

# Semiparametric maximum likelihood estimation in normal transformation models for bivariate survival data

BY YI LI

*Department of Biostatistics, Harvard University, 44 Binney Street, Boston,  
Massachusetts 02115, U.S.A.*

yili@jimmy.harvard.edu

ROSS L. PRENTICE

*Fred Hutchinson Cancer Research Center, 1959 NE Pacific Street, Seattle,  
Washington 98195, U.S.A.*

rprentic@whi.org

AND XIHONG LIN

*Department of Biostatistics, Harvard School of Public Health, 655 Huntington Avenue,  
Boston, Massachusetts 02115, U.S.A.*

xlin@hsph.harvard.edu

## SUMMARY

We consider a class of semiparametric normal transformation models for right-censored bivariate failure times. 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 procedure is developed for estimating the marginal survival distribution and the pairwise correlation parameters. This produces an efficient estimator of the correlation parameter of the semiparametric normal transformation model, which characterizes the dependence of bivariate survival outcomes. In addition, a simple positive-mass-redistribution algorithm can be used to implement the estimation procedures. Since the likelihood function involves infinite-dimensional parameters, empirical process theory is utilized to study the asymptotic properties of the proposed estimators, which are shown to be consistent, asymptotically normal and semiparametric efficient. A simple estimator for the variance of the estimates is derived. Finite sample performance is evaluated via extensive simulations.

*Some key words:* Asymptotic normality; Bivariate failure time; Consistency; Semiparametric efficiency; Semiparametric maximum likelihood estimate; Semiparametric normal transformation.

## 1. INTRODUCTION

Examples of censored bivariate data include the Danish Twin Study (Wienke et al., 2003), the diabetic retinopathy study (Hougaard, 2000), the dual infection kidney dialysis study (van Keilegom & 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, such as twins, is of major interest, the latter because it renders genetic information.

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

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

where  $S(\tilde{t}_1, \tilde{t}_2) = \text{pr}(\tilde{T}_1 > \tilde{t}_1, \tilde{T}_2 > \tilde{t}_2)$ ,  $S_1(\tilde{t}_1) = S(\tilde{t}_1, 0^-)$  and  $S_2(\tilde{t}_2) = S(0^-, \tilde{t}_2)$  are bivariate survival and marginal survival functions, respectively, and  $\theta$  has an interpretation as cross ratio (Oakes, 1989) as well as corresponding to other dependence measures such as Kendall’s  $\tau$ . This model allows for negative dependence when  $-1 < \theta < 0$ . However, for random variables  $\tilde{T}_1$  and  $\tilde{T}_2$  which are marginally absolutely continuous with respect to the Lebesgue measure  $\mu$ , the joint distribution of  $(\tilde{T}_1, \tilde{T}_2)$  is absolutely continuous with respect to the product Lebesgue measure  $\mu \times \mu$  only when  $\theta > -0.5$ . When  $\theta \leq -0.5$ , Oakes (1989) noted 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 is of substantial interest to specify a semiparametric likelihood model that allows for arbitrary modelling of the marginal survival functions, that has 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, for example, Klaassen (1997) and Pitt et al. (2006). However, all the previous efforts in normal copula focused only on noncensoring 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 & Lin (2006) on spatial survival data. Li & Lin 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.

## 2. SEMIPARAMETRIC NORMAL TRANSFORMATION MODELS

Consider a pair of survival times  $(\tilde{T}_1, \tilde{T}_2)$ , 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

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

has a standard normal distribution, where  $\Phi(\cdot)$  is the cumulative distribution function 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  $\rho$  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 joint probability density function 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

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

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

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

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

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

Clearly,  $\rho = 0$  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  $\rho \rightarrow 1^-$ , and the lower Fréchet bound  $\max\{S_1(\tilde{t}_1) + S_2(\tilde{t}_2) - 1, 0\}$  as  $\rho \rightarrow -1^+$ . Indeed, the correlation parameter  $\rho$  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 & Lin \(2006\)](#). Also of interest is that equation (4) 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 parameter  $\rho$  can also be interpreted as 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  $\tilde{X}_{ij} = \tilde{T}_{ij} \wedge \tilde{U}_{ij} = \min(\tilde{T}_{ij}, \tilde{U}_{ij})$ ,  $\delta_{ij} = I(\tilde{T}_{ij} \leq \tilde{U}_{ij})$  ( $j = 1, 2$ ). For simplicity, we assume that censoring is random in that the censoring pair  $(\tilde{U}_{i1}, \tilde{U}_{i2})$  is independent of the survival pair  $(\tilde{T}_{i1}, \tilde{T}_{i2})$ . The likelihood function can then be factorized into the product of contributions from the survival and censoring times, facilitating likelihood-based inference.

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 the constraint of  $\Lambda_1 \equiv \Lambda_2$  in § 3, followed by the case of separate marginal cumulative hazards  $\Lambda_j$  ( $j = 1, 2$ ), which do not have such a constraint, in § 4.

### 3. ESTIMATION WITH A COMMON MARGINAL CUMULATIVE HAZARD

#### 3.1. The likelihood function

This section proposes a semiparametric maximum likelihood estimation procedure for the semiparametric normal transformation model with a common marginal cumulative hazard,  $\Lambda$ , say. 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  $(X_{ij}, \delta_{ij})$  contains the same information as the original  $(\tilde{X}_{ij}, \delta_{ij})$ , facilitating the derivation of a likelihood function that can be factorized into the product of contributions from the survival and censoring times. It follows that the likelihood function for

the unknown parameters  $(\Lambda, \rho)$  can be written, up to a constant, as the product of factors

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

for  $i = 1, \dots, m$ , where  $\Psi_j(x_1, x_2; \rho) = -(\partial/\partial x_j)\Psi(x_1, x_2; \rho)/\phi(x_j)$ , for  $j = 1, 2$ . Indeed,  $\Psi_j(x_1, x_2; \rho) = \text{pr}(T_{3-j} \geq x_{3-j} \mid T_j = x_j)$ , for  $j = 1, 2$ .

Direct maximization of the above likelihood in a space containing continuous hazard  $\Lambda(\cdot)$  is not feasible, as one can always let the likelihood go to infinity by choosing some continuous function  $\Lambda(\cdot)$  with fixed values at each  $\tilde{X}_{ij}$ , while letting  $\Lambda'(\cdot)$  go to infinity at an observed failure time, for example, at some  $\tilde{X}_{ij}$  with  $\delta_{ij} = 1$ . Thus we need to assume that  $\Lambda$  is cadlag, right continuous with left limits, and piecewise constant. It follows that the maximum likelihood estimator 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 semiparametric maximum likelihood estimator is the maximizer of 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  $\ell(\rho, \Lambda) = \log L(\rho, \Lambda)$ .

### 3.2. Theoretical properties of the semiparametric maximum likelihood estimator

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

*Condition 1 (Boundedness).* The parameter  $\rho$  lies in a compact set within  $(-1, 1)$ .

*Condition 2 (Finite interval).* There exist a  $t_0 > 0$  and a constant  $c_0 > 0$  such that  $\text{pr}(\tilde{U}_{ij} \geq t_0) = \text{pr}(\tilde{U}_{ij} = t_0) > c_0$ . In practice,  $t_0$  is usually the duration of the study.

*Condition 3 (Differentiability).* The marginal cumulative hazard  $\Lambda(t)$  is differentiable and  $\Lambda'(t) > 0$  over  $[0, t_0]$ . Moreover,  $\Lambda(t_0) < \infty$ .

Condition 1 ensures 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 2 ensures that the failures for both pair members can be observed over a finite interval  $[0, t_0]$ , which makes it possible to estimate the hazard over  $[0, t_0]$ . Condition 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 3 guarantees the identifiability of the semiparametric normal transformation model specified in equations (1) and (2).

Under Conditions 1–3, we show in a technical report, obtainable at <http://biowww.dfci.harvard.edu/~yili/bikaproof.pdf>, that the semiparametric maximum likelihood estimates do exist and are finite. Furthermore, the next two propositions indicate that the semiparametric maximum likelihood estimator of  $\Lambda$  remains bounded, and that the estimators of  $\{\rho, \Lambda(\cdot)\}$  are consistent and asymptotically normal estimators of the true parameters. The proofs can be found in the aforementioned technical report.

**PROPOSITION 1 (Consistency).** *Denote the true parameters by  $(\rho_0, \Lambda_0)$ . Then  $|\hat{\rho} - \rho_0| \rightarrow 0$  and  $\sup_{t \in [0, t_0]} |\hat{\Lambda}(t) - \Lambda_0(t)| \rightarrow 0$  almost surely.*

PROPOSITION 2 (Asymptotic normality). *The scaled process  $m^{1/2}(\hat{\rho} - \rho_0, \hat{\Lambda} - \Lambda_0)$  converges weakly to a zero-mean Gaussian process in the metric space  $R \times l^\infty[0, t_0]$ , where  $l^\infty[0, t_0]$  is the linear space containing all the bounded functions in  $[0, t_0]$  equipped with the supremum norm. Furthermore,  $\hat{\rho}$  and  $\int_0^{t_0} \eta(s)d\hat{\Lambda}(s)$  are asymptotically efficient, where  $\eta(s)$  is any function of bounded variation over  $[0, t_0]$ .*

Proposition 2 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  $t \in [0, t_0]$ . It further implies that the infinite-dimensional parameter,  $\Lambda(\cdot)$ , can be treated in the same fashion as the finite-dimensional correlation parameter  $\rho$ . Hence the asymptotic covariance matrix can be estimated by inverting the observed information matrix. To be specific, for any constant  $h_1$  and any function  $h_2$  of bounded variation, the asymptotic variance of

$$h_1\hat{\rho} + \int_0^{t_0} h_2(s)d\hat{\Lambda}(s) = h_1\hat{\rho} + \sum_{\{(i,j):\delta_{ij}=1\}} h_2(\tilde{X}_{ij})\Delta\hat{\Lambda}(\tilde{X}_{ij}) \tag{7}$$

can be estimated by  $\hat{h}'\hat{J}^{-1}\hat{h}$ , where  $\hat{h}$  is a column vector comprising  $h_1$  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  $\rho$  and the jump size of  $\Lambda$  at  $\tilde{X}_{ij}$  when  $\delta_{ij} = 1$ . More formally, it can be shown that  $m\hat{h}'\hat{J}^{-1}\hat{h} \rightarrow V(h_1, h_2)$  in probability as  $m \rightarrow \infty$ , where  $V(h_1, h_2)$  is the asymptotic variance of  $m^{1/2}\{h_1\hat{\rho} + \int_0^{t_0} h_2(s)d\hat{\Lambda}(s)\}$ . 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 § 5.

### 3.3. A positive-mass-redistribution algorithm

Consider the following computationally efficient procedure for obtaining the semiparametric maximum likelihood estimates and their variance-covariance matrix. 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  $t_1 < \dots < t_K$  the  $K$  distinct, ordered and pooled  $\tilde{T}_1$  and  $\tilde{T}_2$  failure times. Define  $r(t_1, t_2) = \#\{l \mid \tilde{X}_{1l} \geq t_1, \tilde{X}_{2l} \geq t_2\}$  to be the size of the risk set at  $(t_1, t_2)$  and let  $\tilde{R} = \{(t_1, t_2) \mid r(t_1, t_2) > 0\}$  denote the risk region. We focus on the square grids  $\{(t_1, t_2) \mid t_1 = t_i, t_2 = t_j, (i, j) = 1, \dots, K\}$  formed by the observed  $\tilde{T}_1$  and  $\tilde{T}_2$  pooled 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  $\tilde{T}_1$  and  $\tilde{T}_2$  pooled uncensored failure time, or by zero if there is no corresponding smaller uncensored time, without affecting the log empirical likelihood  $\ell(\rho, \Lambda)$ . We refer to such replacement as positive mass redistribution. Then let  $n_{ij}^{\delta_1\delta_2} = \#\{l \mid \tilde{X}_{1l} = t_i, \tilde{X}_{2l} = t_j, \delta_{1l} = \delta_1, \delta_{2l} = \delta_2\}$  for  $\delta_1, \delta_2 \in \{0, 1\}$  ( $i, j = 1, \dots, K$ ) with  $t_0 = 0$ . Also let  $f_{ij} = f(t_i, t_j)$ ,  $F_i = \prod_{l=1}^i (1 - \lambda_l)$  and  $F_i^- = F_{i-1}$  where  $\lambda_l = \Delta\Lambda(t_l)$ . The log empirical likelihood  $\ell(\rho, \Lambda)$  defined in § 3.1 can now be written as

$$\begin{aligned} \ell = & \sum_{i=1}^K \sum_{j=1}^K n_{ij}^{11} \log f_{im} + n_{ij}^{10} \log \left( \lambda_i F_i^- - \sum_{v=1}^j f_{iv} \right) + n_{ij}^{01} \log \left( \lambda_j F_j^- - \sum_{u=1}^i f_{uj} \right) \\ & + n_{ij}^{00} \log \left( F_i + F_j + \sum_{u=1}^i \sum_{v=1}^j f_{uv} - 1 \right), \end{aligned} \tag{8}$$

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 equation (5) and  $s_i = \Phi^{-1}(1 - F_i)$ . To ensure numerical stability and avoid arguments of 0 for  $\Phi^{-1}$  in computation in finite samples, we use an asymptotically equivalent transformation  $s_i = \Phi^{-1}\{1 - (1 - m^{-1})F_i\}$ . A simple Newton–Raphson procedure, starting with  $\rho = 0$  and the Kaplan–Meier marginal hazard rates derived by treating  $(\tilde{X}_{ij}, \delta_{ij})$  ( $j = 1, 2$  and  $i = 1, \dots, m$ ), as  $2m$  independent observations can be used to compute the semiparametric maximum likelihood estimates  $\hat{\lambda}_i, \hat{\rho}$ . These calculations are less computationally demanding as they do not require the evaluation of bivariate incomplete normal integrals, only the evaluation of the univariate  $\Phi^{-1}$ .

If we follow the arguments in §3.2, the variability of  $\hat{\lambda}_i, \hat{\rho}$  can be assessed by inverting the negative Hessian matrix of equation (8), denoted by  $\hat{J}$ , a  $(K + 1) \times (K + 1)$  matrix. Furthermore, the functional (7) can be rewritten as a linear combination of  $\hat{\rho}$  and the  $\hat{\lambda}_i$ , namely,  $h_1 \hat{\rho} + \sum_{i=1}^K h_2(t_i) \hat{\lambda}_i$ , whose variance function can be easily computed as  $\hat{h}' \hat{J}^{-1} \hat{h}$ , where  $\hat{h} = \{h_1, h_2(t_1), \dots, h_2(t_K)\}'$ . We can easily apply this result to estimate the variance of the estimator of a survival probability. For example, the common marginal survival  $S(u_0) = \text{pr}(\tilde{T}_1 > u_0)$  at any given time  $u_0 \in [0, t_0]$  can be estimated by  $\hat{S}(u_0) = e^{-\hat{\Lambda}(u_0)}$ . With a first-order Taylor expansion,

$$\hat{S}(u_0) - S(u_0) \simeq -S(u_0)\{\hat{\Lambda}(u_0) - \Lambda(u_0)\} = \int_0^{t_0} -S(u_0)I(s \leq u_0)d\hat{\Lambda}(s) + \text{const.}$$

Hence,  $\hat{S}(u_0)$  can be approximated by the functional form (7) with  $h_1 = 0$  and  $h_2(s) = -S(u_0)I(s \leq u_0)$  and application of the above results will render  $\hat{S}^2(u_0) \hat{e}' \hat{J}^{-1} \hat{e}$  as a consistent estimator of the variance of  $\hat{S}(u_0)$ , where  $\hat{e} = \{0, I(t_1 \leq u_0), \dots, I(t_K \leq u_0)\}'$ .

#### 4. ESTIMATION FOR THE STRATIFIED-HAZARD MODEL

##### 4.1. The estimator and its theoretical properties

In this section, we relax the condition of a common marginal hazard and allow each  $\tilde{T}_{ij}$  to have a separate cumulative hazard function  $\Lambda_j(\cdot)$  ( $j = 1, 2$ ). This is often appropriate in practice.

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

$$\begin{aligned} & \tilde{L}_i(\rho, \Lambda_1, \Lambda_2) \\ &= \{e^{g(X_{i1}, X_{i2}; \rho)} \Lambda_1'(\tilde{X}_{i1}) \Lambda_2'(\tilde{X}_{i2}) e^{-\Lambda_1(\tilde{X}_{i1}) - \Lambda_2(\tilde{X}_{i2})}\}^{\delta_{i1} \delta_{i2}} \{\Psi_1(X_{i1}, X_{i2}; \rho) \Lambda_1'(\tilde{X}_{i1}) e^{-\Lambda_1(\tilde{X}_{i1})}\}^{\delta_{i1}(1 - \delta_{i2})} \\ & \quad \times \{\Psi_2(X_{i1}, X_{i2}; \rho) \Lambda_2'(\tilde{X}_{i2}) e^{-\Lambda_2(\tilde{X}_{i2})}\}^{(1 - \delta_{i1}) \delta_{i2}} \times \{\Psi(X_{i1}, X_{i2}; \rho)\}^{(1 - \delta_{i1})(1 - \delta_{i2})}. \end{aligned} \tag{9}$$

Here  $X_{ij} = \Phi^{-1}\{1 - \exp(-\Lambda_j(\tilde{X}_{ij}))\}$  for  $j = 1, 2$ . Again, direct maximization of the likelihood function in a space containing continuous hazards  $\Lambda_1(\cdot)$  or  $\Lambda_2(\cdot)$  is infeasible, as one can always make the likelihood 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 infinity at an observed failure time. Hence, when performing the maximum likelihood estimation, we assume that  $(\Lambda_1, \Lambda_2)$  are cadlag and piecewise constant. It follows that the semiparametric maximum likelihood estimator,  $(\hat{\rho}, \hat{\Lambda}_1, \hat{\Lambda}_2)$ , is the maximizer of the empirical likelihood function  $\ell(\rho, \Lambda_1, \Lambda_2)$ , which is obtained from equation (9) with the derivatives  $\Lambda_1'(\cdot)$  and  $\Lambda_2'(\cdot)$  at the observed failure times, respectively

replaced by their jumps  $\Delta\Lambda_1(\cdot)$  and  $\Delta\Lambda_2(\cdot)$  at the corresponding time-points. We can show that  $(\hat{\rho}, \hat{\Lambda}_1, \hat{\Lambda}_2)$  exist and are finite. Furthermore, under Conditions 1–3, with both  $\Lambda_1$  and  $\Lambda_2$  satisfying Condition 3, the asymptotic properties of the estimators are summarized in the following two theorems, the proofs of which can be found in our technical report.

**PROPOSITION 3 (Consistency).** *Denote the true parameters by  $(\rho_0, \Lambda_{01}, \Lambda_{02})$ . Then  $|\hat{\rho} - \rho_0| \rightarrow 0$ ,  $\sup_{t \in [0, t_0]} |\hat{\Lambda}_1(t) - \Lambda_{01}(t)| \rightarrow 0$  and  $\sup_{t \in [0, t_0]} |\hat{\Lambda}_2(t) - \Lambda_{02}(t)| \rightarrow 0$  almost surely.*

**PROPOSITION 4 (Asymptotic normality).** *The empirical process  $m^{1/2}(\hat{\rho} - \rho_0, \hat{\Lambda}_1 - \Lambda_{01}, \hat{\Lambda}_2 - \Lambda_{02})$  converges weakly to a zero-mean Gaussian process in the metric space  $R \times l^\infty[0, t_0] \times l^\infty[0, t_0]$ , where  $l^\infty[0, t_0]$  is the linear space containing all the bounded functions in  $[0, t_0]$  equipped with the supremum norm. Furthermore,  $\hat{\rho}$ ,  $\int_0^{t_0} \eta_1(s) d\hat{\Lambda}_1(s)$  and  $\int_0^{t_0} \eta_2(s) d\hat{\Lambda}_2(s)$  are asymptotically efficient, where  $\eta_1(s)$  and  $\eta_2(s)$  are any functions of bounded variation over  $[0, t_0]$ .*

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. For any constant  $h_1$  and any function  $h_2$  and  $h_3$  of bounded variation, the asymptotic variance of

$$h_1 \hat{\rho} + \int_0^{t_0} h_2(s) d\hat{\Lambda}_1(s) + \int_0^{t_0} h_3(s) d\hat{\Lambda}_2(s) = h_1 \hat{\rho} + \sum_{\{i: \delta_{i1}=1\}} h_2(\tilde{X}_{i1}) \Delta \hat{\Lambda}(\tilde{X}_{i1}) + \sum_{\{i: \delta_{i2}=1\}} h_2(\tilde{X}_{i2}) \Delta \hat{\Lambda}_2(\tilde{X}_{i2}) \quad (10)$$

can be estimated by  $\hat{h}' \hat{J}^{-1} \hat{h}$ , where  $\hat{h}$  is a column vector comprising  $h_1$ , 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  $\rho$  and the jump sizes of  $\Lambda_j$  at  $\tilde{X}_{ij}$  when  $\delta_{ij} = 1$ . Indeed, following the proof of Theorem 3 in [Parner \(1998\)](#), one can show  $m \hat{h}' \hat{J}^{-1} \hat{h} \rightarrow V(h_1, h_2, h_3)$  in probability as  $m \rightarrow \infty$ , where  $V(h_1, h_2, h_3)$  is the asymptotic variance of  $m^{1/2}\{h_1 \hat{\rho} + \int_0^{t_0} h_2(s) d\hat{\Lambda}_1(s) + \int_0^{t_0} h_3(s) d\hat{\Lambda}_2(s)\}$ .

#### 4.2. Practical implementation for the stratified-hazard model

Denote by  $t_{11} < \dots < t_{1I}$  the  $I$  distinct ordered  $\tilde{T}_1$ -failure times and by  $t_{21} < \dots < t_{2J}$  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  $\{(t_1, t_2) \mid t_1 = t_{1i}, t_2 = t_{2j}, (i = 1, \dots, I, j = 1, \dots, J)\}$  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 times, or by zero if there is no corresponding smaller uncensored time, without affecting the empirical likelihood  $\ell(\rho, \Lambda_1, \Lambda_2)$ . With these replacements, or the so-called positive mass redistributions, let  $n_{ij}^{\delta_1, \delta_2} = \#\{l \mid \tilde{X}_{1l} = t_{1i}, \tilde{X}_{2l} = t_{2j}, \delta_{1l} = \delta_1, \delta_{2l} = \delta_2\}$  for  $\delta_1, \delta_2 \in \{0, 1\}$  and for  $i = 0, \dots, I$  and  $j = 0, \dots, J$ , with  $t_{10} = t_{20} = 0$ . Also let  $f_{ij} = f(t_{1i}, t_{2j})$ ,  $F_{1i} = \prod_{l=1}^i (1 - \lambda_{1l})$ ,  $F_{1i}^- = F_{1, i-1}$ ,  $F_{2j} = \prod_{k=1}^j (1 - \lambda_{2k})$  and  $F_{2j}^- = F_{2, j-1}$ , where  $\lambda_{1l} = \Delta\Lambda_1(t_{1l})$ ,  $\lambda_{2k} = \Delta\Lambda_2(t_{2k})$ . The log empirical likelihood function can now be written as

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

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, we use an asymptotically equivalent transformation  $s_{1i} = \Phi^{-1}\{1 - (1 - m^{-1})F_{1i}\}$  and  $s_{2j} = \Phi^{-1}\{1 - (1 - m^{-1})F_{2j}\}$ . 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 semiparametric maximum likelihood estimators  $\hat{\lambda}_{1i}$ ,  $\hat{\lambda}_{2j}$  and  $\hat{\rho}$ . The likelihood evaluations are computationally less demanding, requiring only the computation of the univariate  $\Phi^{-1}$ .

Similarly, the variability of  $\hat{\lambda}_{1i}$ ,  $\hat{\lambda}_{2j}$  and  $\hat{\rho}$  can be assessed by inverting the negative Hessian matrix of equation (11), denoted by the  $(I + J + 1) \times (I + J + 1)$  matrix  $\hat{J}$ . Moreover, functional (10) can be rewritten as a linear combination of  $\hat{\rho}$  and the  $\hat{\lambda}_{1i}$ ,  $\hat{\lambda}_{2j}$ , 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  $\hat{h} = \{h_1, h_2(t_{11}), \dots, h_2(t_{1I}), h_3(t_{21}), \dots, h_3(t_{2J})\}'$ .

As an illustration of this variance formula, consider the bivariate survival estimator  $\hat{S}(u_0, v_0)$   $S(u_0, v_0)$  at any given time pair  $(u_0, v_0) \in [0, t_0]^2$ . Based on the semiparametric normal transformation model, this is

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

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

$$\begin{aligned} & \hat{S}(u_0, v_0) - S(u_0, v_0) \\ & \simeq \gamma_1(u_0, v_0)(\hat{\rho} - \rho_0) + \gamma_2(u_0, v_0)\{\hat{\Lambda}_1(u_0) - \Lambda_1(u_0)\} + \gamma_3(u_0, v_0)\{\hat{\Lambda}_2(v_0) - \Lambda_2(v_0)\} \\ & = \gamma_1(u_0, v_0)\hat{\rho} + \int_0^{t_0} \gamma_2(u_0, v_0)I(s \leq u_0)d\hat{\Lambda}_1(s) + \int_0^{t_0} \gamma_3(u_0, v_0)I(s \leq v_0)d\hat{\Lambda}_2(s) + \text{const}, \end{aligned}$$

where

$$\begin{aligned} \gamma_1(t_1, t_2) &= \partial \Psi(x_1, x_2; \rho) / \partial \rho, \quad \gamma_2(t_1, t_2) = -\Phi_1(x_1, x_2; \rho_0) \exp\{-\Lambda_1(t_1)\}, \\ \gamma_3(t_1, t_2) &= -\Phi_2(x_1, x_2; \rho_0) \exp\{-\Lambda_2(t_2)\}, \quad x_j = \Phi^{-1}\{1 - e^{-\Lambda_j(t_j)}\}. \end{aligned}$$

Hence,  $\hat{S}(u_0, v_0)$  can be approximated by the functional form (10) with

$$h_1 = \gamma_1(u_0, v_0), \quad h_2(s) = \gamma_2(u_0, v_0)I(s \leq u_0), \quad h_3(s) = \gamma_3(u_0, v_0)I(s \leq v_0),$$

and application of the above variance formula renders a consistent estimator of the variance for  $\hat{S}(u_0, v_0)$ , namely,  $\hat{h}' \hat{J}^{-1} \hat{h}$ , where

$$\begin{aligned} \hat{h} &= \{\hat{\gamma}_1(u_0, v_0), \hat{\gamma}_2(u_0, v_0)I(t_{11} \leq u_0), \dots, \hat{\gamma}_2(u_0, v_0)I(t_{1I} \leq u_0), \\ & \hat{\gamma}_3(u_0, v_0)I(t_{21} \leq v_0), \dots, \hat{\gamma}_3(u_0, v_0)I(t_{2J} \leq v_0)\}' \end{aligned}$$

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 evaluate the finite sample performance of this variance estimator in § 5.



Table 1. Averages and model-based and empirical SEs of the SPMLEs under the semiparametric normal transformation model (3) with a common-hazard function. The true values are  $\rho = 0.5$ ,  $S(0.1625) = 0.85$ ,  $S(0.3566) = 0.70$ ,  $S(0.5978) = 0.55$

Censoring	$\rho$	$SE_e$	$SE_m$	$t = 0.1625$		$t = 0.3566$		$t = 0.5978$	
				$\hat{S}(t)$	$SE_e$	$\hat{S}(t)$	$SE_e$	$\hat{S}(t)$	$SE_e$
Censoring on $T_1$ only	0.502	0.109	0.104	0.842	0.023	0.692	0.035	0.546	0.044
Univariate censoring	0.503	0.122	0.114	0.842	0.025	0.694	0.035	0.546	0.051
Bivariate censoring	0.493	0.141	0.138	0.846	0.027	0.697	0.042	0.555	0.055

SE, standard error; SPMLE, semiparametric maximum likelihood error;  $SE_e$ , empirical standard error;  $SE_m$ , model-based standard error.

## 5. NUMERICAL STUDIES

### 5.1. Preamble

A series of simulation studies was performed to examine the properties of the proposed estimator and to compare it with the existing bivariate survivor estimators, including the Prentice–Cai (Prentice & Cai, 1992), Dabrowska (1988) and repaired nonparametric maximum likelihood (van der Laan, 1996; Moodie & Prentice, 2005) estimators. The simulation set-up mimics those in Prentice et al. (2004). The marginal distributions of  $\tilde{T}_1$  and  $\tilde{T}_2$  were specified as unit exponential, and the censoring time  $\tilde{U}_1$  was taken to be an exponential variate with mean 0.5. Three special cases for  $\tilde{U}_2$  were considered:  $\tilde{U}_2 = \infty$ , corresponding to no  $\tilde{T}_2$  censoring;  $\tilde{U}_2 = \tilde{U}_1$ , corresponding to univariate censoring;  $\tilde{U}_2$  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.

### 5.2. Performance under the correct model

We began by evaluating the finite sample performance of the semiparametric maximum likelihood estimator when the true model follows the semiparametric normal transformation model (3) with  $\rho = 0.5$ . We considered both the common-hazard model and the stratified-hazard model, but, as both models yielded similar results, in Table 1 we report only the simulation results for the common-hazard model. As efficient estimation of the common-hazard function or the common marginal distribution function is of major interest under the common-hazard model, we report only the estimates of the marginal survival function at various time-points in Table 1. The sample averages of the estimates are very close to the true values, and the model-based standard errors, which were computed by applying the results of § 3.3, match very well with the empirical standard errors up to the third decimal point.

### 5.3. Performance under the misspecified model

We next considered the robustness of the semiparametric maximum likelihood estimator when the semiparametric normal transformation model was misspecified. The failure times were generated under the following bivariate Clayton model:

$$S(\tilde{t}_1, \tilde{t}_2) = \{S_1(\tilde{t}_1)^{-\theta} + S_2(\tilde{t}_2)^{-\theta} - 1\}^{-1/\theta}, \tag{12}$$

with  $\theta = 4$ , implying a strong positive dependence between  $\tilde{T}_1$  and  $\tilde{T}_2$ . We compared the performance of the semiparametric maximum likelihood estimator based on the semiparametric normal transformation model,  $\hat{S}_{NT}$  with the other existing nonparametric estimators, including the Prentice–Cai estimator,  $\hat{S}_{PC}$ , the empirical hazard rate estimator,  $\hat{S}_E$  and the redistributed empirical estimator,  $\hat{S}_{RE}$ , which were taken from Tables 1 and 2 of Prentice et al. (2004). As

Table 2. *Averages, SES and mean squared errors for various bivariate survival function estimators at various time pairs  $(t_1, t_2)$  when the correct model (12) with  $\theta = 4$  is misspecified to the semiparametric normal model (3). The true bivariate survival probabilities at these pairs are 0.771, 0.666 and 0.608, respectively*

Censoring		$(t_1, t_2) = (0.1625, 0.1625)$			$(t_1, t_2) = (0.1625, 0.3566)$			$(t_1, t_2) = (0.3566, 0.3566)$		
		Bias (%)	SE ( $\times 10^{-2}$ )	MSE ( $\times 10^{-3}$ )	Bias (%)	SE ( $\times 10^{-2}$ )	MSE ( $\times 10^{-3}$ )	Bias (%)	SE ( $\times 10^{-2}$ )	MSE ( $\times 10^{-3}$ )
Censoring on $T_1$ only	$\hat{S}_E$	0.0	4.6	2.1	0.3	5.5	3.0	0.0	5.7	3.2
	$\hat{S}_{PC}$	0.1	4.0	1.6	0.1	4.3	1.9	-0.1	4.7	2.2
	$\hat{S}_{RE}$	0.1	4.3	1.8	0.3	4.6	2.1	0.0	4.9	2.4
	$\hat{S}_{NT}$	3.2	3.5 (3.3)	1.7	4.4	3.5 (3.0)	2.0	3.3	4.3 (3.8)	2.2
Univariate censoring	$\hat{S}_E$	0.0	5.1	2.6	0.3	5.9	3.5	0.1	6.3	4.0
	$\hat{S}_{PC}$	0.1%	4.1	1.7	0.3%	4.9	2.4	0.0%	5.1	2.6
	$\hat{S}_{RE}$	0.1	4.8	2.3	0.3	5.7	3.2	0.0	5.8	3.4
	$\hat{S}_{NT}$	2.6	3.2 (2.8)	1.4	3.3	3.9 (3.7)	2.0	2.1	4.2 (4.2)	1.9
Bivariate censoring	$\hat{S}_E$	0.0	5.8	3.4	0.3	7.3	5.3	-0.1	7.7	5.9
	$\hat{S}_{PC}$	-0.1	4.1	1.7	-0.1	4.9	2.4	-0.3	5.3	2.8
	$\hat{S}_{RE}$	-0.2	5.6	3.1	0.3	6.6	4.4	-0.3	6.8	4.6
	$\hat{S}_{NT}$	2.2	3.5 (3.3)	1.5	1.9	4.2 (4.6)	1.9	0.1	4.7 (4.8)	2.2

MSE, mean squared error;  $\hat{S}_E$ , the empirical hazard rate estimator;  $\hat{S}_{PC}$ , the Prentice–Cai estimator;  $\hat{S}_{RE}$ , the redistributed empirical estimator;  $\hat{S}_{NT}$ , the semiparametric maximum likelihood estimator; SES, empirical standard errors; for  $\hat{S}_{NT}$  the model-based standard errors are displayed inside the brackets.

Prentice et al. (2004) only considered stratified-hazard models, we focused on the stratified-hazard model to make the resulting estimates comparable.

Tables 2 and 3 contain 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. We also list the summary statistics for the empirical hazard rate,  $\hat{S}_E$ , Prentice–Cai,  $\hat{S}_{PC}$  and redistributed empirical,  $\hat{S}_{RE}$  estimators (Prentice et al., 2004). Finally, we computed the mean squared errors for all the estimators.

Our results show that, even when the underlying model is misspecified, the semiparametric maximum likelihood estimation based on the semiparametric normal transformation model incurred only small biases. Among all the scenarios examined, the relative biases of the semiparametric maximum likelihood estimates ranged from  $-5.7\%$  to  $4\%$ . Compared to the competing nonparametric estimators, the new estimator also achieved high efficiency and had the smallest standard errors in almost all the scenarios examined. It had a smaller mean squared error than the other estimators in most cases considered. In addition, the model-based standard errors agreed well with their empirical counterparts.

#### 5.4. Comparison with the inverse-probability-of-censoring-weighted estimator

We also compared the new estimator with a doubly robust inverse-probability-of-censoring-weighted estimator, derived under univariate censoring, which stipulates that the censoring time is common for both pair members (Lin & Ying, 1993; Tsai & Crowley, 1998; Wang & Wells, 1998; Nan et al., 2006). The detailed derivation can be found in our technical report. We first compared

Table 3. Averages, standard errors and mean squared errors for various estimators of the marginal survival functions  $S_1$  and  $S_2$  when the correct model (12) with  $\theta = 4$  is misspecified to the semiparametric normal model (3). The true marginal survival probabilities for both  $T_1$  and  $T_2$  are 0.850, 0.700 and 0.55, respectively

Censoring		$t_1 = 0.1625$			$t_1 = 0.3566$			$t_1 = 0.5978$		
		Bias (%)	SE ( $\times 10^{-2}$ )	MSE ( $\times 10^{-3}$ )	Bias (%)	SE ( $\times 10^{-2}$ )	MSE ( $\times 10^{-3}$ )	Bias (%)	SE ( $\times 10^{-2}$ )	MSE ( $\times 10^{-3}$ )
Censoring on $T_1$ only	$\hat{S}_E$	0.0	3.6	1.3	0.0	4.9	2.4	0.2	6.3	3.9
	$\hat{S}_{KM}$	0.0	3.6	1.3	0.1	4.9	2.4	0.4	6.4	4.1
	$\hat{S}_{RE}$	0.1	3.6	1.3	0.2	4.8	2.3	1.6	6.0	3.7
	$\hat{S}_{NT}$	0.8	3.1	1.0	2.7	4.7	2.5	3.4	6.5	4.6
			(3.6)		(5.0)			(6.2)		
Univariate censoring	$\hat{S}_E$	0.1	4.3	1.8	0.1	5.7	3.2	0.4	7.1	5.0
	$\hat{S}_{KM}$	0.0	3.6	1.3	0.1	4.9	2.4	0.4	6.4	4.1
	$\hat{S}_{RE}$	0.1	4.1	1.7	0.3	5.4	2.9	1.1	6.9	4.8
	$\hat{S}_{NT}$	0.6	3.0	0.9	2.4	4.6	2.4	3.4	6.1	4.1
			(3.5)		(5.0)			(6.7)		
Bivariate censoring	$\hat{S}_E$	-0.1	4.8	2.3	0.1	7.2	5.1	-0.4	10.0	10.0
	$\hat{S}_{KM}$	0.0	3.5	1.2	0.3	5.1	2.6	0.4	6.4	4.1
	$\hat{S}_{RE}$	0.0	4.7	2.2	0.6	6.5	4.2	2.3	6.9	4.9
	$\hat{S}_{NT}$	0.9	3.1	1.0	1.9	4.7	2.4	1.0	6.3	4.0
			(3.5)		(5.1)			(6.5)		

$\hat{S}_E$ , the empirical hazard rate estimator;  $\hat{S}_{KM}$ , the Kaplan–Meier estimator;  $\hat{S}_{RE}$ , the redistributed empirical estimator;  $\hat{S}_{NT}$ , the semiparametric maximum likelihood estimator. SEs; empirical standard errors; for  $\hat{S}_{NT}$  the model-based standard errors are displayed inside the brackets; only the estimates of the  $T_1$ -survival are reported, as similar patterns were observed for the estimates of the  $T_2$ -survival.

the efficiency of the inverse-probability-of-censoring-weighted estimator with the semiparametric maximum likelihood estimator when the true underlying model indeed followed the semiparametric normal transformation model (3) with  $\rho = 0.5$ . The results are documented in Table 4, and demonstrate that the new estimator has noticeably smaller variances than its competitor. We next considered the robustness of the inverse-probability-of-censoring-weighted estimator when the underlying model was misspecified as the semiparametric normal transformation model, while the true model followed the bivariate Clayton model (12) with  $\theta = 4$ . The results in Table 4 indicate that the inverse-probability-of-censoring-weighted estimator has eliminated the bias caused by the misspecification of the semiparametric normal transformation model, while the new estimator incurs negligible biases and retains smaller variances. In terms of mean squared error, the semiparametric maximum likelihood estimator was superior in the scenarios considered.

### 5.5. Estimation of $\rho$

Finally, we discuss the interpretation and estimation of  $\rho$ , which has a one-to-one correspondence between the common dependence measure for bivariate survival, such as Kendall's  $\tau$ . As indicated in Li & Lin (2006),

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

Table 4. Comparison of the semiparametric normal transformation model (3) based SPMLE estimator and the inverse-probability-of-censoring-weighted estimator at various time pairs  $(t_1, t_2)$  under univariate censoring. The true underlying models are the semiparametric normal transformation model (3) with  $\rho = 0.5$ , i.e. the working model is correctly specified, and Clayton's model (12) with  $\theta = 4$ , i.e. the working model is misspecified. The true bivariate survival probabilities at the specified points are 0.7577, 0.6415 and 0.5568, respectively, and the averages of the relative biases 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 (%)	SE <sub>e</sub>	MSE	Bias (%)	SE <sub>e</sub>	MSE	Bias (%)	SE <sub>e</sub>	MSE
			( $\times 10^{-2}$ )	( $\times 10^{-3}$ )		( $\times 10^{-2}$ )	( $\times 10^{-3}$ )		( $\times 10^{-2}$ )	( $\times 10^{-3}$ )
Semiparametric normal Transformation model	$\hat{S}_{NT}$	0.1	4.0	1.6	0.0	4.9	2.4	0.1	5.0	2.5
	$\hat{S}_{IP}$	-0.1	4.3	1.9	-0.1	5.2	2.7	0.0	5.5	3.0
Clayton model	$\hat{S}_{NT}$	2.6	3.1	1.4	3.3	3.9	2.0	2.1	4.2	1.9
	$\hat{S}_{IP}$	0	4.1	1.7	0.3	5.1	2.6	0	5.3	2.8

$\hat{S}_{NT}$ , the semiparametric maximum likelihood estimator based on the semiparametric normal transformation model;  $\hat{S}_{IP}$ , the inverse-probability-of-censoring-weighted estimator; SE<sub>e</sub>, the empirical standard error; MSE, the mean squared error.

Table 5. Averages and SEs of  $\rho$  and the model-based Kendall's  $\tau$  when the correct model (12) with various true  $\theta$  is misspecified to the semiparametric normal model (3) with parameter  $\rho$  to be estimated

$\theta$	True Kendall's $\tau$	censoring	$\rho$			Model-based $\tau$		
			$\hat{\rho}$	SE <sub>e</sub>	SE <sub>m</sub>	$\hat{\tau}$	SE <sub>e</sub>	SE <sub>m</sub>
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

SE<sub>e</sub>, empirical standard errors; SE<sub>m</sub>, model-based standard errors.

where  $S(\tilde{t}_1, \tilde{t}_2; \rho)$  and  $f(\tilde{t}_1, \tilde{t}_2; \rho)$  are, respectively, the joint bivariate survival and density functions defined in equations (3) and (4). After a change of variables, we have that

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

where  $\Psi(\cdot)$  is the joint tail function for the bivariate normal distribution defined in equation (2),  $g(t_1, t_2; \rho)$  is the cross-term defined in equation (5) and  $\phi$  is the standard normal density function, none of which depends on any specific forms of hazard functions. As shown in Li & Lin (2006),  $\rho$  uniquely determines  $\tau$ , and thus provides a standardized dependence measure for bivariate

survival. Indeed,  $\tau(\hat{\rho})$  yields the model-based estimator of Kendall's  $\tau$ , with a model-based standard error that can be conveniently obtained using the delta method.

We considered the estimation and the interpretation of the estimate of  $\rho$  when the semiparametric normal transformation model was misspecified, and the failure times were generated under the bivariate Clayton model in equation (12). We chose  $\theta$  to be 0.5, 1, 2 and 4, which correspond to Kendall'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  $\rho$  and the model-based Kendall's  $\tau$  are displayed in Table 5, along with the empirical and model-based standard errors. It appears that, when the underlying model is misspecified, the estimate of  $\rho$  itself might not be of interest, as it does not recover the specific dependence structure of the true model. However, the model-based estimator of Kendall's  $\tau$  based on the estimates of  $\rho$  are very comparable to the true values. We envisage that, at least for the scenarios we considered, the estimate of  $\rho$  would lead to a reasonable approximation for Kendall's  $\tau$  even if the model were misspecified.

## 6. DISCUSSION

With the analytical framework established in this article, we are able to extend 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 hazards 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.

## ACKNOWLEDGEMENT

This work was supported in part by grants to each author from the U.S. National Institute of Health. The authors thank Professor D. M. Titterington and the associate editor for their insightful suggestions.

## REFERENCES

- CLAYTON, D. (1978). A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika*, **65**, 141–51.
- DABROWSKA, D. (1988). Kaplan–Meier estimate on the plane. *Ann. Statist.* **16**, 1475–89.
- HOUGAARD, P. (2000). *Analysis of Multivariate Survival Data*. New York: Springer.
- KLAASSEN, C. A. J. & WELLNER, J. A. (1997). Efficient estimation in the bivariate normal copula model: Normal margins are least favourable. *Bernoulli* **3**, 55–77.
- LI, Y. & LIN, X. (2006). Semiparametric normal transformation models for spatially correlated survival data. *J. Am. Statist. Assoc.* **101**, 591–603.
- LIN, D. & YING, Z. (1993). A simple nonparametric estimator of the bivariate survival function under univariate censoring. *Biometrika* **80**, 573–81.
- MOODIE, F. Z. & PRENTICE, R. L. (2005). An adjustment to improve the bivariate survivor function repaired NPMLE. *Lifetime Data Anal.* **11**, 291–307.
- MURPHY, S. A. (1994). Consistency in a proportional hazards model incorporating a random effect. *Ann. Statist.* **22**, 712–31.
- NAN, B., LIN, X., LISABETH, L. D. & HARLOW, S. D. (2006). Piecewise constant cross-ratio estimation for association of age at a marker event and age at menopause. *J. Am. Statist. Assoc.* **101**, 65–77.
- OAKES, D. (1989). Bivariate survival models induced by frailties. *J. Am. Statist. Assoc.* **84**, 487–93.
- PARNER, E. (1998). Asymptotic theory for the correlated gamma-frailty model. *Ann. Statist.* **26**, 183–214.
- PITT, M., CHAN, D. & KOHN, R. (2006). Efficient Bayesian inference for Gaussian copula regression models. *Biometrika* **93**, 537–54.

- PRENTICE, R. L. & CAI, J. (1992). Covariance and survivor function estimation using censored multivariate failure time data. *Biometrika* **79**, 495–512. Correction (1993) **80**, 711–2.
- PRENTICE, R. L., MOODIE, Z. F. & WU, J. (2004). Hazard-based nonparametric survivor function estimation. *J. R. Statist. Soc. B* **66**, 305–19.
- TSAI, W. & CROWLEY, J. (1998). A note on nonparametric estimators of the bivariate survival function under univariate censoring. *Biometrika* **85**, 573–80.
- VAN DER LAAN, M. (1996). Efficient estimation in the bivariate censoring model and repairing NPMLE. *Ann. Statist.* **24**, 596–627.
- VAN KEILEGOM, I. & HETTMANSPERGER, T.P. (2002) Inference on multivariate M-estimators based on bivariate censored data. *J. Am. Statist. Assoc.* **97**, 328–36.
- WANG, W. & WELLS, M. (1998). Nonparametric estimators of the bivariate survival function under simplified censoring conditions. *Biometrika* **84**, 863–80.
- WIENKE, A., LICHTENSTEIN, P. & YASHIN, A. (2003). A bivariate frailty model with a cure fraction for modeling familial correlations in diseases. *Biometrics* **59**, 1178–83.

[Received July 2007. Revised March 2008]