Asymptotic Properties of Semiparametric Maximum Likelihood Estimator in Normal Transformation Models for Bivariate Survival Data

(Technical Report)

Yi Li, Ross L. Prentice and Xihong Lin *

Abstract

In this technical report, we resort to the modern empirical process theory to study the asymptotic properties of the semiparametric maximum likelihood estimator in normal transformation models for bivariate survival data. We prove that the semiparametric maximum likelihood estimators exist, are consistent and asymptotically normal, and reach the semiparametric efficiency bound, under the semiparametric normal transformation model.

KEY WORDS: Bivariate Failure Time; Semiparametric Normal Transformation; Semiparametric Maximum Likelihood Estimate; Consistency; Asymptotic Normality.

RUNNING TITLE: Semiparametric Normal Transformation Models.

*Yi Li is Associate Professor, Department of Biostatistics, Harvard School of Public Health and Dana-Farber Cancer Institute, Boston, MA 02115; Ross L. Prentice is Member, Fred Hutchinson Cancer Research Center, Seattle, WA 98109, and Professor of Biostatistics, University of Washington, Seattle, WA 98195; Xihong Lin is Professor, Department of Biostatistics, Harvard School of Public Health, Boston, MA 02115. This work was partially supported by U. S. National Cancer Institute grants R01 CA95747 (Li), P01 CA53996 (Prentice), R01 CA76404 (Lin). The first author (Li) also thanks Professor David Harrington for many helpful discussions.

1 Introduction

Consider a survival time pair $(\tilde{T}_1, \tilde{T}_2)$, where each $\tilde{T}_j$ marginally has a cumulative hazard $\Lambda_j(t)$. Then $\Lambda_j(\tilde{T}_j)$ marginally follows a unit exponential distribution, and its probit transformation

$$T_j = \Phi^{-1}\left\{1 - e^{-\Lambda_j(\tilde{T}_j)}\right\} \tag{1}$$

has a standard normal distribution, where $\Phi(\cdot)$ is the CDF for $N(0,1)$.

To specify the correlation structure within the survival time pair $(\tilde{T}_1, \tilde{T}_2)$, we assume that the normally transformed survival time pair (T_1, T_2) is jointly normally distributed with correlation coefficient ρ and with a joint tail probability function

$$\Psi(z_1, z_2; \rho) = \int_{z_1}^{\infty}\int_{z_2}^{\infty} \phi(x_1, x_2; \rho) dx_1 dx_2 \tag{2}$$

where $\phi(x_1, x_2; \rho)$ is the pdf for a bivariate normal vector with mean $(0,0)$ and covariance matrix $\begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$. It follows that the bivariate survival function for the original survival time pair $(\tilde{T}_1, \tilde{T}_2)$ is

$$S(\tilde{t}_1, \tilde{t}_2; \rho) = P(\tilde{T}_1 > \tilde{t}_1, \tilde{T}_2 > \tilde{t}_2; \rho) = \Psi[\Phi^{-1}\{F_1(\tilde{t}_1)\}, \Phi^{-1}\{F_2(\tilde{t}_2)\}; \rho] \tag{3}$$

where $F_j(\cdot)$ are the marginal CDFs of $\tilde{T}_j (j = 1, 2)$ respectively. Moreover, the density for the original survival time pair $(\tilde{T}_1, \tilde{T}_2)$ is

$$f(\tilde{t}_1, \tilde{t}_2; \rho) = f_1(\tilde{t}_1) f_2(\tilde{t}_2) e^{g(t_1, t_2)} \tag{4}$$

where $t_i = \Phi^{-1}\left\{1 - e^{-\Lambda_i(\tilde{t}_i)}\right\}$, $f_i(\tilde{t}) = \lambda_i(\tilde{t}) \exp\{-\Lambda_i(\tilde{t})\}$ is the marginal density for $\tilde{T}_i, i = 1, 2$ and

$$g(t_1, t_2) = -0.5 \log(1 - \rho^2) - 0.5(1 - \rho^2)^{-1}(\rho^2 t_1^2 + \rho^2 t_2^2 - 2\rho t_1 t_2). \tag{5}$$

Indeed, the correlation parameter ρ provides a summary measure for the pairwise dependence, whose connection with the other commonly used dependence measures, including the cross ratio, Kendall's *tau* and Spearman's *rho*, can be found in Li and Lin (2006).

We consider estimation of the unknown $(\rho, \Lambda_1, \Lambda_2)$ based on a censored sample of m pairs. That is, we estimate the marginal hazard rate and the correlation parameter on the basis of observed pairs $(\tilde{X}_{i1}, \delta_{i1}, \tilde{X}_{i2}, \delta_{i2})$, where $\tilde{X}_{ij} = \tilde{T}_{ij} \wedge \tilde{U}_{ij} \overset{def}{=} min(\tilde{T}_{ij}, \tilde{U}_{ij}), \delta_{ij} = I(\tilde{T}_{ij} \leq \tilde{U}_{ij})$, for $j = 1, 2$. For simplicity, we assume that the censoring mechanism satisfies the usual random censorship, i.e. the censoring pair $(\tilde{U}_{i1}, \tilde{U}_{i2})$ is independent of the survival pair $(\tilde{T}_{i1}, \tilde{T}_{i2})$. Under this random censorship, the likelihood function can be factorized into the product of contributions from the survival and censoring times, facilitating likelihood-based inferential procedures.

In many applications involving bivariate survival data, including studies of disease occurrence patterns of twins or siblings, it is natural to restrict the marginal cumulative hazard to be common for members of the same pair. Hence, we first consider drawing inference with $\Lambda_1 \equiv \Lambda_2 (= \Lambda)$ in Section 2, followed by the case of distinct marginal cumulative hazards $\Lambda_1 \not\equiv \Lambda_2$ in Section 3. We summarize this report in Section 4.

2 Semiparametric Maximum Likelihood Estimation With A Common Marginal Cumulative Hazard

We begin by noting that the likelihood function for the unknown parameters (Λ, ρ), based on the observed data $(\tilde{X}_{ij}, \delta_{ij}), j = 1, 2, i = 1, \ldots, m$, can be written, up to a constant, as the product of factors $(i = 1, \ldots, m)$

$$\begin{aligned} \tilde{L}_i(\rho, \Lambda) &= \{e^{g(X_{i1}, X_{i2})}\Lambda'(\tilde{X}_{i1})\Lambda'(\tilde{X}_{i2})e^{-\Lambda(\tilde{X}_{i1})-\Lambda(\tilde{X}_{i2})}\}^{\delta_{i1}\delta_{i2}}\{\Psi_1(X_{i1}, X_{i2}; \rho)\Lambda'(\tilde{X}_{i1})e^{-\Lambda(\tilde{X}_{i1})}\}^{\delta_{i1}(1-\delta_{i2})} \\ &\quad \times\{\Psi_2(X_{i1}, X_{i2}; \rho)\Lambda'(\tilde{X}_{i2})e^{-\Lambda(\tilde{X}_{i2})}\}^{(1-\delta_{i1})\delta_{i2}} \times \{\Psi(X_{i1}, X_{i2}; \rho)\}^{(1-\delta_{i1})(1-\delta_{i2})}, \end{aligned} \tag{6}$$

where we denote the transformed observed time by $X_{ij} = \Phi^{-1}\{1-\exp(-\Lambda(\tilde{X}_{ij})\}$ for $j = 1, 2$, $\Psi_j(x_1, x_2; \rho) = -\frac{\partial}{\partial x_j}\Psi(x_1, x_2; \rho)/\phi(x_j)$ for $j = 1, 2$. Indeed, $\Psi_j(x_1, x_2; \rho) = P(T_{3-j} \geq x_{3-j}|T_j = x_j)$ for $j = 1, 2$.

Directly maximizing the above likelihood in a space containing continuous hazard $\Lambda(\cdot)$ is not feasible, as one can always let the likelihood go to ∞ by choosing some continuous function $\Lambda(\cdot)$ with fixed values at each $\tilde{X}_{ij}$ while letting $\Lambda'(\cdot)$ go to ∞ at an observed failure time (i.e. at some $\tilde{X}_{ij}$ with $\delta_{ij} = 1$). Thus we need consider the following parameter space for Λ

$$\{\Lambda : \Lambda \,\text{is cadlag and piecewise constant}\}$$

where by cadlag we mean *right continuous with left hand limit*. It follows that the MLE of $\Lambda(\cdot)$ will be the one which jumps only at distinct observed failure times. We denote the jump size of $\Lambda(\cdot)$ at t by $\Delta\Lambda(t) = \Lambda(t) - \Lambda(t-)$. The SPMLE is the maximizer to the empirical likelihood function $L(\rho, \Lambda)$, which is the product of terms (6) with $\Lambda'(\cdot)$ replaced by $\Delta\Lambda(\cdot)$. We denote the log empirical likelihood function by $\ell(\rho, \Lambda) = \log L(\rho, \Lambda)$.

The main results of the paper are proved under the following set of regularity conditions.

(c.1) (Boundness) ρ lies in a known open interval within $[-1, 1]$.

(c.2) (Finite Interval) There exist a $\tau > 0$ and a constant $c_0 > 0$ such that $P(\tilde{U}_{ij} \geq \tau) = P(\tilde{U}_{ij} = \tau) > c_0$. In practice, τ is usually the duration of the study.

(c.3) (Differentiability) Assume the marginal cumulative hazard $\Lambda(t)$ is differentiable and $\Lambda'(t) > 0$ over $[0, \tau]$. Moreover, $\Lambda(\tau) < \infty$.

Under these conditions, we show that the SPMLEs do exist and are finite. Furthermore, we show in the next two Propositions that, the SPMLEs of $\hat{\Lambda}$, stay bounded and that the SPMLEs of $\{\rho, \Lambda(\cdot)\}$ are consistent and asymptotically normal estimators of the true parameters.

We first present two Lemmas which will be used for proving the asymptotics under the common marginal hazard models.

Lemma 1 *Under the common hazard model and with probability 1,*

$$\begin{aligned} Q_i(\rho, \Lambda) &\overset{def}{=} \{e^{g(X_{i1}, X_{i2})} e^{-\Lambda(\tilde{X}_{i1}) - \Lambda(\tilde{X}_{i2})}\}^{\delta_{i1}\delta_{i2}} \{\Psi_1(X_{i1}, X_{i2}; \rho) e^{-\Lambda(\tilde{X}_{i1})}\}^{\delta_{i1}(1-\delta_{i2})} \\ &\quad \times \{\Psi_2(X_{i1}, X_{i2}; \rho) e^{-\Lambda(\tilde{X}_{i2})}\}^{(1-\delta_{i1})\delta_{i2}} \{\Psi(X_{i1}, X_{i2}; \rho)\}^{(1-\delta_{i1})(1-\delta_{i2})} \\ &< O(1) \prod_{j=1}^{2} \{1 + \Lambda(\tilde{X}_{ij})\}^{-(1+\delta_{ij})}. \end{aligned}$$

where O(1) is a finite constant independent of $\rho, \Lambda()$.

Proof: To avoid interruption, we list below several equalities frequently used in the development. For any $p > 0$ when $x \to \infty$.

$$\phi(x) = o(1)x^{-p}, 1 - \Phi(x) = o(1)x^{-p}, \frac{1 - \Phi(x)}{\phi(x)} = O(1)x^{-1}.$$

These equalities can be conveniently obtained using the L'Hopital rule. Moreover, let $x = \Phi^{-1}(1 - e^{-t})$, then applying the L'Hopital rule,

$$x^2/(1 + t) \to C_0 \tag{7}$$

when $t \to \infty$. Here C_0 is a fixed positive constant (Indeed, numerically $C_0 = 1.324\ldots$).

Note

$$e^{g(X_{i1}, X_{i2})} e^{-\Lambda(\tilde{X}_{i1})} e^{-\Lambda(\tilde{X}_{i2})} = \frac{\phi(x_1, x_2; \rho)}{\phi(x_1)\phi(x_2)} \{1 - \Phi(x_1)\}\{1 - \Phi(x_2)\}, \tag{8}$$

where, for notational ease, we denote by $x_j = X_{ij} = \Phi^{-1}\{1 - e^{-\Lambda(\tilde{X}_{ij})}\}$. In view of (7),

$$\frac{1 - \Phi(x_j)}{\phi(x_j)} = O(1)\{1 + \Lambda(\tilde{X}_{ij})\}^{-1/2} \le O(1).$$

By the boundness condition (c.1), we assume that $\rho \in (-1 + \epsilon_0, 1 - \epsilon_0)$, where $0 < \epsilon_0 < 1$ is a known constant. Then

$$\begin{aligned}
\phi(s_1, s_2; \rho) &= \{2\pi(1-\rho^2)^{1/2}\}^{-1} \exp[-\{2(1-\rho^2)\}^{-1}(s_1^2 + s_2^2 - 2\rho s_1 s_2)] \\
&= \{2\pi(1-\rho^2)^{1/2}\}^{-1} \exp(-\{2(1-\rho^2)\}^{-1}[|\rho|\{s_1 - sign(\rho)s_2\}^2 + (1-|\rho|)s_1^2 + (1-|\rho|)s_2^2]) \\
&\le \{2\pi(1-\rho^2)^{1/2}\}^{-1} \exp[-\{2(1-\rho^2)\}^{-1}\{(1-|\rho|)s_1^2 + (1-|\rho|)s_2^2\}] \\
&\le (2\pi\epsilon_0^{1/2})^{-1} \exp(-s_1^2/4)\exp(-s_2^2/4) \\
&= O(1)\{\phi(s_1)\phi(s_2)\}^{1/2}.
\end{aligned}$$

Consider $\phi(x_j) = o(1)x_j^{-8}$. Hence, in view of (7), (8) $\le \prod_j \{1 + \Lambda(\tilde{X}_{ij})\}^{-2}$.

Next consider

$$\Psi_2(x_1, x_2; \rho)e^{-\Lambda(\tilde{X}_{i2})} = -\frac{\partial}{\partial x_2}\Psi(x_1, x_2, \rho)\frac{1 - \Phi(x_2)}{\phi(x_2)}, \qquad (9)$$

where $-\frac{\partial}{\partial x_2}\Psi(x_1, x_2, \rho) = \int_{x_1}^{\infty} \phi(s_1, x_2; \rho)ds_1$. Hence, (9) $=$

$$\left\{\int_{x_1} \exp(-s_1^2/4)ds_1\right\} e^{-x_2^2/4}\frac{1 - \Phi(x_2)}{\phi(x_2)}. \qquad (10)$$

L'Hopital rule will give for any $p > 0$, $\int_{x_1} \exp(-s_1^2/4)ds_1 = o(1)x_1^{-p}$, $e^{-x_2^2/4} = o(1)x_2^{-p}$ and $\frac{\{1-\Phi(x_2)\}}{\phi(x_2)} = O(1)x_2^{-1}$. Hence, in particular, by choosing $p = 4$, (10) $\le o(1)x_1^{-4}x_2^{-4} \le o(1)\{1 + \Lambda(\tilde{X}_{ij})\}^{-2}$. Similarly, we have that

$$\Psi_1(X_{i1}, X_{i2}; \rho)e^{-\Lambda(\tilde{X}_{i1})} \le O(1)\prod_j \{1 + \Lambda(\tilde{X}_{ij})\}^{-2}.$$

Now consider the inequality

$$\Psi(x_1, x_2; \rho) = P(X_1 \ge x_1, X_2 \ge x_2; \rho) \le \min\{P(X_1 \ge x_1), P(X_2 \ge x_2)\} \le \{P(X_1 \ge x_1)P(X_1 \ge x_1)\}^{1/2}.$$

Since $P(X_j \ge x_j) = 1 - \Phi(x_j) = o(1)x_j^{-p}$ $(j = 1, 2)$ for any $p > 0$, we have that $\Psi(x_1, x_2; \rho) \le \prod_j \{1 + \Lambda(\tilde{X}_{ij})\}^{-1}$ by choosing $p = 4$. Thus, combining all the terms, we have that

$$Q_i(\rho, \Lambda) < O(1)\prod_j \{1 + \Lambda(\tilde{X}_{ij})\}^{-\{1+\delta_{i1}+\delta_{i2}-\delta_{i1}\delta_{i2}\}} < O(1)\prod_j \{1 + \Lambda(\tilde{X}_{ij})\}^{-(1+\delta_{ij})}.$$

□

Next we show that the SPMLEs do exist and are finite.

Lemma 2 *(existence of SPMLEs) Denote by $N_{max} = max_i N_i(\tau)$ the maximum number of observed failures for a pair over $[0,\tau]$., where $N_i(\tau) = \sum_j I(\tilde{T}_{ij} \leq \tau, \delta_{ij} = 1)$. If $N_{max} > 1$, then the maximum likelihood estimators of $\ell(\rho,\Lambda)$, $(\hat{\rho},\hat{\Lambda})$ exist and are finite.*

Proof: Let $\ell_i(\rho,\Lambda)$ be the log of (6) with $\Lambda'(\cdot)$ replaced by the jumpsize $\Delta\Lambda(\cdot)$ and $\ell(\rho,\Lambda) = \sum_{i=1}^m \ell_i(\rho,\Lambda)$ be the log empirical likelihood for the observed data. Following Murphy (Theorem 1, 1994), the maximizer of ρ lies inside the parameter space. Moreover, since $\ell(\rho,\Lambda)$ is also a continuous function of jumpsizes, we only need to show that the maximizing jumpsizes are finite.

First note that

$$\begin{aligned}\ell_i(\rho,\Lambda) \quad < \quad & \delta_{i1}\delta_{i2}\{g(X_{i1},X_{i2}) + \log\Delta\Lambda(\tilde{X}_{i1}) + \log\Delta\Lambda(\tilde{X}_{i2}) - \Lambda(\tilde{X}_{i1}) - \Lambda(\tilde{X}_{i2})\} \\ & +\delta_{i1}(1-\delta_{i2})\{\log\Delta\Lambda(\tilde{X}_{i1}) - \Lambda(\tilde{X}_{i1})\} + (1-\delta_{i1})\delta_{i2}\{\log\Delta\Lambda(\tilde{X}_{i2}) - \Lambda(\tilde{X}_{i2})\}\end{aligned}$$

because $0 \leq \Psi_j(X_{i1},X_{i2};\rho) \leq 1$ for $j = 1,2$ and $0 \leq \Psi(X_{i1},X_{i2};\rho) \leq 1$. Also recall $g(t_1,t_2)$ as defined in (5) and note

$$\rho^2 t_1^2 + \rho^2 t_2^2 - 2\rho t_1 t_2 = \rho^2\{t_1 - sign(\rho)t_2\}^2 + 2\{sign(\rho)\rho^2 - \rho\}t_1 t_2.$$

Hence,

$$g(t_1,t_2) < const + \frac{|\rho|}{1+|\rho|}|t_1 t_2| < const + \frac{1}{2}|t_1 t_2|$$

in view of $\rho \in (-1+\epsilon_0, 1-\epsilon_0)$. In particular, *const* is the finite supremum of $-0.5\log(1-\rho^2)$ over $(-1+\epsilon_0, 1-\epsilon_0)$. Therefore,

$$\begin{aligned}\ell_i(\rho,\Lambda) \quad < \quad & const + \frac{1}{2}\delta_{i1}\delta_{i2}|X_{i1}X_{i2}| + \delta_{i1}\{\log\Delta\Lambda(\tilde{X}_{i1}) - \Lambda(\tilde{X}_{i1})\} + \delta_{i2}\{\log\Delta\Lambda(\tilde{X}_{i2}) - \Lambda(\tilde{X}_{i2})\} \\ \leq \quad & const + \sum_{i=1}^m\sum_{j=1}^2 \delta_{ij}\{X_{ij}^2 + \log\Delta\Lambda(\tilde{X}_{ij}) - \Lambda(\tilde{X}_{i1})\}.\end{aligned}$$

wherein the second inequality follows from

$$\delta_{i1}\delta_{i2}|X_{i1}X_{i2}| \leq \frac{1}{2}(\delta_{i1}^2 X_{i1}^2 + \delta_{i2}^2 X_{i2}^2) = \frac{1}{2}(\delta_{i1}X_{i1}^2 + \delta_{i2}X_{i2}^2).$$

Finally, denote by $\lambda_1, \ldots, \lambda_K$ the jumpsizes of Λ at distinct observed failure times $t_1, \ldots, t_K$ over $[0, \tau]$. Since $N_{max} > 1$, it follows that $K > 1$. Then

$$\begin{aligned} \ell(\rho, \Lambda) &\leq O(1) + \sum_{k=1}^{K} n_{(k)} \left[\frac{1}{4} \{\Phi^{-1}(1 - e^{-\sum_{j=1}^{k} \lambda_j})\}^2 + \log \lambda_k - \sum_{j=1}^{k} \lambda_j \right] \\ &\leq O(1) + \sum_{k=1}^{K} n_{(k)} \left\{ \log \lambda_k - (1 - \frac{C_0}{4}) \sum_{j=1}^{k} \lambda_j \right\} \end{aligned}$$

where $n_k \geq 1$ is the number of ties at $t_k, k = 1, \ldots, K$ and the second inequality comes from (7) and C_0 is the constant defined right after (7). Hence $\ell(\rho, \Lambda)$ will diverge to $-\infty$ if $\lambda_j \to \infty$ for some $j \in \{1, \ldots, K\}$, yielding a contradiction. □

Proposition 1 *(Consistency) Denote by* (ρ_0, Λ_0) *the true parameters. Then* $|\hat{\rho} - \rho_0| \to 0$ *and* $\sup_{t \in [0,\tau]} |\hat{\Lambda}(t) - \Lambda_0(t)| \to 0$ *almost surely.*

Proof: Two major steps are involved in this proof. We first show the SPMLEs stay bounded, followed by showing every convergent subsequence converges to the true parameters.

To proceed, denote by $\hat{\rho}, \hat{\Lambda}$ the SPMLEs for the true parameters. Our immediate goal is to show by contradiction that $\hat{\Lambda}_0(\cdot)$ has an upper bound in $[0, \tau]$ with probability one. The contradiction will be established as follows: we construct a step function $\bar{\Lambda}$ which jumps only at distinct observed failure times, ie $\tilde{X}_{ij}$ for which $\delta_{ij} = 1$ such that $\bar{\Lambda}$ will be close to the true function Λ_0. Since $\hat{\rho}, \hat{\Lambda}$ maximize the likelihood, it follows that

$$0 \leq \ell(\hat{\rho}, \hat{\Lambda}) - \ell(\rho_0, \bar{\Lambda}). \tag{11}$$

Then we prove that, if $\hat{\Lambda}(\tau) \to \infty$ the right-hand side of the inequality will be negative, yielding a contradiction.

Specifically, the step function $\bar{\Lambda}$ can be constructed as follows. Differentiating $\ell(\rho, \Lambda)$ with respect to $\Delta\Lambda(\tilde{X}_{ij})$ and setting it to 0 leads to the following equation

$$\frac{\delta_{ij}}{\Delta\Lambda(\tilde{X}_{ij})} = \sum_{k=1}^{m} R_k(\tilde{X}_{ij}; \hat{\rho}, \hat{\Lambda}), \tag{12}$$

where

$$R_k(t; \rho, \Lambda) = R_{k1}(t; \rho, \Lambda) + R_{k2}(t; \rho, \Lambda)$$

$$
\begin{aligned}
= \quad & \Bigg[\Bigg\{\delta_{k1}\delta_{k2}\frac{\rho^2 X_{k1}-\rho X_{k2}}{1-\rho^2}+(1-\delta_{k1})(1-\delta_{k2})\frac{\Psi_1(X_{k1},X_{k2};\rho)\phi(X_{k1})}{\Psi(X_{k1},X_{k2};\rho)} \\
& -\rho\delta_{k1}(1-\delta_{k2})\frac{\phi(X_{k2}-\rho X_{k1};1-\rho^2)}{1-\Phi(X_{k2}-\rho X_{k1};1-\rho^2)}+\delta_{k2}(1-\delta_{k1})\frac{\phi(X_{k1}-\rho X_{k2};1-\rho^2)}{1-\Phi(X_{k1}-\rho X_{k2};1-\rho^2)}\Bigg\} \\
& \times\frac{1-\Phi(X_{k1})}{\phi(X_{k1})}+\delta_{k1}\Bigg] I(\tilde{X}_{k1}>t) \\
& +\Bigg[\Bigg\{\delta_{k1}\delta_{k2}\frac{\rho^2 X_{k2}-\rho X_{k1}}{1-\rho^2}+(1-\delta_{k1})(1-\delta_{k2})\frac{\Psi_2(X_{k1},X_{k2};\rho)\phi(X_{k2})}{\Psi(X_{k1},X_{k2};\rho)} \\
& -\rho\delta_{k2}(1-\delta_{k1})\frac{\phi(X_{k1}-\rho X_{k2};1-\rho^2)}{1-\Phi(X_{k1}-\rho X_{k2};1-\rho^2)}+\delta_{k1}(1-\delta_{k2})\frac{\phi(X_{k2}-\rho X_{k1};1-\rho^2)}{1-\Phi(X_{k2}-\rho X_{k1};1-\rho^2)}\Bigg\} \\
& \times\frac{1-\Phi(X_{k2})}{\phi(X_{k2})}+\delta_{k2}\Bigg] I(\tilde{X}_{k2}>t).
\end{aligned}
$$

Here $\phi(x;\theta)$ and $\Phi(x;\theta)$ denote the density and distribution functions for the normal random variable with mean 0 and variance θ, respectively. In particular, $\phi(x;1)$ and $\Phi(x;1)$ are the density and distribution functions for the standard normal random variable.

We are now able to construct a step function $\bar{\Lambda}$, which mimics $\hat{\Lambda}$, such that it jumps at distinct observed failure times $\tilde{X}_{ij}$ for which $\delta_{ij}=1$ and its jump size at $\tilde{X}_{ij}$ satisfies

$$
\frac{\delta_{ij}}{\Delta\bar{\Lambda}(\tilde{X}_{ij})}=\sum_{k=1}^{m} R_k(\tilde{X}_{ij};\rho_0,\Lambda_0),
$$

Thus, $\bar{\Lambda}(t)=\sum_{ij} I(\tilde{X}_{ij}\le t)\Delta\bar{\Lambda}(\tilde{X}_{ij})$.

We then show that $\bar{\Lambda}(t)$ converges to $\Lambda_0(t)$ uniformly over $[0,\tau]$ almost surely. To see this, consider the class satisfying the Glivenko-Cantelli property, $\{R_k(t;\rho_0,\Lambda_0): t\in[0,\tau]\}$. Then we obtain

$$
\sup|\frac{1}{m}\sum R_k(t;\rho_0,\Lambda_0)-\mu(t)|\to 0
$$

almost surely, where $\mu(t)=E\{R_k(t;\rho_0,\Lambda_0)\}$. Next denote by $x=\Phi^{-1}\{1-e^{-\Lambda_0(t)}\}$ and consider

$$
\begin{aligned}
& E\{R_{k1}(t;\rho_0,\Lambda_0)\} \\
= \quad & \int_x^\infty \frac{1-\Phi(x_1)}{\phi(x_1)}\int_{-\infty}^\infty \frac{\rho_0^2 x_1-\rho_0 x_2}{1-\rho_0^2}\phi(x_1,x_2;\rho_0)P(U_{i1}\ge x_1,U_{i2}\ge x_2)dx_1dx_2 \\
& +\int_x^\infty \frac{1-\Phi(x_1)}{\phi(x_1)}\int_{-\infty}^\infty \frac{\Psi_1(x_1,x_2;\rho_0)\phi(x_1)}{\Psi(x_1,x_2;\rho_0)}\Psi(x_1,x_2;\rho_0)P(U_{i2}\in dx_2|U_{i1}=x_1)P(U_{i1}\in dx_1)) \\
& -\rho_0\int_x^\infty \frac{1-\Phi(x_1)}{\phi(x_1)}\int_{-\infty}^\infty \frac{\phi(x_2-\rho_0 x_1;1-\rho_0^2)}{1-\Phi(x_2-\rho_0 x_1;1-\rho_0^2)}\frac{-\partial\Psi(x_1,x_2;\rho_0)}{\partial x_1} \\
& \times P(U_{i2}\in dx_2|U_{i1}\ge x_1)P(U_{i1}\ge dx_1)dx_1 \\
& +\int_x^\infty \frac{1-\Phi(x_1)}{\phi(x_1)}\int_{-\infty}^\infty \frac{\phi(x_1-\rho_0 x_2;1-\rho_0^2)}{1-\Phi(x_1-\rho_0 x_2;1-\rho_0^2)}\frac{-\partial\Psi(x_1,x_2;\rho_0)}{\partial x_2}P(U_{i2}\ge x_2|U_{i1}=x_1)dx_2P(U_{i1}\in dx_1)
\end{aligned}
$$

$$
\begin{aligned}
&+\int_x^\infty P(U_{i1}\geq x_1)\phi(x_1)dx_1\\
=\;&\int_x^\infty\{1-\Phi(x_1)\}\int_{-\infty}^\infty\frac{\rho_0^2x_1-\rho_0x_2}{1-\rho_0^2}\phi(x_2-\rho_0x_1;1-\rho_0^2)P(U_{i1}\geq x_1,U_{i2}\geq x_2)dx_1dx_2 &(13)\\
&+\int_x^\infty\{1-\Phi(x_1)\}\int_{-\infty}^\infty\{1-\Phi(x_2-\rho_0x_1;1-\rho_0^2)\}P(U_{i2}\in dx_2|U_{i1}=x_1))P(U_{i1}\in dx_1) &(14)\\
&-\rho_0\int_x^\infty\{1-\Phi(x_1)\}P(U_{i1}\geq x_1)\int_{-\infty}^\infty\phi(x_2-\rho_0x_1;1-\rho_0^2)P(U_{i2}\in dx_2|U_{i1}=x_1)dx_1 &(15)\\
&+\int_x^\infty\{1-\Phi(x_1)\}\int_{-\infty}^\infty\phi(x_1-\rho_0x_2;1-\rho_0^2)\frac{\phi(x_2)}{\phi(x_1)}P(U_{i2}\geq x_2|U_{i1}=x_1)dx_2P(U_{i1}\in dx_1) &(16)\\
&+\int_x^\infty P(U_{i1}\geq x_1)\phi(x_1)dx_1, &(17)
\end{aligned}
$$

wherein the second equality is owing to that

$$
\frac{-\partial\Psi(x_1,x_2;\rho_0)}{\partial x_j}/\phi(x_j)=\Psi_j(x_1,x_2;\rho_0)=1-\Phi(x_{3-j}-\rho_0x_j;1-\rho_0^2).
$$

Now consider the inner integral in (14). Integration by parts gives

$$
\begin{aligned}
&\int_{-\infty}^\infty\{1-\Phi(x_2-\rho_0x_1;1-\rho_0^2)\}P(U_{i2}\in dx_2|U_{i1}=x_1)\\
=\;&-\{1-\Phi(x_2-\rho_0x_1;1-\rho_0^2)\}P(U_{i2}\geq x_2|U_{i1}=x_1)|_{x_2=-\infty}^{x_2=\infty}\\
&-\int_{-\infty}^\infty -P(U_{i2}\geq x_2|U_{i1}=x_1)\frac{d}{dx_2}\{-\Phi(x_2-\rho_0x_1;1-\rho_0^2)\}dx_2\\
=\;&1-\int_{-\infty}^\infty P(U_{i2}\geq x_2|U_{i1}=x_1)\phi(x_2-\rho_0x_1;1-\rho_0^2)dx_2.
\end{aligned}
$$

Using the equality $\phi(x_1-\rho_0x_2;1-\rho_0^2)\frac{\phi(x_2)}{\phi(x_1)}=\phi(x_2-\rho_0x_1;1-\rho_0^2)$, we obtain that (16) is equal to

$$
\int_x^\infty\{1-\Phi(x_1)\}\int_{-\infty}^\infty\phi(x_2-\rho_0x_1;1-\rho_0^2)P(U_{i2}\geq x_2|U_{i1}=x_1)dx_2P(U_{i1}\in dx_1).
$$

Hence, (14) + (16)+ (17) will be

$$
\begin{aligned}
&\int_x^\infty\{1-\Phi(x_1)\}P(U_{i1}\in dx_1)+\int_x^\infty P(U_{i1}\geq x_1)\phi(x_1)dx_1\\
=\;&P(U_{i1}\geq x)\times\{1-\Phi(x)\}=P(U_{i1}\geq x)e^{-\Lambda_0(t)},
\end{aligned}
$$

wherein the first equality is due to integration by parts.

Then consider the inner integral in (15)

$$
\begin{aligned}
&\int_{-\infty}^\infty\phi(x_2-\rho_0x_1;1-\rho_0^2)P(U_{i2}\in dx_2|U_{i1}\geq x_1)\\
=\;&-\phi(x_2-\rho_0x_1;1-\rho_0^2)P(U_{i2}\geq x_2|U_{i1}\geq x_1)|_{x_2=-\infty}^{x_2=\infty}
\end{aligned}
$$

$$
\begin{aligned}
&+\int_{-\infty}^{\infty} P(U_{i2} \geq x_2 | U_{i1} \geq x_1) \frac{d}{dx_2} \phi(x_2 - \rho_0 x_1; 1 - \rho_0^2) dx_2 \\
= \quad & - \int_{-\infty}^{\infty} P(U_{i2} \geq x_2 | U_{i1} \geq x_1) \frac{x_2 - \rho_0 x_1}{1 - \rho_0^2} \phi(x_2 - \rho_0 x_1; 1 - \rho_0^2) dx_2.
\end{aligned}
$$

Hence (13)+ (15) = 0 and

$$
E\{R_{k1}(t; \rho_0, \Lambda_0)\} = P(U_{i1} \geq x) e^{-\Lambda_0(t)}.
$$

Similarly, we can show that

$$
E\{R_{k2}(t; \rho_0, \Lambda_0)\} = P(U_{i2} \geq x) e^{-\Lambda_0(t)},
$$

leading to

$$
\begin{aligned}
\mu(t) \quad &= \quad P(U_{i1} \geq x) e^{-\Lambda_0(t)} + P(U_{i2} \geq x) e^{-\Lambda_0(t)} \\
&= \quad P(\tilde{U}_{i1} \geq t) e^{-\Lambda_0(t)} + P(\tilde{U}_{i2} \geq t) e^{-\Lambda_0(t)},
\end{aligned}
$$

and that $\mu(t)$ is unformly bounded away from 0 over $[0, \tau]$ under condition (c.1). It then follows that

$$
\bar{\Lambda} = \frac{1}{m} \sum_{i=1}^{m} \sum_{j=1}^{2} I(\tilde{X}_{ij} \leq t) \delta_{ij} \left\{ \frac{1}{m} \sum_{k=1}^{m} R_k(\tilde{X}_{ij}; \rho_0, \Lambda_0) \right\}^{-1}
$$

converges uniformly to $E\{\sum_{j=1}^{2} I(\tilde{X}_{ij} \leq t) \delta_{ij} / \mu(\tilde{X}_{ij})\}$ due to the boundness of the indicators, namely $I(\tilde{X}_{ij} \leq t)$ and δ_{ij}. On the other hand, since

$$
\begin{aligned}
E\left(\sum_{j=1}^{2} \frac{I(\tilde{X}_{ij} \leq t) \delta_{ij}}{\mu(\tilde{X}_{ij})} \right) \quad &= \quad \int_0^t \frac{1}{\mu(s)} \sum_{j=1}^{2} P(\tilde{U}_{ij} \geq s) e^{-\Lambda_0(s)} d\Lambda_0(s) \\
&= \quad \Lambda_0(t),
\end{aligned}
$$

it follows that $\bar{\Lambda}$ converges uniformly to Λ_0 over $[0, \tau]$.

Because $\hat{\rho}, \hat{\Lambda}$ maximizes $\ell(\rho, \Lambda)$, clearly $0 \leq \frac{1}{m} \ell(\hat{\rho}, \hat{\Lambda}) - \frac{1}{m} \ell(\rho_0, \bar{\Lambda})$. Furthermore, the construction of $\bar{\Lambda}$ reveals that

$$
\frac{1}{m} \ell(\rho_0, \bar{\Lambda}) = O(1) + \frac{1}{m} \sum_{i=1}^{m} \sum_{j=1}^{2} \delta_{ij} \log(m^{-1}).
$$

Hence, by Proposition 1, we have that

$$
\frac{1}{m} \ell(\hat{\rho}, \hat{\Lambda}) \leq O(1) + \frac{1}{m} \sum_{i=1}^{m} \sum_{j=1}^{2} \delta_{ij} \log\{\Delta \hat{\Lambda}(\tilde{X}_{ij})\} - \frac{1}{m} \sum_{i=1}^{m} \sum_{j=1}^{2} (1 + \delta_{ij}) \log\{1 + \hat{\Lambda}(\tilde{X}_{ij})\}.
$$

Hence,

$$
0 \leq O(1) + \frac{1}{m} \sum_{i=1}^{m} \sum_{j=1}^{2} \delta_{ij} \log\{m \Delta \hat{\Lambda}(\tilde{X}_{ij})\} - \frac{1}{m} \sum_{i=1}^{m} \sum_{j=1}^{2} (1 + \delta_{ij}) \log\{1 + \hat{\Lambda}(\tilde{X}_{ij})\}. \tag{18}
$$

We will show that as $\hat{\Lambda}(\tau) \to \infty$, the right hand side of (18) will eventually turn negative, yielding a contradiction. For this purpose, we adopt a useful partitioning scheme as in Zeng et al. (2005). Specifically, we partition $[0, \tau]$ as follows: let $s_0 = \tau$ and choose $0 \le s_1 < s_0$ such that

$$\frac{1}{2}E\{\sum_{j=1}^{2} I(\tilde{X}_{ij} = s_0)\} = \frac{1}{2}E\{\sum_{j=1}^{2}(1+\delta_{ij})I(\tilde{X}_{ij} = s_0)\} > E\{\sum_{j=1}^{2}\delta_{ij}I(\tilde{X}_{ij} \in [s_1, s_0))\}$$

The existence of such s_1 is guaranteed by regularity condition (c.1). Then define a constant $0 < c < 1$ such that

$$\frac{E\{\sum_{j=1}^{2}\delta_{ij}I(\tilde{X}_{ij} \in [s_1, s_0))\}}{E\{\sum_{j=1}^{2}\delta_{ij}I(\tilde{X}_{ij} \in [0, s_0))\}} \ge \frac{c}{1-c}.$$

If $s_1 > 0$, we consider s_2 such that

$$s_2 = \inf\{0 < s < s_1 : (1-c)E\{\sum_{j=1}^{2}(1+\delta_{ij})I(\tilde{X}_{ij} \in [s_1, s_0))\} \ge E\{\sum_{j=1}^{2}\delta_{ij}I(\tilde{X}_{ij} \in [s, s_1))\}.$$

Such a process can continue. That is, we choose

$$s_{q+1} = \inf\{0 < s < s_q : (1-c)E\{\sum_{j=1}^{2}(1+\delta_{ij})I(\tilde{X}_{ij} \in [s_q, s_{q-1}))\} \ge E\{\sum_{j=1}^{2}\delta_{ij}I(\tilde{X}_{ij} \in [s, s_q))\}, \qquad (19)$$

yielding a sequence of $s_0, s_1, \ldots$ such that

$$\begin{aligned}\frac{1}{2}E\{\sum_{j=1}^{2} I(\tilde{X}_{ij} = s_0)\} = \frac{1}{2}E\{\sum_{j=1}^{2}(1+\delta_{ij})I(\tilde{X}_{ij} = s_0)\} \quad &> \quad E\{\sum_{j=1}^{2}\delta_{ij}I(\tilde{X}_{ij} \in [s_1, s_0))\} \\ (1-c)E\{\sum_{j=1}^{2}(1+\delta_{ij})I(\tilde{X}_{ij} \in [s_q, s_{q-1}))\} \quad &\ge \quad E\{\sum_{j=1}^{2}\delta_{ij}I(\tilde{X}_{ij} \in [s_{q+1}, s_q))\}\end{aligned}$$

for $q \ge 1$.

We show this sequence cannot be infinite, i.e. there exist a finite N such that $s_{N+1} = 0$. Otherwise, we assume $s_q \to s_*$. By the continuity of the distributions involved, (19) implies that

$$(1-c)E\{\sum_{j=1}^{2}(1+\delta_{ij})I(\tilde{X}_{ij} \in [s_q, s_{q-1}))\} = E\{\sum_{j=1}^{2}\delta_{ij}I(\tilde{X}_{ij} \in [s_{q+1}, s_q))\}$$

for $q \ge 1$.

Summing over $q = 1, 2\ldots$, we have that

$$(1-c)E\{\sum_{j=1}^{2}(1+\delta_{ij})I(\tilde{X}_{ij} \in [s_*, s_0))\} = E\{\sum_{j=1}^{2}\delta_{ij}I(\tilde{X}_{ij} \in [s_*, s_1))\}.$$

Hence, $(1-c)E(\sum_{j=1}^{2} I(\tilde{X}_{ij} \in [s_*, s_0)) \le cE(\sum_{j=1}^{2}\delta_{ij}I(\tilde{X}_{ij} \in [s_*, s_1))\}$ which contradicts with the definition of c. Hence there exists a finite N such that $\tau = s_0 > s_1 \ldots > S_{N+1} = 0$.

Now the right hand side of (11) is bounded above by

$$
\begin{aligned}
& O(1)+\sum_{q=0}^{N}\frac{1}{m}\sum_{i,j}I(\tilde{X}_{ij}\in[s_{q+1},s_q))\log\{m\Delta\hat{\Lambda}(\tilde{X}_{ij})\}-\frac{1}{m}\sum_{i,j}I(\tilde{X}_{ij}=\tau)(1+\delta_{ij})\log\{1+\hat{\Lambda}(\tau)\} \\
& -\sum_{q=0}^{N}\frac{1}{m}\sum_{i,j}I(\tilde{X}_{ij}\in[s_{q+1},s_q))(1+\delta_{ij})\log\{1+\hat{\Lambda}(\tilde{X}_{ij})\} \\
\leq\ & O(1)+\sum_{q=0}^{N}\frac{1}{m}\sum_{i,j}I(\tilde{X}_{ij}\in[s_{q+1},s_q))\log\{m\Delta\hat{\Lambda}(\tilde{X}_{ij})\}-\frac{1}{m}\sum_{i,j}I(\tilde{X}_{ij}=\tau)(1+\delta_{ij})\log\{1+\hat{\Lambda}(\tau)\} \\
& -\sum_{q=0}^{N}\frac{1}{m}\sum_{i,j}I(\tilde{X}_{ij}\in[s_{q+1},s_q))(1+\delta_{ij})\log\{1+\hat{\Lambda}(s_{q+1})\}.
\end{aligned}
$$

The concavity of log() yields that

$$
\begin{aligned}
& \frac{1}{m}\sum_{i,j}I(\tilde{X}_{ij}\in[s_{q+1},s_q))\delta_{ij}\log\{m\Delta\hat{\Lambda}(\tilde{X}_{ij})\} \\
\leq\ & \frac{1}{m}\sum_{i,j}I(\tilde{X}_{ij}\in[s_{q+1},s_q))\delta_{ij}\log\left\{m\frac{\sum_{i,j}I(\tilde{X}_{ij}\in[s_{q+1},s_q))\delta_{ij}\Delta\hat{\Lambda}(\tilde{X}_{ij})}{\sum_{i,j}\delta_{ij}I(\tilde{X}_{ij}\in[s_{q+1},s_q))}\right\} \\
\leq\ & \frac{1}{m}\sum_{i,j}I(\tilde{X}_{ij}\in[s_{q+1},s_q))\delta_{ij}\left\{\log\hat{\Lambda}(s_q)-\log\frac{\sum_{i,j}\delta_{ij}I(\tilde{X}_{ij}\in[s_{q+1},s_q))}{m}\right\} \\
\leq\ & O(1)+\frac{1}{m}\sum_{i,j}\delta_{ij}I(\tilde{X}_{ij}\in[s_{q+1},s_q))\log\hat{\Lambda}(s_q).
\end{aligned}
$$

Therefore the above is bounded from above by

$$
\begin{aligned}
& O(1)+\sum_{q=0}^{N}\frac{1}{m}\sum_{i,j}I(\tilde{X}_{ij}\in[s_{q+1},s_q))\log\hat{\Lambda}(s_q)-\frac{1}{m}\sum_{i,j}I(\tilde{X}_{ij}=\tau)(1+\delta_{ij})\log\{1+\hat{\Lambda}(\tau)\} \\
& -\sum_{q=0}^{N}\frac{1}{m}\sum_{i,j}I(\tilde{X}_{ij}\in[s_{q+1},s_q))(1+\delta_{ij})\log\{1+\hat{\Lambda}(s_{q+1})\}.
\end{aligned}
$$

After rearranging items of the above, this bound is equal to, apart from a bounded constant,

$$
\begin{aligned}
& -\frac{1}{2m}\sum_{i,j}I(\tilde{X}_{ij}=\tau)(1+\delta_{ij})\log\{1+\hat{\Lambda}(\tau)\} \\
& -\Bigg[\frac{1}{2m}\sum_{i,j}I(\tilde{X}_{ij}=\tau)(1+\delta_{ij})\log\{1+\hat{\Lambda}(\tau)\} \\
& -\frac{1}{m}\sum_{i,j}I(\tilde{X}_{ij}\in[s_1,\tau))\log\hat{\Lambda}(\tau)\Bigg] \\
& -\sum_{q=1}^{m}\Bigg[\frac{1}{m}\sum_{i,j}I(\tilde{X}_{ij}\in[s_q,s_{q-1}))(1+\delta_{ij})\log\{1+\hat{\Lambda}(s_q)\}
\end{aligned}
$$

$$-\frac{1}{m}\sum_{i,j} I(\tilde{X}_{ij} \in [s_{q+1}, s_q))\delta_{ij} \log\{1 + \hat{\Lambda}(s_q)\}\Bigg] .$$

The first term diverges to $-\infty$ with probability 1 when $\hat{\Lambda}(\tau) \to \infty$ with probability 1. The second term is negative almost surely due to the choice of s_1. By the choice of the sequence of s_q, the third term will also be nonnegative, when m is large. Hence the right hand side of (11) diverges to ∞ almost surely, establishing the contradiction. Hence, we have shown that with probability 1 $\hat{\Lambda}$ has an upper bound. By the Helly selection theorem and with an abuse of notation, we can assume that $\hat{\rho} \to \rho_*$, and $\hat{\Lambda}$ converges pointwise to some increasing function, say, Λ_*.

The second step involves showing $\rho_* = \rho_0, \Lambda_* = \Lambda_0$. We consider

$$\begin{aligned} 0 &\leq \frac{1}{m}\ell(\hat{\rho}, \hat{\Lambda}) - \frac{1}{m}\ell(\rho_0, \bar{\Lambda}) \\ &= \frac{1}{m}\sum \log Q_i(\hat{\rho}, \hat{\Lambda}) - \frac{1}{m}\sum \log Q_i(\rho_0, \bar{\Lambda}) \\ &\quad + \frac{1}{m}\sum_{i=1}^{m}\sum_{j=1}^{2} \delta_{ij} \log\left\{\frac{\Delta\hat{\Lambda}(\tilde{X}_{ij})}{\Delta\bar{\Lambda}(\tilde{X}_{ij})}\right\}. \end{aligned}$$

We can easily see that $\hat{\Lambda}(t)$ is absolutely continuous with respect to $\bar{\Lambda}(t)$ and

$$\hat{\Lambda}(t) = \int_0^t \frac{E_m\{R(O; \rho_0, \Lambda_0, s)\}}{E_m\{R(O; \hat{\rho}, \hat{\Lambda}, s)\}} d\bar{\Lambda}(s) \tag{20}$$

where O represents the underlying random variables for the observed data $(\tilde{X}_{ij}, \delta_{ij})$, E_m denotes the empirical average and $R(O_k; \rho, \Lambda, t) \stackrel{def}{=} R_k(t; \rho, \Lambda)$. Direct calculation yields

$$\begin{aligned} & |R(O_k; \hat{\rho}, \hat{\Lambda}, t) - R(O_k; \rho_*, \Lambda_*, t)| \\ < \; & O(1)\{|\hat{\rho} - \rho_*| + \sum_{j=1}^{2} |\hat{\Lambda}(\tilde{X}_{ij}) - \Lambda_*(\tilde{X}_{ij})| + \int_0^\tau |\hat{\Lambda}(s) - \Lambda_*(s)| ds\}. \end{aligned}$$

Because of the boundness of $\hat{\Lambda}$, dominated convergence yields $\int_0^\tau |\hat{\Lambda}(s) - \Lambda_*(s)| ds \to 0$. Hence

$$\sup_{t \in [0,\tau]} |E_m\{R(O; \hat{\rho}, \hat{\Lambda}, t)\} - E_m\{R(O; \rho_*, \Lambda_*, t)\}| \to 0$$

almost surely. Both classes

$$\{R(O; \rho_*, \Lambda_*, t), t \in [0, \tau]\}, \{R(O; \rho_0, \Lambda_0, t), t \in [0, \tau]\}$$

consist of the multiplication of a monotone indicator function and a t-independent random variable, e.g $\sum_{j=1}^{2} I(\tilde{X}_{ij} \geq t) B_{ij}$ and hence both are Glivenko-Cantelli, leading to

$$\sup_{t \in [0,\tau]} |E_m\{R(O; \rho_0, \Lambda_0, t)\} - E\{R(O; \rho_0, \Lambda_0, t)\}| \to 0,$$

and

$$\sup_{t\in[0,\tau]} |E_m\{R(O;\hat{\rho},\hat{\Lambda},t)\} - E\{R(O;\rho_*,\Lambda_*,t)\}| \to 0.$$

Furthermore, as $E\{(R(O;\rho_*,\Lambda_*,t)\}$ is uniformly bounded away from 0, taking the limit on both sides of (20) will give

$$\Lambda_*(t) = \int_0^t \frac{E\{R(O;\rho_0,\Lambda_0,t)\}}{E\{R(O;\rho_*,\Lambda_*,t)\}} d\Lambda_0(t).$$

Thus, $\Lambda_*(t)$ is absolutely continuous with respect to $\Lambda_0(t)$ and hence is differentiable with respect to t. Using similar argument we will have

$$E_m\{\log Q_i(\hat{\rho},\hat{\Lambda})\} \to E\{\log Q_i(\rho_*,\Lambda_*)\}$$

and

$$E_m\{\log Q_i(\rho_0,\bar{\Lambda})\} \to E\{\log Q_i(\rho_0,\Lambda_0)\}$$

in probability. Hence we have

$$\begin{aligned} 0 &\leq \frac{1}{m}\ell(\hat{\rho},\hat{\Lambda}) - \frac{1}{m}\ell(\rho_0,\bar{\Lambda}) \\ &\to E\log\left\{\frac{Q_i(\rho_*,\Lambda_*)\Lambda_*'(\tilde{X}_{ij})}{Q_i(\rho_0,\Lambda_0)\Lambda_0'(\tilde{X}_{ij})}\right\} \end{aligned}$$

which is the negative Kulback-Leibler information, implying $\rho_* = \rho_0, \Lambda_* = \Lambda_0$. Hence we conclude, with probability 1, $\hat{\rho} \to \rho_0$ and pointwise $\hat{\Lambda} \to \Lambda_0$. Then the uniform convergence of $\hat{\Lambda}$ follows immediately since Λ_0 is a continuous monotone function. □

Proposition 2 *(Asymptotic Normality) The scaled process $\sqrt{m}(\hat{\rho}-\rho_0, \hat{\Lambda}-\Lambda_0)$ converges weakly to a zero-mean Gaussian process in the metric space $R \times l^\infty[0,\tau]$, where $l^\infty[0,\tau]$ is the linear space containing all the bounded functions in $[0,\tau]$ equipped with the supremum norm. Furthermore, $\hat{\rho}$ and $\int_0^\tau \eta(s)d\hat{\Lambda}(s)$ are asymptotically efficient, where $\eta(s)$ is any function of bounded variation over $[0,\tau]$.*

Proof: The proof involves invoking Theorem 3.3.1 of van der Vaart and Wellner (1996), which requires 4 sufficient conditions, namely, weak convergence of the empirical process at the truth, approximation of the score operator, Fréchet differentiability of the asymptotic score and invertibility of the information operator. To proceed, we need to define a random map ξ_m and a fixed map ξ in a set, say, A, containing the true parameter and the possible values of the estimator (or at least asymptotically). Specifically, denoting by

$\theta = (\rho, \Lambda)$ the unknown parameters and defining the true parameters θ_0 and their estimates $\hat{\theta}$ likewise, we define a small neighborhood containing the true parameter θ_0 as

$$A = \{\theta = (\rho, \Lambda) : ||\theta - \theta_0|| \leq \epsilon\},$$

where $||\theta - \theta_0|| \overset{def}{=} |\rho - \rho_0| + \sup_{t \in [0,\tau]} |\Lambda(t) - \Lambda_0(t)|$ and ϵ is a small positive constant. It then follows by Proposition 1 that, when the sample size is sufficiently large, the estimates, $\hat{\theta}$, fall into A almost surely. We also define a set

$$\mathcal{H} = \{h = (h_1, h_2) : h_1 \in R, and\, h_2\, \mathrm{is\, a function\, of\, bounded\, variation; |h_1| \leq 1, ||h_2||_V \leq 1}\},$$

where for a function of bounded variation h_2, $||h_2||_V$ denotes its total variation over $[0, \tau]$, that is, $||h_2||_V = \int_0^\tau |dh_2| + |h_2(0)|$. We are ready to define ξ_m and ξ as random and fixed maps respectively from A to $l^\infty(\mathcal{H})$, consisting of all the bounded functionals on $\mathcal{H}$. Specifically

$$\xi_m(\rho, \Lambda)[h_1, h_2] = E_m(\psi_{\theta,h}), \;\; \xi(\rho, \Lambda)[h_1, h_2] = E(\psi_{\theta,h}),$$

where E_m denotes the empirical average, $\psi_{\theta,h} = h_1 \ell_\rho(\rho, \Lambda) + \ell_\Lambda(\rho, \Lambda)[\int_0^\cdot h_2 d\Lambda]$, ℓ_ρ is the score for ρ and $\ell_\Lambda[\int_0^\cdot h_2 d\Lambda]$ is the score for Λ along the submodel $\Lambda(\cdot) + \epsilon \int_0^\cdot h_2 d\Lambda$. The MLE's and the true parameters thus satisfy $\xi_m(\hat{\rho}, \hat{\Lambda}) = 0$ and $\xi(\rho_0, \Lambda_0) = 0$. Examining the forms of ℓ_ρ, ℓ_Λ (given later) will yield $\psi_{\theta_0,h}$ is a P-Donsker class when h varies over $\mathcal{H}$. Hence, $\sqrt{m}\{\xi_m(\rho_0, \Lambda_0) - \xi(\rho_0, \Lambda_0)\}$ will converge in distribution to a tight random element, satisfying the first sufficient condition (weak convergence of the empirical process at the truth) of Theorem 3.3.1 of van der Vaart and Wellner (1996). It can be further shown that the class $\psi_{\theta,h} - \psi_{\theta_0,h}$ is a P-Donsker on $A \times \mathcal{H}$ and $\sup_{h \in \mathcal{H}} E(\psi_{\theta,h} - \psi_{\theta_0,h})^2 \to 0$ when $\theta \to \theta_0$. Since $\hat{\theta} \to \theta$ uniformly and almost surely by Proposition 1, Lemma 3.3.5 of van der Vaart and Wellner (1996) implies that

$$||\sqrt{m}\{\xi_m(\hat{\rho}, \hat{\Lambda}) - \xi(\hat{\rho}, \hat{\Lambda})\} - \sqrt{m}\{\xi_m(\rho_0, \Lambda_0) - \xi(\rho_0, \Lambda_0)\}||_{\mathcal{H}} = o_p(1 + \sqrt{m}||\hat{\theta} - \theta_0||), \tag{21}$$

where the norm $|| \cdot ||_{\mathcal{H}}$ is defined for any map ξ on $\mathcal{H}$ as $||\xi||_{\mathcal{H}} = \sup_{h \in \mathcal{H}} |\xi[h]|$. Indeed, (21) is the second sufficient condition (approximation of the score operator) of Theorem 3.3.1 of van der Vaart and Wellner (1996). We next consider the condition of Fréchet-differentiability of ξ at θ_0, namely,

$$||\xi(\theta) - \xi(\theta_0) - \dot{\xi}_{\theta_0}(\theta - \theta_0)||_{\mathcal{H}} = o(||\theta - \theta_0||).$$

Indeed, this follows from the finiteness of the moments of the joint normal distribution [in view of (2)].

We finally show the invertibility of $\dot{\xi}_{\theta_0}$. After linearizing the score equation $\xi_m(\hat{\rho}, \hat{\Lambda}) - \xi(\rho_0, \Lambda_0) = 0$, standard computation would yield

$$\begin{aligned}\sqrt{m}\{\xi(\hat{\theta})[h] - \xi(\theta_0)[h]\} &= \sqrt{m}E(\psi_{\hat{\theta},h} - \psi_{\theta_0,h}) \\ &= -\sqrt{m}(E_m - E)\psi_{\theta_0,h} + o_p(1 + \sqrt{m}(||\theta - \theta_0||). \end{aligned} \quad (22)$$

Also note that

$$\begin{aligned}\xi(\hat{\theta})[h] - \xi(\theta_0)[h] &= \dot{\xi}_{\theta_0}(\hat{\theta} - \theta_0)[h] + o_p(||\hat{\theta} - \theta_0|| \\ &= W_1[h_1, h_2](\hat{\rho} - \rho_0) + \int W_2[h_1, h_2]d(\hat{\Lambda} - \Lambda_0) + o_p(||\hat{\theta} - \theta_0||),\end{aligned}$$

where $W_i, i = 1, 2$, are linear operators on $\mathcal{H}$ satisfying

$$\begin{aligned}& W_1[h_1, h_2]\tilde{h}_1 + \int W_2[h_1, h_2]\tilde{h}_2 d\Lambda_0 \\ =& \frac{d}{d\epsilon}E\left\{h_1\ell_\rho(\rho_\epsilon, \Lambda_\epsilon) + \ell_\Lambda(\rho_\epsilon, \Lambda_\epsilon)[\int_0^\cdot h_2 d\Lambda_\epsilon]\right\}\end{aligned} \quad (23)$$

with $\rho_\epsilon = \rho_0 + \epsilon\tilde{h}_1$, $\Lambda_\epsilon = \Lambda_0(\cdot) + \int_0^\cdot \epsilon\tilde{h}_2 d\Lambda_0$. If $\dot{\xi}_{\theta_0}$ is invertible, the same argument in the proof of Theorem 3.3.1 leads to the $\sqrt{m}$ consistency of $\hat{\theta}$. Hence the remainder term $o_p(1 + \sqrt{m}(||\theta - \theta_0||)$ in (22) can be replaced by $o_p(1)$. We now prove that $\dot{\xi}_{\theta_0}$ is invertible. Specifically, suppose, there exist h_1, h_2 such that $W_i[h_1, h_2] = 0$ for $i = 1, 2$, we wish to show $h_1 = 0, h_2 = 0$. Consider (23). Choosing $\tilde{h}_1 = h_1$, $\tilde{h}_2 = h_2$, we have that

$$E\left\{h_1\ell_\rho(\rho_0, \Lambda_0) + \ell_\Lambda(\rho_0, \Lambda_0)[\int_0^\cdot h_2 d\Lambda_0]\right\}^2 = 0.$$

Hence,

$$h_1\ell_\rho(\rho_0, \Lambda_0) + \ell_\Lambda(\rho_0, \Lambda_0)[\int_0^\cdot h_2 d\Lambda_0] = 0 \quad (24)$$

almost surely.

We next show that (24) implies $h_1 = 0, h_2 = 0$. Note that the left hand side of (24) is the score function at the true value along a one-dimensional submodel $\{\rho_0 + \epsilon h_1, \Lambda_0(\cdot) + \epsilon\int_0^\cdot h_2 d\Lambda_0\}$. Denote by $x_j = \Phi^{-1}\left[1 - e^{-\Lambda_0(\tilde{X}_{ij})}\right]$ for $j = 1, 2$. Direct calculation yields

$$\begin{aligned}& \ell_\Lambda(\theta_0)[\int_0^\cdot h_2 d\Lambda_0] \\ =& \left\{(\delta_{i1}\delta_{i2}\frac{\rho_0}{1 - \rho_0^2}(x_2 - \rho_0 x_1) + \delta_{i1}(1 - \delta_{i2})\frac{\rho_0\phi(x_2 - \rho_0 x_1; 1 - \rho_0^2)}{1 - \Phi(x_2 - \rho_0 x_1; 1 - \rho_0^2)}\right.\end{aligned}$$

$$
\begin{aligned}
&+\delta_{i2}(1-\delta_{i1})\frac{-\phi(x_1-\rho_0 x_2;1-\rho_0^2)}{1-\Phi(x_1-\rho_0 x_2;1-\rho_0^2)} \\
&+(1-\delta_{i2})(1-\delta_{i1})\frac{-\{1-\Phi(x_2-\rho_0 x_1;1-\rho_0^2)\}\phi(x_1)}{\Psi(x_1,x_2;\rho_0)})\Big\}\frac{\exp\{-\Lambda_0(\tilde{X}_{i1})\}}{\phi(x_1)}\int_0^{\tilde{X}_{i1}} h_2(s)d\Lambda_0(s) \\
&+\Big\{(\delta_{i1}\delta_{i2}\frac{\rho_0}{1-\rho_0^2}(x_1-\rho_0 x_2)+\delta_{i2}(1-\delta_{i1})\frac{\rho_0\phi(x_1-\rho_0 x_2;1-\rho_0^2)}{1-\Phi(x_1-\rho_0 x_2;1-\rho_0^2)} \\
&+\delta_{i1}(1-\delta_{i2})\frac{-\phi(x_2-\rho_0 x_1;1-\rho_0^2)}{1-\Phi(x_2-\rho_0 x_1;1-\rho_0^2)} \\
&+(1-\delta_{i2})(1-\delta_{i1})\frac{-\{1-\Phi(x_1-\rho_0 x_2;1-\rho_0^2)\}\phi(x_2)}{\Psi(x_1,x_2;\rho_0)})\Big\}\frac{\exp\{-\Lambda_0(\tilde{X}_{i2})\}}{\phi(x_2)}\int_0^{\tilde{X}_{i2}} h_2(s)d\Lambda_0(s) \\
&+\delta_{i1}(Z_{i1}(h_2(\tilde{X}_{i1})-\int_0^{\tilde{X}_{i1}} h_2(s)d\Lambda_0(s)+\delta_{i2}(h_2(\tilde{X}_{i2})-\int_0^{\tilde{X}_{i2}} h_2(s)d\Lambda_0(s)
\end{aligned}
$$

and

$$
\begin{aligned}
\ell_\rho(\theta_0) &= \delta_{i1}\delta_{i2}\frac{\rho_0(1-\rho_0^2)+(1+\rho_0^2)x_1x_2-\rho_0(x_1^2+x_2^2)}{(1-\rho_0^2)^2} \\
&\quad+\delta_{i1}(1-\delta_{i2})\frac{\int_{x_2}^\infty a(t_2,x_1;\rho_0)dt_2}{1-\Phi(x_2-\rho_0 x_1;1-\rho_0^2)} \\
&\quad+\delta_{i2}(1-\delta_{i1})\frac{\int_{x_1}^\infty a(t_1,x_2;\rho_0)dt_1}{1-\Phi(x_1-\rho_0 x_2;1-\rho_0^2)} \\
&\quad+(1-\delta_{i2})(1-\delta_{i1})\int_{x_1}^\infty\int_{x_2}^\infty a(t_1,t_2;\rho_0)\phi(t_2)dt_2dt_1
\end{aligned}
$$

where $a(t_1,t_2;\rho)=\frac{\partial}{\partial\rho}\phi(t_1-\rho t_2;1-\rho^2)$.

Consider (24) and set $\delta_{i1}=\delta_{i2}=0$ and $\tilde{X}_{ij}=s_j\in[0,\tau]$ for $j=1,2$. Then we obtain the following integral equation

$$
\begin{aligned}
&\left\{1-\Phi(s_2-\rho_0 s_1;1-\rho_0^2)\right\}e^{-\Lambda_0(\tilde{s}_1)}\int_0^{\tilde{s}_1} h_2(s)d\Lambda_0(s) \\
&+\left\{1-\Phi(s_1-\rho_0 s_2;1-\rho_0^2)\right\}e^{-\Lambda_0(\tilde{s}_2)}\int_0^{\tilde{s}_2} h_2(s)d\Lambda_0(s) \\
=\;& h_1\Psi(s_1,s_2;\rho_0)\int_{s_1}^\infty\int_{s_2}^\infty a(t_1,t_2;\rho_0)\phi(t_2)dt_2dt_1
\end{aligned}
$$

where $s_j=\Phi^{-1}(1-e^{-\Lambda_0(\tilde{s}_j)})$ for $j=1,2$. In the above integral equation of h_2, fix $\tilde{s}_1$ and let $\tilde{s}_2\to 0$, we conclude $e^{-\Lambda_0(\tilde{s}_1)}\int_0^{\tilde{s}_1} h_2(s)d\Lambda_0(s)=0$ for any $\tilde{s}_1\in[0,\tau]$. Hence $h_2\equiv 0$ on $[0,\tau]$, which also implies $h_1=0$. Therefore, we have shown the invertibility of $\dot{\xi}_{\theta_0}$, which, in junction with the other proved conditions, implies the weak convergence of $\sqrt{m}(\hat{\theta}-\theta_0)$. Specifically, for any $h\in\mathcal{H}$

$$
\sqrt{m}\left\{h_1(\hat{\rho}-\rho_0)+\int_0^\tau h_2 d(\hat{\Lambda}-\Lambda_0)\right\}=-\sqrt{m}(E_m-E)\psi_{\theta_0,\tilde{h}}+o_p(1),
$$

where $\tilde{h} \overset{def}{=} (\tilde{h}_1, \tilde{h}_2) = (W_1, W_2)^{-1}(h_1, h_2)$. Setting $h_2 = 0$ reveals that $\hat{\rho}$ is an asymptotically linear estimator for ρ_0, with the influence function lying in the space spanned by the score functions, and hence in the tangent space of the semiparametric normal transformation model. Similarly, setting $h_1 = 0, h_2 = \eta$ (without loss of generality, we assume here that $||\eta||_V = 1$; otherwise, we can easily apply a normalization $\eta/||\eta||_V$) reveals that $\int_0^\tau \eta(s)d\hat{\Lambda}$ is also an asymptotically linear estimator for $\int_0^\tau \eta(s)d\Lambda_0$, with corresponding influence functions on the space spanned by the score functions. Therefore, the semiparametric efficiency theory [see e.g. Proposition 1 in (Bickel et al., 1993, ch3.2)] both $\hat{\rho}$, $\int_0^\tau \eta(s)d\hat{\Lambda}$ are efficient estimators. □

3 Semiparametric Maximum Likelihood Estimation With Stratified Hazards

In this section, we relax the condition of a common marginal hazard and allow each member of the pair to have a distinct hazard. That is, each $\tilde{T}_{ij}$ has a cumulative hazard $\Lambda_j(\cdot)$ such that $\Lambda_1(\cdot) \not\equiv \Lambda_1(\cdot)$.

We consider a joint maximum likelihood estimation for inference. The ensuing development is parallel to that in the common hazard model. Specifically, our inference stems from the log likelihood function of unknown parameters $(\Lambda_1, \Lambda_2, \rho)$ based on the observed data $(\tilde{X}_{ij}, \delta_{ij}), j = 1, 2, i = 1, \ldots, m$, which can be written, up to a constant, as the product over $i = 1, \ldots, m$ of terms

$$\begin{aligned}\tilde{L}_i(\rho, \Lambda_1, \Lambda_2) \quad = \quad & \{e^{g(X_{i1}, X_{i2})}\Lambda_1'(\tilde{X}_{i1})\Lambda_2'(\tilde{X}_{i2})e^{-\Lambda_1(\tilde{X}_{i1})-\Lambda_2(\tilde{X}_{i2})}\}^{\delta_{i1}\delta_{i2}}\{\Psi_1(X_{i1}, X_{i2}; \rho)\Lambda_1'(\tilde{X}_{i1})e^{-\Lambda_1(\tilde{X}_{i1})}\}^{\delta_{i1}(1-\delta_{i2})} \\ & \times\{\Psi_2(X_{i1}, X_{i2}; \rho)\Lambda_2'(\tilde{X}_{i2})e^{-\Lambda_2(\tilde{X}_{i2})}\}^{(1-\delta_{i1})\delta_{i2}} \times \{\Psi(X_{i1}, X_{i2}; \rho)\}^{(1-\delta_{i1})(1-\delta_{i2})}. \qquad (25)\end{aligned}$$

Here $X_{ij} = \Phi^{-1}\{1 - \exp(-\Lambda_j(\tilde{X}_{ij})\}$ for $j = 1, 2$. Again, directly maximizing the likelihood function (25) in a space containing continuous hazards $\Lambda_1(\cdot)$ or $\Lambda_2(\cdot)$ is infeasible, as one can always make the likelihood be arbitrarily large by constructing some continuous functions $\Lambda_1(\cdot)$ and $\Lambda_2(\cdot)$ with fixed values at each $\tilde{X}_{ij}$ while letting $\Lambda_1'(\cdot)$ or $\Lambda_2'(\cdot)$ go to ∞ at an observed failure time. Hence, when performing the maximum likelihood estimation, we need to consider the following parameter space for (Λ_1, Λ_2):

$$\{(\Lambda_1, \Lambda_2) : \Lambda_1, \Lambda_2 \text{ are cadlag and piecewise constant}\}.$$

It follows that the SPMLE, $(\hat{\rho}, \hat{\Lambda}_1, \hat{\Lambda}_2)$, is the maximizer to the empirical likelihood function $\ell(\rho, \Lambda_1, \Lambda_2)$, which is obtained from (25) with the derivatives $\Lambda_1'(\cdot)$ and $\Lambda_2'(\cdot)$ at the observed failure times replaced by their jumps $\Delta\Lambda_1(\cdot)$ and $\Delta\Lambda_2(\cdot)$ at the corresponding time points, respectively. Using the similar arguments for Lemma 2 in Section A.1, we know that $(\hat{\rho}, \hat{\Lambda}_1, \hat{\Lambda}_2)$ do exist and are finite. Furthermore, under conditions (c.1)-(c.3) [we let both Λ_1 and Λ_2 satisfy (c.3)], the asymptotic properties of the SPMLEs are summarized in the following two theorems, namely, the consistency theorem, followed by the asymptotic normality theorem.

Proposition 3 *(Consistency) Denote by* $(\rho_0, \Lambda_{01}, \Lambda_{02})$ *the true parameters. Then* $|\hat{\rho}-\rho_0| \to 0$, $\sup_{t\in[0,\tau]} |\hat{\Lambda}_1(t)-\Lambda_{01}(t)| \to 0$ *and* $\sup_{t\in[0,\tau]} |\hat{\Lambda}_2(t) - \Lambda_{02}(t)| \to 0$ *almost surely.*

Proof: This proposition can be proved along the lines for the cases of common marginal hazard model. That is, we show by contradiction that the SPMLEs $\hat{\Lambda}_1, \hat{\Lambda}_2$ stay bounded, followed by showing every convergent subsequence converge to the true parameters. Specifically, we construct two step functions $\bar{\Lambda}_j (j = 1, 2)$, which jump only at distinct observed failure times, ie $\tilde{X}_{ij}$ for which $\delta_{ij} = 1$ such that $\bar{\Lambda}_j$ will be close to the true function Λ_{0j}. Differentiating $\ell(\rho, \Lambda_1, \Lambda_2)$ with respect to $\Delta\Lambda_j(\tilde{X}_{ij})$ and setting it to 0, we have the following equations [compare to (12)]

$$\frac{\delta_{ij}}{\Delta\Lambda_j(\tilde{X}_{ij})} = \sum_{k=1}^{m} R_{kj}(\tilde{X}_{ij}; \hat{\rho}, \hat{\Lambda}_1, \hat{\Lambda}_2)$$

where

$$\begin{aligned} R_{k1}(t; \rho, \Lambda_1, \Lambda_2) \quad = \quad & \Bigg[\Bigg\{\delta_{k1}\delta_{k2}\frac{\rho^2 X_{k1} - \rho X_{k2}}{1-\rho^2} + (1-\delta_{k1})(1-\delta_{k2})\frac{\Psi_1(X_{k1}, X_{k2}; \rho)\phi(X_{k1})}{\Psi(X_{k1}, X_{k2}; \rho)} \\ & -\rho\delta_{k1}(1-\delta_{k2})\frac{\phi(X_{k2} - \rho X_{k1}; 1-\rho^2)}{1-\Phi(X_{k2}-\rho X_{k1}; 1-\rho^2)} + \delta_{k2}(1-\delta_{k1})\frac{\phi(X_{k1} - \rho X_{k2}; 1-\rho^2)}{1-\Phi(X_{k1}-\rho X_{k2}; 1-\rho^2)}\Bigg\} \\ & \times \frac{1-\Phi(X_{k1})}{\phi(X_{k1})} + \delta_{k1}\Bigg] I(\tilde{X}_{k1} > t) \end{aligned}$$

and

$$\begin{aligned} R_{k2}(t; \rho, \Lambda_1, \Lambda_2) \quad = \quad & \Bigg[\Bigg\{\delta_{k1}\delta_{k2}\frac{\rho^2 X_{k2} - \rho X_{k1}}{1-\rho^2} + (1-\delta_{k1})(1-\delta_{k2})\frac{\Psi_2(X_{k1}, X_{k2}; \rho)\phi(X_{k2})}{\Psi(X_{k1}, X_{k2}; \rho)} \\ & -\rho\delta_{k2}(1-\delta_{k1})\frac{\phi(X_{k1} - \rho X_{k2}; 1-\rho^2)}{1-\Phi(X_{k1}-\rho X_{k2}; 1-\rho^2)} + \delta_{k1}(1-\delta_{k2})\frac{\phi(X_{k2} - \rho X_{k1}; 1-\rho^2)}{1-\Phi(X_{k2}-\rho X_{k1}; 1-\rho^2)}\Bigg\} \\ & \times \frac{1-\Phi(X_{k2})}{\phi(X_{k2})} + \delta_{k2}\Bigg] I(\tilde{X}_{k2} > t). \end{aligned}$$

We let the jump size of $\bar{\Lambda}_j$, which mimics $\hat{\Lambda}_j$, satisfy

$$\frac{\delta_{ij}}{\Delta\bar{\Lambda}_j(\tilde{X}_{ij})} = \sum_{k=1}^{m} R_{kj}(\tilde{X}_{ij}; \rho_0, \Lambda_{01}, \Lambda_{02}).$$

Thus, $\bar{\Lambda}_j(t) = \sum_{ij} I(\tilde{X}_{ij} \leq t)\Delta\bar{\Lambda}_j(\tilde{X}_{ij})$ for $j = 1, 2$.

Since $\hat{\rho}, \hat{\Lambda}_1, \hat{\Lambda}_2$ maximize the likelihood, it follows that

$$0 \leq \ell(\hat{\rho}, \hat{\Lambda}_1, \hat{\Lambda}_2) - \ell(\rho_0, \bar{\Lambda}_1, \bar{\Lambda}_2). \tag{26}$$

Then we can use similar arguments as in the common hazard cases to show that if $\hat{\Lambda}_1(\tau) \to \infty$ or $\hat{\Lambda}_2(\tau) \to \infty$, the right-hand side of inequality (26) will be negative, yielding a contradiction. The proof can be completed by using the same arguments as in the common hazard cases to show every convergent subsequence converges to the true parameters. □

Proposition 4 *(Asymptotic Normality) The empirical process* $\sqrt{m}(\hat{\rho} - \rho_0, \hat{\Lambda}_1 - \Lambda_{01}, \hat{\Lambda}_2 - \Lambda_{02})$ *converges weakly to a zero-mean Gaussian process in the metric space* $R \times l^\infty[0, \tau] \times l^\infty[0, \tau]$*, where* $l^\infty[0, \tau]$ *is the linear space containing all the bounded functions in* $[0, \tau]$ *equipped with the supremum norm. Furthermore,* $\hat{\rho}$*,* $\int_0^\tau \eta_1(s) d\hat{\Lambda}_1(s)$ *and* $\int_0^\tau \eta_2(s) d\hat{\Lambda}_2(s)$ *are asymptotically efficient, where* $\eta_1(s), \eta_2(s)$ *are any functions of bounded variation over* $[0, \tau]$.

Proof: The arguments are parallel to those in the proof of Proposition 2, and thus are omitted. □

4 A Simple One-Step double-robust IPCW Estimator

When the censoring time, for example, the drop-out time or the time from study entry to the end of the study is common for both pair members (Lin and Ying, 1993; Tsai and Crowley, 1998; Wang and Wells, 1998; Nan et al., 2006), one may apply the ideas of Robins and Rotnitzky (1992) and van der Laan et al. (2002) to construct a simple one-step estimator for estimating the multivariate survival functions at any fixed time points, say, (t_1, t_2). Such an estimator might be appealing because, when the underlying semiparametric normal transformation model is correctly specified, it is locally semiparametric efficient in the sense that its asymptotic variance reaches the supremum of the Cramer-Rao bounds among all the parametric sub-models which pass through the true model (van der Laan et al, 2002). Moreover, this estimator enjoys a double robustness property such that it is consistent even if the underlying semiparametric normal transformation model is misspecified.

The idea is to use an IPCW (inverse-probability-of-censoring-weighted) estimator (Robins and Rotnitzky, 1992) as an initial estimator, and add to it an estimate of the empirical mean of the estimated efficient influence function in the class of semiparametric models, wherein the bivariate survival $S(u, v) = P(\tilde{T}_1 >$

$u, \tilde{T}_2 > v)$ is nonparametric and the censoring time is independent from the true survival time. Indeed, van der Laan et al. (2002) allowed the censoring mechanism to satisfy coarsening at random, including independent censoring as an important special case.

To facilitate the ensuing development, we denote by $B = I(\tilde{T}_1 > t_1, \tilde{T}_2 > t_2), \tilde{T}^* = \tilde{T}_1 \vee \tilde{T}_2 \stackrel{def}{=} max(\tilde{T}_1, \tilde{T}_2), t^* = t_1 \vee t_2, \tilde{X}^* = \tilde{X}_1 \vee \tilde{X}_2, \Delta = I(\mathrm{B\,is\,observed})$ and by V the earliest time at which B is observed, in which case $V = \tilde{T}^* \wedge t^*$. Also, we denote by $Y_i = (\tilde{X}_{i1}, \tilde{X}_{i2}, \delta_{i1}, \delta_{i2})$ the observation from each pair, which is an iid realization of $Y = (\tilde{X}_1, \tilde{X}_2, \delta_1, \delta_2)$, and let G be the survival function for the common censoring time $\tilde{U}$.

The IPCW estimator stems from the observation (see, e.g. van der Laan et al., 2002)

$$E\left\{\frac{\Delta B}{G(V)}\right\} = E\left\{\frac{\Delta B}{G(t^*)}\right\} = S(t_1, t_2),$$

which naturally leads to the following estimator

$$\mu_m^0 = \frac{1}{m}\sum_{i=1}^{m}\frac{\Delta_i B_i}{G_m(t^*)} = \frac{1}{m}\sum_{i=1}^{m}\frac{I(\tilde{X}_{i1} > t_1, \tilde{X}_{i2} > t_2)}{G_m(t^*)}.$$

Here, G_m is the Kaplan-Meier estimator for G based on the m observations $\{\tilde{U}_i \wedge \tilde{T}_i^*, I(\tilde{U}_i \leq \tilde{T}_i^*)\}$, where $\tilde{T}_i^*$ plays the role of censoring $\tilde{U}_i$.

The local efficient estimator will be constructed by subtracting from the influence curve of μ_m^0 (when $G_m = G$ is known), $IC_0(Y|G, \mu_0) = \frac{\Delta B}{G(V)} - \mu_0$, its projection IC^*_{nu}onto the tangent space of the nuisance parameter G, which is given by

$$IC^*_{nu}(Y|\mathcal{Q}, G) = -\int_0^{t^*} Q(u)\frac{dM(u)}{G(u)}$$

where $M(u) = I\{\tilde{U}_i \leq u, I(\tilde{U} \leq \tilde{T}^*)\} - \int_0^u I(\tilde{X}^* \geq s)d\{-logG(s)\}$, $\mathcal{Q} = E\{B|\tilde{X}^* \geq u, I(\tilde{T}_1 > s), I(\tilde{T}_2 > s), 0 < s < u\}$. We write $\mathcal{Q}$ in the conditioning part owing to that the projection of IC_0 onto the tangent space of the nuisance G depends on the true law of $S(\cdot, \cdot)$ through $\mathcal{Q}$. Some algebra would give the closed-form expression of $\mathcal{Q}$ as follows

$$\mathcal{Q}(u) = \begin{cases} I(\tilde{T}_1 > t_1)P(\tilde{T}_2 > t_2 \vee u|\tilde{T}_1)/P(\tilde{T}_2 > t_2 \vee u|\tilde{T}_1) & \text{if } \tilde{T}_1 \leq u, \tilde{T}_2 > u, \\ I(\tilde{T}_2 > t_2)P(\tilde{T}_1 > t_1 \vee u|\tilde{T}_2)/P(\tilde{T}_1 > t_1 \vee u|\tilde{T}_1) & \text{if } \tilde{T}_1 > u, \tilde{T}_2 \leq u, \\ P(\tilde{T}_1 > t_1 \vee u, \tilde{T}_2 > t_2)/P(\tilde{T}_1 > t_1 \vee u, \tilde{T}_2 > t_2 \vee u) & \text{if } \tilde{T}_1 > u, \tilde{T}_2 > u. \end{cases}$$

Hence, the efficient influence curve IC^* is given by

$$IC^*(Y|\mathcal{Q}, G, \mu) = IC_0(Y|G, \mu) - IC^*_{nu}(Y|\mathcal{Q}, G)$$

and the optimal estimating equation, corresponding to the efficient influence curve, is

$$0 = \frac{1}{m}\sum_{i=1}^{m} IC^*(Y_i|\mathcal{Q}_m, G_m, \mu)$$

where $\mathcal{Q}_m$ is the estimator for $\mathcal{Q}$, with all the bivariate survivals and conditional survivals involving in $\mathcal{Q}$ estimated based on the semiparametric normal transformation model.

Thus, a one-step Newton-Raphson algorithm yields a solution

$$\mu_m^1 = \mu_m^0 + \frac{1}{m}\sum_{i=1}^{m} IC^*(Y_i|\mathcal{Q}_m, G_m, \mu_m^0) = \mu_m^0 - \frac{1}{m}\sum_{i=1}^{m} IC^*_{nu}(Y_i|\mathcal{Q}_m, G_m),$$

which is a locally efficient estimator when the semiparametric normal transformation model is correctly specified. On the other hand, even if the underlying semiparametric normal transformation model is not true, because of the (double) unbiasedness property (van der Laan et al., 2002)

$$E\{IC^*(Y|\mathcal{Q}_1, G_1, \mu)\} = 0, \text{ if } G_1 = G \; or \; \mathcal{Q}_1 = \mathcal{Q}\{S(\cdot,\cdot)\},$$

this one-step estimator will also be consistent. Furthermore, a simple variance estimator for μ_m^1 is given by

$$\frac{1}{m^2}\sum_{i=1}^{m}\{IC^*(Y_i|\mathcal{Q}_m, G_m, \mu_m^0)\}^2, \tag{27}$$

which is valid even when the semiparametric normal transformation model is misspecified.

In theory, one would be able to extend the results from the univariate censoring mechanism to accommodate arbitrary bivariate censoring (Robins, Rotnitzky and van der Laan, 1999). However, bivariate censoring will result in loss of the monotonicity of the missingness, as opposed to univariate censoring. Therefore, the closed-form projections onto the tangent space for the censoring distributions are in general not available and, in practice, more complicated successive approximations or Monte Carlo simulations (see, e.g., Keles et al., 2004), are needed to obtain the desired projections. We may explore this in a separate project.

5 Conclusion

In this report, we have proposed a semiparametric maximum likelihood estimation procedure for normal transformation models for bivariate failure time data. As the likelihood function involves infinite-dimensional parameters, we resort to modern asymptotic techniques to establish the asymptotic results. Specifically, we have shown that the SPMLEs are consistent, asymptotically normal and semiparametric efficient, under the semiparametric normal transformation model.

Reference

Keles, S., van der Laan, M. and Robins, J. (2004) Estimation of the Bivariate Survival Function with Generalized Bivariate Right Censored Data Structures. *Advances in Survival Analysis, (ed. N. Balakrishnan and C.R. Rao), Handbook of Statistics, Elsevier North Holland,* Chap. 8, 143-175.

Li, Y., and Lin, X. (2006) Semiparametric Normal Transformation Models for Spatially Correlated Survival Data. *Journal of the American Statistical Association,* 101, 591-603.

Lin, D. and Ying, Z. (1993) A simple nonparametric estimator of the bivariate survival function under univariate censoring. *Biometrika,* 80, 573-581.

Murphy, S. A. (1994) Consistency in a Proportional Hazards Model Incorporating a Random Effect. *Annals of Statistics*, 22, 712-731.

Nan, B., Lin, X., Lisabeth, L.D., and Harlow, S.D. (2006) Piecewise Constant Cross-ration Estimation for Association of Age at a Marker Event and Age at Menopause. *Journal of the American Statistical Association,* 101, 65-77.

Robins, J. and Rotnitzky, A. (1992) Recovery of information and adjustment for dependent censoring using surrogate markers. *Aids Epidemiology, Methodological issues*, Birkhauser.

Tsai, W. and Crowley, J. (1998) A Note on Nonparametric Estimators of the Bivariate Survival Function Under Univariate Censoring. *Biometrika*, 85, 573-580.

van der Laan, M., Hubbard, A.E., and Robins, J. (2002) Locally Efficient Estimation of a Multivariate Survival Function in Longitudinal Studies. *Journal of the American Statistical Association,* 97. 494-508.

van der Vaart, A. W., and Wellner, J. A. (1996) *Weak convergence and empirical processes.* Springer-Verlag: New York.

Wang, W. and Wells, M. (1998) Nonparametric Estimators of the Bivariate Survival Function Under Simplified Censoring Conditions. *Biometrika*, 84, 863-880.

Zeng, D., Lin, D.Y. and Yin, G. (2005) Maximum Likelihood Estimation for the Proportional Odds Model with Random Effects. *Journal of the American Statistical Association,* 100, 470-483.