Semiparametric Maximum Likelihood Estimation in Normal Transformation Models for Bivariate Survival Data

YI LI

Department of Biostatistics

Harvard School of Public Health and Dana-Farber Cancer Institute

Ross L. Prentice

Fred Hutchinson Cancer Research Center and University of Washington

Xihong Lin

Department of Biostatistics

Harvard School of Public Health

Abstract

We consider a class of semiparametric normal transformation models for right censored bivariate failure times. Specifically, nonparametric hazard rate models are transformed to a standard normal model and a joint normal distribution is assumed for the bivariate vector of transformed variates. A semiparametric maximum likelihood estimation (SPMLE) procedure is developed for estimating the marginal survival distribution and the pairwise correlation parameters. This model and its SPMLE estimation procedure are advantageous. First, the proposed SPMLE produces an efficient estimator of the correlation parameter of the semiparametric normal transformation model, which characterizes the bivariate dependence of bivariate survival outcomes. Secondly, a simple *positive-mass-redistribution algorithm* can be used to implement the SPMLE procedures. On the theoretical aspect, since the likelihood function involves infinite-dimensional parameters, this paper utilizes the empirical process theory to study the asymptotic properties of the proposed estimator. The SPMLEs are shown to be consistent, asymptotically normal and semiparametric efficient. A simple estimator for the variance of the estimates is also derived. The finite sample performance is evaluated via extensive simulations.

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

RUNNING TITLE: Bivariate Normal Transformation Models.

1 Introduction

The development of methods for the analysis of censored bivariate failure times is an essential component of multivariate survival analysis as it typically leads to representations that generalize readily to higher dimensions. Bivariate data are often of substantive interest in their own right with well-known examples including the Danish Twin Study (see, Wienke et al., 2002), the diabetic retinopathy study (Hougaard, 2000), the dual infection kidney dialysis study (Van Keilegom and Hettmansperger, 2002), and the reproductive health study of the association of age at a marker event and age at menopause (Nan et al., 2006). In all these studies, the assessment of marginal distribution as well as dependence among dependent individuals (e.g. twins) is of major interest, the latter because it renders genetic information.

Few existing bivariate distributions for non-negative random variables accommodate semiparametric specifications of marginal distribution and unrestricted pairwise dependence. Consider Clayton's (1978) model for a pair of survival times $(\tilde{T}_1, \tilde{T}_2)$

$$S(\tilde{t}_1, \tilde{t}_2) = [\max\{S_1(\tilde{t}_1)^{-\theta} + S_2(\tilde{t}_2)^{-\theta} - 1, 0\}]^{-\theta^{-1}}, \tag{1}$$

where $S(\tilde{t}_1, \tilde{t}_2) = P(\tilde{T}_1 > \tilde{t}_1, \tilde{T}_2 > \tilde{t}_2), S_1(\tilde{t}_1) = S(\tilde{t}_1, 0^-), S_2(\tilde{t}_2) = S(0^-, \tilde{t}_2)$ are bivariate survival and marginal survival functions respectively, and θ has an interpretation as cross ratio (Oakes, 1989) and also corresponds to other dependence measures such as Kendall's *tau*. This model allows for negative dependence when $-1 < \theta < 0$. But for random variables $\tilde{T}_1$ and $\tilde{T}_2$ which are marginally absolutely continuous (w.r.t say Lebesgue measure μ), the joint distribution of $(\tilde{T}_1, \tilde{T}_2)$ is absolutely continuous (w.r.t the product Lebesgue measure $\mu \times \mu$) only when $\theta > -0.5$. When $\theta \leq -0.5$, Oakes (1989) noticed that the distribution is no longer absolutely continuous, but has a mass along the curve given by $\{(\tilde{t}_1, \tilde{t}_2) : S_1(\tilde{t}_1)^{-\theta} + S_2(\tilde{t}_2)^{-\theta} - 1 = 0\}$. Hougaard (2000) further noted that frailty models cannot yield unrestricted marginal distributions with unrestricted pairwise parameters.

Hence it will be of substantial interest to specify a semiparametric likelihood model that allows for arbitrary modeling of the marginal survival functions, that allows for a flexible and interpretable correlation structure, and that retains a likelihood so that an efficient and simple estimating procedure is possible. For this purpose, we study a class of semiparametric normal transformation models for right censored bivariate failure times. Specifically, nonparametric marginal hazard rate models are transformed to a standard normal model and a joint normal distribution is imposed on the bivariate vector of transformed variates. The induced joint distribution is closely related to the normal copula model developed by, e.g., Klaassen and Wellner (1997) and Pitt, Chan, Kohn (2006). However, all the previous efforts in normal copula focused only on non-censoring situations, and it is unclear whether these existing results can be generalized to censoring situations.

This paper is motivated by a recent work of Li and Lin (2006) on spatial survival data. Li and Lin (2006) only considered estimating equation approaches in spatial settings and their estimators are not efficient under the bivariate normal transformation model. In contrast, we focus this paper on semiparametric likelihood based inference for bivariate survival data. Our major contributions are: (i) we propose a semiparametric efficient survivor function estimator under semiparametric normal transformation model for censored survival data and study its asymptotic properties. This work fills the gap of a lack of semiparametric efficient survivor function estimator for normal copula models. For example, Klaassen and Wellner's (1997) estimator only handles non-censored data and is not efficient for estimating marginal survivals. (ii) Our work manifests the potential to improve the efficiency of marginal survivor function estimators for one variable based on its dependent pair member, which leads to an easily implementable algorithm. Currently, there is no practical method for doing so. (iii) We perform extensive simulations to examine robustness to departures from bivariate normal transformation models, and compare it with a double- robust estimator. (iv) Our bivariate framework sets up a theoretical stage for general regression extensions, which will come in a subsequent communication.

2 Semiparametric Normal Transformation Models

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

$$T_j = \Phi^{-1}\left\{1 - e^{-\Lambda_j(\tilde{T}_j)}\right\} \tag{2}$$

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

$$\Psi(z_1, z_2; \rho) = \int_{z_1}^{\infty} \int_{z_2}^{\infty} \phi(x_1, x_2; \rho) dx_1 dx_2 \tag{3}$$

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

$$S(\tilde{t}_1, \tilde{t}_2; \rho) = P(\tilde{T}_1 > \tilde{t}_1, \tilde{T}_2 > \tilde{t}_2; \rho) = \Psi[\Phi^{-1}\{F_1(\tilde{t}_1)\}, \Phi^{-1}\{F_2(\tilde{t}_2)\}; \rho] \tag{4}$$

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

$$f(\tilde{t}_1, \tilde{t}_2; \rho) = f_1(\tilde{t}_1) f_2(\tilde{t}_2) e^{g(t_1, t_2; \rho)} \tag{5}$$

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

$$g(t_1, t_2; \rho) = -0.5 \log(1 - \rho^2) - 0.5(1 - \rho^2)^{-1}(\rho^2 t_1^2 + \rho^2 t_2^2 - 2\rho t_1 t_2). \tag{6}$$

It is obvious that $\rho = 0$ results in $f(\tilde{t}_1, \tilde{t}_2; \rho = 0) = f_1(\tilde{t}_1) f_2(\tilde{t}_2)$, corresponding to the independent case. One can easily show that the bivariate survival function approaches the upper Fréchet bound $\min\{S_1(\tilde{t}_1), S_2(\tilde{t}_2)\}$ as $\rho \to 1^-$, and the lower Fréchet bound $\max\{S_1(\tilde{t}_1) + S_2(\tilde{t}_2) - 1, 0\}$ as $\rho \to -1^+$. Indeed, the correlation parameter ρ provides a summary measure for the pairwise dependence, whose connection with the other commonly used dependence measures, including Kendall's *tau*, Spearman's *rho* and the cross ratio, can be found in Li and Lin (2006). Also of interest is to note that (5) can be rewritten as

$$\frac{f(\tilde{t}_1 | \tilde{t}_2)}{f(\tilde{t}_1 | \mathcal{O})} = \frac{f(\tilde{t}_2 | \tilde{t}_1)}{f(\tilde{t}_2 | \mathcal{O})} = e^{g(t_1, t_2; \rho)},$$

where $f(\cdot|\cdot)$ denotes a conditional density function and $\mathcal{O}$ is the empty set. Hence, function g or ρ also has interpretations of a Bayes factor for a dependence model against an independence model.

We are now in a position to consider estimation 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 $\tilde{X}_{ij} = \tilde{T}_{ij} \wedge \tilde{U}_{ij} \stackrel{def}{=} min(\tilde{T}_{ij}, \tilde{U}_{ij}), \delta_{ij} = I(\tilde{T}_{ij} \leq \tilde{U}_{ij})$, 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 factored into the product of contributions from the survival and censoring times, facilitating likelihood-based inferential procedures.

In some 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 §3, followed by the case of distinct marginal cumulative hazards $\Lambda_1 \not\equiv \Lambda_2$ in §4.

3 Semiparametric Maximum Likelihood Estimation With A Common Marginal Cumulative Hazard

3.1 The Likelihood Function

This section proposes a semiparametric maximum likelihood estimation (SPMLE) procedure for the semiparametric normal transformation model with a common marginal cumulative hazard, say, Λ. We define the normally transformed observed time $X_{ij} = \Phi^{-1}\{1 - \exp(-\Lambda(\tilde{X}_{ij})\}$ for $j = 1, 2$. As this transformation is monotone, it can easily accommodate right censored data as the transformed outcome (X_{ij}, δ_{ij}) contains the same information as the original $(\tilde{X}_{ij}, \delta_{ij})$, facilitating the derivation of a likelihood function that can be factored into the product of contributions from the survival and censoring times. It follows that the likelihood function for the unknown parameters (Λ, ρ) can be written, up to a constant, as the product of factors $(i = 1, \ldots, m)$

$$\begin{aligned} \tilde{L}_i(\rho, \Lambda) &= \{e^{g(X_{i1}, X_{i2}; \rho)} \Lambda'(\tilde{X}_{i1}) \Lambda'(\tilde{X}_{i2}) e^{-\Lambda(\tilde{X}_{i1}) - \Lambda(\tilde{X}_{i2})}\}^{\delta_{i1}\delta_{i2}} \{\Psi_1(X_{i1}, X_{i2}; \rho) \Lambda'(\tilde{X}_{i1}) e^{-\Lambda(\tilde{X}_{i1})}\}^{\delta_{i1}(1-\delta_{i2})} \\ &\quad \times \{\Psi_2(X_{i1}, X_{i2}; \rho) \Lambda'(\tilde{X}_{i2}) e^{-\Lambda(\tilde{X}_{i2})}\}^{(1-\delta_{i1})\delta_{i2}} \times \{\Psi(X_{i1}, X_{i2}; \rho)\}^{(1-\delta_{i1})(1-\delta_{i2})}, \end{aligned} \tag{7}$$

where $\Psi_j(x_1, x_2; \rho) = -\frac{\partial}{\partial x_j} \Psi(x_1, x_2; \rho) / \phi(x_j)$ 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 of the empirical likelihood function $L(\rho, \Lambda)$, which is the product of terms (7) with $\Lambda'(\cdot)$ replaced by $\Delta\Lambda(\cdot)$. We denote the log empirical likelihood function by $\ell(\rho, \Lambda) = \log L(\rho, \Lambda)$.

3.2 Theoretical Properties of the SPMLE

The main results of the paper are proved under the following set of regularity conditions. Namely, (c.1) (Boundedness) ρ lies in an 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, $\Lambda(\tau) < \infty$.

Condition (c.1) is assumed to ensure the existence and consistency of the estimators. A similar boundedness condition on the frailty parameter was assumed by Murphy (1994) in the context of frailty models for the same purpose. Condition (c.2) ensures that the failures for both pair members can be observed over a finite interval $[0, \tau]$, entailing an estimate of the hazard over $[0, \tau]$. Condition (c.3) implies absolute continuity of the cumulative hazard, which is useful in the consistency proof, and that we can work with the supremum norm on the space of cumulative hazard functions. Also, condition (c.3) guarantees the identifiability of the semiparametric normal transformation model specified in (2) and (3).

Under conditions (c.1)-(c.3), we show in our technical report that the SPMLEs do exist and are finite. Furthermore, the next two Propositions indicate 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. The proofs can be found in our technical report (Li, Prentice and Lin, 2006).

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.*

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 $[0,\tau]$.*

Proposition 2 is of significance as it implies that both $\hat{\rho}$ and $\hat{\Lambda}(t)$ (and, hence, the estimator of the marginal survival) are asymptotically efficient by taking $\eta(s) = I(s \leq t)$ for any $t \in [0,\tau]$. It further implies that the infinite dimensional parameter, $\Lambda(\cdot)$, can be treated in the same fashion as the finite dimensional correlation parameter ρ. Hence the asymptotic covariance matrix can be estimated by inverting the observed information matrix. Specifically, for any constant h_1 and any function h_2 of bounded variation, the asymptotic covariance of

$$h_1\hat{\rho} + \int_0^{\tau} h_2(s) d\hat{\Lambda}(s) = h_1\hat{\rho} + \sum_{\{(i,j):\delta_{ij}=1\}} h_2(\tilde{X}_{ij}) \Delta\hat{\Lambda}(\tilde{X}_{ij}) \tag{8}$$

can be estimated by $\hat{h}'\hat{J}^{-1}\hat{h}$, where $\hat{h}$ is a column vector comprising of h_1 and $h_2(\tilde{X}_{ij})$ for which $\delta_{ij} = 1$, and $\hat{J}$ is the negative Hessian matrix of $\ell(\rho,\Lambda)$ with respect to ρ and the jump size of Λ at $\tilde{X}_{ij}$ when $\delta_{ij} = 1$. More formally, it can be shown that $m\hat{h}'\hat{J}^{-1}\hat{h} \xrightarrow{p} V(h_1, h_2)$ as $m \to \infty$, where $V(h_1, h_2)$ is the asymptotic variance of $\sqrt{m}\{h_1\hat{\rho} + \int_0^{\tau} h_2(s) d\hat{\Lambda}(s)\}$. The justification follows the proof of Theorem 3 in Parner (1998), who argued that the empirical information operator based on $\hat{J}$ approximates the true invertible information operator. We will evaluate the finite sample performance of this variance estimator in the simulation section.

3.3 A Positive-Mass-Redistribution Algorithm

Consider the following computationally efficient procedure for obtaining the SPMLEs and their variance. Since $\tilde{T}_1$ and $\tilde{T}_2$ have the same distribution function, whose estimator has masses at the distinct failure times of $\tilde{T}_1$ and $\tilde{T}_2$, we denote by $t_1 < \ldots < t_K$ the K distinct, ordered and pooled $\tilde{T}_1$ and $\tilde{T}_2$ failure times. Define $r(t_1, t_2) = \#\{l | \tilde{X}_{1l} \geq t_1, \tilde{X}_{2l} \geq t_2\}$ as the size of the risk set at (t_1, t_2) and let $\tilde{R} = \{(t_1, t_2) | r(t_1, t_2) > 0\}$ denote the risk region. We focus on the square grids $\{(t_1, t_2) | t_1 = t_i, t_2 = t_j, 1 \leq i \leq K, 1 \leq j \leq K\}$ formed by the observed $\tilde{T}_1$ and $\tilde{T}_2$ pooled failure times. This is due to the fact that the censored values in $\tilde{T}_1$ (or $\tilde{T}_2$) in the sample can be replaced by censored values at the immediately smaller $\tilde{T}_1$ and $\tilde{T}_2$ pooled uncensored failure time, or by zero if there are no corresponding smaller uncensored times, without affecting the log empirical likelihood $\ell(\rho, \Lambda)$. We term such replacement as *positive mass redistribution.* Then let $n_{ij}^{\delta_1\delta_2} = \#\{l|\tilde{X}_{1l} = t_i, \tilde{X}_{2l} = t_j, \delta_{1l} = \delta_1, \delta_{2l} = \delta_2\}$ for $\delta_1, \delta_2 \in \{0,1\}$ and for $0 \leq i \leq K$ and $0 \leq j \leq K$, with $t_0 = 0$. Also denote $f_{ij} = f(t_i, t_j)$, $F_i = \prod_{l=1}^{i}(1-\lambda_l), F_i^- = F_{i-1}$ for $i = 1, \ldots, K$, where $\lambda_l = \Delta\Lambda(t_l)$. The log empirical likelihood $\ell(\rho, \Lambda)$ defined in §3.1 can now be written

$$\begin{aligned}\ell = \sum_{i=1}^{K}\sum_{j=1}^{K} n_{ij}^{11}\log f_{lm} \quad &+ \quad n_{ij}^{10}\log\{\lambda_i F_i^- - \sum_{v=1}^{j} f_{iv}\} + n_{ij}^{01}\log\{\lambda_j F_j^- - \sum_{u=1}^{i} f_{uj}\} \\ &+ \quad n_{ij}^{00}\log\{F_i + F_j + \sum_{u=1}^{i}\sum_{v=1}^{j} f_{uv} - 1\}, \end{aligned} \tag{9}$$

which involves only the marginal hazard rates at uncensored $\tilde{T}_1$ and $\tilde{T}_2$ times, and the joint density at grid points in the risk region. The latter can be rewritten as

$$f_{ij} = \lambda_i F_i^- \lambda_j F_j^- e^{g(s_i, s_j; \rho)}$$

with $g(\cdot,\cdot)$ defined in (6) and $s_i = \Phi^{-1}(1 - F_i)$. To ensure numerical stability and avoid arguments of 0 for Φ^{-1} in computation in finite samples, we use an asymptotically equivalent transformation $s_i = \Phi^{-1}\{1 - (1 - \frac{1}{m})F_i\}$. A simple Newton-Raphson procedure, starting with $\rho = 0$ and the Kaplan-Meier marginal hazard rates derived by treating $(\tilde{X}_{ij}, \delta_{ij}), j = 1, 2, i = 1, \ldots, m$, as $2m$ independent observations, can be used to compute the SPMLEs $\hat{\lambda}_i, \hat{\rho}$. These calculations are less computationally demanding as they do not require the evaluation of bivariate incomplete normal integrals, only the evaluation of the univariate Φ^{-1}.

Following the arguments in §3.2, the variability of $\hat{\lambda}_i, \hat{\rho}$ can be assessed by inverting the negative Hessian matrix of (9), denoted by $\hat{J}$ [a $(K+1) \times (K+1)$ matrix]. Furthermore, the functional (8) can be rewritten as a linear combination of $\hat{\lambda}_i, \hat{\rho}$, namely,

$$h_1\hat{\rho} + \sum_{i=1}^{K} h_2(t_i)\hat{\lambda}_i,$$

whose variance function can be easily computed by $\hat{h}'\hat{J}^{-1}\hat{h}$, where $\hat{h} = \{h_1, h_2(t_1), \ldots, h_2(t_K)\}'$. We can easily apply this result to estimate the variance of the estimate of a survival probability.

For example, the common marginal survival $S(u_0) = P(\tilde{T}_1 > u_0)$ at any given time $u_0 \in [0, \tau]$ can be estimated by $\hat{S}(u_0) = e^{-\hat{\Lambda}(u_0)}$. With a first order Taylor expansion,

$$\hat{S}(u_0) - S(u_0) \doteq -S(u_0)\{\hat{\Lambda}(u_0) - \Lambda(u_0)\} = \int_0^\tau -S(u_0)I(s \leq u_0)d\hat{\Lambda}(s) + const$$

Hence, $\hat{S}(u_0)$ can be approximated by the functional form (8) with $h_1 = 0$ and $h_2(s) = -S(u_0)I(s \leq u_0)$ and applying the above results will render a consistent estimator of the variance of $\hat{S}(u_0)$ as $\hat{S}^2(u_0)\hat{e}'\hat{J}^{-1}\hat{e}$, where $\hat{e} = \{0, I(t_1 \leq u_0), \ldots, I(t_K \leq u_0)\}'$.

4 SPMLE for the Stratified Hazard Model

So far we have considered the cases where the pair members are identically distributed. But in many applications, it would be unnatural to assume a common distribution or hazard for each member of the pair, for example, when considering husband-wife pairs or proband-control pairs. 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 separate cumulative hazard function $\Lambda_j(\cdot), j = 1, 2$.

4.1 The SPMLE and Its Theoretical Properties

We consider 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

$$\begin{aligned} & \tilde{L}_i(\rho, \Lambda_1, \Lambda_2) \\ = \; & \{e^{g(X_{i1}, X_{i2}; \rho)}\Lambda_1'(\tilde{X}_{i1})\Lambda_2'(\tilde{X}_{i2})e^{-\Lambda_1(\tilde{X}_{i1}) - \Lambda_2(\tilde{X}_{i2})}\}^{\delta_{i1}\delta_{i2}}\{\Psi_1(X_{i1}, X_{i2}; \rho)\Lambda_1'(\tilde{X}_{i1})e^{-\Lambda_1(\tilde{X}_{i1})}\}^{\delta_{i1}(1-\delta_{i2})} \\ & \times\{\Psi_2(X_{i1}, X_{i2}; \rho)\Lambda_2'(\tilde{X}_{i2})e^{-\Lambda_2(\tilde{X}_{i2})}\}^{(1-\delta_{i1})\delta_{i2}} \times \{\Psi(X_{i1}, X_{i2}; \rho)\}^{(1-\delta_{i1})(1-\delta_{i2})}. \end{aligned} \tag{10}$$

Here $X_{ij} = \Phi^{-1}\{1 - \exp(-\Lambda_j(\tilde{X}_{ij})\}$ for $j = 1, 2$. Again, directly maximizing the likelihood function (10) 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):$\{(\Lambda_1, \Lambda_2) : \Lambda_1, \Lambda_2$ are cadlag and piecewise constant$\}$. It follows that the SPMLE, $(\hat{\rho}, \hat{\Lambda}_1, \hat{\Lambda}_2)$, is the maximizer of the empirical likelihood function $\ell(\rho, \Lambda_1, \Lambda_2)$, which is obtained from (10) 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. We can show 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, the proofs of which can be found in the technical report (Li, Prentice and Lin, 2007).

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.*

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* $[0,\tau]$.

As in the case of a common hazard model, the asymptotic covariance matrix of the estimators of the unknown (finite dimensional and infinite dimensional) parameters can be estimated by inverting the observed information matrix. Specifically, for any constant h_1 and any function h_2 and h_3 of bounded variation, the asymptotic covariance of

$$h_1\hat{\rho}+\int_0^\tau h_2(s)d\hat{\Lambda}_1(s)+\int_0^\tau h_3(s)d\hat{\Lambda}_2(s) = h_1\hat{\rho}+\sum_{\{i:\delta_{i1}=1\}} h_2(\tilde{X}_{i1})\Delta\hat{\Lambda}(\tilde{X}_{i1})+\sum_{\{i:\delta_{i2}=1\}} h_2(\tilde{X}_{i2})\Delta\hat{\Lambda}_2(\tilde{X}_{i2}) \tag{11}$$

can be estimated by $\hat{h}'\hat{J}^{-1}\hat{h}$, where $\hat{h}$ is a column vector comprising of h_1, the $h_2(\tilde{X}_{i1})$ for which $\delta_{i1} = 1$ and the $h_2(\tilde{X}_{i2})$ for which $\delta_{i2} = 1$, and $\hat{J}$ is the negative Hessian matrix of $\ell(\rho, \Lambda_1, \Lambda_2)$ with respect to ρ and the jump sizes of Λ_j at $\tilde{X}_{ij}$ when $\delta_{ij} = 1$. Indeed, following the proof of Theorem 3 in Parner (1998), one can show $m\hat{h}'\hat{J}^{-1}\hat{h} \xrightarrow{p} V(h_1, h_2, h_3)$ as $m \to \infty$, where $V(h_1, h_2, h_3)$ is the asymptotic variance of $\sqrt{m}\{h_1\hat{\rho} + \int_0^\tau h_2(s)d\hat{\Lambda}_1(s) + \int_0^\tau h_3(s)d\hat{\Lambda}_2(s)\}$.

4.2 Practical Implementation of the SPMLE Procedure

We develop in this section a simple procedure to implement the SPMLE procedure for the stratified hazard model. Denote by $t_{11} < \ldots < t_{1I}$ the I distinct ordered $\tilde{T}_1$-failure times and by $t_{21} < \ldots < t_{2J}$ the J distinct $\tilde{T}_2$-failure times in the observed sample. As defined in §3.2, let $r(t_1, t_2)$ be the size of the risk set at (t_1, t_2) and let $\tilde{R}$ be the risk region. We only consider the rectangular grids $\{(t_1, t_2) | t_1 = t_{1i}, t_2 = t_{2j}, 1 \le i \le I, 1 \le j \le J\}$ formed by the observed $\tilde{T}_1$ and $\tilde{T}_2$ failure times. This is because the censored values in $\tilde{T}_1$ (or $\tilde{T}_2$) in the sample can be replaced by censored values at the immediately smaller $\tilde{T}_1$ (or $\tilde{T}_2$) uncensored failure time, or by zero if there no corresponding smaller uncensored times, without affecting the empirical likelihood $\ell(\rho, \Lambda_1, \Lambda_2)$. With these replacements (or so-called *positive-mass-redistributions*), let $n_{ij}^{\delta_1\delta_2} = \#\{l | \tilde{X}_{1l} = t_{1i}, \tilde{X}_{2l} = t_{2j}, \delta_{1l} = \delta_1, \delta_{2l} = \delta_2\}$ for $\delta_1, \delta_2 \in \{0, 1\}$ and for $0 \le i \le I$ and $0 \le j \le J$, with $t_{10} = t_{20} = 0$. Also denote $f_{ij} = f(t_{1i}, t_{2j})$, $F_{1i} = \prod_{l=1}^{i}(1 - \lambda_{1l}), F_{1i}^{-} = F_{1,i-1}, F_{2j} = \prod_{k=1}^{j}(1 - \lambda_{2k}), F_{2j}^{-} = F_{2,j-1}$, where $\lambda_{1l} = \Delta\Lambda_1(t_{1l}), \lambda_{2k} = \Delta\Lambda_2(t_{2k})$. The log- empirical likelihood function can now be written

$$
\begin{aligned}
\ell = \sum_{i=1}^{I}\sum_{j=1}^{J} n_{ij}^{11} \log f_{lm} \quad &+ \quad n_{ij}^{10} \log\{\lambda_{1i} F_{1i}^{-} - \sum_{v=1}^{j} f_{ij}\} + n_{ij}^{01} \log\{\lambda_{2j} F_{2j}^{-} - \sum_{u=1}^{i} f_{uj}\} \\
&+ \quad n_{ij}^{00} \log\{F_{1i} + F_{2j} + \sum_{u=1}^{i}\sum_{v=1}^{j} f_{uv} - 1\}, \qquad (12)
\end{aligned}
$$

which involves only the marginal hazard rates at uncensored T_1 and T_2 times, and the joint density at grid points in the risk region, namely,

$$
f_{ij} = \lambda_{1i} F_{1i}^{-} \lambda_{2j} F_{2j}^{-} e^{g(s_{1i}, s_{2j}; \rho)}
$$

with $s_{1i} = \Phi^{-1}(1 - F_{1i})$ and $s_{2j} = \Phi^{-1}(1 - F_{2j})$. In practice, to avoid arguments of 0 for Φ^{-1} in computation for a finite sample size, we use an asymptotically equivalent transformation $s_{1i} = \Phi^{-1}\{1 - (1 - \frac{1}{m})F_{1i}\}$ and $s_{2j} = \Phi^{-1}\{1 - (1 - \frac{1}{m})F_{2j}\}$. A simple Newton-Raphson procedure, starting with $\rho = 0$, and Kaplan-Meier marginal hazard rates $\lambda_{1i} = \sum_{j=1}^{J}(n_{ij}^{11} + n_{ij}^{10})/r(t_{1i}, 0), \lambda_{2j} = \sum_{i=1}^{I}(n_{ij}^{11} + n_{ij}^{01})/r(0, t_{2j})$, can be used to compute the SPMLEs $\hat{\lambda}_{1i}, \hat{\lambda}_{2j}, \hat{\rho}$. We again note that the likelihood evaluations are less computationally demanding, requiring only the computation of the univariate Φ^{-1}.

Similarly, the variability of $\hat{\lambda}_{1i}, \hat{\lambda}_{2j}, \hat{\rho}$ can be assessed by inverting the negative Hessian matrix of (12), denoted by $\hat{J}$ [a $(I+J+1) \times (I+J+1)$ matrix]. Moreover, the functional (11) can be rewritten as a linear combination of $\hat{\lambda}_{1i}, \hat{\lambda}_{2j}\hat{\rho}$, namely,

$$h_1\hat{\rho} + \sum_{i=1}^{I} h_2(t_{1i})\hat{\lambda}_{1i} + \sum_{j=1}^{J} h_3(t_{2j})\hat{\lambda}_{2j}$$

whose variance can be easily computed by $\hat{h}'\hat{J}^{-1}\hat{h}$, where $\hat{h} = \{h_1, h_2(t_{11}), \ldots, h_2(t_{1I}), h_3(t_{21}), \ldots, h_3(t_{2J}\}'$.

We now illustrate a practical usage of this variance formula. For example, consider the bivariate survival estimates of $S(u_0, v_0)$ at any given time $(u_0, v_0) \in [0, \tau]^2$, which can be obtained, based on the semiparametric normal transformation model, by

$$\hat{S}(u_0, v_0) = \Psi\left[\Phi^{-1}\left\{1 - e^{-\hat{\Lambda}_1(u_0)}\right\}, \Phi^{-1}\left\{1 - e^{-\hat{\Lambda}_2(v_0)}\right\}; \hat{\rho}\right].$$

To evaluate the variability of $\hat{S}(u_0, v_0)$, we perform a first order Taylor expansion yielding

$$\begin{aligned}
& \hat{S}(u_0, v_0) - S(u_0, v_0) \\
\doteq\ & \gamma_1(u_0, v_0)\{\hat{\rho} - \rho_0\} + \gamma_2(u_0, v_0)\{\hat{\Lambda}_1(u_0) - \Lambda_1(u_0)\} + \gamma_3(u_0, v_0)\{\hat{\Lambda}_2(v_0) - \Lambda_2(v_0)\} \\
=\ & \gamma_1(u_0, v_0)\hat{\rho} + \int_0^\tau \gamma_2(u_0, v_0) I(s \le u_0) d\hat{\Lambda}_1(s) + \int_0^\tau \gamma_3(u_0, v_0) I(s \le v_0) d\hat{\Lambda}_2(s) + const
\end{aligned}$$

where $\gamma_1(t_1, t_2) = \partial\Psi(x_1, x_2; \rho)/\partial\rho$, $\gamma_2(t_1, t_2) = -\Phi_1(x_1, x_2; \rho_0)\exp(-\Lambda_1(t_1))$, $\gamma_3(t_1, t_2) = -\Phi_2(x_1, x_2; \rho_0)\exp(-\Lambda_2(t_2))$ and $x_j = \Phi^{-1}\left\{1 - e^{-\Lambda_j(t_j)}\right\}$. Hence, $\hat{S}(u_0, v_0)$ can be approximated by the functional form (11) with $h_1 = \gamma_1(u_0, v_0)$, $h_2(s) = \gamma_2(u_0, v_0)I(s \le u_0)$, $h_3(s) = \gamma_3(u_0, v_0)I(s \le v_0)$, and applying the above variance formula will render a consistent estimate of the variance for $\hat{S}(u_0, v_0)$, namely, $\hat{h}'\hat{J}^{-1}\hat{h}$, where $\hat{h} = \{\hat{\gamma}_1(u_0, v_0), \hat{\gamma}_2(u_0, v_0)I(t_{11} \le u_0), \ldots, \hat{\gamma}_2(u_0, v_0)I(t_{1I} \le u_0), \hat{\gamma}_3(u_0, v_0)I(t_{21} \le u_0), \ldots, \hat{\gamma}_3(u_0, v_0)I(t_{2J} \le u_0)\}'$ and $\hat{\gamma}_j(\cdot, \cdot)$ is obtained from $\gamma_j(\cdot, \cdot)$, for $j = 1, 2, 3$, with all the unknown parameters replaced by their estimators. We will evaluate the finite sample performance of this variance estimator in the next simulation section.

5 Numerical Studies

A series of simulation studies were performed to examine the properties of the proposed estimator and to compare it with the existing bivariate survivor estimators, including the Prentice-Cai

(Prentice and Cai, 1992), Dabrowska (1988) and repaired Nonparametric MLE (van der Laan, 1996; Moodie et al, 2005) estimators. The simulation setup mimics those in Prentice et al. (2004). Specifically, the marginal distributions of $\tilde{T}_1$ and $\tilde{T}_2$ were specified as unit exponential. The censoring time $\tilde{U}_1$ was taken to be an exponential variate with mean 0.5 whereas three special cases for $\tilde{U}_2$ were considered: (i) $\tilde{U}_2 = \infty$, corresponding to no $\tilde{T}_2$ censoring; (ii) $\tilde{U}_2 = \tilde{U}_1$, corresponding to univariate censoring; (iii) $\tilde{U}_2$ is independent of $\tilde{U}_1$ and is an exponential variate with mean 0.5. A sample size of 120 (pairs) was considered with 1000 repetitions at a given configuration.

Finite Sample Performance Under the Correct Model: We began by evaluating the finite sample performance of the SPMLE when the true model follows the *semiparametric normal transformation model (4) with* $\rho = 0.5$. When calculating the SPMLE, we considered both the common hazard model and the stratified hazard model. As both models yield similar results, we only reported in Table 1 the summary simulation results for the common hazard model. As efficiently estimating the common hazard function or the common marginal distribution function is of major interest under the common hazard model, we reported only the estimates of the marginal survival function at various time points in Table 1. Our results showed that the sample averages of the estimates were very close to the true values, and the model-based standard errors, which were computed by applying the results of §3.3, matched very well with the empirical SEs.

Finite Sample Performance Under the Misspecified Model: We next considered the robustness of the SPMLE when the semiparametric normal transformation model was misspecified, and the failure times were generated under the following *bivariate Clayton model*

$$S(\tilde{t}_1, \tilde{t}_2) = \{S_1(\tilde{t}_1)^{-\theta} + S_2(\tilde{t}_2)^{-\theta} - 1\}^{-\theta^{-1}}, \tag{13}$$

with $\theta = 4$, implying a strong positive dependence between $\tilde{T}_1$ and $\tilde{T}_2$. We compared the performance of the SPMLE based on the semiparametric normal transformation model ($\hat{S}_{NT}$) with the other existing nonparametric estimators, including the Prentice-Cai estimator ($\hat{S}_{PC}$), empirical hazard rate estimator($\hat{S}_E$), redistributed empirical estimator($\hat{S}_{RE}$), which were taken from Tables 1 and 2 of Prentice et al. (2004). As Prentice et al. (2004) only considered stratified hazard models, we focused on the *stratified hazard model* to make the resulting estimates comparable.

The sample averages of the relative biases of the bivariate survival estimates and marginal survival estimates at selected time points and the average model-based standard errors (calculated by applying the results of §4.2) for the point estimates, along with the empirical standard errors are displayed in Tables 2 and 3. For the comparison purpose, we also list the summary statistics for the empirical hazard rate ($\hat{S}_E$), Prentice-Cai ($\hat{S}_{PC}$), redistributed empirical ($\hat{S}_{RE}$) estimators (see, Prentice et al., 2004). Finally, we computed the mean squared errors (MSEs), which are the sum of the square of the bias and the empirical variance, for all the estimators.

Our results show that, even when the underlying model is misspecified, the SPMLE based on the semiparametric normal transformation model incurred only small biases. Among all the scenarios examined, the relative biases, i.e. (point estimate -true value)/true value, of the semiparametric normal transformation model based SPMLE ranged from -5.7% to 4%. Compared to the competing non-parametric estimators, the semiparametric normal transformation model based estimator also achieved high efficiency and had the smallest standard errors in all most all the scenarios examined. Using the MSE as a measure of overall performance, the semiparametric normal transformation model based SPMLE had a smaller MSE than the other estimators in most cases considered. In addition, the model based standard errors were in a good agreement with their empirical counterparts.

Further Comparison with IPCW in Efficiency and Robustness: To further explore the efficiency and the robustness of the SPMLE, we also compared it with a double-robust IPCW (inverse-probability-of-censoring-weighted) estimator, derived under univariate censoring, which stipulates that the censoring time is common for both pair members (Lin and Ying, 1993; Tsai and Crowley, 1998; Wang and Wells, 1998; Nan et al., 2006). The detailed derivation can be found in our technical report (Li, Prentice and Lin, 2007). We first compared the efficiency of the IPCW estimator with the semiparametric normal transformation model based SPMLE when the true underlying model indeed followed the *semiparametric normal transformation model (4) with* $\rho = 0.5$. The results are documented in Table 4, which demonstrate that the normal transformation SPMLE has noticeably smaller variances than its IPCW counterparts. We next considered the robustness of the IPCW one-step estimator when the underlying model model was misspecified as the semiparametric normal transformation model, while the true model followed the *bivariate Clayton* *model (13) with* $\theta = 4$. The results are also reported in Table 4. Our results indicate that the IPCW estimator has eliminated the bias caused by the misspecification of the semiparametric normal transformation model, while the SPMLE estimator incurs negligible biases and retains smaller variances. Using the mean squared error (MSE) reported in Table 4 as the criterion, it appeared that the SPMLE estimator outperformed its IPCW counterpart in the scenarios considered.

Estimation of ρ: Finally, we discuss the interpretation and the estimation of ρ. We note that ρ has a one-one correspondence between the common dependence measure for bivariate survival, for example, Kendall's coefficient of concordance (Kendall's *tau*). We exemplify Kendall's *tau* as it is the most commonly used global measure for bivariate survival. As indicated in Li and Lin (2006), Kendall's *tau* can be evaluated by

$$tau = 4 \int_0^\infty \int_0^\infty f(\tilde{t}_1, \tilde{t}_2; \rho) S(\tilde{t}_1, \tilde{t}_2; \rho) d\tilde{t}_1 d\tilde{t}_2 - 1$$

where $S(\tilde{t}_1, \tilde{t}_2; \rho)$ and $f(\tilde{t}_1, \tilde{t}_2; \rho)$ are the joint bivariate survival and density functions defined in (4) and (5), respectively. By changing the variables, the notion above can be simplified to

$$tau(\rho) = 4 \int_0^\infty \int_0^\infty \Psi(t_1, t_2; \rho) e^{g(t_1, t_2; \rho)} \phi(t_1) \phi(t_2) dt_1 dt_2 - 1,$$

where $\Psi(\cdot)$ is the joint tail function for the bivariate normal distribution defined in (3), $g(t_1, t_2; \rho)$ is the 'cross' term defined in (6), and ϕ is the standard normal density function, none of which depends on any specific forms of hazard functions. As shown in Li and Lin (2006), ρ uniquely determines τ, and thus provides a standardized dependence measure for bivariate survival. Indeed, $tau(\hat{\rho})$ yields the model-based estimate of Kendall's *tau*, whose model-based standard error can be conveniently obtained using the delta method.

We considered the estimation and the interpretation of the estimate of ρ when the semiparametric normal transformation model was misspecified, and the failure times were generated under the following bivariate Clayton model (13). We varied θ to be 0.5, $1, 2$ and 4, which correspond to Kendal's tau of 0.199, 0.333, 0.5 and 0.613, respectively, using formula (4.4) of Hougaard (2000).

The sample averages of the estimates of ρ and the model-based Kendall's *tau*, along with the empirical as well as model-based standard errors are displayed in Table 5. It appeared that when the underlying model is misspecified, the estimate of ρ itself might not be of interest, as it would not recover the specific dependence structure of the true model. However, the model-based Kendall's *tau* using the estimates ρ were indeed very comparable to the true Kendall's *tau*, as they incurred very small biases when compared to the true Kendall's *tau* (based on the correct Clayton's model). We envision that, at least for the scenarios we considered, the estimate of ρ would lead to a reasonable approximation for Kendall's *tau* even when the model is misspecified.

6 Discussion

In this paper, we have proposed a class of semiparametric normal transformation models for bivariate failure time data. The theoretical properties of the semiparametric maximum likelihood estimation procedure in this model have been explored. We note that unlike the conventional bivariate survival models, e.g. the Clayton family, the correlation parameter in our proposed model can be unrestricted. Secondly, as opposed to the existing nonparametric estimating approaches for bivariate survival data, the proposed semiparametric MLE also produces an efficient estimate for the correlation parameter, which characterizes the bivariate dependence of survival pairs. Finally, 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. Computationally efficient algorithms have been developed to implement the inference procedures. Our simulation studies have shown that the SPMLEs are more efficient than the existing nonparametric bivariate survival estimator under the semiparametric normal transformation model, and have good robustness to departure from these modeling assumptions and generally better efficiency in MSEs compared to their nonparametric competitors. Also as commented by a reviewer, when $\rho = 0$, the likelihoods in (7) (for the unstratified model) and in (10) (for the stratified model) reduce to the nonparametric likelihood for independent survival data. As a result, the MLE that maximizes (7) or (10) indeed reduces to the Kaplan-Meier estimator.

With the analytical framework established in this article, our future work lies in extending the results to multivariate data, where clusters are allowed to have varying cluster sizes and where each pair of failure times may have a distinct correlation parameter. A key feature of this transformation model is that it can easily accommodate covariates in such a way that survival outcomes marginally follow a common Cox proportional hazard model, and their joint distribution is specified by a joint normal distribution. Hence, the regression coefficients have population level interpretations, a feature not shared by conditional frailty models.

Acknowledgements

This work was supported in part by the US NIH grants (Li, R01CA95747; Prentice, P01CA53996; Lin, R37CA76404). The authors thank Dr. Titterington (Editor) and the AE for their relevant expertise and insightful suggestions, which significantly improved the presentation of this work.

Reference

Hougaard, P. (2000) *Analysis of Multivariate Survival Data.* New York: Springer-Verlag.

van Keilegom, I. and Hettmansperger, T.P. (2002) Inference on multivariate M-estimators based on bivariate censored data. *J. Amer. Statist. Assoc.,* 97, 328-336.

Klaassen, C. A. J. and Wellner, J. A. (1997) Efficient estimation in the bivariate normal copula model: Normal margins are least favourable. *Bernoulli,* 3, 55-77.

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

Li, Y., Prentice, R. and Lin, X. (2007) Asymptotic Properties of Maximum Likelihood Estimator in Semiparametric Normal Transformation Models for Bivariate Survival Data. Department of Biostatistics, Harvard University, Technical Report.
http://biowww.dfci.harvard.edu/∼yili/bikaproof.pdf

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

Moodie, F.Z., and Prentice, R.L. (2005) An Adjustment to Improve the Bivariate Survivor Function Repaired NPMLE. *Lifetime Data Analysis*, 11, 291-307.

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.

Oakes, D. (1989) Bivariate survival models induced by frailties. *Journal of the American Statistical Association*, 84, 487-493.

Parner, E. (1998) Asymptotic theory for the correlated gamma-frailty model. *Ann Statist*, 26, 183-214.

Pitt, M., Chan, D. and Kohn, R. (2006) Efficient Bayesian inference for Gaussian copula regression models. *Biometrika,* 93, 537-554.

Prentice, R. L., and Cai, J. (1992) Covariance and survivor function estimation using censored multivariate failure time data. (Corr: 93V80 p.711-712) *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.

Prentice, R, L., Moodie, Z. F., and Wu, J. (2004) Hazard-based nonparametric survivor function estimation. *Journal of the Royal Statistical Society, Series B*, 66, 305-319.

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. (1996) Efficient estimation in the bivariate censoring model and repairing NPMLE *The Annals of Statistics,* 24, 596-627.

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

Wienke, A., Lichtenstein, P., and Yashin, A. (2003) A Bivariate Frailty Model with a Cure Fraction for Modeling Familial Correlations in Diseases. *Biometrics,* 59, 1178-1183.

Table 1: Averages and model based and empirical SEs of the SPMLEs under the **semiparametric normal transformation model (4) with a common hazard function.** *The true values are:* $\rho = 0.5, S(0.1625) = 0.85, S(0.3566) = 0.70, S(0.5978) = 0.55$. *$SE_e$ and SE_m are empirical and model based standard errors, respectively.*

				t=0.1625			t=0.3566		
Censoring	ρ	SE_e	SE_m	$\hat{S}(t)$	$\mathrm{SE_e}$	$\mathrm{SE_m}$	$\hat{S}(t)$	$\mathrm{SE_e}$	$\mathrm{SE_m}$
Censoring on T_1 only	0.502	0.109	0.104	0.842	0.023	0.023	0.692	0.035	0.036
Univariate censoring	0.503	0.122	0.114	0.842	0.025	0.026	0.694	0.035	0.035
Bivariate censoring	0.493	0.141	0.138	0.846	0.027	0.028	0.697	0.042	0.040

	t=0.5978		
Failure Model	$\hat{S}(t)$	$\mathrm{SE_e}$	$\mathrm{SE_m}$
censoring on T_1 only	0.546	0.044	0.046
univariate censoring	0.546	0.051	0.053
bivariate censoring	0.555	0.055	0.049

Table 2: Averages, SEs and mean squared errors (MSE) for various bivariate survival function estimators at various time pairs (t_1, t_2) when the **correct model (13) with $\theta = 4$ is misspecified to the semiparametric normal model (4)**. *$\hat{S}_E, \hat{S}_{PC}, \hat{S}_{RE}$ and $\hat{S}_{NT}$ are the empirical hazard rate, Prentice-Cai, redistributed empirical, and semiparametric normal transformation based SPMLE estimator respectively. SE are empirical standard errors, while for $\hat{S}_{NT}$ the model based standard errors are displayed inside the brackets. The true bivariate survival probabilities at these pairs are 0.771, 0.666, 0.608, 0.516 and 0.468, respectively.*

		(t_1, t_2) =(0.1625, 0.1625)			(t_1, t_2) =(0.1625, 0.3566)			(t_1, t_2) =(0.3566, 0.3566)		
Censoring		bias	SE	MSE ($\times 10^{-3}$)	bias	SE	MSE ($\times 10^{-3}$)	bias	SE	MSE ($\times 10^{-3}$)
Censoring on	$\hat{S}_E$	0.0%	0.046	2.1	0.3%	0.055	3.0	0.0%	0.057	3.2
T_1 only	$\hat{S}_{PC}$	0.1%	0.040	1.6	0.1%	0.043	1.9	-0.1%	0.047	2.2
	$\hat{S}_{RE}$	0.1%	0.043	1.8	0.3%	0.046	2.1	0.0%	0.049	2.4
	$\hat{S}_{NT}$	3.2%	0.035	1.7	4.4%	0.035	2.0	3.3%	0.043	2.2
			(0.033)			(0.030)			(0.038)	
Univariate	$\hat{S}_E$	0.0%	0.051	2.6	0.3%	0.059	3.5	0.1%	0.063	4.0
censoring	$\hat{S}_{PC}$	0.1%	0.041	1.7	0.3%	0.049	2.4	0.0%	0.051	2.6
	$\hat{S}_{RE}$	0.1%	0.048	2.3	0.3%	0.057	3.2	0.0%	0.058	3.4
	$\hat{S}_{NT}$	2.6%	0.032	1.4	3.3%	0.039	2.0	2.1%	0.042	1.9
			(0.028)			(0.037)			(0.042)	
Bivariate	$\hat{S}_E$	0.0%	0.058	3.4	0.3%	0.073	5.3	-0.1%	0.077	5.9
censoring	$\hat{S}_{PC}$	-0.1%	0.041	1.7	-0.1%	0.049	2.4	-0.3%	0.053	2.8
	$\hat{S}_{RE}$	-0.2%	0.056	3.1	0.3%	0.066	4.4	-0.3%	0.068	4.6
	$\hat{S}_{NT}$	2.2%	0.035	1.5	1.9%	0.042	1.9	0.1%	0.047	2.2
			(0.033)			(0.046)			(0.048)	

		(t_1, t_2) =(0.3566 0.5978)			(t_1, t_2) =(0.5978, 0.5978)		
Censoring		bias	SE	MSE ($\times 10^{-3}$)	bias	SE	MSE ($\times 10^{-3}$)
Censoring on T_1 only	$\hat{S}_E$	1.0%	0.065	4.2	0.6%	0.067	4.5
	$\hat{S}_{PC}$	0.6%	0.046	2.1	0.2%	0.053	2.8
	$\hat{S}_{RE}$	0.8%	0.048	2.3	0.4%	0.054	2.9
	$\hat{S}_{NT}$	2.9%	0.042	2.0	0.1%	0.055	3.0
			(0.044)			(0.051)	
Univariate censoring	$\hat{S}_E$	1.0%	0.070	4.9	0.9%	0.073	5.3
	$\hat{S}_{PC}$	0.8%	0.062	3.8	0.4%	0.064	4.1
	$\hat{S}_{RE}$	0.8%	0.066	4.4	0.6%	0.068	4.6
	$\hat{S}_{NT}$	0.8%	0.049	2.4	-3.6%	0.064	4.3
			(0.049)			(0.061)	
Bivariate censoring	$\hat{S}_E$	0.2%	0.096	9.2	-0.2%	0.102	10.4
	$\hat{S}_{PC}$	-0.8%	0.062	3.8	-0.8%	0.065	4.2
	$\hat{S}_{RE}$	-0.4%	0.068	4.6	-2.7%	0.071	5.0
	$\hat{S}_{NT}$	0.1%	0.056	3.1	-5.7%	0.058	4.0
			(0.059)			(0.059)	

Table 3: Averages and SEs for various estimators of the marginal survival functions S_1 and S_2 when the **correct model (13) with $\theta = 4$ is misspecified to the semiparametric normal model (4)**. *$\hat{S}_E, \hat{S}_{KM}, \hat{S}_{RE}$ and $\hat{S}_{NT}$ are empirical hazard rate, Kaplan-Meier, redistributed empirical, and semiparametric normal transformation based SPMLE estimators, respectively. SE are empirical standard errors, while for $\hat{S}_{NT}$ the model based standard errors are displayed inside the brackets. The true marginal survival probabilities for both T_1 and T_2 are 0.850, 0.700 and 0.55, respectively.*

		$t_1 = 0.1625$			$t_1 = 0.3566$			$t_1 = 0.5978$		
Censoring		bias	SE	MSE ($\times 10^{-3}$)	bias	SE	MSE ($\times 10^{-3}$)	bias	SE	MSE ($\times 10^{-3}$)
Censoring on	$\hat{S}_E$	0.0%	0.036	1.3	0.0%	0.049	2.4	0.2%	0.063	3.9
T_1 only	$\hat{S}_{KM}$	0.0%	0.036	1.3	0.1%	0.049	2.4	0.4%	0.064	4.1
	$\hat{S}_{RE}$	0.1%	0.036	1.3	0.2%	0.048	2.3	1.6%	0.060	3.7
	$\hat{S}_{NT}$	0.8%	0.031	1.0	2.7%	0.047	2.5	3.4%	0.065	4.6
			(0.036)			(0.050)			(0.062)	
Univariate	$\hat{S}_E$	0.1%	0.043	1.8	0.1%	0.057	3.2	0.4%	0.071	5.0
censoring	$\hat{S}_{KM}$	0.0%	0.036	1.3	0.1%	0.049	2.4	0.4%	0.064	4.1
	$\hat{S}_{RE}$	0.1%	0.041	1.7	0.3%	0.054	2.9	1.1%	0.069	4.8
	$\hat{S}_{NT}$	0.6%	0.030	0.9	2.4%	0.046	2.4	3.4%	0.061	4.1
			(0.035)			(0.050)			(0.067)	
Bivariate	$\hat{S}_E$	-0.1%	0.048	2.3	0.1%	0.072	5.1	-0.4%	0.100	10.0
censoring	$\hat{S}_{KM}$	0.0%	0.035	1.2	0.3%	0.051	2.6	0.4%	0.064	4.1
	$\hat{S}_{RE}$	0.0%	0.047	2.2	0.6%	0.065	4.2	2.3%	0.069	4.9
	$\hat{S}_{NT}$	0.9%	0.031	1.0	1.9%	0.047	2.4	1.0%	0.063	4.0
			(0.035)			(0.051)			(0.065)	

		$t_2 = 0.1625$			$t_2 = 0.3566$			$t_3 = 0.5978$		
Censoring		bias	SE	MSE ($\times 10^{-3}$)	bias	SE	MSE ($\times 10^{-3}$)	bias	SE	MSE ($\times 10^{-3}$)
Censoring on	$\hat{S}_E$	0.0%	0.042	1.8	0.3%	0.055	3.0	1.0%	0.066	4.4
T_1 only	$\hat{S}_{KM}$	0.1%	0.033	1.1	0.1%	0.041	1.7	0.6%	0.044	1.9
	$\hat{S}_{RE}$	0.1%	0.037	1.4	0.1%	0.045	2.0	0.6%	0.046	2.1
	$\hat{S}_{NT}$	0.6%	0.030	0.9	3.0%	0.040	2.0	4.0%	0.044	2.4
			(0.034)			(0.045)			(0.048)	
Univariate	$\hat{S}_E$	0.0%	0.042	1.8	0.1%	0.056	3.1	1.0%	0.069	4.9
censoring	$\hat{S}_{KM}$	0.1%	0.036	1.3	0.3%	0.049	2.4	1.0%	0.064	4.1
	$\hat{S}_{RE}$	0.0%	0.041	1.7	0.3%	0.055	3.0	1.0%	0.067	4.5
	$\hat{S}_{NT}$	0.3%	0.034	1.1	2.2%	0.045	2.2	2.7%	0.061	3.9
			(0.037)			(0.052)			(0.064)	
Bivariate	$\hat{S}_E$	0.1%	0.046	2.1	0.4%	0.071	5.0	0.4%	0.097	9.4
censoring	$\hat{S}_{KM}$	0.0%	0.034	1.1	0.0%	0.050	2.5	-0.4%	0.063	4.0
	$\hat{S}_{RE}$	-0.1%	0.047	2.2	0.6%	0.063	4.0	2.0%	0.068	4.7
	$\hat{S}_{NT}$	0.2%	0.030	0.9	1.1%	0.046	2.2	2.0%	0.063	4.1
			(0.036)			(0.048)			(0.066)	

Table 4: Comparison of the semiparametric normal transformation model (4) based SPMLE estimator and the IPCW estimator at various time pairs (t_1, t_2) under univariate censoring. The true underlying models are the **semiparametric normal transformation model (4) with** $\rho = 0.5$ *(i.e. the working model is correctly specified) and* **Clayton's model (13) with** $\theta = 4$ *(i.e. the working model is misspecified).* $\hat{S}_{NT}$, $\hat{S}_{IP}$ *are semiparametric normal transformation based SPMLE estimators, and IPCW estimator respectively.* SE_e *are the empirical standard errors, while* MSE *are the mean squared errors. The true bivariate survival probabilities at the specified points are 0.7577, 0.6415, 0.5568, 0.4574, and 0.3847, respectively and the averages of the relative biases (based on 1000 runs) are listed in the table.*

True Underlying Model		$(t_1, t_2) =$ (0.1625, 0.1625)			$(t_1, t_2) =$ (0.1625, 0.3566)			$(t_1, t_2) =$ (0.3566, 0.3566)		
		bias	$\mathrm{SE_e}$	MSE ($\times 10^{-3}$)	bias	$\mathrm{SE_e}$	MSE ($\times 10^{-3}$)	bias	$\mathrm{SE_e}$	MSE ($\times 10^{-3}$)
Semiparametric Normal	$\hat{S}_{NT}$	0.1%	0.040	1.60	0.0%	0.049	2.40	0.1%	0.050	2.50
Tranformation Model	$\hat{S}_{IP}$	-0.1%	0.043	1.85	-0.1%	0.052	2.70	0.0%	0.055	3.03
Clayton Model	$\hat{S}_{NT}$	2.6%	0.031	1.41	3.3%	0.039	1.98	2.1%	0.042	1.90
	$\hat{S}_{IP}$	0%	0.041	1.68	0.3%	0.051	2.61	0%	0.053	2.81

True Underlying Model		$(t_1, t_2) =$ (0.3566 0.5978)			$(t_1, t_2) =$ (0.5978, 0.5978)		
		bias	$\mathrm{SE_e}$	MSE ($\times 10^{-3}$)	bias	$\mathrm{SE_e}$	MSE ($\times 10^{-3}$)
Semiparametric Normal	$\hat{S}_{NT}$	0.5%	0.059	3.48	0.5%	0.056	3.13
Transformation Model	$\hat{S}_{IP}$	0.1%	0.061	3.72	0.1%	0.062	3.84
Clayton Model	$\hat{S}_{NT}$	0.8%	0.049	2.41	-3.6%	0.064	4.28
	$\hat{S}_{IP}$	-0.4%	0.062	3.84	-0.1%	0.066	4.36

Table 5: Averages and SEs of ρ and the model-based Kendall's tau when the **correct model (13) with various true θ is misspecified to the semiparametric normal model (4) with parameter ρ to estimate**. *SE_e and SE_m are empirical and model based standard errors, respectively.*

θ	True Kendall's *tau*	Censoring	ρ			Model-based *tau*		
			$\hat{\rho}$	$\mathrm{SE_e}$	$\mathrm{SE_m}$	$\widehat{tau}$	$\mathrm{SE_e}$	$\mathrm{SE_m}$
4	**0.614**	Censoring on T_1 only	0.840	0.020	0.014	**0.620**	0.024	0.016
		Univariate censoring	0.839	0.026	0.021	**0.619**	0.028	0.022
		Bivariate censoring	0.820	0.033	0.030	**0.608**	0.033	0.026
2	**0.500**	Censoring on T_1 only	0.696	0.042	0.038	**0.490**	0.035	0.030
		Univariate censoring	0.685	0.063	0.057	**0.487**	0.050	0.041
		Bivariate censoring	0.693	0.064	0.057	**0.492**	0.052	0.043
1	**0.333**	Censoring on T_1 only	0.500	0.060	0.059	**0.339**	0.044	0.041
		Univariate censoring	0.483	0.076	0.077	**0.320**	0.056	0.053
		Bivariate censoring	0.493	0.077	0.072	**0.327**	0.058	0.051
0.5	**0.199**	Censoring on T_1 only	0.313	0.067	0.072	**0.209**	0.046	0.048
		Univariate censoring	0.291	0.084	0.078	**0.194**	0.054	0.062
		Bivariate censoring	0.297	0.087	0.079	**0.196**	0.057	0.064