Scandinavian Journal of Statistics, Vol. 36: 481–500, 2009
doi: 10.1111/j.1467-9469.2008.00635.x

Marginal Regression Analysis for Semi-Competing Risks Data Under Dependent Censoring

A. ADAM DING and GUANGKAI SHI
Department of Mathematics, Northeastern University

WEIJING WANG
Institute of Statistics, National Chiao-Tung University

JIN-JIAN HSIEH
Department of Mathematics, National Chung Cheng University

ABSTRACT. Multiple events data are commonly seen in medical applications. There are two types of events, namely terminal and non-terminal. Statistical analysis for non-terminal events is complicated due to dependent censoring. Consequently, joint modelling and inference are often needed to avoid the problem of non-identifiability. This article considers regression analysis for multiple events data with major interest in a non-terminal event such as disease progression. We generalize the technique of artificial censoring, which is a popular way to handle dependent censoring, under flexible model assumptions on the two types of events. The proposed method is applied to analyse a data set of bone marrow transplantation.

Key words: artificial censoring, log-rank statistic, multiple events data, transformation model

1. Introduction

Multiple events data are commonly seen in medical studies. One type of events is terminal in the sense that its occurrence precludes the occurrence of other events. The other type is non-terminal and subject to censoring by the terminal event. In this article, we use death and disease progression as examples of terminal and non-terminal events respectively. Let T_1 be the time to progression, T_2 be the time to death, C be the time to external censoring and $\mathbf{Z}=(Z_1,\ldots,Z_K)'$ be the vector of covariates. Observed variables become $X=T_1\wedge T_2\wedge C$, $Y=T_2\wedge C$, $\delta_1=I(T_1\le T_2\wedge C)$, $\delta_2=I(T_2\le C)$ and $\mathbf{Z}$. Throughout the paper, we will assume that C is independent of T_1 and T_2 given $\mathbf{Z}$.

Let $(X_i, Y_i, \delta_{1i}, \delta_{2i}, \mathbf{Z}_i), i=1,\ldots,n$, be independent and identically distributed random replications of $(X, Y, \delta_1, \delta_2, \mathbf{Z})$. Such data are called semi-competing risks data by Fine *et al.* (2001). Inference about the progression time T_1 is complicated due to dependent censoring by T_2. There has been increasing attention in developing statistical methods for analysing semi-competing risks data under dependent censoring. In the absence of covariates, Day *et al.* (1997), Fine *et al.* (2001) and Wang (2003) studied the association relationship between T_1 and T_2 in the region of $T_1\le T_2$.

We are interested in how the covariate $\mathbf{Z}$ affects T_1. Regression analysis based on $T_1\wedge T_2$ is straightforward. Alternatively, semi-parametric regression models based on the cumulative incidence function, $\Pr(T_1\le t, T_1\le T_2)$ have been proposed by Fine & Gray (1999) and Fine (2001). Another popular approach to analysing semi-competing risks data is via a multistate framework, or more specifically an illness–death model in which no recovery is possible (Andersen *et al.*, 1993). This approach usually imposes model assumptions on the transition

rates between the states. For example, if the interest is in progression, one may assume that the progression intensities are proportional among different covariate groups (see Andersen *et al.* 1993, example VII.1.1).

Here, we take a different approach by directly modelling the covariate effect on T_1. Specifically, we assume the linear regression model

$$h_1(T_1) = \boldsymbol{\theta}_1' \mathbf{Z} + \epsilon_1, \tag{1}$$

where $h_1(t)$ is a known monotone function and ϵ_1 is an error term. The distribution of ϵ_1 is not specified. For example, when $h(t) = t$, the model becomes a location-shift (LS) model, and when $h(t) = \log(t)$, the model is the accelerated failure time (AFT) model. Our primary interest is to estimate $\boldsymbol{\theta}_1$ based on semi-competing risks data $(X_i, Y_i, \delta_{1i}, \delta_{2i}, \mathbf{Z}_i), i = 1, \ldots, n$. However, due to dependent censoring by T_2, the marginal distribution of T_1 is not identifiable. This implies that model (1) is not sufficient and additional assumptions are needed. We make two extra assumptions as follows. First, we assume that T_2 follows the regression model

$$h_2(T_2) = \boldsymbol{\theta}_2' \mathbf{Z} + \epsilon_2, \tag{2}$$

where $h_2(t)$ is a monotone function which may be known or unknown and ϵ_2 is another error variable. When $h_2(t)$ is unknown but the distribution of ϵ_2 is completely specified, T_2 is said to follow a transformation model. Examples of this class of models include the proportional hazards (PH) model, where ϵ_2 has the extreme value distribution, and the proportional odds model, where ϵ_2 has the standard logistic distribution. Secondly, we assume that the correlation structure between the failure times T_1 and T_2 is unknown but is common for all covariate values. That is,

$$\Pr(\epsilon_1 > s, \epsilon_2 > t \mid \mathbf{Z}) = \Pr(\epsilon_1 > s, \epsilon_2 > t) \tag{3}$$

but the form of the joint distribution is not specified. Compared with the existing methods, our proposal in (1) contains a flexible class of regression models and allows one to study the 'net' covariate effect on T_1 after separating its effect on T_2. The additional model assumption in (2) often by itself is of some interest. Condition (3) states that after removing the marginal covariate effects, the bivariate error distribution is independent of $\mathbf{Z}$. This condition has been discussed by Robins (1995a, lemma 2) under the simple two-sample case ($Z = 0$ or 1).

In deriving the inference procedure in this paper, we generalize the method of artificial censoring to remove the bias due to dependent censoring. The artificial censoring technique has been frequently used in related problems. For example, Robins (1995a,b) specified a model which describes the joint effect of $\mathbf{Z}$ and T_2 on T_1 and then proposed an estimation method for discrete covariates. Under a two-sample setting with $Z = 0, 1$, Lin *et al.* (1996) assumed models (1) and (2) with $h_j(t) = \log(t), j = 1, 2$, and condition (3). Chang (2000), Ghosh & Lin (2003) and Lin & Ying (2003) further extended the idea of Lin *et al.* (1996) to recurrent events and time-varying covariates with the progression event being replaced by recurrent events but death (or informative drop-out) still being the terminal event. It is important to mention that these papers all assume that the types of marginal models in (1) and (2) are the same. That is $h_1(t) = h_2(t)$.

In this paper, we extend the method of artificial censoring to a much more general regression setting that allows $h_1(t)$ and $h_2(t)$ to be different. Through this extension, the mechanism of artificial censoring can be better understood. A nice consequence is that the model assumption on T_2 is extended to the class of transformation models with $h_2(t)$ being unspecified. Thus, the commonly used proportional hazards and proportional odds models can be included as possible model alternatives for T_2. Such an extension greatly increases the

applicability of the artificial censoring technique. The proposed ideas are described in section 2. A model checking procedure is proposed in section 3. Section 4 contains simulation and data analyses. Section 5 provides some concluding remarks. Proofs are given in the Appendices.

2. The proposed methodology

2.1. Inference of $\boldsymbol{\theta}_1$ when T_2 follows a linear regression model

The proposed method was motivated by the log-rank statistic (Kalbfleisch & Prentice, 2002, p. 231). First consider the simplified situation of comparing the marginal distributions of T_2 for two samples with $Z=0,1$. The log-rank statistic can be used to test the equivalence of the two samples. Specifically, let

$$L_2=\frac{1}{n}\sum_{i=1}^{n}\delta_{2i}\left\{I(Z_i=1)-\frac{\sum_{j=1}^{n}I(Y_j\geq Y_i,Z_j=1)}{\sum_{j=1}^{n}I(Y_j\geq Y_i)}\right\},$$

which is an estimator of

$$\int_{t=0}^{\infty}\Pr(C>t)\left\{\mathrm{d}\Pr(T_2\leq t,Z=1)-\frac{\Pr(T_2\geq t,Z=1)}{\Pr(T_2\geq t)}\mathrm{d}\Pr(T_2\leq t)\right\}.$$

When $\Pr(T_2>t\,|\,Z=1)=\Pr(T_2>t\,|\,Z=0)$, L_2 converges to zero. The test statistic L_2 can be inverted to construct an estimating function of θ_2. Notice that $\tilde{T}_2(\theta_2)=h_2(T_2)-\theta_2 Z$ is independent of Z only at the true value $\theta_2=\theta_2^0$. That is,

$$\Pr(\tilde{T}_2(\theta_2)>t\,|\,Z=1)=\Pr(\tilde{T}_2(\theta_2)>t\,|\,Z=0)$$

at the true value $\theta_2=\theta_2^0$. Hence, one can estimate θ_2 by solving

$$U_2(\theta_2)=\frac{1}{n}\sum_{i=1}^{n}\delta_{2i}\left\{Z_i-\frac{\sum_{j=1}^{n}I(\tilde{Y}_j(\theta_2)\geq\tilde{Y}_i(\theta_2))Z_j}{\sum_{j=1}^{n}I(\tilde{Y}_j(\theta_2)\geq\tilde{Y}_i(\theta_2))}\right\}=0, \tag{4}$$

where $\tilde{Y}_i(\theta_2)=\tilde{T}_{2i}(\theta_2)\wedge\tilde{C}_i(\theta_2)=h_2(Y_i)-\theta_2 Z$ and $\tilde{C}(\theta_2)=h_2(C)-\theta_2 Z$. Note that $\delta_2=I(T_2=Y)=I(\tilde{T}_2=\tilde{Y})$. It follows that $U_2(\theta_2^0)$ converges to

$$\int_{t=0}^{\infty}\Pr(\tilde{C}(\theta_2^0)>t)\left\{\mathrm{d}\Pr(\tilde{T}_2(\theta_2^0)\leq t,Z=1)-\frac{\Pr(\tilde{T}_2(\theta_2^0)\geq t,Z=1)}{\Pr(\tilde{T}_2(\theta_2^0)\geq t)}\,\mathrm{d}\Pr(\tilde{T}_2(\theta_2^0)\leq t)\right\}=0.$$

The above idea can be easily extended to test equivalence of the distributions of T_1 for two samples. Specifically, consider

$$L_1=\frac{1}{n}\sum_{i=1}^{n}\delta_{1i}\left\{I(Z_i=1)-\frac{\sum_{j=1}^{n}I(X_j\geq X_i,Z_j=1)}{\sum_{j=1}^{n}I(X_j\geq X_i)}\right\},$$

which is an estimator of

$$\int_{0}^{\infty}\Pr(C>s)\left\{\mathrm{d}\Pr(T_1\leq s,T_1\leq T_2,Z=1)-\frac{\Pr(T_1\geq s,T_1\leq T_2,Z=1)}{\Pr(T_1\geq s,T_1\leq T_2)}\,\mathrm{d}\Pr(T_1\leq s,T_1\leq T_2)\right\}.$$

It is important to note that L_1 converges to zero if

$$\Pr(T_1>s,T_2>t\,|\,Z=1)=\Pr(T_1>s,T_2>t\,|\,Z=0). \tag{5}$$

Our main purpose is to estimate θ_1 under model (1) for general forms of $\mathbf{Z}$. To modify L_1 as an estimating equation of $\boldsymbol{\theta}_1$, we consider the transformed variables $\tilde{T}_j(\boldsymbol{\theta}_j)=$

$h_j(T_j) - \boldsymbol{\theta}_j'\mathbf{Z}, j=1,2$. Notice that $(\tilde{T}_1(\boldsymbol{\theta}_1^0), \tilde{T}_2(\boldsymbol{\theta}_2^0))$ has the same joint distribution as (ϵ_1, ϵ_2), where $\boldsymbol{\theta}_j^0$ is the true value of $\boldsymbol{\theta}_j$ for $j=1,2$. We will use the property that

$$\Pr(T_1(\boldsymbol{\theta}_1^0) > s, T_2(\boldsymbol{\theta}_2^0) > t \mid \mathbf{Z}) = \Pr(\epsilon_1 > s, \epsilon_2 > t)$$

does not depend on $\mathbf{Z}$. Recall that T_1 is subject to censoring by T_2. Hence, the transformed $\tilde{T}_1(\boldsymbol{\theta}_1)$ is subject to censoring by

$$h_1(T_2) - \boldsymbol{\theta}_1'\mathbf{Z} = h_1 \circ h_2^{-1}(\tilde{T}_2(\boldsymbol{\theta}_2) + \boldsymbol{\theta}_2'\mathbf{Z}) - \boldsymbol{\theta}_1'\mathbf{Z},$$

which depends on $\mathbf{Z}$. Under the new scale of $(\tilde{T}_1(\boldsymbol{\theta}_1), \tilde{T}_2(\boldsymbol{\theta}_2))$, the observable region can be written as

$$R_T^{\mathbf{Z}}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = \{(s,t) : s \le h_1 \circ h_2^{-1}(t + \boldsymbol{\theta}_2'\mathbf{Z}) - \boldsymbol{\theta}_1'\mathbf{Z}\}.$$

As the proposed log-rank-type statistics require that observed variables in the analysis have a common distribution, we can only use those which lie in intersection of $R_T^{\mathbf{Z}}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ for all possible values of $\mathbf{Z}$. To attain this, the final observable region becomes

$$\left\{(s,t) : s \le \min_{\text{all}\,\mathbf{z}}[h_1 \circ h_2^{-1}(t + \boldsymbol{\theta}_2'\mathbf{z}) - \boldsymbol{\theta}_1'\mathbf{z}]\right\}.$$

The mechanism of deleting originally non-censored observations, which are located outside the intersection, is called artificial censoring.

Figures 1 and 2 provide graphical illustrations for two-sample comparison without external censoring under $h_1(t) = t, h_2(t) = \log(t)$ and $\theta_1^0 = \theta_2^0 = 1$. That is, we assume a combination of LS and AFT models on $T_1 \mid Z$ and $T_2 \mid Z$ respectively. Under the scale of the transformed variables, the horizontal axis is $h_1(T_1) - \theta_1^0 Z$ and the vertical is $h_2(T_2) - \theta_2^0 Z$. Figure 1 shows the censoring lines for the two groups with $Z=0$ and 1. Notice that the censoring lines are not straight lines if $h_1(t) \neq h_2(t)$. For semi-competing risks data, the observable region is

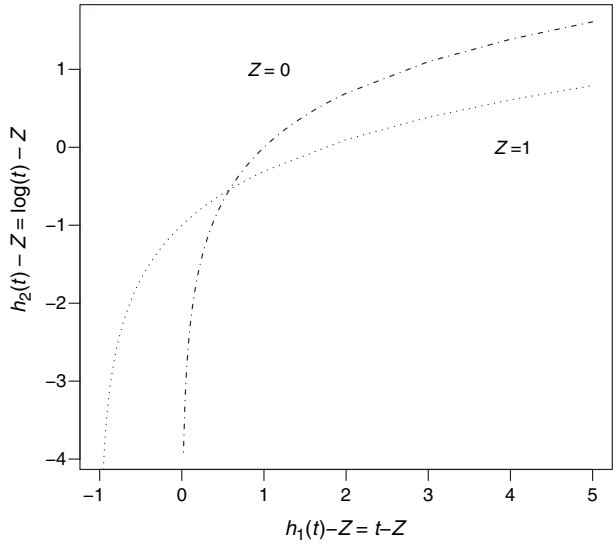

Fig. 1. The censoring lines for the two groups drawn in the transformed scale with $h_1(t) = t$, $h_2(t) = \log(t)$ and $\theta_1^0 = \theta_2^0 = 1$.

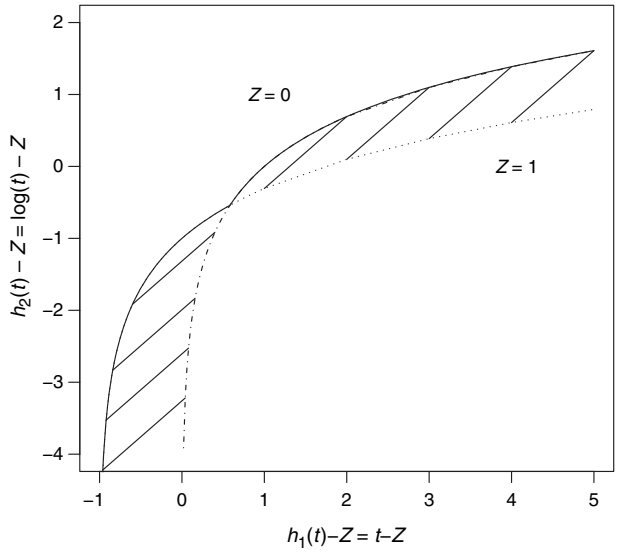

Fig. 2. The region of being artificially censored (shaded) drawn in the transformed scale.

located above the censoring line. We shall use the solid curve shown in Fig. 2 as the artificial censoring line as only observations located above this line would have a common distribution that is independent of Z and hence can be analysed by non-parametric methods which require independent and identically distributed (i.i.d.) observations. The shaded region in Fig. 2 is the part that has been artificially censored. Note that Lin *et al.* (1996) assumed that $h_1(t) = h_2(t)$ which results in parallel straight lines for different groups. Therefore, the artificial censoring formula becomes very simple in their case. Later work, such as Chang (2000), Ghosh & Lin (2003), Lin & Ying (2003) and Peng & Fine (2006) are all based on this simple formula for artificial censoring derived under $h_1(t) = h_2(t)$.

Finally, we reorganize the above idea in an algebraic way. To let $(\tilde{T}_1(\boldsymbol{\theta}_1), \tilde{T}_2(\boldsymbol{\theta}_2))$ fall into the intersection, $\tilde{T}_1(\boldsymbol{\theta}_1)$ should be subject to further right censoring by

$$\begin{aligned} H_{\boldsymbol{\theta}_1,\boldsymbol{\theta}_2}(\tilde{T}_2(\boldsymbol{\theta}_2)) &= \inf_{\mathbf{z}\in\Omega} h_1 \circ h_2^{-1}(\tilde{T}_2(\boldsymbol{\theta}_2) + \boldsymbol{\theta}_2'\mathbf{z}) - \boldsymbol{\theta}_1'\mathbf{z} \\ &= \inf_{\mathbf{z}\in\Omega} h_1 \circ h_2^{-1}(h_2(T_2) - \boldsymbol{\theta}_2'\mathbf{Z} + \boldsymbol{\theta}_2'\mathbf{z}) - \boldsymbol{\theta}_1'\mathbf{z}, \end{aligned} \tag{6}$$

where Ω denotes the set of all possible values of $\mathbf{Z}$. Because the joint distribution of $(\tilde{T}_1(\boldsymbol{\theta}_1^0), \tilde{T}_2(\boldsymbol{\theta}_2^0))$ and the distribution of $H_{\boldsymbol{\theta}_1,\boldsymbol{\theta}_2}(\tilde{T}_2(\boldsymbol{\theta}_2^0))$ do not depend on $\mathbf{Z}$, the distribution of the observed (after artificial censoring) variable $\tilde{T}_1(\boldsymbol{\theta}_1) \wedge H_{\boldsymbol{\theta}_1,\boldsymbol{\theta}_2}(\tilde{T}_2(\boldsymbol{\theta}_2))$ also does not depend on $\mathbf{Z}$ at the true parameter values $(\boldsymbol{\theta}_1^0, \boldsymbol{\theta}_2^0)$.

Now, we propose an estimating function for $\boldsymbol{\theta}_1$ based on semi-competing risks data $(X_i, Y_i, \delta_{1i}, \delta_{2i}, \mathbf{Z}_i)$, $i = 1, \ldots, n$. Define $\tilde{C}(\boldsymbol{\theta}_2) = h_2(C) - \boldsymbol{\theta}_2'\mathbf{Z}$. We let $\tilde{T}_1(\boldsymbol{\theta}_1) = h_1(T_1) - \boldsymbol{\theta}_1'\mathbf{Z}$ be subject to censoring by $H_{\boldsymbol{\theta}_1,\boldsymbol{\theta}_2}(\tilde{T}_2(\boldsymbol{\theta}_2)) \wedge H_{\boldsymbol{\theta}_1,\boldsymbol{\theta}_2}(\tilde{C}(\boldsymbol{\theta}_2))$, where

$$H_{\boldsymbol{\theta}_1,\boldsymbol{\theta}_2}(t) = \inf_{\mathbf{z}\in\Omega} h_1 \circ h_2^{-1}(t + \boldsymbol{\theta}_2'\mathbf{z}) - \boldsymbol{\theta}_1'\mathbf{z}.$$

Define the following hypothetical observations

$$\check{Y}_i(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = H_{\boldsymbol{\theta}_1,\boldsymbol{\theta}_2}(\tilde{T}_{2i}(\boldsymbol{\theta}_2) \wedge \tilde{C}_i(\boldsymbol{\theta}_2)) = H_{\boldsymbol{\theta}_1,\boldsymbol{\theta}_2}(h_2(Y_i) - \boldsymbol{\theta}_2'\mathbf{Z}_i) = H_{\boldsymbol{\theta}_1,\boldsymbol{\theta}_2}(\tilde{Y}_i(\boldsymbol{\theta}_2)),$$

$$\check{X}_i(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = \{h_1(T_{1i}) - \boldsymbol{\theta}_1'\mathbf{Z}_i\} \wedge \check{Y}_i(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = \{h_1(X_i) - \boldsymbol{\theta}_1'\mathbf{Z}_i\} \wedge \check{Y}_i(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$$

and $\tilde{\delta}_i = I(\tilde{X}_i = \tilde{T}_{1i}) = I(\tilde{T}_{1i} \leq \check{Y}_i)$. Note that $\tilde{\delta}_i = 1$ implies that $\delta_{1i} = 1$ but not vice versa. Those observations with $\delta_{1i} = 1$ but $\tilde{\delta}_i = 0$ are artificially censored in the analysis. To estimate $\boldsymbol{\theta} = (\boldsymbol{\theta}_1', \boldsymbol{\theta}_2')'$, we first solve $\mathbf{U}_2(\boldsymbol{\theta}_2) = 0$ to get $\hat{\boldsymbol{\theta}}_2$. Then we solve

$$\mathbf{U}_1(\boldsymbol{\theta}_1, \hat{\boldsymbol{\theta}}_2) = \frac{1}{n}\sum_{i=1}^{n} \tilde{\delta}_i \left\{ \mathbf{Z}_i - \frac{\sum_{j=1}^{n} I(\tilde{X}_j(\boldsymbol{\theta}_1, \hat{\boldsymbol{\theta}}_2) \geq \tilde{X}_i(\boldsymbol{\theta}_1, \hat{\boldsymbol{\theta}}_2))\mathbf{Z}_j}{\sum_{j=1}^{n} I(\tilde{X}_j(\boldsymbol{\theta}_1, \hat{\boldsymbol{\theta}}_2) \geq \tilde{X}_i(\boldsymbol{\theta}_1, \hat{\boldsymbol{\theta}}_2))} \right\} = 0 \tag{7}$$

to get the estimate $\hat{\boldsymbol{\theta}}_1$.

Denote $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\theta}}_1', \hat{\boldsymbol{\theta}}_2')'$ and $\boldsymbol{\theta}^0 = (\boldsymbol{\theta}_1^{0\prime}, \boldsymbol{\theta}_2^{0\prime})'$ as the true parameter values. In Appendix 1, we show that $n^{1/2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^0)$ converges to a mean zero normal distribution. To estimate the limiting covariance matrix, we use a resampling technique as in Lin *et al.* (1996) which can avoid tedious derivation of ϕ given in (A.7). The resampling algorithm is summarized as follows. Given the observed data, we generate a set of i.i.d. standard normal random variables G_i, $i = 1, 2, \ldots, n$, and solve the estimating equations

$$\mathbf{U}_2^*(\boldsymbol{\theta}_2) = n^{-1/2}\sum_{i=1}^{n} \mathbf{W}_{2i}(\boldsymbol{\theta}_2)G_i = 0$$

and

$$\mathbf{U}_1^*(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2^*) = n^{-1/2}\sum_{i=1}^{n} \mathbf{W}_{1i}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2^*)G_i = 0$$

to obtain $\boldsymbol{\theta}^*$, where $\mathbf{W}_{li}; l = 1, 2$; are given in (A.5) and (A.6). The above procedure is repeated, say, B times and the empirical distribution of $\boldsymbol{\theta}_j^*, j = 1, \ldots, B$, where $\boldsymbol{\theta}_j^*$ is the solution at the jth time, would provide a good approximation for the distribution of $\hat{\boldsymbol{\theta}}$ when B is large. Confidence region of $\boldsymbol{\theta}^0$ can be constructed either using the normality result, which requires calculating the sample covariance matrix of $\boldsymbol{\theta}_j^*$, $j = 1, \ldots, B$, or based on the percentiles of $\boldsymbol{\theta}_j^*$, $j = 1, \ldots, B$.

2.2. Inference of $\boldsymbol{\theta}_1$ when T_2 follows a transformation model

Now, we consider the situation that T_2 follows a transformation model in which $h_2(t)$ becomes unknown but the distribution of ϵ_2 is completely specified. The difficulty of this extension comes from the fact that $H_{\boldsymbol{\theta}_1, \boldsymbol{\theta}_2}(t)$ involves specification of $h_2(t)$ or $h_2^{-1}(t)$ which is unknown. Now, we demonstrate how to recover this information. Denote the known survival function of ϵ_2 as $\tilde{S}_2(t) = \Pr(\epsilon_2 > t)$ and define the baseline survival function of T_2 as $S_2(t) = \Pr(T_2 > t \mid Z = 0)$. As $S_2(h_2^{-1}(t)) = \tilde{S}_2(t)$, it follows that $h_2^{-1}(t) = S_2^{-1} \circ \tilde{S}_2(t)$ and $h_2(t) = \tilde{S}_2^{-1} \circ S_2(t)$. Therefore,

$$H_{\boldsymbol{\theta}_1, \boldsymbol{\theta}_2}(t) = \inf_{\mathbf{z} \in \Omega} h_1(S_2^{-1}(\tilde{S}_2(t + \boldsymbol{\theta}_2'\mathbf{z}))) - \boldsymbol{\theta}_1'\mathbf{z}$$

which is still unknown as $S_2(t)$ is unknown. However, if we replace $H_{\boldsymbol{\theta}_1, \boldsymbol{\theta}_2}(t)$ by a uniformly consistent estimator $\hat{H}_{\boldsymbol{\theta}_1, \boldsymbol{\theta}_2}(t)$, then the previous proposed methods are still valid asymptotically. Therefore, when T_2 follows a transformation model, we first find $\hat{S}_2(t)$, which is a uniformly consistent estimator for $S_2(t)$, by a conventional univariate method. Let $\tilde{\mathbf{U}}_2(\boldsymbol{\theta}_2)$ denote $\mathbf{U}_2(\boldsymbol{\theta}_2)$ with $h_2(t)$ being replaced by $\hat{h}_2(t) = \tilde{S}_2^{-1} \circ \hat{S}_2(t)$, and let $\tilde{\mathbf{U}}_1(\boldsymbol{\theta})$ denote $\mathbf{U}_1(\boldsymbol{\theta})$ with $H_{\boldsymbol{\theta}_1, \boldsymbol{\theta}_2}(t)$ being replaced by

$$\hat{H}_{\boldsymbol{\theta}_1, \boldsymbol{\theta}_2}(t) = \inf_{\mathbf{z} \in \Omega} h_1(\hat{S}_2^{-1}(\tilde{S}_2(t + \boldsymbol{\theta}_2'\mathbf{z}))) - \boldsymbol{\theta}_1'\mathbf{z}.$$

Then, we can follow the approach in the previous section by solving the estimating equations $\tilde{\mathbf{U}}_2(\boldsymbol{\theta}_2) = 0$ and $\tilde{\mathbf{U}}_1(\boldsymbol{\theta}) = 0$.

Theorem 1
Assume that when $n \to \infty$, $|\hat{S}_2(t) - S_2(t)| \to 0$ *in probability uniformly on the interval* $[0, t]$ *for all* $t > 0$. *Moreover, assume that the covariate values are bounded. Let* $\tilde{\boldsymbol{\theta}}_1$ *and* $\tilde{\boldsymbol{\theta}}_2$ *denote the solutions to estimating equations* $\tilde{\mathbf{U}}_2(\boldsymbol{\theta}_2) = 0$ *and* $\tilde{\mathbf{U}}_1(\boldsymbol{\theta}) = 0$. *Then* $\tilde{\boldsymbol{\theta}}_1$ *and* $\tilde{\boldsymbol{\theta}}_2$ *has the same asymptotic distribution given in the previous sections where in the variance formula* $h_2(t)$ *and* $H_{\boldsymbol{\theta}_1,\boldsymbol{\theta}_2}(t)$ *are replaced correspondingly by* $\hat{h}_2(t)$ *and* $\hat{H}_{\boldsymbol{\theta}_1,\boldsymbol{\theta}_2}(t)$.

The estimator $\hat{S}_2(t)$ can be obtained using existing methods developed for the imposed transformation model. For example, under the proportional hazards model, one may estimate $S_2(t)$ by

$$\hat{S}_2(t) = \prod_{Y_i \le t} \left\{ 1 - \frac{\exp(\hat{\boldsymbol{\theta}}_2' \mathbf{Z}_i)}{\sum_{j=1}^n I(Y_j \ge Y_i) \exp(\hat{\boldsymbol{\theta}}_2' \mathbf{Z}_j)} \right\}^{\delta_{2i} \exp(-\hat{\boldsymbol{\theta}}_2' \mathbf{Z}_i)} \tag{8}$$

at the uncensored death times Y_i (Kalbfleisch & Prentice, 2002, p. 116). Between these jump points, the estimator can be connected linearly to get a continuous monotone estimator. Under the proportional odds model, one may estimate $S_2(t)$ by the maximum likelihood estimator of Murphy *et al.* (1997). In the special case of two-sample comparison with $\mathbf{Z}$ taking values of 0 and 1, a simple consistent estimator valid for all transformation models is the product-limit estimator (Kalbfleisch & Prentice, 2002, p. 16) using only the observations in the baseline group:

$$\hat{S}_2(t) = \prod_{u \le t} \left\{ 1 - \frac{\sum_{j=1}^n I(Y_j = u, \delta_{2i} = 1, Z_j = 0)}{\sum_{j=1}^n I(Y_j \ge Y_i, Z_j = 0)} \right\}.$$

2.3. Examples

Now, we discuss some examples to illustrate the formula for constructing the artificial censoring line.

Example 1. Both are LS models ($h_j(t) = t, j = 1, 2$). It follows that $H_{\boldsymbol{\theta}_1,\boldsymbol{\theta}_2}(t) = \inf_{\mathbf{z} \in \Omega}(t - \{\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\}'\mathbf{z})$. Lin *et al.* (1996) considered the case of two-sample comparison under this model assumption. It is obvious to see that when $Z = 0, 1$, $H_{\theta_1,\theta_2}(t) = \min\{t + (\theta_2 - \theta_1), t\}$.

Example 2. Both are accelerated failure time models ($h_j(t) = \log(t), j = 1, 2$). It follows that $H_{\boldsymbol{\theta}_1,\boldsymbol{\theta}_2}(t) = \inf_{\mathbf{z} \in \Omega}(t - \{\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\}'\mathbf{z})$. Chang (2000) considered the case of two-sample comparison under this model assumption. When $Z = 0, 1$, $H_{\theta_1,\theta_2}(t) = \min\{t + (\theta_2 - \theta_1), t\}$.

Example 3. T_1 *follows an LS model and* T_2 *follows an AFT model* ($h_1(t) = t; h_2(t) = \log(t)$). It follows that $H_{\boldsymbol{\theta}_1,\boldsymbol{\theta}_2}(t) = \inf_{\mathbf{z} \in \Omega}\{e^{t + \boldsymbol{\theta}_2'\mathbf{z}} - \boldsymbol{\theta}_1'\mathbf{z}\}$. When $Z = 0, 1$, $H_{\theta_1,\theta_2}(t) = \min\{e^{t+\theta_2} - \theta_1, e^t\}$.

Example 4. T_1 *follows an AFT model and* T_2 *follows an LS model* ($h_1(t) = \log(t); h_2(t) = t$). It follows that $H_{\boldsymbol{\theta}_1,\boldsymbol{\theta}_2}(t) = \inf_{\mathbf{z} \in \Omega}\{\log(t + \boldsymbol{\theta}_2'\mathbf{z}) - \boldsymbol{\theta}_1'\mathbf{z}\}$. When $Z = 0, 1$, $H_{\theta_1,\theta_2}(t) = \min\{\log(t + \theta_2) - \theta_1, \log(t)\}$.

Example 5. T_1 *follows an LS model and* T_2 *follows a proportional hazards model.* We have $h_1(t) = t$ and ϵ_2 follows the extreme value distribution with $h_2(t)$ being unspecified. The

survival function of ϵ_2 is $\tilde{S}_2(t)=\exp(-\exp(t))$. The most well-known estimator of $S_2(t)$, under the PH model, is the Nelson–Aalen estimator (8). Therefore,

$$\hat{H}_{\theta_1,\theta_2}(t)=\inf_{\mathbf{z}\in\Omega}\hat{S}_2^{-1}\left[\exp\{-\exp(t+\boldsymbol{\theta}_2'\mathbf{z})\}\right]-\boldsymbol{\theta}_1'\mathbf{z}.$$

2.4. Computational method for multiple covariates

We discuss the computational problem of solving the estimating equations when there are multiple covariates. When there is only one covariate, the root $\hat{\theta}_j$, $j=1,2$, can be found by a simple linear search using the monotonicity of θ in the estimating statistic. When there are multiple covariates, searching for the root in a high-dimensional space may be time consuming. Here, we modify the techniques proposed by Jin *et al.* (2003) who considered solving the weighted version of the following estimating equations

$$\mathbf{U}_2(\boldsymbol{\theta}_2)=\frac{1}{n}\sum_{i=1}^{n}\delta_{2i}Q_i(\boldsymbol{\theta}_2)\left\{\mathbf{Z}_i-\frac{\sum_{j=1}^{n}I(\tilde{Y}_j(\boldsymbol{\theta}_2)\geq\tilde{Y}_i(\boldsymbol{\theta}_2))\mathbf{Z}_j}{\sum_{j=1}^{n}I(\tilde{Y}_j(\boldsymbol{\theta}_2)\geq\tilde{Y}_i(\boldsymbol{\theta}_2))}\right\}=0,$$

where the $Q_i(\boldsymbol{\theta}_2)$ are the weights. For Gehan weights with

$$Q_i(\boldsymbol{\theta}_2)=\frac{1}{n}\sum_{j=1}^{n}I(\tilde{Y}_j(\boldsymbol{\theta}_2)\geq\tilde{Y}_i(\boldsymbol{\theta}_2)),$$

Jin *et al.* (2003) showed that this is equivalent to minimizing the quantity

$$\sum_{i=1}^{n}\sum_{j=1}^{n}\delta_{2i}|\tilde{Y}_i(\boldsymbol{\theta}_2)-\tilde{Y}_j(\boldsymbol{\theta}_2)|+\left|M-\boldsymbol{\theta}_2'\sum_{i=1}^{n}\sum_{j=1}^{n}\delta_{2i}(\mathbf{Z}_i-\mathbf{Z}_j)\right|$$

where M is an extremely large number. This minimization can be achieved by using existing methods of quantile regression.

Our proposed estimating equation for $\boldsymbol{\theta}_2$ corresponds to $Q_i(\boldsymbol{\theta}_2)=1$. Jin *et al.* (2003) suggested to consider

$$\tilde{Q}_i(\boldsymbol{\theta}_2)=\frac{Q_i(\boldsymbol{\theta}_2)n}{\sum_{j=1}^{n}I(\tilde{Y}_j(\boldsymbol{\theta}_2)\geq\tilde{Y}_i(\boldsymbol{\theta}_2))}$$

which is the ratio of the weights versus the Gehan weights. Then finding the root of the estimating equation is equivalent to minimizing

$$\sum_{i=1}^{n}\sum_{j=1}^{n}\delta_{2i}\tilde{Q}_i|\tilde{Y}_i(\boldsymbol{\theta}_2)-\tilde{Y}_j(\boldsymbol{\theta}_2)|+\left|M-\boldsymbol{\theta}_2'\sum_{i=1}^{n}\sum_{j=1}^{n}\delta_{2i}\tilde{Q}_i(\mathbf{Z}_i-\mathbf{Z}_j)\right|$$

for $\tilde{Q}_i$ independent of $\boldsymbol{\theta}_2$. When $\tilde{Q}_i(\boldsymbol{\theta}_2)$ depends on $\boldsymbol{\theta}_2$, Jin *et al.* (2003) proposed the following iterative algorithm: first start with an approximate solution $\hat{\boldsymbol{\theta}}_2^{(0)}$ (for example, the root using Gehan weights) and then update $\hat{\boldsymbol{\theta}}_2^{(k)}$ by minimizing

$$\sum_{i=1}^{n}\sum_{j=1}^{n}\delta_{2i}\tilde{Q}_i(\hat{\boldsymbol{\theta}}_2^{(k-1)})|\tilde{Y}_i(\boldsymbol{\theta}_2)-\tilde{Y}_j(\boldsymbol{\theta}_2)|+\left|M-\boldsymbol{\theta}_2'\sum_{i=1}^{n}\sum_{j=1}^{n}\delta_{2i}\tilde{Q}_i(\hat{\boldsymbol{\theta}}_2^{(k-1)})(\mathbf{Z}_i-\mathbf{Z}_j)\right|.$$

In our case, the estimation equation for $\boldsymbol{\theta}_1$ is more complicated as it contains $\boldsymbol{\theta}_2$ as well as $\boldsymbol{\theta}_1$ and $\tilde{\delta}_i$ depends on both parameters. We propose to modify the iterative algorithm suggested by Jin *et al.* (2003) as follows. First, fit the estimating equation for $\boldsymbol{\theta}_2$ and then

treat $\hat{\boldsymbol{\theta}}_2$ as known in the following steps. Secondly, start with an approximate solution $\hat{\boldsymbol{\theta}}_1^{(0)}$ and then update $\hat{\boldsymbol{\theta}}_1^{(k)}$ by minimizing

$$\sum_{i=1}^{n}\sum_{j=1}^{n}\tilde{\delta}_i(\hat{\boldsymbol{\theta}}_1^{(k-1)})\tilde{Q}_i(\hat{\boldsymbol{\theta}}_1^{(k-1)})|\tilde{X}_i(\hat{\boldsymbol{\theta}}_1)-\tilde{X}_j(\hat{\boldsymbol{\theta}}_1)|+\left|M-\boldsymbol{\theta}_1'\sum_{i=1}^{n}\sum_{j=1}^{n}\tilde{\delta}_{1i}(\hat{\boldsymbol{\theta}}_1^{(k-1)})\tilde{Q}_i(\hat{\boldsymbol{\theta}}_1^{(k-1)})(\mathbf{Z}_i-\mathbf{Z}_j)\right|.$$

In the updating, notice that we have calculated values of $\tilde{\delta}_i$ and $\tilde{Q}_i$ at the previous estimation $\hat{\boldsymbol{\theta}}_1^{(k-1)}$ and fixed their values. In the minimization, the value of $\boldsymbol{\theta}_1$ affects only quantities like $\tilde{X}_i$ and $\tilde{X}_j$.

Similar to the estimator in Jin *et al.* (2003), if $\hat{\boldsymbol{\theta}}_1^{(k)}$ converges when $k\to\infty$, then it converges to the solution of (7). And for any fixed k, $\hat{\boldsymbol{\theta}}_1^{(k)}$ is consistent and asymptotically normal. These results can be proved similarly as in Jin *et al.* (2003). Notice that the main difference here is that we are using $\tilde{\delta}_i$ which also depends on $\boldsymbol{\theta}_1$ unlike δ_{1i}. However, we can write $\tilde{\delta}_i=\delta_{1i}\check{\delta}_i$ where $\check{\delta}_i=I\{\{h_1(X_i)-\boldsymbol{\theta}_1'\mathbf{Z}_i\}\le\check{Y}_i(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2)\}$ is the indicator for artificial censoring. Then solving (7) is related to minimizing

$$\sum_{i=1}^{n}\sum_{j=1}^{n}\delta_{1i}\check{\delta}_i\tilde{Q}_i|\tilde{X}_i-\tilde{X}_j|+\left|M-\boldsymbol{\theta}_1'\sum_{i=1}^{n}\sum_{j=1}^{n}\delta_{1i}\check{\delta}_i\tilde{Q}_i(\mathbf{Z}_i-\mathbf{Z}_j)\right|.$$

We can then consider $\check{\delta}_i\tilde{Q}_i$ as the new weight function replacing the quantity $\tilde{Q}_i$ in each step of the proofs in Jin *et al.* (2003).

3. Model checking and model selection

The model assumptions (1)–(3) can be examined using the residual martingales defined in (A.3) and (A.4). Specifically, one can estimate M_{1i} and M_{2i} by

$$\hat{M}_{1i}(t;\hat{\boldsymbol{\theta}})=N_{1i}(t;\hat{\boldsymbol{\theta}})-\int_{-\infty}^{t}I\{\tilde{X}_i(\hat{\boldsymbol{\theta}})\ge u\}\,\mathrm{d}\hat{\Lambda}_1(u;\hat{\boldsymbol{\theta}}),$$

$$\hat{M}_{2i}(t;\hat{\boldsymbol{\theta}}_2)=N_{2i}(t;\hat{\boldsymbol{\theta}}_2)-\int_{-\infty}^{t}I\{\tilde{Y}_i(\hat{\boldsymbol{\theta}}_2)\ge u\}\,\mathrm{d}\hat{\Lambda}_2(u;\hat{\boldsymbol{\theta}}_2),$$

where

$$\hat{\Lambda}_1(t;\boldsymbol{\theta})=\int_{-\infty}^{t}\frac{\sum_{i=1}^{n}\mathrm{d}N_{1i}(t;\boldsymbol{\theta})}{\sum_{j=1}^{n}I\{\tilde{X}_j(\boldsymbol{\theta})\ge u\}},\quad \hat{\Lambda}_2(t;\boldsymbol{\theta}_2)=\int_{-\infty}^{t}\frac{\sum_{i=1}^{n}\mathrm{d}N_{2i}(t;\boldsymbol{\theta}_2)}{\sum_{j=1}^{n}I\{\tilde{Y}_j(\boldsymbol{\theta}_2)\ge u\}}$$

are the Nelson–Aalen estimators for the cumulative hazard functions of $g_1(t)$ and $g_2(t)$ given in (A.1) and (A.2) in Appendix 1, respectively. The ideas presented here are motivated by the papers of Lin *et al.* (1996) and Lin *et al.* (1993, 2002).

When the model assumptions are correctly specified, the martingale residuals $(\hat{M}_{1i}(t;\hat{\boldsymbol{\theta}})$, $\hat{M}_{2i}(t;\hat{\boldsymbol{\theta}}_2))$, $i=1,\ldots,n$, are distributed around zero. Define the following vector of functions based on the marginal residuals

$$\mathbf{U}(t_1,t_2;\boldsymbol{\theta})=\begin{pmatrix}\mathbf{U}_1(t_1;\boldsymbol{\theta})\\ \mathbf{U}_2(t_2;\boldsymbol{\theta}_2)\end{pmatrix}=n^{-1/2}\sum_{i=1}^{n}\begin{pmatrix}\mathbf{Z}_i\hat{M}_{1i}(t_1;\boldsymbol{\theta})\\ \mathbf{Z}_i\hat{M}_{2i}(t_2;\boldsymbol{\theta}_2)\end{pmatrix}.$$

Under the assumed models, $\mathbf{U}(t_1,t_2;\hat{\boldsymbol{\theta}})$ converges to a mean zero Gaussian process. The corresponding limiting Gaussian process distribution can be approximated by applying similar resampling techniques as introduced earlier. Let

$$\hat{\mathbf{U}}(t_1,t_2)=\mathbf{U}(t_1,t_2;\hat{\boldsymbol{\theta}})-\mathbf{U}(t_1,t_2;\boldsymbol{\theta}^*)$$
$$+n^{-1/2}\sum_{i=1}^{n}G_i\begin{pmatrix}\displaystyle\int_{-\infty}^{t_1}\left\{\mathbf{Z}_i-\frac{\sum_{j=1}^{n}I(\tilde{X}_j(\hat{\boldsymbol{\theta}})\geq u)\mathbf{Z}_j}{\sum_{j=1}^{n}I(\tilde{X}_j(\hat{\boldsymbol{\theta}})\geq u)}\right\}\mathrm{d}\hat{M}_{1i}(u;\hat{\boldsymbol{\theta}})\\ \displaystyle\int_{-\infty}^{t_2}\left\{\mathbf{Z}_i-\frac{\sum_{j=1}^{n}I(\tilde{Y}_j(\hat{\boldsymbol{\theta}}_2)\geq u)\mathbf{Z}_j}{\sum_{j=1}^{n}I(\tilde{Y}_j(\hat{\boldsymbol{\theta}}_2)\geq u)}\right\}\mathrm{d}\hat{M}_{2i}(u;\hat{\boldsymbol{\theta}}_2),\end{pmatrix},$$

where the definitions of G_i and $\boldsymbol{\theta}^*$ are given at the end of section 2.1.

Graphic diagnostics can be conducted by plotting $\mathbf{U}(t_1,t_2;\hat{\boldsymbol{\theta}})$ together with a few, say 20–30, re-sampled realizations of $\hat{\mathbf{U}}(t_1,t_2)$. For the simulations presented in section 4, we used formal lack-of-fit tests based on the deviation statistics $\sup_t|\mathbf{U}_1(t;\hat{\boldsymbol{\theta}})|$ and $\sup_t|\mathbf{U}_2(t;\hat{\boldsymbol{\theta}}_2)|$. We obtain approximate P-values of the tests based on the corresponding empirical probabilities by resampling the process $\hat{\mathbf{U}}(t_1,t_2)$ many times. Theoretical arguments about the validity of this approach are given in appendix 1.

In practice, we may be more interested in selecting the best fit versions of (1) and (2) among several model candidates. We propose a two-stage procedure for selecting the best fit covariate models based on the P-values of the above lack-of-fit tests. Assume that the candidate covariate models for T_1 are $\{M_1^{(1)},\ldots,M_1^{(K_1)}\}$, and the candidate covariate models for T_2 are $\{M_2^{(1)},\ldots,M_2^{(K_2)}\}$, where $M_1^{(k)}$ and $M_2^{(k)}$ have forms of models (1) and (2) respectively. Let $p_2^{(k)}$ denote the P-value based on $\sup_t|\mathbf{U}_2(t;\hat{\boldsymbol{\theta}}_2)|$ for testing model $M_2^{(k)}$. First, we select the covariate model $M_2^{(k)}$ such that $p_2^{(k)}=\max_{i=1}^{K_2}p_2^{(i)}$. For the second step, we compute the lack-of-fit tests for models $\{M_1^{(1)},\ldots,M_1^{(K_1)}\}$ assuming model $M_2^{(k)}$ and denote the computed P-values as $\{p_1^{(1)},\ldots,p_1^{(K_1)}\}$ respectively. We then select the covariate model $M_1^{(j)}$ such that $p_1^{(j)}=\max_{i=1}^{K_1}p_1^{(i)}$.

The following arguments show that the proposed two-stage procedure is consistent for model selection. When $M_2^{(i)}$ is the correct model, the asymptotic analysis in appendix 1 shows that $p_2^{(i)}$ follows a standard uniform distribution. When $M_2^{(i)}$ is not the correct model, the asymptotic analysis shows that $\mathbf{U}_2(t;\hat{\theta}_2)$ is the sum of a non-zero function plus the martingale of a higher order so that $p_2^{(i)}\to 0$ as $n\to\infty$. Hence, the selected model $M_2^{(k)}$ is consistent. Notice that there may be more than one correct model. For example, with the survival time following an exponential distribution, the accelerated failure time model $M_2^{(1)}$ and the PH model $M_2^{(2)}$ are both correct. In this example, the consistency means that with probability one, as $n\to\infty$, $k=1$ or $k=2$. Similarly, the selected model in the second step $M_1^{(j)}$ is also consistent.

The consistency condition means that the model selection procedure will select the correct model if at least one of the candidate model is indeed correct. When all models are incorrect, then all P-values go to zero asymptotically. In other words, if we finally chose a model with a small P-value, it is still a better fit model among all the incorrect alternatives.

4. Numerical results

4.1. Simulation studies

Monte Carlo simulations were conducted to check finite-sample performance of the proposed procedures. The joint distribution of (ϵ_1,ϵ_2) was generated from the Clayton family of the form,

$$S(t_1,t_2)=\Pr(\epsilon_1>t_1,\epsilon_2>t_2)=(S_1(t_1)^{-\alpha}+S_2(t_2)^{-\alpha}-1)^{-1/\alpha}$$

where $S_1(t_1)=\Pr(\epsilon_1>t_1)$, $S_2(t_2)=\Pr(\epsilon_2>t_2)$ and $\alpha>0$. The parameter α is related to Kendall's tau (τ) such that $\tau=\alpha/(2+\alpha)$. The covariate Z was generated from $U(0,1)$. With Z and (ϵ_1,ϵ_2), (T_1,T_2) were obtained after specifying the forms of $h_j(\cdot)$, $j=1,2$, and the parameter

values. The independent censoring variable C was generated from $U(0, 20)$. Finally with $(T_{1i}, T_{2i}, C_i, Z_i)$, we can create semi-competing risks data $(X_i, Y_i, \delta_{1i}, \delta_{2i}, Z_i)$ for $i = 1, \ldots, n$. Two levels of association are considered, $\tau = 0.25$ and 0.50. Two sample sizes with $n = 100$ and 500 are evaluated.

In Table 1, we evaluate the performance of the proposed method and compare it with existing methods. In the first part of Table 1, T_1 follows the LS model and T_2 follows both AFT and PH models. Specifically for the progression time T_1, we set $h_1(t) = t$, $\theta_1^0 = 1$ and $\epsilon_1 \sim \text{Exp}(1)$: the exponential distribution with parameter 1. For the survival time T_2, we set $h_2(t) = \log(t)$, $\theta_2^0 = 1$ and $T_2 \,|\, Z = 0$ to follow Exp(2.1). Hence, ϵ_2 follows an extreme value distribution and T_2 are both AFT and PH models. We compare our proposed inference procedure with those appeared previously in the literature. As the procedure proposed by Robins (1995b) cannot deal with continuous covariates, we consider the method by Lin *et al.* (1996). Specifically, we evaluate three fits for (T_1, T_2), namely (A): (LS, AFT), (B): (AFT, AFT) and (C): (LS, PH). To analyse the second fit (B), we adopt the method proposed by Lin *et al.* (1996) which assumes $h_1(t) = h_2(t) = t$ but, here, is a wrong specification for T_1. The proposed methods presented in sections 2.1 and 2.2 are applied to analyse fits (A) and (C), respectively, both of which are correctly specified.

The results of these three fits based on 1000 simulation runs are summarized in the first half of Table 1. For estimating θ_2, fit (A) and fit (B) yield the same and correct results. However, for estimating θ_1, fit (B) assumes a wrong model on T_1 which results in larger bias and variance and also lower coverage probability compared with the correct fit (A). As the sample size increases from $n = 100$ to 500, the difference between the two fits increases too.

Table 1. *Finite-sample performance of the proposed inference procedures based on* 1000 *simulation runs*

Distribution (correlation)	n	True model		Fitted model		Bias		Variance		Coverage of C.I. (%)	
		T_1	T_2	T_1	T_2	θ_1	θ_2	θ_1	θ_2	θ_1	θ_2
Exponential	100	LS	AFT	LS	AFT	−0.007	−0.013	0.139	0.165	96.3	96.0
($\tau = 0.25$)			(PH)	AFT	AFT	−0.265	−0.013	0.087	0.165	88.8	95.5
				LS	PH	−0.002	0.042	0.138	0.187	95.5	94.3
Exponential		LS	AFT	LS	AFT	−0.037	−0.010	0.136	0.176	97.0	92.5
($\tau = 0.50$)			(PH)	AFT	AFT	−0.321	−0.009	0.077	0.185	85.3	93.5
				LS	PH	−0.025	−0.056	0.136	0.195	95.3	92.8
Exponential	500	LS	AFT	LS	AFT	0.016	0.021	0.027	0.031	96.5	94.5
($\tau = 0.25$)			(PH)	AFT	AFT	−0.256	0.022	0.017	0.031	56.3	95.3
				LS	PH	0.019	0.040	0.027	0.031	95.0	95.0
Exponential		LS	AFT	LS	AFT	0.015	−0.009	0.028	0.035	95.3	93.5
($\tau = 0.50$)			(PH)	AFT	AFT	−0.311	−0.009	0.017	0.035	35.0	93.0
				LS	PH	−0.009	0.007	0.028	0.037	95.5	93.8
Log-normal	100	AFT	PH	AFT	PH	0.057	−0.087	0.231	0.193	94.3	94.5
($\tau = 0.25$)				AFT	AFT	0.069	0.124	0.369	0.130	95.0	92.3
Log-normal		AFT	PH	AFT	PH	0.077	−0.065	0.248	0.191	91.5	91.5
($\tau = 0.50$)				AFT	AFT	0.069	0.157	0.483	0.122	94.3	90.3
Log-normal	500	AFT	PH	AFT	PH	0.009	−0.005	0.051	0.030	95.5	94.3
($\tau = 0.25$)				AFT	AFT	−0.015	0.126	0.055	0.021	96.8	86.0
Log-normal		AFT	PH	AFT	PH	0.013	−0.008	0.055	0.032	95.3	93.5
($\tau = 0.50$)				AFT	AFT	−0.035	0.121	0.064	0.024	98.3	89.0

Comparing fit (A) with fit (C) which are both correct, we see that fit (C) produces larger bias in the estimation of θ_2. A possible explanation is that fit (C) involves extra estimation of the baseline survival function $S_2(t)$. Nevertheless this difference does not carry over to the estimation of θ_1. In fact for estimating θ_1, the performances of fits (A) and (C) are about the same. When the sample size increases, the difference in the biases of θ_2 decreases.

In the second half of Table 1, the error distributions (ϵ_1, ϵ_2) were chosen to make $(T_1, T_2)\,|\,Z=0$ to follow log-normal distributions marginally. The regression model for T_1 follows the AFT model with $h_1(t)=\log(t)$ and $\theta_1^0=1$ and the model for T_2 follows the PH model with $\theta_2^0=1$. The censoring variable C follows $U(0,20)$. Note that under this setting, the PH model is no longer equivalent to the AFT model. Two model fits are considered. Fit (D) assumes AFT on T_1 and the PH model on T_2 which is the correct specification. Fit (E) assumes AFT model on both T_1 and T_2 which is a wrong specification but the best possible fit under the construction of Lin *et al.* (1996).

According to the second half of Table 1, we can see that fit (E), conducted under a wrong model specification on T_2, results in a much larger bias of $\hat{\theta}_2$ than the correct fit (D). The coverage probability of the confidence interval for θ_2 is much lower than the nominal level 95% and becomes worse as the sample size increases from $n=100$ to 500. The model mis-specification on T_2 has little effect on the estimation of θ_1 when the sample size is small. When $n=100$, fits (D) and (E) have similar performance. Fits (D) and (E) both specify the correct regression models on T_1 and hence the estimations of θ_1 are essentially the same for the two fits. As the sample sizes increases to $n=500$, the difference in the ways of artificial censoring becomes more obvious. The model mis-specification on T_2 by fit (E) also leads to a larger bias of $\hat{\theta}_1$ and less reliable results in interval estimation. For the case with $n=500$ and $\tau=0.50$, the 95% confidence interval for θ_1 has a coverage of 98.3%. This exceeds the nominal level 95% more than the simulation variation, indicating that the variance estimator of $\hat{\theta}_1$ in fit (E) is too big.

We also evaluate the performance of the proposed model testing method. Similar to Lin *et al.* (1996), we first test model assumption (2) based on $\sup_t |U_2(t;\hat{\theta}_2)|$ and then test assumption (1) based on $\sup_t |U_1(t;\hat{\boldsymbol{\theta}})|$. As mentioned earlier, fit (B) mis-specifies model (1) while correctly selects model (2). At $\alpha=0.05$ level of significance and $n=100$, the test using $\sup_t |U_1(t;\hat{\boldsymbol{\theta}})|$ rejects this wrong assumption on model (1) 18.3% of the times for $\tau=0.25$ and 19.8% of the times for $\tau=0.50$. Although the performance looks unsatisfactory at $n=100$, the rejection rate becomes 100% for both cases of $\tau=0.25$ and 0.50 when the sample size increases to $n=500$. Fit (E) violates the model assumption (2) instead. At $\alpha=0.05$ and $n=100$, the test using $\sup_t |U_2(t;\hat{\theta}_2)|$ rejects this wrong assumption on model (2) 1.8% of the times for $\tau=0.25$ and 1.8% of the times for $\tau=0.50$. When the sample size increases to $n=500$, the rejection rate becomes 29.0% of the times for $\tau=0.25$ and 24.3% of the times for $\tau=0.50$. So, we can see that it is harder to detect the violation of assumption (2) by fit (E). Although the PH model and AFT model are different for log-normal marginal, the difference is not very big and require large sample size to detect. The results for fits (B) and (E) both showed that the power to detect model violation improves when the sample size increases from $n=100$ to 500.

We then evaluate the model selection procedure proposed in section 3 and summarize the results in Table 2. The first step is to select a model for T_2 based on $\sup_t |U_2(t;\hat{\theta}_2)|$ and then a model for T_1 based on $\sup_t |U_1(t;\hat{\boldsymbol{\theta}})|$. In each step, the model which gives the highest P-value is chosen. The results shows that the method proposed can pick the correct model most of the times. When $n=100$, the wrong fit (B) is selected only 4.3% ($\tau=0.25$) and 2.8% ($\tau=0.50$) of the times. When $n=500$, the wrong fit (B) is never chosen. The wrong fit (E) is selected 31.2% ($\tau=0.25$) and 31.0% ($\tau=0.50$) of the times when $n=100$. The error rate is reduced to 9.0% ($\tau=0.25$) and 13.0% ($\tau=0.50$) when $n=500$.

Table 2. *Finite-sample performance of the proposed model selection method, which picks the models with highest P-values, based on 1000 simulation runs*

		True model		Fitted model		Proportion (%) chosen as the
Distribution (correlation)	n	T_1	T_2	T_1	T_2	true model
Exponential ($\tau=0.25$)	100	LS	AFT	LS	AFT	36.4
			(PH)	AFT	AFT	4.3
				LS	PH	59.3
Exponential ($\tau=0.50$)		LS	AFT	LS	AFT	31.7
			(PH)	AFT	AFT	2.8
				LS	PH	65.5
Exponential ($\tau=0.25$)	500	LS	AFT	LS	AFT	30.2
			(PH)	AFT	AFT	0.0
				LS	PH	69.8
Exponential ($\tau=0.50$)		LS	AFT	LS	AFT	32.0
			(PH)	AFT	AFT	0.0
				LS	PH	68.0
Log-normal ($\tau=0.25$)	100	AFT	PH	AFT	PH	68.8
				AFT	AFT	31.2
Log-normal ($\tau=0.50$)		AFT	PH	AFT	PH	69.0
				AFT	AFT	31.0
Log-normal ($\tau=0.25$)	500	AFT	PH	AFT	PH	91.0
				AFT	AFT	9.0
Log-normal ($\tau=0.50$)		AFT	PH	AFT	PH	87.0
				AFT	AFT	13.0

4.2. Data analysis

The proposed methodology was applied to analyse the data set about bone marrow transplant given in the book by Klein & Moeschberger (2003). The study included 137 patients who received bone marrow transplant in four hospitals from 1984 to 1989. After transplantation, some patients experienced relapse, while others died without relapse. The patients were classified into three groups: acute lymphoblastic leukaemia (ALL), acute myelocytic leukaemia low risk (AML-Low) and AML high risk (AML-High). A more detailed description of the data is given in the book.

To code the three risk groups, we created two covariates: $Z_1=1$ for patients in the ALL group and $Z_1=0$ otherwise; $Z_2=1$ for AML-Low patients and $Z_2=0$ otherwise. Therefore, the AML-High group was chosen as the baseline group with $Z_1=Z_2=0$. Let T_1 denote the time to relapse, T_2 the time to death and $T_1\wedge T_2$ the time to the first event of relapse or death. We would like to investigate whether the three groups differed in T_1, T_2 and $T_1\wedge T_2$. Note that regression analysis for the latter two variables can be performed using univariate methods without dependent censoring. The analysis based on T_1 allows us to investigate whether the relapse time for patients in the three groups were different when the group effect on T_2 had been taken into account.

We fitted six combinations of models (1) and (2). The estimators of $\boldsymbol{\theta}$ were obtained using the methods discussed in section 2.4. Then the method proposed in section 3 was applied to select the best model combination. For each fitted model, we report the P-values (in the parentheses) by using the proposed resampling algorithm based on $\sup_t|\mathbf{U}_1(t;\hat{\boldsymbol{\theta}})|$ and $\sup_t|\mathbf{U}_2(t;\hat{\boldsymbol{\theta}}_2)|$. The results are model (i): LS on T_1 (0.828) and LS on T_2 (0.350); model (ii): LS on T_1 (0.988) and AFT on T_2 (0.275); model (iii): LS on T_1 (0.968) and PH on T_2 (0.880); model (iv): AFT on T_1 (0.970) and LS on T_2 (0.350); model (v): AFT on T_1 (0.788) and AFT on T_2 (0.255); model (vi): AFT on T_1 (0.750) and PH on T_2 (0.880). As the sample size is small, we see that none of the six combinations was rejected by the model checking procedure. Based on the P-value for each fitted model as a measure for the goodness of fit,

we selected for the proportional hazard model for T_2 and the LS model for T_1. That is, we chose model combination (iii).

Under the proportional hazards model for T_2, we obtained $\hat{\boldsymbol{\theta}}_2' = (\hat{\theta}_{21}, \hat{\theta}_{22}) = (0.42, 1.12)$ which implies that AML high-risk patients tended to have the shortest survival time (with 1.5 and 3.1 times higher risk of death compared with ALL and AML-Low patients respectively). Applying the bootstrap procedure, the resulting 95% confidence interval for θ_{22} does not contain zero, indicating that the difference in survival time between AML-Low and AML-High patients is statistically significant. Under the LS model for T_1, $\hat{\boldsymbol{\theta}}_1' = (\hat{\theta}_{11}, \hat{\theta}_{12}) = (-3.17, 30.95)$, but both estimates are not significantly different from zero.

For illustration, we also analysed $T_1 \wedge T_2$ by directly modifying the methodology developed for T_2. Two regression models were fitted. The first one assumed a proportional hazards model on $T_1 \wedge T_2$ and the estimated parameters are given by (0.49, 1.51). The 95% confidence interval for the second component does not contain zero, indicating a significant effect: the risk of experiencing either relapse or death for AML-High patients was 4.5 times higher than that for AML-Low patients. The second analysis assumed an LS model on $T_1 \wedge T_2$ with estimated parameters (20, 111). The 95% confidence interval for the second component does not contain zero either. We see that the analysis based on $T_1 \wedge T_2$ is similar to that based on T_2 as previous analysis shows that the group effect on T_1 is less obvious than that on T_2. In situations where a covariate has opposite effects on the two events, the analysis based on $T_1 \wedge T_2$ may be misleading. As mentioned in the Introduction, we can fit covariate effect model directly on the identifiable quantities such as $T_1 \wedge T_2$ or T_1 only for those $T_1 \leq T_2$. While these quantities have clear clinical meaning, the covariate effect on them compounds the effects on both T_1 and T_2. As this example shows, we do not know if covariate effect on $T_1 \wedge T_2$ is simply a reflection of its effect on T_2. For better biological understanding, we are also interested in the 'net' effect on T_1 with effect on T_2 removed. This 'net' effect has to come from analysis directly on T_1.

The data were re-analysed by merging the ALL and AML-High patients into one category. Specifically, let $Z = 1$ be the indicator for AML low-risk patients and $Z = 0$ for the rest. The two-group analysis chose the PH model for T_2 with $P = 0.846$ and the AFT model for T_1 with $P = 0.706$. Note that the P-value for the previous LS model for T_1 became 0.213. For this model combination we obtained $\hat{\theta}_1 = 1.66$ with 95% confidence interval (0.96, 3.32) and $\hat{\theta}_2 = 0.91$ with 95% confidence interval (0.46, 1.46). Notice the bootstrap confidence intervals do not centre at the point estimates. Both intervals do not contain zero which shows that the differences of the two groups in both relapse time and survival time were significant. The AML-Low patients had longer survival time T_2 with an average of 40% risk of death of other patients. They also tended to have longer relapse time T_1. Relapse times for AML-Low patients on average were 66% longer than those of other patients. Several regression models (i.e. PH, LS and AFT) were also applied to analyse $T_1 \wedge T_2$, all the results show significant covariate effects. However, without the information provided by the analysis of T_1, it is hard to trace the source of the effect. The covariate may affect T_2 only and the effect shows up for $T_1 \wedge T_2$ also.

We also evaluated whether patient's age, a continuous covariate, affected the relapse time and survival time. Again, we considered the six covariate effects model combination (i)–(vi) above for patient age. We selected the AFT model with $P = 0.86$ for measuring the age effect on T_2. Note that the P-values for the LS model and PH model were 0.69 and 0.54 respectively. So, they are also good fits for this data set. Then, we chose the AFT model for the relapse time which gives $P = 0.97$. Note that the LS model also fits well with $P = 0.95$. Under model combination (v), the estimated parameters are $\hat{\theta}_1 = -0.027$ and $\hat{\theta}_2 = -0.029$, but the

95% confidence intervals for the two parameters contain zero. So, age does not have statistically significant effects for this data set.

5. Concluding remarks

Traditionally, semi-competing risks data are analysed under the framework of multi-state models in which the transition rates between the states are the quantities of interest. If the failure time to the intermediate event is of interest, one may study $\Pr(T_1 \wedge T_2 > t)$ or $\Pr(T_1 \le t, T_1 \le T_2)$, both of which, however, are also functions of T_2. In practice, covariates may affect T_1 and T_2 in a different way. Our method estimates the 'net' covariate effect on T_1 that provides information not available through the multi-state model. To illustrate, consider a covariate that reduces the survival time T_2 but does not affect occurrence of the disease time T_1. Then the multi-state analysis will show that the quantities $\Pr(T_1 \wedge T_2 > t)$, $\Pr(T_1 \le t, T_1 \le T_2)$ and $\Pr(T_1 \le T_2)$ are all reduced. We cannot see from the multi-state analysis the fact that the covariate does not affect the biological process related to the occurrence of the disease time T_1. However, zero 'net' covariate effect will provide an indication of this fact. For another example, suppose practitioners are concerned about the side effect of some treatment (say, high-dose chemotherapy) on cancer patients which may increase the risk of death due to other reasons. If the side effect is too serious, the overall survival time T_2 may be shortened, indicating that the treatment is not clinically beneficial in its current form. However, how this treatment affects the cancer recurrence time T_1 after the surgery would influence the decision making on whether the treatment has potential for future improvement. If the 'net' treatment effect shows that T_1 is indeed prolonged, the treatment carries some potential value and future research may be devoted to reduce the side effects. Otherwise, there is no need to study the treatment further.

We have provided a rich class of model choices for describing the covariate effects on both events. A two-stage model checking procedure is also proposed to justify the model assumptions which have nice large sample properties. When the sample size is small, the test may accept several models at the same time but the *P*-value for a fitted model provides a reliable measure for the goodness-of-fit. In our simulations, such a selection strategy seems to work quite well even for small samples.

The purpose of artificial censoring is to create homogeneous observations that can be used in non-parametric statistics such as the suggested log-rank-type statistic $\mathbf{U}_1(\boldsymbol{\theta})$. To obtain a set of observations with the same distribution, what we actually need for $\mathbf{U}_1$ is that the hazard function $g_1(t)$ in (A.1) does not depend on $\mathbf{Z}$ which is a weaker condition than the assumption in (3). As discussed in other related papers, artificial censoring may result in substantial loss in efficiency if the range of $\mathbf{Z}$ spreads too widely. For our method when the artificial censoring variable $H_{\theta_1,\theta_2}(\tilde{T}_2(\boldsymbol{\theta}_2))$ becomes too small, the proportion of artificially censored observations will be excessive. Lin & Ying (2003) suggest using stratification to alleviate the problem of loss in efficiency. Peng & Fine (2006) proposed to apply artificial censoring to a different type of statistic constructed by pairwise comparison that can keep more available data than the log-rank statistics. Note that the method of Peng & Fine (2006) is originally developed under assumptions (1) and (2) with $h_1(t) = h_2(t) = t$ and (3). We think that the proposed idea of generalizing the artificial censoring technique to more flexible regression settings can also be applied to their method. We will leave this extension as a future project.

To simplify the notations, we have used $\mathbf{Z}$ to denote the covariates that affect both T_1 and T_2. Consider the situation that T_1 and T_2 are affected by different covariates $\mathbf{Z}_1$ and $\mathbf{Z}_2$. Let $\mathbf{Z}$ denote the union of $\mathbf{Z}_1$ and $\mathbf{Z}_2$. When solving the estimating equation (4) for T_2, we only

solve for the components of $\boldsymbol{\theta}_2$ that correspond to the covariates $\mathbf{Z}_2$ by restricting the other components to be zero. Similar arguments apply to the estimation of $\boldsymbol{\theta}_1$.

References

Andersen, P. K., Borgan, Ø., Gill, R. D. & Keiding, N. (1993). *Statistical models based on counting processes*. Springer, New York.

Chang, S.-H. (2000). A two-sample comparison for multiple ordered event data. *Biometrics* **56**, 183–189.

Day, R., Bryant, J. & Lefkopolou, M. (1997). Adaptation of bivariate frailty models for prediction, with applications to biological markers as prognostic indicators. *Biometrika* **84**, 45–56.

Fine, J. P. (2001). Regression modeling of competing crude failure probabilities. *Biostatistics* **2**, 85–97.

Fine, J. P. & Gray, R. J. (1999). A proportional hazards model for the subdistribution of a competing risk. *J. Amer. Statist. Assoc.* **94**, 496–509.

Fine, J. P., Jiang, H. & Chappell, R. (2001). On semi-competing risks data. *Biometrika* **88**, 907–919.

Ghosh, D. & Lin, D. Y. (2003). Semiparametric analysis of recurrent events data in the presence of dependent censoring. *Biometrics* **59**, 877–885.

Jin, Z., Lin, D. Y., Wei, L. J. & Ying, Z. (2003). Rank-based inference for the accelerated failure time model. *Biometrika* **90**, 341–353.

Kalbfleisch, J. D. & Prentice, R. L. (2002). *The Statistical Analysis of Failure Time Data*, 2nd edn. Wiley, Hoboken, NJ.

Klein, J. P. & Moeschberger, M. L. (2003). *Survival Analysis: Techniques for Censored and Truncated Data*, 2nd edn. Springer, New York.

Lin, D. Y. & Ying, Z. (2003). Semiparametric regression analysis of longitudinal data with informative drop-outs. *Biostatistics* **4**, 385–398.

Lin, D. Y., Wei, L. J. & Ying, Z. (1993). Checking the Cox model with cumulative sums of martingale-based residuals. *Biometrika* **80**, 557–572.

Lin, D. Y., Robins, J. M. & Wei, L. J. (1996). Comparing two failure time distributions in presence of dependent censoring. *Biometrika* **83**, 381–393.

Lin, D. Y., Wei, L. J. & Ying, Z. (2002). Model-checking techniques based on cumulative residuals. *Biometrics* **58**, 1–12.

Murphy, S. A., Rossini, A. J. & van der Vaart, A. W. (1997). MLE in the proportional odds model. *J. Amer. Statist. Assoc.* **92**, 968–976.

Peng, L. & Fine, J. P. (2006). Rank estimation of accelerated lifetime models with dependent censoring. *J. Amer. Statist. Assoc.* **101**, 1085–1093.

Robins, J. M. (1995a). An analytic method for randomized trials with informative censoring. Part 1. *Lifetime Data Anal.* **1**, 241–254.

Robins, J. M. (1995b). An analytic method for randomized trials with informative censoring. Part II. *Lifetime Data Anal.* **1**, 417–434.

Wang, W. (2003). Estimation of the association parameter for copula models under dependent censoring. *J. Roy. Statist. Soc. Ser. B Statist. Methodol.* **65**, 257–273.

Received December 2006, in final form November 2008

A. Adam Ding, Department of Mathematics, Northeastern University, Boston, MA 02115, USA.
E-mail: a.ding@neu.edu

Appendix 1: asymptotic analysis under the linear regression model

We will extend the proof of Lin *et al.* (1996) to our case. As the proofs are very similar, we only provide an abbreviated proof here. Denote counting processes $N_{1i}(t;\boldsymbol{\theta}) = \tilde{\delta}_i(\boldsymbol{\theta}) I\{\tilde{X}_i(\boldsymbol{\theta}) \le t\}$ and $N_{2i}(t;\boldsymbol{\theta}_2) = \delta_{2i} I\{\tilde{Y}_i(\boldsymbol{\theta}_2) \le t\}$, where $\boldsymbol{\theta} = (\boldsymbol{\theta}_1', \boldsymbol{\theta}_2')'$. The predictable compensators for $\mathrm{d}N_{1i}(t;\boldsymbol{\theta}_1^0, \boldsymbol{\theta}_2^0)$ and $\mathrm{d}N_{2i}(t;\boldsymbol{\theta}_2^0)$ are $I\{\tilde{X}_i(\boldsymbol{\theta}_1^0, \boldsymbol{\theta}_2^0) \ge t\} g_1(t)\,\mathrm{d}t$ and $I\{\tilde{Y}_i(\boldsymbol{\theta}_2^0) \ge t\} g_2(t)\,\mathrm{d}t$, where

$$g_1(t) = \lim_{\Delta t \to 0} \Pr(t \le \tilde{T}_1(\boldsymbol{\theta}_1^0, \boldsymbol{\theta}_2^0) < t + \Delta t \,|\, \tilde{T}_1(\boldsymbol{\theta}_1^0, \boldsymbol{\theta}_2^0) \ge t, H_{\boldsymbol{\theta}_1^0, \boldsymbol{\theta}_2^0}(\tilde{T}_2(\boldsymbol{\theta}_2^0)) \ge t)/\Delta t, \tag{A.1}$$

and

$$g_2(t) = \lim_{\Delta t \to 0} \Pr(t \le \tilde{T}_2(\boldsymbol{\theta}_2^0) < t + \Delta t \,|\, \tilde{T}_2(\boldsymbol{\theta}_2^0) \ge t)/\Delta t, \tag{A.2}$$

which can be considered as the hazard functions of the transformed variables. Notice that $g_j(t)(j=1,2)$ do not depend on the covariate value of $\mathbf{Z}$. The resulting martingales of $N_{1i}(t;\boldsymbol{\theta}^0)$ and $N_{2i}(t;\boldsymbol{\theta}_2^0)$ become

$$M_{1i}(t) = N_{1i}(t;\boldsymbol{\theta}_1^0,\boldsymbol{\theta}_2^0) - \int_{-\infty}^{t} I\{\tilde{X}_i(\boldsymbol{\theta}_1^0,\boldsymbol{\theta}_2^0) \ge u\} g_1(u)\,\mathrm{d}u \tag{A.3}$$

and

$$M_{2i}(t) = N_{2i}(t;\boldsymbol{\theta}_2^0) - \int_{-\infty}^{t} I\{\tilde{Y}_i(\boldsymbol{\theta}_2^0) \ge u\} g_2(u)\,\mathrm{d}u \tag{A.4}$$

respectively, both of which are of mean zero.

In terms of the above counting processes, the proposed estimating functions can be re-expressed as

$$\mathbf{U}_1(\boldsymbol{\theta}) = n^{-1/2} \sum_{i=1}^{n} \int_{-\infty}^{\infty} \{\mathbf{Z}_i - \bar{\mathbf{Z}}^{(1)}(u;\boldsymbol{\theta})\}\,\mathrm{d}N_{1i}(u;\boldsymbol{\theta}),$$

$$\mathbf{U}_2(\boldsymbol{\theta}_2) = n^{-1/2} \sum_{i=1}^{n} \int_{-\infty}^{\infty} \{\mathbf{Z}_i - \bar{\mathbf{Z}}^{(2)}(u;\boldsymbol{\theta}_2)\}\,\mathrm{d}N_{2i}(u;\boldsymbol{\theta}_2),$$

where

$$\bar{\mathbf{Z}}^{(1)}(u;\underline{\theta}) = \frac{\sum_{j=1}^{n} I\{\tilde{X}_j(\underline{\theta}) \ge u\}\mathbf{Z}_j}{\sum_{j=1}^{n} I\{\tilde{X}_j(\underline{\theta}) \ge u\}}, \quad \bar{\mathbf{Z}}^{(2)}(u;\theta_2) = \frac{\sum_{j=1}^{n} I\{\tilde{Y}_j(\boldsymbol{\theta}_2) \ge u\}\mathbf{Z}_j}{\sum_{j=1}^{n} I\{\tilde{Y}_j(\boldsymbol{\theta}_2) \ge u\}}.$$

By the martingale central limit theorem,

$$\mathbf{U}_1(\boldsymbol{\theta}^0) = n^{-1/2} \sum_{i=1}^{n} \int_{-\infty}^{\infty} \{\mathbf{Z}_i - \bar{\mathbf{z}}^{(1)}(u)\}\,\mathrm{d}M_{1i}(u) + o_p(1)$$

and

$$\mathbf{U}_2(\boldsymbol{\theta}_2^0) = n^{-1/2} \sum_{i=1}^{n} \int_{-\infty}^{\infty} \{\mathbf{Z}_i - \bar{\mathbf{z}}^{(2)}(u)\}\,\mathrm{d}M_{2i}(u) + o_p(1)$$

converge to a multivariate mean zero normal distribution, where $\bar{\mathbf{z}}_1(u)$ and $\bar{\mathbf{z}}_2(u)$ are the limits of $\bar{\mathbf{Z}}^{(1)}(u;\boldsymbol{\theta}^0)$ and $\bar{\mathbf{Z}}^{(2)}(u;\boldsymbol{\theta}_2^0)$, respectively, as $n \to \infty$. That is, the vector $\mathbf{U}(\boldsymbol{\theta}^0) = \{\mathbf{U}_1(\boldsymbol{\theta}^0)', \mathbf{U}_2(\boldsymbol{\theta}_2^0)'\}'$ asymptotically has a multivariate normal distribution with zero mean, and its covariance matrix can be estimated consistently by

$$\hat{V} = \frac{1}{n} \sum_{i=1}^{n} \mathbf{W}_i(\hat{\boldsymbol{\theta}}) \mathbf{W}_i(\hat{\boldsymbol{\theta}})',$$

where

$$\begin{aligned}
\mathbf{W}_{1i}(\hat{\boldsymbol{\theta}}) &= \int_{-\infty}^{\infty} \{\mathbf{Z}_i - \bar{\mathbf{Z}}^{(1)}(u;\hat{\boldsymbol{\theta}})\}\,\mathrm{d}\hat{M}_{1i}(u) \\
&= \tilde{\delta}_i(\hat{\boldsymbol{\theta}}) \left\{ \mathbf{Z}_i - \frac{\sum_{j=1}^{n} I(\tilde{X}_j(\hat{\boldsymbol{\theta}}) \ge \tilde{X}_i(\hat{\boldsymbol{\theta}}))\mathbf{Z}_j}{\sum_{j=1}^{n} I(\tilde{X}_j(\hat{\boldsymbol{\theta}}) \ge \tilde{X}_i(\hat{\boldsymbol{\theta}}))} \right\} \\
&\quad - \sum_{l=1}^{n} \frac{\tilde{\delta}_l(\hat{\boldsymbol{\theta}}) I(\tilde{X}_i(\hat{\boldsymbol{\theta}}) \ge \tilde{X}_l(\hat{\boldsymbol{\theta}}))}{\sum_{j=1}^{n} I(\tilde{X}_j(\hat{\boldsymbol{\theta}}) \ge \tilde{X}_l(\hat{\boldsymbol{\theta}}))} \left\{ \mathbf{Z}_i - \frac{\sum_{j=1}^{n} I(\tilde{X}_j(\hat{\boldsymbol{\theta}}) \ge \tilde{X}_l(\hat{\boldsymbol{\theta}}))\mathbf{Z}_j}{\sum_{j=1}^{n} I(\tilde{X}_j(\hat{\boldsymbol{\theta}}) \ge \tilde{X}_l(\hat{\boldsymbol{\theta}}))} \right\},
\end{aligned} \tag{A.5}$$

$$\begin{aligned}\mathbf{W}_{2i}(\hat{\boldsymbol{\theta}}_2) &= \int_{-\infty}^{\infty} \{\mathbf{Z}_i - \bar{\mathbf{Z}}^{(2)}(u;\hat{\boldsymbol{\theta}}_2)\}\, \mathrm{d}\hat{M}_{2i}(u) \\ &= \delta_{2i}\left\{\mathbf{Z}_i - \frac{\sum_{j=1}^n I(\tilde{Y}_j(\hat{\boldsymbol{\theta}}_2) \geq \tilde{Y}_i(\hat{\boldsymbol{\theta}}_2))\mathbf{Z}_j}{\sum_{j=1}^n I(\tilde{Y}_j(\hat{\boldsymbol{\theta}}_2) \geq \tilde{Y}_i(\hat{\boldsymbol{\theta}}_2))}\right\} \\ &\quad - \sum_{l=1}^n \frac{\delta_{2l} I(\tilde{Y}_i(\hat{\boldsymbol{\theta}}_2) \geq \tilde{Y}_l(\hat{\boldsymbol{\theta}}_2))}{\sum_{j=1}^n I(\tilde{Y}_j(\hat{\boldsymbol{\theta}}_2) \geq \tilde{Y}_l(\hat{\boldsymbol{\theta}}_2))}\left\{\mathbf{Z}_i - \frac{\sum_{j=1}^n I(\tilde{Y}_j(\hat{\boldsymbol{\theta}}_2) \geq \tilde{Y}_l(\hat{\boldsymbol{\theta}}_2))\mathbf{Z}_j}{\sum_{j=1}^n I(\tilde{Y}_j(\hat{\boldsymbol{\theta}}_2) \geq \tilde{Y}_l(\hat{\boldsymbol{\theta}}_2))}\right\}, \end{aligned} \tag{A.6}$$

and $\mathbf{W}_i(\hat{\boldsymbol{\theta}}) = (\mathbf{W}_{1i}(\hat{\boldsymbol{\theta}})', \mathbf{W}_{2i}(\hat{\boldsymbol{\theta}}_2)')'$.

In a small neighbourhood of $\boldsymbol{\theta}^0$, we have

$$\mathbf{U}(\boldsymbol{\theta}) = \mathbf{U}(\boldsymbol{\theta}^0) + \phi n^{1/2}(\boldsymbol{\theta} - \boldsymbol{\theta}^0) + o_p(1), \tag{A.7}$$

where ϕ is a $(2K) \times (2K)$ matrix of constants. It follows that $n^{1/2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^0)$ is asymptotically normal with zero mean and covariance matrix $\phi^{-1} V \phi^{-1}$.

Then the same arguments as those in Lin *et al.* (1996) ensures that the re-sampled estimator $\boldsymbol{\theta}^*$ has the same asymptotical distribution as $\boldsymbol{\theta}$. Moreover, $\mathbf{U}(t_1, t_2; \hat{\boldsymbol{\theta}})$ and $\hat{\mathbf{U}}(t_1, t_2)$ have the same asymptotical distribution. Hence, the confidence intervals and the model checking procedure described in sections 2.2 and 3 are valid.

Appendix 2: proof of theorem 1

From the proof in appendix 1, it suffices to prove that $\mathbf{U}_1(\boldsymbol{\theta}) - \tilde{\mathbf{U}}_1(\boldsymbol{\theta}) = o_p(1)$ and $\mathbf{U}_2(\boldsymbol{\theta}_2) - \tilde{\mathbf{U}}_2(\boldsymbol{\theta}_2) = o_p(1)$. Here, we only show $\mathbf{U}_1(\boldsymbol{\theta}) - \tilde{\mathbf{U}}_1(\boldsymbol{\theta}) = o_p(1)$ as the second claim $\mathbf{U}_2(\boldsymbol{\theta}_2) - \tilde{\mathbf{U}}_2(\boldsymbol{\theta}_2) = o_p(1)$ is simpler and can be proved in a similar way.

Let $\dot{Y}_i(\boldsymbol{\theta}) = \hat{H}_{\theta_1, \theta_2}(\hat{h}_2(Y_i) - \boldsymbol{\theta}_2' \mathbf{Z}_i)$, $\dot{X}_i(\boldsymbol{\theta}) = \{h_1(X_i) - \boldsymbol{\theta}_1' \mathbf{Z}_i\} \wedge \dot{Y}_i(\boldsymbol{\theta}) = \tilde{T}_{1i}(\boldsymbol{\theta}_1) \wedge \dot{Y}_i(\boldsymbol{\theta})$ and $\dot{\delta}_i = I(\tilde{T}_{1i}(\boldsymbol{\theta}_1) \leq \dot{Y}_i(\boldsymbol{\theta}))$ denote the plug-in versions of variables $\tilde{Y}_i$, $\tilde{X}_i$ and $\tilde{\delta}_i$ by replacing estimator $\hat{S}_2(t)$ for $S_2(t)$ in them. Then we can write $\tilde{\mathbf{U}}_1(\boldsymbol{\theta})$ as

$$\frac{1}{n}\sum_{i=1}^n \dot{\delta}_i \left\{\mathbf{Z}_i - \frac{\sum_{j=1}^n I(\dot{X}_j(\boldsymbol{\theta}) \geq \dot{X}_i(\boldsymbol{\theta}))\mathbf{Z}_j}{\sum_{j=1}^n I(\dot{X}_j(\boldsymbol{\theta}) \geq \dot{X}_i(\boldsymbol{\theta}))}\right\} = \frac{1}{n}\sum_{i=1}^n \dot{\delta}_i \{\mathbf{Z}_i - \dot{\mathbf{Z}}(\dot{X}_i(\boldsymbol{\theta}))\}$$

where

$$\dot{\mathbf{Z}}(t) = \frac{\sum_{j=1}^n I(\dot{X}_j(\boldsymbol{\theta}) \geq t)\mathbf{Z}_j}{\sum_{j=1}^n I(\dot{X}_j(\boldsymbol{\theta}) \geq t)}.$$

Correspondingly, we rewrite $\mathbf{U}_1(\boldsymbol{\theta})$ as

$$\frac{1}{n}\sum_{i=1}^n \tilde{\delta}_i \left\{\mathbf{Z}_i - \frac{\sum_{j=1}^n I(\tilde{X}_j(\boldsymbol{\theta}) \geq \tilde{X}_i(\boldsymbol{\theta}))\mathbf{Z}_j}{\sum_{j=1}^n I(\tilde{X}_j(\boldsymbol{\theta}) \geq \tilde{X}_i(\boldsymbol{\theta}))}\right\} = \frac{1}{n}\sum_{i=1}^n \tilde{\delta}_i \{\mathbf{Z}_i - \tilde{\mathbf{Z}}(\tilde{X}_i(\boldsymbol{\theta}))\}$$

where

$$\tilde{\mathbf{Z}}(t) = \frac{\sum_{j=1}^n I(\tilde{X}_j(\boldsymbol{\theta}) \geq t)\mathbf{Z}_j}{\sum_{j=1}^n I(\tilde{X}_j(\boldsymbol{\theta}) \geq t)}.$$

Because the covariates are bounded, $|\mathbf{Z}_i| \le M < \infty$ for $i = 1, \ldots, n$. Hence,

$$\begin{aligned}
|\mathbf{U}_1(\boldsymbol{\theta}) - \tilde{\mathbf{U}}_1(\boldsymbol{\theta})| &= \Big|\frac{1}{n}\sum_{i=1}^{n} \tilde{\delta}_i\{\mathbf{Z}_i - \tilde{\mathbf{Z}}(\tilde{X}_i(\boldsymbol{\theta}))\} - \frac{1}{n}\sum_{i=1}^{n} \dot{\delta}_i\{\mathbf{Z}_i - \dot{\mathbf{Z}}(\dot{X}_i(\boldsymbol{\theta}))\}\Big| \\
&\le \Big|\frac{1}{n}\sum_{i:\tilde{\delta}_i \ne \dot{\delta}_i} [\tilde{\delta}_i\{\mathbf{Z}_i - \tilde{\mathbf{Z}}(\tilde{X}_i(\boldsymbol{\theta}))\} - \dot{\delta}_i\{\mathbf{Z}_i - \dot{\mathbf{Z}}(\dot{X}_i(\boldsymbol{\theta}))\}]\Big| \\
&\quad + \Big|\frac{1}{n}\sum_{i:\tilde{\delta}_i = \dot{\delta}_i} [\tilde{\delta}_i\{\mathbf{Z}_i - \tilde{\mathbf{Z}}(\tilde{X}_i(\boldsymbol{\theta}))\} - \dot{\delta}_i\{\mathbf{Z}_i - \dot{\mathbf{Z}}(\dot{X}_i(\boldsymbol{\theta}))\}]\Big| \\
&\le \frac{1}{n}\sum_{i:\tilde{\delta}_i \ne \dot{\delta}_i} 2M + \frac{1}{n}\sum_{i:\tilde{\delta}_i = \dot{\delta}_i} \tilde{\delta}_i |\tilde{\mathbf{Z}}(\tilde{X}_i(\boldsymbol{\theta})) - \dot{\mathbf{Z}}(\dot{X}_i(\boldsymbol{\theta}))| \\
&\le \frac{2M}{n}\sum_{i=1}^{n} |\tilde{\delta}_i - \dot{\delta}_i| + \frac{1}{n}\sum_{i=1}^{n} \tilde{\delta}_i |\tilde{\mathbf{Z}}(\tilde{X}_i(\boldsymbol{\theta})) - \dot{\mathbf{Z}}(\tilde{X}_i(\boldsymbol{\theta}))|. \qquad \text{(A.8)}
\end{aligned}$$

The $\dot{X}_i$ was replaced by $\tilde{X}_i$ in the second term of the last inequality because, when $\tilde{\delta}_i = \dot{\delta}_i = 1$, $\tilde{X}_i(\boldsymbol{\theta}) = \tilde{T}_{1i}(\boldsymbol{\theta}_1) = h_1(T_{1,i}) - \boldsymbol{\theta}_1'\mathbf{Z}_i = \dot{X}_i(\boldsymbol{\theta})$.

The quantity

$$\frac{1}{n}\sum_{i=1}^{n} |\tilde{\delta}_i - \dot{\delta}_i|$$

denotes the proportion of observations that are artificially censored differently under $\dot{Y}_i(\boldsymbol{\theta})$ and $\tilde{Y}_i(\boldsymbol{\theta})$. As

$$\begin{aligned}
\dot{\delta}_i &= I(\tilde{T}_{1i}(\boldsymbol{\theta}_1) \le \inf_{z\in\Omega} [h_1(\hat{S}_2^{-1}(\tilde{S}_2(\hat{h}_2(T_{2,i} \wedge C_i) - \boldsymbol{\theta}_2'\mathbf{Z}_i + \boldsymbol{\theta}_2\mathbf{z}))) - \boldsymbol{\theta}_1'\mathbf{z}]) \\
&= I(\sup_{\mathbf{z}\in\Omega}\{\tilde{T}_{1i}(\boldsymbol{\theta}_1) + \boldsymbol{\theta}_1'\mathbf{z} - h_1[\hat{S}_2^{-1} \circ \tilde{S}_2(\tilde{S}_2^{-1} \circ \hat{S}_2(T_{2,i} \wedge C_i) - \boldsymbol{\theta}_2'(\mathbf{Z}_i - \mathbf{z}))]\} \le 0) \\
&= I(\sup_{\mathbf{z}\in\Omega}\{h_1^{-1}[\tilde{T}_{1i}(\boldsymbol{\theta}_1) + \boldsymbol{\theta}_1'\mathbf{z}] - \hat{S}_2^{-1} \circ \tilde{S}_2[\tilde{S}_2^{-1} \circ \hat{S}_2(T_{2,i} \wedge C_i) - \boldsymbol{\theta}_2'(\mathbf{Z}_i - \mathbf{z})]\} \le 0) \\
&= I(\sup_{\mathbf{z}\in\Omega}\{\tilde{S}_2^{-1} \circ \hat{S}_2 \circ h_1^{-1}[\tilde{T}_{1i}(\boldsymbol{\theta}_1) + \boldsymbol{\theta}_1'\mathbf{z}] - \tilde{S}_2^{-1} \circ \hat{S}_2(T_{2,i} \wedge C_i) + \boldsymbol{\theta}_2'(\mathbf{Z}_i - \mathbf{z})\} \le 0)
\end{aligned}$$

and

$$\tilde{\delta}_i = I(\sup_{\mathbf{z}\in\Omega}\{\tilde{S}_2^{-1} \circ S_2 \circ h_1^{-1}[\tilde{T}_{1i}(\boldsymbol{\theta}_1) + \boldsymbol{\theta}_1'\mathbf{z}] - \tilde{S}_2^{-1} \circ S_2(T_{2,i} \wedge C_i) + \boldsymbol{\theta}_2'(\mathbf{Z}_i - \mathbf{z})\} \le 0),$$

$\hat{S}_2$ converges to S_2 uniformly, the first term in (A.8), namely

$$\frac{2M}{n}\sum_{i=1}^{n} |\tilde{\delta}_i - \dot{\delta}_i|,$$

does converge to zero.

For the second term in (A.8), we can re-express it as the sum of

$$\frac{1}{n}\sum_{\tilde{X}_i(\boldsymbol{\theta}) \le t} \tilde{\delta}_i |\tilde{\mathbf{Z}}(\tilde{X}_i(\boldsymbol{\theta})) - \dot{\mathbf{Z}}(\tilde{X}_i(\boldsymbol{\theta}))| \quad \text{and} \quad \frac{1}{n}\sum_{\tilde{X}_i(\boldsymbol{\theta}) > t} \tilde{\delta}_i |\tilde{\mathbf{Z}}(\tilde{X}_i(\boldsymbol{\theta})) - \dot{\mathbf{Z}}(\tilde{X}_i(\boldsymbol{\theta}))|$$

for any given fixed time t. As, for any t, we have uniform convergence on the interval $[0, t]$ of $\tilde{\mathbf{Z}}(u)$ and $\dot{\mathbf{Z}}(u)$ to the same limiting process $\bar{\mathbf{z}}^{(1)}(u)$ in appendix 1,

$$\frac{1}{n}\sum_{\tilde{X}_i(\boldsymbol{\theta}) \le t} \tilde{\delta}_i |\tilde{\mathbf{Z}}(\tilde{X}_i(\boldsymbol{\theta})) - \dot{\mathbf{Z}}(\tilde{X}_i(\boldsymbol{\theta}))| \to 0$$

in probability. The remainder

$$\frac{1}{n}\sum_{\tilde{X}_i(\boldsymbol{\theta})>t}\tilde{\delta}_i|\tilde{\mathbf{Z}}(\tilde{X}_i(\boldsymbol{\theta}))-\dot{\mathbf{Z}}(\tilde{X}_i(\boldsymbol{\theta}))|\le\frac{2M}{n}\#(\tilde{X}_i(\boldsymbol{\theta})>t)\to 2M\ \Pr(\tilde{X}_i(\boldsymbol{\theta})>t).$$

So, the last expression in (A.8) does not exceed $3M\ \Pr(\tilde{X}_i(\boldsymbol{\theta})>t)$ for large n in probability. As $\Pr(\tilde{X}_i(\boldsymbol{\theta})>t)$ goes to zero as t increases, the last expression in (A.8) converges to zero in probability.

Therefore, $\mathbf{U}_1(\boldsymbol{\theta})-\tilde{\mathbf{U}}_1(\boldsymbol{\theta})=o_p(1)$ and the derivations in appendix 1 still work with $S_2(t)$ being replaced by its estimator $\hat{S}_2(t)$ in all related terms.