I N S T I T U T D E

S T A T I S T I Q U E

UNIVERSITÉ CATHOLIQUE DE LOUVAIN

D I S C U S S I O N

P A P E R

0109

NONPARAMETRIC ANALYSIS OF COVARIANCE FOR CENSORED DATA

Y. DU, M.G. AKRITAS and I. VAN KEILEGOM

http://www.stat.ucl.ac.be

Nonparametric analysis of covariance for censored data

YUNLING DU

Division of Biostatistics, Columbia University School of Public Health,

New York, New York 10032, U.S.A.

ydu@biostat.cpmc.columbia.edu

MICHAEL G. AKRITAS

Department of Statistics, Penn State University, University Park,

Pennsylvania 16802, U.S.A.

mga@stat.psu.edu

INGRID VAN KEILEGOM

Université Catholique de Louvain

vankeilegom@stat.ucl.ac.be

20 March 2001

SUMMARY

The fully nonparametric model for nonlinear analysis of covariance, proposed in Akritas, Arnold & Du (2000), is considered in the context of censored observations. Under this model, the distributions for each factor level combination and covariate value are not restricted to comply to any parametric or semiparametric model. The data can be continuous or ordinal categorical. The possibility of different shapes of covariate effect in different factor level combinations is also allowed. This generality is useful whenever modeling assumptions such as additive risks, proportional hazards or proportional odds appear suspect. Test statistics for the nonparametric hypotheses of no main effects, and no interaction effects which adjust for the presence of a covariate are obtained. They are quadratic forms based on averages over the covariate values of Beran estimators of the conditional distribution of the survival time given each covariate value. The derivation of the asymptotic χ^2-distribution of the test statistics uses a recently obtained asymptotic representation of the Beran estimator as average of independent random variables. A real data set is analyzed and results of simulation studies are reported.

Some key words: Asymptotic representation; Nonparametric hypothesis; Nonparametric regression; Ordinal data; Ties; Unbalanced design.

1 Introduction

Exploring the effect that factors and covariates have on a response variable is central to many scientific investigations. When the response is subject to random censoring, this is mainly done with the use of various semiparametric models. The proportional hazards model (Cox 1972, 1975) is probably the most widely used model in the field of biostatistics. However, when the assumption of proportional hazards is violated, the parameters lose their interpretation. This has motivated interest in other models such as additive risks (Aalen, 1980; Huffer & McKeague, 1991; McKeague & Sasieni, 1994; Lin & Ying, 1994), and the proportional odds model (Bennett, 1983; Pettitt, 1984; Cheng, Wei & Ying, 1995; Yang & Prentice, 1999). While it might be hoped that one of these semiparametric models will hold in any given situation, there is no guarantee for it. In addition, the theory of partial likelihood for the proportional hazards model breaks down for left-censored data. Left censoring is common in ecology and astronomy where water or soil samples may have lower concentrations of a certain hazardous substance than can be reliably detected in laboratories, or the luminosity of a celestial object, from a preselected sample, needs longer than the allowed exposure time for accurate measurement. Motivated by a similar breakdown of the partial likelihood theory for right truncated data, Lagakos, Barraj & De Gruttola (1988) proposed the reverse time proportional hazards model. While it is now possible to use nonparametric regression methods to fit semiparametric models to any form of incomplete data (Akritas, 1996; Grigoletto & Akritas, 2000) the possibility that none of the common semiparametric models fit the data exists.

The purpose of this paper is to extend to censored data the fully nonparametric approach to analysis of covariance which was introduced in Akritas, Arnold & Du (2000). This method makes no modeling assumptions and thus it should be particularly useful in cases where the available semiparametric models are inappropriate. A fully nonparametric approach to the analysis of variance with censored data was developed in Akritas & Brunner (1997). The present extension to analysis of covariance requires a more general nonparametric model which allows for continuous covariates, as well as new asymptotic methods. In particular, while Akritas & Brunner (1997) used martingale-based arguments, the present theory makes use of asymptotic representations of the Beran (1981) estimator of the conditional distribution. In order to include all types of ordinal data it was necessary to develop a suitable asymptotic representation of the Beran estimator valid for arbitrary distributions; see Du and Akritas (2000) or Du (1999). It should be stressed that the results of this paper carry over to any incomplete data setting, e.g. truncated, censored and truncated, interval censored etc, for which a similar asymptotic representation for the conditional distribution can be established.

To review the nonparametric model and hypotheses, let (X_{ij}, Y_{ij}), $j = 1, \ldots, n_i$, denote the set of covariates and responses from cell i, $i = 1, \ldots, k$. The single index i can be used to enumerate the factor-level combinations of several factors, but for now consider a one-way analysis of covariance design. The nonparametric model specifies only that given $X_{ij} = x$,

$$Y_{ij}|X_{ij} = x \sim F_{ix}, \tag{1}$$

some distribution function that depends on i and x. We will be concerned with testing

$$H_0: \text{ the } F_{i\cdot}(y) \text{ are all equal,} \tag{2}$$

where

$$F_{i\cdot}(y) == \int F_{ix}(y)\, dG(x), \tag{3}$$

and G is a chosen weight function. Most of the development we will take G to be the overall distribution function of the covariate, as recommended in Akritas, Arnold & Du (2000). Additional choices, depending on the censoring pattern, are discussed in Remark 1 in the next section. A different perspective of the above nonparametric hypothesis, is offered by the unique decomposition

$$F_{ix}(y) = M(y) + A_i(y) + D_x(y) + C_{ix}(y), \tag{4}$$

where $\sum_{i=1}^k A_i(y) = 0$, for all y, $\int D_x(y)\, dG(x) = 0$, for all y, $\sum_{i=1}^k C_{ix}(y) = 0$, for all x and y, and $\int C_{ix}(y)\, dG(x) = 0$, for all i and y. Thus, $M(y) = k^{-1}\sum_{i=1}^k F_i(y)$, $A_i(y) = F_i(y) - M(y)$, $D_x(y) = k^{-1}\sum_{i=1}^k F_{ix}(y) - M(y)$, and $C_{ix}(y) = F_{ix}(y) - M(y) - A_i(y) - D_x(y)$. The hypothesis (2) is equivalent to the statement that the covariate-adjusted treatment effects A_i are all zero.

The above extends easily to multi-factor analysis of covariance designs. For example consider the case of two-factors, so $F_{ix} = F_{rcx}$, where r enumerates the rows and c enumerates the columns. Decomposition (4) is written as $F_{rcx}(y) = M(y) + A_{rc}(y) + D_x(y) + C_{rcx}(y)$. The nonparametric main and interaction effects, adjusted for the presence of a covariate, are obtained from the further decomposition of the covariate-adjusted effect of cell (r, c) as $A_{rc}(y) = A_r(y) + B_c(y) + (AB)_{rc}(y)$; see also Akritas, Arnold & Du (2000).

In the next section we outline the definitions and conventions that must be adopted for the censored data setting. The main results are stated in Section 3. Simulations and the data analysis are presented in Sections 4, 5, respectively, while the assumptions and proofs are contained in the Appendix.

2 Conventions

Let (X_{ij}, Y_{ij}) be as in the previous section and suppose that each Y_{ij} is censored by a random variable C_{ij} which is conditionally independent of Y_{ij} given X_{ij}. Thus, from cell i, $i = 1, \ldots, k$, we only observe $(X_{ij}, Z_{ij}, \Delta_{ij})$, $j = 1, \ldots, n_i$, where $Z_{ij} = \min(Y_{ij}, C_{ij})$ and $\Delta_{ij} = I(Y_{ij} \le C_{ij})$. Recall that the single index i is used for enumerating the factor-level combinations of possibly several factors.

Denote $G_i(x) = \mathrm{pr}\,(X_i \le x)$, $F_{ix}(y) = F_i(y|x) = \mathrm{pr}\,(Y_i \le y|X_i = x)$, $S_i(y|x) = 1 - F_i(y|x)$, $H_i(y|x) = \mathrm{pr}\,(Z_i \le y|X_i = x)$, and $H_{i1}(y|x) = \mathrm{pr}\,(Z_i \le y, \Delta_i = 1|X_i = x)$. Finally, denote $\hat{\lambda}_i = n_i/N$, where $N = \sum_{i=1}^k n_i$, and $\lambda_i = \lim_{n_i,N\to\infty} \hat{\lambda}_i$.

The nonparametric hypotheses described in Section 1 can all be written as

$$H_0 : CF.(y) = 0, \quad \text{for all } y,$$

for some full-rank contrast matrix C, where $F.(y) = (F_1.(y), \ldots, F_k.(y))'$. For example, the hypothesis (2) can be written in the above form with $C = (1_{k-1}| - I_{k-1})$, where 1_a is the a-dimensional vector of 1's, and I_a is the $a \times a$ identity matrix. For randomly censored data, however, this hypothesis must be modified to

$$H_0 : CF.(y) = 0, \quad \text{for all } y \le T, \tag{5}$$

where T is some value satisfying $T < \inf_{i,x} \tau_{ix}$, where $\tau_{ix} = \inf\{y; H_i(y|x) = 1\}$ is the right endpoint of the support of $H_i(y|x)$. This restriction in the range of y-values in the comparisons is necessitated by the fact that the tails of F_{ix} cannot be estimated beyond the right endpoint of the support of $H_i(y|x)$, and the fact that the asymptotic representations we use in the derivation of the asymptotic theory hold up to a point less than the right endpoint.

Remark 1. Many studies, including the real data example we analyze in Section 5, have a termination point and censoring occurs if the event under study has not been observed for a particular subject by the end of the study. Typically there exist staggered entries and/or drop-outs so the censoring is actually random. In such cases it is often appropriate to take as T_0 the length of the study. It could happen, however, that either censoring is heavy and/or the time to the event of interest is particularly small for some region of the covariate values. In such cases the choice of $T_0 < \inf_{ix} \tau_{ix}$ can be too small for covariate values outside that region, and this compromises the interpretation of the hypotheses (5). To define more meaningful hypotheses in such cases, we can consider weight functions that depend on y. To illustrate how this can be done, suppose that for a partition of the half interval $y > T = T_0$ into $m+1$ regions $T_0 < y \leq T_1, T_1 < y \leq T_2, \ldots, T_m < y$, there are regions $R(T_\kappa)$, $\kappa = 1, \ldots, m$, of the covariate values such that, if $x \in R(T_\kappa)$, the conditional distribution $F_{ix}(y)$ can be estimated until T_κ, for all i; thus, $R(T_\kappa) \subseteq R(T_{\kappa-1})$. Then, for $y \leq T_0$ define $F_{i\cdot}(y) = \int F_{ix}(y) dG(x)$, as before, while for $T_{\kappa-1} < y \leq T_\kappa$ define $F_{i\cdot}(y) = \int F_{ix}(y) dG_{R(T_\kappa)}(x)$, where $G_{R(T_\kappa)}$ is the overall distribution function of the covariate conditioned on the region $R(T_\kappa)$. Thus (3) is replaced by

$$F_{i\cdot}(y) = \int F_{ix}(y) \sum_{\kappa=0}^{m} I(T_{\kappa-1} < y \leq T_\kappa) dG_{R(T_\kappa)}(x), \tag{6}$$

where $T_{-1} = -\infty$, and $G_{R(T_0)} = G$. For example, suppose that the function $T_x = \min_i \tau_{ix}$ is as in Figure 1. Then for $y \leq T$ we use the weight function G, for $T < y \leq T_1$ we use G_{x_1}, the conditional distribution of the covariate given its value is larger than x_1, and for $T_1 < y \leq T_2$ we use G_{x_2}, the conditional distribution of the covariate given its value is larger than x_2. For simplicity, we present the derivations for testing the null hypotheses (5). We believe that this case covers the majority of the applications, but for completeness in the next section we do indicate the changes in the formulas that are necessary for the more complicated null hypotheses with weight function depending on y.

Insert Figure 1 about here

The proposed statistics for testing hypotheses of the general form (5) are based on estimating

$$T_C = C \int_{-\infty}^{T} F. dH = C \left(\int_{-\infty}^{T} F_1. dH, \ldots, \int_{-\infty}^{T} F_k. dH \right)', \tag{7}$$

where H is defined to be the limit of the empirical distribution function evaluated from all Z_{ij}. In order to estimate T_C, we need to estimate H, $F_{ix}(y)$ and $G(x)$, where the last two are needed for the estimation of $F_{i\cdot}(y) = \int F_{ix}(y) dG(x)$. Clearly, H, G are estimated by the empirical distribution functions from all Z_{ij}, all X_{ij}, respectively. Let

$$W_{ij}(x, a_{n_i}) = \frac{K\left(\frac{x - X_{ij}}{a_{n_i}}\right)}{\sum_{j'=1}^{n_i} K\left(\frac{x - X_{ij'}}{a_{n_i}}\right)},$$

where K is a known probability density function (kernel) and $\{a_{n_i}\}$ is a sequence of positive constants, tending to zero as $n_i \to \infty$, called a bandwidth sequence. To express the estimator of $F_i(y|x)$ in a general form that also applies to tied data, let $t_{i1} < t_{i2} < \ldots < t_{im_i}$ be the distinct observations in the ith group, and define $\bar{W}_{ij}(x, a_{n_i}) = \sum_{Z_{ij'} = t_{ij}} W_{ij'}(x, a_{n_i})$ and $\bar{W}_{1ij}(x, a_{n_i}) = \sum_{Z_{ij'} = t_{ij}} W_{ij'}(x, a_{n_i}) \Delta_{ij'}$, for $i = 1, \ldots, k$ and $j = 1, \ldots, m_i$. Then, the estimator of $F_i(y|x)$ is given by

$$\hat{F}_{ix}(y) \;=\; 1 - \prod_{j: d_{ij} \leq y} \left(1 - \frac{\bar{W}_{1ij}(x, a_{n_i})}{\sum_{\ell=j}^{m_i} \bar{W}_{i\ell}(x, a_{n_i})}\right).$$

This estimator was introduced by Beran (1981) and has been studied by a number of authors including McKeague & Utikal (1990), Akritas (1994), Li & Doss (1995), Van Keilegom & Veraverbeke (1997) among others. It reduces to the Kaplan-Meier (1958) estimator if all weights equal n_i^{-1}. Define $\hat{F}_{i\cdot}(y) = \int \hat{F}_i(y|x)\, d\hat{G}(x)$, and set

$$\hat{T}_C = C \int_{-\infty}^{T} \hat{F}. d\hat{H} \;, \tag{8}$$

where $\hat{F}.(y) = (\hat{F}_{1\cdot}(y), \ldots, \hat{F}_{k\cdot}(y))'$. In order to obtain the asymptotic distribution of (8) it is very useful to note that, under $H_0 : CF. = 0$, $\hat{T}_C = C \int_{-\infty}^{T} (\hat{F}. - F.) d\hat{H}$.

3 Main Results

Theorem 3.1 *Let assumptions A1 through A4 hold. If the covariate is continuous, additionally require the bandwidth sequence a_{n_i} to satisfy $(\log n_i)^2/(n_i a_{n_i})^2 = o(1)$, and $n_i a_{n_i}^4 \to 0$. Then,*

$$N^{1/2} \int_{-\infty}^{T} \left\{\hat{F}_{i\cdot}(y) - F_{i\cdot}(y)\right\} d\left\{\hat{H}(y) - H(y)\right\} \to 0,$$

in probability.

Theorem 3.2 *Suppose that the assumptions of Theorem 3.1 are valid. Then,*

$$\int_{-\infty}^{T} \left\{\hat{F}_{i\cdot}(y) - F_{i\cdot}(y)\right\} dH(y) = \frac{1}{n_i}\sum_{j=1}^{n_i} \int_{-\infty}^{T} \xi_i(Z_{ij}, \Delta_{ij}, y|X_{ij})\, dH(y) \frac{g(X_{ij})}{g_i(X_{ij})} + \int \left\{ \int_{-\infty}^{T} F_i(y|x)\, dH(y) \right\} d(\hat{G}(x) - G(x)) + o_P(n_i^{-1/2}), \quad (9)$$

where $g_i(x)$ is the probability density (mass) function of the covariate in the ith group, and $g(x)$ that in the overall population.

Corollary 3.3 *Suppose that the assumptions of Theorem 3.1 are valid. Then, under $H_0 : CF_{\cdot} = 0$, where C is an $r \times k$ contrast matrix of rank r,*

$$N^{1/2}\hat{T}_C = N^{1/2} C \int_{-\infty}^{T} \hat{F}_{\cdot}(y)\, d\hat{H}(y) \to N(0, C\Sigma C'),$$

in distribution, where $\mathbf{\Sigma} = (\sigma^2_{i_1 i_2})_{i_1 i_2}$ and

$$\begin{aligned} \sigma^2_{i_1 i_2} &= \operatorname{cov}\left\{ \int_{-\infty}^{T} F_{i_1}(y|X)\, dH(y), \int_{-\infty}^{T} F_{i_2}(y|X)\, dH(y) \right\} \\ &\quad + I(i_1 = i_2)\lambda_{i_1}^{-1} \operatorname{var}\left\{ \int_{-\infty}^{T} \xi_{i_1}(Z_{i_1 j}, \Delta_{i_1 j}, y|X_{i_1 j})\, dH(y) \frac{g(X_{i_1 j})}{g_{i_1}(X_{i_1 j})} \right\}, \end{aligned} \quad (10)$$

where

$$\xi_i(z, \delta, y|x) = S_i(y|x) \left\{ \frac{S_i(z-|x) I(z \le y, \delta = 1)}{S_i(z|x)(1 - H_i(z-|x))} - \int_{-\infty}^{y \wedge z} \frac{S_i(s-|x) dH_{i1}(s|x)}{S_i(s|x)(1 - H_i(s-|x))^2} \right\}.$$

Proposition 3.4 *Suppose that the assumptions of Theorem 3.1 are valid. For $i_1, i_2 = 1, \ldots, k$ define*

$$\begin{aligned} \hat{\sigma}^2_{i_1 i_2} &= N^{-1} \sum_{\ell=1}^{N} \int_{-\infty}^{T} \int_{-\infty}^{T} \hat{F}_{i_1}(y_1|X_\ell) \hat{F}_{i_2}(y_2|X_\ell)\, d\hat{H}(y_1)\, d\hat{H}(y_2) \\ &\quad - \left\{ N^{-1} \sum_{\ell=1}^{N} \int_{-\infty}^{T} \hat{F}_{i_1}(y|X_\ell)\, d\hat{H}(y) \right\} \left\{ N^{-1} \sum_{\ell=1}^{N} \int_{-\infty}^{T} \hat{F}_{i_2}(y|X_\ell)\, d\hat{H}(y) \right\} \\ &\quad + I(i_1 = i_2) \hat{\lambda}_{i_1}^{-1} \left[n_{i_1}^{-1} \sum_{j=1}^{n_{i_1}} \left\{ \int_{-\infty}^{T} \hat{\xi}_{i_1}(Z_{i_1 j}, \Delta_{i_1 j}, y|X_{i_1 j})\, d\hat{H}(y) \frac{n_{i_1}}{N} \sum_{\ell=1}^{N} W_{i_1 j}(X_\ell, a_{n_{i_1}}) \right\}^2 \right. \\ &\quad \left. - \left\{ n_{i_1}^{-1} \sum_{j=1}^{n_{i_1}} \int_{-\infty}^{T} \hat{\xi}_{i_1}(Z_{i_1 j}, \Delta_{i_1 j}, y|X_{i_1 j})\, d\hat{H}(y) \frac{n_{i_1}}{N} \sum_{\ell=1}^{N} W_{i_1 j}(X_\ell, a_{n_{i_1}}) \right\}^2 \right], \end{aligned}$$

where

$$\hat{\xi}_i(z,\delta,y|x) = \hat{S}_i(y|x)\left\{\frac{\hat{S}_i(z-|x)I(z\le y,\delta=1)}{\hat{S}_i(z|x)(1-\hat{H}_i(z-|x))} - \int_{-\infty}^{y\wedge z}\frac{\hat{S}_i(s-|x)d\hat{H}_{i1}(s|x)}{\hat{S}_i(s|x)(1-\hat{H}_i(s-|x))^2}\right\}.$$

Then,

$$\hat{\sigma}^2_{i_1 i_2} - \sigma^2_{i_1 i_2} \to 0 \quad a.s.$$

Corollary 3.3 and Proposition 3.4 immediately yield the following corollary.

Corollary 3.5 *Suppose that the assumptions of Theorem 3.1 are valid, and let $\hat{\Sigma}$ be the matrix whose (i_1, i_2)-th element is $\hat{\sigma}^2_{i_1 i_2}$ given in Proposition 3.4. Then,*

$$N\hat{T}'_C(C\hat{\Sigma}C')^{-1}\hat{T}_C \to \chi^2_r,$$

in distribution.

We close this section by indicating the change in the representation given in Theorem 3.2 when $F_{i\cdot}$ is defined by (6). Using the notation given in Remark 1, relation (9) becomes

$$\begin{aligned}
&\int_{-\infty}^{T_m}(\hat{F}_{i\cdot}(y) - F_{i\cdot}(y))\,dH(y) \qquad (11)\\
=\ &\frac{1}{n_i}\sum_{j=1}^{n_i}\int \xi_i(Z_{ij},\Delta_{ij},y|X_{ij})\left(\sum_{\kappa=0}^{m} I(T_{\kappa-1}<y\le T_\kappa, X_{ij}\in R(T_\kappa))\frac{n_i}{n_{\kappa i}}\frac{g_\kappa(X_{ij})}{g_{\kappa i}(X_{ij})}\right)dH(y)\\
&+\frac{1}{N}\sum_{i'=1}^{k}\sum_{j=1}^{n_{i'}}\int F_i(y|X_{i'j})\left(\sum_{\kappa=0}^{m} I(T_{\kappa-1}<y\le T_\kappa, X_{i'j}\in R(T_\kappa))\frac{N}{N_\kappa}\right)dH(y)\\
&-E\left[\int F_i(y|X_\ell)\left(\sum_{\kappa=0}^{m} I(T_{\kappa-1}<y\le T_\kappa, X_\ell\in R(T_\kappa))\frac{N}{N_\kappa}\right)dH(y)\right] + o_p(N^{-1/2}),
\end{aligned}$$

where $g_{\kappa i}$ is the probability density function of X_{ij} conditional on the region $R(T_\kappa)$, g_κ is the corresponding overall probability density function, $n_{\kappa i}$ is the number of observations from group i with $X_{ij}\in R(T_\kappa)$ and and N_κ is the corresponding number from all groups. The corresponding changes in the variance and covariance formulas (10) should be obvious.

4 Simulations

The S-Plus statistical package and the C programming language were used for the numerical work. The kernel $K(x) = \frac{15}{16}(1 - x^2)^2$ was used throughout, and the bandwidth sequence was $a_n = cn^{-0.26}$ for some constant c.

The first two simulations considered the achieved alpha levels, with and without small sample correction, in a one-way analysis of covariance design with three groups and one covariate. We simulated the situation in which the total length of study is 10 with approximately half of the participants entering at the beginning of the study, and the remaining half entering the study at random uniformly over the period from 0 to 5. The covariate was generated from the interval $[10, 20]$, and the survival time from the exponential distribution with the inverse of its mean linear in the covariate. The mean survival times at covariate values 10 and 20 were chosen to be $10/\log(\frac{10}{3}) = 8.31$ and $10/\log(10) = 4.34$ respectively; thus, the censoring proportions at covariate values 10 and 20 were 0.1 and 0.3, respectively, for observations that enter at the beginning of the study. A sample size of 50 in each group was used and 3000 runs were performed. As was done with uncensored data in Akritas, Arnold & Du (2000), a modified Box-type approximation procedure was implemented to improve the performance of the testing procedure in small samples.

For the first simulation the covariate follows the uniform distribution on the interval [10,20] in all three groups. This yields an overall censoring proportion of 0.2332. The achieved alpha levels for a selected number of constants c are reported in Table 1; the numbers in parentheses correspond to the small-sample correction. It is seen that the achieved Type I error probabilities get better as the bandwidth increases, but are quite reasonable even with $c = 15$.

Insert Table 1 about here

Remark 2. The fact that the achieved Type I error probabilities improve with the sample size has the following explanation. When the covariate has the same distribution in all factor level combinations, $F_{i\cdot}$ is the marginal distribution function of the response variable in group i, and thus the hypotheses (5) are identical to those considered in Akritas & Brunner (1997). Under the additional assumption of independent censoring, $F_{i\cdot}$ can also be estimated by the usual Kaplan-Meier estimator from the i-th category, and thus the test statistics developed in Akritas & Brunner (1997) can be applied. These statistics are a special case of the present statistics when the bandwidth is infinite. Thus the choice $c = 10000$ in Table 1 corresponds to the Akritas & Brunner (1997) procedure which is also valid in the simulated case. Estimating $F_{i\cdot}$ by the Kaplan-Meier estimator, however, ignores existing information on the covariate. To investigate the advantages of using the additional information, we conducted a small simulation study using n=150 with 1000 simulation runs. We considered only $c = 15$ and $c = 10000$ with no small-sample correction. As above, the covariate is uniform in $[10, 20]$ for all three groups. At the null hypothesis, the survival times are exponential with $\lambda(x) = log(2)/10 + \{log(15)/100\}(x - 10)$, where λ is the inverse of the mean. For nominal levels of 0.01, 0.05 and 0.1, the achieved Type I error probabilities for $c = 15$ are 0.009, 0.043 and 0.091, respectively, while for $c = 10000$ they are 0.012, 0.054 and 0.099, respectively. For the alternative hypothesis, the survival times are exponential with rates $\lambda_1(x) = log(2)/10 + \{log(15)/100\}(x - 10)$, $\lambda_2(x) = log(3)/10 + \{log(15)/100\}(x - 10)$ and $\lambda_3(x) = log(4)/10 + \{log(15)/100\}(x - 10)$, for groups 1, 2 and 3, respectively. For nominal levels of 0.01, 0.05 and 0.1, he achieved powers for $c = 15$ are 0.317, 0.536 and 0.656, respectively, while for $c = 10000$ they are 0.273, 0.504 and 0.626, respectively. Thus with $c = 15$ the procedure is more powerful even though it is somewhat conservative.

For the second simulation the covariate from uniform [10,20] in the first group but from different distributions in the second and third groups. To describe it, let $B_1 \sim Be(3, 3)$, $B_2 \sim Be(5, 3)$. Then the covariate was generated as $8 + 14 \times B_1$ restricted to [10,20] for the second group, and as $8 + 14 \times B_2$ restricted to [10,20] for the third. This yields an overall censoring proportion of 0.2227. Table 2 contains the achieved Type I error probabilities; as in Table 1, the numbers in parentheses correspond to the small-sample correction. The best choice for the constant appears to be 16, while the test procedure tends to become more liberal as the constant increases. Evidently, this is due to the fact that the Kaplan-Meier estimator does not estimate $F_{i\cdot}$ consistently. We remark that if the censoring variable depends on the covariate, a similar bias of the testing procedure of Akritas & Brunner (1997) is expected even when the covariate has the same distribution in all groups.

Insert Table 2 about here

For the next set of simulations, we considered the settings of Tables 1 and 2 and increased the sample size to 150 in each group to investigate (i) whether the small-sample correction is still needed and (ii) whether the choice of the constant c remains the same. The results from 1000 simulation runs are reported in Table 3. It is seen that the asymptotic theory of Section 3 provides good approximation to the null distribution of the test statistic, so there is no need for small-sample correction. Moreover, while the best choices of c from Tables 1 and 2 continue to give good results, it is seen that the achieved alpha levels are less sensitive to the choice of c. Finally, it can be seen that the bias of the procedure with $c = 10000$ is more noticeable for large samples.

Insert Table 3 about here

Next, we compared the power of the nonparametric method with that of a proportional hazards model. Two groups with sample size 30 were considered, and thus the small-sample correction was used for the nonparametric method. For both groups the covariate follows the uniform distribution on $[0, 1]$. Given the covariate value $X = x$, the survival and censoring times for the fist group were generated independently from exponential distributions with means $a_1 + a_2 x + a_3 x^2$ and $c_1 + c_2 x + c_3 x^2$ respectively, with $a_1 = 5, a_2 = -30, c_1 = 5, c_2 = -10$

and $c_3 = 8$. The coefficient a_3 is specified from $p(x)$, the desired censoring proportion at $X = x$, through the relationship

$$p(x) = P(\Delta = 0|X = x) = \frac{a_1 + a_2 x + a_3 x^2}{(a_1 + c_1) + (a_2 + c_2)x + (a_3 + c_3)x^2}.$$

We used a constant censoring proportion $p(x) = p$, for all x, and considered various values of p. It can be verified that $a_1 + a_2 x + a_3(x)x^2 > 0$ and $c_1 + c_2 x + c_3 x^2 > 0$, for all $p, x \in (0, 1)$. For the second group, the censoring variable was generated in exactly the same way, but the survival time was sampled from a Gamma distribution with the same scale parameter as in the fist group, i.e. $(a_1 + a_2 x + a_3 x^2)^{-1}$, and shape parameter τ. When $\tau = 1$, it reduces to the baseline situation, i.e. no group effect. As τ increases, the difference between the two groups becomes larger; notice, however, that the chosen alternatives do not conform to the proportional hazards model. The proportional hazards model incorporated linear effects of group indicator and the covariate on the hazard as commonly done in practice. Under the alternative hypothesis, the proportional hazards assumption was violated. The results from 3000 simulation runs are summarized in Table 4. At baseline, the nonparametric method tended to be slightly liberal. The proportional hazards method achieved quite accurate alpha levels for $p = 0.25$ and $p = 0.5$, but became quite conservative for $p = 0.75$. It is seen that the nonparametric method is more powerful than the proportional hazards method and the power advantage increases with τ. In the presence of heavy censoring, the power of proportional hazards method dropped considerably.

Insert Table 4 about here

5 Data Analysis

We now apply the proposed method to the bone marrow transplantation study which is described, and used in various examples, in Klein & Moeschberger (1997). Details of the study can be found in Copelan et al. (1991). 137 patients who had bone marrow transplants from March 1, 1984 to June 30, 1989 were enrolled. The maximum follow-up was 7 years. By the end of the study, 83 patients relapsed or died with treatment-related disease, with the remaining patients still alive without an episode of leukemia relapse. We will consider the disease-free survival time as the response variable and will analyze the effect that two factors have on it adjusting for one covariate, the donor age. The two factors are (a) a binary variable indicating whether or not the patient received methotrexate, in addition to a regular graft-versus-host prophylactic, and (b) the initial disease category with three groups: acute lymphoblastic leukemia, low-risk acute myeloctic leukemia, and high-risk acute myeloctic leukemia. The censoring proportion was 0.394.

To apply the nonparametric procedure, we first performed a simulation study to determine a reasonable choice for the constant in the bandwidth of the form $a_n = \text{Const.} \times n^{-0.26}$. In each simulation run, six groups with the same sizes as in the data set were generated. The covariate was sampled with replacement group-wise to retain the different shapes of the covariate in different groups. To simulate the null hypothesis, the range of the covariate values in the data set, 2-56 years old, was equally divided up into five subsets. Then the disease-free survival time and the censoring indicator were randomly selected with replacement within each of the five subsets and allocated to the six groups. Note that this resampling implies that the responses are generated according to the same conditional distributions in each group. Moreover the degree and nature of correlation between the response and the covariate in the data set is approximately maintained in the simulations. We performed 5000 simulation runs testing, with small-sample corrections, for the covariate-adjusted main effects of the two factors as well as their interaction. The Type I error probabilities resulting from various constants are reported in Du (1999). The constant values 50 and 60 seem to give the best results, and both were used in data analysis. The resulting p-values are reported in Table 5.

For comparison, proportional hazards models were considered. We first considered the issue of validity of the proportional hazards assumptions by testing for interaction between each of the explanatory variables and the artificial time-dependent variable $\log t$. Significance of an interaction indicates violation of the proportional hazards assumption for the corresponding variable. The p-values from the tests were 0.42, 0.0376 and 0.0588 for disease, methotrexate and donor age, respectively. Thus there is some evidence that the proportional hazards assumption is violated for methotrexate and donor age. For Figure 2 (a) we combined all the data that correspond to each of the two methotrexate levels, computed the corresponding Nelson-Aalen estimators $\hat{\Lambda}(t)$, and plotted the difference of their logarithms versus time. For Figure 2 (b) the donor age we discretized into four groups, with approximately the same number of observations in each group, calculated the Nelson-Aalen estimators for each group and plotted the differences of their logarithms. These figures also reveal the non-proportionality of hazards in the two methotrexate levels, and also that the non-proportionality in the donor age is mainly due to the oldest group.

Insert Figure 2 about here

Next we stratified on methotrexate. To investigate the assumptions of the stratified proportional hazards model, we computed the likelihood ratio statistic by fitting the stratified model, and also separate proportional hazards models for each of the two strata. The Helmert scheme for coding the factor levels was used; see Akritas, Arnold & Du (2000) for a discussion of the effects of the coding scheme. The likelihood ratio statistic was 8.109; with five degrees of freedom this corresponds to a p-value of 0.15. Therefore, we concluded that the stratified model would be appropriate. It gave a p-value of 0.0016 for the effect of initial disease group.

However, we were not able to test for the effect of methotrexate and its interaction with disease group adjusted for donor age when stratifying on methotrexate. For purely comparison purposes we applied the non-stratified model, even though the model assumptions appear to be violated. Again we used the Helmert scheme for coding the factor levels. The

p-values are reported in Table 5, together with the two sets of p-values from the nonparametric method. It is seen that the naive, in the sense that the assumptions are violated, application of the proportional hazards model suggests that none of the effects is significant. The nonparametric method concluded that the difference among disease categories is significant at level 0.05, and also that the effect of methotrexate groups was on the borderline of being significant.

Insert Table 5 about here

As an illustration of the graphical capabilities offered by the nonparametric approach, Figure 3 presents the nonparametric main and interaction effect for this data set. The plots of main effects are equivalent to the plots of $\hat{F}_{i\cdot}$ and $\hat{F}_{\cdot j}$ presented in Figure 4(a), (b). It is easier to see from these figures that low-risk acute myeloctic leukemia patients have the best disease-free survival rate, followed by the acute lymphoblastic leukemia patients, and the high-risk acute myeloctic leukemia patients; also, patients receiving additionally methotrexate appear to have higher risk of relapse or disease-related death, though the evidence of that in terms of p-values is not very strong. These figures re-enforce the findings of the nonparametric analysis.

Insert Figures 3, 4 about here

Appendix

The main results will be derived under the following assumptions:

A1 (i) The kernel function K is a symmetric probability density function with compact support.

(ii) K has a bounded second derivative.

A2 (i) If the covariate is discrete, then

a) The set of atoms finite and is the same for all k cells.

b) Let the atoms be denoted as $d_1 < d_2 < \ldots$, then $\min_\ell g_i(d_\ell) > 0$.

(ii) If the covariate is continuous, then

a) The covariate has the same bounded compact support $R_X = [x_\ell, x_u]$ for all k groups.

b) The distribution function $G_i(x)$ of the covariate X_i has bounded third derivative.

c) The density function $g_i(x)$ of the covariate X_i satisfies $\inf_x g_i(x) > 0$.

A3 (i) The continuous parts of the (sub-)distribution functions $H_i(y|x)$ and $H_{i1}(y|x)$, $H_{ic}(y|x)$ and $H_{i1c}(y|x)$ respectively, are uniformly Lipschitz of order 1 in the following sense:

$$\begin{aligned} |H_{ic}(y_1|x) - H_{ic}(y_2|x)| &\leq const.|y_1 - y_2|, \text{ for all } x, y_1, y_2. \\ |H_{i1c}(y_1|x) - H_{i1c}(y_2|x)| &\leq const.|y_1 - y_2|, \text{ for all } x, y_1, y_2. \end{aligned}$$

If X_i is continuous we further assume (ii), (iii) and (iv) below:

(ii) The first derivatives of $H_i(y|x)$ and $H_{i1}(y|x)$ with respect to x, $\dot{H}_i(y|x)$, $\dot{H}_{i1}(y|x)$, exist and are bounded uniformly for all $y \leq T$, x. In addition the continuous parts of these first derivatives, $\dot{H}_{ic}(y|x)$ and $\dot{H}_{i1c}(y|x)$, are uniformly Lipschitz of order 1 in the sense described above.

(iii) The second derivatives of $H_i(y|x)$ and $H_{i1}(y|x)$ with respect to x, $\ddot{H}_i(y|x)$, $\ddot{H}_{i1}(y|x)$, exist and are bounded uniformly for all $y \leq T$, x.

(iv) The first two derivatives of $H_{i1}(y|x)$ with respect to x are of bounded variation, i.e.

$$\sup_{\triangle} \| \dot{H}_{i1}(y|x) \|_{\triangle} < \infty, \text{ and } \sup_{\triangle} \| \ddot{H}_{i1}(y|x) \|_{\triangle} < \infty,$$

where $\triangle$ ranges over all partitions of $[x_l, x_u]$, and for any partition $\triangle : x_l = x_0 < x_1 <$

$\ldots < x_m = x_u$, the norms are defined as

$$\begin{aligned} \parallel \dot{H}_{i1}(y|x) \parallel_{\triangle} &= \sum_{j=1}^{m} \Big| \dot{H}_{i1}(y|x_j) - \dot{H}_{i1}(y|x_{j-1}) \Big|, \text{ and} \\ \parallel \ddot{H}_{i1}(y|x) \parallel_{\triangle} &= \sum_{j=1}^{m} \Big| \ddot{H}_{i1}(y|x_j) - \ddot{H}_{i1}(y|x_{j-1}) \Big|. \end{aligned}$$

A4 (i) For each $i = 1, \ldots, k$, the atoms of $C_i(y|x)$, the conditional distribution of the censoring time for group i, are the same for all x.

(ii) For each $i = 1, \ldots, k$, the support and the atoms of $F_i(y|x)$ are the same for all x.

(iii) The first two derivative of $F_i(y|x)$ with respect to x, $\dot{F}_i(y|x)$ and $\ddot{F}_i(y|x)$ exist and are uniformly bounded for all $y \leq T$, x.

(iv) For any atom $d_i \leq T$, define $s_i = \sup_x |F_i(d_i|x) - F_i(d_i - |x)|$, and $s_i' = \sup_x \Big| \dot{F}_i(d_i|x)$ $- \dot{F}_i(d_i - |x) \Big|$, then $\sum_{d_i \leq T} s_i < \infty$ and $\sum_{d_i \leq T} s_i' < \infty$.

We start with an auxiliary result needed in the proof of Theorem 3.1.

Lemma A.1 *Let X be a continuous variable on $R_X = [x_\ell, x_u]$ with distribution function $G(x)$ and density function $g(x)$ satisfying A2(ii), and $K(x)$ be a kernel function satisfying A1. For a sample $X_1, \ldots, X_n$, and a bandwidth sequence a_n with $\lim_{n\to\infty} a_n = 0$, $na_n^5/\log n = O(1)$ and $\log n/(na_n) = O(1)$, we have*

$$\sup_x |\hat{g}(x) - g(x)| = O\Big\{(na_n)^{-1/2}(\log n)^{1/2}\Big\} \text{ almost surely,}$$

where $\hat{g}(x) = \sum_{i=1}^{n} K(xa_n^{-1} - X_i a_n^{-1})/(na_n)$ is an estimator of the density function $g(x)$.

Proof. First write

$$\begin{aligned} & |\hat{g}(x) - g(x)| \\ & \leq \Big| g(x) - \int \frac{1}{a_n} K\Big(\frac{x-y}{a_n}\Big) dG(y) \Big| + \Big| \int \frac{1}{a_n} K\Big(\frac{x-y}{a_n}\Big) d\Big\{\hat{G}(y) - G(y)\Big\} \Big|. \qquad (1) \end{aligned}$$

Using a change of variable and a Taylor expansion,

$$\int \frac{1}{a_n} K\Big(\frac{x-y}{a_n}\Big) dG(y) = g(x) + a_n^2 \int K(u) u^2 \frac{\partial^2 g(x(u))}{\partial x^2} du, \qquad (2)$$

where $x(u)$ is between x and $x+ua_n$. Since $|\partial^2 g(x)/\partial x^2|$ is bounded, the first term of (1) is $O(a_n^2)$ uniformly over x, which implies uniform $O\left\{(na_n)^{-1/2}(\log n)^{1/2}\right\}$ by the condition $na_n^5/\log n = O(1)$. To deal with the second term of (1), partition $[x_l, x_u)$ into subintervals $[x_i, x_{i+1})$ of length $O(a_n^{3/2})$, $i = 1, \ldots, k_n = O(a_n^{-3/2})$. For any x, let x_i be such that $x \in [x_i, x_{i+1}]$, thus $x - x_i \le C a_n^{3/2}$, for some $C > 0$. The second term of (1) is bounded by, uniformly in x,

$$\max_{1\le i\le k_n}\left|\frac{1}{a_n}\int K\Big(\frac{x_i-y}{a_n}\Big)d\Big\{\hat{G}(y)-G(y)\Big\}\right| \tag{3}$$
$$+\max_{1\le i\le k_n}\sup_{x_i\le x\le x_{i+1}}\left|\int\frac{1}{a_n}\left\{K\Big(\frac{x-y}{a_n}\Big)-K\Big(\frac{x_i-y}{a_n}\Big)\right\}d\Big\{\hat{G}(y)-G(y)\Big\}\right|.$$

Using a change of variable and a Taylor expansion, we have

$$\left|\int\frac{1}{a_n}\left\{K\Big(\frac{x-y}{a_n}\Big)-K\Big(\frac{x_i-y}{a_n}\Big)\right\}d\Big\{\hat{G}(y)-G(y)\Big\}\right|$$
$$=\left|\int K\big(x(u)\big)\frac{x_i-x}{a_n^2}d\Big\{\hat{G}(x_i+ua_n)-G(x_i+ua_n)\Big\}\right|,$$

where $x(u)$ is between u and $u+\dfrac{x_i-x}{a_n}$. Applying integration by parts and the fact that $\sup_x|\hat{G}(x)-G(x)| = O\Big\{n^{-1/2}(\log n)^{1/2}\Big\}$ almost surely, it is easily seen that the second term of (3) is $O\Big\{(na_n)^{-1/2}(\log n)^{1/2}\Big\}$ almost surely uniformly in x. Since $\sum_{j=1}^n \mathrm{var}\Big\{K\Big(x_i a_n^{-1} - X_j a_n^{-1}\Big)\Big\} = O(na_n)$, it follows from Bernstein's inequality (Uspensky, 1937) that

$$\mathrm{pr}\left[\max_{1\le i\le k_n}\left|\frac{1}{na_n}\sum_{j=1}^n\left\{K\left(\frac{x_i-X_j}{a_n}\right)-\int K\left(\frac{x_i-y}{a_n}\right)dG(y)\right\}\right| > c(na_n)^{-1/2}(\log n)^{1/2}\right]$$
$$\le k_n\,\mathrm{pr}\left[\left|\sum_{j=1}^n\left\{K\left(\frac{x_i-X_j}{a_n}\right)-\int K\left(\frac{x_i-y}{a_n}\right)dG(y)\right\}\right| > c(na_n\log n)^{1/2}\right]$$
$$\le 2k_n\exp\left[-\frac{\Big\{c(na_n\log n)^{1/2}\Big\}^2}{2\sum_{j=1}^n\mathrm{var}\Big\{K\Big(\frac{x_i-X_j}{a_n}\Big)\Big\}+\frac{2}{3}c\sup_u K(u)(na_n\log n)^{1/2}}\right]$$
$$= 2k_n\exp\left\{-\frac{c^2na_n\log n}{c'na_n+c''(na_n\log n)^{1/2}}\right\}$$
$$\le K_1k_nn^{-K_2},$$

which is summable by choosing an appropriate constant $c > 0$. This shows that the second term of (1) is also $O\left\{(na_n)^{-1/2}(\log n)^{1/2}\right\}$ almost surely uniformly over all x.

Proof of Theorem 3.1. Let us write the sum $\sum_{i=1}^{k}\sum_{j=1}^{n_i}$ in the definition of $\hat{F}_{i\cdot}(y)$ as one big sum $\sum_{\ell=1}^{N}$. Then,

$$\begin{aligned}\hat{F}_{i\cdot}(y) - F_{i\cdot}(y) &= N^{-1}\sum_{\ell=1}^{N}\hat{F}_i(y|X_\ell) - \int F_i(y|x)dG(x)\\ &= N^{-1}\sum_{\ell=1}^{N}\Big\{\hat{F}_i(y|X_\ell) - F_i(y|X_\ell)\Big\} + \int F_i(y|x)\,d\Big\{\hat{G}(x) - G(x)\Big\}\\ &= T_{1i}(y) + T_{2i}(y).\end{aligned}$$

Hence,

$$\begin{aligned}&\int_{-\infty}^{T}\Big\{\hat{F}_{i\cdot}(y) - F_{i\cdot}(y)\Big\}\,d\Big\{\hat{H}(y) - H(y)\Big\}\\ &= \int_{-\infty}^{T} T_{1i}(y)\,d\Big\{\hat{H}(y) - H(y)\Big\} + \int_{-\infty}^{T} T_{2i}(y)\,d\Big\{\hat{H}(y) - H(y)\Big\}\\ &= A_{1i} + A_{2i}.\end{aligned}$$

A_{2i} equals

$$\begin{aligned}&\int_{-\infty}^{T}\int F_i(y|x)\,d\Big\{\hat{G}(x) - G(x)\Big\}\,d\Big\{\hat{H}(y) - H(y)\Big\}\\ =\;& \sum_{i_1=1}^{k}\sum_{i_2=1}^{k}\hat{\lambda}_{i_1}\hat{\lambda}_{i_2}\int_{-\infty}^{T}\int F_i(y|x)\,d\Big\{\hat{G}_{i_2}(x) - G_{i_2}(x)\Big\}\,d\Big\{\hat{H}_{i_1}(y) - H_{i_1}(y)\Big\} + o_P(n_i^{-1/2}),\end{aligned}$$

using the fact that $\hat{\lambda}_i \to \lambda_i$. Each summand is a V-statistic of degree 2 with degenerate kernel. Hence, $A_{2i} = O_P(N^{-1}) = o_P(N^{-1/2})$. For A_{1i}, use is made of the representation

$$\hat{F}_i(y|X_\ell) - F_i(y|X_\ell) = \sum_{j=1}^{n_i} W_{ij}(X_\ell, a_{n_i})\xi_i(Z_{ij}, \Delta_{ij}, y|X_\ell) \; + \; r_{n_i,X_\ell}(y),$$

where $\sup_x\sup_{y\le T}|r_{n_i,x}(y)| = O\Big\{(na_{n_i})^{-3/4}(\log n_i)^{3/4}\Big\}$ almost surely, holds for any $T < \inf_x \tau_{ix}$; see Theorem 3.2 in Du & Akritas (2000). We have

$$\begin{aligned}A_{1i} &= N^{-1}\sum_{\ell=1}^{N}\sum_{j=1}^{n_i} W_{ij}(X_\ell, a_{n_i})\int_{-\infty}^{T}\xi_i(Z_{ij}, \Delta_{ij}, y|X_\ell)\,d\Big\{\hat{H}(y) - H(y)\Big\} + o_P(n_i^{-1/2})\\ &= (Nn_i a_{n_i})^{-1}\sum_{\ell=1}^{N}\sum_{j=1}^{n_i}\frac{K(\frac{X_{ij}-X_\ell}{a_{n_i}})}{\hat{g}_i(X_\ell)}\int_{-\infty}^{T}\xi_i(Z_{ij}, \Delta_{ij}, y|X_\ell)\,d\Big\{\hat{H}(y) - H(y)\Big\} + o_P(n_i^{-1/2}),\end{aligned}$$

where $\hat{g}_i(x) = (n_i a_{n_i})^{-1} \sum_{j=1}^{n_i} K\left(x a_{n_i}^{-1} - X_{ij} a_{n_i}^{-1}\right)$. First assume that the covariate is continuous, then $\hat{g}_i(x)$ is an estimator of the probability density function $g_i(x)$ of the covariate in the i-th cell, and by Lemma A.1,

$$\sup_x |\hat{g}_i(x) - g_i(x)| = O\Big\{(n_i a_{n_i})^{-1/2} (\log n_i)^{1/2}\Big\}, \text{ almost surely.} \tag{4}$$

Further, it can be verified that $\xi_i(Z_{ij}, \Delta_{ij}, y|x)$ can be also written as

$$\begin{aligned} \xi_i(Z_{ij}, \Delta_{ij}, y|x) &= \int_{-\infty}^{y} \frac{S_i(s-|x)}{S_i(s|x)} \cdot \frac{d\{I(Z_{ij} \le s, \delta_{ij} = 1) - H_{i1}(s|x)\}}{1 - H_i(s-|x)} \\ &\quad - \int_{-\infty}^{y} \frac{S_i(s-|x)}{S_i(s|x)} \cdot \frac{I(Z_{ij} < s) - H_i(s-|x)}{\Big\{1 - H_i(s-|x)\Big\}^2} dH_{i1}(s|x), \end{aligned}$$

which gives that

$$\begin{aligned} & \sum_{j=1}^{n_i} W_{ij}(x, a_{n_i}) \xi_i(Z_{ij}, \Delta_{ij}, y|x) \\ =\; & \int_{-\infty}^{y} \frac{S_i(s-|x)}{S_i(s|x)} \cdot \frac{d\Big\{\hat{H}_{i1}(s|x) - H_{i1}(s|x)\Big\}}{1 - H_i(s-|x)} \\ & - \int_{-\infty}^{y} \frac{S_i(s-|x)}{S_i(s|x)} \cdot \frac{\hat{H}_i(s-|x) - H_i(s-|x)}{\Big\{1 - H_i(s-|x)\Big\}^2} dH_{i1}(s|x) \\ =\; & O\Big((n_i a_{n_i})^{-1/2} (\log n_i)^{1/2}\Big) \text{ almost surely,} \end{aligned} \tag{5}$$

uniformly in x by Lemma 4.2 in Du & Akritas (2000). Using (4), (5) and the assumption $\inf_x g_i(x) > 0$ yields that

$$\begin{aligned} & \frac{1}{N n_i a_{n_i}} \sum_{\ell=1}^{N} \sum_{j=1}^{n_i} \left\{ \frac{K(\frac{X_{ij} - X_\ell}{a_{n_i}})}{\hat{g}_i(X_\ell)} - \frac{K(\frac{X_{ij} - X_\ell}{a_{n_i}})}{g_i(X_\ell)} \right\} \int_{-\infty}^{T} \xi_i(Z_{ij}, \Delta_{ij}, y|X_\ell)\, d\Big\{\hat{H}(y) - H(y)\Big\} \\ =\; & N^{-1} \sum_{\ell=1}^{N} \sum_{j=1}^{n_i} W_{ij}(X_\ell, a_{n_i}) \frac{g_i(X_\ell) - \hat{g}_i(X_\ell)}{g_i(X_\ell)} \int_{-\infty}^{T} \xi_i(Z_{ij}, \Delta_{ij}, y|X_\ell)\, d\Big\{\hat{H}(y) - H(y)\Big\} \\ =\; & O_P\Big\{(n_i a_{n_i})^{-1} (\log n_i)\Big\}, \end{aligned}$$

which is $o_P(n_i^{-1/2})$ by the condition $(\log n_i)^2 / \Big(n_i a_{n_i}^2\Big) = o(1)$. Therefore, A_{1i} can be written as

$$\frac{1}{N n_i a_{n_i}} \sum_{\ell=1}^{N} \sum_{j=1}^{n_i} \frac{K\left(\frac{X_\ell - X_{ij}}{a_{n_i}}\right)}{g_i(X_\ell)} \int_{-\infty}^{T} \xi_i(Z_{ij}, \Delta_{ij}, y|X_\ell)\, d\Big\{\hat{H}(y) - H(y)\Big\} + o_P(n_i^{-1/2})$$

$$
\begin{aligned}
&= (Nn_i a_{n_i})^{-1} \sum_{\ell=1}^{N} \sum_{j=1}^{n_i} g_i^{-1}(X_\ell) K\left(\frac{X_\ell - X_{ij}}{a_{n_i}}\right) \\
&\quad \times \int_{-\infty}^{T} \{\xi_i(Z_{ij}, \Delta_{ij}, y|X_\ell) - \xi_i(Z_{ij}, \Delta_{ij}, y|X_{ij})\}\, d\Big\{\hat{H}(y) - H(y)\Big\} \\
&\quad + (Nn_i a_{n_i})^{-1} \sum_{\ell=1}^{N} \sum_{j=1}^{n_i} g_i^{-1}(X_\ell) K\left(\frac{X_\ell - X_{ij}}{a_{n_i}}\right) \\
&\quad \times \int_{-\infty}^{T} \xi_i(Z_{ij}, \Delta_{ij}, y|X_{ij})\, d\Big\{\hat{H}(y) - H(y)\Big\} + o_P(n_i^{-1/2}) \\
&= A_{11i} + A_{12i} + o_P(n_i^{-1/2}).
\end{aligned}
$$

Let $E(\cdot|X)$ denote the conditional expectation on $X_1, \ldots, X_N$. Also write $d_{i\ell j}(y) = \xi_i(Z_{ij}, \Delta_{ij}, y|X_\ell) - \xi_i(Z_{ij}, \Delta_{ij}, y|X_{ij})$ and $c_\ell(y) = I(Z_\ell \le y)$. Note that assumptions A3(ii), A3(iv) and A4(iii) imply that $\partial\xi_i(z, \delta, y|x)/\partial x$ is bounded uniformly over $x, \delta, y \le T$ and $z \le T$. Therefore, $d_{i\ell j}(y) = \partial\xi_i(Z_{ij}, \Delta_{ij}, y|x = \xi_{i\ell j})/\partial x \times (X_\ell - X_{ij}) = O(a_{n_i})$, uniformly over $y \le T$ and over i, j and ℓ for $K\left(X_\ell a_{n_i}^{-1} - X_{ij} a_{n_i}^{-1}\right) \neq 0$, where $\xi_{i\ell j}$ is some value between X_ℓ and X_{ij}. We obtain

$$
\begin{aligned}
&NE(A_{11i}^2|X) \\
&= N(Nn_i a_{n_i})^{-2} \sum_{\ell_1,\ell_2=1}^{N} \sum_{j_1,j_2=1}^{n_i} g_i^{-1}(X_{\ell_1}) g_i^{-1}(X_{\ell_2}) K\left(\frac{X_{\ell_1} - X_{ij_1}}{a_{n_i}}\right) K\left(\frac{X_{\ell_2} - X_{ij_2}}{a_{n_i}}\right) \\
&\quad \times E\left[\int_{-\infty}^{T}\int_{-\infty}^{T} d_{i\ell_1 j_1}(y_1) d_{i\ell_2 j_2}(y_2)\, d\Big\{\hat{H}(y_1) - H(y_1)\Big\} d\Big\{\hat{H}(y_2) - H(y_2)\Big\}\,\Big|\, X\right] \\
&= N(Nn_i a_{n_i})^{-2} \sum_{\ell_1,\ell_2=1}^{N} \sum_{j_1,j_2=1}^{n_i} g_i^{-1}(X_{\ell_1}) g_i^{-1}(X_{\ell_2}) K\left(\frac{X_{\ell_1} - X_{ij_1}}{a_{n_i}}\right) K\left(\frac{X_{\ell_2} - X_{ij_2}}{a_{n_i}}\right) \\
&\quad \times N^{-2} E\left[\sum_{\ell_3=1}^{N} \int_{-\infty}^{T}\int_{-\infty}^{T} d_{i\ell_1 j_1}(y_1) d_{i\ell_2 j_2}(y_2)\, d\Big\{c_{\ell_3}(y_1) - H(y_1)\Big\} d\Big\{c_{\ell_3}(y_2) - H(y_2)\Big\}\,\Big|\, X\right] \\
&\le C(Nn_i a_{n_i})^{-2} \sum_{\ell_1,\ell_2=1}^{N} \sum_{j_1,j_2=1}^{n_i} g_i^{-1}(X_{\ell_1}) g_i^{-1}(X_{\ell_2}) K\left(\frac{X_{\ell_1} - X_{ij_1}}{a_{n_i}}\right) K\left(\frac{X_{\ell_2} - X_{ij_2}}{a_{n_i}}\right) a_{n_i}^2,
\end{aligned}
$$

for some constant C, and

$$
\begin{aligned}
&NE(A_{11i}^2) = E(NE(A_{11i}^2|X)) \\
&\le C(Nn_i)^{-2} \sum_{\ell_1,\ell_2=1}^{N} \sum_{j_1,j_2=1}^{n_i} E\left\{g_i^{-1}(X_{\ell_1}) g_i^{-1}(X_{\ell_2}) K\left(\frac{X_{\ell_1} - X_{ij_1}}{a_{n_i}}\right) K\left(\frac{X_{\ell_2} - X_{ij_2}}{a_{n_i}}\right)\right\} \\
&= O(a_{n_i}^2) = o(1),
\end{aligned}
$$

using the assumption $\inf_x g_i(x) > 0$ and the result $E\left\{K(xa_{n_i}^{-1} - X_i a_{n_i}^{-1})\right\} = O(a_{n_i})$ uniformly in x by relation 2. By Chebyshev's inequality, $N^{1/2}A_{11i} \rightarrow 0$ in probability. Similarly for the term A_{12i}, consider

$$
\begin{aligned}
& NE(A_{12i}^2) \\
= \; & N(Nn_i a_{n_i})^{-2} E\Bigg[\sum_{\ell_1,\ell_2=1}^{N} \sum_{j_1,j_2=1}^{n_i} g_i^{-1}(X_{\ell_1}) g_i^{-1}(X_{\ell_2}) K\left(\frac{X_{\ell_1} - X_{ij_1}}{a_{n_i}}\right) K\left(\frac{X_{\ell_2} - X_{ij_2}}{a_{n_i}}\right) \\
& \times \int_{-\infty}^{T}\int_{-\infty}^{T} \xi_i(Z_{ij_1}, \Delta_{ij_1}, y_1|X_{ij_1}) \xi_i(Z_{ij_2}, \Delta_{ij_2}, y_2|X_{ij_2})\, d\Big\{\hat{H}(y_1) - H(y_1)\Big\} d\Big\{\hat{H}(y_2) - H(y_2)\Big\}\Bigg] \\
= \; & N^{-1}(Nn_i a_{n_i})^{-2} E\Bigg[\sum_{\ell_1,\ell_2=1}^{N} \sum_{j_1,j_2=1}^{n_i} \sum_{\ell=1}^{N} g_i^{-1}(X_{\ell_1}) g_i^{-1}(X_{\ell_2}) K\left(\frac{X_{\ell_1} - X_{ij_1}}{a_{n_i}}\right) K\left(\frac{X_{\ell_2} - X_{ij_2}}{a_{n_i}}\right) \\
& \times \int_{-\infty}^{T}\int_{-\infty}^{T} \xi_i(Z_{ij_1}, \Delta_{ij_1}, y_1|X_{ij_1}) \xi_i(Z_{ij_2}, \Delta_{ij_2}, y_2|X_{ij_2})\, d\Big\{c_\ell(y_1) - H(y_1)\Big\} d\Big\{c_\ell(y_2) - H(y_2)\Big\}\Bigg] \\
= \; & N^{-1}(Nn_i a_{n_i})^{-2} E\Bigg[\sum_{\ell_1,\ell_2=1}^{N} \sum_{j=1}^{n_i} \sum_{\ell=1}^{N} g_i^{-1}(X_{\ell_1}) g_i^{-1}(X_{\ell_2}) K\left(\frac{X_{\ell_1} - X_{ij}}{a_{n_i}}\right) K\left(\frac{X_{\ell_2} - X_{ij}}{a_{n_i}}\right) \\
& \times \int_{-\infty}^{T}\int_{-\infty}^{T} \xi_i(Z_{ij}, \Delta_{ij}, y_1|X_{ij}) \xi_i(Z_{ij}, \Delta_{ij}, y_2|X_{ij})\, d\Big\{c_\ell(y_1) - H(y_1)\Big\} d\Big\{c_\ell(y_2) - H(y_2)\Big\}\Bigg] \\
= \; & O\Big(\frac{1}{n_i}\Big) = o(1).
\end{aligned}
$$

Thus, the result $N^{1/2}A_{12i} \rightarrow 0$ in probability follows again from Chebychev's inequality. For a discrete covariate, $K\left(X_\ell a_{n_i}^{-1} - X_{ij} a_{n_i}^{-1}\right) = 0$ for any $X_{ij} \neq X_\ell$ as N increases, which reduced A_{1i} to

$$
\frac{1}{n_i}\sum_{j=1}^{n_i} \frac{\frac{1}{N}\sum_{\ell=1}^{N} I(X_\ell = X_{ij})}{\frac{1}{n_i}\sum_{j'=1}^{n_i} I(X_{ij'} = X_{ij})} \int_{-\infty}^{T} \xi_i(Z_{ij}, \Delta_{ij}, y|X_{ij})\, d\Big\{\hat{H}(y) - H(y)\Big\} + o_P(n_i^{-1/2}).
$$

That this term is $o_P(N^{-1/2})$ can be shown in a very similar way as for A_{12i}.

Proof of Theorem 3.2. Decompose (9) as

$$
\frac{1}{N}\sum_{\ell=1}^{N}\int_{-\infty}^{T}\left\{\hat{F}_i(y|X_\ell) - F_i(y|X_\ell)\right\} dH(y) + \frac{1}{N}\sum_{\ell=1}^{N}\int_{-\infty}^{T}\left\{F_i(y|X_\ell) - F_{i\cdot}(y)\right\} dH(y). \qquad (6)
$$

The second term of (6) equals

$$\int \left\{ \int_{-\infty}^{T} F_i(y|x)\, dH(y) \right\} d\Big\{\hat{G}(x) - G(x)\Big\},$$

which is the second term on the right hand side of (9). Applying Theorem 3.2 in Du & Akritas (2000), the first term in (6) is written as

$$\frac{1}{N}\sum_{\ell=1}^{N}\sum_{j=1}^{n_i} W_{ij}(X_\ell, a_{n_i}) \int_{-\infty}^{T} \xi_i(Z_{ij}, \Delta_{ij}, y|X_\ell)\, dH(y) + o(n_i^{-1/2}), \text{ almost surely.} \tag{7}$$

Assume first that the covariate is continuous. Using (4), (5) and the assumption $\inf_x g_i(x) > 0$ as we did for the term A_{1i}, (7) becomes

$$\begin{aligned}
&(Nn_ia_{n_i})^{-1}\sum_{\ell=1}^{N}\sum_{j=1}^{n_i} g_i^{-1}(X_\ell) K\left(\frac{X_\ell - X_{ij}}{a_{n_i}}\right) \int_{-\infty}^{T} \xi_i(Z_{ij}, \Delta_{ij}, y|X_\ell)\, dH(y) + o(n_i^{-1/2}) \\
&= (Nn_ia_{n_i})^{-1}\sum_{\ell=1}^{N}\sum_{j=1}^{n_i} K\left(\frac{X_\ell - X_{ij}}{a_{n_i}}\right) \int_{-\infty}^{T} \Big\{\eta_i(Z_{ij}, \Delta_{ij}, y|X_\ell) - \eta_i(Z_{ij}, \Delta_{ij}, y|X_{ij})\Big\}\, dH(y) \\
&\quad +(Nn_ia_{n_i})^{-1}\sum_{\ell=1}^{N}\sum_{j=1}^{n_i} K\left(\frac{X_\ell - X_{ij}}{a_{n_i}}\right) \int_{-\infty}^{T} \eta_i(Z_{ij}, \Delta_{ij}, y|X_{ij})\, dH(y) + o(n_i^{-1/2}),
\end{aligned} \tag{8}$$

where $\eta_i(z, \delta, y|x) = \xi_i(z, \delta, y|x) g_i^{-1}(x)$. Using a two-term Taylor expansion around X_{ij}, and letting $\xi_{ij\ell}$ denote a number between X_{ij} and X_ℓ, the first term on the right hand side of (8) can be written as

$$\begin{aligned}
&(Nn_ia_{n_i})^{-1}\sum_{\ell=1}^{N}\sum_{j=1}^{n_i} K\left(\frac{X_\ell - X_{ij}}{a_{n_i}}\right) (X_\ell - X_{ij}) \int_{-\infty}^{T} \eta_i'(Z_{ij}, \Delta_{ij}, y|X_{ij})\, dH(y) \\
&+\frac{1}{2}(Nn_ia_{n_i})^{-1}\sum_{\ell=1}^{N}\sum_{j=1}^{n_i} K\left(\frac{X_\ell - X_{ij}}{a_{n_i}}\right) (X_\ell - X_{ij})^2 \int_{-\infty}^{T} \eta_i''(Z_{ij}, \Delta_{ij}, y|\xi_{ij\ell})\, dH(y) \\
= \;& (n_ia_{n_i})^{-1}\sum_{j=1}^{n_i} \int K\left(\frac{u - X_{ij}}{a_{n_i}}\right) (u - X_{ij})\, d\Big\{\hat{G}(u) - G(u)\Big\} \\
&\times \int_{-\infty}^{T} \eta_i'(Z_{ij}, \Delta_{ij}, y|X_{ij})\, dH(y) + O(a_{n_i}^2) \\
&+\frac{1}{2}(Nn_ia_{n_i})^{-1}\sum_{\ell=1}^{N}\sum_{j=1}^{n_i} K\left(\frac{X_\ell - X_{ij}}{a_{n_i}}\right) (X_\ell - X_{ij})^2 \int_{-\infty}^{T} \eta_i''(Z_{ij}, \Delta_{ij}, y|\xi_{ij\ell})\, dH(y) \\
= \;& (n_ia_{n_i})^{-1}\sum_{j=1}^{n_i} \int K\left(\frac{u - X_{ij}}{a_{n_i}}\right) (u - X_{ij})\, d\Big\{\hat{G}(u) - G(u)\Big\} \\
&\times \int_{-\infty}^{T} \eta_i'(Z_{ij}, \Delta_{ij}, y|X_{ij})\, dH(y) + O(a_{n_i}^2), \text{ almost surely ,}
\end{aligned} \tag{9}$$

where the first $O(a_{n_i}^2)$ follows from the fact that $a_{n_i}^{-1} \int K\left(ua_{n_i}^{-1} - X_{ij}a_{n_i}^{-1}\right)(u - X_{ij})\, dG(u) = O(a_{n_i}^2)$, while the second term resulting from the Taylor expansion is $O(a_{n_i}^2)$ almost surely because assumptions A2(ii) b) and c), A3(ii)-(iv) and A4(iii) guarantee that $\partial\eta_i(z, \delta, y|x)/\partial x$ and $\partial^2\eta_i(z, \delta, y|x)/\partial x^2$ are bounded uniformly in $y, z \le T$, and $x \in R_X$. Arguing as in the proof of Lemma 3.2 of Akritas (1994), we set $\alpha_{n_i}(u) = n_i^{1/2}\{\hat{G}(u) - G(u)\}$ and write the term in (9) as

$$
\begin{aligned}
& n_i^{-3/2} a_{n_i}^{-1} \sum_{j=1}^{n_i} \int K\left(\frac{u - X_{ij}}{a_{n_i}}\right)(u - X_{ij})\, d\Big\{\alpha_{n_i}(u) - \alpha_{n_i}(X_{ij})\Big\} \int_{-\infty}^{T} \eta_i'(Z_{ij}, \Delta_{ij}, y|X_{ij})\, dH(y) \\
&= n_i^{-3/2} \sum_{j=1}^{n_i} \int K(v) v\, d\Big\{\alpha_{n_i}(X_{ij} + va_{n_i}) - \alpha_{n_i}(X_{ij})\Big\} \int_{-\infty}^{T} \eta_i'(Z_{ij}, \Delta_{ij}, y|X_{ij})\, dH(y) \\
&\le C n_i^{-3/2} n_i \sup_{|x_2 - x_1| \le a_{n_i}} |\alpha_{n_i}(x_2) - \alpha_{n_i}(x_1)| \;=\; o(n_i^{-1/2}), \qquad (10)
\end{aligned}
$$

where the inequality is derived by integrating by parts in the first integral. Relations (9) and (10) imply that the first term of (8) is $o(n_i^{-1/2})$ almost surely. The second term of (8) can be written in the following way :

$$
n_i^{-1} \sum_{j=1}^{n_i} \int_{-\infty}^{T} \eta_i(Z_{ij}, \Delta_{ij}, y|X_{ij})\, dH(y) \left\{(Na_{n_i})^{-1} \sum_{\ell=1}^{N} K\left(\frac{X_\ell - X_{ij}}{a_{n_i}}\right)\right\}.
$$

Next it will be shown that

$$
V_{n_i} = n_i^{-1} \sum_{j=1}^{n_i} \int_{-\infty}^{T} \eta_i(Z_{ij}, \Delta_{ij}, y|X_{ij})\, dH(y) \left\{(Na_{n_i})^{-1} \sum_{\ell=1}^{N} K\left(\frac{X_\ell - X_{ij}}{a_{n_i}}\right) - g(X_{ij})\right\}
$$

is $o_P(n_i^{-1/2})$. This, together with above results implies that the term in (7) is asymptotically equivalent to the first term in (9) and completes the proof. To show that $n_i^{1/2} V_{n_i} = o_P(1)$ it suffices to show that its conditional variance given X tends to zero. This, however, follows from

$$
n_i \mathrm{var}(V_{n_i}|X) = n_i^{-1} \sum_{j=1}^{n_i} \mathrm{var}\left\{\int_{-\infty}^{T} \eta_i(Z_{ij}, \Delta_{ij}, y|X_{ij})\, dH(y) \middle| X_{ij}\right\} \{\hat{g}(X_{ij}) - g(X_{ij})\}^2,
$$

and the fact that $\sup_x |\hat{g}(x) - g(x)| \to 0$, almost surely by Lemma A.1.

As in the proof of Theorem 3.1, if the covariate is discrete, (7) reduces to

$$\frac{1}{n_i}\sum_{j=1}^{n_i}\frac{\frac{1}{N}\sum_{\ell=1}^{N} I[X_\ell = X_{ij}]}{\frac{1}{n_i}\sum_{j'=1}^{n_i} I[X_{ij'} = X_{ij}]}\int_{-\infty}^{T}\xi_i(Z_{ij},\Delta_{ij},y|X_{ij})\,dH(y) + o(n_i^{-1/2}),\ \text{almost surely,}$$

for N large enough. A similar argument as was used to show that $V_{n_i} = o_P(n_i^{-1/2})$ can be applied to replace $\sum_{\ell=1}^{N} I[X_\ell = X_{ij}]/N$ and $\sum_{j'=1}^{n_i} I[X_{ij'} = X_{ij}]/n_i$ by the probability mass functions $g(X_{ij})$ and $g_i(X_{ij})$ in the above expression, yielding the desired result in the discrete case.

Proof of Corollary 3.3. First note that under H_0,

$$\hat{T}_C = C\int_{-\infty}^{T}\Big\{\hat{F}_{\cdot}(y) - F_{\cdot}(y)\Big\}\,d\hat{H}(y).$$

Theorems 3.1 and 3.2, imply that

$$\begin{aligned}
&N^{1/2}\int_{-\infty}^{T}(\hat{F}_i(y) - F_i(y))\,d\hat{H}(y) \qquad (11)\\
&= N^{1/2}\int\int_{-\infty}^{T} F_i(y|x)\,dH(y)\,d(\hat{G}(x) - G(x))\\
&\quad + \frac{N^{1/2}}{n_i}\sum_{j=1}^{n_i}\int_{-\infty}^{T}\xi_i(Z_{ij},\Delta_{ij},y|X_{ij})\,dH(y)\frac{g(X_{ij})}{g_i(X_{ij})} + o_P(1)\\
&= T_{i1} + T_{i2} + o_P(1).
\end{aligned}$$

It is readily seen that the covariance between T_{i_11} and T_{i_21} is equal to the first term of $\sigma^2_{i_1i_2}$. Since T_{i_12} and T_{i_22} are independent for $i_1 \neq i_2$, its covariance can be expressed as

$$I(i_1 = i_2)\frac{N}{n_{i_1}}\mathrm{var}\left\{\int_{-\infty}^{T}\xi_{i_1}(Z_{i_1j},\Delta_{i_1j},y|X_{i_1j})\,dH(y)\frac{g(X_{i_1j})}{g_{i_1}(X_{i_1j})}\right\}$$

which tends to the second term of $\sigma^2_{i_1i_2}$. By conditioning on $\mathbf{X}$ first, it is seen that T_{i_11} and T_{i_22} are uncorrelated for any i_1, i_2. Thus, the variance of $N^{1/2}\int_{-\infty}^{T}(\hat{F}_i(y) - F_i(y))\,d\hat{H}(y)$ is given by $\mathrm{var}(T_{i1}) + \mathrm{var}(T_{i2})$, which tends to σ^2_{ii} as $N \to \infty$. For $i_1 \neq i_2$,

$$\begin{aligned}
&\mathrm{cov}\left[N^{1/2}\int_{-\infty}^{T}\hat{H}(y)\,d\Big\{\hat{F}_{i_1}(y) - F_{i_1}(y)\Big\}, N^{1/2}\int_{-\infty}^{T}\hat{H}(y)\,d\Big\{\hat{F}_{i_2}(y) - F_{i_2}(y)\Big\}\right]\\
&= \mathrm{cov}(T_{i_11} + T_{i_12}, T_{i_21} + T_{i_22}) = N\mathrm{cov}(T_{i_11}, T_{i_21})
\end{aligned}$$

and this equals σ_{ij}^2. The asymptotic normality for $N^{1/2}\hat{\mathbf{T}}_{\mathbf{C}}$ follows from an application of the multivariate central limit theorem.

Proof of Proposition 3.4. From the uniform consistency of $\hat{H}_i(y|x)$ and $\hat{H}_{i1}(y|x)$ and the uniform consistency of $\hat{F}_i(y|x)$, see Lemma A.1, Remark A.1 and Proposition 4.1 in Van Keilegom & Akritas (1997), together with integration by parts, the uniform consistency of $\hat{\xi}_i(z,\delta,y|x)$ for $\xi_i(z,\delta,y|x)$ can be derived. These, coupled with the uniform consistency of $\hat{H}(y)$, imply the consistency of $\int_{-\infty}^{T}\hat{\xi}_i(z,\delta,y|x)\,d\hat{H}(y)$ for $\int_{-\infty}^{T}\xi_i(z,\delta,y|x)\,dH(y)$. Since

$$\max_j\left|\frac{n_i}{N}\sum_{\ell=1}^{N}W_{ij}(X_\ell,a_{n_i})-\frac{g(X_{ij})}{g_i(X_{ij})}\right|\to 0,\ \text{almost surely;}$$

see Section 5 in Akritas, Arnold & Du, (2000). The second term of $\hat{\sigma}_{i_1i_2}^2$ is a consistent estimator of the second term of $\sigma_{i_1i_2}^2$ by the strong law of large numbers. The consistency of the first term in $\hat{\sigma}_{ij}^2$ can be shown similarly.

References

Akritas, M. G. (1994). Nearest neighbor estimation of a bivariate distribution under random censoring. *Ann. Statist.*, **22**, 1299-1327.

Akritas, M. G. (1996). On the use of nonparametric regression techniques for fitting parametric regression models. *Biometrics* **52**, 1342-1362.

Akritas, M. G. & Brunner, E. (1997). Nonparametric methods for factorial designs with censored data. *J. Amer. Statist. Assoc.* **92**, 568-576.

Akritas, M. G., Arnold, S. F. & Du, Y. (2000). Nonparametric models and methods for nonlinear analysis of covariance. *Biometrika* **87**, 507-526.

Beran, R. (1981). Nonparametric regression with randomly censored survival data. Technical report, Univ. California, Berkeley.

Bennett, S. (1983). Analysis of survival data by the proportional odds model. *Statistics in Medicine* **2**, 273–277.

Brunner, E. Dette, H. & Munk, A. (1997). Box-type approximations in nonparametric factorial designs. *J. Amer. Statist. Assoc.* **92**, 1494-1502.

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

Copelan, E. A., Biggs, J. C., Thompson, J. M., Crilley, P., Sver, J., Klein, J. P., Kapoor, N., Avalos, B. R., Cunningham, I., Atkinson, K., Downs, K., Harmon, G. S., Daly, M. B., Brodsky, I., Bulova, S.I. & Tutschka, P. J. (1991). Treatment for acute myelocytic leukemia with allogeneic bone marrow transplantation following preparation with Bu/Cy. *Blood*, **78**, 838-843.

Du, Y. (1999). *Fully nonparametric ANCOVA models for censored data* Ph.D. Thesis, Department of Statistics, Penn State University.

Du, Y. & Akritas, M. G. (2000). Strong representation of the conditional Kaplan-Meier process. Preprint.

Dvoretzky, A., Kiefer, J. & Wolfowitz, J. (1956). Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator. *Ann. Math. Statist.*, **27**, 642-669.

Grigoletto, M. & Akritas, M. G. (2000). Analysis of covariance with incomplete data via semiparametric model transformations. *Biometrics* **55**, 1177-1187.

Huffer, F. W. & McKeague, I. W. (1991). Weighted least squares estimation for Aalen's additive risk model. *Journal of the American Statistical Association* **86**, 114–129.

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

Klein J. P. & Moeschberger, M. L. (1997). *Survival Analysis, Techniques for Censored and Truncated Data.* Springer-Verlag New York.

Lagakos, S. W., Barraj, L. M. & De Gruttola, V. (1988). Nonparametric analysis of truncated survival data, with application to AIDS. *Biometrika* **75**, 515–523.

Li, G. & Doss, H. (1995). An approach to nonparametric regression for life history data using local linear fitting. *Ann. Statist.*, **23**, 787-823.

Lin, D. Y. & Ying, Z. (1994). Semiparametric analysis of the additive risk model. *Biometrika* **81**, 61–71.

McKeague, I. W. & Sasieni, P. D. (1994). A partly parametric additive risk model. *Biometrika* **81**, 501–514.

McKeague, I. W. & Utikal, K. J. (1990). Inference for a nonlinear counting process regression model. *Ann. Statist.*, **18**, 1172-1187.

Pettitt, A. N. (1984). Proportional odds models for survival data and estimates using ranks. *Applied Statistics* **33**, 169–175.

Uspensky, J. V. (1937). *Introduction to Mathematical Probability.* McGraw-Hill, New York.

Van Keilegom, I. & Veraverbeke, N. (1997). Estimation and bootstrap with censored data in fixed design nonparametric regression. *Ann. Inst. Statist. Math.*, **49**, 467-491.

Table 1. Bandwidth Calibration for Covariate with Same Distributions (the numbers between parentheses are obtained by using the small sample correction procedure)

Constant	1%	5%	10%
7.5	0.0207 (0.0160)	0.0740 (0.0667)	0.1267 (0.1227)
10	0.0187 (0.0140)	0.0767 (0.0627)	0.1207 (0.1127)
13	0.0193 (0.0147)	0.0673 (0.0593)	0.1193 (0.1120)
14	0.0193 (0.0153)	0.0653 (0.0593)	0.1180 (0.1107)
15	0.0193 (0.0147)	0.0660 (0.0560)	0.1180 (0.1113)
16	0.0187 (0.0140)	0.0667 (0.0553)	0.1180 (0.1107)
17	0.0187 (0.0133)	0.0660 (0.0553)	0.1173 (0.1087)
20	0.0200 (0.0133)	0.0660 (0.0567)	0.1180 (0.1087)
30	0.0187 (0.0127)	0.0633 (0.0540)	0.1180 (0.1087)
10000	0.0193 (0.0120)	0.0620 (0.0540)	0.1120 (0.1033)

Table 2. Bandwidth Calibration for Covariate with Different Distributions (the numbers between parentheses are obtained by using the small sample correction procedure)

Constant	1%	5%	10%
7.5	0.0207 (0.0167)	0.0767 (0.0700)	0.1373 (0.1333)
10	0.0153 (0.0133)	0.0700 (0.0620)	0.1260 (0.1173)
13	0.0147 (0.0113)	0.0680 (0.0600)	0.1133 (0.1093)
14	0.0120 (0.0113)	0.0667 (0.0593)	0.1180 (0.1107)
15	0.0127 (0.0113)	0.0653 (0.0580)	0.1180 (0.1127)
16	0.0140 (0.0100)	0.0646 (0.0580)	0.1193 (0.1113)
17	0.0153 (0.0120)	0.0700 (0.0593)	0.1260 (0.1120)
20	0.0153 (0.0127)	0.0667 (0.0587)	0.1200 (0.1133)
30	0.0200 (0.0160)	0.0693 (0.0620)	0.1253 (0.1147)
10000	0.0233 (0.0153)	0.0773 (0.0660)	0.1420 (0.1167)

Table 3. Large-sample Performance of the Nonparametric Procedure

	Same Covariate Distribution			Different Covariate Distribution		
Constant	1%	5%	10%	1%	5%	10%
10	0.0110	0.0450	0.0990	0.0130	0.0500	0.0990
15	0.0080	0.0420	0.1010	0.0120	0.0520	0.1000
30	0.0090	0.0510	0.0940	0.0150	0.0650	0.1190
10000	0.0100	0.0500	0.0970	0.0280	0.1050	0.1710

Table 4. Power Study (the first line shows the power of the nonparametric method, the second line the power of the proportional hazards method)

τ	$p = 0.25$			$p = 0.50$			$p = 0.75$		
	1%	5%	10%	1%	5%	10%	1%	5%	10%
1.0	0.0110	0.0523	0.1073	0.0120	0.0530	0.1057	0.0103	0.0600	0.1120
	0.0120	0.0520	0.0990	0.0110	0.0490	0.1040	0.0040	0.0450	0.0883
1.2	0.0360	0.1327	0.2153	0.0310	0.1230	0.2120	0.0227	0.0987	0.1837
	0.0307	0.1180	0.1947	0.0207	0.1160	0.1863	0.0070	0.0640	0.1413
1.4	0.1333	0.3267	0.4577	0.1220	0.2963	0.4170	0.0810	0.2433	0.3633
	0.0983	0.2750	0.3887	0.0907	0.2430	0.3717	0.0203	0.1623	0.2777
1.6	0.3207	0.5717	0.6913	0.2567	0.5057	0.6363	0.1620	0.3863	0.5360
	0.2410	0.4913	0.6140	0.1893	0.4317	0.5623	0.0327	0.2547	0.4200
1.8	0.5293	0.7670	0.8557	0.4480	0.6950	0.7910	0.2723	0.5503	0.6757
	0.4223	0.6700	0.7737	0.3283	0.6070	0.7233	0.0573	0.3260	0.4937
2.0	0.7210	0.8850	0.9393	0.6030	0.8243	0.8950	0.3507	0.6657	0.7943
	0.5940	0.8110	0.8913	0.4670	0.7527	0.8517	0.0600	0.3523	0.5297

Table 5. Data Analysis Results

	Proportional	Nonparametric model	
Effect	hazards model	Const.= 50	Const.= 60
Methotrexate	0.8173	0.0824	0.0663
Disease	0.1170	0.0487	0.0392
Interaction	0.2525	0.3916	.4193

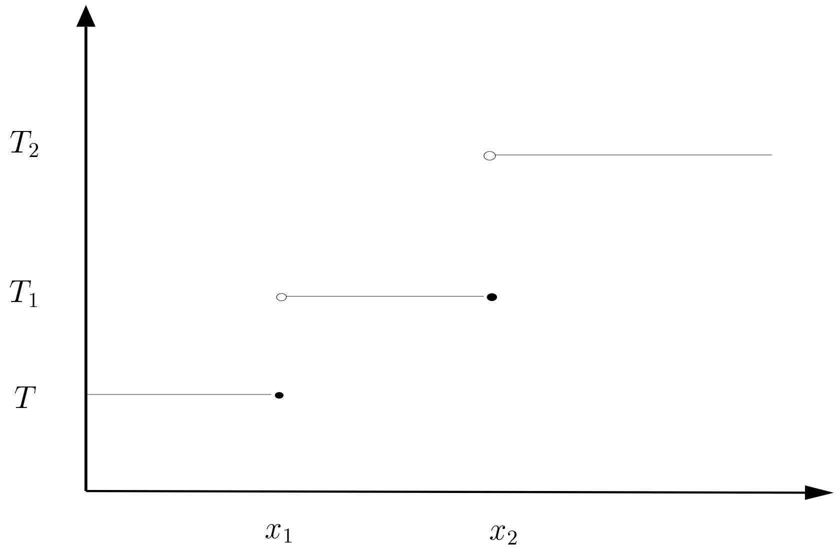


Figure 1: An Example of Variable Censoring Pattern

(a)

(b)

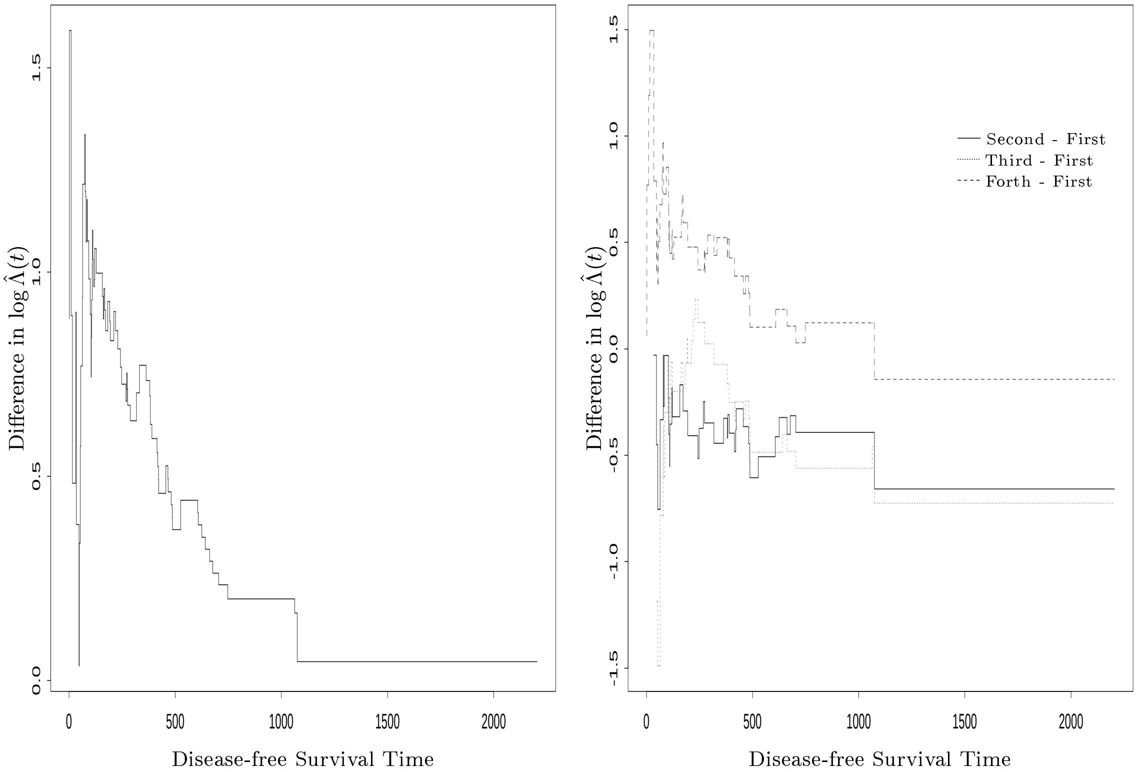
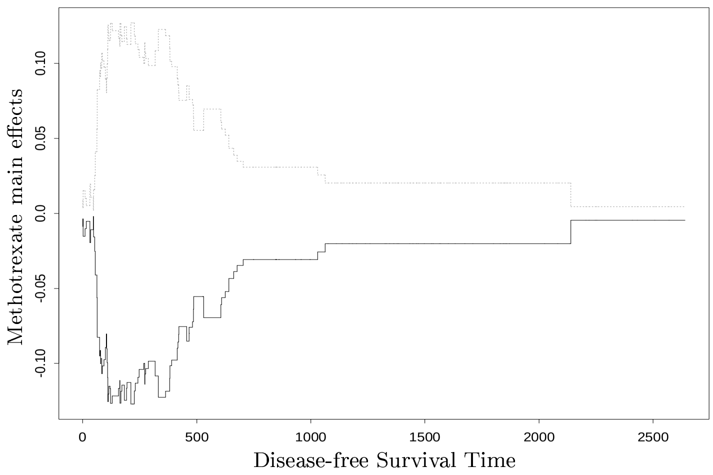
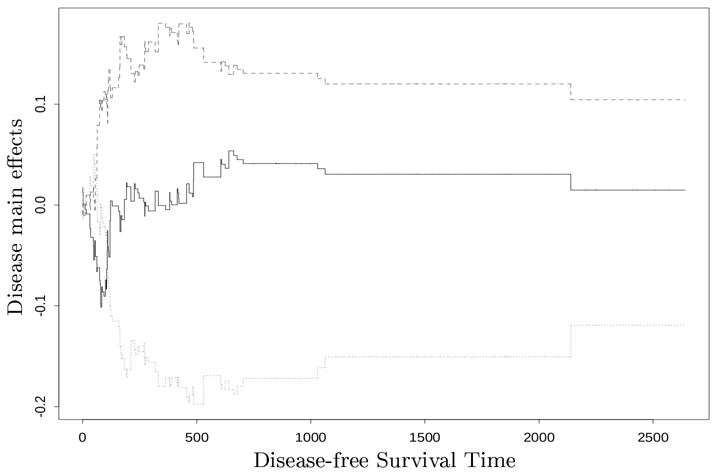


Figure 2: Difference in Log Cumulative Hazard Functions

(a)



(b)



(b)

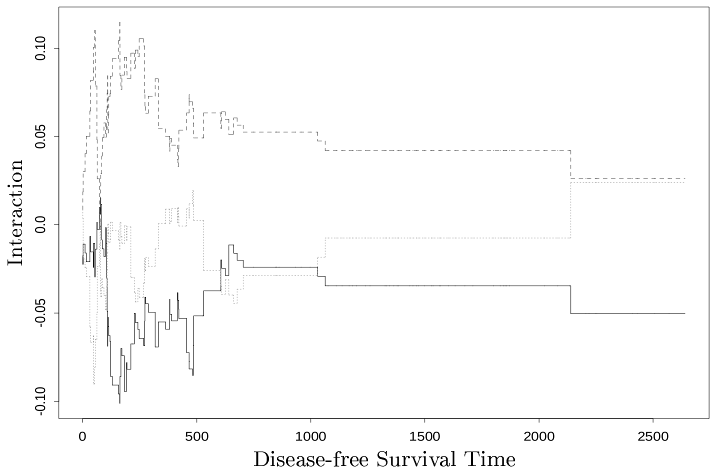


Figure 3: Nonparametric Effects

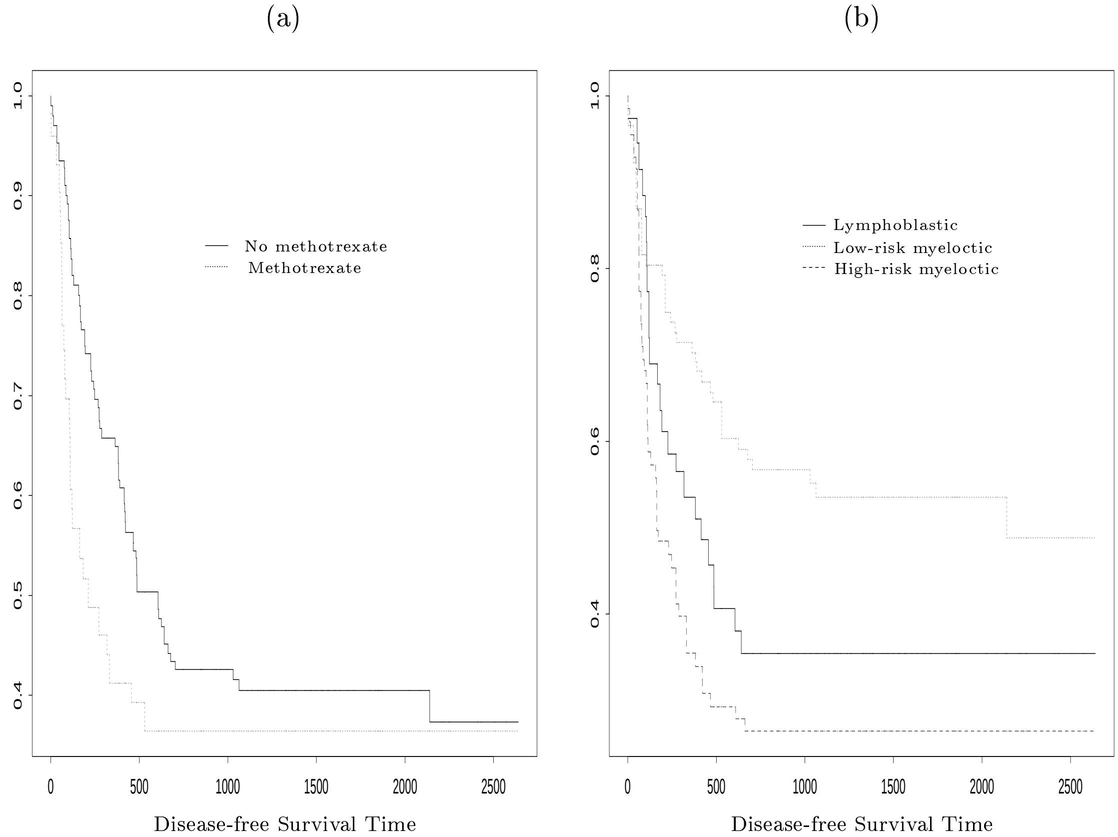


Figure 4: Averaged Beran Estimates of the Disease-free Survival Times