Statistica Sinica **24** (2014), 121-146
doi:http://dx.doi.org/10.5705/ss.2012.091

SEMIPARAMETRIC ESTIMATION OF TREATMENT EFFECTS IN TWO SAMPLE PROBLEMS WITH CENSORED DATA

Fangfang Bai[1,2], Jian Huang[3] and Yong Zhou[2,4]

[1] *University of International Business and Economics,*
[2] *Chinese Academy of Sciences,* [3] *University of Iowa*
and [4] *Shanghai University of Finance and Economics*

Abstract: The problem of estimating treatment effects with censored two-sample data is of importance in survival analysis and has received much attention in the literature. A common procedure for dealing with censoring is the inverse probability weighted method. However, this method only uses information from uncensored data and can suffer from loss of efficiency. In this paper, we propose a unified semiparametric estimating equation approach to estimate various types of treatment effects with censored data, including the mean difference between two populations, the difference between two survival times at a given point, the probability that the survival time from one population is greater than that from the other, and the difference in mean residual life times, among others. Our approach uses all the available data, thus it typically leads to gains in efficiency as compared with the existing methods. We study the theoretical properties of the proposed estimator and derive its consistent variance estimator. Our simulation studies demonstrate that the proposed method tends to work better than the existing ones in finite sample settings. We also analyze a data set to illustrate its application.

Key words and phrases: Treatment effect, semiparametric model, estimating equation, censored data, two sample problem.

1. Introduction

The estimation of treatment effects based on two sample data is a common and important problem in survival analysis and biomedical studies. For example, in comparing the efficacy of two treatments, we are often interested in the difference between the probabilities of survival for at least a certain length of time, or whether the probability that the survival time of the patients under treatment 1 is greater than that under treatment 2. Such questions are not directly addressed by the existing two-sample test methods, such as log-rank and other rank-based tests for censored data (Fleming and Harrington (1991)), since they are constructed for testing general differences between two survival functions, not for these specific questions.

There has been much work on estimating treatment effects with two sample data and a variety of methods have been proposed. For example, Goddard and Hinberg (1990) proposed a method assuming data are normally distributed; Campbell and Ratnaparkhi (1993) studied a receiver operating characteristic curve (ROC) based procedure when data follow a Lomax distribution; Li, Tiwari, and Wells (1996) proposed a control percentile test, a chi-squared test, and a Kolmogorov-type test for comparing two distributions with censored survival data; Hollander and Korwar (1982) developed a nonparametric Bayesian estimation of the horizontal distance between the distributions of two populations.

Several authors have considered a semiparametric approach for estimating certain specific treatment effects based on two-sample data. In this approach, a parametric model is assumed for one sample and a nonparametric one is used for the other. This is attractive in many two-sample problems. For instance, when evaluating the effect of a new treatment with respect to a standard treatment in medical studies, more is often known about the standard treatment, since it may have been used for a long time and a parametric model can often be formulated based on historic data. In contrast, because there is limited data on a new treatment, a nonparametric model is more appropriate. Qin (1997) studied various differences in the two populations by an empirical likelihood approach in the semiparametric model. Li, Tiwari, and Wells (1999) considered such an approach for two-sample problems with one group modeled parametrically and the other nonparametrically. Li and Lin (2009) conducted a semiparametric maximum likelihood estimation of the parametric and nonparametric components based on a semiparametric mixture model. Hsieh and Turnbull (1996) considered semiparametric estimation of the ROC assuming one of the samples follows a normal distribution. Zhou and Liang (2005) proposed a general approach for treatment estimation based on estimating equations with inverse probability weights for censored data. They studied the inference of treatment effects based on normal approximation and empirical likelihood-based methods. They observed that the empirical likelihood-based method performs better than the normal approximation method in their simulation studies.

We propose a new approach for estimating treatment effects based on semiparametric estimating equations without resorting to inverse probability weighting, and using all the observed data. Our method is reminiscent of the Buckley-James method for the accelerated failure time models (Buckley and James (1979)), in which the censored observations are estimated by their conditional expectations given all the observed data. In this way, the information contained in the uncensored observations as well as censored observations is fully utilized. We show that the proposed estimator is consistent and asymptotically normal under mild regularity conditions. We also demonstrate via simulation studies that the proposed method outperforms existing ones.

The rest of the paper is organized as follows. In Section 2 we describe the proposed method. In Section 3 we discuss its asymptotic properties. In Section 4 we use simulation studies to evaluate the finite sample performance of the proposed method and compare it with existing ones. In Section 5, a data set is analyzed to illustrate the application of the proposed method. Concluding remarks are given in Section 6. The proofs of the theorems are given in the appendix.

2. Semiparametric Estimation of Treatment Effects

2.1. The model and motivating examples

Let X^0 and Y^0 be independent nonnegative random variables with survival functions S and D, respectively. Consider the random censorship model, in which X^0 and Y^0 are subject to right censoring. Denote the censoring variables by U and V, assumed to be independent with survival functions K and Q and independent of X^0 and Y^0. Suppose we can only observe (X, δ) and (Y, ϖ), where $X = \min\{X^0, U\}$, $\delta = 1\{X^0 \leq U\}$, $Y = \min\{Y^0, V\}$ and $\varpi = 1\{Y^0 \leq V\}$. Denote the distribution functions of X^0 and Y^0 by $F(\cdot) = 1 - S(\cdot)$ and $G(\cdot) = 1 - D(\cdot)$, respectively.

A semiparametric approach is often reasonable when there is more information on one of the two samples. Here we assume that $G(\cdot)$ is of known form depending on a d-dimensional parameter θ, $G(y) = G_\theta(y)$. Let $g_\theta(y)$ be its density function. Suppose the distribution of the censoring variable V does not depend on θ.

The observed data consists of two samples, $(X_i, \delta_i), 1 \leq i \leq n$, which are independent and identically distributed as (X, δ); and $(Y_j, \varpi_j), 1 \leq j \leq m$, which are independent and identically distributed as (Y, ϖ). Let $\varsigma = m/n$ and $\tau_1 = \sup\{s : S(s) > 0\}$.

Let Δ be the parameter representing the treatment effect of interest. Assume that there exists a function φ such that

$$E_F[\varphi(X^0, \Delta_0; \theta_0)] = 0, \tag{2.1}$$

where Δ_0 and θ_0 are the true values of parameters Δ and θ, respectively. $\varphi(\cdot, \cdot, \cdot)$ is called an unbias estimating function. We can estimate Δ by constructing estimating equations based on $\varphi(X^0, \Delta, \theta_0)$, and $\varphi(X^0, \Delta; \theta_0)$ can be easily constructed in most applications.

Example 1 (Estimation of the mean difference)**.** Here $\Delta = E(X^0) - E(Y^0)$ and we can take $\varphi(X^0, \Delta, \theta_0) = X^0 - EY^0 - \Delta$. When Y^0 follows a exponential distribution with mean θ_0, then $\varphi(X^0, \Delta, \theta_0) = X^0 - \theta_0 - \Delta$. When Y^0 follows a Weibull distribution with parameter $\theta_0 = (\theta_1, \theta_2)$, then $\varphi(X^0, \Delta; \theta_0) = X^0 - \theta_1 \Gamma_1(1 + 1/\theta_2) - \Delta$, where $\Gamma_1(\cdot)$ is the gamma function.

Example 2 (Estimation of the difference between two survival functions)**.** For a given time point t_0, let $\Delta = S(t_0) - (1 - G_{\theta_0}(t_0))$. Then $\varphi(X^0, \Delta; \theta_0) = I(X^0 > t_0) - (1 - G_{\theta_0}(t_0)) - \Delta$. When Y^0 follows a exponential distribution with mean θ_0, then $\varphi(X^0, \Delta; \theta_0) = I(X^0 > t_0) - \exp(-t_0/\theta_0) - \Delta$. When Y^0 follows a Weibull distribution with parameter $\theta_0 = (\theta_1, \theta_2)$, then $\varphi(X^0, \Delta; \theta_0) = I(X^0 > t_0) - \exp(-(t_0/\theta_1)^{\theta_2}) - \Delta$.

Example 3 (Estimation of the probability that the failure time in one population is greater than that in the other)**.** For this problem, $\Delta = P(X^0 < Y^0)$. We construct the estimating equation as $\varphi(X^0, \Delta; \theta_0) = 1 - G_{\theta_0}(X^0) - \Delta$. When Y^0 follows a exponential distribution with mean θ_0, then $\varphi(X^0, \Delta; \theta_0) = \exp(-X^0/\theta_0) - \Delta$. When Y^0 follows a Weibull distribution with parameter $\theta_0 = (\theta_1, \theta_2)$, then $\varphi(X^0, \Delta; \theta_0) = \exp(-(X^0/\theta_1)^{\theta_2}) - \Delta$.

Example 4 (Difference of mean residual life times)**.** Mean residual life function (MRLF) is an important characteristic of a failure time, describing the remaining life expectancy of an individual who has survived up to time t. We consider the problem of estimating the difference between the residual life times in two samples in the presence of a covariate. Consider the multiplicative model

$$E\{X^0 - Y^0 | X^0 > t, Y^0 > t, W) = \alpha(t) g(\Delta' W),$$

where $g(\cdot)$ is a known function, W is a covariate. If $\Delta = 0$, then there is no difference between the treatment group and control group after adjusting for the covariate effect. Note that we have

$$\begin{aligned} &E\{(X^0 - t) | X^0 > t, W\} - E\{(Y^0 - t) | Y^0 > t, W\} \\ &\quad = E\{X^0 - Y^0 | X^0 > t, Y^0 > t, W) = \alpha(t) g(\Delta' W). \end{aligned}$$

Therefore, we can construct an estimating function as

$$\varphi(X^0, \Delta; \theta_0) = I(X^0 > t) I(Y^0 > t)[(X^0 - t) - m(t, \theta_0) - \alpha(t) g(\Delta' W)]$$

where $m(t, \theta_0) = E_{\theta_0}(Y^0 - t | Y^0 > t, W)$ is a known function, up to an unknown parameter θ_0.

2.2. A semiparametric estimating equation approach

To describe the proposed approach, we first assume θ_0 is known. If X^0 is fully observed, then by (2.1), we can use an estimating equation to estimate Δ :

$$\frac{1}{n} \sum_{i=1}^{n} \varphi(X_i^0, \Delta; \theta_0) = 0. \tag{2.2}$$

For an unknown θ_0, we replace it by a consistent estimator. However, with censored data, the value of X_i^0 associated with $\delta_i = 0$ is unobservable, so we cannot estimate Δ directly based on (2.2).

A commonly used procedure for dealing with censoring is based on the inverse probability weighted method, or IPW (Robins, Rotnitzky, and Zhao (1995)). Zhou and Liang (2005) provided an estimator based on this method. However, the IPW method makes use of the uncensored data only, and generally results in a loss of efficiency.

Rather than only using the uncensored data, we make use of all the information from the uncensored and censored data. In this way, a more efficient estimator can be obtained. We note that, in the presence of censoring, $\varphi(X_i^0, \Delta, \theta_0)$ is not an unbiased estimating function if only uncensored data is used. However, we can construct an unbiased estimating function based on all the observations, as follows. For simplicity, let $S =: S(\cdot)$ and

$$\phi(X_i, \delta_i, \Delta; \theta_0, S) = E\big[\varphi(X_i^0, \Delta; \theta_0)\big|X_i, \delta_i\big], \tag{2.3}$$

and then the statistical inference for the treatment effect parameter Δ can based on $\phi(X_i, \delta_i, \Delta, \theta_0, S)$. This is valid since

$$E\big[\phi(X, \delta, \Delta; \theta_0, S)\big] = E\big[\varphi(X^0, \Delta; \theta_0)\big] = 0,$$

and ϕ is an unbiased estimating function for censored data using all the observations.

Expression (2.3) can be rewritten as

$$\phi(X_i, \delta_i, \Delta; \theta_0, S) = \delta_i \varphi(X_i^0, \Delta; \theta_0) + (1 - \delta_i)\frac{\int_{X_i}^{\tau_1} \varphi(s, \Delta; \theta_0) dF(s)}{1 - F(X_i)}. \tag{2.4}$$

The estimating equation is

$$\frac{1}{n}\sum_{i=1}^{n} \phi(X_i, \delta_i, \Delta; \theta_0, S) = 0. \tag{2.5}$$

Here both θ_0 and S need to be estimated. When $\delta_i = 1$, that is, X_i^0 is observed, the term in (2.3) is $\varphi(X_i^0, \Delta; \theta_0)$; when $\delta_i = 0$, the term in (2.3) is estimated by

$$E\big[\varphi(X_i^0, \Delta; \theta_0)\big|X_i, \delta_i = 0\big] = \frac{\int_{X_i}^{\tau_1} \varphi(s, \Delta; \theta_0) dF(s)}{1 - F(X_i)}.$$

This is similar to the Buckley-James estimator in the accelerated failure time models (Buckley and James (1979)). See also Honoré, Khan, and Powell (2002), Zhou, Wan, and Wang (2008) for a similar approach to deal with missing data problems in the context of estimating equations.

In (2.5), Δ is the parameter of main interest, S and θ_0 are nuisance parameters. Since S and θ_0 are unknown, we need to estimate them. This can be done with the maximum likelihood estimate.

The likelihood function for the observed data is

$$\ell(F,\theta) = \prod_{i=1}^{n} \{F(X_i) - F(X_i-)\}^{\delta_i}(1-F(X_i))^{(1-\delta_i)} \prod_{j=1}^{m} g_\theta(Y_j)^{\varpi_j}(1-G_\theta(Y_j))^{(1-\varpi_j)}.$$

By maximizing the likelihood function $\ell(F,\theta)$, we can obtain the maximum likelihood estimator of (F,θ_0), denoted by $(\widehat{F},\widehat{\theta}_{MLE})$. It can be easily shown that $\widehat{\theta}_{MLE}$ is the maximum likelihood estimator based on the sample $\{(Y_j,\varpi_j), j = 1,\ldots,m\}$, and $\widehat{F}$ is the Kaplan-Meier estimator (Kaplan and Meier (1958)),

$$\widehat{F}(t) = 1 - \prod_{u\le t}\Big(1 - \frac{dN(u)}{Y(u)}\Big),$$

where $N(u) = \sum(X_i \le u, \delta_i = 1)$ and $Y(u) = \sum(X_i \ge u)$. Let $\widehat{S}(t) = 1 - \widehat{F}(t)$ be the Kaplan-Meier estimator of the survival function. By substituting $\widehat{S}$ and $\widehat{\theta}_{MLE}$ into (2.5), a feasible estimator of Δ can be obtained by solving

$$\frac{1}{n}\sum_{i=1}^{n} \phi(X_i,\delta_i,\Delta;\widehat{\theta}_{MLE},\widehat{S}) = 0. \tag{2.6}$$

This can be done with existing methods for estimating equations. In Examples 1-3, (2.6) can be solved explicitly, while in Example 4, (2.6) needs to be solved numerically.

3. Large Sample Properties of the Estimator

In (2.5) there are two unknown nuisance parameters, θ_0 and S, which make it difficult to derive the large sample properties of $\widehat{\Delta}$ directly. So we proceed in two steps. We first assume θ_0 is known, and replace the parameter S by the Kaplan-Meier estimator $\widehat{S}$. The large sample properties of $\widehat{\Delta}$ can be obtained. In the second step, we replace θ_0 by its maximum likelihood estimate and derive the properties of estimator $\widehat{\Delta}$ based on the result with a known θ_0.

Consider the situation where θ_0 is known. There is only one nuisance parameter S, which can be estimated by its Kaplan-Meier estimator $\widehat{S}$ and then, by solving (2.5), we can obtain the estimator $\widehat{\Delta}$. Under this assumption, the estimator is consistent and asymptotically normal.

Proposition 1. *Suppose θ_0 is known and the conditions* (A1)$-$(A4) *in the Appendix hold. Let the true value of Δ be Δ_0, $\widehat{S}$ be the Kaplan-Meier estimator of*

S, and $\widehat{\Delta}$ be the solution of the estimating equation

$$\frac{1}{n}\sum_{i=1}^{n}\phi(X_i,\delta_i,\Delta;\theta_0,\widehat{S})=0. \tag{3.1}$$

Then $\widehat{\Delta}\xrightarrow{\mathcal{P}}\Delta_0$.

Throughout, for a function $f(x,y)$, let $\dot{f}_x(x,y)\triangleq\partial f(x,y)/\partial x$. Let $M(s)=I\{X\le t,\delta=1\}-\int_0^t I(X\ge t)d\Lambda(s)$, $h(s)=S(s)K(s-)$ and

$$\eta=\int_0^{\tau_1}(1-K(s-))h^{-1}(s)\big[\varphi(s)S(s)+\int_s^{\tau_1}\varphi(t)dS(t)\big]dM(s).$$

Proposition 2. *Suppose θ_0 is known and the conditions* (A1)$-$(A4) *in the Appendix hold. Let the true value of Δ be Δ_0, $\widehat{S}$ be the Kaplan-Meier estimator of S, and $\widehat{\Delta}$ be the solution of the estimating equation*

$$\frac{1}{n}\sum_{i=1}^{n}\phi(X_i,\delta_i,\Delta;\theta_0,\widehat{S})=0. \tag{3.2}$$

Then

$$\sqrt{n}(\widehat{\Delta}-\Delta_0)\xrightarrow{\mathcal{L}}N(0,A^{-1}\Gamma A'^{-1}),$$

where $A=E[\dot{\phi}_\Delta(x_i,\delta_i,\Delta_0;\theta_0,S)]$, $\Gamma=E\big[(\phi(X,\delta,\Delta_0;\theta_0,S)+\eta)(\phi(X,\delta,\Delta_0;\theta_0,S)+\eta)'\big]$.

Next, we consider the more realistic situation where both θ_0 and S are unknown. In this case, we estimate θ_0 by its maximum likelihood estimator $\widehat{\theta}_{MLE}$ and estimate S by the Kaplan-Meier estimator $\widehat{S}$. Then by solving (2.6), we obtain the final estimator of Δ, the parameter of our main interest.

Theorem 1. *Suppose conditions* (A1)$-$(A8) *in the Appendix hold. Let $\widehat{\Delta}$ be the solution of the estimating equation*

$$\frac{1}{n}\sum_{i=1}^{n}\phi(X_i,\delta_i,\Delta;\widehat{\theta}_{MLE},\widehat{S})=0.$$

Then $\widehat{\Delta}$ is consistent and asymptotically normal with

$$\sqrt{n}(\widehat{\Delta}-\Delta_0)\xrightarrow{\mathcal{L}}N(0,\Sigma),$$

where $\Sigma=A^{-1}(\Gamma+B\Sigma_1B')A'^{-1}$, $A=E[\dot{\phi}_\Delta(x_i,\delta_i,\Delta_0;\theta_0,S)]$, $\Gamma=E\big[(\phi(X,\delta,\Delta_0,\theta_0,S)+\eta)(\phi(X,\delta,\Delta_0;\theta_0,S)+\eta)'\big]$, $\Sigma_1=I(\theta_0)^{-1}/\varsigma$ with $I(\theta)$ defined in (A8) *in the Appendix, and $B=E\big[\dot{\phi}_\theta(X,\delta,\Delta_0,\theta_0,S)\big]$.*

The proofs of Propositions 1 and 2 and Theorem 1 are given in the Appendix.

These results provide the theoretical justification for the proposed method, and can be used as the basis for statistical inference. For example, to construct confidence intervals for Δ, we first obtain consistent estimator of the variance Σ using the method described in Honoré, Khan, and Powell (2002): take $\widehat{\Sigma} = \widehat{A}^{-1}(\widehat{\Gamma} + \widehat{B}\widehat{\Sigma}_1\widehat{B}')\widehat{A}'^{-1}$, where $\widehat{A}, \widehat{B}, \widehat{\Gamma}$, and $\widehat{\Sigma}_1$ are consistent estimators of A, B, Γ and Σ_1, respectively. Here, $\widehat{K}$ is the Kalpan-Meier estimator of K, $\widehat{A}$ and $\widehat{B}$ are just sample analogous of A and B,

$$\widehat{A} = \frac{1}{n}\sum_{i=1}^{n}\dot{\phi}_{\Delta}(x_i, \delta_i, \widehat{\theta}_{MLE}, \widehat{\Delta}, \widehat{S}) \quad \text{and} \quad \widehat{B} = \frac{1}{n}\sum_{i=1}^{n}\dot{\phi}_{\theta}(x_i, \delta_i, \widehat{\theta}_{MLE}, \widehat{\Delta}, \widehat{S}).$$

For Γ and Σ_1, let

$$\begin{aligned}
\widehat{\phi}_i &= \delta_i\varphi(x_i, \widehat{\theta}_{MLE}, \widehat{\Delta}) + (1-\delta_i)\frac{\int_{x_i}^{\tau_1}\varphi(s, \widehat{\Delta}, \widehat{\theta}_{MLE})d\widehat{F}(s)}{1-\widehat{F}(x_i)}, \\
\widehat{\Lambda}(s) &= \Big(\frac{1}{n}\sum_{i=1}^{n}\delta_i I(x_i \le s)\Big)\Big(\frac{1}{n}\sum_{i=1}^{n} I(x_i \ge s)\Big)^{-1}, \\
\widehat{h}(s) &= \frac{1}{n}\sum_{i=1}^{n}\big(\delta_i I(x_i > s) + (1-\delta_i)I(x_i \ge s)\big), \\
\widehat{M}_i(s) &= I(x_i \le t, \delta = 1) - \int_0^t I(x_i \ge s)d\widehat{\Lambda}(s), \\
\widehat{\eta}_i &= \int_0^{\tau_1}(1-\widehat{K}(s-))\widehat{h}^{-1}(s)\big[\varphi(s, \widehat{\theta}_{MLE}, \widehat{\Delta})\widehat{S}(s) \\
&\quad + \int_s^{\tau_1}\varphi(t, \widehat{\theta}_{MLE}, \widehat{\Delta})d\widehat{S}(t)\big]d\widehat{M}_i(s).
\end{aligned}$$

We estimate Γ by $\widehat{\Gamma} = \frac{1}{n}\sum_{i=1}^{n}(\widehat{\phi}_i + \widehat{\eta}_i)(\widehat{\phi}_i + \widehat{\eta}_i)'$. Let $\ell_g(\theta, y_j) = g_\theta(y_j)^{\varpi_j}(1 - G_\theta(y_j))^{(1-\varpi_j)}$, then the information matrix $I(\theta_0)$ can be estimated by

$$\widehat{I}(\widehat{\theta}_{MLE}) = -\frac{1}{m}\sum_{j=1}^{m}\frac{\partial^2 \log \ell_g(\theta, y_j)}{\partial\theta\partial\theta^\tau}\Big|_{\widehat{\theta}_{MLE}}.$$

We estimate Σ_1 by $\widehat{\Sigma}_1 = n\widehat{I}(\widehat{\theta}_{MLE})/m$.

4. Simulation Studies

We conducted simulation studies to evaluate the finite sample performance of the proposed method and compare it with two methods studied in Zhou and Liang (2005): a method based on the normality assumption, and a method based on the empirical likelihood.

Table 1. Simulation results based on the Norm, EL, and New methods for exponential distributions. Here Norm refers to the method based on normal distribution, EL refers to the method based on empirical likelihood method, and New refers to the proposed method. Bias: the bias of the estimator; SD: the mean of the standard deviation of the estimator; SE: the standard error of the estimator; ECP(95%): the empirical 95% coverage probability; C(%): the censoring probability in the two samples.

Bias			SD			SE			ECP(95%)			C(%)	
Norm	EL	New	Norm	EL	New	Norm	EL	New	Norm	EL	New	X^0	Y^0
$n = 100$													
0.018	0.018	-0.003	0.0408	0.0408	0.0397	0.0442	0.0442	0.0405	90.4	94.0	94.2	37	33
0.019	0.019	-0.002	0.0402	0.0402	0.0392	0.0438	0.0438	0.0401	90.0	93.4	93.7	37	30
0.014	0.014	-0.002	0.0401	0.0401	0.0394	0.0423	0.0423	0.0402	91.6	93.2	94.2	31	33
0.014	0.014	-0.001	0.0395	0.0395	0.0388	0.0418	0.0418	0.0400	91.1	93.0	93.6	31	30
$n = 200$													
0.012	0.012	-0.002	0.0293	0.0293	0.0286	0.0305	0.0305	0.0292	92.2	93.4	94.0	37	33
0.011	0.011	-0.003	0.0289	0.0289	0.0282	0.0302	0.0302	0.0285	92.6	93.2	94.3	37	30
0.009	0.009	-0.001	0.0286	0.0286	0.0282	0.0295	0.0295	0.0283	92.8	94.0	94.4	31	33
0.007	0.007	-0.002	0.0282	0.0282	0.0278	0.0288	0.0288	0.0276	93.2	94.8	93.7	31	30

Simulation study 1. Let X^0 and Y^0 be independent random variables from the exponential distributions with mean 6 and 10, respectively. The censored random variables U and V were generated from the exponential distribution with mean c_x and c_y, respectively. The parameter c_x and c_y were used to control the censoring percentages of the two samples. Take the parameter of interest to be $\Delta = P(X^0 < Y^0)$ with estimating function $\psi(x, \Delta, \theta) = \exp(-x/\theta) - \Delta$. Simulation results are given in Table 1.

Simulation study 2. Let $W(\theta_1, \theta_2)$ denote a two-parameter Weibull distribution. We took $X^0 \sim W(10, 5)$ and $Y^0 \sim W(8, 2)$, with uniformly distributed censoring variables. The parameter of interest here is $\Delta = EX^0 - EY^0$ and the corresponding estimating function is $\psi(x, \Delta, \theta_1, \theta_2) = x - \theta_1\Gamma_1(1 + 1/\theta_2) - \Delta$, where $\Gamma_1(\cdot)$ is the gamma function. Different censoring percentages were considered. Similar to the first example, the parameter was estimated and then the results were compared with the methods in Zhou and Liang (2005). The simulation results are given in Table 2.

We considered sample sizes, $n = m = 100$ and 200. The number of replications was 1,000. We evaluated the performance of the proposed estimator, the Norm and the EL estimators. See Zhou and Liang (2005) for detailed description of the last two.

It can be seen that all three estimators are unbiased and there is a good agreement between SD and SE. The empirical 95% coverage probabilities are

Table 2. Simulation results based on the Norm, EL, and New methods for Weibull distributions. Here Norm refers to the method based on normal distribution, EL refers to the method based on empirical likelihood method, and New refers to the proposed method. Bias: the bias of the estimator; SD: the mean of the standard deviations of the estimator; SE: the standard error of the estimator; ECP(95%): the empirical 95% coverage probability; C(%): the censoring probability in the two samples.

Bias			SD			SE			ECP(95%)			C(%)	
Norm	EL	New	Norm	EL	New	Norm	EL	New	Norm	EL	New	X^0	Y^0
$n = 100$													
-0.012	-0.012	-0.015	0.5154	0.5154	0.5125	0.5234	0.5234	0.5239	94.5	92.9	94.3	47	35
-0.009	-0.009	-0.012	0.4896	0.4896	0.4865	0.4959	0.4959	0.4966	95.2	92.8	95.1	47	25
-0.006	-0.006	-0.007	0.5095	0.5095	0.5077	0.5147	0.5147	0.5145	95.0	93.2	95.0	40	35
-0.003	-0.003	-0.004	0.4834	0.4834	0.4815	0.4869	0.4869	0.4868	95.3	93.7	95.3	40	25
$n = 200$													
-0.013	-0.013	-0.014	0.3646	0.3646	0.3635	0.3653	0.3653	0.3652	95.4	95.5	95.1	47	35
-0.010	-0.010	-0.011	0.3473	0.3473	0.3461	0.3452	0.3452	0.3451	95.1	95.1	95.1	47	25
-0.012	-0.012	-0.013	0.3602	0.3602	0.3595	0.3580	0.3580	0.3581	95.3	95.2	95.2	40	35
-0.009	-0.009	-0.010	0.3425	0.3425	0.3417	0.3458	0.3458	0.3459	95.3	94.8	95.0	40	25

reasonably close to the nominal level. As expected, the SD and SE decrease when the sample size increases from 100 to 200. The SD of the Norm procedure is similar to that of the EL procedure. This is consistent with the results of Qin and Lawless (1994). Furthermore, the NEW method is relatively robust to the censoring percentages when comparing with the other two methods, especially in Simulation Study 1.

In general, the NEW approach performs better, with SD smaller than those from the Norm and EL procedures. In particular, the proposed method tends to have more accurate coverage probabilities that are closer to the nominal 95% level. Although improvements are generally modest, the pattern is consistent in almost all the cases we considered.

The simulation results suggest that the proposed method performs well in finite sample situations and provides a useful alternative to the existing methods for estimating treatment effects in two-sample problems with censored data.

5. Data Example

The data set is from a study on Primary biliary cirrhosis of the liver (PBC). PBC is a rare but fatal chronic disease, with prevalence about 50 cases per million population. A double-blinded randomized trial of PBC was conducted between January 1974 and May 1984 at the Mayo clinic. One aim of the trial was to study the effect of the drug D-penicillamine (DPCA) on PBC, compared with a

Table 3. Estimation results of PBC data. Norm, EL and New represent the Norm-based, EL-based and the proposed methods, respectively. Est: estimator of the probability of the survival time of the DPCA group less than that of the placebo group; Sd: standard deviation of the estimator; CI: confidence interval.

	Norm	EL	New
Est	0.4574	0.4574	0.4574
Sd	0.0420	0.0420	0.0384
95% CI	(0.37, 0.54)	(0.37, 0.54)	(0.38, 0.53)

placebo. At the time of registration, there were 424 patients, among whom 312 agreed to participate in the randomized trial. A detailed description of the data can be found in Fleming and Harrington (1991).

Let X^0 and Y^0 denote the survival time of the population with DPCA and placebo, respectively. We take the parameter of our interest to be $\Delta = P(X^0 < Y^0)$.

After examining the Q-Q plot of the observations from the placebo group (figure not shown), we used the exponential model for the placebo group in our analysis. We compared our proposed method with the Norm-based and EL-based methods. The estimated values of Δ and their standard deviations are listed in Table 3.

The 95% confidence intervals based on each of the methods contain 0.50, which suggest that the drug (DPCA) has no significant effect on this disease. This is consistent with the conclusion in Fleming and Harrington (1991). The results from the methods are similar, with the proposed method providing a slightly shorter confidence interval.

6. Concluding Remarks

The problem of comparing two treatment effects arises in many areas of statistical applications. This paper focuses on semiparametric inference of treatment effect when data are subject to censoring. It develops a general method based on estimating equations that can deal with a large class of two-sample problems. The proposed method makes full use of the data. Asymptotic properties of the proposed method are developed and simulation studies show that the proposed estimator performs well when compared with existing methods. A data example is also used to illustrate the application of the proposed method.

For simplicity, our method requires that the function $\varphi(x, \Delta; \theta_0)$ is continuous at θ_0 and Δ. If it is discontinuous at θ_0, our approach cannot be applied directly. For example, for estimating the ROC curve, we can use the function $\varphi(X^0, \theta_0, \Delta) = 1 - I\{X^0 \le G_{\theta_0}^{-1}(1-p)\} - \Delta$. However, this function is not differentiable. The theoretical results proved here are not applicable, and further work is needed to deal with such non-smooth estimating functions.

Acknowledgement

We wish to thank two anonymous referees, the associate editor, and Editor for their helpful comments that led to considerable improvement in the paper. Zhou's work was partially supported by National Natural Science Foundation of China (NSFC) (No.71271128), National Natural Science Funds for Distinguished Young Scholar (No. 70825004), Science Fund for Creative Research Groups (11021161), NCMIS.

Appendix

In the Appendix, we prove the main results stated in the Section 3. We require some technical assumptions.

(A1) The true parameter Δ_0 is an interior point of the parameter space Θ, which is compact.

(A2) $\tau_1 \leq \tau_K$, where $\tau_K = \sup\{s : K(s) > 0\}$, and $F(X^0 = \tau_1) - F(X^0 = \tau_1-) > 0$.

(A3) $\varphi(x, \Delta, \theta)$ is continuous differentiable at Δ, with $\int_0^{\tau_1} \varphi^2(s)/K(s)dF(s) < \infty$, and for an integral function $L(s)$,

$$|\varphi(s, \Delta_1, \theta) - \varphi(s, \Delta_2, \theta)| \leq L(s)|\Delta_1 - \Delta_2|,$$

with $\int L(s)dF(s) < \infty$ and $\int L^2(s)/K(s)dF(s) < \infty$.

(A4) $\dot{\varphi}_\Delta(X, \Delta, \theta)$ is continuous and bounded by some function $d(x)$ in a neighborhood of the true value Δ_0 and θ_0, and $\int d(s)dF(s) < \infty$, and $E\dot{\varphi}_\Delta(X, \Delta_0, \theta_0) \neq 0$.

(A5) $\varphi(x, \Delta, \theta)$ and $\dot{\varphi}_\theta(x, \Delta, \theta)$ are continuous and bounded by some integral function $d_2(x)$ in a neighborhood of the true value θ_0 for any Δ.

(A6) $\dot{\varphi}_\theta(x, \Delta, \theta)$ is continuous in some neighborhood of the true value Δ_0 for any θ, and $E\dot{\varphi}_\theta(X, \Delta, \theta) \neq 0$.

(A7) The density function $g_\theta(y)$ is three times differentiable with respect to θ on $A = \{y : g_\theta(y) > 0\}$. For any $y \in A$ and θ, there exists a function $M_1(y)$ satisfying $E_\theta|M_1(Y)| < \infty$.

(A8) The element $I_{i,j}$ in the information matrix $I(\theta)$ has an expression

$$I_{i,j} = -\int_0^{\tau_2} \frac{\partial^2 \log g_\theta(y)}{\partial\theta_i \partial\theta_j} Q(y) g_\theta(y) dy - \int_0^{\tau_2} \frac{\partial^2 \log\{1 - G_\theta(y)\}}{\partial\theta_i \partial\theta_j} \{1 - G_\theta(y)\} dQ(y),$$

for $i, j = 1, \ldots, d$, where $\tau_2 = \sup\{t : D(t)Q(t) > 0\}$. $I(\theta)$ is continuous and positive definite.

The compactness condition on the parameter space usually holds in applications. The condition $\tau_1 \leq \tau_K$ on the upper support of the survival function S ensures that S can be estimated properly on $[0, \tau_1]$. The positive mass on the upper boundary of its support, ensures that $(1-\delta)/\widehat{S}(X_i)$ behaves well in the analysis of the large sample properties, because $\widehat{S}(X_i) > 0$ with $\delta = 0$. In fact, this condition can be satisfied by artificially censoring all observations at some point in the observed support of $\{X_i\}$. The Lipschitz condition (A3) simplifies the proof and is satisfied in most applications. The conditions $\int_0^{\tau_1} \varphi^2(s)/K(s)dF(s) < \infty$ guarantees that the variance of our proposed estimator exists. Conditions (A4)-(A6) are common in the literatures and hold for general cases. Assumptions (A7)$-$(A8) are regularity conditions needed for the maximum likelihood estimator of the parameter θ when data are right censored.

Before proving the main results, we state several lemmas.

Lemma A.1. *For any* $0 < r < 1/2$,

$$\sup_{t \in \mathbf{R}^+} \left| \widehat{S}(t) - S(t) \right| = o_p(n^{-r}).$$

Proof. This is Lemma A.4. of Honoré, Khan, and Powell (2002).

Lemma A.2. *Let* $\widehat{\Delta}$ *be a solution of the estimating equation of the form*

$$\frac{1}{n} \sum_{i=1}^{n} \psi(X_i, \Delta) = 0,$$

where $X_1, \ldots, X_n$ *are i.i.d.* X. *Let* $q(\Delta) = E\psi(X, \Delta)$. *Suppose that* Δ_0 *is the unique value such that* $q(\Delta_0) = 0$. *Furthermore assume that the parameter space* Θ *is compact and* $q(\Delta)$ *is continuous with respect to* Δ. *If*

$$\sup_{\Delta \in \Theta} \left\| \frac{1}{n} \sum_{i=1}^{n} \psi(X_i, \Delta) - q(\Delta) \right\| = o_p(1),$$

then $\widehat{\Delta} \xrightarrow{p} \Delta_0$.

Proof. Since

$$\sup_{\Delta \in \Theta} \left\| \frac{1}{n} \sum_{i=1}^{n} \psi(X_i, \Delta) - q(\Delta) \right\| = o_p(1),$$

if $\mathscr{N}(\Delta_0, \delta^*)$ is a neighborhood of Δ_0 for any given $\delta^* > 0$, we have

$$\sup_{\Delta \notin \mathscr{N}(\Delta_0, \delta^*)} \left\| \frac{1}{n} \sum_{i=1}^{n} \psi(X_i, \Delta) - q(\Delta) \right\| = o_p(1).$$

It follows that

$$\inf_{\Delta\notin\mathscr{N}(\Delta_0,\delta^*)}\Big\|\frac{1}{n}\sum_{i=1}^{n}\psi(X_i,\Delta)\Big\|\geq\inf_{\Delta\notin\mathscr{N}(\Delta_0,\delta^*)}\|q(\Delta)\|-o_p(1)>M,$$

where M is some positive constant. Hence $\widehat{\Delta}\in\mathscr{N}(\Delta_0,\delta^*)$, so $\widehat{\Delta}$ is consistent.

Proof of Proposition 1. Let $q(\Delta)=E[\varphi(X^0,\theta_0,\Delta)]$, and suppose $\widehat{\Delta}$ solves the estimating equation

$$\frac{1}{n}\sum_{i=1}^{n}\phi(X_i,\delta_i,\theta_0,\Delta,\widehat{S})=0.$$

In order to get the consistency of $\widehat{\Delta}$, by Lemma A.2, it suffices to show that

$$\sup_{\Delta\in\Theta}\Big\|\frac{1}{n}\sum_{i=1}^{n}[\phi(X_i,\delta_i,\theta_0,\Delta,\widehat{S})-q(\Delta)]\Big\|\xrightarrow{\mathcal{P}}0. \tag{A.1}$$

First note that the Kaplan-Meier estimator satisfies Lemma A.1 and, for all $\Delta\in\Theta$,

$$\begin{aligned}&\sup_{\Delta\in\Theta}\Big\|\frac{1}{n}\sum_{i=1}^{n}[\phi(X_i,\delta_i,\theta_0,\Delta,\widehat{S})-\phi(X_i,\delta_i,\theta_0,\Delta,S)]\Big\|\\&\leq\sup_{\Delta\in\Theta}\Big\|\frac{1}{n}\sum_{i=1}^{n}(1-\delta_i)\frac{\int_{X_i}^{\tau_1}\varphi(s,\Delta,\theta_0)d(\widehat{F}(s)-F(s))}{S(X_i)}\Big\|\\&\quad+\sup_{\Delta\in\Theta}\Big\|\frac{1}{n}\sum_{i=1}^{n}(1-\delta_i)\int_{X_i}^{\tau_1}\varphi(s,\Delta,\theta_0)dF(s)S(X_i)^{-2}(S(X_i)-\widehat{S}(X_i))\Big\|+o_p(1).\\&=:J_{11}+J_{12}\end{aligned}\tag{A.2}$$

Partition Θ into m nonoverlapping regions $\Theta_1^m,\Theta_2^m,\ldots,\Theta_m^m$ and choose m enough large such that $\max_{1\leq j\leq m}\rho(\Theta_j^m)<\epsilon_1$, for any $\epsilon_1>0$, where $\rho(\cdot)$ is the distance between any two points in parameter space Θ_j^m. Let $\Delta_1,\Delta_2,\ldots,\Delta_m$ be an arbitrary sequence such that $\Delta_i\in\Theta_i^m$, as similar to the proof of Theorem 4.2.1 of Amemiya (1985). Then we have, for any $\varepsilon>0$,

$$\begin{aligned}&P\Big(\sup_{\Delta\in\Theta}\Big\|\frac{1}{n}\sum_{i=1}^{n}(1-\delta_i)\frac{\int_{X_i}^{\tau_1}\varphi(s,\Delta,\theta_0)d(\widehat{F}(s)-F(s))}{S(X_i)}\Big\|>\varepsilon\Big) \qquad \text{(A.3)}\\&\leq\sum_{j=1}^{m}P\Big(\frac{1}{n}\sum_{i=1}^{n}\Big\|(1-\delta_i)\frac{\int_{X_i}^{\tau_1}\varphi(s,\Delta_j,\theta_0)d(\widehat{F}(s)-F(s))}{S(X_i)}\Big\|>\frac{\varepsilon}{2}\Big)\\&+\sum_{j=1}^{m}P\Big(\frac{1}{n}\sum_{i=1}^{n}\sup_{\Delta\in\Theta_j^m}\Big\|(1-\delta_i)\frac{\int_{X_i}^{\tau_1}(\varphi(s,\Delta,\theta_0)-\varphi(s,\Delta_j,\theta_0))d(\widehat{F}(s)-F(s))}{S(X_i)}\Big\|>\frac{\varepsilon}{2}\Big).\end{aligned}$$

Now we show the first part in (A.3) is $o_p(1)$.

Note that for all observations with $\delta = 0$, $S(X)$ is bounded away from zero, so

$$\begin{aligned}\Big[\Big\|(1-\delta)\frac{\int_X^{\tau_1}\varphi(s,\Delta_j,\theta_0)d(\widehat{F}(s)-F(s))}{S(X)}\Big\|\Big]\\ \le C\sup_{s\in(0,\tau_1]}\big|\varphi(s,\Delta_j,\theta_0)\big|\cdot\bigvee_0^{\tau_1}(\widehat{F}(s)-F(s)),\end{aligned}$$

where C is some constant and $\bigvee_0^{\tau_1}(\widehat{F}(s)-F(s))$ denotes the total variation of $(\widehat{F}(s)-F(s))$ on the interval $[0,\tau_1]$. Since $(\widehat{F}(s)-F(s))$ is a function of bounded variation, we can choose a partition of $[0,\tau_1]: 0 = z_1,\ldots,z_{m_0} = \tau_1$ such that $m_0 = O(n^{1/4})$ and

$$\begin{aligned}\bigvee_0^{\tau_1}(\widehat{F}(s)-F(s)) &= \sum_{j=1}^{m_0-1}\Big|\widehat{F}(z_{j+1})-F(z_{j+1})-\widehat{F}(z_j)+F(z_j)\Big|\\ &\le 2m_0\cdot\sup_{s\in[0,\tau_1]}\big|\widehat{F}(s)-F(s)\big|.\end{aligned}$$

By Lemma A.1, $\sup_{s\in[0,\tau_1]}\big|\widehat{F}(s)-F(s)\big| = o_p(n^{-1/3})$ for $r=\frac{1}{3}$, so we find

$$\bigvee_0^{\tau_1}(\widehat{F}(s)-F(s)) = o_p(1).$$

On the other hand it is easy to show that $\sup_{s\in(0,\tau_1]}\big|\varphi(s,\Delta_j,\theta_0)\big|$ is bounded almost sure everywhere, therefore

$$\Big\|\frac{1}{n}\sum_{i=1}^{n}(1-\delta_i)\frac{\int_{X_i}^{\tau_1}\varphi(s,\Delta_j,\theta_0)d(\widehat{F}(s)-F(s))}{S(X_i)}\Big\| = o_p(1). \tag{A.4}$$

Next we will show the second part in (A.3) is $o_p(1)$.

Let

$$\begin{aligned}u(s,\Delta) &= (\varphi(s,\Delta,\theta_0)-\varphi(s,\Delta_j,\theta_0))I(X_i<s),\ F_n^*(s)=\widehat{F}(s)-F(s),\\ u_1(s,\Delta) &= \varphi(s,\Delta,\theta_0)-\varphi(s,\Delta_j,\theta_0).\end{aligned}$$

Under assumption (A3) and the property of Θ_j^m, we have

$$\sup_{\Delta\in\Theta_j^m}\big|u(s,\Delta)\big|\to 0,\quad \sup_{\Delta\in\Theta_j^m}\big|u_1(s,\Delta)\big|\to 0, \tag{A.5}$$

in probability.

Divide the interval $[0, \tau_1]$ into subintervals $[r_k, r_{k+1}]$, $k = 1, \ldots, k_n + 1$ where $k_n = O(n^{1/3})$. Then

$$\begin{aligned}
&\sup_{\Delta\in\Theta_j^m} \Big\| \int_0^{\tau_1} u(s,\Delta) dF_n^*(s) \Big\| = \sup_{\Delta\in\Theta_j^m} \Big\| \sum_{i=1}^{k_n} \int_{r_i}^{r_{i+1}} u(s,\Delta) dF_n^*(s) \Big\| \\
&\quad \le 2 \sup_{\Delta\in\Theta_j^m} \sup_i \sup_{r_i \le s \le r_{i+1}} \Big| u_1(s,\Delta) - u_1(r_{i+1},\Delta) \Big| + 4 \sup_{\Delta\in\Theta_j^m} \sup_i \Big| u_1(r_{i+1},\Delta) \Big| \\
&\qquad + \sup_{\Delta\in\Theta_j^m} k_n \sup_i \Big| F_n^*(r_{i+1}) - F_n^*(r_i) \Big| \sup_i \Big| u_1(r_{i+1},\Delta) \Big|,
\end{aligned}$$

By Lemma A.1, we get $k_n \sup_i \big| F_n^*(r_{i+1}) - F_n^*(r_i) \big| = O_p(1)$, because $S(X_i) > 0$ with $\delta_i = 0$ and (A.5), and then

$$\sup_{\Delta\in\Theta_j^m} \Big\| \frac{(1-\delta_i) \int_{X_i}^{\tau_1} (\varphi(s,\Delta,\theta_0) - \varphi(s,\Delta_j,\theta_0)) d(\widehat{F}(s) - F(s))}{S(X_i)} \Big\| \xrightarrow{\mathcal{P}} 0.$$

Hence,

$$\frac{1}{n} \sum_{i=1}^n \sup_{\Delta\in\Theta_j^m} \Big\| (1-\delta_i) \frac{\int_{X_i}^{\tau_1} (\varphi(s,\Delta,\theta_0) - \varphi(s,\Delta_j,\theta_0)) d(\widehat{F}(s) - F(s))}{S(X_i)} \Big\| = o_p(1). \tag{A.6}$$

Combining (A.3), (A.4), and (A.6),

$$J_{11} = \sup_{\Delta\in\Theta} \Big\| \frac{1}{n} \sum_{i=1}^n (1-\delta_i) \frac{\int_{X_i}^{\tau_1} \varphi(s,\Delta,\theta_0) d(\widehat{F}(s) - F(s))}{S(X_i)} \Big\| \xrightarrow{\mathcal{P}} 0. \tag{A.7}$$

Note that $S(X_i) > 0$ when $\delta_i = 0$ and

$$\begin{aligned}
&\sup_{\Delta\in\Theta} \Big\| \frac{1}{n} \sum_{i=1}^n (1-\delta_i) \int_{X_i}^{\tau_1} \varphi(s,\Delta,\theta_0) dF(s) S(X_i)^{-2} \Big\| \\
&\quad \xrightarrow{\mathcal{P}} \sup_{\Delta\in\Theta} \Big\| E\big[(1-\delta_i) E(I(X_i < X)\varphi(X,\Delta,\theta_0)) S(X_i)^{-2}\big] \Big\| \\
&\quad \le C_1 \sup_{\Delta\in\Theta} E \big| \varphi(X,\Delta,\theta_0) \big| = O(1),
\end{aligned}$$

where C_1 is some positive constant. Since $\sup_{s\in R^+} \big| S(s) - \widehat{S}(s) \big| = o_p(1)$, we get

$$J_{12} = \sup_{\Delta\in\Theta} \Big\| \frac{1}{n} \sum_{i=1}^n (1-\delta_i) \int_{X_i}^{\tau_1} \varphi(s,\Delta,\theta_0) dF(s) S(X_i)^{-2} (S(X_i) - \widehat{S}(X_i)) \Big\| \xrightarrow{\mathcal{P}} 0. \tag{A.8}$$

Hence, (A.7) together with (A.8) yield

$$\sup_{\Delta\in\Theta}\Big\|\frac{1}{n}\sum_{i=1}^{n}[\phi(X_i,\delta_i,\theta_0,\Delta,\widehat{S})-\phi(X_i,\delta_i,\theta_0,\Delta,S)]\Big\| \xrightarrow{\mathcal{P}} 0.$$

The conditions of a Uniform Law of Large Numbers is satisfied by assumptions, and this implies $\sup_{\Delta\in\Theta}\|(1/n)\sum_{i=1}^{n}\phi(X_i,\delta_i,\theta_0,\Delta,\widehat{S})-q(\Delta)\| \xrightarrow{\mathcal{P}} 0$. Now (A.1) holds, so the estimator $\widehat{\Delta}$ is consistent.

Lemma A.3. *With $X = X^0 \wedge U$, $\delta = I(X^0 \le U)$, $H(t) = P(X \ge t)$, and $\Lambda(\cdot)$ the cumulative hazard function of X^0, if*

$$N(t)=\sum_{i=1}^{n} I\{X_i\le t,\delta_i=1\}, M_j(t)=I\{x_j\le t,\delta_j=1\}-\int_0^t I(x_j\ge s)d\Lambda(s),$$

$$Y(t)=\sum_{i=1}^{n} I(X_i\ge t),\quad \overline{M}(t)=\sum_{i=1}^{n} M_j(t)=N(t)-\int_0^t Y(s)d\Lambda(s),$$

then for the Kaplan-Meier estimator,

$$\begin{aligned}\widehat{S}(t)-S(t)&=-S(t)\frac{1}{n}\int_0^t(1-\Delta\Lambda(s))^{-1}H(s)^{-1}d\overline{M}(s)+R_n(t)\\&=-S(t)\frac{1}{n}\int_0^t h(s)^{-1}d\overline{M}(s)+R_n(t),\end{aligned}$$

where $h(s) = S(s)K(s-)$, and $\sup_{0\le t<\infty}|R_n(t)| = o_p(n^{-1/2})$.

Proof. The proof follows from Lemma A.2 of Honoré, Khan, and Powell (2002).

Lemma A.4. *Under* (A2)*, we have*

$$\frac{1}{n}\sum_{i=1}^{n}[\phi(X_i,\delta_i,\theta_0,\Delta_0,\widehat{S})-\phi(X_i,\delta_i,\theta_0,\Delta_0,S)]=\frac{1}{n}\sum_{j=1}^{n}\eta_j+o_p(n^{-1/2}),\quad (A.9)$$

where

$$\eta_j=\int_0^{\tau_1}(1-K(s-))h(s)^{-1}\big[\varphi(s)S(s)+\int_s^{\tau_1}\varphi(t)dS(t)\big]dM_j(s).$$

Proof. The left-hand side of (A.9) is

$$\begin{aligned}LHS&=\frac{1}{n}\sum_{i=1}^{n}(1-\delta_i)\left\{\frac{\int_{X_i}^{\tau_1}\varphi(s,\Delta_0,\theta_0)d\widehat{F}(s)}{\widehat{S}(X_i)}-\frac{\int_{X_i}^{\tau_1}\varphi(s,\Delta_0,\theta_0)dF(s)}{S(X_i)}\right\}\\&=\frac{1}{n}\sum_{i=1}^{n}(1-\delta_i)\big\{\frac{\int_{X_i}^{\tau_1}\varphi(s,\Delta_0,\theta_0)d(\widehat{F}(s)-F(s))}{S(X_i)}\\&\quad+\int_{X_i}^{\tau_1}\varphi(s,\Delta_0,\theta_0)dF(s)S(X_i)^{-2}[S(X_i)-\widehat{S}(X_i)]\big\}+o_p(n^{-1/2}).\quad (A.10)\end{aligned}$$

For suitable U, V, and W,

$$\begin{cases} U(t)V(t) = U(0)V(0) + \int_0^t U(s-)dV(s) + \int_0^t V(s)dU(s), \\ d\{W(s)\}^{-1} = -\{W(s)W(s-)\}dW(s). \end{cases}$$

If $U(t) = \widehat{F}(t) - F(t), V(t) = S(t)^{-1}, W(t) = S(t) = V(t)^{-1}$, then we have

$$d(\widehat{F}(t) - F(t)) = S(t)d\left(\frac{\widehat{F}(t) - F(t)}{S(t)}\right) + \frac{S(t-) - \widehat{S}(t-)}{S(t-)}dS(t).$$

Lemma A.3 implies that

$$d(\widehat{F}(t) - F(t)) = \frac{1}{n}S(t)h(t)^{-1}d\overline{M}(t) + \frac{S(t-) - \widehat{S}(t-)}{S(t-)}dS(t) + o_p(n^{-1/2}). \quad \text{(A.11)}$$

Thus, (A.10) is

$$\begin{aligned} LHS = &\frac{1}{n}\int_0^{\tau_1} \Big\{\frac{1}{n}\sum_{i=1}^{n}(1-\delta_i)S(X_i)^{-1}I(X_i < s)\Big\}\varphi(s, \Delta_0, \theta_0)S(s)h(s)^{-1}d\overline{M}(s) \\ &+\frac{1}{n}\sum_{i=1}^{n}(1-\delta_i)S(X_i)^{-1}\int_{X_i}^{\tau_1}\varphi(s, \Delta_0, \theta_0)\Big\{\frac{\widehat{S}(X_i)}{S(X_i)} - \frac{\widehat{S}(s-)}{S(s-)}\Big\}dS(s) \\ &+o_p(n^{-1/2}). \end{aligned} \quad \text{(A.12)}$$

Since $\widehat{S}(X_i)/S(X_i) - \widehat{S}(t-)/S(t-) = (1/n)\int_{X_i}^{t-} h(s)^{-1}d\overline{M}(s) + o_p(n^{-1/2})$, so the second term in last equation can be written as

$$\frac{1}{n}\sum_{i=1}^{n}(1-\delta_i)S(X_i)^{-1}\frac{1}{n}\int_{X_i}^{\tau_1} h(s)^{-1}\int_s^{\tau_1}\varphi(t, \Delta_0, \theta_0)dS(t)d\overline{M}(s).$$

Noting that the $i = j$ term is asymptotically negligible, (A.12) yields a U statistic:

$$\frac{1}{n}\sum_{i=1}^{n}(1-\delta_i)S(X_i)^{-1}\frac{1}{n-1}\int_{X_i}^{\tau_1}h(s)^{-1}\big[\varphi(s)S(s)+\int_s^{\tau_1}\varphi(t)dS(t)\big]d\overline{M}(s)+o_p(n^{-1/2}).$$

The kernel function of the U statistic is denoted by $\mathcal{F}_1(Z_{1i}, Z_{1j})$ where $Z_{1i} = (X_i, \delta_i)$.
Note that

$$\begin{aligned} &\frac{2}{n}\sum_{j=1}^{n}E\big[\mathcal{F}_1(Z_{1i}, Z_{1j})|Z_{1j}\big] \\ &\quad= \frac{1}{n}\int_0^{\tau_1} E\Big[(1-\delta_i)S^{-1}(X_i)I(X_i < s)\Big]\Big[\varphi(s)S(s) + \int_s^{\tau_1}\varphi(t)dS(t)\Big]d\overline{M}(s). \end{aligned}$$

Under (A2) and (A3) one has $E\big[\mathcal{F}_1(Z_{1i}, Z_{1j})\big] = 0$, $E^2\big[\mathcal{F}_1(Z_{1i}, Z_{1j})\big] < \infty$; by a standard projection theorem for U statistics, $LHS = (1/n)\sum_{j=1}^{n} \eta_j + o_p(n^{-1/2})$, where

$$\begin{aligned}\eta_j &= \int_0^{\tau_1} E\big[(1-\delta)S(X)^{-1}I(X<s)\big]h(s)^{-1}\big[\varphi(s)S(s) + \int_s^{\tau_1}\varphi(t)dS(t)\big]dM_j(s)\\ &= \int_0^{\tau_1}(1-K(s-))h(s)^{-1}\big[\varphi(s)S(s) + \int_s^{\tau_1}\varphi(t)dS(t)\big]dM_j(s).\end{aligned}$$

The proof is completed.

Lemma A.5. *Under* (A1)$-$(A4)*, for any* $r \in (0, 1/2)$*, then we have* $|\widehat{\Delta} - \Delta_0| = o_p(n^{-r})$*, where* $\widehat{\Delta}$ *is the solution of* (3.1).

Proof. It follows from a Taylor expansion of (3.1) that

$$\frac{1}{n}\sum_{i=1}^{n}\phi(X_i,\delta_i,\theta_0,\Delta_0,\widehat{S}) + \Big(\frac{1}{n}\sum_{i=1}^{n}\dot{\phi}_\Delta(X_i,\delta_i,\theta_0,\Delta_0,\widehat{S}) + o_p(1)\Big)(\widehat{\Delta}-\Delta_0) = 0.$$

After transformation the last equation is:

$$\begin{aligned}&\frac{1}{n}\sum_{i=1}^{n}\phi(X_i,\delta_i,\theta_0,\Delta_0,S) + \frac{1}{n}\sum_{i=1}^{n}\Big[\phi(X_i,\delta_i,\theta_0,\Delta_0,\widehat{S}) - \phi(X_i,\delta_i,\theta_0,\Delta_0,S)\Big]\\ &\quad + \Big(\frac{1}{n}\sum_{i=1}^{n}\dot{\phi}_\Delta(X_i,\delta_i,\theta_0,\Delta_0,\widehat{S}) + o_p(1)\Big)(\widehat{\Delta}-\Delta_0) = 0. \qquad \text{(A.13)}\end{aligned}$$

Lemma A.4 and the limit theory for i.i.d random variables imply that, for any $r \in (0, 1/2)$,

$$\Big\|\frac{1}{n}\sum_{i=1}^{n}\big[\phi(X_i,\delta_i,\theta_0,\Delta_0,\widehat{S}) - \phi(X_i,\delta_i,\theta_0,\Delta_0,S)\big]\Big\| = o_p(n^{-r}).$$

Similar to the proof of Proposition 1, we can show that

$$\Big\|\frac{1}{n}\sum_{i=1}^{n}\big[\dot{\phi}_\Delta(X_i,\delta_i,\theta_0,\Delta_0,\widehat{S}) - \dot{\phi}_\Delta(X_i,\delta_i,\theta_0,\Delta_0,S)\big]\Big\| = o_p(1),$$

and, together with the fact that $(1/n)\sum_{i=1}^{n}\dot{\phi}_{\Delta_0}(X_i,\delta_i,\theta_0,\Delta_0,S) \xrightarrow{\mathcal{P}} E\big[\dot{\varphi}_\Delta(X_i, \Delta_0, S)\big]$, we get $(1/n)\sum_{i=1}^{n}\dot{\phi}_\Delta(X_i,\delta_i,\theta_0,\Delta_0,\widehat{S}) \xrightarrow{\mathcal{P}} E\big[\dot{\varphi}_\Delta(X_i,\Delta_0,S)\big]$. In addition the first term in $(A.13) = O_p(n^{-1/2})$ and $E\big[\dot{\varphi}_\Delta(X_i,\Delta_0,S)\big] \neq 0$, so $|\widehat{\Delta}-\Delta_0| = o_p(n^{-r})$.

Lemma A.6. *Under* (A1)−(A4)*, we have*

$$\frac{1}{n}\sum_{i=1}^{n}\big[\phi(X_i,\delta_i,\theta_0,\widehat{\Delta},\widehat{S})-\phi(X_i,\delta_i,\theta_0,\widehat{\Delta},S)-\phi(X_i,\delta_i,\theta_0,\Delta_0,\widehat{S}) +\phi(X_i,\delta_i,\theta_0,\Delta_0,S)\big]=o_p(n^{-1/2}). \tag{A.14}$$

Proof. It is easy to show that the representation (A.14) is bounded by

$$\begin{aligned}&\Big\|\frac{1}{n}\sum_{i=1}^{n}\frac{1-\delta_i}{S(X_i)}\Big(\int_{X_i}^{\tau_1}\varphi(s,\widehat{\Delta},\theta_0)d\widehat{F}(s)-\int_{X_i}^{\tau_1}\varphi(s,\widehat{\Delta},\theta_0)dF(s)\\ &\quad-\int_{X_i}^{\tau_1}\varphi(s,\Delta_0,\theta_0)\widehat{F}(s)+\int_{X_i}^{\tau_1}\varphi(s,\Delta_0,\theta_0)dF(s)\Big)\Big\|\\ &\quad+\Big\|\frac{1}{n}\sum_{i=1}^{n}(1-\delta_i)S(X_i)^{-2}\int_{X_i}^{\tau_1}[\varphi(s,\widehat{\Delta},\theta_0)-\varphi(s,\Delta_0,\theta_0)]dF(s)\big(\widehat{S}(X_i)-S(X_i)\big)\Big\|\\ &\quad+o_p(n^{-1/2})\\ &=: I_1+I_2+o_p(n^{-1/2}).\end{aligned}$$

So it suffices to show that both I_1 and I_2 are $o_p(n^{-1/2})$.

We first show that $I_2=o_p(n^{-1/2})$. Note that $S(X_i)>0$ with $\delta_i=0$. Since $\varphi(X;\theta_0,\Delta)$ is Lipschitz with respect to Δ, it is Euclidean for the envelope $|\varphi(\cdot,\theta_0,\Delta^*)+M_1L(\cdot)|$ by Lemma 2.13 in Pakes and Pollard (1989), where Δ^* is some point of Θ and $M_1=2\sup_\Theta|\Delta-\Delta_0|$. Therefore, by Lemma 5 in Sherman (1994), the class of functions of X_i,δ_i, indexed by Δ,

$$(1-\delta_i)S(X_i)^{-2}\int_{X_i}^{\tau_1}(\varphi(s,\Delta,\theta_0)-\varphi(s,\Delta_0,\theta_0))dF(s), \tag{A.15}$$

is Euclidean for a constant envelope.

Next, we will show that $(1-\delta_i)S(X_i)^{-2}\int_{X_i}^{\tau_1}(\varphi(s,\Delta,\theta_0)-\varphi(s,\Delta_0,\theta_0))dF(s)$ is continuous with respect to Δ in L^2 space. Here

$$\begin{aligned}&\lim_{\widehat{\Delta}\to\Delta_0}E\Big((1-\delta_i)S(X_i)^{-2}\int_{X_i}^{\tau_1}(\varphi(s,\widehat{\Delta},\theta_0)-\varphi(s,\Delta_0,\theta_0))dF(s)\Big)^2\\ &\quad\le C\lim_{\widehat{\Delta}\to\Delta_0}\Big[\int_0^{\tau_1}L(s)dF(s)\Big]^2|\widehat{\Delta}-\Delta_0|^2,\end{aligned}$$

where C is a constant. This functional class is $\mathcal{L}^2(p)$ continuous with Δ. Let $\delta^*=1/3$. By Lemma 2.17 of Pakes and Pollard (1989), we have

$$\begin{aligned}&\sup_{\|\Delta-\Delta_0\|\le n^{-\delta^*}}\Big\|\frac{1}{n}\sum_{i=1}^{n}\Big((1-\delta_i)S(X_i)^{-2}\int_{X_i}^{\tau_1}(\varphi(s,\Delta,\theta_0)-\varphi(s,\Delta_0,\theta_0))dF(s)\Big)\\ &\quad-E\Big((1-\delta_i)S(X_i)^{-2}\int_{X_i}^{\tau_1}(\varphi(s,\Delta,\theta_0)-\varphi(s,\Delta_0,\theta_0))dF(s)\Big)\Big\|=o_p(n^{-1/2}).\end{aligned}$$

On the other hand, by the Lipschitz assumption on $\varphi(x, \Delta, \theta)$,

$$\sup_{\|\Delta-\Delta_0\|\le n^{-\delta*}} \left\| E((1-\delta_i)S(X_i)^{-2} \int_{X_i}^{\tau_1} (\varphi(s,\Delta,\theta_0) - \varphi(s,\Delta_0,\theta_0))dF(s)) \right\|$$
$$= O_p(\|\Delta - \Delta_0\|).$$

With $r = 1/3$, by Lemma A.5 we have

$$\sup_{\|\widehat{\Delta}-\Delta_0\|\le n^{-\delta*}} \left\| \frac{1}{n} \sum_{i=1}^{n} ((1-\delta_i)S(X_i)^{-2} \int_{X_i}^{\tau_1} (\varphi(s,\widehat{\Delta},\theta_0) - \varphi(s,\Delta_0,\theta_0))dF(s) \right\|$$
$$= o_p(n^{-r}). \tag{A.16}$$

Combine the fact that $\sup_{s\in R^+} \left|\widehat{S}(s) - S(s)\right| = o_p(n^{-r})$, and $(A.16)$ to obtain $I_2 = o_p(n^{-1/2})$.

Now have to show that $I_1 = o_p(n^{-1/2})$. Using (A.11),

$$\begin{aligned} I_1 &= \left\| \frac{1}{n} \sum_{i=1}^{n} \frac{(1-\delta_i)}{S(X_i)} (\int_{X_i}^{\tau_1} (\varphi(s,\widehat{\Delta},\theta_0) - \varphi(s,\Delta_0,\theta_0))d(\widehat{F}(s) - F(s))) \right\| \\ &\le \left\| \frac{1}{n} \sum_{i=1}^{n} \frac{(1-\delta_i)}{S(X_i)} \int_{X_i}^{\tau_1} [\varphi(s,\widehat{\Delta},\theta_0) - \varphi(s,\Delta_0,\theta_0)] n^{-1} h(s)^{-1} S(s) d\overline{M}(s) \right\| \\ &\quad + \left\| \frac{1}{n} \sum_{i=1}^{n} \frac{(1-\delta_i)}{S(X_i)} (\int_{X_i}^{\tau_1} [\varphi(s,\widehat{\Delta},\theta_0) - \varphi(s,\Delta_0,\theta_0)] \big(1 - \frac{\widehat{S}(s-)}{S(s-)}\big) dS(s)) \right\| \\ &\quad + o_p(n^{-1/2}) \\ &=: I_{11} + I_{12} + o_p(n^{-1/2}). \end{aligned}$$

We first show $I_{12} = o_p(n^{-1/2})$. Let $r = 1/3$, since $\sup_{s\in R^+} |S(s) - \widehat{S}(s)| = o_p(n^{-r})$, $1-\delta_i$, $S(s-) > 0$ with $s < \tau_1$ and $(1-\delta_i)S(X_i)^{-1}$ are both bounded, it suffices to show that

$$\sup_{\|\Delta-\Delta_0\|\le n^{-\delta*}} \left\| \frac{1}{n} \sum_{i=1}^{n} \int_{X_i}^{\tau_1} (\varphi(s,\Delta,\theta_0) - \varphi(s,\Delta_0,\theta_0)) dS(s) \right\| = o_p(n^{-r}).$$

This holds by a similar method as used in the proof of equation (A.16).

Next we show $I_{11} = o_p(n^{-1/2})$. Note that it is a U statistic since the own observation terms are asymptotically negligible. Let $Z_i = (X_i, \delta_i, \Delta)$ and $\mathcal{F}(Z_i, Z_j, \Delta)$ be the kernel function of the U statistic and

$$\begin{aligned} \mathcal{F}(Z_i, Z_j, \Delta) &= \frac{1}{2} \frac{(1-\delta_i)}{S(X_i)} \int_{X_i}^{\tau_1} [\varphi(s,\Delta,\theta_0) - \varphi(s,\Delta_0,\theta_0)] h(s)^{-1} S(s) dM_j(s) \\ &\quad + \frac{1}{2} \frac{(1-\delta_j)}{S(x_j)} \int_{x_j}^{\tau_1} [\varphi(s,\Delta,\theta_0) - \varphi(s,\Delta_0,\theta_0)] h(s)^{-1} S(s) dM_i(s). \end{aligned}$$

Since

$$I_{11} \leq \Big\| \frac{2}{n} \sum_{j=1}^{n} E[\mathcal{F}(Z_i, Z_j, \widehat{\Delta})|Z_j] \Big\| + \Big\| \frac{2}{n(n-1)} \sum_{i<j} \mathcal{F}(Z_i, Z_j, \widehat{\Delta}) - E[\mathcal{F}(Z_i, Z_j, \widehat{\Delta})|Z_j] \Big\|,$$

with Lemma A.5, it is sufficient to show that

$$J_1 =: \sup_{\|\Delta-\Delta_0\|\leq n^{-\delta^*}} \Big\| \frac{2}{n} \sum_{j=1}^{n} E[\mathcal{F}(Z_i, Z_j, \Delta)|Z_j] \Big\| = o_p(n^{-1/2}), \tag{A.17}$$

$$J_2 =: \sup_{\|\Delta-\Delta_0\|\leq n^{-\delta^*}} \Big\| \frac{2}{n(n-1)} \sum_{i<j} \mathcal{F}(Z_i, Z_j, \Delta) - E[\mathcal{F}(Z_i, Z_j, \Delta)|Z_j] \Big\| = o_p(n^{-1/2}), \tag{A.18}$$

where $\delta^* = 1/3$.

Let $\xi_j(X, \Delta) = \int_0^{\tau_1} (1 - K(s-))\varphi(s, \Delta, \theta_0)h(s)^{-1}S(s)dM_j(s)$, so $\xi_j(X, \Delta)$ is a zero mean process. Then $J_1 = \|(1/n)\sum_{j=1}^{n} [\xi_j(x, \Delta) - \xi_j(x, \Delta_0)]\|$. By (A3), $\varphi(X; \theta_0, \Delta)$ satisfies a Lipschitz condition, and $(1 - K(s-)), h(s)^{-1}S(s), M(\cdot)$ is Euclidean for a constant envelope under (A1)$-$(A4), so $\xi_j(X, \Delta)$ is Euclidean for a constant envelope. Similar arguments as above can be used to establish an analogous $\mathcal{L}^2(p)$-continuity of $\xi_j(x, \Delta)$, specifically,

$$E\big[\xi_j(x, \Delta) - \xi_j(x, \Delta_0)\big]^2 \leq E\Big(\int_0^{\tau_1} (1 - K(s-))^2 K(s)^{-1} L^2(s) dF(s)\Big)(\Delta - \Delta_0)^2.$$

Hence, as $\Delta \to \Delta_0$, we get $\mathcal{L}^2(p)$-continuity of $\xi_j(x, \Delta)$, and by Lemma 2.17 of Pakes and Pollard (1989), we have

$$\sup_{\|\Delta-\Delta_0\|\leq n^{-\delta^*}} \Big\| \frac{1}{n} \sum_{j=1}^{n} \big[\xi_j(x, \Delta) - \xi_j(x, \Delta_0)\big] \Big\| = o_p(n^{-1/2}),$$

so (A.17) holds.

Now we show (A.18). The left hand side of (A.18) can be written as

$$J_2 = \sup_{\|\Delta-\Delta_0\|\leq n^{-\delta^*}} \Big\| \frac{1}{n} \sum_{j=1}^{n} \int_0^{\tau_1} \frac{1}{n-1} \sum_{i=1}^{j-1} \frac{1-\delta_i}{S(X_i)} I(X_i < s)(\varphi(s, \Delta, \theta_0) - \varphi(s, \Delta_0, \theta_0)) h(s)^{-1}S(s)dM_j(s) + \frac{1}{n}\sum_{i=1}^{n} \int_0^{\tau_1} \frac{1}{n-1} \sum_{j=i+1}^{n} \frac{1-\delta_j}{S(x_j)} I(x_j < s)(\varphi(s, \Delta, \theta_0) - \varphi(s, \Delta_0, \theta_0))h(s)^{-1}S(s)dM_i(s) - \frac{2}{n}\sum_{j=1}^{n} E[\mathcal{F}(Z_i, Z_j, \Delta)|Z_j] \Big\|$$

$$= \sup_{\|\Delta-\Delta_0\|\le n^{-\delta^*}} \Big\| \frac{1}{n}\sum_{j=1}^n \int_0^{\tau_1} \big(\frac{1}{n-1}\sum_{i=1,i\neq j}^n \frac{1-\delta_i}{S(X_i)} I(X_i<s) - (1-K(s-))\big)$$
$$\big(\varphi(s,\Delta,\theta_0)-\varphi(s,\Delta_0,\theta_0)\big)h(s)^{-1}S(s)dM_j(s)\Big\|.$$

Furthermore, $(1-\delta_i)/S(X_i)I(X_i<s)$ is Euclidean for a constant envelope by Lemma 2.8 in Pakes and Pollard (1989). We have for fixed j,

$$\sup_{s\in[0,\tau_1]} \Big\| \frac{1}{n-1}\sum_{i=1,i\neq j}^n \frac{1-\delta_i}{S(X_i)} I(X_i<s) - (1-K(s-)\Big\| = o_p(1).$$

Similar arguments as used for $\xi_j(X,\Delta)$ can be used to show that

$$\sup_{\|\Delta-\Delta_0\|\le n^{-\delta^*}} \Big\| \frac{1}{n}\sum_{j=1}^n \int_0^{\tau_1} \big(\varphi(s,\Delta,\theta_0)-\varphi(s,\Delta_0,\theta_0)\big)h(s)^{-1}S(s)dM_j(s)\Big\| = o_p(n^{-1/2}),$$

therefore $(A.18)$ holds. Hence, $I_1=o_p(n^{-1/2})$. This leads to the conclusion of Lemma A.6.

Proof of Proposition 2. First,

$$0=\frac{1}{n}\sum_{i=1}^n \phi(X_i,\delta_i,\theta_0,\widehat{\Delta},\widehat{S}) = \frac{1}{n}\sum_{i=1}^n \phi(X_i,\delta_i,\theta_0,\widehat{\Delta},S)$$
$$+\frac{1}{n}\sum_{i=1}^n \big[\phi(X_i,\delta_i,\theta_0,\Delta_0,\widehat{S})-\phi(X_i,\delta_i,\theta_0,\Delta_0,S)\big] + \frac{1}{n}\sum_{i=1}^n \Big[\phi(X_i,\delta_i,\theta_0,\widehat{\Delta},\widehat{S})$$
$$-\phi(X_i,\delta_i,\theta_0,\widehat{\Delta},S)-\phi(X_i,\delta_i,\theta_0,\Delta_0,\widehat{S})+\phi(X_i,\delta_i,\theta_0,\Delta_0,S)\Big].$$

By a Taylor expansion

$$\frac{1}{n}\sum_{i=1}^n \phi(X_i,\delta_i,\theta_0,\widehat{\Delta},S)$$
$$=\frac{1}{n}\sum_{i=1}^n \phi(X_i,\delta_i,\theta_0,\Delta_0,S) + \frac{1}{n}\sum_{i=1}^n \dot{\phi}_\Delta(X_i,\delta_i,\theta_0,\widetilde{\Delta},S)(\widehat{\Delta}-\Delta_0),$$

where $\widetilde{\Delta}$ lies between Δ_0 and $\widehat{\Delta}$. Together with Lemma A.4 and Lemma A.6, we get

$$\sqrt{n}(\widehat{\Delta}-\Delta_0)=\Big(\frac{1}{n}\sum_{i=1}^n \dot{\phi}_\Delta(X_i,\delta_i,\theta_0,\widetilde{\Delta},S)\Big)^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^n\big[\phi(X_i,\delta_i,\theta_0,\Delta_0,S)+\eta_i\big]+o_p(1).$$

Under (A4), it follows from the Law of Large Numbers and the consistency of $\widehat{\Delta}$ that

$$\frac{1}{n}\sum_{i=1}^n \dot{\phi}_\Delta(X_i,\delta_i,\theta_0,\widetilde{\Delta},S) \xrightarrow{\mathcal{P}} E[\dot{\phi}_\Delta(X_i,\delta_i,\theta_0,\Delta_0,S)].$$

Let $A = E[\dot{\phi}_\Delta(X_i, \delta_i, \theta_0, \Delta_0, S)]$ and $\Gamma = E\big[(\phi(X, \delta, \theta_0, \Delta_0, S)+\eta\big)(\phi(X, \delta, \theta_0, \Delta_0, S) + \eta\big)'\big]$. The Central Limit theorem of i.i.d random variables yields

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\big[\phi(X_i, \delta_i, \theta_0, \Delta_0, S) + \eta_i\big] \xrightarrow{\mathcal{L}} N(0, \Gamma),$$

so $\sqrt{n}(\widehat{\Delta} - \Delta_0) \xrightarrow{\mathcal{L}} N(0, A^{-1}\Gamma A'^{-1})$.

Proof of Theorem 1. Since θ_0 is unknown, we replace θ_0 with the MLE of θ_0, $\widehat{\theta}_{MLE}$ in the estimating equation. This implies $\widehat{\Delta}$ is the solution to

$$\frac{1}{n}\sum_{i=1}^{n}\phi(X_i, \delta_i, \widehat{\theta}_{MLE}, \widehat{\Delta}, \widehat{S}) = 0.$$

The observations for the parameter θ are $(Y_j,\ \varpi_j), j = 1, \ldots, m$, and the likelihood function about θ is $L(\theta; y) = \Pi_{j=1}^{m} g_\theta(y_j)^{I(\varpi_j=1)}(1 - G_\theta(y_j))^{I(\varpi_j=0)}$, with

$$\widehat{\theta}_{MLE} = \underset{\theta\in\Omega}{\operatorname{argmax}} \log L(\theta; y),$$

where Ω is the parameter space of θ.

From maximum likelihood theory, $\sqrt{m}(\widehat{\theta}_{MLE} - \theta_0) \xrightarrow{\mathcal{L}} N(0, I(\theta)^{-1})$, where $I(\theta)$ is defined in (A8).

To obtain the consistency of $\widehat{\Delta}$, it suffices to show that

$$\sup_{\Delta\in\Theta}\Big\|\frac{1}{n}\sum_{i=1}^{n}[\phi(X_i, \delta_i, \widehat{\theta}_{MLE}, \Delta, \widehat{S}) - h_1(\Delta)]\Big\| = o_p(1),$$

where $h_1(\Delta) = E\big[\phi(X_i, \delta_i, \theta_0, \Delta, S)\big]$.

After transformation and a Taylor expansion, the left hand side of the last equation is

$$\sup_{\Delta\in\Theta}\ \Big\|\frac{1}{n}\sum_{i=1}^{n}\big(\phi(X_i, \delta_i, \theta_0, \Delta, \widehat{S}) - h_1(\Delta) + [\dot{\phi}_\theta(X_i, \delta_i, \theta_0, \Delta, S) + \big(\dot{\phi}_\theta(X_i, \delta_i, \theta_0, \Delta, \widehat{S}) - \dot{\phi}_\theta(X_i, \delta_i, \theta_0, \Delta, S)\big) + o_p(1)](\widehat{\theta}_{MLE} - \theta_0)\big)\Big\|. \quad (A.19)$$

By (A.1) and (A5),

$$(A.19) \le \sup_{\Delta\in\Theta}\Big\|\frac{1}{n}\sum_{i=1}^{n}\big\{\phi(X_i, \delta_i, \theta_0, \Delta, \widehat{S}) - h_1(\Delta)\big\}\Big\| + \Big\{\sup_{\Delta\in\Theta}\Big\|\frac{1}{n}\sum_{i=1}^{n}\dot{\phi}_\theta(X_i, \delta_i, \theta_0, \Delta, S)\Big\| + \sup_{\Delta\in\Theta}\Big\|\frac{1}{n}\sum_{i=1}^{n}\big\{\dot{\phi}_\theta(X_i, \delta_i, \theta_0, \Delta, \widehat{S})$$

$$-\dot{\phi}_\theta(X_i,\delta_i,\theta_0,\Delta,S)\}\Big\|\Big\}\|\widehat{\theta}_{MLE}-\theta_0\|$$
$$=o_p(1)+(C_2+o_p(1))\|\widehat{\theta}_{MLE}-\theta_0\|,$$

where C_2 is some positive constant, since $\widehat{\theta}_{MLE}$ is a consistent estimator of θ_0. So $(A.19)=o_p(1)$, and $\widehat{\Delta}$ is consistent.

Note that both $\varphi(x,\theta,\delta)$, and $\dot{\varphi}_\theta(x,\theta,\delta)$ are continuous in some neighborhood of θ_0, thus we have

$$\begin{aligned}0&=\frac{1}{n}\sum_{i=1}^n\phi(X_i,\delta_i,\widehat{\theta}_{MLE},\widehat{\Delta},\widehat{S})\\&=\frac{1}{n}\sum_{i=1}^n\phi(X_i,\delta_i,\theta_0,\widehat{\Delta},\widehat{S})+\frac{1}{n}\sum_{i=1}^n\dot{\phi}_\theta(X_i,\delta_i,\widetilde{\theta},\widehat{\Delta},\widehat{S})(\widehat{\theta}_{MLE}-\theta_0),\end{aligned}$$

where $\widetilde{\theta}$ lies between $\widehat{\theta}_{MLE}$ and θ. Let $\Sigma_1=(1/\varsigma)I(\theta)^{-1}$, and $B=E\big[\dot{\phi}_\theta(X,\delta,\theta_0,\Delta_0,S)\big]$. Since $\widehat{\theta}_{MLE}$ is a consistent estimator of θ_0, with (A5) and (A6), we can show that

$$\frac{1}{n}\sum_{i=1}^n\dot{\phi}_\theta(X_i,\delta_i,\widetilde{\theta},\widehat{\Delta},\widehat{S})\xrightarrow{\mathcal{P}}B.$$

Again using a Taylor expansion with respect to Δ_0, we have

$$\begin{aligned}\sqrt{n}(\widehat{\Delta}-\Delta_0)=&-A^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^n[\phi(X_i,\delta_i,\theta_0,\Delta_0,S)+\eta_i]-A^{-1}\sqrt{n}B(\widehat{\theta}_{MLE}-\theta_0)\\&+o_p(1).\end{aligned}$$

Since X^0 and Y^0 are independent, the terms on the right hand side are uncorrelated. Hence

$$\sqrt{n}(\widehat{\Delta}-\Delta_0)\xrightarrow{\mathcal{L}}N(0,A^{-1}(\Gamma+B\Sigma_1B')A'^{-1}).$$

This completes the proof of **Theorem 1.**

References

Amemiya, T. (1985). *Advanced Econometrics.* Harvard University Press, Cambridge, MA.

Buckley, J. and James, I. (1979). Linear regression with censored data. *Biometrika* **66**, 429-436.

Campbell, G. and Ratnaparkhi, M. V. (1993). An application of Lomax distributions in receiver operating characteristic (ROC) curve analysis. *Comm. Statist.* **22**, 1681-1697.

Fleming, T. R. and Harrington, D. P. (1991). *Counting Process and Survival Analysis.* Wiley, New York.

Goddard, M. J. and Hinberg, I. (1990). Receiver operating characteristic (ROC) curves and non-normal data: an empirical study. *Statist. Medicine* **9**, 325-337.

Hsieh, F. and Turnbull, B. W. (1996). Nonparametric and semiparametric estimation of the receiver operating characteristic curve. *Ann. Statist.* **24**, 25-40.

Hollander, M. J. and Korwar, R. W. (1982). Nonparametric bayesian estimation of the horizontal distance between two populations. In *Nonparametric Statistical Inference I* (Edited by B. V. Gnedenko, M. L. Puri and I. Vincze), 409-416. North Holland, New York.

Honoré, B., Khan, S., and Powell, J. L. (2002). Quantile regression under random censoring. *J. Econometrics* **109**, 67-105.

Kaplan, E. L. and Meier, P. (1958). Nonparametric estimation from incomplete observation. *J. Amer. Statist. Assoc.* **53**, 457-481.

Li, G. and Lin, C. (2009). Analysis of two-sample censored data using a semiparametric mixture model. *Acta Math. Appl. Sinica* **25**, 389-398.

Li, G., Tiwari, R. C. and Wells, M. T. (1996). Quantile comparison function in two-sample problems: with applications to comparisons of diagnostic markers. *J. Amer. Statist. Assoc.* **91**, 689-698.

Li, G., Tiwari, R. C. and Wells, M. T. (1999). Semiparametric inference for a quantile comparison function with applications to receiver operating characteristic curves. *Biometrika* **86**, 487-502.

Pakes, A. and Pollard, D. (1989). Simulation and the asymptotic of optimization estimators. *Econometrica* **57**, 1027-1058.

Qin, J. and Lawless, J. (1994). Empirical likelihood and general estimating equations. *Ann. Statist.* **22**, 300-325.

Qin, Y. S. (1997). Semi-parametric likelihood ratio confidence intervals for various differences of two populations. *Statist. Probab. Lett.* **33**, 135-143.

Robins, J. M., Rotnitzky, A. and Zhao, L. P. (1995). Analysis of semiparametric regression models for repeated outcomes in the presence of missing data. *J. Amer. Statist. Assoc.* **90**, 106-121.

Sherman, R. P. (1994). Maximal inequalities for degenerate U-processes with applications to optimization estimators. *Ann. Statist.* **22**, 439-459.

Zhou, Y. and Liang, H. (2005). Empirical likilihood-based semi-parametric inference of the treatment effect in the two-sample problem with censoring. *Biometrika* **92**, 271-282.

Zhou, Y., Wan, A. T. K. and Wang, X. (2008). Estimating equation inference with missing data. *J. Amer. Statist. Assoc.* **103**, 1187-1199.

School of Statistics, University of International Business and Economics, China.

Academy of Mathematics and Systems Science, Chinese Academy of Sciences, China.

E-mail: bff03021@163.com

Department of Statistics and Actuarial Science, University of Iowa, USA.

E-mail: jian-huang@uiowa.edu

Academy of Mathematics and Systems Science, Chinese Academy of Sciences, China.

School of Statistics and Management, Shanghai University of Finance and Economics, China.

E-mail: yzhou@amss.ac.cn

(Received March 2012; accepted November 2012)