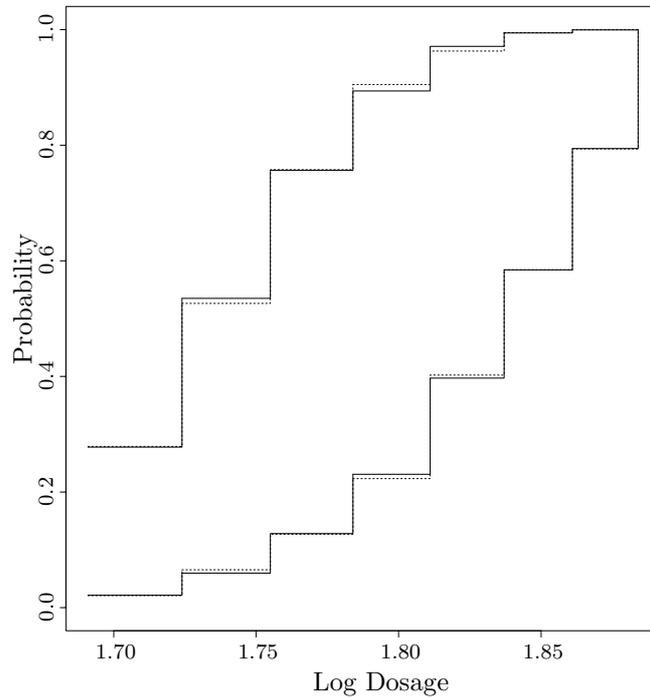


Figure 4. Example 4.2: Estimated cumulative distribution functions  $\tilde{G}$ ,  $\hat{G}$ ,  $\tilde{H}$  and  $\hat{H}$ . The solid curve and dashed curve on the upper left represent  $\tilde{G}$  and  $\hat{G}$ , respectively, whereas the solid curve and dashed curve on the lower right stand for  $\tilde{H}$  and  $\hat{H}$ , respectively.

## 5. Confidence Bands for $G$

In this section we demonstrate that the results of Theorems 3.2 and 4.1 can be adapted to construct confidence bands on  $G$  under model (1.2). To this end, we first establish the weak convergence of the stochastic process  $\sqrt{n}(\tilde{G} - G)$  and its bootstrap counterpart.

**Theorem 5.1.** *Suppose that model (1.2) and Assumptions (A1)–(A4) hold, and  $A$  is positive definite.*

- (a) *As  $n \rightarrow \infty$ ,  $\sqrt{n}(\tilde{G} - G) \xrightarrow{\mathcal{D}} U$  in  $D[-\infty, \infty]^p$ , where  $U$  is a Gaussian process with continuous sample paths and satisfies, for  $-\infty \leq s, t \leq \infty$ ,*

$$EU(t) = 0, \quad EU(s)U(t) = (1 + \rho) \left[ G(s \wedge t) - G(s)G(t) - \rho A_{11}(s \wedge t) + \rho(A_{11}(s), A_{21}^T(s))A^{-1} \begin{pmatrix} A_{11}(t) \\ A_{21}(t) \end{pmatrix} \right].$$

- (b) *With probability one as  $n \rightarrow \infty$ ,  $\sqrt{n}(\tilde{G}^* - \tilde{G}) \xrightarrow{\mathcal{D}} U$  in  $D[-\infty, \infty]^p$ .*

Theorem 5.1 indicates that the limit process of  $\sqrt{n}(\tilde{G}^* - \tilde{G})$  agrees with that of  $\sqrt{n}(\tilde{G} - G)$ , and thus forms the basis for constructing confidence bands for  $G$  under model (1.2). According to the Continuous Mapping Theorem, we have  $\sup_{-\infty \leq t \leq \infty} \sqrt{n}|\tilde{G}(t) - G(t)| \xrightarrow{d} \sup_{-\infty \leq t \leq \infty} |U(t)|$  and  $\sup_{-\infty \leq t \leq \infty} \sqrt{n}|\tilde{G}^*(t) - \tilde{G}(t)| \xrightarrow{d} \sup_{-\infty \leq t \leq \infty} |U(t)|$  almost surely. As a result, we can approximate the quantiles of the distribution of  $\sup_{-\infty \leq t \leq \infty} \sqrt{n}|\tilde{G}(t) - G(t)|$  by those of the distribution of  $\sup_{-\infty \leq t \leq \infty} \sqrt{n}|\tilde{G}^*(t) - \tilde{G}(t)|$ . For  $0 < \alpha < 1$ , let  $u_{1-\alpha}^n = \inf\{t; P^*(\sup_{-\infty \leq t \leq \infty} \sqrt{n}|\tilde{G}^*(t) - \tilde{G}(t)| \leq t) \geq 1 - \alpha\}$ , then a level  $1 - \alpha$  bootstrap confidence band for  $G$  under model (1.2) is given by

$$\left( \tilde{G}(\cdot) - \frac{u_{1-\alpha}^n}{\sqrt{n}}, \tilde{G}(\cdot) + \frac{u_{1-\alpha}^n}{\sqrt{n}} \right). \tag{5.1}$$

The bootstrap confidence bands in (5.1) are forced to be symmetric and will have the same width at all points regardless of the amount of data-support. Alternatively, non-forced symmetric bootstrap confidence intervals can be calculated as follows. For each bootstrap replicate, estimate the covariance matrix of  $U$  in part (a) of Theorem 5.1, so that we have an estimate of  $\text{Var}(\tilde{G}(T_i))_j$  for each  $i = 1, \dots, n$ . Then for each  $i$  we average these variance estimates across the bootstrap samples to get an overall estimate of each  $\text{Var}(\tilde{G}(T_i))$ . Then a  $1 - \alpha$  bootstrap pointwise confidence interval for  $G(T_i)$  is

$$\left( \tilde{G}(T_i) - z_{1-\frac{\alpha}{2}} \sqrt{\text{Var}(\tilde{G}(T_i))}, \tilde{G}(T_i) + z_{1-\frac{\alpha}{2}} \sqrt{\text{Var}(\tilde{G}(T_i))} \right), \tag{5.2}$$

where  $z_{1-\frac{\alpha}{2}}$  satisfies  $P(Z \leq z_{1-\frac{\alpha}{2}}) = 1 - \frac{\alpha}{2}$  with  $Z \sim N(0, 1)$ . Another alternative is the  $1 - \alpha$  Bonferroni simultaneous confidence intervals (Johnson and Wichern (1998, p.250)) for  $\{G(T_i) : i = 1, \dots, n\}$  given by

$$\left( \tilde{G}(T_i) - t_{1-\frac{\alpha}{2n}}(n-1) \sqrt{\text{Var}(\tilde{G}(T_i))}, \tilde{G}(T_i) + t_{1-\frac{\alpha}{2n}}(n-1) \sqrt{\text{Var}(\tilde{G}(T_i))} \right), \quad i = 1, \dots, n, \tag{5.3}$$

where  $t_{1-\frac{\alpha}{2n}}(n-1)$  satisfies  $P(T \leq t_{1-\frac{\alpha}{2n}}(n-1)) = 1 - \frac{\alpha}{2n}$  with  $T \sim t_{n-1}$ . We adopt the convention that when the left endpoint in (5.1), (5.2), or (5.3) is negative, it is replaced by 0, and when the right endpoint in (5.1), (5.2), or (5.3) is greater than 1, it is replaced by 1.

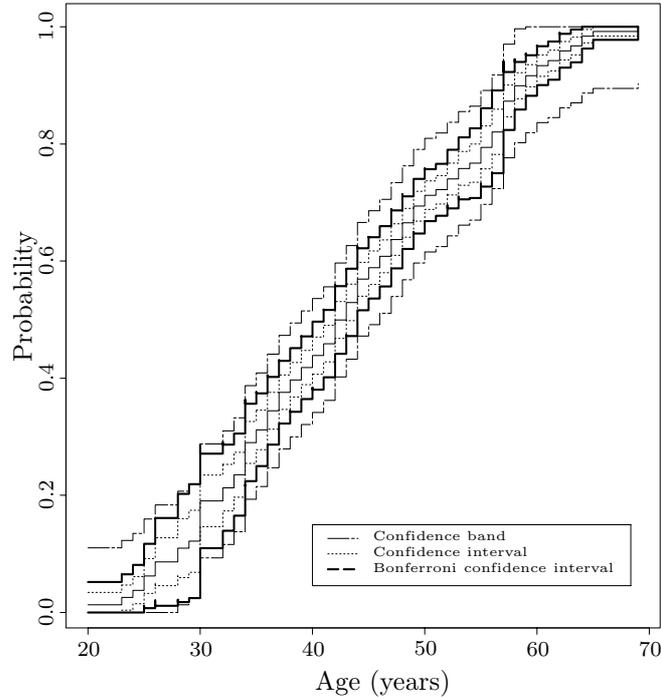


Figure 2. Example 4.1: Estimated cumulative distribution function  $\tilde{G}$  (solid curve), 95% confidence band (5.1) (dashed curve), 95% pointwise confidence interval (5.2) (dotted curve), and 95% Bonferroni simultaneous confidence intervals (5.3) (thick curve).

For the odds-linear model and the data set described in Example 4.1, Figure 2 shows the curve of  $\tilde{G}$  along with the 95% confidence band (5.1), 95% pointwise confidence interval (5.2), and the 95% Bonferroni simultaneous confidence intervals (5.3), whereas Figure 3 shows the curve of  $\hat{G}$  together with the 95% standard confidence band and pointwise and Bonferroni confidence intervals for  $G$  constructed from the control data  $X_1, \dots, X_{n_0}$ . Similarly, Figures 5 and 6 display these curves, bands, and intervals for model (4.4) and the data set described in Example 4.2. The confidence bands and intervals in Figures 2, 3, 5, and 6 are constructed based on 1000 bootstrap samples. In both examples, we found that 500 bootstrap samples or more are needed to get reliable variance estimates and confidence bands.

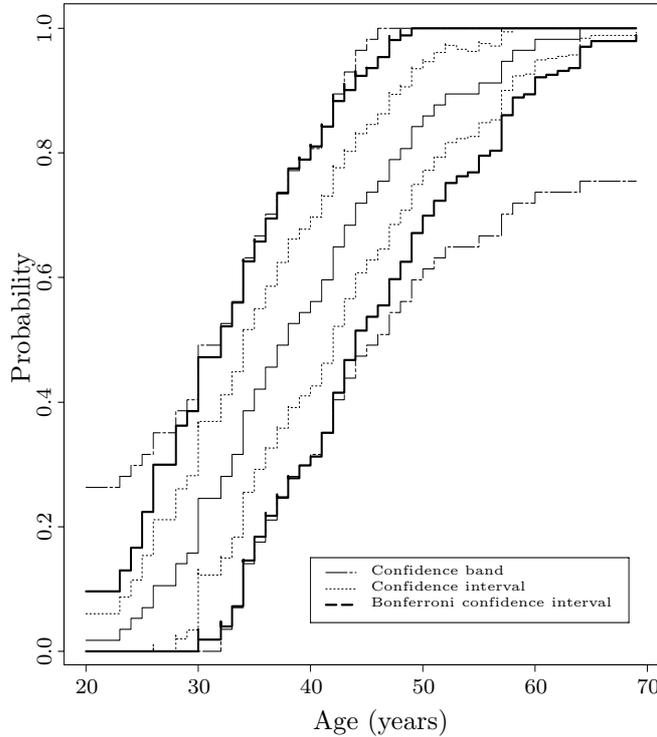


Figure 3. Example 4.1: Estimated cumulative distribution function  $\hat{G}$  (solid curve), 95% standard confidence band (dashed curve), 95% standard pointwise confidence interval (dotted curve), and 95% standard Bonferroni simultaneous confidence intervals (thick curve).

For Example 4.1, the pointwise confidence interval (5.2) is narrower than the Bonferroni simultaneous confidence intervals (5.3), which in turn are narrower than the confidence band (5.1). For Example 4.2, the pointwise confidence interval (5.2) is narrower than the Bonferroni simultaneous confidence intervals (5.3) and the confidence band (5.1), and yet the Bonferroni simultaneous confidence intervals (5.3) are wider (narrower) than the confidence band (5.1) for lower (higher) log dosage.

**6. A Simulation Study**

We now assess, via simulation, the finite sample performance of the maximum semiparametric likelihood estimator  $(\hat{\theta}, \hat{\beta})$  in (2.3) for the odds-linear model described in Example 4.1 with  $s(x; \beta) = \log(1 + x\beta)$ . Considering that there are other methods that could be used to estimate  $(\theta, \beta, G)$ , we may, for example, estimate  $G$  by  $\hat{G}(x) = \frac{1}{n_0} \sum_{i=1}^{n_0} I_{[X_i \leq x]}$  based on the control sample  $X_1, \dots, X_{n_0}$ ,

and then estimate  $(\theta, \beta)$  by the case sample  $Z_1, \dots, Z_{n_1}$ . This approach is *ad hoc* but is computationally simple. Specifically, if we equate the sample mean of  $Z_1, \dots, Z_{n_1}$  to the population mean of  $H$  with  $G$  replaced by  $\hat{G}$  along with the fact that  $\int dH(x) = \int \exp[\theta + s(x; \beta)] dG(x) = 1$ , we may estimate  $(\theta, \beta)$  by  $(\hat{\theta}, \hat{\beta})$  defined to be a solution to the following system of equations:

$$\frac{1}{n_0} \sum_{i=1}^{n_0} \exp[\theta + s(X_i; \beta)] = 1, \quad \frac{1}{n_0} \sum_{i=1}^{n_0} X_i \exp[\theta + s(X_i; \beta)] = \bar{Z}.$$

Then  $\hat{\theta} = -\log(1 + \hat{\beta}\bar{X})$  and  $\hat{\beta} = \frac{\bar{Z} - \bar{X}}{S^2 - \bar{X}\bar{Z}}$ , where  $S^2 = \frac{1}{n_0} \sum_{i=1}^{n_0} X_i^2$ . It can be shown that  $\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} N(0, \sigma_\theta^2)$  and  $\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{d} N(0, \sigma_\beta^2)$ , where  $\sigma_\theta^2 = \mu_1^2 e^{2\theta} \sigma_\beta^2$  and  $\sigma_\beta^2 = \frac{1+\rho}{\rho} \frac{\rho\sigma_X^2 + \sigma_Z^2}{(\mu_2 - \mu_1\mu_Z)^2}$  with  $\mu_k = \int x^k dG(x)$  for  $k = 1, 2, 3$ ,  $\mu_Z = e^\theta(\mu_1 + \beta\mu_2)$ ,  $\sigma_X^2 = \text{Var}(X_1)$ , and  $\sigma_Z^2 = \text{Var}(Z_1) = e^\theta(\mu_2 + \beta\mu_3) - \mu_Z^2$ . In our simulation study, we assume that  $g(x)$  is the standard exponential density function. Let  $\beta = 0.1$  be fixed so that  $\theta = -0.09531$ . Our aim is to compare the performance of  $(\tilde{\theta}, \tilde{\beta})$  with that of  $(\hat{\theta}, \hat{\beta})$  by examining their biases, variances, and relative efficiencies.

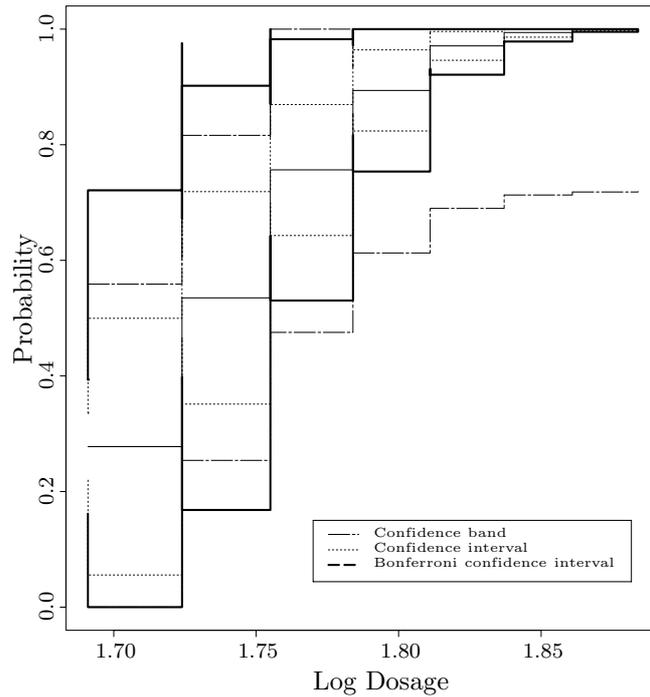


Figure 5. Example 4.2: Estimated cumulative distribution function  $\tilde{G}$  (solid curve), 95% confidence band (5.1) (dashed curve), 95% pointwise confidence

interval (5.2) (dotted curve), and 95% Bonferroni simultaneous confidence intervals (5.3) (thick curve).

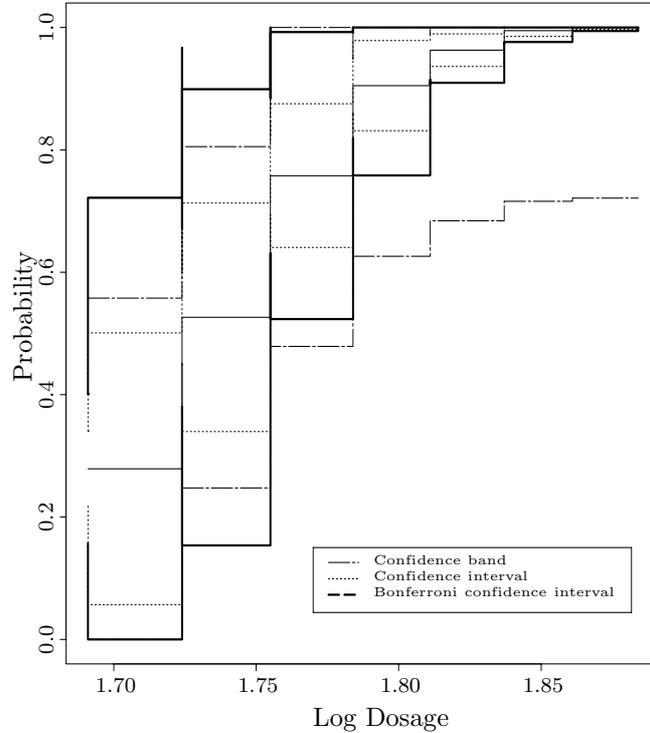


Figure 6. Example 4.2: Estimated cumulative distribution function  $\hat{G}$  (solid curve), 95% standard confidence band (dashed curve), 95% standard pointwise confidence interval (dotted curve), and 95% standard Bonferroni simultaneous confidence intervals (thick curve).

In our simulations, we considered sample sizes of  $(n_0, n_1) = (80, 80), (80, 120), (120, 80), (120, 120)$ . For each pair  $(n_0, n_1)$ , we generated 1000 independent sets of combined random samples from  $g(x)$  and  $h(x)$ . Simulation results are summarized in Tables 1 and 2.

In Table 1,  $\text{Bias}(\hat{\theta})$  and  $\text{Var}(\hat{\theta})$  stand for, respectively, the average of 1000 biases of  $\hat{\theta}$  and the sample variance of 1000 estimates  $\hat{\theta}$ , whereas  $\text{Bias}(\tilde{\theta})$  and  $\text{Var}(\tilde{\theta})$  stand for, respectively, the average of 1000 biases of  $\tilde{\theta}$  and the sample variance of 1000 estimates  $\tilde{\theta}$ . In addition, we use  $\widehat{\text{Var}}(\hat{\theta})$  and  $\widehat{\text{Var}}(\tilde{\theta})$  to represent, respectively, the averages of 1000 estimated asymptotic variances of  $\hat{\theta}$  and that of  $\tilde{\theta}$ . The estimated asymptotic variances of  $\tilde{\theta}$  are obtained from  $\Sigma$  in Theorem 3.3 with  $\tilde{G}$  in place of  $G$ , whereas the estimated asymptotic variances of  $\hat{\theta}$  are obtained from  $\sigma_{\theta}^2$  with  $G$  replaced by  $\hat{G}$ . Moreover,  $e(\hat{\theta}, \tilde{\theta})$  stands for the relative

efficiency of  $\hat{\theta}$  relative to  $\tilde{\theta}$ , i.e.,  $e(\hat{\theta}, \tilde{\theta}) = \text{Var}(\tilde{\theta}) / \text{Var}(\hat{\theta})$ . In Table 2, the notations for  $\hat{\beta}$  and  $\tilde{\beta}$  are similar to those of  $\hat{\theta}$  and  $\tilde{\theta}$  in Table 1.

Tables 1 and 2 reveal that the biases and variances of  $(\tilde{\theta}, \tilde{\beta})$  are all smaller than those of  $(\hat{\theta}, \hat{\beta})$ , with relative efficiencies ranging from 0.48 to 0.80. In addition, both tables indicate that the estimated asymptotic variances of  $\tilde{\theta}$  and  $\tilde{\beta}$  work well on average.

Table 1. Biases, variances, and efficiencies of  $\tilde{\theta}$  and  $\hat{\theta}$ .

$(n_0, n_1)$	$\rho$	Bias( $\hat{\theta}$ )	Bias( $\tilde{\theta}$ )	Var( $\hat{\theta}$ )	Var( $\tilde{\theta}$ )	$\widehat{\text{Var}}(\hat{\theta})$	$\widehat{\text{Var}}(\tilde{\theta})$	$e(\hat{\theta}, \tilde{\theta})$
(80, 80)	1.00	-0.05746	-0.03489	0.05250	0.03756	0.02877	0.03965	0.71533
(80, 120)	1.50	-0.04994	-0.02670	0.04648	0.02975	0.02537	0.03161	0.64004
(120, 80)	0.67	-0.05288	-0.03745	0.04125	0.03297	0.02407	0.03361	0.79928
(120, 120)	1.00	-0.04671	-0.02879	0.03609	0.02529	0.01886	0.02535	0.70094

Table 2. Biases, variances, and efficiencies of  $\tilde{\beta}$  and  $\hat{\beta}$ .

$(n_0, n_1)$	$\rho$	Bias( $\hat{\beta}$ )	Bias( $\tilde{\beta}$ )	Var( $\hat{\beta}$ )	Var( $\tilde{\beta}$ )	$\widehat{\text{Var}}(\hat{\beta})$	$\widehat{\text{Var}}(\tilde{\beta})$	$e(\hat{\beta}, \tilde{\beta})$
(80, 80)	1.00	0.12544	0.08277	0.15621	0.09059	0.04883	0.10724	0.57994
(80, 120)	1.50	0.11383	0.06801	0.14207	0.06885	0.06201	0.07605	0.48464
(120, 80)	0.67	0.10423	0.07756	0.10114	0.06826	0.03894	0.07667	0.67497
(120, 120)	1.00	0.09333	0.06114	0.08542	0.04713	0.02867	0.05206	0.55175

### 7. Proofs

We present six lemmas which will be used in the proof of the main results. The proofs of Lemmas 7.1 and 7.3 are straightforward and are omitted. Lemma 7.2 can be proved by employing Example 2.10.10 of van der Vaart and Wellner (1996, p.192). Throughout this section, the norm of a  $n_1 \times n_2$  matrix  $D = (d_{ij})_{n_1 \times n_2}$  is defined by  $\|D\| = (\sum_{i=1}^{n_1} \sum_{j=1}^{n_2} d_{ij}^2)^{1/2}$  for  $n_1, n_2 \geq 1$ . Further, write

$$\begin{aligned}
 s_{n11} &= -\frac{\partial^2 \ell(\theta_0, \beta_0)}{\partial \theta^2} = \sum_{i=1}^n \frac{\rho \exp[\theta + s(T_i; \beta)]}{\{1 + \rho \exp[\theta + s(T_i; \beta)]\}^2}, \\
 s_{n21} &= -\frac{\partial^2 \ell(\theta_0, \beta_0)}{\partial \theta \partial \beta} = \sum_{i=1}^n \frac{\rho \exp[\theta + s(T_i; \beta)]}{\{1 + \rho \exp[\theta + s(T_i; \beta)]\}^2} s_1(T_i; \beta), \\
 s_{n22} &= -\frac{\partial^2 \ell(\theta_0, \beta_0)}{\partial \beta \partial \beta^\tau} = \sum_{i=1}^n \frac{\rho \exp[\theta + s(T_i; \beta)]}{\{1 + \rho \exp[\theta + s(T_i; \beta)]\}^2} s_1(T_i; \beta) s_1^\tau(T_i; \beta) \\
 &\quad + \sum_{i=1}^n \frac{\rho \exp[\theta + s(T_i; \beta)]}{1 + \rho \exp[\theta + s(T_i; \beta)]} s_2(T_i; \beta) - \sum_{j=1}^{n_1} s_2(Z_j; \beta), \quad S_n = \begin{pmatrix} s_{n11} & s_{n21}^\tau \\ s_{n21} & s_{n22} \end{pmatrix},
 \end{aligned}$$

$$\begin{aligned}
 H_0(t; \theta, \beta) &= \frac{1}{n_0} \sum_{i=1}^n \frac{\rho \exp[\theta + s(T_i; \beta)]}{\{1 + \rho \exp[\theta + s(T_i; \beta)]\}^2} I_{[T_i \leq t]}, & H_0(t) &= H_0(t; \theta_0, \beta_0), \\
 H_3(t; \theta, \beta) &= \frac{1}{n_0} \sum_{i=1}^n \frac{\rho \exp[\theta + s(T_i; \beta)]}{\{1 + \rho \exp[\theta + s(T_i; \beta)]\}^2} s_1(T_i; \beta) I_{[T_i \leq t]}, \\
 H_3(t) &= H_3(t; \theta_0, \beta_0), \\
 H_4(t) &= \frac{1}{n_0} \sum_{i=1}^{n_0} \frac{\rho \exp[\theta_0 + s(X_i; \beta_0)]}{1 + \rho \exp[\theta_0 + s(X_i; \beta_0)]} I_{[X_i \leq t]}, \\
 H_{12}(t) &= \frac{1}{n_1} \sum_{j=1}^{n_1} \frac{\rho I_{[Z_j \leq t]}}{1 + \rho \exp[\theta_0 + s(Z_j; \beta_0)]}. \tag{7.1}
 \end{aligned}$$

**Lemma 7.1.** *Suppose that model (1.2) holds and  $A$  is positive definite. If  $B = \frac{1}{n} \text{Var} \left[ \begin{pmatrix} \frac{\partial \ell(\theta_0, \beta_0)}{\partial \theta} \\ \frac{\partial \ell(\theta_0, \beta_0)}{\partial \beta} \end{pmatrix} \right]$ , then*

$$B = \frac{\rho}{1+\rho} A - \rho \begin{pmatrix} A_{11} \\ A_{21} \end{pmatrix} (A_{11}, A_{21}^T), \quad S^{-1} B S^{-1} = \frac{1+\rho}{\rho} \left[ A^{-1} - \begin{pmatrix} 1+\rho & 0 \\ 0 & 0 \end{pmatrix} \right] = \Sigma.$$

**Lemma 7.2.** *Let  $Y_1, \dots, Y_n$  be independent  $p$ -dimensional random vectors with distribution  $F_Y$ . Suppose that  $q$  is a real function such that  $E|q(Y_1)| < \infty$ . Let  $\Lambda_n(y) = \frac{1}{n} \sum_{i=1}^n q(Y_i) I_{[Y_i \leq y]}$  and  $\Lambda(y) = E[q(Y_1) I_{[Y_1 \leq y]}]$ . Then  $\sup_{-\infty \leq y \leq \infty} |\Lambda_n(y) - \Lambda(y)| \xrightarrow{\text{a.s.}} 0$ .*

**Lemma 7.3.** *Suppose that model (1.2) holds and  $A$  is positive definite. For  $-\infty \leq s, t \leq \infty$ , we have*

$$\begin{aligned}
 &\text{Cov}(\sqrt{n}[H_1(s) - \hat{G}(s)], \sqrt{n}[H_1(t) - \hat{G}(t)]) \\
 &\quad = \rho(1 + \rho) A_{11}(s \wedge t) - \rho(1 + \rho)^2 A_{11}(s) A_{11}(t), \\
 &\text{Cov}(\sqrt{n}[H_1(s) - \hat{G}(s)], \sqrt{n}H_2(t)) = \text{Cov}(\sqrt{n}H_2(s), \sqrt{n}H_2(t)) \\
 &\quad = \rho(1 + \rho)(A_{11}(t), A_{21}^T(t)) A^{-1} \begin{pmatrix} A_{11}(s) \\ A_{21}(s) \end{pmatrix} - \rho(1 + \rho)^2 A_{11}(s) A_{11}(t),
 \end{aligned}$$

where  $H_1(t)$  and  $H_2(t)$  are defined in (3.3).

**Lemma 7.4.** *Suppose that model (1.2) and Assumption (A2) hold. If  $A$  is positive definite, then the stochastic process  $\{\sqrt{n}[H_1(t) - \hat{G}(t) - H_2(t)], -\infty \leq t \leq \infty\}$  is tight in  $D[-\infty, \infty]^p$ .*

**Proof.** First, it is easy to see from (7.1) that  $\sqrt{n}[H_1(t) - \hat{G}(t) - H_2(t)] = \sqrt{n}[H_{12}(t) - \rho A_{11}(t)] - \sqrt{n}[H_4(t) - \rho A_{11}(t)] - \sqrt{n}H_2(t)$ . Let  $\mathcal{F} = \{I_{(-\infty, t]} : t \in \mathbf{R}^p\}$  be the collection of all indicator functions of lower rectangles  $(-\infty, t]$  in  $\mathbf{R}^p$ .

According to classical empirical process theory,  $\mathcal{F}$  is a  $P_X$ -Donsker class as well as a  $P_Z$ -Donsker class, where  $P_X$  is the law of  $X_1$  and  $P_Z$  is the law of  $Z_1$ . Let

$$f_0(y) = \frac{\rho \exp[\theta_0 + s(y; \beta_0)]}{1 + \rho \exp[\theta_0 + s(y; \beta_0)]} \quad \text{and} \quad f_1(y) = \frac{\rho}{1 + \rho \exp[\theta_0 + s(y; \beta_0)]}.$$

Then it is seen that both  $f_0$  and  $f_1$  are uniformly bounded functions. According to Example 2.10.10 of van der Vaart and Wellner (1996, p.192), we can conclude that  $\mathcal{F} \cdot f_0$  is a  $P_X$ -Donsker class and  $\mathcal{F} \cdot f_1$  is a  $P_Z$ -Donsker class. Let  $P_{n_0} = \frac{1}{n_0} \sum_{i=1}^{n_0} \delta_{X_i}$  be the empirical measure of  $X_1, \dots, X_{n_0}$  and  $P_{n_1} = \frac{1}{n_1} \sum_{j=1}^{n_1} \delta_{Z_j}$  be the empirical measure of  $Z_1, \dots, Z_{n_1}$ , where  $\delta_x$  is the measure with mass one at  $x$ . Then it can be shown, after some algebra, that  $\sqrt{n_0}(P_{n_0} - P_X)(I_{(-\infty, t]} f_0) = \sqrt{n_0}[H_4(t) - \rho A_{11}(t)]$  and  $\sqrt{n_1}(P_{n_1} - P_Z)(I_{(-\infty, t]} f_1) = \sqrt{n_1}[H_{12}(t) - \rho A_{11}(t)]$ . As a result, there exist two zero-mean Gaussian processes  $B_0$  and  $B_1$  such that

$$\sqrt{n_0}(H_4 - \rho A_{11}) \xrightarrow{\mathcal{D}} B_0 \text{ in } D[-\infty, \infty]^p, \quad \sqrt{n_1}(H_{12} - \rho A_{11}) \xrightarrow{\mathcal{D}} B_1 \text{ in } D[-\infty, \infty]^p. \tag{7.2}$$

Thus, the stochastic processes  $\{\sqrt{n_0}[H_4(t) - \rho A_{11}(t)], -\infty \leq t \leq \infty\}$  and  $\{\sqrt{n_1}[H_{12}(t) - \rho A_{11}(t)], -\infty \leq t \leq \infty\}$  are tight in  $D[-\infty, \infty]^p$ . To complete the proof, it suffices to show that the stochastic process  $\{\sqrt{n}H_2(t), -\infty \leq t \leq \infty\}$  is tight in  $D[-\infty, \infty]^p$ . For  $k = 0, 1, \dots, p$ , let

$$\begin{aligned} A_{21k}(t) &= \int \frac{\exp[\theta_0 + s(y; \beta_0)]}{1 + \rho \exp[\theta_0 + s(y; \beta_0)]} s_{1k}(y; \beta_0) I_{[y \leq t]} dG(y), \\ L_{1k} &= \sum_{i=1}^{n_0} \frac{\rho \exp[\theta_0 + s(X_i; \beta_0)]}{1 + \rho \exp[\theta_0 + s(X_i; \beta_0)]} s_{1k}(X_i; \beta_0) \\ &\quad - n_0 \int \frac{\rho \exp[\theta_0 + s(y; \beta_0)]}{1 + \rho \exp[\theta_0 + s(y; \beta_0)]} s_{1k}(y; \beta_0) dG(y), \\ L_{2k} &= \sum_{j=1}^{n_1} \frac{1}{1 + \rho \exp[\theta_0 + s(Z_j; \beta_0)]} s_{1k}(Z_j; \beta_0) \\ &\quad - n_0 \int \frac{\rho \exp[\theta_0 + s(y; \beta_0)]}{1 + \rho \exp[\theta_0 + s(y; \beta_0)]} s_{1k}(y; \beta_0) dG(y). \end{aligned}$$

Then  $EL_{1k} = EL_{2k} = 0$  for  $k = 0, 1, \dots, p$ . Furthermore, it is seen from (2.3) that

$$(A_{11}(t), A_{21}^T(t)) = (A_{210}(t), A_{211}(t), \dots, A_{21p}(t)), \quad \begin{pmatrix} \frac{\partial \ell(\theta_0, \beta_0)}{\partial \theta} \\ \frac{\partial \ell(\theta_0, \beta_0)}{\partial \beta} \end{pmatrix} = \begin{pmatrix} L_{20} - L_{10} \\ L_{21} - L_{11} \\ \vdots \\ L_{2p} - L_{1p} \end{pmatrix}.$$

Consequently, if  $S^{-1} = (s^{jk})$ , we can write

$$\begin{aligned} \sqrt{n}H_2(t) &= \frac{\rho}{\sqrt{n}}(A_{11}(t), A_{21}^\tau(t))S^{-1} \begin{pmatrix} \frac{\partial \ell(\theta_0, \beta_0)}{\partial \theta} \\ \frac{\partial \ell(\theta_0, \beta_0)}{\partial \beta} \end{pmatrix} \\ &= \frac{\rho}{\sqrt{n}} \sum_{k=0}^p \sum_{m=0}^p s^{km} A_{21k}(t)(L_{2m} - L_{1m}). \end{aligned}$$

Let  $\mathcal{U} = \{u_t(\cdot) : t \in \mathbf{R}^p\} \cup \{U_k(\cdot) : k = 0, 1, \dots, p\}$  and  $\mathcal{V} = \{v_t(\cdot) : t \in \mathbf{R}^p\} \cup \{V_k(\cdot) : k = 0, 1, \dots, p\}$  be two classes of real functions on  $\mathbf{R}^p$ , where

$$\begin{aligned} U_k(y) &= \frac{\rho^2 \exp[\theta_0 + s(y; \beta_0)]}{1 + \rho \exp[\theta_0 + s(y; \beta_0)]} s_{1k}(y; \beta_0), \\ V_k(y) &= \frac{\rho^2}{1 + \rho \exp[\theta_0 + s(y; \beta_0)]} s_{1k}(y; \beta_0), \end{aligned} \quad k = 0, 1, \dots, p,$$

$u_t(y) = \sum_{m=0}^p w_m(t)U_m(t)$ ,  $v_t(y) = \sum_{m=0}^p w_m(t)V_m(t)$ , and  $w_m(t) = \sum_{k=0}^p s^{km} A_{21k}(t)$  for  $m = 0, 1, \dots, p$ . Since each member of  $\mathcal{U}$  can be expressed as a linear combination of a fixed, finite set of functions  $U_0, U_1, \dots, U_p$  in  $\mathcal{U}$  and each member of  $\mathcal{V}$  can be expressed as a linear combination of a fixed, finite set of functions  $V_0, V_1, \dots, V_p$  in  $\mathcal{V}$ , both  $\mathcal{U}$  and  $\mathcal{V}$  are finite-dimensional vector spaces of dimension at most  $p + 1$ . According to Lemma 2.6.15 of van der Vaart and Wellner (1996, p.146), both  $\mathcal{U}$  and  $\mathcal{V}$  are VC-subgraph classes of index smaller than or equal to  $p + 3$ . Consequently, according to Theorems 2.6.7 and 2.5.2 of van der Vaart and Wellner (1996, pp.141, 127), we can conclude that  $\mathcal{U}$  is a  $P_X$ -Donsker class and  $\mathcal{V}$  is a  $P_Z$ -Donsker class. Since it can be shown, after some algebra, that  $\sqrt{n_1}(P_{n_1} - P_Z)(v_t) - \sqrt{n_0}(P_{n_0} - P_X)(u_t) = \sqrt{1 + \rho}\sqrt{n}H_2(t)$ , we can conclude by employing a similar argument as in (7.2) that the stochastic process  $\{\sqrt{n}H_2(t), -\infty \leq t \leq \infty\}$  is tight in  $D[-\infty, \infty]^p$ , and this completes the proof of Lemma 7.4.

**Lemma 7.5.** *Under the conditions of Theorem 3.1, we have  $\sup_{-\infty \leq t \leq \infty} |\tilde{G}(t) - \hat{G}(t)| \xrightarrow{\text{a.s.}} 0$  as  $n \rightarrow \infty$ .*

**Proof.** Since  $\text{E}H_4(t) = \text{E}H_{12}(t) = \rho A_{11}(t)$  and  $(\tilde{\theta}, \tilde{\beta})$  is strongly consistent by part (a) of Theorem 3.1, applying a first-order Taylor expansion yields

$$\begin{aligned} \tilde{G}(t) - \hat{G}(t) &= [H_{12}(t) - \text{E}H_{12}(t)] - [H_4(t) - \text{E}H_4(t)] - H_0(t; \theta_{0*}, \beta_{0*})(\tilde{\theta} - \theta_0) \\ &\quad - H_3(t; \theta_{0*}, \beta_{0*})(\tilde{\beta} - \beta_0), \end{aligned} \tag{7.3}$$

where  $(\theta_{0*}, \beta_{0*})$  satisfies  $|\theta_{0*} - \theta_0|^2 + \|\beta_{0*} - \beta_0\|^2 \leq |\tilde{\theta} - \theta_0|^2 + \|\tilde{\beta} - \beta_0\|^2$ . According to Lemma 7.2, we have

$$\sup_{-\infty \leq t \leq \infty} |H_{12}(t) - \text{E}H_{12}(t)| \xrightarrow{\text{a.s.}} 0, \quad \sup_{-\infty \leq t \leq \infty} |H_4(t) - \text{E}H_4(t)| \xrightarrow{\text{a.s.}} 0. \tag{7.4}$$

Furthermore, we have

$$\sup_{-\infty \leq t \leq \infty} |H_0(t; \theta_{0*}, \beta_{0*})(\tilde{\theta} - \theta_0)| \leq (1 + \rho)|\tilde{\theta} - \theta_0| \xrightarrow{\text{a.s.}} 0. \tag{7.5}$$

Moreover, by Assumption (A2), we have

$$\sup_{-\infty \leq t \leq \infty} |H_3(t; \theta_{0*}, \beta_{0*})(\tilde{\beta} - \beta_0)| \leq \|\tilde{\beta} - \beta_0\| \left( \frac{1}{n_0} \sum_{i=1}^n Q_1(T_i) \right) \stackrel{\text{a.s.}}{=} \|\tilde{\beta} - \beta_0\| O(1) \xrightarrow{\text{a.s.}} 0. \tag{7.6}$$

Therefore, Lemma 7.5 follows from (7.3)–(7.6). The proof is completed.

**Lemma 7.6.** *Suppose that conditions of Theorem 3.1 hold.*

- (a) *Let  $q(t, \theta, \beta)$  be a function such that  $\frac{\partial q(t, \theta, \beta)}{\partial \theta}$  and  $\frac{\partial q(t, \theta, \beta)}{\partial \beta}$  exist for  $(\theta, \beta) \in \Theta_1 \times \Theta_0$ , where  $\Theta_1$  is some neighborhood of  $\theta_0$ . Suppose that there exists a function  $Q$  such that*

$$|q(t, \theta, \beta)| \leq Q(t), \quad \left| \frac{\partial q(t, \theta, \beta)}{\partial \theta} \right| \leq Q(t), \quad \left| \frac{\partial q(t, \theta, \beta)}{\partial \beta_k} \right| \leq Q(t)$$

for all  $(\theta, \beta) \in \Theta_1 \times \Theta_0$  and  $k = 1, \dots, p$ .

Suppose further that  $\int Q(y)[1 + Q_1(y)]\{1 + \rho \exp[\theta_0 + s(y; \beta_0)]\}dG(y) < \infty$ .

Then  $E_{\tilde{G}}[q(X_1^*, \tilde{\theta}, \tilde{\beta})] \xrightarrow{\text{a.s.}} E_G[q(X_1, \theta_0, \beta_0)] = \int q(y, \theta_0, \beta_0)dG(y)$  as  $n \rightarrow \infty$ .

- (b) *If  $\int |q(y)|[1 + Q_1(y)]\{1 + \rho \exp[\theta_0 + s(y; \beta_0)]\}dG(y) < \infty$ , then  $E_{\tilde{G}}[q(X_1^*)] \xrightarrow{\text{a.s.}} E_G[q(X_1)] = \int q(y)dG(y)$  as  $n \rightarrow \infty$ .*
- (c) *If  $\int Q_1^2(y)Q_2(y)\{1 + \rho \exp[\theta_0 + s(y; \beta_0)]\}dG(y) < \infty$ , then  $A_n \stackrel{\text{a.s.}}{=} A + o(1)$  as  $n \rightarrow \infty$  and  $A_n$  is positive definite for sufficiently large  $n$  with probability 1.*

**Proof.** For part (a), since  $(\tilde{\theta}, \tilde{\beta})$  is strongly consistent by part (a) of Theorem 3.1, applying a first-order Taylor expansion yields

$$\begin{aligned} E_{\tilde{G}}[q(X_1^*, \tilde{\theta}, \tilde{\beta})] &= \sum_{i=1}^n \tilde{p}_i q(T_i, \tilde{\theta}, \tilde{\beta}) = \frac{1}{n_0} \sum_{i=1}^n \frac{q(T_i, \tilde{\theta}, \tilde{\beta})}{1 + \rho \exp[\tilde{\theta} + s(T_i; \tilde{\beta})]} \\ &= D_{n1} + D_{n2} + D_{n3}, \end{aligned} \tag{7.7}$$

where

$$\begin{aligned} D_{n1} &= \frac{1}{n_0} \sum_{i=1}^n \frac{q(T_i, \theta_0, \beta_0)}{1 + \rho \exp[\theta_0 + s(T_i; \beta_0)]}, \\ D_{n2} &= \left[ \frac{1}{n_0} \sum_{i=1}^n \frac{\left\{ 1 + \rho \exp[\theta_{0*} + s(T_i; \beta_{0*})] \right\} \frac{\partial q(T_i, \theta_{0*}, \beta_{0*})}{\partial \theta} - q(T_i, \theta_{0*}, \beta_{0*}) \rho \exp[\theta_{0*} + s(T_i; \beta_{0*})]}{\left\{ 1 + \rho \exp[\theta_{0*} + s(T_i; \beta_{0*})] \right\}^2} \right] (\tilde{\theta} - \theta_0), \\ D_{n3} &= \left[ \frac{1}{n_0} \sum_{i=1}^n \frac{\left\{ 1 + \rho \exp[\theta_{0*} + s(T_i; \beta_{0*})] \right\} \frac{\partial q(T_i, \theta_{0*}, \beta_{0*})}{\partial \beta^T} - \rho \exp[\theta_{0*} + s(T_i; \beta_{0*})] q(T_i, \theta_{0*}, \beta_{0*}) s_1^T(T_i; \beta_{0*})}{\left\{ 1 + \rho \exp[\theta_{0*} + s(T_i; \beta_{0*})] \right\}^2} \right] (\tilde{\beta} - \beta_0) \end{aligned}$$

with  $(\theta_{0*}, \beta_{0*})$  satisfying  $|\theta_{0*} - \theta_0|^2 + \|\beta_{0*} - \beta_0\|^2 \leq |\tilde{\theta} - \theta_0|^2 + \|\tilde{\beta} - \beta_0\|^2$ . Applying part (a) of Theorem 3.1 and the Strong Law of Large Numbers gives

$$\begin{aligned}
 D_{n1} &= \frac{1}{n_0} \sum_{i=1}^{n_0} \frac{q(X_i, \theta_0, \beta_0)}{1 + \rho \exp[\theta_0 + s(X_i; \beta_0)]} + \rho \frac{1}{n_1} \sum_{j=1}^{n_1} \frac{q(Z_j, \theta_0, \beta_0)}{1 + \rho \exp[\theta_0 + s(Z_j; \beta_0)]} \\
 &\xrightarrow{\text{a.s.}} \int_{-\infty}^{\infty} q(y, \theta_0, \beta_0) dG(y), \\
 |D_{n2}| &\leq \left[ \frac{1}{n_0} \sum_{i=1}^n \left( \left| \frac{\partial q(T_i, \theta_{0*}, \beta_{0*})}{\partial \theta} \right| + |q(T_i, \theta_{0*}, \beta_{0*})| \right) \right] |\tilde{\theta} - \theta_0| \\
 &\leq 2 \left[ \frac{1}{n_0} \sum_{i=1}^n Q(T_i) \right] |\tilde{\theta} - \theta_0| \xrightarrow{\text{a.s.}} 0, \\
 |D_{n3}| &\leq \left[ \frac{1}{n_0} \sum_{i=1}^n \left( \left\| \frac{\partial q(T_i, \theta_{0*}, \beta_{0*})}{\partial \beta} \right\| + |q(T_i, \theta_{0*}, \beta_{0*})| \|s_1(T_i; \beta_{0*})\| \right) \right] \|\tilde{\beta} - \beta_0\| \\
 &= \sqrt{p} \left[ \frac{1}{n_0} \sum_{i=1}^n Q(T_i) [1 + Q_1(T_i)] \right] \|\tilde{\beta} - \beta_0\| \xrightarrow{\text{a.s.}} 0. \tag{7.8}
 \end{aligned}$$

Combining (7.7) with (7.8) completes the proof of part (a). Parts (b) and (c) are straightforward consequences of part (a). The proof is completed.

**Proof of Theorem 3.1.** Part (a) can be proved by employing a similar approach as in the proof of the consistency of  $(\tilde{\theta}, \tilde{\beta})$  in the logistic regression model in Prentice and Pyke (1979).

For part (b), since  $(\tilde{\theta}, \tilde{\beta})$  is strongly consistent by part (a), expanding  $\frac{\partial \ell(\tilde{\theta}, \tilde{\beta})}{\partial \theta}$  and  $\frac{\partial \ell(\tilde{\theta}, \tilde{\beta})}{\partial \beta}$  at  $(\theta_0, \beta_0)$  gives

$$\begin{aligned}
 0 &= \frac{\partial \ell(\tilde{\theta}, \tilde{\beta})}{\partial \theta} = \frac{\partial \ell(\theta_0, \beta_0)}{\partial \theta} + \frac{\partial^2 \ell(\theta_0, \beta_0)}{\partial \theta^2} (\tilde{\theta} - \theta_0) + \frac{\partial^2 \ell(\theta_0, \beta_0)}{\partial \theta \partial \beta \tau} (\tilde{\beta} - \beta_0) + o_p(\delta_n), \\
 0 &= \frac{\partial \ell(\tilde{\theta}, \tilde{\beta})}{\partial \beta} = \frac{\partial \ell(\theta_0, \beta_0)}{\partial \beta} + \frac{\partial^2 \ell(\theta_0, \beta_0)}{\partial \theta \partial \beta} (\tilde{\theta} - \theta_0) + \frac{\partial^2 \ell(\theta_0, \beta_0)}{\partial \beta \partial \beta \tau} (\tilde{\beta} - \beta_0) + o_p(\delta_n),
 \end{aligned}$$

where  $\delta_n = |\tilde{\theta} - \theta_0| + \|\tilde{\beta} - \beta_0\| = o_p(1)$ . Thus,  $S_n \begin{pmatrix} \tilde{\theta} - \theta_0 \\ \tilde{\beta} - \beta_0 \end{pmatrix} = \begin{pmatrix} \frac{\partial \ell(\theta_0, \beta_0)}{\partial \theta} \\ \frac{\partial \ell(\theta_0, \beta_0)}{\partial \beta} \end{pmatrix} + o_p(\delta_n)$ . Since

$S_n = nS + o_p(n)$  by the Weak Law of Large Numbers and  $\frac{1}{\sqrt{n}} \begin{pmatrix} \frac{\partial \ell(\theta_0, \beta_0)}{\partial \theta} \\ \frac{\partial \ell(\theta_0, \beta_0)}{\partial \beta} \end{pmatrix} = O_p(1)$  by the Central Limit Theorem, it follows that

$$\begin{aligned}
 \begin{pmatrix} \tilde{\theta} - \theta_0 \\ \tilde{\beta} - \beta_0 \end{pmatrix} &= \frac{1}{n} S^{-1} \begin{pmatrix} \frac{\partial \ell(\theta_0, \beta_0)}{\partial \theta} \\ \frac{\partial \ell(\theta_0, \beta_0)}{\partial \beta} \end{pmatrix} + o_p(n^{-1/2}) \frac{1}{\sqrt{n}} \begin{pmatrix} \frac{\partial \ell(\theta_0, \beta_0)}{\partial \theta} \\ \frac{\partial \ell(\theta_0, \beta_0)}{\partial \beta} \end{pmatrix} + o_p(n^{-1} \delta_n) \\
 &= \frac{1}{n} S^{-1} \begin{pmatrix} \frac{\partial \ell(\theta_0, \beta_0)}{\partial \theta} \\ \frac{\partial \ell(\theta_0, \beta_0)}{\partial \beta} \end{pmatrix} + o_p(n^{-1/2}),
 \end{aligned}$$

thus establishing (3.1). To prove (3.2), it suffices to show that  $\frac{1}{\sqrt{n}}S^{-1}\left(\frac{\partial\ell(\theta_0,\beta_0)}{\partial\theta}, \frac{\partial\ell(\theta_0,\beta_0)}{\partial\beta}\right) \xrightarrow{d} N_{p+1}(\mathbf{0}, \Sigma)$ . Since each term in  $\frac{\partial\ell(\theta_0,\beta_0)}{\partial\theta}$  and  $\frac{\partial\ell(\theta_0,\beta_0)}{\partial\beta}$  has mean 0, it follows from the Multivariate Central Limit Theorem that  $\frac{1}{\sqrt{n}}B^{-1/2}\left(\frac{\partial\ell(\theta_0,\beta_0)}{\partial\theta}, \frac{\partial\ell(\theta_0,\beta_0)}{\partial\beta}\right) \xrightarrow{d} N_{p+1}(\mathbf{0}, I_{p+1})$ , where  $B$  is defined in Lemma 7.1. By Slutsky's Theorem and Lemma 7.1, we have

$$\begin{aligned} \frac{1}{\sqrt{n}}S^{-1}\left(\frac{\partial\ell(\theta_0,\beta_0)}{\partial\theta}, \frac{\partial\ell(\theta_0,\beta_0)}{\partial\beta}\right) &= S^{-1}B^{1/2}\frac{1}{\sqrt{n}}B^{-1/2}\left(\frac{\partial\ell(\theta_0,\beta_0)}{\partial\theta}, \frac{\partial\ell(\theta_0,\beta_0)}{\partial\beta}\right) \xrightarrow{d} S^{-1}B^{1/2}N_{p+1}(\mathbf{0}, I_{p+1}) \\ &\stackrel{d}{=} N_{p+1}(\mathbf{0}, \Sigma). \end{aligned}$$

The proof is complete.

**Proof of Theorem 3.2.** Since  $EH_0(t) = \rho A_{11}(t)$ ,  $EH_3(t) = \rho A_{21}(t)$ , and  $(\tilde{\theta}, \tilde{\beta})$  is strongly consistent, applying a first-order Taylor expansion and Theorem 3.1 gives, uniformly in  $t$ ,

$$\begin{aligned} \tilde{G}(t) &= \frac{1}{n_0} \sum_{i=1}^n \frac{I_{[T_i \leq t]}}{1 + \rho \exp[\tilde{\theta} + s(T_i; \tilde{\beta})]} \\ &= H_1(t) - H_0(t)(\tilde{\theta} - \theta_0) - H_3^T(t)(\tilde{\beta} - \beta_0) + o_p(\delta_n) \\ &= H_1(t) - (EH_0(t), EH_3^T(t)) \begin{pmatrix} \tilde{\theta} - \theta_0 \\ \tilde{\beta} - \beta_0 \end{pmatrix} - R_{n1}(t) + o_p(\delta_n) \\ &= H_1(t) - \frac{\rho}{n} (A_{11}(t), A_{21}^T(t)) S^{-1} \begin{pmatrix} \frac{\partial\ell(\theta_0,\beta_0)}{\partial\theta} \\ \frac{\partial\ell(\theta_0,\beta_0)}{\partial\beta} \end{pmatrix} + o_p(n^{-1/2}) - R_{n1}(t) + o_p(\delta_n) \\ &= H_1(t) - H_2(t) + R_n(t), \end{aligned}$$

where  $\delta_n = |\tilde{\theta} - \theta_0| + \|\tilde{\beta} - \beta_0\|$  and

$$\begin{aligned} R_{n1}(t) &= [H_0(t) - EH_0(t), H_3^T(t) - EH_3^T(t)] \begin{pmatrix} \tilde{\theta} - \theta_0 \\ \tilde{\beta} - \beta_0 \end{pmatrix}, \\ R_n(t) &= o_p(n^{-1/2}) - R_{n1}(t) + o_p(\delta_n). \end{aligned}$$

It follows from part (b) of Theorem 3.1 that  $\delta_n = O_p(n^{-1/2})$ . Moreover, it can be shown by Assumption (A2) and Lemma 7.2 that  $\sup_{-\infty \leq t \leq \infty} |R_{n1}(t)| = o_p(n^{-1/2})$ . As a result,  $\sup_{-\infty \leq t \leq \infty} |R_n(t)| = o_p(n^{-1/2})$ . To prove weak convergence, it suffices to show that  $\sqrt{n}(H_1 - \hat{G} - H_2) \xrightarrow{D} W$  in  $D[-\infty, \infty]^p$ . According to (3.4) and Lemma 7.3, we have

$$E\{\sqrt{n}[H_1(t) - \hat{G}(t) - H_2(t)]\} = 0,$$

$$\begin{aligned}
 & \text{Cov}(\sqrt{n}[H_1(s) - \hat{G}(s) - H_2(s)], \sqrt{n}[H_1(t) - \hat{G}(t) - H_2(t)]) \\
 &= \text{Cov}(\sqrt{n}[H_1(s) - \hat{G}(s)], \sqrt{n}[H_1(t) - \hat{G}(t)]) - \text{Cov}(\sqrt{n}H_2(s), \sqrt{n}H_2(t)) \\
 &= \rho(1 + \rho)A_{11}(s \wedge t) - \rho(1 + \rho)^2 A_{11}(s)A_{11}(t) \\
 &\quad - \rho(1 + \rho)(A_{11}(s), A_{21}^\tau(s))A^{-1} \begin{pmatrix} A_{11}(t) \\ A_{21}(t) \end{pmatrix} + \rho(1 + \rho)^2 A_{11}(s)A_{11}(t) \\
 &= \rho(1 + \rho)A_{11}(s \wedge t) - \rho(1 + \rho)(A_{11}(s), A_{21}(s))A^{-1} \begin{pmatrix} A_{11}(t) \\ A_{21}(t) \end{pmatrix} = EW(s)W(t).
 \end{aligned}$$

It then follows from the Central Limit Theorem for sample means and the Cramer-Wold device that the finite-dimensional distributions of  $\sqrt{n}(H_1 - \hat{G} - H_2)$  converge weakly to those of  $W$ . Thus, in order to prove  $\sqrt{n}(H_1 - \hat{G} - H_2) \xrightarrow{\mathcal{D}} W$  in  $D[-\infty, \infty]^p$ , it is enough to show that the process  $\{\sqrt{n}[H_2(t) - \hat{G}(t) - H_2(t)], -\infty \leq t \leq \infty\}$  is tight in  $D[-\infty, \infty]^p$ . But this has been established by Lemma 7.4. The proof is complete.

**Proof of Theorem 4.1.** For part (a), since  $(\tilde{\theta}, \tilde{\beta})$  is strongly consistent for estimating  $(\theta_0, \beta_0)$  by part (a) of Theorem 3.1,  $\Theta_0$  in Assumption (A1) will contain  $\tilde{\beta}$  for sufficiently large  $n$ , with probability 1. Furthermore, the Strong Law of Large Numbers yields  $\tilde{q}_{1j} \equiv \int_{-\infty}^{\infty} Q_1^j(y)\{1 + \rho \exp[\tilde{\theta} + s(y; \tilde{\beta})]\}d\tilde{G}(y) = \frac{1}{n_0} \sum_{i=1}^n Q_1^j(T_i) \xrightarrow{\text{a.s.}} q_{1j}$  for  $j = 1, 2, 3$ . Similarly, we have  $\tilde{q}_{2j} \equiv \int_{-\infty}^{\infty} Q_2^j(y)\{1 + \rho \exp[\tilde{\theta} + s(y; \tilde{\beta})]\}d\tilde{G}(y) \xrightarrow{\text{a.s.}} q_{2j}$  for  $j = 1, 2$  and  $\tilde{q}_3 \equiv \int_{-\infty}^{\infty} Q_3(y)\{1 + \rho \exp[\tilde{\theta} + s(y; \tilde{\beta})]\}d\tilde{G}(y) \xrightarrow{\text{a.s.}} q_3$ . Thus,  $\tilde{q}_{11}, \tilde{q}_{12}, \tilde{q}_{13}, \tilde{q}_{21}, \tilde{q}_{22}$ , and  $\tilde{q}_3$  are finite for sufficiently large  $n$ , with probability 1, and hence Assumptions (A1)–(A4) hold almost surely if we replace  $\beta_0$  by  $\tilde{\beta}$  and  $G$  by  $\tilde{G}$ . Moreover, according to part (c) of Lemma 7.6,  $A_n$  is positive definite for sufficiently large  $n$  with probability 1, and  $\Sigma_n \xrightarrow{\text{a.s.}} \Sigma$  as  $n \rightarrow \infty$ . Consequently, it follows from part (b) of Theorem 3.1 and the Bootstrap Central Limit Theorem for sample means (Bickel and Freedman (1981) or Singh (1981)) that (4.1) and (4.2) hold with probability one.

For part (b), the representation with remainder can be proved by employing Theorem 3.1 and Lemmas 7.5 and 7.6. To prove weak convergence, it is enough to show that with probability one as  $n \rightarrow \infty$ ,  $\sqrt{n}(H_1^* - \hat{G}^* - H_2^*) \xrightarrow{\mathcal{D}} W$  in  $D[-\infty, \infty]^p$ . It can be shown after extensive algebra that

$$\begin{aligned}
 & E_{\tilde{G}, \tilde{H}}\{\sqrt{n}[H_1^*(t) - \hat{G}^*(t) - H_2^*(t)]\} = 0, \quad -\infty \leq t \leq \infty, \\
 & \text{Cov}_{\tilde{G}, \tilde{H}}(\sqrt{n}[H_1^*(s) - \hat{G}^*(s) - H_2^*(s)], \sqrt{n}[H_1^*(t) - \hat{G}^*(t) - H_2^*(t)]) \\
 &= \rho(1 + \rho) \left[ A_{n11}(s \wedge t) - (A_{n11}(s), A_{n21}^\tau(s))A_n^{-1} \begin{pmatrix} A_{n11}(t) \\ A_{n21}(t) \end{pmatrix} \right], \quad -\infty \leq s, t \leq \infty.
 \end{aligned}$$

Since  $\text{Cov}_{\tilde{G}, \tilde{H}}(\sqrt{n}[H_1^*(s) - \hat{G}^*(s) - H_2^*(s)], \sqrt{n}[H_1^*(t) - \hat{G}^*(t) - H_2^*(t)]) \xrightarrow{\text{a.s.}} EW(s)W(t)$  by (3.4) and Lemma 7.6, it follows from the Bootstrap Central

Limit Theorem for sample means and the Cramer-Wold device that the finite-dimensional distributions of  $\sqrt{n}(H_1^* - \hat{G}^* - H_2^*)$  converge weakly to those of  $W$  almost surely. Thus, in order to show  $\sqrt{n}(H_1^* - \hat{G}^* - H_2^*) \xrightarrow{\mathcal{D}} W$  in  $D[-\infty, \infty]^p$ , it suffices to show that the process  $\{\sqrt{n}[H_1^*(t) - \hat{G}^*(t) - H_2^*(t)], -\infty \leq t \leq \infty\}$  is tight in  $D[-\infty, \infty]^p$  almost surely. But this can be proved by making use of a similar approach as in the proof of Lemma 7.4. The proof of Theorem 4.1 is complete.

### Acknowledgement

I am grateful to two referees and the Chair-Editor for their many valuable comments and constructive suggestions that have greatly improved my original submission.

### References

- Adler, R. J. (1981). *The Geometry of Random Fields*. Wiley, New York.
- Agresti, A. (1990). *Categorical Data Analysis*. Wiley, New York.
- Anderson, J. A. (1972). Separate sample logistic discrimination. *Biometrika* **59**, 19-35.
- Anderson, J. A. (1979). Robust inference using logistic models. *Internat. Statist. Inst. Bull.* **48**, 35-53.
- Bickel, P. J. and Freedman, D. A. (1981). Some asymptotic theory for the bootstrap. *Ann. Statist.* **9**, 1196-1217.
- Bickel, P. J. and Krieger, A. M. (1989). Confidence bands for a distribution function using the bootstrap. *J. Amer. Statist. Assoc.* **84**, 95-100.
- Billingsley, P. (1968). *Convergence of Probability Measures*. Wiley, New York.
- Bliss, C. I. (1935). The calculation of the dosage-mortality curve. *Ann. Appl. Biol.* **22**, 134-67.
- Dabrowska, D. M. and Doksum, K. A. (1988). Estimation and testing in a two-sample generalized odds-rate model. *J. Amer. Statist. Assoc.* **83**, 744-749.
- Farewell, V. (1979). Some results on the estimation of logistic models based on retrospective data. *Biometrika* **66**, 27-32.
- Gilbert, P., Lele, S. and Vardi, Y. (1999). Maximum likelihood estimation in semiparametric selection bias models with application to AIDS vaccine trials. *Biometrika* **86**, 27-43.
- Gill, R. D., Vardi, Y. and Wellner, J. A. (1988). Large sample theory of empirical distributions in biased sampling models. *Ann. Statist.* **16**, 1069-1112.
- Hosmer, D. J. and Lemeshow, S. (1989). *Applied Logistic Regression*. Wiley, New York.
- Hsieh, D. A., Manski, C. F. and McFadden, D. (1985). Estimation of response probabilities from augmented retrospective observations. *J. Amer. Statist. Assoc.* **80**, 651-662.
- Johnson, R. A. and Wichern, D. W. (1998). *Applied Multivariate Statistical Analysis*, 4th edition. Prentice Hall, Upper Saddle River, New Jersey.
- Murphy, S. A., Rossini, A. J. and van der Vaart, A. W. (1997). Maximum likelihood estimation in the proportional odds model. *J. Amer. Statist. Assoc.* **92**, 968-976.
- Owen, A. B. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika* **75**, 237-249.
- Owen, A. B. (1990). Empirical likelihood confidence regions. *Ann. Statist.* **18**, 90-120.
- Prentice, R. L. and Pyke, R. (1979). Logistic disease incidence models and case-control studies. *Biometrika* **66**, 403-411.

- Qin, J. (1993). Empirical likelihood in biased sample problems. *Ann. Statist.* **21**, 1182-1196.
- Qin, J. and Lawless, J. (1994). Empirical likelihood and general estimating equations. *Ann. Statist.* **22**, 300-325.
- Qin, J. and Zhang, B. (1997). A goodness of fit test for logistic regression models based on case-control data. *Biometrika* **84**, 609-618.
- Singh, K. (1981). On the asymptotic accuracy of Efron's bootstrap. *Ann. Statist.* **9**, 1187-1195.
- Tsirel'son, V. S. (1975). The density of the distribution of the maximum of a Gaussian process. *Theory Probab. Appl.* **20**, 847-856.
- van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes with Applications to Statistics*. Springer, New York.
- Vardi, Y. (1982). Nonparametric estimation in presence of length bias. *Ann. Statist.* **10**, 616-620.
- Vardi, Y. (1985). Empirical distribution in selection bias models. *Ann. Statist.* **13**, 178-203.
- Wacholder, S. and Weinberg, C. R. (1994). Flexible maximum likelihood methods for assessing joint effects in case-control studies with complex sampling. *Biometrics* **50**, 350-357.
- Weinberg, C. R. and Sandler, D. P. (1991). Randomized recruitment in case-control studies. *Amer. J. Epidemiology* **134**, 421-433.
- Weinberg, C. R. and Wacholder, S. (1993). Prospective analysis of case-control data under general multiplicative-intercept risk models. *Biometrika* **80**, 461-465.

Department of Mathematics, The University of Toledo, Toledo, OH 43606, U.S.A.

E-mail: bzhang@math.utoledo.edu

(Received September 1997; accepted November 1999)