Statistica Sinica **12**(2002), 689-705

COVARIATE-ADJUSTED DEPENDENCE ESTIMATION ON A FINITE BIVARIATE FAILURE TIME REGION

Juanjuan Fan and Ross L. Prentice

University of California at Davis and
Fred Hutchinson Cancer Research Center

Abstract: Recently Fan and colleagues have proposed two measures of the strength of dependency between failure time variates over a finite region of the sample space; namely, an average relative risk measure and a finite region version of Kendall's tau. Here, these dependency measures are generalized to accommodate regression effects on marginal hazard functions. Specifically, the dependency measures previously proposed are applied to possibly censored, estimated cumulative hazard variates, calculated under Cox model marginal hazard function models. The resulting dependency estimators use a nonparametric estimator of the bivariate survivor function and are shown to be consistent and asymptotically normally distributed, with consistent bootstrap variance estimators, for certain classes of covariates. The small sample properties of the estimators, and their variance estimators, are examined in simulation studies and the estimators are compared to corresponding homogeneous dependency estimators that do not condition on covariates.

Key words and phrases: Bootstrap, censoring, cox model, cross ratio function, Kendall's tau, relative risk.

1. Introduction

The statistical analysis of multivariate failure time data typically includes characterization of the pairwise dependence between failure time variates. For example, a difference in strength of disease association between monozygotic (MZ) and dizygotic (DZ) twins may suggest a genetic component to disease risk. Most previous approaches to estimating a summary measure of association (Clayton and Cuzick (1985), Oakes (1982, 1986), Hsu and Prentice (1996)) assume a semiparametric bivariate failure time model in which a single parameter is assumed to govern the association between the two failure times. The model of Clayton (1978) is the most popular and most studied in this category, but it makes the rather strong assumption that the so-called cross ratio (Clayton (1978), Oakes (1989)) is constant over time. Hence, there is a need for nonparametric measures of association that do not impose assumptions on the form of the bivariate survivor function. In the presence of independent right censoring, the support for failure time variates (T_1, T_2) may be restricted, and measures that express the dependence between T_1 and T_2 over a finite region $[0, t_1] \times [0, t_2]$ are of particular interest. One approach to the development of finite region dependency measures is to consider a weighted average of some readily interpreted local dependency measure over such a region.

Recently Fan, Hsu and Prentice (2000a) have proposed two finite region dependency measures. One, an average relative risk measure, is given by

$$C(t_1, t_2) = \int_0^{t_1} \int_0^{t_2} c(s_1, s_2)\ F(ds_1, ds_2) \ / \int_0^{t_1} \int_0^{t_2} F(ds_1, ds_2), \tag{1}$$

where $F(s_1, s_2) = P(T_1 > s_1, T_2 > s_2)$ is the joint survivor function for (T_1, T_2), $F(ds_1, ds_2)$ is the corresponding density function and

$$c = F(ds_1, s_2^-)\ F(s_1^-, ds_2)/\{F(ds_1, ds_2)\ F(s_1^-, s_2^-)\} \tag{2}$$

is the reciprocal of the cross ratio function (Clayton (1978), Oakes (1989)). It can also be written $c = \lambda_1(s_1 \mid T_2 \geq s_2)/\lambda_1(s_1 \mid T_2 = s_2) = \lambda_2(s_2 \mid T_1 \geq s_1)/\lambda_2(s_2 \mid T_1 = s_1)$, where λ_1 and λ_2 are hazard functions, explaining the relative risk terminology. The other is an average concordance measure

$$\begin{aligned}\mathcal{T}(t_1, t_2) &= \frac{\int_0^{t_1}\int_0^{t_2} F(ds_1, ds_2)F(s_1^-, s_2^-) - \int_0^{t_1}\int_0^{t_2} F(ds_1, s_2^-)F(s_1^-, ds_2)}{\int_0^{t_1}\int_0^{t_2} F(ds_1, ds_2)F(s_1^-, s_2^-) + \int_0^{t_1}\int_0^{t_2} F(ds_1, s_2^-)F(s_1^-, ds_2)} \\ &= E\{\text{sign}(T_{11} - T_{21})(T_{12} - T_{22}) \mid T_{11} \wedge T_{21} \leq t_1,\ T_{12} \wedge T_{22} \leq t_2\}, \end{aligned} \tag{3}$$

where (T_{11}, T_{12}) and (T_{21}, T_{22}) are independent variates with survivor function F, and "$\wedge$" denotes minimum. Note that $\mathcal{T}(t_1, t_2)$ takes values in $[-1, 1]$ and approaches Kendall's τ as $t_1, t_2 \to \infty$ in the absence of cersoring. Nonparametric dependence estimators were obtained by inserting nonparametric estimators (e.g., Dabrowska (1988), Prentice and Cai (1992)) of the bivariate survivor function and empirical estimators of the bivariate cumulative hazard functions into (1) and (3). These estimators, as a function of (t_1, t_2), were shown to be strongly consistent and weakly convergent to a Gaussian process, and the asymptotic validity of the bootstrap was shown. Corresponding results for a more general class of estimators are given in Fan, Prentice and Hsu (2000b).

In this article we generalize the dependence measures of Fan et al. (2000a) to allow for covariate effects on marginal hazard rates. Beyond the usual independent censorship assumption, our principal assumption is that the joint distribution of marginal cumulative hazard variates, which have unit exponential marginal distributions if the failure times are absolutely continuous, is independent of covariates. We comment in the Discussion section that this assumption may be relaxed. Cox (1972) model marginal hazard functions are also assumed for convenience. The average relative risk and finite region concordance measures of Fan et al. (2000a) are then applied to the estimated cumulative hazard variates, following the estimation of Cox model regression parameters and baseline hazard functions. The resulting estimators have an attractive interpretation as estimators of interpretable nonparametric dependency measures for the standandized (unit exponential) variates that, by their construction, have accommodated regression effects on marginal hazard rates. These estimators are shown to have similar asymptotic properties to those previously given for homogeneous failure times under i.i.d. sampling and standard regularity conditions, provided the support of the covariates is a finite set. Simulation results suggest that this latter assumption is unlikely to be necessary, and indicate that the small sample performance of the covariate-accommodated dependence estimates is similar to that for the corresponding homogeneous data estimators. Finally, simulation results are used to illustrate the major differences that may arise between dependence measure estimates that do or do not allow for covariate effects on marginal hazard rates.

2. Marginal Proportional Hazards Models

Suppose that the observed data consist of $\{X_{ki}, \delta_{ki}, Z_{ki}\}$, $i = 1, 2$; $k = 1, \ldots, K$, where $X_{ki} = T_{ki} \wedge U_{ki}$, the minimum of the failure time and the censoring time; $\delta_{ki} = I_{\{X_{ki}=T_{ki}\}}$; and Z_{ki} is a fixed vector of covariates. If T_{ki} is missing, let $U_{ki} = 0$. This implies that $X_{ki} = 0$ and $\delta_{ki} = 0$. By doing so, we allow the rate of missingness to depend on covariate values at time zero. As usual the failure times $T_k = (T_{k1}, T_{k2})$ are assumed to be independent of the potential censoring times $U_k = (U_{k1}, U_{k2})$ given covariate values $Z'_K = (Z'_{k1}, Z'_{k2})$. Furthermore, we assume that $(T_{k1}, T_{k2}, U_{k1}, U_{k2}, Z'_{k1}, Z'_{k2})'$, $k = 1, \ldots, K$, are independent replicates of random variables $(T_1, T_2, U_1, U_2, Z'_1, Z'_2)$. Suppose that the marginal hazard rate for the ith individual in the kth pair follows a proportional hazards model (Cox, 1972)

$$\lambda_{ki}(t; Z_k) = \lambda_{0i}(t)e^{\beta_0' Z_{ki}}, \qquad t \geq 0 \tag{4}$$

where $Z'_{ki} = (Z_{ki1}, \ldots, Z_{kip})$ is a p-dimensional covariate vector, λ_{0i} is an unknown and unspecified baseline hazard pertaining to the ith member of each pair, and β_0 is a p-vector of unknown regression coefficients. Note that this model does allow distinct failure-specific regression coefficients (different β_0's for $i = 1, 2$) by a suitable definition of Z_{ki} on the right side of (4) and that a straightforward extension of the methods presented here, in conjunction with the results of Lee, Wei and Amato (1992), would apply if the baseline hazard functions λ_{01} and λ_{02} were restricted to be equal.

If, in addition, T_{k1} and T_{k2}, given Z_K, were assumed to be statistically independent of each other for each $k = 1, \ldots, K$, the partial likelihood function for β based on data from the K pairs of individuals could be written in counting process notation as

$$L(\beta) = \prod_{k=1}^{K} \prod_{i=1}^{2} \prod_{t \geq 0} [\frac{e^{\beta' Z_{ki}}}{\sum_{l=1}^{K} Y_{li}(t) e^{\beta' Z_{li}}}]^{N_{ki}(dt)},$$

where $Y_{ki}(t) = I\{X_{ki} \geq t\}$ and $N_{ki}(t) = I\{X_{ki} \leq t,\, \delta_{ki} = 1\}$ for $t \geq 0$, $i = 1, 2$ and $k = 1, \ldots, K$, and where $I\{\cdot\}$ is an indicator function. Wei, Lin and Weissfeld (1989) proposed that, in the dependent failure time case, β_0 still be estimated by $\hat{\beta}$ that maximizes $L(\beta)$. It can be shown that this estimator $\hat{\beta}$ is consistent and asymptotically Gaussian (e.g., Wei, Lin and Weissfeld (1989), Cai and Prentice (1995)) under departure from the independence of T_{k1} and T_{k2} given Z_k, though it is typically not semiparametrically efficient.

Under (4), the cumulative marginal hazard function $\Lambda_{ki}(t)$ can be estimated by $\hat{\Lambda}_{ki}(t) = e^{\hat{\beta}' Z_{ki}} \hat{\Lambda}_{0i}(t; \hat{\beta})$, where $\hat{\Lambda}_{0i}$ is the Breslow (1972, 1974) cumulative hazard estimator

$$\hat{\Lambda}_{0i}(t; \beta) = \int_0^t \frac{\sum_{k=1}^{K} N_{ki}(du)}{\sum_{k=1}^{K} Y_{ki}(u) e^{\beta' Z_{ki}}}. \tag{5}$$

It can be shown that the estimator $\hat{\Lambda}_{0i}(t; \hat{\beta})$ is uniformly consistent for $\Lambda_{0i}(t)$ for $t \leq r$, where $r = (r_1, r_2)$ satisfies $P\{X_1 \geq r_1, X_2 \geq r_2) > 0$. The proof follows the arguments of Andersen and Gill (1982) and is not presented here.

3. Covariate-Adjusted Average Relative Risk Estimator and a Finite Region Version of Kendall's τ

The average relative risk dependency measure (1) can be adjusted to accommodate covariate effects on marginal hazard rates by defining

$$C_\Lambda(t_1, t_2) = \int_0^{t_1} \int_0^{t_2} c_\Lambda F_\Lambda(ds_1, ds_2) / \int_0^{t_1} \int_0^{t_2} F_\Lambda(ds_1, ds_2), \tag{6}$$

where $F_\Lambda(s_1, s_2) = P\{\Lambda_1(T_1) > s_1,\ \Lambda_2(T_2) > s_2\}$. Note that $\Lambda_i(T_i) = \int_0^{T_i} \lambda_{0i}(u_i) e^{\beta_0' Z_i} du_i$ is the "standardized" cumulative hazard variate that is distributed as exponential(1) for absolutely continuous T_i, and that $\{\Lambda_1(T_1), \Lambda_2(T_2)\}$ is assumed to be independent of $Z' = (Z_1', Z_2')$ (see Discussion for relaxation of this assumption). Also c_Λ is the cross ratio function in (2) with F replaced by F_Λ, and (t_1, t_2) is in the support of $\{\Lambda_1(X_1), \Lambda_2(X_2)\}$, so that $P\{\Lambda_1(T_1) \geq t_1$, $\Lambda_2(T_2) \geq t_2$, $\Lambda_1(U_1) \geq t_1$, $\Lambda_2(U_2) \geq t_2\} > 0$. Define $H_\Lambda = (H_\Lambda^{11}, H_\Lambda^{10}, H_\Lambda^{01})$ with

$$H_\Lambda^{11}(t_1, t_2) = \int_0^{t_1} \int_0^{t_2} P\{\Lambda(T) \in [s, s + ds) \mid \Lambda(T \wedge U) \geq s\},$$

$$H_\Lambda^{10}(t_1, t_2) = \int_0^{t_1} P\{\Lambda_1(T_1) \in [s_1, s_1 + ds_1) \mid \Lambda(T \wedge U) \geq s\},$$

$$H_\Lambda^{01}(t_1,t_2) = \int_0^{t_2} P\{\Lambda_2(T_2) \in [s_2, s_2+ds_2) \mid \Lambda(T \wedge U) \geq s\},$$

where $s = (s_1, s_2)$, $\Lambda(T) \in [s, s+ds)$ represents $\Lambda_1(T_1) \in [s_1, s_1+ds_1)$ and $\Lambda_2(T_2) \in [s_2, s_2+ds_2)$, etc. Then (6) can be rewritten as

$$C_\Lambda(t_1,t_2) = \int_0^{t_1}\int_0^{t_2} H_\Lambda^{10}(ds_1, s_2) H_\Lambda^{01}(s_1, ds_2) F_\Lambda \{\int_0^{t_1}\int_0^{t_2} F_\Lambda(s_1,s_2)\}^{-1}.$$

Since F_Λ is a Hadamard differentiable functional of H_Λ (Dabrowska (1989), Gill, van der Laan and Wellner (1995)), one has $C_\Lambda = \Phi(H_\Lambda)$, where the functional Φ is a special case of that considered in Fan et al. (2000a, 2000b).

A natural estimator of C_Λ is given by $\hat{C}_\Lambda = \Phi(\hat{H}_\Lambda)$, where $\hat{H}_\Lambda = (\hat{H}_\Lambda^{11}, \hat{H}_\Lambda^{10}, \hat{H}_\Lambda^{01})$ with

$$\hat{H}_\Lambda^{11}(t_1,t_2) = \int_0^{t_1}\int_0^{t_2} \frac{\frac{1}{K}\sum_{k=1}^K I(\hat{\Lambda}_k(T_k) \in [s, s+ds), \hat{\Lambda}_k(U_k) \geq s)}{\frac{1}{K}\sum_{k=1}^K I(\hat{\Lambda}_k(T_k) \geq s, \hat{\Lambda}_k(U_k) \geq s)}$$

and similar expressions for $\hat{H}_\Lambda^{10}(t_1,t_2)$ and $\hat{H}_\Lambda^{01}(t_1,t_2)$, and where $\hat{\Lambda}_{ki}(t) = e^{\hat{\beta}' Z_{ki}} \hat{\Lambda}_{0i}(t;\hat{\beta})$ and $\hat{\Lambda}_{0i}$ is the Breslow estimator as given in (5). Note that in the absence of covariates, one has $C_\Lambda(t_1,t_2) = C\{\Lambda_1^{-1}(t_1), \Lambda_2^{-1}(t_2)\}$ and $\hat{C}_\Lambda(t_1,t_2) = \hat{C}\{\hat{\Lambda}_1^{-1}(t_1), \hat{\Lambda}_2^{-1}(t_2)\}$.

A regression generalization can also be developed for the finite region concordance measure (3) by considering a general hazard function H_Λ with respect to $\Lambda_1(T_1)$ and $\Lambda_2(T_2)$. Similarly, $\mathcal{T}_\Lambda = \Psi(H_\Lambda)$, with Ψ a uniformly Hadamard differentiable function, which can be estimated by $\hat{\mathcal{T}}_\Lambda = \Psi(\hat{H}_\Lambda)$. In the absence of covariates one has $\mathcal{T}_\Lambda(t_1,t_2) = \mathcal{T}\{\Lambda_1^{-1}(t_1), \Lambda_2^{-1}(t_2)\}$ and $\hat{\mathcal{T}}_\Lambda(t_1,t_2) = \hat{\mathcal{T}}\{\hat{\Lambda}_1^{-1}(t_1), \hat{\Lambda}_2^{-1}(t_2)\}$.

4. Asymptotic Properties of the Proposed Estimators

In this section we state strong consistency, weak convergence and asymptotic validity of the bootstrap results for the covariate-adjusted average relative risk estimator $\hat{C}_\Lambda$ and outline the proofs, with much of the technical details deferred to the Appendices. The asymptotic properties and corresponding proofs for $\hat{\mathcal{T}}_\Lambda$ are essentially the same and are omitted for the sake of brevity.

Theorem 1. [Strong consistency of $\hat{C}_\Lambda$] *Suppose*

1. *The absolutely continuous failure times (T_1, T_2) and the censoring times (U_1, U_2) are independent given the covariate vectors (Z_1, Z_2);*
2. *The joint survivor function for the unit exponential cumulative hazard variates $\Lambda_1(T_1)$ and $\Lambda_2(T_2)$, denoted by F_Λ, is independent of (Z_1, Z_2). To avoid singular distributions F_Λ is required to possess a density $F_\Lambda(ds_1, ds_2)$ relative to Lebesgue measure;*

3. $(T_{k1}, T_{k2}, U_{k1}, U_{k2}, Z'_{k1}, Z'_{k2})'$, $k = 1, \ldots, K$, *are independent with the same distribution as* $(T_1, T_2, U_1, U_2, Z'_1, Z'_2)$*;*
4. *There exist* $\delta = (\delta_1, \delta_2)$ *and* $\gamma = (\gamma_1, \gamma_2)$ *with* $0 < \delta_i < \gamma_i$, $i = 1, 2$, *such that*

$$\int_0^{\delta_1} \int_0^{\delta_2} F_\Lambda(ds_1, ds_2) > 0, \tag{7}$$

$$P\{\Lambda_1(T_1) \geq \gamma_1, \Lambda_2(T_2) \geq \gamma_2, \Lambda_1(U_1) \geq \gamma_1, \Lambda_2(U_2) \geq \gamma_2\} > 0; \tag{8}$$

5. *The marginal hazard rates* $\lambda_1(t_1 \mid Z)$ *and* $\lambda_2(t_2 \mid Z)$ *are of Cox model form with* $\lambda_i(t_i \mid Z) = \lambda_{0i}(t_i)e^{\beta_0' Z_i}$, *where* Z_1 *and* Z_2 *are time-independent and take values only in a finite set.*

Then $\hat{C}_\Lambda(t_1, t_2)$ *converges to* $C_\Lambda(t_1, t_2)$ *almost surely and uniformly for* $t \in [\delta, \gamma]$.

To prove this strong consistency result for $\hat{C}_\Lambda$, it suffices to prove the strong consistency of $\hat{H}_\Lambda$ since $\hat{C}_\Lambda = \Phi(\hat{H}_\Lambda)$ and the functional Φ is Hadamard differentiable (and thus continuous). Note in the definition of $\hat{H}_\Lambda$ that the K indicator functions are not independent because all K pairs are used in estimating $\hat{\Lambda}_{ki}$. Thus the usual laws of large numbers (or central limit theorems) do not apply here. The indicator functions also complicate the asymptotic distribution theory for $\hat{C}_\Lambda$. Since the indicator function is not smooth, Taylor expansion type methods also do not apply. The key to the asymptotic results for $\hat{C}_\Lambda$ is the lemma given in Appendix A, which establishes asymptotic equivalence between dependent sums and independent sums. The strong consistency of $\hat{H}_\Lambda$ then follows from the Glivenko-Cantelli Theorem and the Continuous Mapping Theorem.

Theorem 2. [Weak convergence of $\hat{C}_\Lambda$ and $C^{\#}_\Lambda$] *Under the conditions of Theorem* 1, $\sqrt{K}(\hat{C}_\Lambda - C_\Lambda)$ *converges weakly on* $D[\delta, \gamma]$ *to a Gaussian process and, given observed data, the bootstrapped process* $\sqrt{K}(C^{\#}_\Lambda - \hat{C}_\Lambda)$ *converges weakly to the same Gaussian process almost surely, where* $D[\delta, \gamma]$ *denotes all cadlag functions on* $[\delta, \gamma]$.

Using the lemma in Appendix A, it can be shown that $\sqrt{K}(\hat{H}_\Lambda - H_\Lambda)$ is a uniformly Hadamard differentiable functional of empirical processes. Thus the weak convergence of $\hat{H}_\Lambda$ and $H^{\#}_\Lambda$ (a.s.) follows from the weak convergence of empirical processes, by using the functional delta-method (e.g., Gill et al. (1995), van der Vaart and Wellner (1996)). Applying the functional delta-method to Φ in turn gives the weak convergence of $\hat{C}_\Lambda$ and $C^{\#}_\Lambda$. See Appendix B for details.

This method does not give an explicit variance formula for $\hat{C}_\Lambda$ due to the complexity of Φ, which inherits from the complexity of the bivariate survivor function representations (e.g., Gill et al. (1995)). However, the convergence of $\hat{C}_\Lambda$ and $C^{\#}_\Lambda$ to the same limiting process implies that a bootstrap procedure can be used for variance estimation of $\hat{C}_\Lambda$.

5. Simulation Study

5.1. Small sample properties of $\hat{C}_\Lambda$ and $\hat{\mathcal{T}}_\Lambda$

The performances of the average relative risk estimator $\hat{C}_\Lambda$ and finite region concordance estimator $\hat{\mathcal{T}}_\Lambda$ were studied in samples of moderate size, when there were covariate effects on marginal hazards. For simplicity, we restricted ourselves to the situation where there is only one time-independent covariate for each individual. The small sample properties of $\hat{C}_\Lambda$ were investigated when (i) $Z_i, i = 1, 2$, take values independently in $\{0, 1\}$ with $P(Z_1 = 1) = P(Z_2 = 0) = 0.2$; and (ii) $Z_i, i = 1, 2$, follow independent exponential distributions with the same means as in (i).

Failure times were generated under independence, Clayton (1978) and Frank (e.g., Genest (1987)) models with unit exponential margins. For example, the extension of the Frank family model with covariates can be written as

$$F(t_1,t_2) = e^{-t_1 e^{\beta' Z_1}} + e^{-t_2 e^{\beta' Z_2}} - 1 + \log_\alpha\{1 + \frac{(\alpha^{1-e^{-t_1 e^{\beta' Z_1}}} - 1)(\alpha^{1-e^{-t_2 e^{\beta' Z_2}}} - 1)}{\alpha - 1}\}.$$

The failure times (t_1, t_2) were generated from independent uniform $(0, 1)$ variates (v_1, v_2) using transformations $t_1 = -e^{-\beta_0 z_1} \log\, v_1, t_2 = -e^{-\beta_0 z_2} \log\, v_2$ for the independence configuration; $t_1 = (\theta - 1)^{-1}\, e^{-\beta_0 z_1} \log\, \{(1 - a) + a v_1^{-(\theta-1)/\theta}\}, t_2 = -e^{-\beta_0 z_2} \log\, v_2$, where $a = v_2^{-(\theta-1)}$ for the Clayton model with $\theta = 4$; and $t_1 = -e^{-\beta_0 z_1} \log\, v_1, t_2 = -e^{-\beta_0 z_2} \log(\log_\alpha[a/\{a + (1 - \alpha)v_2\}])$, where $a = \alpha^{v_1} + (\alpha - \alpha^{v_1})v_2$ for the Frank family model with $\alpha = 0.0023$. The value $\alpha = 0.0023$ was chosen for the Frank family to give $C_\Lambda(0.6931, 0.6931) = 0.25$ in order to match the constant $c_\Lambda(t_1, t_2) = C_\Lambda(t_1, t_2) = 0.25$ for all t_1, t_2 under the Clayton model, 0.6931 being the median of the unit exponential distribution. Covariate values were obtained by transforming independent uniform $(0, 1)$ variates.

The censoring distribution was allowed to depend on covariates according to $P(U_1 > u_1, U_2 > u_2; Z_1, Z_2) = e^{-\frac{u_1}{2} e^{\beta' Z_1}} e^{-\frac{u_2}{2} e^{\beta' Z_2}}$, so that each failure time still has a marginal probability of 1/3 of being censored. Three sets of 200 simulations were carried out with sample size $K = 100$ in each configuration. When the performance of the estimator was poor at $K = 100$, larger sample sizes were also considered. The bootstrap approximation to the variance was evaluated based on 50 bootstrap samples. The Newton-Raphson method was used to find the root of the score equation $\partial \log L(\beta)/\partial\beta = 0$ (see Section 2). The true β_0 was specified as 1 throughout, which represents a 2.72-fold increase in the hazard rate on each margin, for each unit increase in the covariate value.

Table 1 shows simulation summary statistics for the average relative risk estimator $\hat{C}_\Lambda$ when there is a single exponentially distributed covariate for each subject. The performance of $\hat{C}_\Lambda$ with Bernoulli distributed covariates was similar, and is not shown for the sake of brevity. The entries of Table 1 are the true values of covariate-adjusted average relative risk C_Λ, sample means and sample standard deviations for $\hat{C}_\Lambda$, the means of bootstrap standard deviation estimates, and corresponding estimated coverage rates for nominal 95% confidence intervals, for values of (t_1, t_2) in the grid formed by the 25th, 50th and 75th percentiles of the marginal distributions for $\Lambda_1(T_1)$ and $\Lambda_2(T_2)$.

Table 1. Simulation summary statistics for the covariate-adjusted average relative risk estimator $\hat{C}_\Lambda$ at selected percentiles of the unit exponential marginal distributions for $\Lambda_1(T_1)$ and $\Lambda_2(T_2)$. Also shown are the average of bootstrap estimates of standard deviation of $\hat{C}_\Lambda$, based on 50 bootstrap samples, and corresponding estimated 95% confidence interval coverage probabilities. Each entry is based on 200 simulation runs of $K = 100$ pairs of failure times, with independent exponential censoring times having a mean of two and time-independent exponentially distributed covariates with $E(Z_1) = 0.2$ and $E(Z_2) = 0.8$. Summary statistics are shown for (T_1, T_2) independent, as well as for positively dependent failure times from Clayton and Frank family models.

$\Lambda_1(T_1)$ Percentile	(T_1, T_2) Distribution	$\Lambda_2(T_2)$ Percentile 25		$\Lambda_2(T_2)$ Percentile 50		$\Lambda_2(T_2)$ Percentile 75	
25	Indep.	1.00	1.22 (1.016,2.143,0.950)*	1.00	1.08 (0.514,1.030,0.945)	1.00	1.00 (0.309,0.463,0.935)
	Clayton	0.25	0.27 (0.108,0.129,0.945)	0.25	0.27 (0.080,0.086,0.955)	0.25	0.27 (0.085,0.081,0.960)
	Frank	0.20	0.22 (0.068,0.084,0.955)	0.22	0.23 (0.065,0.073,0.955)	0.23	0.24 (0.070,0.076,0.935)
50	Indep.	1.00	1.06 (0.504,0.911,0.920)	1.00	1.01 (0.278,0.315,0.940)	1.00	0.96 (0.209,0.219,0.910)
	Clayton	0.25	0.26 (0.077,0.080,0.960)	0.25	0.27 (0.063,0.063,0.945)	0.25	0.27 (0.062,0.065,0.960)
	Frank	0.22	0.23 (0.059,0.072,0.960)	0.25	0.26 (0.057,0.063,0.945)	0.28	0.28 (0.063,0.071,0.945)
75	Indep.	1.00	0.99 (0.312,0.517,0.915)	1.00	0.97 (0.227,0.225,0.910)	1.00	0.92 (0.177,0.153,0.845)
	Clayton	0.25	0.26 (0.073,0.074,0.945)	0.25	0.27 (0.062,0.063,0.965)	0.25	0.28 (0.064,0.069,0.975)
	Frank	0.23	0.24 (0.069,0.077,0.935)	0.28	0.28 (0.064,0.073,0.930)	0.33	0.33 (0.075,0.080,0.955)

* Entries are the true value of covariate-adjusted average relative risk C_Λ, the mean of covariate-adjusted average relative risk estimates $\hat{C}_\Lambda$, and in parentheses the sample standard deviation of $\hat{C}_\Lambda$, the average of bootstrap standard deviation estimates, and the estimated coverage probabilities for nominal 95% confidence intervals for C_Λ given by $\hat{C}_\Lambda \pm 1.96$ (bootstrap standard deviation).

One sees that the covariate-adjusted average relative risk estimators, bootstrap variance estimators and 95% coverage probabilities seem to be fairly accurate under the positively dependent Clayton and Frank models for (T_1, T_2). The slight overestimation by $\hat{C}_\Lambda$ under the Clayton model can be mostly corrected by increasing sample size to $K = 200$ (not shown). As was the case without covariates (Fan et al. (2000a)), $\hat{C}_\Lambda$ appears to be positively biased at smaller values of (t_1, t_2) and negatively biased at larger values of (t_1, t_2) if T_1 and T_2 are independent. The inadequacy of the asymptotic approximation to the distribution of $\hat{C}_\Lambda$

under independence seems to be related to the small number of double failures in $[0, t_1] \times [0, t_2]$ for small t_1 and t_2 and to the high probability of empty risk set at large t_1 and t_2, as discussed in detail in Fan et al. (2000a). The performance of $\hat{C}_\Lambda$ under independence can be improved by increasing sample size. Note that continuously distributed covariates were used in these simulations. Hence, the simulation results seem to suggest that the Theorem 1 and 2 condition that the covariates take values only from a finite set may be able to be relaxed.

A logarithmic transformation to $\hat{C}_\Lambda$ seems beneficial for enhancing asymptotic distributional approximations at small (t_1, t_2) values. For example at $K = 100$ and T_1 and T_2 independent, the sample mean of $\log \hat{C}_\Lambda(0.2877, 0.2877)$, 0.2877 being the 25th percentile of the distribution for $\Lambda_1(T_1)$ and $\Lambda_2(T_2)$, was 0.02, in agreement with the theoretical value of 0.00; the sample standard deviation was 0.55, in agreement with the bootstrap standard deviation estimate of 0.55; and the estimated 95% confidence interval coverage probability was 0.940, based on 200 simulation runs and 50 bootstrap samples. As a "rule of thumb" based on our simulation studies, the logarithmic transformation may lead to useful improvements to asymptotic approximations when the number of doubly uncensored failures in the integration region is under ten.

Simulations were also carried out with binary covariates when there was no covariate effect on the marginal survivor functions, in which case the true regression coefficient β_0 was specified as 0. The sample means and variances (not shown) for $\hat{C}_\Lambda$ when the regression coefficient β_0 was estimated from the data, were comparable to the corresponding sample means and variances (Fan et al. (2000a), Table 1) for $\hat{C}$, the corresponding non-regression dependence estimate, under all three configurations.

As discussed in Section 1, an important feature of our approach is that it can quantify association between two failure times for any region in the support of the observed failure times, while most previous approaches only estimate a single association parameter. For comparison, the association between T_1 and T_2 was also estimated in simulations, assuming that they follow the Clayton model (Clayton (1978), Clayton and Cuzick (1985))

$$F(t_1, t_2) = \{e^{-t_1(1-\theta)} + e^{-t_2(1-\theta)} - 1\}^{1/(1-\theta)}. \tag{9}$$

The concordance estimator (Oakes (1986)) was used for the estimation of the association parameter in (9). In particular, let (T_{i1}, T_{i2}) and (T_{j1}, T_{j2}) be the underlying failure times for the ith and jth pairs and let $\Delta_{ij} = \text{sign}(T_{i1} - T_{j1})(T_{i2} - T_{j2})$, i.e., $\Delta_{ij} = 1$ for concordant pairs and 0 for discordant pairs. The concordance estimator of the association parameter in (9) can be written as

$$\frac{\sum_{i<j} \Delta_{ij} Z_{ij}}{\sum_{i<j} (1 - \Delta_{ij}) Z_{ij}}, \tag{10}$$

where $Z_{ij} = 1$ if the sign of $(T_{i1} - T_{j1})(T_{i2} - T_{j2})$ can be determined by the observed data and 0 otherwise.

The sample mean and sample standard deviation of estimator (10) were calculated in the same simulations as in Tables 1 and 2. The results corresponding to Table 1 are 1.00 (0.217), 0.26 (0.060) and 0.25 (0.051), respectively, under independence, Clayton and Frank family models. Hence, the estimates seem to be accurate when the assumed model is correct, with somewhat smaller standard deviations than our nonparametric estimators. However, when the true underlying model is not of the Clayton model form (9), one arrives at some kind of average of the true cross ratios which typically will depend on follow-up durations. Moreover, the association parameter estimate will depend on the censoring distribution when the data are not from the Clayton's model, as is evident from simulation results based on the same data as for Table 2. Specifically, the sample means and sample standard deviations are 0.51 (0.048), 0.60 (0.068) and 0.65 (0.097), respectively, under our no, light and heavy censoring configurations.

Table 2. Simulation summary statistics for $\hat{\mathcal{T}}_\Lambda$, the covariate-adjusted finite region version of Kendall's τ, at various percentiles of the unit exponential marginal distributions for $\Lambda_1(T_1)$ and $\Lambda_2(T_2)$. Also shown are average of bootstrap estimates of the standard deviation of $\hat{\mathcal{T}}_\Lambda$, based on 50 bootstrap samples, and corresponding estimated 95% confidence interval coverage probabilities. Each entry is based on 200 simulation runs of $K = 100$ pairs of failure times, with time-independent exponentially distributed covariates having means of 0.2 and 0.8. Summary statistics are shown for (T_1, T_2) from a Frank family model under three censoring conditions: no censoring, "light" censoring in which T_1 and T_2 have a probability of 1/3 of being censored, and "heavy" censoring under which these marginal censoring probabilities are 2/3.

$\Lambda_1(T_1)$		$\Lambda_2(T_2)$ Percentile					
Percentile	Censoring	25		50		75	
25	No	0.67	0.65 (0.076,0.087,0.955)*	0.65	0.63 (0.071,0.073,0.940)	0.64	0.62 (0.072,0.072,0.955)
	Light		0.64 (0.088,0.098,0.965)		0.62 (0.085,0.082,0.930)		0.62 (0.081,0.080,0.945)
	Heavy		0.63 (0.111,0.129,0.985)		0.61 (0.115,0.114,0.970)		0.60 (0.112,0.114,0.965)
50	No	0.65	0.64 (0.066,0.072,0.945)	0.61	0.60 (0.057,0.056,0.940)	0.59	0.58 (0.056,0.052,0.920)
	Light		0.63 (0.079,0.081,0.960)		0.60 (0.069,0.065,0.920)		0.58 (0.064,0.062,0.930)
	Heavy		0.63 (0.096,0.110,0.975)		0.59 (0.094,0.103,0.970)		0.57 (0.093,0.107,0.970)
75	No	0.64	0.63 (0.068,0.072,0.940)	0.59	0.58 (0.055,0.054,0.935)	0.56	0.55 (0.049,0.046,0.915)
	Light		0.62 (0.083,0.080,0.930)		0.58 (0.070,0.063,0.885)		0.55 (0.061,0.056,0.910)
	Heavy		0.62 (0.096,0.108,0.970)		0.57 (0.093,0.105,0.970)		0.55 (0.092,0.108,0.975)

* Entries are the theoretical value $\mathcal{T}_\Lambda$, the average of $\hat{\mathcal{T}}_\Lambda$ values, and in parentheses the sample standard deviation of $\hat{\mathcal{T}}_\Lambda$, the average of bootstrap standard deviation estimates, and the estimated coverage probabilities for nominal 95% confidence intervals for $\mathcal{T}_\Lambda$ given by $\hat{\mathcal{T}}_\Lambda \pm 1.96$ (bootstrap standard deviation).

To evaluate the covariate-adjusted finite region version of Kendall's τ, failure times were generated according to the extended Frank model as described earlier. Three censoring schemes were used, namely, no censoring; light censoring as above with independent censoring variates arising from an exponential regression model giving marginal censoring probabilities of 1/3; and heavier censoring with independent exponential regression censoring variates giving a censoring rate of 2/3 on each margin. Table 2 shows summary statistics based on 200 simulation runs of $K = 100$ pairs of failure times. The theoretical value of finite region concordance measure $\mathcal{T}_\Lambda(t_1, t_2)$ decreases as a function of t_1 or t_2 from 0.67 at the 25th percentiles of the $\Lambda_1(T_1)$ and $\Lambda_2(T_2)$ distributions to 0.56 at the 75th percentiles of the marginal distributions of $\Lambda_1(T_1)$ and $\Lambda_2(T_2)$, and further to the usual Kendall's τ value of 0.52 as t_1 and t_2 approach infinity. From Table 2, the asymptotic approximations appear to be adequate under these sampling configurations.

5.2. Comparison between $\hat{C}$ and $\hat{C}_\Lambda$

Some additional simulation studies were carried out to compare the relative risk summary estimate $\hat{C}$, that does not accommodate regression effects on marginal hazard rates, to $\hat{C}_\Lambda$ proposed here, that does make such accommodation.

First, independent unit exponential failure times (T_1, T_2) were generated with a single covariate affecting Cox model marginal hazards as in Section 5.1. The censoring times were also independent exponential variates with covariate effects incorporated as above so that the marginal censoring probabilities are each one-third. The regression variable was taken to be common within a pair, so that $Z_1 = Z_2$, and Z_1 was sampled from a uniform $(0, 3)$ distribution with corresponding regression coefficients $\beta_{01} = \beta_{02} = 1$. From Table 3 we wee that $\hat{C}$ estimates average about 0.5 implying a positive dependence between T_1 and T_2 when the marginal relationship between T_1 and T_2 is considered without conditioning on Z. Also in Table 3 we see that $\hat{C}_\Lambda$ estimates average about 1.0 (independence) indicating that accommodating the regression effects on marginal hazard rates can yield the more insightful inference that the regression variables appear to affect the marginal hazards only, with correspondng cumulative hazard variates approximately independent.

A second simulation was carried out under the Clayton model with $\theta = 2$, so that $C_\Lambda \equiv 0.5$, with Z_1 and Z_2 independent uniform $(0, 3)$ variates and with other specifications as before. From the lower portion of Table 3 one sees that $\hat{C}_\Lambda$ values average about 0.5 while $\hat{C}$ averages are in the 0.60 to 0.65 range, indicating that a stronger dependence can be identified between cumulative hazard variates, conditional on covariates, than can be identified between T_1 and T_2 unconditionally.

Table 3. Simulation summary statistics for $\hat{C}_\Lambda$ and $\hat{C}$, the relative risk summary estimates that do and do not adjust for covariate effects on marginal hazard rates, at various percentiles of the unit exponential marginal distributions for $\Lambda_1(T_1)$ and $\Lambda_2(T_2)$. Each entry is based on 200 simulation runs of $K = 100$ pairs of failure times, with independent exponential regression censoring variates giving a marginal censoring rate of 1/3. Summary statistics are shown for (T_1, T_2) under independence, in which case the covariate is common within a pair and is from a uniform $(0, 3)$ distribution; and for (T_1, T_2) from the Clayton model with positive dependence, in which case the covariates are independent uniform $(0, 3)$ variates.

	$T_1, \Lambda_1(T_1)$		$T_2, \Lambda_2(T_2)$ Percentile					
Scheme	Percentile	Estimator	25		50		75	
Independence	25	$\hat{C}$	1.00	0.53 (0.10)*	1.00	0.52 (0.10)	1.00	0.51 (0.09)
		$\hat{C}_\Lambda$		1.22 (1.03)		1.02 (0.36)		0.96 (0.28)
	50	$\hat{C}$	1.00	0.53 (0.10)	1.00	0.52 (0.09)	1.00	0.50 (0.08)
		$\hat{C}_\Lambda$		1.05 (0.37)		0.99 (0.25)		0.95 (0.22)
	75	$\hat{C}$	1.00	0.51 (0.10)	1.00	0.50 (0.08)	1.00	0.47 (0.08)
		$\hat{C}_\Lambda$		1.01 (0.31)		0.96 (0.22)		0.92 (0.17)
Positive dependence	25	$\hat{C}$	0.50	0.63 (0.19)	0.50	0.65 (0.13)	0.50	0.64 (0.14)
		$\hat{C}_\Lambda$		0.55 (0.30)		0.54 (0.20)		0.53 (0.17)
	50	$\hat{C}$	0.50	0.63 (0.18)	0.50	0.65 (0.14)	0.50	0.63 (0.13)
		$\hat{C}_\Lambda$		0.54 (0.21)		0.53 (0.14)		0.54 (0.12)
	75	$\hat{C}$	0.50	0.63 (0.19)	0.50	0.65 (0.14)	0.50	0.62 (0.11)
		$\hat{C}_\Lambda$		0.53 (0.17)		0.52 (0.12)		0.53 (0.11) Γ

$*$ Entries are the theoretical value C (or C_Λ), the sample mean and, in parentheses, sample variance of $\hat{C}$ (or $\hat{C}_\Lambda$).

6. Discussion

The assumption that the joint distribution of marginal cumulative hazard variates is independent of covariates could be relaxed by allowing $\Lambda_1(T_1)$ and $\Lambda_2(T_2)$ to have a distribution that is stratified, with strata defined by the Z_1 and Z_2 values. Provided the number of strata is finite and the stratum probabilities are nonzero the Appendix arguments could be applied to each stratum, and a summary dependence estimator could be formed as a weighted linear combination of the stratum-specific estimators. This stratification approach can be expected to work well when the sample size is large enough to allow reliable estimation in each stratum.

Acknowledgement

The authors would like to thank Li Hsu for helpful related discussions and Jon A. Wellner for help with the proof in Appendix A.

Appendix A

Lemma. *Suppose that the assumptions in the statement of Theorem* 1 *hold, so that* $(T_{k1}, T_{k2}, U_{k1}, U_{k2}, Z'_{k1}, Z'_{k2})'$, $k = 1, \cdots, K$, *are independent and identically distributed, while the covariates* Z_{ki} *are time-independent and take values only in a finite set* $\{c_1, \cdots, c_d\}$. *Then*

$$\begin{aligned} &K^{-\frac{1}{2}} \sum_{k=1}^{K} I\{e^{\hat{\beta}' Z_{k1}} \hat{\Lambda}_{01}(T_{k1}) \geq s_1, e^{\hat{\beta}' Z_{k2}} \hat{\Lambda}_{02}(T_{k2}) \geq s_2, \\ &\qquad\qquad e^{\hat{\beta}' Z_{k1}} \hat{\Lambda}_{01}(U_{k1}) \geq s_1, e^{\hat{\beta}' Z_{k2}} \hat{\Lambda}_{02}(U_{k2}) \geq s_2\} \\ &- K^{-\frac{1}{2}} \sum_{k=1}^{K} I\{e^{\beta_0' Z_{k1}} \Lambda_{01}(T_{k1}) \geq s_1, e^{\beta_0' Z_{k2}} \Lambda_{02}(T_{k2}) \geq s_2, \\ &\qquad\qquad e^{\beta_0' Z_{k1}} \Lambda_{01}(U_{k1}) \geq s_1, e^{\beta_0' Z_{k2}} \Lambda_{02}(U_{k2}) \geq s_2\} \to_p 0 \end{aligned} \tag{A.1}$$

as $K \to \infty$ *uniformly for* $s_1, s_2 \in \mathcal{R}$.

Proof. Many subsequent citations and notation are from van der Vaart and Wellner (1996), hereafter, VW. We deal with the indicator function with only one of the four inequalities first. Let $\mathcal{M}_i = \{1_{[h(x_i;\beta,\Lambda_i) \geq s]} : s \in \mathcal{R}, \| \beta - \beta_0 \| \leq \eta, \| \Lambda_i - \Lambda_{0i} \| \leq \eta\}$, where $x_i = (t_i, z_i)$ or (u_i, z_i), and $h(x_i; \beta, \Lambda_i) = e^{\beta^t z_i} \Lambda_i(t_i)$. Note that the $\hat{\Lambda}_{0i}(t)$ in (A.1) is uniformly consistent for $\Lambda_{0i}(t)$ for t in the support of $\{X_{ki}\}$ and $i = 1, 2$. We show that $\mathcal{M}_i$ is a Donsker class.

For ease of notation, we omit all the i's in the function h for the moment. Also set $f(z) = e^{\beta^t z}$, $g(t) = \Lambda(t)$, $\mathcal{F} = \{f : \| \beta - \beta_0 \| \leq \eta\}$ and $\mathcal{G} = \{g : \| \Lambda - \Lambda_0 \| \leq \eta\}$. Note that $\mathcal{G}$ is a VC-major class since it is a set of bounded monotone functions from $\mathcal{R}$ to $\mathcal{R}$. We claim that $\mathcal{F} \cdot \mathcal{G} = \{fg : f \in \mathcal{F}, g \in \mathcal{G}\}$ is VC-major, that is, the sets $\{(t, z) : f(z)g(t) > r\}$ with f ranging over $\mathcal{F}$, g over $\mathcal{G}$ and r over $\mathcal{R}$ form a VC-class of sets $\{(t, z) : f(z)g(t) > r\} = \bigcup_{i=1}^{d} \{(t, c_i) : f(c_i)g(t) > r\}$. For each fixed f and c_i, the sign of $f(c_i)$ is fixed: it is greater than, equal to, or less than zero. The sets $\{(t, c_i) : f(c_i)g(t) > r\}$ with f ranging over $\mathcal{F}$, g over $\mathcal{G}$ and r over $\mathcal{R}$ form a VC-class in each case. For example, if $f(c_i) < 0$, then $\{(t, c_i) : f(c_i)g(t) > r\} = \{(t, c_i) : g(t) < \frac{r}{f(c_i)}\}$. The sets $\{(t, c_i) : g(t) \geq \frac{r}{f(c_i)}\}$ with f ranging over $\mathcal{F}$, g over $\mathcal{G}$ and r over $\mathcal{R}$ form a VC-class, since $\mathcal{G}$ is a VC-major class. The sets $\{(t, c_i) : g(t) < \frac{r}{f(c_i)}\}$ are the complements of $\{(t, c_i) : g(t) \geq \frac{r}{f(c_i)}\}$ and hence form a VC-class. Finally, $\{(t, z) : f(z)g(t) > r\}$ with f ranging over $\mathcal{F}$, g over $\mathcal{G}$ and r over $\mathcal{R}$ form a VC-class of sets following Lemma 2.6.17 (iii) (VW, p.147).

By Lemma 2.6.19 (VW, p.148), the class of functions $\mathcal{M}_i$ is VC-major. Since a bounded VC-major class is a VC-hull class, $\mathcal{M}_i$ is Donsker by Corollary 2.6.12 (VW, p.145). The fact that $\mathcal{M}_i$ are uniformly bounded gives that

$$\{1_{[e^{\beta^t z_1} \Lambda_1(t_1) \geq s_1, e^{\beta^t z_2} \Lambda_2(t_2) \geq s_2, e^{\beta^t z_1} \Lambda_1(u_1) \geq s_1, e^{\beta^t z_2} \Lambda_2(u_2) \geq s_2]} :$$

$$s_1, s_2 \in \mathcal{R}, \quad \| \beta - \beta_0 \| \leq \eta, \quad \| \Lambda_i - \Lambda_{0i} \| \leq \eta, \quad i = 1, 2\}$$

forms a Donsker class.

Using Lemma 3.3.5 of van der Vaart and Wellner (p.311), the lemma is established if $\sup_{s_1,s_2\in\mathcal{R}} P(\psi_{\beta,\Lambda_1,\Lambda_2,s_1,s_2} - \psi_{\beta_0,\Lambda_{01},\Lambda_{02},s_1,s_2})^2 \to 0$ as $\beta \to \beta_0$, $\Lambda_i \to \Lambda_{0i}$, $i = 1, 2$, where $Qf \equiv \int f dQ$ for a given measurable function f and signed measure Q, and $\psi_{\beta,\Lambda_1,\Lambda_2,s_1,s_2} = \prod_{i=1}^{2} 1_{[e^{\beta^t Z_i}\Lambda_i(T_i)\geq s_i]} 1_{[e^{\beta^t Z_i}\Lambda_i(U_i)\geq s_i]}$. This is true as can be seen in the following simplified case: for all s, $P(1_{[h(X,\beta,\Lambda)\geq s]} - 1_{[h(X,\beta_0,\Lambda_0)\geq s]})^2 = P\,(\,1_{[h(X,\beta,\Lambda)\geq s]}\, 1_{[h(X,\beta_0,\Lambda_0)<s]}\,) + P\,(\,1_{[h(X,\beta,\Lambda)<s]}\, 1_{[h(X,\beta_0,\Lambda_0)\geq s]}\,) \to 0$ as $\beta \to \beta_0$, $\Lambda_i \to \Lambda_{0i}$, $i = 1, 2$. The corresponding proof for the product of the four indicator variables proceeds in the same way.

Appendix B. Weak Convergence of $\hat{C}_\Lambda$ and $C_\Lambda^{\#}$

First, we want to show $\sqrt{K}(\hat{H}_\Lambda - H_\Lambda)$ is asymptotically equivalent to a uniformly Hadamard differentiable functional of empirical processes under the conditions of Theorem 1.

Let us consider $\sqrt{K}(\hat{H}_\Lambda^{11} - H_\Lambda^{11})$ first.

$$\sqrt{K}\{\hat{H}_\Lambda^{11}(t_1,t_2) - H_\Lambda^{11}(t_1,t_2)\}$$

$$= \sqrt{K}\int_0^{t_1}\int_0^{t_2} \frac{1}{W_k(s_1,s_2)}\Bigg[\frac{1}{K}\sum_{k=1}^{K} I\{\hat{\Lambda}_{k1}(T_{k1}) = s_1,\ \hat{\Lambda}_{k2}(T_{k2}) = s_2,\ \hat{\Lambda}_{k1}(U_{k1}) \geq s_1,$$

$$\hat{\Lambda}_{k2}(U_{k2}) \geq s_2\} - P\{\Lambda_1(T_1) = s_1,\ \Lambda_2(T_2) = s_2,\ \Lambda_1(U_1) \geq s_1,\ \Lambda_2(U_2) \geq s_2\}\Bigg] \quad \text{(B.1)}$$

$$+ \sqrt{K}\int_0^{t_1}\int_0^{t_2} \frac{P\{\Lambda_1(T_1) = s_1, \Lambda_2(T_2) = s_2, \Lambda_1(U_1) \geq s_1, \Lambda_2(U_2) \geq s_2\}}{W_k W(s_1,s_2)}$$

$$\times\{Wk(s_1,s_2)\}, \quad \text{(B.2)}$$

where $\hat{\Lambda}_{ki}(T_{ki}) = s_i$ in (B.1) is an abbreviation of $\hat{\Lambda}_{ki}(T_{ki}) \in [s_i, s_i + ds_i)$, $i = 1, 2$, and similarly for $\Lambda_i(T_i) = s_i$ in (B.1) and (B.2), and where

$$W_k = \frac{1}{K}\sum_{k=1}^{K} I\{\hat{\Lambda}_{k1}(T_{k1}) \geq s_1,\ \hat{\Lambda}_{k2}(T_{k2}) \geq s_2,\ \hat{\Lambda}_{k1}(U_{k1}) \geq s_1,\ \hat{\Lambda}_{k2}(U_{k2}) \geq s_2\},$$

$$W = P\{\Lambda_1(T_1) \geq s_1,\ \Lambda_2(T_2) \geq s_2,\ \Lambda_1(U_1) \geq s_1,\ \Lambda_2(U_2) \geq s_2\}.$$

Term (B.1) is asymptotically equivalent to

$$\sqrt{K}\int_0^{t_1}\int_0^{t_2} \frac{1}{W(s_1,s_2)}\Bigg[\frac{1}{K}\sum_{k=1}^{K} I\{\hat{\Lambda}_{k1}(T_{k1}) = s_1,\ \hat{\Lambda}_{k2}(T_{k2}) = s_2,\ \hat{\Lambda}_{k1}(U_{k1}) \geq s_1,$$

$$\hat{\Lambda}_{k2}(U_{k2}) \geq s_2\} - P\{\Lambda_1(T_1) = s_1,\ \Lambda_2(T_2) = s_2,\ \Lambda_1(U_1) \geq s_1,\ \Lambda_2(U_2) \geq s_2\}\Bigg] \quad \text{(B.3)}$$

since the difference between these two terms (B.1) and (B.3) is

$$\sqrt{K}\int_0^{t_1}\int_0^{t_2}\{\frac{1}{W_k(s_1,s_2)}-\frac{1}{W(s_1,s_2)}\}$$
$$\times\Bigg[\frac{1}{K}\sum_{k=1}^{K}I\{\hat\Lambda_{k1}(T_{k1})=s_1,\ \hat\Lambda_{k2}(T_{k2})=s_2,\ \hat\Lambda_{k1}(U_{k1})\geq s_1,\ \hat\Lambda_{k2}(U_{k2})\geq s_2\}$$
$$-P\{\Lambda_1(T_1)=s_1,\Lambda_2(T_2)=s_2,\ \Lambda_1(U_1)\geq s_1,\ \Lambda_2(U_2)\geq s_2\}\Bigg],$$

which converges in probability to zero following the Continuous Mapping Theorem since $W_k^{-1}-W^{-1}\to_p 0$ uniformly by the lemma in Appendix A (hereafter, Lemma) and the Glivenko-Cantelli Theorem, and since

$$\sqrt{K}\Bigg[\frac{1}{K}\sum_{k=1}^{K}I\{\hat\Lambda_{k1}(T_{k1})\geq s_1,\hat\Lambda_{k2}(T_{k2})\geq s_2,\hat\Lambda_{k1}(U_{k1})\geq s_1,\hat\Lambda_{k2}(U_{k2})\geq s_2\}$$
$$-P\{\Lambda_1(T_1)\geq s_1,\Lambda_2(T_2)\geq s_2,\Lambda_1(U_1)\geq s_1,\Lambda_2(U_2)\geq s_2\}\Bigg]$$
$$\overset{a}{=}\sqrt{K}\Bigg[\frac{1}{K}\sum_{k=1}^{K}I\{\Lambda_{k1}(T_{k1})\geq s_1,\Lambda_{k2}(T_{k2})\geq s_2,\Lambda_{k1}(U_{k1})\geq s_1,\Lambda_{k2}(U_{k2})\geq s_2\}$$
$$-P\{\Lambda_1(T_1)\geq s_1,\Lambda_2(T_2)\geq s_2,\Lambda_1(U_1)\geq s_1,\Lambda_2(U_2)\geq s_2\}\Bigg]\quad\text{by Lemma}$$
$$\Rightarrow Z$$

where $\overset{a}{=}$ denotes "asymptotic equivalence" and Z is a Gaussian process.

Term (B.3) is in turn asymptotically equivalent to

$$\sqrt{K}\int_0^{t_1}\int_0^{t_2}\frac{1}{W(s_1,s_2)}\Bigg[\frac{1}{K}\sum_{k=1}^{K}I\{\Lambda_{k1}(T_{k1})=s_1,\ \Lambda_{k2}(T_{k2})=s_2,\ \Lambda_{k1}(U_{k1})\geq s_1,$$
$$\Lambda_{k2}(U_{k2})\geq s_2\}-P\{\Lambda_1(T_1)=s_1,\ \Lambda_2(T_2)=s_2,\ \Lambda_1(U_1)\geq s_1,\ \Lambda_2(U_2)\geq s_2\}],\tag{B.4}$$

which is a sum of i.i.d. random variables.

By similar arguments, term (B.2) is asymptotically equivalent to

$$-\sqrt{K}\int_0^{t_1}\int_0^{t_2}\frac{P\{\Lambda_1(T_1)=s_1,\Lambda_2(T_2)=s_2,\Lambda_1(U_1)\geq s_1,\Lambda_2(U_2)\geq s_2\}}{W^2(s_1,s_2)}$$
$$\times[\frac{1}{K}\sum_{k=1}^{K}I\{\Lambda_{k1}(T_{k1})\geq s_1,\Lambda_{k2}(T_{k2})\geq s_2,\Lambda_{k1}(U_{k1})\geq s_1,\Lambda_{k2}(U_{k2})\geq s_2\}$$
$$-P\{\Lambda_1(T_1)\geq s_1,\Lambda_2(T_2)\geq s_2,\Lambda_1(U_1)\geq s_1,\Lambda_2(U_2)\geq s_2\}],\tag{B.5}$$

a sum of i.i.d. random variables.

The asymptotic equivalence between $\sqrt{K}(\hat{H}_\Lambda^{10} - H_\Lambda^{10})$, $\sqrt{K}(\hat{H}_\Lambda^{01} - H_\Lambda^{01})$ and sums of i.i.d. random variables can be established in the same way.

Define

$$B^{11}(t_1,t_2) = \int_0^{t_1}\int_0^{t_2} P\{\Lambda_1(T_1) = s_1, \Lambda_2(T_2) = s_2, \Lambda_1(U_1) \geq s_1, \Lambda_2(U_2) \geq s_2\},$$
$$B^{10}(t_1,t_2) = \int_0^{t_1} P\{\Lambda_1(T_1) = s_1, \Lambda_2(T_2) \geq t_2, \Lambda_1(U_1) \geq s_1, \Lambda_2(U_2) \geq t_2\},$$
$$B^{01}(t_1,t_2) = \int_0^{t_2} P\{\Lambda_1(T_1) \geq t_1, \Lambda_2(T_2) = s_2, \Lambda_1(U_1) \geq t_1, \Lambda_2(U_2) \geq s_2\},$$
$$B^{00}(t_1,t_2) = P\{\Lambda_1(T_1) \geq t_1, \Lambda_2(T_2) \geq t_2, \Lambda_1(U_1) \geq t_1, \Lambda_2(U_2) \geq t_2\}.$$

Let $B = (B^{11}, B^{10}, B^{01}, B^{00})$. Denote the empirical distribution of B by $\hat{B}_K = (\hat{B}_K^{11}, \hat{B}_K^{10}, \hat{B}_K^{01}, \hat{B}_K^{00})$, with $\hat{B}_K^{11}(t_1,t_2) = \int_0^{t_1}\int_0^{t_2} \frac{1}{K}\sum_{k=1}^K I\{\hat{\Lambda}_{k1}(T_{k1}) = s_1, \hat{\Lambda}_{k2}(T_{k2}) = s_2, \hat{\Lambda}_{k1}(U_{k1}) \geq s_1, \hat{\Lambda}_{k2}(U_{k2}) \geq s_2\}$, where $\hat{\Lambda}_{k1}(T_{k1}) = e^{\hat{\beta}' Z_{k1}}\hat{\Lambda}_{01}(T_{k1})$, etc.

After re-writing the i.i.d. sums in terms of B and $\hat{B}_K$, we get

$$\begin{aligned}
\sqrt{K}(\hat{H}_\Lambda^{11} - H_\Lambda^{11}) &\overset{a}{=} \int_0^{t_1}\int_0^{t_2} \frac{1}{W(s_1,s_2)}\sqrt{K}\{\hat{B}_K^{11}(ds_1,ds_2) - B^{11}(ds_1,ds_2)\} \\
&\quad - \int_0^{t_1}\int_0^{t_2} \frac{B^{11}(ds_1,ds_2)}{W^2(s_1,s_2)}\sqrt{K}\{\hat{B}_K^{00} - B^{00}(s_1,s_2)\}, \\
\sqrt{K}(\hat{H}_\Lambda^{10} - H_\Lambda^{10}) &\overset{a}{=} \int_0^{t_1} \frac{1}{W(s_1,t_2)}\sqrt{K}\{\hat{B}_K^{10}(ds_1,t_2) - B^{10}(ds_1,t_2)\} \\
&\quad - \int_0^{t_1} \frac{B^{10}(ds_1,t_2)}{W^2(s_1,t_2)}\sqrt{K}\{\hat{B}_K^{00}(s_1,t_2) - B^{00}(s_1,t_2)\}, \\
\sqrt{K}(\hat{H}_\Lambda^{01} - H_\Lambda^{01}) &\overset{a}{=} \int_0^{t_2} \frac{1}{W(t_1,s_2)}\sqrt{K}\{\hat{B}_K^{01}(t_1,ds_2) - B^{01}(t_1,ds_2)\} \\
&\quad - \int_0^{t_2} \frac{B^{01}(t_1,ds_2)}{W^2(t_1,s_2)}\sqrt{K}\{\hat{B}_K^{00}(t_1,s_2) - B^{00}(t_1,s_2)\}.
\end{aligned}$$

Thus we have established that $\sqrt{K}(\hat{H}_\Lambda - H_\Lambda)$ is asymptotically equivalent to a functional, say φ, of empirical process $\sqrt{K}(\hat{B}_K - B)$, where φ is defined as above. The functional φ is uniformly Hadamard differentiable by the differentiability of $W \mapsto 1/W$ when $W > 0$ uniformly on $[\delta,\gamma]$ (this is guaranteed by condition (8) and by Lemma 5.1 of Gill et al. (1995)). The weak convergence of $\hat{C}_\Lambda$ and $C_\Lambda^{\#}$ follows from the weak convergence of empirical processes and by applying the functional delta-method (e.g., Gill et al. (1995), van der Vaart and Wellner (1996)) first to φ, then to Φ.

References

Breslow, N. E. (1972). Contribution to Discussion of Paper by D. R. Cox. *J. Roy. Statist. Soc. Ser. B* **34**, 216-217.

Breslow, N. E. (1974). Covariance Analysis of Censored Survival Data. *Biometrics* **30**, 89-99.

Cai, J. and Prentice, R. L. (1995). Estimating equations for hazard ratio parameters based on correlated failure time data. *Biometrika* **82**, 151-164.

Clayton, D. G. (1978). A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika* **65**, 141-151.

Clayton, D. G. and Cuzick, J. (1985). Multivariate Generalizations of the Proportional Hazards Model (with discussion). *J. Roy. Statist. Soc. Ser. A* **148**, 82-117.

Cox, D. R. (1972). Regression models and life-tables. *J. Roy. Statist. Soc. Ser. B* **34**, 187-202.

Dabrowska D. M. (1988). Kaplan-Meier estimate on the plane. *Ann. Statist.* **16**, 1475-1489.

Dabrowska D. M. (1989). Kaplan-Meier estimate on the plane: weak convergence, LIL, and the bootstrap. *J. Multivariate Anal.* **29**, 308-325.

Duffy, D. L., Martin, N. G. and Matthews, J. D. (1990). Appendectomy in Australian twins. *Amer. J. Hum. Genetics* **47**, 590-592.

Fan, J. J., Hsu, L. and Prentice, R. L. (2000a). Dependence Estimation over a Finite Bivariate Failure Time Region. *Lifetime Data Anal.* **6**, 343-355.

Fan, J. J., Prentice, R. L. and Hsu, L. (2000b). A Class of Weighted Dependence Measures for Bivariate Failure Time Data. *J. Roy. Statist. Soc. Ser. B* **62**, 181-190.

Genest, C. (1987). Frank's family of bivariate distributions. *Biometrika* **74**, 549-555.

Gill, R. D., van der Laan, M. J. and Wellner, J. A. (1995). Inefficient estimators of the bivariate survival function for three models. *Ann. Inst. Henri Poincare* **31**, 545-597.

Hsu, L and Prentice, R. L. (1996). On assessing the strength of dependency between failure time variates. *Biometrika* **83**,491-506.

Kendall, M. G. (1938). A new measure of rank correlation. *Biometrika* **30**, 81-93.

Lee, E. W., Wei, L. J. and Amato, D. A. (1992). Cox-type regression analysis for large numbers of small groups of correlated failure time observations. In *Survival Analysis: State of the Art* (Edited by J. P. Klein and P. K. Goel), 237-247. Kluwer Academic Publishers.

Oakes, D. (1982). A model for association in bivariate survival data. *J. Roy. Statist. Soc. Ser. B* **44**, 414-422.

Oakes, D. (1986). Semiparametric inference in a model for association in bivariate survival data. *Biometrika* **73**, 353-361.

Oakes, D. (1989). Bivariate survival models induced by frailties. *J. Amer. Statist. Assoc.* **84**, 487-493.

Prentice, R. L. and Cai, J. (1992). Covariance and survivor function estimation using censored multivariate failure time data. *Biometrika* **79**, 495-512.

Prentice, R. L. and Hsu, L. (1997). Regression on hazard ratios and cross ratios in multivariate failure time analysis. *Biometrika* **84**, 349-363.

van der Vaart, A. W. and Wellner, J. A. (1996). Weak Convergence and Empirical Processes. Springer, New York.

Wei, L. J., Lin, D. Y. and Weissfeld, L. (1989). Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. *J. Amer. Statist. Assoc.* **84**, 1065-1073.

Division of Statistics, University of California at Davis, One Shields Avenue, Davis, CA 95616, U.S.A.

E-mail: jjfan@wald.ucdavis.edu

Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue North, MP-1002, Seattle, WA 98109, U.S.A.

E-mail: rprentic@fhcrc.org

(Received January 2000; accepted November 2001)