Estimation of the Bivariate Survival Function with Generalized Bivariate Right Censored Data Structures

Sunduz Keles* Mark J. van der Laan† James M. Robins‡

*Department of Statistics, University of Wisconsin, Madison
†Division of Biostatistics, School of Public Health, University of California, Berkeley
‡Department of Epidemiology, Harvard School of Public Health

Estimation of the Bivariate Survival Function with Generalized Bivariate Right Censored Data Structures

Sunduz Keles, Mark J. van der Laan, and James M. Robins

Abstract

We propose a bivariate survival function estimator for a general right censored data structure that includes a time dependent covariate process. Firstly, an initial estimator that generalizes Dabrowska's (1988) estimator is introduced. We obtain this estimator by a general methodology of constructing estimating functions in censored data models. The initial estimator is guaranteed to improve on Dabrowska's estimator and remains consistent and asymptotically linear under informative censoring schemes if the censoring mechanism is estimated consistently. We then construct an orthogonalized estimating function which results in a more robust and efficient estimator than our initial estimator. A simulation study demonstrates the performance of the proposed estimators.

1 Introduction

Bivariate survival data arise when study units are paired such as child and parent, or twins or paired organs of the same individual. This paper addresses the survival function estimation in a general data structure which includes time independent and/or dependent covariate processes which are subject to right censoring. Consider a time dependent process $X(t) = (X_1(t), X_2(t))$ where $X_k(t)$ includes a component $R_k(t) = I(T_k \leq t), k = 1, 2$. Let the full data be $X = (\bar{X}_1(T_1), \bar{X}_2(T_2))$, where $\bar{X}_k = \{X_k(s) : s \in [0, t]\}$. We will denote the maximum of T_1 and T_2 with T so that we can represent the full data with $\bar{X}(T)$. Let C_1 and C_2 be two censoring variables. Define $\tilde{T}_k = \min(T_k, C_k)$ and $\Delta_k = I(C_k > T_k), k = 1, 2$. Then the observed data is given by

$$Y = (\tilde{T}_1, \Delta_1, \bar{X}_1(\tilde{T}_1), \tilde{T}_2, \Delta_2, \bar{X}_2(\tilde{T}_2)).$$

Such data structures easily arise in longitudinal studies where study units are monitored over a period of time. In this paper, we are interested in estimating

$$\mu = S(t_1, t_2) = P(T_1 \geq t_1, T_2 \geq t_2)$$

based on n i.i.d. $Y_1, \cdots, Y_n$ copies of Y. Let F_X denote the distribution of the full data X and $G(. \mid X)$ denote the distribution of bivariate censoring variables (C_1, C_2) conditional on X. Then, the distribution of the observed data Y is a function of F_X and $G(. \mid X)$, which we will denote with $P_{F_X,G}$.

There is no previous work on the estimation of such marginal parameter μ with the generalized bivariate right censored data structure. However, estimation of the bivariate distribution of survival times when both study units are subject to random censoring in marginal data structures (no associated covariate process) has received a considerable attention in statistical literature. Some of the proposed nonparametric estimators are Dabrowska (1988), Prentice and Cai (1992), Pruitt (1991) and van der Laan (1996). These estimators employ the independent censoring assumption. Dabrowska, Prentice-Cai and Pruitt estimators are not, in general, efficient estimators. van der Laan's (1996) SOR-NPMLE is globally efficient and typically needs a larger sample size for good performance. A review of most of the available estimators can be found in Pruitt (1993) and van der Laan (1997). Recently, Quale et al. (2001) proposed a new estimator of the bivariate survival function based on the locally efficient estimation theory. Their approach guesses semiparametric models for F_X and $G(. \mid X)$ and the estimator proposed is consistent if either one of the models is correctly specified and locally efficient if both are correctly specified.

The generalized bivariate right censored data structure has two important aspects. Firstly, by utilizing the associated covariate process of the data structure one can allow informative censoring. Secondly, again through utilization of the covariate processes, one can gain efficiency in estimation of the parameter of the interest. This paper is concerned with achieving these properties in estimation of the bivariate survival function. Estimation of the parameter of interest with these type of general bivariate right censored data structures are addressed in great details in Chapter 5 of van der Laan (2002).

Firstly, we will propose an initial estimator for μ that is a generalization of Dabrowska's (1988) estimator. Dabrowska's (1988) estimator, which is developed for marginal data structures, is widely used and depends on a smart representation of the bivariate survival function. It is only efficient under complete independence when survival times and censoring times are all independent of each other (Gill et al., 1995) and becomes inconsistent when there is informative censoring. Our generalization of it deals with informative censoring through utilization of the covariate processes. In our model, we leave the full data distribution completely unspecified and assume a model for $G(. \mid X)$ that will allow dependent censoring. One crucial assumption that we make on the censoring mechanism is that $\bar{G}(T_1, T_2 \mid X) > \delta > 0, F_X - a.e.$ This assumption can be arranged by artificially censoring the data as in van der Laan (1996). For a given τ_1, τ_2 satisfying $\bar{G}(\tau_1, \tau_2 \mid X) > 0$, artificial censoring sets $\tilde{T}_i = \tau_i$ and $\Delta_i = 1$ if $\tilde{T}_i > \tau_i$, $i = 1, 2$. Our initial estimator remains consistent under informative censoring if the censoring mechanism is either known or estimated consistently. Subsequently, we will provide an orthogonalized estimating function that will result in a robust and more efficient estimator.

The general organization of the paper is as follows. In the next section, we will describe the methods to estimate the censoring mechanism in a way that allows dependent censoring. In Section 3, we will briefly review a general methodology of constructing mappings from full data estimating functions to observed data estimating functions, and introduce a new way of obtaining such mappings by using the influence curve of a given regular asymptotically linear (RAL) estimator. In Section 4, we use this method to obtain a generalized Dabrowska's estimator. We will introduce an orthogonalized estimating function and discuss its corresponding estimator in Section 5. The practical performances of the proposed estimators are demonstrated with a simulation study in Section 6. Finally, we end the paper with a summary of conclusions.

2 Modeling the Censoring Mechanism

We will represent the bivariate censoring variable as a bivariate time-dependent process. Let $A_k(t) = I(C_k \leq t)$, $k = 1, 2$, and we define $C_k = \infty$ if $C_k > T_k$, $k = 1, 2$. For a given $A = (A_1, A_2)$ we define $X_A = (\bar{X}_1(C_1), \bar{X}_2(C_2))$. Moreover, let $\bar{X}_A(t) = (\bar{X}_1(C_1 \wedge t), \bar{X}_2(C_2 \wedge t))$ be the part of X_A which is observed by time t. Here $\bar{X}_A(t)$ only depends on A through $\bar{A}(t^-)$. Now, we can represent the observed data as

$$Y = (A, X_A), \tag{1}$$

which corresponds with observing $\bar{Y}(t) = (\bar{A}(t), \bar{X}_A(t))$ over time t. The distribution of the observed data Y is thus indexed by the distribution F_X of X and the conditional distribution of A, given X.

We now consider the modeling and estimation of the bivariate time dependent censoring process A in the discrete and continuous case. Let $g(A \mid X)$ denote the conditional distribution of this bivariate process given the full data X. Firstly, we will assume that $A_k(t)$, $k = 1, 2$, only change value at time points $j = 1, \ldots, p$ (indicating the true chronological time points at which A_k can jump). We will assume

$$g(A(j) \mid \bar{A}(j-1), X) = g(A(j) \mid \bar{A}(j-1), \bar{X}_A(j)), \tag{2}$$

for all $j \in \{1, \ldots, p\}$. This assumption is the analogue of the sequential randomization assumption (SRA) in the causal inference literature (e.g. Robins, 1989b; Robins, 1989a; Robins, 1992; Robins et al., 1994; Robins, 1998; Robins, 1999). Then, we have

$$g(A \mid X) = \prod_{j=1}^{p} g(A(j) \mid \bar{A}(j-1), X)$$

$$= \prod_{j=1}^{p} g_1(A_1(j) \mid \bar{A}(j-1), \bar{X}_A(j)) \prod_{j=1}^{p} g_2(A_2(j) \mid A_1(j), \bar{A}(j-1), \bar{X}_A(j)). \tag{3}$$

Let $\mathcal{F}_1(j) = (\bar{A}(j-1), \bar{X}_A(j))$ and let $\mathcal{F}_2(j) = (A_1(j), \bar{A}(j-1), \bar{X}_A(j))$. Moreover, define $\lambda_k(j \mid \mathcal{F}_k(j)) = P(C_k = j \mid C_k \geq j, \mathcal{F}_k(j))$ to be the conditional hazard of C_k with respect to the history $\mathcal{F}_k$, $k = 1, 2$. Then,

$$\alpha_k(j \mid \mathcal{F}_k(j)) \equiv P(A_k(j) = 1 \mid \mathcal{F}_k(j)) = Y_k(j)\lambda_k(j \mid \mathcal{F}_k(j)),$$

and

$$g_k(A_k(j) \mid \mathcal{F}_k(j)) = \alpha_k(j)^{A_k(j)}(1 - \alpha_k(j))^{1-A_k(j)}, \ k = 1, 2,$$

where $Y_k(j) = I(\tilde{T}_k \geq j)$.

We propose to model the discrete intensities α_k, $k = 1, 2$, with separate models. For example, we could assume a logistic regression model

$$\lambda_k(j \mid \mathcal{F}_k(j)) = \frac{1}{1 + \exp(m(j, W_k(j) \mid \gamma_k))}, \tag{4}$$

where $W_k(j)$ are functions of the observed past $\mathcal{F}_k(j)$. One can model the effect of time j as nonparametric as possible so that this model contains, in particular, the independent censoring model which assumes that (C_1, C_2) is independent of X. If the grid is fine, then the multiplicative intensity model $\lambda_k(j \mid \mathcal{F}_k(j)) = \lambda_0(t)\exp(\gamma_k W_k(j))$ is also appropriate for $k = 1, 2$.

The sequential randomization is a stronger assumption than the well known coarsening at random assumption (CAR) (Heitjan and Rubin, 1991; Jacobsen and Keiding, 1995; Gill et al., 1997). Under CAR, the likelihood $P_{F_X,G}(dy)$ of Y factorizes in an F_X and G-part. Consecutively, the maximum likelihood estimator of $\gamma = (\gamma_1, \gamma_2)$ is given by:

$$\gamma_n = \max_{\gamma}^{-1} \prod_{i=1}^{n} \prod_{j=1}^{C_i} g_{1,\gamma_1}(A_{1i}(j) \mid \mathcal{F}_{1i}(j)) g_{2,\gamma_2}(A_{2i}(j) \mid \mathcal{F}_{2i}(j)).$$

If the models for g_1 and g_2 have no common parameters, then

$$\gamma_{1n} = \max_{\gamma_1}^{-1} \prod_{i=1}^{n} \prod_{j=1}^{C_i} \alpha_{1,\gamma_1}(j \mid \mathcal{F}_{1i}(j))^{dA_1(j)} \{1 - \alpha_{1,\gamma_1}(j \mid \mathcal{F}_{1i}(j))\}^{1-dA_1(j)}$$

and

$$\gamma_{2n} = \max_{\gamma_2}^{-1} \prod_{i=1}^{n} \prod_{j=1}^{C_i} \alpha_{2,\gamma_2}(j \mid \mathcal{F}_{2i}(j))^{dA_2(j)} \{1 - \alpha_{\gamma_2}(j \mid \mathcal{F}_{2i}(j))\}^{1-dA_2(j)}.$$

If we assume the logistic regression model given in (4), then γ_{kn} can be obtained by applying the Splus-function `glm()` or `gam()` with logit link to the pooled sample $(A_{ki}(j), j, W_i(j))$, $i = 1, \ldots, n$, $j = 1, \ldots, m_{ki} \equiv \min(C_{2i}, T_{2i})$, treating it as $N = \sum_i m_{ki}$ i.i.d. observations on a Bernoulli random variable A_k with covariates time t and W.

If $A(t)$ is *continuous*, then one can formally define $g(\bar{A} \mid X)$ as the partial likelihood of the bivariate counting process $A = (A_1, A_2)$ with respect to the observed history $\mathcal{F}(t) = \sigma(\bar{Y}(t-))$ (Andersen et al., 1993)

$$g(\bar{A} \mid X) = \prod_{t,k} \alpha_k(t)^{\Delta A_k(t)} \prod_t (1 - \alpha.(t)dt)^{1-\Delta A.(t)}, \tag{5}$$

where

$$\alpha_k(t) = E(dA_k(t) \mid \mathcal{F}(t))$$

is the intensity of A_k with respect to $\mathcal{F}(t)$ and $\alpha.(t) = \sum_{k=1}^{2} \alpha_k(t)$ is the intensity of $A. = A_1 + A_2$. To estimate $g(\bar{A} \mid X)$ one could assume a multiplicative intensity model ((Andersen et al., 1993)):

$$\alpha_k(t) = Y_k(t)\lambda_k(t \mid \mathcal{F}_k(t)) \equiv Y_k(t)\lambda_{0k}(t)\exp(\gamma W_k(t)),$$

where $Y_k(t)$ is the indicator that A_k is at risk of jumping at time t.

To summarize, by treating the bivariate censoring variable (C_1, C_2) as a bivariate time-dependent process (A_1, A_2) indexed by the same time t as the full-data and assuming sequential randomization, we have succeeded in presenting a flexible modeling framework that allows dependent censoring. Moreover, parameters of these models can be estimated using the standard software.

3 Constructing an Initial Mapping From Full Data Estimating Functions to Observed Data Estimating Functions

In this section, we will briefly review the main ideas of the locally efficient estimation methodology which includes full data estimating functions and mappings into observed data estimating functions (Robins and Rotnitzky, 1992; van der Laan, 2002). Consecutively, we will define a new way of constructing observed data estimating functions using the influence curve of a given RAL estimator.

In order to construct an estimator for the parameter of interest μ based on the observed data $Y_1, \cdots, Y_n$, the estimation problem is firstly considered in the full data model since this class of estimating functions is the foundation of the estimating functions in the observed data model. We will firstly go over estimating functions of the full data model and then link these to the observed data estimating functions.

Estimating functions in the full data model: Let $\mu(F_X)$ be the parameter of interest. We will denote the model for F_X by $\mathcal{M}^F$. Typically, we are interested in estimating functions whose asymptotic behavior is not affected by the choice of the estimators of nuisance parameters. Finding such class of estimating functions requires finding the so called orthogonal complement of the nuisance tangent space at F_X for each $F_X \in \mathcal{M}^F$. The full data nuisance tangent space at F_X, $T^F_{nuis}(F_X)$, is a subspace of Hilbert space $L^2_0(F_X)$ (space of functions of X with finite variance and mean zero endowed with the covariance inner product $< f, g >_{F_X} = E_{F_X} f(X)g(X)$) defined as the linear space spanned by all nuisance scores. A nuisance score is a score function which is obtained by only varying the the nuisance parameters within one dimensional sub-models of F_X (i.e. varying one dimensional sub-models F_ϵ through F_X at $\epsilon = 0$ for which $\frac{d}{d\epsilon}\mu(F_\epsilon) \mid_{\epsilon=0} = 0$). We refer to Bickel et al. (1993) for the general theory of tangent spaces. Let $T^{F,\perp}_{nuis}(F_X)$ be the orthogonal complement of $T^F_{nuis}(F_X)$. The representation of $T^{F,\perp}_{nuis}(F_X), \forall F_X$ plays an important role in constructing the full data estimating functions since this representation generally hints the form of a class of estimating functions. Mainly, one tries to find a class of estimating functions $\{D_h(X \mid \mu, \rho) : h \in \mathcal{H}\}$ such that $D_h, h \in \mathcal{H}$ falls into $T^{F,\perp}_{nuis}(F_X), \forall F_X$ when evaluated at the true parameter values $(\mu(F_X), \rho(F_X))$. Here, $\rho(F_X)$ is a possible nuisance parameter of the full data distribution F_X and $\mathcal{H}$ represents an index set for this class of estimating functions. A template for constructing such a class is given in van der Laan (2002). Ideally, one would like this class to be rich and cover the whole $T^{F,\perp}_{nuis}(F_X)$. In our model, since we will leave the full data distribution completely nonparamateric, there is only one full data estimating function, namely $D(X \mid \mu) = I(T_1 \geq t_1, T_2 \geq t_2) - S(t_1, t_2)$, where $\mu = S(t_1, t_2)$.

Estimating functions in the observed data model: Defining a class of observed data estimating functions requires the notion of orthogonal complement of the observed data nuisance tangent space as in the full data model. Let $\mathcal{G}(CAR)$ be the set of all conditional bivariate distributions $G(. \mid X)$ satisfying CAR. We will then represent the observed data model for the distribution of Y as $\mathcal{M}(CAR) = \{P_{F_X,G} : F_X \in \mathcal{M}^F, G \in \mathcal{G}(CAR)\}$. Next, define $T_{CAR}(P_{F_X,G})$ as the tangent space for G in the model $\mathcal{M}(CAR)$ at $P_{F_X,G}$. $T_{CAR}(P_{F_X,G})$ consists of all functions of the observed data Y that have mean zero given the full data X. Let $D_h \rightarrow IC_0(Y \mid Q_0, G, D_h)$ be an initial mapping from full data estimating functions into observed data estimating functions which satisfies $E_G(IC_0(. \mid Q_0, G, D_h(. \mid \mu, \rho)) \mid X) = D_h(X \mid \mu, \rho), \forall Q_0$. Here, Q_0 refers to a nuisance parameters of the full data model other than ρ. As established in Robins and Rotnitzky (1992), the orthogonal complement of the nuisance tangent space $T^{\perp}_{nuis}(P_{F_X,G})$ in the observed data model $\mathcal{M}(CAR)$ at $P_{F_X,G}$ is given by

$$T^{\perp}_{nuis}(P_{F_X,G}) = \{IC_0(. \mid Q_0(F_X), G, D_h) - \Pi(IC_0(. \mid Q_0(F_X), G, D_h) \mid T_{CAR}(P_{F_X,G})) : D_h \in T^{F,\perp}_{nuis}(F_X)\},$$

where $\Pi(IC_0(. \mid Q_0(F_X), G, D_h) \mid T_{CAR}(P_{F_X,G}))$ represents the projection of $IC_0(. \mid Q_0(F_X), G, D_h)$ onto $T_{CAR}(P_{F_X,G})$. As in the full data model, this representation of $T^{\perp}_{nuis}$ can be used to construct a mapping $IC(Y \mid Q(F_X, G), G, D_h)$ from full data estimating functions $\{D_h(. \mid \mu, \rho), h \in \mathcal{H}\}$ into observed data estimating functions with the property that it falls into $T^{\perp}_{nuis}(P_{F_X,G})$ if evaluated at the true parameter values of the data generating distribution. If the set of the full data estimating functions with the index set $\mathcal{H}$ covers all of the $T^{\perp,F}_{nuis}$ then the set of the corresponding observed data mappings does not exclude any regular asymptotically linear estimator in the model $\mathcal{M}(CAR)$. These mappings result in estimators that are more efficient that the estimators of the initial mappings and are protected against misspecification of either the censoring mechanism or the full data distribution (Robins et al., 2000; van der Laan and Yu, 2001; van der Laan, 2002). This particular way of constructing observed data estimating functions relies on projections onto T_{CAR} which can sometimes be burdensome. For the marginal bivariate right censored data structure, this projection operator does not exist in closed form but it is still possible to implement it algorithmically and this was done by Quale et al. (2001). However, for the general bivariate right censored data structure, the projection operator does not exist in closed form and is computationally much more complicated to implement. Therefore, we propose to project onto $T_{SRA} \subset T_{CAR}$ that is defined as the tangent space for G in the model assuming only SRA. In essence, we will be orthogonalizing the initial mapping with respect to T_{SRA} instead of T_{CAR}. There are two key aspects of these orthogonalized estimating functions. Firstly, they will provide more efficient estimators than the corresponding initial mappings $IC_0(Y \mid Q_0(F_X), G, D)$. However, since one is not projecting onto T_{CAR}, this class of estimating functions will exclude some estimators, and which estimators are included in the class will depend on the choice of initial mapping $D \rightarrow IC_0(. \mid Q_0(F_X), G, D)$. Therefore, we propose a method to construct initial mappings that result in RAL estimators of a specific choice, and hence guarantee that our class of estimators will include the specified RAL estimators with good practical performances.

Initial mappings that correspond with a specified RAL estimator: In order to obtain a mapping $D_h \rightarrow IC_0(Y \mid Q_0(F_X), G, D_h)$ from full data estimating functions into observed data estimating functions that would result in an estimator asymptotically equivalent to a specified RAL estimator for a particular choice h, we use the influence curve, $IC(Y \mid Q_{0,1}(F_X), G)$, of the specified RAL estimator. Influence curve, $IC(Y \mid Q_{0,1}(F_X), G)$, of a RAL estimator μ_n of μ is defined by

$$\sqrt{n}(\mu_n - \mu) = \frac{1}{\sqrt{n}} \sum_i^n IC(Y_i \mid Q_{0,1}(F_X), G) + o_p(1).$$

The parameter $Q_{0,1}(F_X)$ indicates that this influence curve depends on F_X only through a function $Q_{0,1}(F_X)$ of F_X. Since $IC(Y \mid Q_{0,1}(F_X), G)$ is an influence curve it is an element of $T^{\perp}_{nuis}(P_{F_X,G})$. Consecutively, it satisfies

$$E_G(IC(Y \mid Q_{0,1}(F_X), G) \mid X) \in T^{F,\perp}_{nuis}(F_X) \quad \forall F \in \mathcal{M}^F.$$

Let h^* be such that $E_G(IC(Y \mid Q_{0,1}(F_X), G) \mid X) = D_{h^*}(X \mid \mu(F_X), \rho(F_X))$. Let $D_h \rightarrow U(Y \mid Q_{0,2}(F_X), G, D_h)$ be a mapping from full data estimating functions into observed data estimating functions which satisfies $E_G(U(Y \mid Q_{0,2}(F_X), G, D_h) \mid X) = D_h(X \mid \mu, \rho), \forall F_X$. An example of such a mapping would be an inverse probability of censoring weighted mapping and we will use this as an example below. We define

$$IC_{CAR}(Y \mid Q_0(F_X), G) = IC(Y \mid Q_{0,1}(F_X), G) - U(Y \mid Q_{0,2}(F_X), G, D_{h^*}(. \mid \mu(F_X), \rho(F_X))), \quad (6)$$

where $Q_0(F_X) \equiv (Q_{0,1}(F_X), Q_{0,2}(F_X))$. Note that $E_G(IC_{CAR}(Y \mid Q_0(F_X), G) \mid X) = 0 \ \forall F_X \in \mathcal{M}^F$. We now propose the following as an initial mapping from full data estimating functions into observed data estimating functions.

$$IC_0(Y \mid Q_0(F_X), G, D_h(. \mid \mu, \rho)) = U(Y \mid Q_{0,2}(F_X), G, D_h(. \mid \mu, \rho)) + IC_{CAR}(Y \mid Q_0(F_X), G).$$

Note that $E_G(IC_0(Y \mid Q_0(F_X), G, D_h(. \mid \mu, \rho)) \mid X) = D_h(X \mid \mu, \rho)$, $\forall F_X, G$. Then, the corresponding estimating equation is

$$0 = \frac{1}{n} \sum_{i=1}^n IC_0(Y_i \mid Q_{0,n}, G_n, D(. \mid \mu, \rho_n)),$$

where $Q_{0,n}$, ρ_n and G_n are estimates of Q_0, ρ and G respectively. We can then construct a one-step estimator

$$\mu_n^1 = \mu_n^0 + \frac{1}{n} \sum_{i=1}^n IC_0(Y_i \mid Q_{0,n}, G_n, c_n, D(. \mid \mu_n^0, \rho_n)) \quad (7)$$

where $IC_0(Y \mid Q_{0,n} G_n, c_n, D(. \mid \mu_n^0, \rho_n))$ equals

$$\left[- \frac{d}{d\mu} \frac{1}{n} \sum_{i=1}^n IC(Y_i \mid Q_{0,n}, G_n, D(. \mid \mu, \rho_n)) \Big|_{\mu=\mu_n^0} \right]^{-1} IC_0(Y \mid Q_{0,n}, G_n, D(. \mid \mu_n^0, \rho_n)),$$

and μ_n^0 is a $\sqrt{n}$ consistent initial estimator. This is the classical one-step estimator defined in Bickel et al. (1993), i.e. first step in the Newton-Raphson algorithm for solving the estimating equation of interest. The general asymptotic linearity Theorem 8.1 in the Appendix can now be applied to this one-step estimator. Under the regularity conditions of this theorem, if $Q_{0,n}$ converges to a Q_0^1, ρ_n converges to $\rho(F_X)$ and G_n is an efficient estimator of G in the model $\mathcal{G} \subset \mathcal{G}(SRA)$, where $\mathcal{G}(SRA)$ is the set of all conditional bivariate distributions $G(. \mid X)$ satisfying SRA, with tangent space $T_2(P_{F_X,G})$ then μ_n^1 is asymptotically linear with influence curve

$$IC(Y \mid Q_0^1, G, D(. \mid \mu, \rho)) - \Pi(IC(Y \mid Q_0^1, G, D(. \mid \mu, \rho)) \mid T_2(P_{F_X,G})).$$

If $h = h^*$ and Q_0^1 equals $Q_0(F_X)$ then we have the following properties of the one-step estimator. Firstly, if the model $\mathcal{G}$ used for G is a sub model of the model $\mathcal{G}^*$ that the RAL estimator poses for G, μ_n^1 is asymptotically equivalent to the RAL estimator, i.e. it has the same influence curve since $\Pi(IC(Y \mid Q_0(F_X), G, D(. \mid \mu, \rho)) \mid T_2(P_{F_X,G}))$ is zero. Moreover, if $T_2(P_{F_X,G})$ contains scores of submodels which are not in $\mathcal{G}^*$, then μ_n^1 is a more efficient estimator than the RAL estimator.

Consider the following example with the parameter of interest $\mu = S(t_1, t_2)$ in the general bivariate right censored data structure. Since the full data model is nonparametric the only full data estimating function is $I(T_1 \geq t_1, T_2 \geq t_2) - \mu$. Let the $U(Y \mid G, D(. \mid \mu)) = I(T \geq t)\Delta/\bar{G}(T \mid X) - \mu$ be a mapping from full data estimating functions to observed data estimating functions. We use shorthand notation $I(T \geq t)$ to denote $I(T_1 \geq t_1, T_2 \geq t_2)$. Let μ_n be a RAL estimator with the influence curve $IC(Y \mid Q_0(F_X), G)$ and satisfy $E_G(IC(Y \mid Q_0(F_X), G) \mid X) = I(T \geq t) - \mu, \forall Q_0(F_X)$. Then, the corresponding mapping with h^* indexed full data function is,

$$U(Y \mid G, D_{h^*}(. \mid \mu)) = \frac{I(T \geq t)\Delta}{\bar{G}(T \mid X)} - \mu.$$

We have,

$$IC_{CAR}(Y \mid Q_0(F_X), G) = IC(Y \mid Q_0(F_X), G) - \frac{I(T \geq t)\Delta}{\bar{G}(T \mid X)} + \mu(F_X),$$

where we assume that $\mu(F_X)$ depends on F_X only through $Q_0(F_X)$ (i.e. $\mu(F_X) = \mu(Q_0(F_X))$). Then, the proposed initial mapping equals

$$\begin{aligned} IC_0(Y \mid Q_0(F_X), G, D(. \mid \mu)) &= U(Y \mid G, D(. \mid \mu)) + IC_{CAR}(Y \mid Q_0(F_X), G) \\ &= \frac{I(T \geq t)\Delta}{\bar{G}(T \mid X)} - \mu + IC(Y \mid Q_0(F_X), G) - \frac{I(T \geq t)\Delta}{\bar{G}(T \mid X)} + \mu(F_X) \\ &= \mu(F_X) - \mu + IC(Y \mid Q_0(F_X), G). \end{aligned}$$

We solve this estimating equation for μ by setting its empirical mean to zero and replacing $Q_0(F_X)$ and G by their estimates $Q_{0,n}$ and G_n. Here, G_n is a consistent estimate of G according to a model $\mathcal{G} \subset \mathcal{G}(SRA)$.

In the next section, we will construct an observed data estimating function which has $\mu(Q_{0,n})$ equal to the Dabrowska's (1988) estimator and $IC(Y \mid Q_0(F_X), G)$ equal to its influence curve.

4 Generalized Dabrowska's Estimator

A well known estimator of $\mu = S(t_1, t_2)$ based on marginal bivariate right-censored data in the independent censoring model $\mathcal{G}^*$ for G is the Dabrowska's estimator (Dabrowska, 1988; Dabrowska, 1989). The influence curve $IC_{Dabr}(Y \mid F,G)$ of Dabrowska's estimator of $S(t_1,t_2)$ derived in Gill et al. (1995) and van der Laan (1997) is given by

$$
\begin{aligned}
IC_{Dabr}(Y \mid F,G) \;\; = \;\; & S(t_1,t_2)\Bigg\{ - \int_0^{t_1} \frac{I(\tilde{T}_1 \in du, \Delta_1 = 1) - I(\tilde{T}_1 \geq u)P(T_1 \in du \mid T_1 \geq u)}{P_{F,G}(\tilde{T}_1 \geq u)} \\
& - \int_0^{t_2} \frac{I(\tilde{T}_2 \in du, \Delta_2 = 1) - I(\tilde{T}_2 \geq u)P(T_2 \in du \mid T_2 \geq u)}{P_{F,G}(\tilde{T}_2 \geq u)} \\
& + \int_0^{t_1}\int_0^{t_2} \frac{I(\tilde{T}_1 \in du, \tilde{T}_2 \in dv, \Delta_1 = 1, \Delta_2 = 1)}{P_{F,G}(\tilde{T}_1 \geq u, \tilde{T}_2 \geq v)} \\
& - \int_0^{t_1}\int_0^{t_2} \frac{I(\tilde{T}_1 \geq u, \tilde{T}_2 \geq v)P(T_1 \in du, T_2 \in dv \mid T_1 \geq u, T_2 \geq v)}{P_{F,G}(\tilde{T}_1 \geq u, \tilde{T}_2 \geq v)} \\
& - \int_0^{t_1}\int_0^{t_2} \frac{I(\tilde{T}_1 \in du, \tilde{T}_2 \geq v, \Delta_1 = 1)P(T_2 \in dv \mid T_1 \geq u, T_2 \geq v)}{P_{F,G}(\tilde{T}_1 \geq u, \tilde{T}_2 \geq v)} \\
& + \int_0^{t_1}\int_0^{t_2} \frac{I(\tilde{T}_1 \geq u, \tilde{T}_2 \geq v)P(T_1 \in du \mid T_1 \geq u, T_2 \geq v)P(T_2 \in dv \mid T_1 \geq u, T_2 \geq v)}{P_{F,G}(\tilde{T}_1 \geq u, \tilde{T}_2 \geq v)} \\
& - \int_0^{t_1}\int_0^{t_2} \frac{I(\tilde{T}_1 \geq u, \tilde{T}_2 \in dv, \Delta_2 = 1)P(T_1 \in du \mid T_1 \geq u, T_2 \geq v)}{P_{F,G}(\tilde{T}_1 \geq u, \tilde{T}_2 \geq v)} \\
& + \int_0^{t_1}\int_0^{t_2} \frac{I(\tilde{T}_1 \geq u, \tilde{T}_2 \geq v)P(T_1 \in du \mid T_1 \geq u, T_2 \geq v)P(T_2 \in dv \mid T_1 \geq u, T_2 \geq v)}{P_{F,G}(\tilde{T}_1 \geq u, \tilde{T}_2 \geq v)}\Bigg\}.
\end{aligned}
$$

Here F represents the bivariate distribution of (T_1,T_2) and $P_{F,G}(\tilde{T}_1 \geq s, \tilde{T}_2 \geq t) = S(s,t)\bar{G}(s,t)$. Firstly, we note that, as expected by the theory, $E_G(IC_{Dabr}(Y \mid F,G) \mid X) = I(T_1 \geq t_1, T_2 \geq t_2) - \mu$ for all independent censoring distributions $G \in \mathcal{G}^*$ satisfying $\bar{G}(t_1,t_2) > \delta > 0, F_X - a.e.$ In addition, if we replace G in this influence curve by any G satisfying CAR and $\bar{G}(t_1,t_2 \mid X) > 0, F_X - a.e.$, then we still have

$$E_G(IC_{Dabr}(Y \mid F,G) \mid X) = I(T_1 \geq t_1, T_2 \geq t_2) - \mu \tag{8}$$

for all bivariate distributions F, as we show in the Appendix. This explicitly corresponds to replacing $P_{F,G}(\tilde{T}_1 \geq s, \tilde{T}_2 \geq t) = S(s,t)\bar{G}(s,t)$ by $S(s,t)\bar{G}(s,t \mid X)$. We will refer to this resulting influence curve as the modified Dabrowska's influence curve. Note that $Q_{0,1}(F_X) \equiv F$ for this influence curve. Also, note that this modification will enable us to use covariate processes when estimating the censoring mechanism.

Let $U(Y \mid G, D(. \mid \mu)) = I(T_1 \geq t_1, T_2 \geq t_2)\Delta/\bar{G}(T_1,T_2 \mid X) - \mu$. We now define

$$IC_{CAR}(Y \mid F,G) \equiv IC_{Dabr}(Y \mid F,G) - U(Y \mid D(. \mid \mu(F))).$$

By (8), $E_G(IC_{CAR}(Y \mid F,G) \mid X) = 0$ for all $G \in \mathcal{G}(CAR)$ and bivariate distributions F.

We now propose the following observed data estimating function for μ indexed by the true censoring mechanism G and the bivariate distribution F of (T_1,T_2)

$$IC_0(Y \mid F,G,D(\cdot \mid \mu)) = U(Y \mid G, D(\cdot \mid \mu)) + IC_{CAR}(Y \mid F,G).$$

In this estimating equation bivariate distribution F of T_1,T_2 plays role of the nuisance parameter $Q_0(F_X)$ of full data distribution. Note that this estimating function for μ satisfies (8) and at the true μ and F it reduces to the modified Dabrowska's influence curve (and to Dabrowska's influence curve at $G(. \mid X) = G(.)$).

Given consistent estimators F_n of F and G_n of G, let μ_n^0 be the solution of

$$0 = \frac{1}{n}\sum_{i=1}^{n} IC_0(Y_i \mid F_n, G_n, D(\cdot \mid \mu)).$$

Moreover, we have the following closed form solution of this estimating equation

$$\mu_n^0 = \mu(F_n) + \frac{1}{n}\sum_{i=1}^{n} IC_{Dabr}(Y_i \mid F_n, G_n), \tag{9}$$

where $\mu(F_n)$ which will be denoted with μ_n^{Dab} is the Dabrowska's (1988) estimator. We will refer to μ_n^0 as the generalized Dabrowska's estimator. We estimate F nonparametrically by Dabrowska's estimator and this corresponds to replacing hazards in the numerator of IC_{Dabr} by their empirical estimates and $S(t_1, t_2)$ by μ_n^{Dab}. Consequently integrals in this expression simply become sums. Conditional bivariate survival function $\bar{G}(. \mid X)$ of (C_1, C_2) can be estimated by low dimensional models such as frailty models or by methods proposed in Section 2.

Under regularity conditions of Theorem 8.1, μ_n^0 is asymptotically linear with influence curve $IC(Y) \equiv \Pi(IC(\cdot \mid F, G, D(\cdot \mid \mu)) \mid T_2^{\perp}(P_{F_X,G}))$, where $T_2^{\perp}(P_{F_X,G}) \subset T_{CAR}$ is the orthogonal complement of the observed data tangent space of G under the posed model $\mathcal{G}$ for G. Two results emerge from the analysis of this estimator. Firstly, if the posed model for G is the independent censoring model or a submodel of it, then the resulting generalized estimator is asymptotically equivalent to Dabrowska's (1988) estimator since $IC_0(Y \mid F, G, D(. \mid \mu))$ is already orthogonal to the tangent space, T_{indep}, in this model. In fact, under this scenario $\sum_{i=1}^{n} \widehat{IC}_{Dabr}(Y_i \mid F_n, G_n)$ algebraically equals to 0, thus resulting a μ_n^0 that is exactly equal to Dabrowska's (1988) estimator. Secondly, if the tangent space $T_2(P_{F_X,G})$ contains scores which are not in the tangent space of the G in the model posed by the RAL estimator then the generalized estimator is more efficient than the Dabrowska's (1988) estimator even when (C_1, C_2) is independent of X.

5 Orthogonalized Estimating Function and Corresponding Estimator

In this section we discuss the orthogonalization of our initial estimating function $IC_0(Y \mid F, G, D(. \mid \mu)) \equiv IC_0(Y \mid Q_0(F_X), G, \mu)$ to improve efficiency and gain robustness. We define a new estimating function at $G_1 \in \mathcal{G}(SRA)$

$$IC^*(Y \mid Q(F_X, G_1), G_1, \mu) = IC_0(Y \mid Q_0(F_X), G_1, \mu) - IC_{SRA}(Y \mid Q_1(F_X, G_1), G_1), \tag{10}$$

where $IC_{SRA}(Y \mid Q_1(F_X, G_1), G_1) = \Pi(IC_0(Y \mid Q_0(F_X), G_1, \mu) \mid T_{SRA}(P_{F_X,G_1}))$ represents the projection of $IC_0(Y \mid Q_0(F_X), G_1, \mu(F_X))$ onto T_{SRA} at P_{F_X,G_1}. This orthogonalized estimating function has the so called double robustness property (Robins et al., 2000; van der Laan and Yu, 2001; van der Laan, 2002). The double robustness property allows misspecification of either the censoring mechanism $G(. \mid X)$ or the full data distribution F_X. Let F_X^1 and $G_1 \in \mathcal{G}(SRA)$ be guesses of F_X and $G(. \mid X)$, respectively. Then, we have

$$E_{P_{F_X,G}} IC^*(Y \mid Q(F_X^1, G_1), G_1, \mu(F_X)) = 0$$

if either $G_1 = G(. \mid X)$ and $G(. \mid X)$ satisfies the identifiability condition $G(. \mid X) > \delta > 0, F_X - a.e.$ or $F_X^1 = F_X$ and without any further assumptions on $G(. \mid X)$. We refer to the General Double Robustness Theorem in Chapter 1 of van der Laan (2002) for details and the proof of this result. In order to obtain the double robustness property, special care must be paid to estimating the nuisance parameter $Q_1(F_X, G)$ for which we will have an explicit representation below. As we will see shortly, specifying $Q_1(F_X, G)$ correctly is usually a much harder task due to the nature of the projections of $IC_{Dabr}(Y \mid F, G)$ onto T_{SRA}. Therefore, in this section we focus on the scenario where G is estimated consistently and satisfies the identifiability condition. We describe how to obtain an estimator from the orthogonalized estimating function by estimating $Q_1(F_X, G)$ with a regression approach. Then, in the subsequent subsection 5.1, we discuss an alternative way of estimating $Q_1(F_X, G)$ in the form of Monte-Carlo simulations that would allow misspecification of $G(. \mid X)$ when $Q_1(F_X, G)$ is correctly specified.

Given a consistent estimate G_n of G, and an estimate Q_n of $Q(F_X, G)$, we propose to estimate μ with the solution of

$$0 = \frac{1}{n} \sum_{i=1}^{n} IC^*(Y_i \mid Q_n, G_n, \mu). \tag{11}$$

The closed form solution of this estimating equation equals

$$\mu_n^1 = \mu_n^0 - \frac{1}{n} \sum_{i=1}^{n} IC_{SRA}(Y_i \mid Q_{1,n}, G_n), \tag{12}$$

where μ_n^0 is the generalized Dabrowska's estimator introduced in Section 4. This is also equivalent to the one step estimator one would obtain from (11) by using generalized Dabrowska's estimator as the initial estimator. This estimator is asymptotically linear and consistent under the regularity conditions of Theorem 8.1 if G is estimated consistently.

We now present the explicit representation of the projections onto T_{SRA}.

Lemma 5.1 *Define the following functions of* $Y = (\tilde{T}_1, \tilde{T}_2, \Delta_1, \Delta_2, \bar{X}_1(\tilde{T}_1), \bar{X}_2(\tilde{T}_2))$, $A_j(t) = I(C_j \leq t), j = 1, 2$ *where* C_1, C_2 *are two discrete time-variables with finite support contained in* $j = 1, ..., p$*:*

$$dM_{G,k}(u) = I(C_k \in u, \Delta_k = 0) - \lambda_k(u \mid \mathcal{F}_k(u)) I(\tilde{T}_k \geq u),$$

where

$$\mathcal{F}_1(j) = (\bar{A}(j-1), \bar{X}_A(j)) \tag{13}$$

$$\mathcal{F}_2(j) = (A_1(j), \bar{A}(j-1), \bar{X}_A(j)) \tag{14}$$

$$\lambda_k(j \mid \mathcal{F}_k(j)) = P(C_k = j \mid C_k \geq j, \mathcal{F}_k(j)), k = 1, 2. \tag{15}$$

Then, the nuisance tangent space of G *at* $P_{F_X,G}$ *under SRA is given by:*

$$T_{SRA}(P_{F_X,G}) = T_{SRA,1}(P_{F_X,G}) \oplus T_{SRA,2}(P_{F_X,G}),$$

where

$$T_{SRA,k}(P_{F_X,G}) = \left\{ \sum_{j=1}^{p} H(j, \mathcal{F}_k(j)) dM_{G,k}(j) : H \right\}.$$

Subsequently, the projection of any function $V(Y) \in L_0^2(P_{F_X,G})$ *onto* $T_{SRA}(P_{F_X,G})$ *is given by*

$$\Pi(V(Y) \mid T_{SRA}(P_{F_X,G})) = \sum_{k=1}^{2}\sum_{j=1}^{p} H_k(j, \mathcal{F}(j)) dM_{G,k}(j),$$

where

$$H_k(j, \mathcal{F}(j)) = E(V(Y) \mid dA_k(j) = 1, \mathcal{F}_k(j)) - E(V(Y) \mid dA_k(j) = 0, \mathcal{F}_k(j)).$$

Proof: Firstly, by factorization of $g(\bar{A} \mid X)$ into two products under the assumption (2), we have that $T_{SRA}(G) = T_{SRA,1}(G) \oplus T_{SRA,2}(G)$. By the same argument we have that $T_{SRA,k}(G) = T_{SRA,k,1}(G) \oplus \ldots \oplus T_{SRA,k,p}(G)$ where $T_{SRA,k,j}$ is the tangent space for the $j-$th component of the $k-$th product of (3). We will now derive the tangent space $T_{SRA,k,j}$. Let $\mathcal{F}_1(j) = (\bar{A}(j-1), \bar{X}_A(j))$ and $\mathcal{F}_2(j) = (A_1(j), \bar{A}(j-1), \bar{X}_A(j))$ be the observable histories. Let $\alpha_k(j \mid \mathcal{F}(j)) = E(dA_k(j) \mid \mathcal{F}_k(j)), k = 1, 2$. Then the $k-$th product, $k = 1, 2$ in (3) can be represented as:

$$\prod_{j=1}^{p} \alpha_k(j \mid \mathcal{F}_k(j))^{dA_k(j)} \{1 - \alpha_k(j \mid \mathcal{F}_k(j))\}^{1-dA_k(j)}.$$

Note that $\alpha_k(j \mid \mathcal{F}(j)) = \lambda_k(j \mid \mathcal{F}_k(j)) I(\tilde{T}_k \geq j)$ where $\lambda_k(j \mid \mathcal{F}_k(j))$ is the conditional hazard of C_k as defined in (15). Since $\alpha_k(j \mid \mathcal{F}_k(j))^{dA_k(j)} \{1 - \alpha_k(j \mid \mathcal{F}_k(j))\}^{1-dA_k(j)}$ is just a Bernoulli likelihood for the random variable $dA_k(j)$ with probability $\alpha_k(j \mid \mathcal{F}_k(j))$, it follows that the tangent space of $\alpha_k(j \mid \mathcal{F}_k(j))$ is the space of all functions of $(dA_k(j), \mathcal{F}_k(j))$ with conditional mean zero given $\mathcal{F}_k(j)$. It can be shown that any such function V can be written as

$$V(dA_k(j), \mathcal{F}_k(j)) - E[V(dA_k(j), \mathcal{F}_k(j)) \mid \mathcal{F}_k(j)) = \{V(1, \mathcal{F}_k(j)) - V(0, \mathcal{F}_k(j))\} dM_{G,k}(j), \quad (16)$$

where

$$\begin{aligned} dM_{G,k}(j) &= dA_k(j) - \alpha_k(j \mid \mathcal{F}_k(j)) \\ &= I(C_k = j) - \lambda_k(j \mid \mathcal{F}_k(j)) I(\tilde{T}_k \geq j). \end{aligned}$$

Note that $I(C_k = j) = I(C_k = j, \Delta_k = 0)$.

Thus the tangent space of $\alpha_k(j \mid \mathcal{F}_k(j))$ for a fixed j equals

$$T_{SRA,k,j}(P_{F_X,G}) \equiv \{H(\mathcal{F}_k(j))\{dA_k(j) - \alpha_k(j \mid \mathcal{F}(j))\} : H\},$$

where H ranges over all functions of $\mathcal{F}_k(j)$ for which the right-hand side are elements have finite variance. By factorization of the likelihood we have that

$$T_{SRA,k}(P_{F_X,G}) = T_{SRA,k,1} \oplus T_{SRA,k,2} \ldots \oplus T_{SRA,k,p}. \quad (17)$$

Equivalently,

$$T_{SRA,k}(P_{F_X,G}) = \left\{ \sum_{j=1}^{p} H_k(j, \mathcal{F}_k(j)) dM_{G,k}(j) : H \right\}.$$

The projection of any function $V(Y) \in L_0^2(P_{F_X,G})$ onto $T_{SRA,k,j}(P_{F_X,G})$ is obtained by first projecting on all functions of $dA_k(j)$ and the subtracting from this its conditional expectation given $\mathcal{F}_k(j)$. Thus, we have

$$\begin{aligned}\Pi(V(Y) \mid T_{SRA}(P_{F_X,G})) &= \sum_{k=1}^{2}\sum_{j=1}^{p}\{E(V(Y) \mid dA_k(j)=1,\mathcal{F}_k(j)) \\ &- E(V(Y) \mid dA_k(j)=0,\mathcal{F}_k(j))\}\, dM_{G,k}(j).\end{aligned}$$

This completes the proof. □

Application of this lemma with $V(Y) = IC_0(Y \mid Q_0(F_X), G, \mu)$ gives

$$\begin{aligned}\Pi(IC_0(Y \mid Q_0(F_X), G, \mu(F_X)) \mid T_{SRA}) &= \sum_{k=1}^{2}\sum_{j=1}^{p}\{E(IC_{Dabr}(Y \mid F,G) \mid dA_k(j)=1,\mathcal{F}_k(j)) &(18)\\ &- E(IC_{Dabr}(Y \mid F,G) \mid dA_k(j)=0,\mathcal{F}_k(j))\}\, dM_{G,k}(j). &(19)\end{aligned}$$

Let $Q_k(dA_k(j), \mathcal{F}_k(j)) \equiv E(IC_{Dabr}(Y \mid F,G) \mid dA_k(j), \mathcal{F}_k(j))$ for $k=1,2$. Note that $Q_1(F_X,G)$ in (10) represents $Q_k(dA_k(j),\mathcal{F}_k(j))$, $k=1,2$, $j=1,\cdots,p$. Then, following the representation of the projections onto T_{SRA}, the explicit form of the one-step estimator given in (12) becomes

$$\mu_n^1 = \mu_n^0 - \sum_{k=1}^{2}\sum_{j=1}^{p}\left\{\hat{Q}_k(dA_k(j)=1,\mathcal{F}_k(j)) - \hat{Q}_k(dA_k(j)=0,\mathcal{F}_k(j))\right\} dM_{G_n,k}(j),$$

where $\hat{Q}_k(dA_k(j),\mathcal{F}_k(j))$ is an estimate of $Q_k(dA_k(.),\mathcal{F}_k(.))$, $k=1,2$ at j. One way to obtain such estimates is to estimate the corresponding conditional expectations parametrically or semiparametrically by regressing the estimate of $IC_{Dabr}(Y \mid F,G)$ onto time variable j and covariates extracted from the past $(A_k(t),\mathcal{F}_k(t))$. Making this conditional expectation dependent on time covariate j allows us to evaluate it at all required j. Note that in order to avoid technical conditions such as measurability in establishing the projections onto T_{SRA}, we assumed that C_1 and C_2 are discrete on a grid $\{1,\cdots,p\}^2$. Since the time points can be chosen to be the grid points of an arbitrarily fine partition this assumption can be made without loss of practical applicability.

We have observed in our simulation studies that models for conditional expectations in projection terms are often misspecified, leading to inconsistent estimates for $Q_1(F_X,G)$. It is then desirable to prevent a possible efficiency loss due to projections. We apply the general modification proposed by Robins and Rotnitzky (1992) and use the following estimating function

$$IC^*(Y \mid Q(F_X,G), c_{nu}, G, \mu) = IC_0(Y \mid Q_0(F_X), G, \mu) - c_{nu} IC_{SRA}(Y \mid Q_1(F_X,G), G), \qquad (20)$$

where c_{nu} is defined as

$$\frac{E_{P_{F_X,G}}\{IC_0(Y \mid Q_0(F_X),G,\mu) IC_{SRA}(Y \mid Q_1(F_X,G),G)\}}{E_{P_{F_X,G}}\{IC_{SRA}(Y \mid Q_1(F_X,G),G)^2\}}$$

so that $c_{nu} IC_{SRA}(Y \mid Q_1(F_X,G),G) = \Pi(IC_0(Y \mid Q_0(F_X),G,\mu) \mid T_{SRA})$. Note that when $IC_{SRA}(Y \mid Q_1(F_X,G),G)$ is the projection of $IC_0(Y \mid Q_0(F_X),G,\mu)$ onto T_{SRA}, c_{nu} equals 1. In other words, this adjustment will only have an effect when $Q_1(F_X,G)$ is misspecified. Moreover, it guarantees that the resulting estimating function $IC^*(Y \mid Q(F_X, G), c_{nu}, G, \mu)$ is more efficient than the initial estimating function $IC_0(Y \mid Q_0(F_X), G, \mu)$ even when $Q_1(F_X, G)$ is estimated inconsistently (Robins and Rotnitzky, 1992; van der Laan, 2002). c_{nu} can be estimated by taking empirical expectations of the estimated $IC_0(Y \mid Q_0(F_X), G, \mu)$ and $IC_{SRA}(Y \mid Q_1, G))$. Specifically, estimate $c_{nu,n}$ of c_{nu} is given by

$$c_{nu,n} = \frac{\sum_{i=1}^n IC_0(Y_i \mid Q_0(F_n), G_n, \mu(F_n)) IC_{SRA}(Y_i \mid Q_{1,n}, G_n)}{\sum_{i=1}^n IC_{SRA}(Y_i \mid Q_{1,n}, G_n) IC_{SRA}(Y_i \mid Q_{1,n}, G_n)^T},$$

where $Q_{1,n}$ is the estimate of $Q_1(F_X, G)$. One practical aspect of this adjustment parameter is that it provides a way of monitoring the goodness of fit of the projection term $IC_{SRA}(Y \mid Q_{1,n}, G_n)$. Since $c_{nu,n}$ will be approximately 1 at the best fit of the projection term, one can use this property to choose the best fit.

5.1 Estimation of $Q_1(F_X, G)$ by Monte-Carlo Simulations

In this subsection, we will discuss a Monte-Carlo simulation method to estimate the nuisance parameter $Q_1(F_X, G)$. This approach requires guessing a low dimensional model for the full data distribution F_X and the censoring mechanism G, respectively. As a result, the corresponding estimator of the orthogonalized estimating function (10) remains consistent if either of the guessed models is correctly specified. We will use the longitudinal representation of observed data with the notation of Section 2 over a discrete time axes ($j = 1, \cdots, p$) given as

$$\begin{aligned} Y \; = \; & X_1(1), X_2(1), A_1(1), A_2(1), X_{1,\bar{A}(1)}(2), X_{2,\bar{A}(1)}(2), A_1(2), A_2(2), \\ & \cdots, X_{1,\bar{A}(p-1)}(p), X_{2,\bar{A}(p-1)}(p), A_1(p), A_2(p). \end{aligned}$$

Define $L_1(j) = X_{1,\bar{A}(j-1)}(j)$ and $L_2(j) = X_{2,\bar{A}(j-1)}(j)$, then

$$Y = L_1(1), L_2(1), A_1(1), A_2(1), \cdots, L_1(p), L_2(p), A_1(p), A_2(p).$$

Under SRA, the likelihood of the observed data is given by

$$\begin{aligned} dP_{F_X,G}(Y) \; = \; & \prod_{j=1}^p \left[f_1(L_1(j) \mid \bar{L}(j-1), \bar{A}(j-1)) f_2(L_2(j) \mid L_1(j), \bar{L}(j-1), \bar{A}(j-1)) \right. \\ & \left. g_1(A_1(j) \mid \bar{A}(j-1), \bar{L}(j)) g_2(A_2(j) \mid A_1(j), \bar{A}(j-1), \bar{L}(j)) \right], \end{aligned}$$

where $\bar{L}(j) = (\bar{L}_1(j), \bar{L}_2(j))$. Since the likelihood factorizes under SRA, we have that the F_X and G part of the likelihood are given by

$$Q(F_X) \; = \; \prod_{j=1}^p f_1(L_1(j) \mid \bar{L}(j-1), \bar{A}(j-1)) \prod_{j=1}^p f_2(L_2(j) \mid L_1(j), \bar{L}(j-1), \bar{A}(j-1)), \quad (21)$$

$$g(\bar{A} \mid X) \; = \; \prod_{j=1}^p g_1(A_1(j) \mid \bar{A}(j-1), \bar{L}(j)) \prod_{j=1}^p g_2(A_2(j) \mid A_1(j), \bar{A}(j-1), \bar{L}(j)). \quad (22)$$

The modeling and estimation strategies proposed for the censoring mechanism in Section 2 applies to both of these likelihood parts. Let $(f_{1,\theta_1}, f_{2,\theta_2})$ and $(g_{1,\eta_1}, g_{2,\eta_2})$ be parametric or semi-parametric models for F_X and G part of the likelihood. Let $(\theta_{1,n}, \theta_{2,n})$ and $(\eta_{1,n}, \eta_{2,n})$ be the maximum likelihood estimators and $Q_n = Q(\theta_{1,n}, \theta_{2,n})$ and $G_n = G_{\eta_{1,n},\eta_{2,n}}$ be the corresponding estimators of $Q(F_X)$ and G, respectively. Now, one can evaluate the conditional expectations in the projection terms (18) and (19) under the known law P_{Q_n,G_n} with a Monte-Carlo simulation. Consider a particular observation Y and let j be fixed. The following is the algorithm for performing Monte-Carlo simulation on this observation:

- **SIMULATE**: This step simulates the complete observation from a fixed history. Set $b = 1$ and $m = j + 1$.
  1. *With history* $(dA_1(j) = 1, \mathcal{F}_1(j))$*:* Set $dA_1(j) = 1$ and
     - (A) Generate $A_2(m-1)$ from $g_{2,\eta_{2,n}}(. \mid A_1(m-1) = 1, \bar{A}(m-2), \bar{L}(m-1))$.
     - (B) Generate $L_1(m)$ from $f_{1,\theta_{1,n}}(. \mid \bar{L}(m-1), \bar{A}(m-1))$.
     - (C) Generate $L_2(m)$ from $f_{2,\theta_{2,n}}(. \mid L_1(m), \bar{L}(m-1), \bar{A}(m-1))$.
     - Set $m = m + 1$ and repeat steps (A), (B), (C) until the complete data structure denoted by $Y_{1,b}^{1,*}$ is observed.
  2. *With history* $(dA_1(j) = 0, \mathcal{F}_1(j))$*:*
     - (A) Generate $A_2(m-1)$ from $g_{2,\eta_{2,n}}(. \mid A_1(m-1) = 0, \bar{A}(m-2), \bar{L}(m-1))$.
     - (B) Generate $L_1(m)$ from $f_{1,\theta_{1,n}}(. \mid \bar{L}(m-1), \bar{A}(m-1))$.
     - (C) Generate $L_2(m)$ from $f_{2,\theta_{2,n}}(. \mid L_1(m), \bar{L}(m-1), \bar{A}(m-1))$.
     - (D) Generate $A_1(m)$ from $g_{1,\eta_{1,n}}(. \mid \bar{A}(m-1), \bar{L}(m))$.
     - Set $m = m+1$ and repeat steps (A), (B), (C), (D) until the complete data structure denoted by $Y_{0,b}^{1,*}$ is observed.
  3. *With history* $(dA_2(j) = 1, \mathcal{F}_2(j))$*:* Set $dA_2(j) = 1$ and
     - (A) Generate $L_1(m)$ from $f_{1,\theta_{1,n}}(. \mid \bar{L}(m-1), \bar{A}(m-1))$.
     - (B) Generate $L_2(m)$ from $f_{2,\theta_{2,n}}(. \mid L_1(m), \bar{L}(m-1), \bar{A}(m-1))$.
     - (C) Generate $A_1(m)$ from $g_{1,\eta_{1,n}}(. \mid \bar{A}(m-1), \bar{L}(m))$.
     - Set $m = m + 1$ and repeat steps (A), (B), (C) until the complete data structure denoted by $Y_{2,b}^{1,*}$ is observed.
  4. *With history* $(dA_2(j) = 0, \mathcal{F}_2(j))$*:*
     - (A) Generate $L_1(m)$ from $f_{1,\theta_{1,n}}(. \mid \bar{L}(m-1), \bar{A}(m-1))$.
     - (B) Generate $L_2(m)$ from $f_{2,\theta_{2,n}}(. \mid L_1(m), \bar{L}(m-1), \bar{A}(m-1))$.
     - (C) Generate $A_1(m)$ from $g_{1,\eta_{1,n}}(. \mid \bar{A}(m-1), \bar{L}(m))$.
     - (D) Generate $A_2(m)$ from $g_{2,\eta_{2,n}}(. \mid A_1(m), \bar{A}(m-1), \bar{L}(m))$.
     - Set $m = m+1$ and repeat steps (A), (B), (C), (D) until the complete data structure denoted by $Y_{0,b}^{2,*}$ is observed.
- **EVALUATE:** Evaluate $IC_{Dabr}(Y \mid F_n, G_n)$ at $Y = Y_{i,b}^{k,*}, k = 1, 2, i = 0, 1$.
- **REPEAT:** Repeat the steps SIMULATE and EVALUATE B times and report

$$IC_{SRA}(Y \mid Q_{1,n}, G_n)(j) = \frac{1}{B}\sum_{b=1}^{B}\sum_{k=1}^{2}\left[IC_{Dabr}(Y_{1,b}^{k,*} \mid F_n, G_n) - IC_{Dabr}(Y_{0,b}^{k,*} \mid F_n, G_n)\right] dM_{G_{k,n}}(j)$$

$IC_{SRA}(Y \mid Q_{1,n}, G_n)(j)$ is now an estimate of the projection of $IC_{Dabr}(Y \mid F, G)$ onto $T_{SRA,1,j} \oplus T_{SRA,2,j}, j = 1, \cdots, p$. Note that for each observation j runs up to the corresponding $\max(\tilde{T}_1, \tilde{T}_2)$. This way of estimating the projection terms guarantees that the resulting estimator of the orthogonalized estimating function (10) is consistent if either $(f_{1,\theta_1}, f_{2,\theta_2})$ or $(g_{1,\eta_1}, g_{2,\eta_2})$ is correctly specified.

5.2 Confidence Intervals

In this subsection, we will briefly discuss the ways of constructing Wald-type confidence intervals for the proposed one step estimator given in (12). In particular, we will consider the case where we assume that the model $\mathcal{G}$ posed for censoring mechanism is correct. Application of Lemma 8.1 in Appendix shows that μ_n^1 is asymptotically linear with influence curve $IC^*(Y \mid Q(F_X, G), G, \mu) - \Pi(IC^*(Y \mid Q(F_X, G), G, \mu) \mid T(\mathcal{G}))$ where $T(\mathcal{G})$ is the tangent space of G for the chosen model. Therefore, one can use

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} IC(Y_i \mid Q_n, G_n, \mu_n^0)$$

as a conservative estimate of the asymptotic variance of μ_n^1, and this can be used to construct a conservative 95% confidence interval for μ:

$$\mu_n^1 \pm 1.96\frac{\hat{\sigma}}{\sqrt{n}}.$$

This confidence interval is asymptotically correct if Q_n is a consistent estimate of Q, i.e. conditional expectations in the projection terms are estimated consistently. Moreover, it is freely obtained after having computed μ_n^1.

6 Simulations

We performed a simulation study to assess the relative performance of μ_n^0, μ_n^{Dab} and μ_n^1. In our simulations, we generated bivariate survival and censoring times from frailty models with and without covariates. Frailty models are a subclass of Copula models. The theory of Copulas dates back to Sklar (1959) but their application in statistical modeling is a more recent phenomenon (e.g. Genest and MacKay, 1986; Genest and Rivest, 1993; Oakes, 1989; Clayton, 1978; Clayton and Cuzick, 1985; Hougaard, 1987). We have two main simulation setups. Below we describe these in details, and the explicit formulas for data generation is provided in Appendix.

- Simulation I (Informative censoring): We generated binary baseline covariates $Z1, Z2 \sim Bernoulli(p)$ for each pair of subject. Consecutively, both censoring and survival times were made dependent on these baseline covariates to enforce informative censoring. Survival times T_1 and T_2 are generated from a gamma frailty model with truncated baseline hazard. This assumes a proportional hazards model of the type

$$\lambda_i(t \mid W = w, Z_i = z_i) = \lambda_0(t) w e^{\beta_t z_i}, \qquad i = 1, 2,$$

where w represents a realization from the hidden gamma random variable. Truncated exponential baseline hazard was chosen to ensure that $\bar{G}(t_1, t_2 \mid X) > \delta > 0\ \forall t_1, t_2$ in the support of F_X. Similarly, C_1, C_2 were generated from a gamma frailty model with covariates Z_1, Z_2 using a constant baseline hazard. We adjusted the amount of dependence between survival and censoring times through the coefficient in front of the Zs (β_t for (T_1, T_2) and β_c for (C_1, C_2)).

- Simulation II (Independent censoring): We generated survival times as in simulation setup I, but enforced censoring times to be independent of T_1, T_2. This simply corresponds to setting $\beta_c = 0$ in the conditional hazard functions of C_1, C_2.

6.1 Comparison of μ_n^0 with μ_n^{Dab}

We firstly report the mean squared error ratios for μ_n^{Dab} and μ_n^0 from a simulation study of setup I for moderate informative censoring in Table 1. The two estimators are evaluated on a 4×4 grid. We observe that μ_n^0 outperforms μ_n^{Dab} at all grid points. This result indicates that our generalization of the Dabrowska's estimator is truly accounting for informative censoring as expected.

[Table 1 about here.]

In Table 2, we report relative performance of the two estimators when censoring times are independent of the failure times i.e. $G(. \mid X) = G(.)$ (generating from setup II). In this simulation, when constructing μ_n^0, $G(. \mid X)$ is still estimated by a bivariate frailty model with covariates ignoring independence structure. We observe from Table 2 that both estimators perform about the same under this scenario. There is no efficiency loss since our posed model for $G(. \mid X)$ includes the independent censoring model as a sub model.

[Table 2 about here.]

6.2 Comparison of μ_n^0, μ_n^{Dab} and μ_n^1

We compare the performances of the three estimators on a simulated data set of sample size 250 generated from the simulation setup I. We estimated the quantity

$$E\left[IC_{Dabr}(Y \mid F_n, G_n) \mid dA_i(t), \mathcal{F}_i(t)\right]$$

by a linear regression model based on covariates extracted from the supplied history $\mathcal{F}_i(t)$. This corresponds to using the regression approach described in Section 5. Covariates such as $I(T_i \leq t)$, t, $I(T_i \leq t) \times T_i$, Z_i, $I(C_i \leq t) \times C_i$, $i = 1, 2$ and some interactions with the time variable t are used and standard model selection techniques are employed. Moreover, the conditional hazards in the projections are estimated by fitting a cox-proportional hazards model. Survival function estimates at different grid points are given in Table 3. Firstly, since there is informative censoring, μ_n^0 outperforms μ_n^{Dab} at all grid points. Secondly, we observe that the one step estimator μ_n^1 provides some improvement over the initial estimator μ_n^0 (i.e. the change in the estimator is in the desired direction), however it is not big of a improvement overall. This is not a surprising result if we look at the estimated adjustment parameter, $c_{nu,n}$, reported in this table. $c_{nu,n}$ is away from 1 at all grid points indicating that we are doing a poor job when estimating the projections (i.e. nuisance parameter $Q_1(F_X, G)$ is misspecified). It would be worthwhile to put effort in making the projection constant close to 1 with real data applications. We also report the conservative 95% intervals for the one-step estimator in column 6 of the Table 3.

[Table 3 about here.]

7 Discussion

We firstly presented a general method of constructing mappings from full data estimating functions to observed data estimating functions which results in estimators asymptotically equivalent to a specified RAL estimator. This is a powerful method and application of it in general bivariate right censored data structure resulted in a generalized estimator of Dabrowska's (1988) estimator. This proposed generalized estimator overcomes the deficiencies of the commonly used Dabrowska's estimator by allowing informative censoring and incorporating covariate processes. Secondly, we constructed an orthogonalized estimating function that has the double robustness property. We mainly considered the scenario where the censoring mechanism is specified correctly and constructed a one-step estimator that improves on our initial estimator. We have shown with a simulation study that generalized estimator is superior to Dabrowska's estimator when censoring mechanism is estimated consistently and the results are dramatic in favor of the generalized estimator when there is dependent censoring. We used the one-step estimator together with Dabrowska's estimator and generalized Dabrowska's estimator on a simulated data set that included informative censoring. In this example dataset, one-step estimator did not improve much on the generalized Dabrowska's estimator since we did a poor job on estimating the projections onto T_{SRA}. We were able to monitor this by the estimated adjustment parameter. One future research direction would be implementing the Monte-Carlo simulation method of Section 5.1 for estimating the projection terms. This would provide the desired flexibility to misspecify $G(. \mid X)$.

8 APPENDIX

Both the influence curve lemma and the asymptotic linearity theorem of Subsections 8.1 and 8.2 require the following Hilbert space terminology: $L_0^2(P_{F_X,G})$ is the Hilbert space of functions of Y with finite variance and mean zero endowed with the covariance inner product $< v_1, v_2 >_{P_{F_X,G}} \equiv \sqrt{\int v_1 v_2 dP_{F_X,G}}$.

8.1 Influence curve of a asymptotically linear estimator when censoring mechanism is estimated efficiently

The following lemma is from van der Laan et al. (2000).

Lemma 8.1 *Let Y be observed data from $P_{F_X,G}$ where G satisfies coarsening at random. Denote the tangent space for the parameter F_X with $T_1(P_{F_X,G})$. Consider the parameter μ which is a real valued functional of F_X. Let $\mu_n(G)$ be an asymptotically linear estimator of μ with influence curve $IC_0(. \mid F_X, G)$ which uses the true G. Assume that for an estimator G_n*

$$\mu_n(G_n) - \mu = \mu_n(G) - \mu + \Phi(G_n) - \Phi(G) + o_P(1/\sqrt{n}) \tag{23}$$

for some functional Φ of G_n. Assume that $\Phi(G_n)$ is an asymptotically efficient estimator of $\Phi(G)$ for a given model $\{G_\eta : \eta \in \Gamma\}$ with tangent space $T_2(P_{F_X,G})$. Then, $\mu_n(G_n)$ is asymptotically linear with influence curve

$$IC_1(. \mid F_X, G) = IC_0(. \mid F_X, G) - \Pi(IC_0(. \mid F_X, G) \mid T_2(P_{F_X,G})).$$

Proof: We decompose $L_0^2(P_{F_X,G})$ orthogonally in $T_1(P_{F_X,G}) \oplus T_2(P_{F_X,G}) \oplus T_\perp(P_{F_X,G})$, where $T_\perp(P_{F_X,G})$ is the orthogonal complement of $T_1(P_{F_X,G}) \oplus T_2(P_{F_X,G})$. By (23), $\mu_n(G_n)$ is asymptotically linear with with influence curve $IC = IC_0 + IC_{nu}$, where IC_{nu} is an influence curve corresponding with an estimator of the nuisance parameter $\Phi(G)$ under the model with nuisance tangent space $T_1(P_{F_X,G})$. Let $IC_0 = a_0 + b_0 + c_0$ and $IC_{nu} = a_{nu} + b_{nu} + c_{nu}$ according to the orthogonal decomposition of $L_0(P_{F_X,G})$. We will now use two general facts about the influence curves. Firstly, an influence curve is orthogonal to the nuisance tangent space, and secondly efficient influence curve lies in the tangent space. Since IC_{nu} is an influence curve of $\Phi(G)$ in the model where F_X is not specified, it is orthogonal to $T_1(P_{F_X,G})$, i.e. $a_{nu} = 0$. Moreover, since $\Phi(G_n)$ is efficient, IC_{nu} lies in the tangent space $T_2(P_{F_X,G})$ and hence $c_{nu} = 0$. We also have that $IC_0 + IC_{nu}$ is influence curve of $\mu_n(G_n)$ thus it is orthogonal to $T_2(P_{F_X,G})$, i.e. $b_0 + b_{nu} = 0$. Consequently, we have that

$$IC_1 + IC_{nu} = a_0 + c_0 = \Pi(IC_0 \mid T_2^\perp(P_{F_X,G})) \equiv IC_0 - \Pi(IC_0 \mid T_2(P_{F_X,G}))$$

This completes the proof. □

8.2 Asymptotics assuming consistent estimation of the censoring mechanism.

The following theorem (van der Laan, 2002) provides a template for proving asymptotic linearity with specified influence curve of the one-step estimator μ_n^1 given by (7, 12) (i.e., set $c_{nu,n} = c_{nu} = 1$)

or of the one-step solution of the estimating function (20) (if one uses the adjustment constant $c_{nu,n}$). The tangent space $T_2 = T_2(P_{F_X,G})$ for the parameter G is the closure of the linear extension in $L_0^2(P_{F_X,G})$ of the scores at $P_{F_X,G}$ from all correctly specified parametric sub-models (i.e., sub-models of the assumed semiparametric model $\mathcal{G}$) for the distribution G.

Theorem 8.1 *Consider the observed data model $\mathcal{M}(\mathcal{G}) = \{\mathcal{P}_{\mathcal{F}_{\mathcal{X}},\mathcal{G}} : \mathcal{F}_{\mathcal{X}} \in \mathcal{M}^{\mathcal{F}}, \mathcal{G} \in \mathcal{G} \subset \mathcal{G}(\mathcal{SRA})\}$. Let $Y_1, \ldots, Y_n$ be n i.i.d. observations of $Y \sim P_{F_X,G} \in \mathcal{M}(\mathcal{G})$. Consider a one-step estimator of the parameter $\mu \in \mathbb{R}^1$ of the form $\mu_n^1 = \mu_n^0 + c_n^{-1} P_n IC(\cdot \mid Q_n, G_n, c_{nu,n}, D_{h_n}(\mu_n^0, \rho_n))$. We will refer to $c_n^{-1} IC(\cdot \mid Q_n, G_n, c_{nu,n}, D_{h_n}(\mu_n^0, \rho_n))$ also by $IC(\cdot \mid Q_n, G_n, c_{nu,n}, c_n, D_{h_n}(\mu_n^0, \rho_n))$. Assume that the limit of $IC(. \mid Q_n, G_n, c_{nu,n}, D_{h_n}(\mu_n^0, \rho_n))$ specified in (ii) below satisfies:*

$$
\begin{aligned}
E_G(IC(Y \mid Q^1, G, c_{nu}, D_h(\cdot \mid \mu, \rho)) \mid X) &= D_h(X \mid \mu, \rho) \; F_X - a.e. && (24)\\
D_h(\cdot \mid \mu, \rho) &\in T_{nuis}^{F,\perp}(F_X). && (25)
\end{aligned}
$$

Assume (we write $f \approx g$ for $f = g + o_P(1/\sqrt{n})$)

$$
c_n^{-1} P_n \left\{ IC(\cdot \mid Q_n, G_n, c_{nu,n}, D_{h_n}(\mu_n^0, \rho_n)) - IC(\cdot \mid Q_n, G_n, c_{nu,n}, D_{h_n}(\mu, \rho_n)) \right\} \approx \mu - \mu_n^0. \quad (26)
$$

and

$$
E_{P_{F_X,G}} IC(Y \mid Q_n, G, c_{nu,n}, D_{h_n}(\mu, \rho_n)) = o_P(1/\sqrt{n}). \quad (27)
$$

where the G-component of ρ_n is set equal to G as well.

In addition, assume

(i) $IC(\cdot \mid Q_n, G_n, c_{nu,n}, c_n, D_{h_n}(\cdot \mid \mu_n^0, \rho_n))$ falls in a $P_{F_X,G}$-Donsker class with probability tending to 1.

(ii) For some (h, Q^1) we have:

$$
\| IC(\cdot \mid Q_n, G_n, c_{nu,n}, c_n, D_{h_n}(\cdot \mid \mu_n^0, \rho_n)) - IC(\cdot \mid Q^1, G, c_{nu}, c, D_h(\cdot \mid \mu, \rho)) \|_{P_{F_X,G}} \rightarrow 0,
$$

where the convergence is in probability. Here (suppressing the dependence of the estimating functions on parameters) $c_{nu} = \langle IC_0, IC_{nu}^\top \rangle \langle IC_{nu}, IC_{nu}^\top \rangle^{-1}$ is such that $c_{nu} IC_{nu}$ equals the projection of IC_0 onto the k-dimensional space $< IC_{nu,j}, j = 1, \ldots, k >$ in $L_0^2(P_{F_X,G})$.

(iii) Define for a G_1

$$
\Phi(G_1) = P_{F_X,G} IC(\cdot \mid Q^1, G_1, c_{nu}, c, D_h(\mu, \rho)).
$$

For notational convenience, let

$$
\begin{aligned}
IC_n(G) &\equiv IC(\cdot \mid Q_n, G, c_{nu,n}, c_n, D_{h_n}(\mu, \rho_n))\\
IC(G) &\equiv IC(\cdot \mid Q^1, G, c_{nu}, c, D_h(\mu, \rho)).
\end{aligned}
$$

Assume

$$
P_{F_X,G}\{IC_n(G_n) - IC_n(G)\} \approx \Phi(G_n) - \Phi(G).
$$

(iv) $\Phi(G_n)$ is an asymptotically efficient estimator of $\Phi(G)$ for the SRA-model $\mathcal{G}$ containing the true G with tangent space $T_2(P_{F_X,G}) \subset T_{SRA}(P_{F_X,G})$.

Then μ_n^1 is asymptotically linear with influence curve given by

$$
IC \equiv \Pi(IC(\cdot \mid Q^1, G, c_{nu}, c, D_h(\cdot \mid \mu, \rho)) \mid T_2^\perp(P_{F_X,G})).
$$

If $Q^1 = Q(F_X, G)$ and $IC(Y \mid Q(F_X, G), G, c_{nu}, D_h(\cdot \mid \mu, \rho)) \perp T_2(P_{F_X,G})$, then this influence curve equals $IC(\cdot \mid Q(F_X, G), G, c_{nu} = 1, c, D_h(\mu, \rho))$.

8.3 Proof of Theorem 8.1.

For notational convenience, we will give the proof for $c_{nu,n} = 1$ and use appropriate short hand notation. We have

$$\begin{aligned}\mu_n^1 &= \mu_n^0 + c_n^{-1} P_n \left\{ IC(Q_n, G_n, D_{h_n}(\mu_n^0, \rho_n)) - IC(Q_n, G_n, D_{h_n}(\mu, \rho_n)) \right\} \\ &\quad + c_n^{-1} P_n IC(Q_n, G_n, D_{h_n}(\mu, \rho_n)).\end{aligned}$$

By condition (26) the difference on the right hand side equals $\mu - \mu_n^0 + o_P(1/\sqrt{n})$. Thus we have:

$$\begin{aligned}\mu_n^1 - \mu &= (P_n - P) c_n^{-1} IC(Q_n, G_n, D_{h_n}(\mu, \rho_n)) \\ &\quad + c_n^{-1} P IC(Q_n, G_n, D_{h_n}(\mu, \rho_n)).\end{aligned}$$

For empirical process theory we refer to van der Vaart and Wellner (1996). Condition (i) and (ii) in the theorem imply that the empirical process term on the right hand side is asymptotically equivalent with $(P_n - P_{F_X,G}) c^{-} IC(\cdot \mid Q^1, G^1, D_h(\mu, \rho))$. So it remains to analyze the term

$$c_n^{-1} P IC(Q_n, G_n, D_{h_n}(\mu, \rho_n)).$$

Now, we write this term as a sum of two terms $A + B$, where

$$\begin{aligned}A &= c_n^{-1} P \left\{ IC(Q_n, G_n, D_{h_n}(\mu, \rho_n)) - IC(Q^1, G, D_h(\mu, \rho) \right\} \\ B &= c_n^{-1} P IC(Q^1, G, D_h(\mu, \rho)),\end{aligned}$$

By (24) and (25) we have $B = 0$. As in the theorem, let

$$\begin{aligned}IC_n(G) &\equiv IC(\cdot \mid Q_n, G, D_{h_n}(\mu, \rho_n(G))) \\ IC(G) &\equiv IC(\cdot \mid Q^1, G, D_h(\mu, \rho)).\end{aligned}$$

We decompose $A = A_1 + A_2$ as follows:

$$A = P_{F_X,G}\{IC_n(G_n) - IC(G)\} = P_{F_X,G}\{IC_n(G) - IC(G)\} + P_{F_X,G}\{IC_n(G_n) - IC_n(G)\}.$$

By assumption (27) we have that $A_1 = o_P(1/\sqrt{n})$. By assumption (iii)

$$A_2 = \Phi_2(G_n) - \Phi_2(G) + o_P(1/\sqrt{n}).$$

By assumption (iv), we can conclude that μ_n^1 is asymptotically linear with influence curve $IC(\cdot \mid Q^1, G, c, c_{nu}, D_h(\mu, \rho)) + IC_{nuis}$, where IC_{nuis} is the influence curve of $\Phi_2(G_n)$. Now, the same argument as given in the proof of Lemma 8.1 proves that this influence curve of μ_n^1 is given by:

$$\Pi(IC(\cdot \mid Q^1, G, c, c_{nu}, D_h(\mu, \rho)) \mid T_2^{\perp}).$$

This completes the proof. □

8.4 Data Generation For The Simulation Study

Let W be a gamma random variable with mean 1 and variance α_t. Let Z_1 and Z_2 be Bernoulli random variables with probability p. We assume the following proportional hazards model for T_1 and T_2:

$$\lambda_i(t \mid W = w, Z_i = z_i) = \lambda_{0,T}(t) w e^{\beta_t z_i}, \qquad i = 1, 2,$$

where w represents a realization from the hidden gamma random variable W. The baseline hazard $\lambda_{0,T}(t)$ is set to the hazard of a truncated exponential distribution and is given by

$$\lambda_0(t) = \frac{\lambda_t e^{-\lambda_t t}}{e^{-\lambda_t t} - e^{-\lambda_t \tau}},$$

where λ_t is the rate and τ is the truncation constant of the distribution. The bivariate distribution of T_1 and T_2 conditional on $Z \equiv (Z_1, Z_2)$ is given by

$$S(t_1, t_2 \mid Z) = \left(S_1(t_1 \mid Z)^{-\alpha_t} + S_2(t_2 \mid Z)^{-\alpha_t} - 1\right)^{-\frac{1}{\alpha_t}}, \tag{28}$$

where

$$S_i(t \mid Z) = \left(1 + \alpha_t e^{\beta_t Z_i} \Lambda_0(t)\right)^{-\frac{1}{\alpha_t}}, \qquad i = 1, 2.$$

We use a similar frailty model with constant baseline hazard, $\lambda_{0,C}(t) = \lambda_c$, for the censoring mechanism and denote the variance of the corresponding hidden gamma variable by α_c. We now provide the explicit formulas for generating data from the above defined structures. Let U_1, U_2 be random draws from uniform distribution on the interval $[0, 1]$. Let Z_1 and Z_2 denote random draws from *Bernoulli(p)*. Then, T_1 given (Z_1, Z_2) and T_2 given (T_1, Z_1, Z_2) can be generated as

$$\begin{aligned}
\phi_1 &= \frac{(1 - U_1)^{-\alpha_t} - 1}{\alpha_t e^{\beta_t Z_1}}, \\
T_1 &= -\frac{1}{\lambda_t} \log\left[e^{\left(\log(1 - e^{-\lambda_t \tau}) - \phi_1\right)} + e^{-\lambda_t \tau}\right], \\
\phi_2 &= \left[U_1^{-\frac{\alpha_t}{1+\alpha_t}} S_1(T_1 \mid Z_1)^{-\alpha_t} - S_1(T_1 \mid Z_1)^{-\alpha_t} + 1\right]^{-\frac{1}{\alpha_t}}, \\
\phi_3 &= \frac{\phi_2^{-\alpha_t} - 1}{\alpha_t e^{\beta_t Z_2}}, \\
T_2 &= -\frac{1}{\lambda_t} \log\left[e^{\left(\log(1 - e^{-\lambda_t \tau}) - \phi_3\right)} + e^{-\lambda_t \tau}\right].
\end{aligned}$$

Similarly, we generate the censoring times C_1 given Z_1, Z_2 and C_2 given (C_1, Z_1, Z_2) as follows

$$\begin{aligned}
C_1 &= \frac{1}{\lambda_c}\left[\frac{(1 - U_2)^{-\alpha_c} - 1}{\alpha_c e^{\beta_c Z_1}}\right], \\
\phi_4 &= \left[U_2^{-\frac{\alpha_c}{1+\alpha_c}} S_1(C_1 \mid Z_1)^{-\alpha_c} - S_1(C_1 \mid Z_1)^{-\alpha_c} + 1\right]^{-\frac{1}{\alpha_c}}, \\
C_2 &= \frac{1}{\lambda_c} \frac{\phi_4^{-\alpha_c} - 1}{\alpha_c e^{\beta_c Z_2}}.
\end{aligned}$$

8.5 Computational Remarks

We will now go over a few computational details that are required for estimation of $G(. \mid Z)$. R function coxph is used to estimate $G(. \mid Z)$ by a frailty model. This is a straight forward procedure and is explained quite well in the help menu of R. Below we provide a piece of code for extracting cumulative baseline hazard function from the R objects generated by coxph. datafr is a data frame of the data.

```
fra1_coxph(Surv(time,cstatus)~frailty(id)+strata(strat)+Z,data=datafr)
fra1.sf_survfit(fra1) #survfit gives the survival estimates at
                        #uncensored points with mean covariates.

fra1.sum_ summary(fra1.sf)

S1_fra1.sum\$surv[fra1.sum\$strata=="strat=0"] #P(T_1 >= tt1 | Z_1)
S2_fra1.sum\$surv[fra1.sum\$strata=="strat=1"] #P(T_2 >= tt2 | Z_2)
tt1_fra1.sum\$time[fra1.sum\$strata=="strat=0"]#tt1
tt2_fra1.sum\$time[fra1.sum\$strata=="strat=1"]#tt2
alph_fra1\$history\$"frailty(id)"\$theta #extracts the variance of the gamma frailty.

#Extracting the cumulative baseline hazard at tt1 and tt2 including time 0:

ch1_c(0,(S1^(-alph)-1)/(alph*exp(fra1\$coef*mean(Z1))))
ch2_c(0,(S2^(-alph)-1)/(alph*exp(fra1\$coef*mean(Z2))))
```

Once we have the estimates of the baseline cumulative hazard for various time points, we can estimate $\bar{G}(. \mid Z)$ using eq. 28.

8.6 Proving $E\left(IC_D(Y \mid F,G) \mid X\right) = I(T_1 \geq t_1, T_2 \geq t_2) - S(t_1,t_2)$

We are going to first show that $E(IC_{Dabr} \mid X) = I(T_1 \geq t_1, T_2 \geq t_2) - \bar{F}(t_1,t_2)$ where IC_{Dabr} is the influence curve of Dabrowska's (1988) estimator (without any modification) and $X \equiv (T_1, T_2)$. Then, it is easily seen that conditional expectation of modified Dabrowska's influence curve given X also reduces to $I(T_1 \geq t_1, T_2 \geq t_2) - \bar{F}(t_1,t_2)$. Note that we are using $\bar{F}(t_1,t_2) \equiv S(t_1,t_2)$. Basically, $\bar{G}(. \mid X)$ terms in denominator and numerator cancel out. Influence curve of Dabrowska's bivariate survival function estimator in the random censoring model is given by

$$IC(t_1,t_2) =$$
$$\bar{F}(t_1,t_2)\left\{ -\int_0^{t_1} \frac{I(\tilde{T}_1 \in du, \Delta_1 = 1) - I(\tilde{T}_1 \geq u)P(T_1 \in du \mid T_1 \geq u)}{P(\tilde{T}_1 \geq u)} \right. \tag{29}$$
$$- \int_0^{t_2} \frac{I(\tilde{T}_2 \in du, \Delta_2 = 1) - I(\tilde{T}_2 \geq u)P(T_2 \in du \mid T_2 \geq u)}{P(\tilde{T}_2 \geq u)} \tag{30}$$
$$+ \int_0^{t_1}\int_0^{t_2} \frac{I(\tilde{T}_1 \in du, \tilde{T}_2 \in dv, \Delta_1 = 1, \Delta_2 = 1)}{P(\tilde{T}_1 \geq u, \tilde{T}_2 \geq v)} \tag{31}$$

$$
\begin{aligned}
- & \int_0^{t_1}\int_0^{t_2} \frac{I(\tilde T_1 \ge u, \tilde T_2 \ge v) P(T_1 \in du, T_2 \in dv \mid T_1 \ge u, T_2 \ge v)}{P(\tilde T_1 \ge u, \tilde T_2 \ge v)} && (32)\\
- & \int_0^{t_1}\int_0^{t_2} \frac{I(\tilde T_1 \in du, \tilde T_2 \ge v, \Delta_1 = 1) P(T_2 \in dv \mid T_1 \ge u, T_2 \ge v)}{P(\tilde T_1 \ge u, \tilde T_2 \ge v)} && (33)\\
+ & \int_0^{t_1}\int_0^{t_2} \frac{I(\tilde T_1 \ge u, \tilde T_2 \ge v) P(T_1 \in du \mid T_1 \ge u, T_2 \ge v) P(T_2 \in dv \mid T_1 \ge u, T_2 \ge v)}{P(\tilde T_1 \ge u, \tilde T_2 \ge v)} && (34)\\
- & \int_0^{t_1}\int_0^{t_2} \frac{I(\tilde T_1 \ge u, \tilde T_2 \in dv, \Delta_2 = 1) P(T_1 \in du \mid T_1 \ge u, T_2 \ge v)}{P(\tilde T_1 \ge u, \tilde T_2 \ge v)} && (35)\\
+ & \int_0^{t_1}\int_0^{t_2} \frac{I(\tilde T_1 \ge u, \tilde T_2 \ge v) P(T_1 \in du \mid T_1 \ge u, T_2 \ge v) P(T_2 \in dv \mid T_1 \ge u, T_2 \ge v)}{P(\tilde T_1 \ge u, \tilde T_2 \ge v)} \Bigg\}. && (36)
\end{aligned}
$$

Firstly, we will show that $E(IC(t_1,t_2) \mid X) = I(T_1 \ge t_1, T_2 \ge t_2) - \bar F(t_1,t_2)$ where $X \equiv (T_1,T_2)$. We will take the conditional expectations of the terms (29), (30), (31), (32),(33), (34), (35), (36) separately.

Term (29):

$$
\begin{aligned}
& E\left[- \int_0^{t_1} \frac{I(\tilde T_1 \in du, \Delta_1 = 1) - I(\tilde T_1 \ge u) P(T_1 \in du \mid T_1 \ge u)}{P(\tilde T_1 \ge u)} \middle| X \right] \\
&= -\int_0^{t_1} \frac{E\left[I(\tilde T_1 \in du, \Delta_1 = 1) \mid X\right]}{P(\tilde T_1 \ge u)} + \int_0^{t_1} \frac{E\left[I(\tilde T_1 \ge u) \mid X\right] P(T_1 \in du \mid T_1 \ge u)}{P(\tilde T_1 \ge u)} \\
&= -\int_0^{t_1} \frac{I(T_1 \in du) P(C_1 \ge u \mid X)}{P(T_1 \ge u) P(C_1 \ge u \mid T_1 \ge u)} + \int_0^{t_1} \frac{I(T_1 \ge u) P(C_1 \ge u \mid T_1 \ge u) P(T_1 \in du \mid T_1 \ge u)}{P(T_1 \ge u) P(C_1 \ge u \mid T_1 \ge u)} \\
&= -\frac{I(T_1 \le t_1)}{\bar F(T_1,0)} + \frac{1}{\bar F(u,0)}\bigg|_{u=0}^{u=T_1\wedge t_1} \\
&= -\frac{I(T_1 \le t_1)}{\bar F(T_1,0)} + \frac{1}{\bar F(T_1 \wedge t_1,0)} - 1 \\
&= -\frac{I(T_1 \le t_1)}{\bar F(T_1,0)} + \frac{I(T_1 \le t_1)}{\bar F(T_1,0)} + \frac{I(T_1 \ge t_1)}{\bar F(t_1,0)} - 1 = \frac{I(T_1 \ge t_1)}{\bar F(t_1,0)} - 1.
\end{aligned}
$$

Term (30):

$$
\begin{aligned}
& E\left[- \int_0^{t_2} \frac{I(\tilde T_2 \in dv, \Delta_2 = 1) - I(\tilde T_2 \ge v) P(T_2 \in dv \mid T_2 \ge v)}{P(\tilde T_2 \ge v)} \middle| X \right] \\
&= -\int_0^{t_2} \frac{E\left[I(\tilde T_2 \in dv, \Delta_2 = 1) \mid X\right]}{P(\tilde T_2 \ge v)} + \int_0^{t_2} \frac{E\left[I(\tilde T_2 \ge v) \mid X\right] P(T_2 \in dv \mid T_2 \ge v)}{P(\tilde T_2 \ge v)} \\
&= -\int_0^{t_2} \frac{I(T_2 \in dv) P(C_2 \ge v \mid X)}{P(T_2 \ge v) P(C_2 \ge v \mid T_2 \ge v)} + \int_0^{t_2} \frac{I(T_2 \ge v) P(C_2 \ge v \mid T_2 \ge v) P(T_2 \in dv \mid T_2 \ge v)}{P(T_2 \ge v) P(C_2 \ge v \mid T_2 \ge v)} \\
&= -\frac{I(T_2 \le t_2)}{\bar F(0,T_2)} + \frac{1}{\bar F(0,v)}\bigg|_{v=0}^{v=T_2\wedge t_2} \\
&= -\frac{I(T_2 \le t_2)}{\bar F(0,T_2)} + \frac{1}{\bar F(0,T_2 \wedge t_2)} - 1
\end{aligned}
$$

$$= -\frac{I(T_2 \le t_2)}{\bar{F}(0,T_2)} + \frac{I(T_2 \le t_2)}{\bar{F}(0,T_2)} + \frac{I(T_2 \ge t_2)}{\bar{F}(0,t_2)} - 1 = \frac{I(T_2 \ge t_2)}{\bar{F}(0,t_2)} - 1.$$

Term (31):

$$\begin{aligned}
& E\left[\int_0^{t_1}\int_0^{t_2} \frac{I(\tilde{T}_1 \in du, \tilde{T}_2 \in dv, \Delta_1 = 1, \Delta_2 = 1)}{P(\tilde{T}_1 \ge u, \tilde{T}_2 \ge v)}\right| X\Bigg] \\
&= \int_0^{t_1}\int_0^{t_2} \frac{E\left[I(\tilde{T}_1 \in du, \tilde{T}_2 \in dv, \Delta_1 = 1, \Delta_2 = 1) \mid X\right]}{P(\tilde{T}_1 \ge u, \tilde{T}_2 \ge v)} \\
&= \int_0^{t_1}\int_0^{t_2} \frac{I(T_1 \in du, T_2 \in dv) P(C_1 \ge u, C_2 \ge v \mid X)}{P(T_1 \ge u, T_2 \ge v) P(C_1 \ge u, C_2 \ge v \mid T_1 \ge u, T_2 \ge v)} \\
&= \frac{I(T_1 \le t_1, T_2 \le t_2)}{\bar{F}(T_1,T_2)}.
\end{aligned}$$

Term (32):

$$\begin{aligned}
& E\left[-\int_0^{t_1}\int_0^{t_2} \frac{I(\tilde{T}_1 \ge u, \tilde{T}_2 \ge v) P(T_1 \in du, T_2 \in dv \mid T_1 \ge u, T_2 \ge v)}{P(\tilde{T}_1 \ge u, \tilde{T}_2 \ge v)}\right| X\Bigg] \\
&= -\int_0^{t_1}\int_0^{t_2} \frac{E\left[I(\tilde{T}_1 \ge u, \tilde{T}_2 \ge v) \mid X\right] P(T_1 \in du, T_2 \in dv \mid T_1 \ge u, T_2 \ge v)}{P(\tilde{T}_1 \ge u, \tilde{T}_2 \ge v)} \\
&= -\int_0^{t_1}\int_0^{t_2} \frac{I(T_1 \ge u, T_2 \ge v) P(C_1 \ge u, C_2 \ge v \mid X) P(T_1 \in du, T_2 \in dv \mid T_1 \ge u, T_2 \ge v)}{P(T_1 \ge u, t_2 \ge v) P(C_1 \ge u, C_2 \ge v \mid T_1 \ge u, T_2 \ge v)} \\
&= -\int_0^{t_1 \wedge T_1}\int_0^{t_2 \wedge T_2} \frac{\bar{F}(du,dv)}{\bar{F}(u,v)^2}.
\end{aligned}$$

Term (33):

$$\begin{aligned}
& E\left[-\int_0^{t_1}\int_0^{t_2} \frac{I(\tilde{T}_1 \in du, \tilde{T}_2 \ge v, \Delta_1 = 1) P(T_2 \in dv \mid T_1 \ge u, T_2 \ge v)}{P(\tilde{T}_1 \ge u, \tilde{T}_2 \ge v)}\right| X\Bigg] \\
&= -\int_0^{t_1}\int_0^{t_2} \frac{E\left[I(\tilde{T}_1 \in du, \tilde{T}_2 \ge v, \Delta_1 = 1) \mid X\right] P(T_2 \in dv \mid T_1 \ge u, T_2 \ge v)}{P(\tilde{T}_1 \ge u, \tilde{T}_2 \ge v)} \\
&= -\int_0^{t_1}\int_0^{t_2} \frac{I(T_1 \in du, T_2 \ge v) P(C_1 \ge u, C_2 \ge v \mid X) P(T_2 \in dv \mid T_1 \ge u, T_2 \ge v)}{P(T_1 \ge u, T_2 \ge v) P(C_1 \ge u, c_2 \ge v \mid T_1 \ge u, T_2 \ge v)} \\
&= -\int_0^{t_2} \frac{I(T_1 \le t_1, T_2 \ge v) P(T_2 \in dv \mid T_1 \ge T_1, T_2 \ge v)}{P(T_1 \ge T_1, t_2 \ge v)} \\
&= -I(T_1 \le t_1) \int_0^{t_2 \wedge T_2} \frac{\bar{F}(T_1, dv)}{\bar{F}(T_1,v)^2} \\
&= -I(T_1 \le t_1) \left.\frac{1}{\bar{F}(T_1,v)}\right|_{v=0}^{v=t2 \wedge T_2} \\
&= -\frac{I(T_1 \le t_1)}{\bar{F}(T_1, t_2 \wedge T_2)} + \frac{I(T_1 \le t_1)}{\bar{F}(T_1,0)} = -\frac{I(T_1 \le t_1, T_2 \le t_2)}{\bar{F}(T_1,T_2)} - \frac{I(T_1 \le t_1, T_2 \ge t_2)}{\bar{F}(T_1,t_2)} + \frac{I(T_1 \le t_1)}{\bar{F}(T_1,0)}.
\end{aligned}$$

Term (34):

$$E\left[\int_0^{t_1}\int_0^{t_2}\frac{I(\tilde{T}_1\ge u,\tilde{T}_2\ge v)P(T_1\in du\mid T_1\ge u,t_2\ge v)P(T_2\in dv\mid T_1\ge u,T_2\ge v)}{P(\tilde{T}_1\ge u,\tilde{T}_2\ge v)}\middle| X\right]$$

$$\begin{aligned}
&= \int_0^{t_1}\int_0^{t_2}\frac{E\left[I(\tilde{T}_1\ge u,\tilde{T}_2\ge v)\mid X\right]P(T_1\in du\mid T_1\ge u,T_2\ge v)P(T_2\in dv\mid T_1\ge u,T_2\ge v)}{P(\tilde{T}_1\ge u,\tilde{T}_2\ge v)}\\
&= \int_0^{t_1}\int_0^{t_2}\frac{I(T_1\ge u,T_2\ge v)P(C_1\ge u,C_2\ge v\mid X)P(T_1\in du\mid T_1\ge u,T_2\ge v)P(T_2\in dv\mid T_1\ge u,T_2\ge v)}{P(T_1\ge u,T_2\ge v)P(C_1\ge u,C_2\ge v\mid T_1\ge u,T_2\ge v)}\\
&= \int_0^{t_1\wedge T_1}\int_0^{t_2\wedge T_2}\frac{\bar{F}(du,v)\bar{F}(u,dv)}{\bar{F}(u,v)^3}.
\end{aligned}$$

Term (35):

$$E\left[-\int_0^{t_1}\int_0^{t_2}\frac{I(\tilde{T}_1\ge u,\tilde{T}_2\in dv,\Delta_2=1)P(T_1\in du\mid T_1\ge u,T_2\ge v)}{P(\tilde{T}_1\ge u,\tilde{T}_2\ge v)}\middle| X\right]$$

$$\begin{aligned}
&= -\int_0^{t_1}\int_0^{t_2}\frac{E\left[I(\tilde{T}_1\ge u,\tilde{T}_2\in dv,\Delta_2=1)\mid X\right]P(T_1\in du\mid T_1\ge u,T_2\ge v)}{P(\tilde{T}_1\ge u,\tilde{T}_2\ge v)}\\
&= -\int_0^{t_1}\int_0^{t_2}\frac{I(T_1\ge u,T_2\in dv)P(C_1\ge u,C_2\ge v\mid X)P(T_1\in du\mid T_1\ge u,T_2\ge v)}{P(T_1\ge u,T_2\ge v)P(C_1\ge u,C_2\ge v\mid T_1\ge u,T_2\ge v)}\\
&= -\int_0^{t_1}\frac{I(T_1\ge u,T_2\le t_2)P(T_1\in du\mid T_1\ge u,T_2\ge T_2)}{P(T_1\ge u,T_2\ge T_2)}\\
&= -I(T_2\le t_2)\int_0^{t_1\wedge T_1}\frac{\bar{F}(du,T_2)}{\bar{F}(u,T_2)^2}\\
&= -I(T_2\le t_2)\left.\frac{1}{\bar{F}(u,T_2)}\right|_{u=0}^{u=t1\wedge T_1}\\
&= -\frac{I(T_2\le t_2)}{\bar{F}(T_1\wedge t_1,T_2)}+\frac{I(T_2\le t_2)}{\bar{F}(0,T_2)}=-\frac{I(T_1\le t_1,T_2\le t_2)}{\bar{F}(T_1,T_2)}-\frac{I(T_1\ge t_1,T_2\le t_2)}{\bar{F}(t_1,T_2)}+\frac{I(T_2\le t_2)}{\bar{F}(0,T_2)}.
\end{aligned}$$

Term (36) (same as the term (34)):

$$E\left[\int_0^{t_1}\int_0^{t_2}\frac{I(\tilde{T}_1\ge u,\tilde{T}_2\ge v)P(T_1\in du\mid T_1\ge u,T_2\ge v)P(T_2\in dv\mid T_1\ge u,T_2\ge v)}{P(\tilde{T}_1\ge u,\tilde{T}_2\ge v)}\middle| X\right]$$

$$\begin{aligned}
&= \int_0^{t_1}\int_0^{t_2}\frac{E\left[I(\tilde{T}_1\ge u,\tilde{T}_2\ge v)\mid X\right]P(T_1\in du\mid T_1\ge u,T_2\ge v)P(T_2\in dv\mid T_1\ge u,T_2\ge v)}{P(\tilde{T}_1\ge u,\tilde{T}_2\ge v)}\\
&= \int_0^{t_1}\int_0^{t_2}\frac{I(T_1\ge u,T_2\ge v)P(C_1\ge u,C_2\ge v\mid X)P(T_1\in du\mid T_1\ge u,T_2\ge v)P(T_2\in dv\mid T_1\ge u,T_2\ge v)}{P(T_1\ge u,T_2\ge v)P(C_1\ge u,C_2\ge v\mid T_1\ge u,T_2\ge v)}\\
&= \int_0^{t_1\wedge T_1}\int_0^{t_2\wedge T_2}\frac{\bar{F}(du,v)\bar{F}(u,dv)}{\bar{F}(u,v)^3}.
\end{aligned}$$

Note that

$$\frac{1}{du}\left(\frac{1}{dv}\left(\frac{1}{\bar{F}(u,v)}\right)\right)=\frac{-\bar{F}(du,dv)}{\bar{F}(u,v)^2}+2\frac{\bar{F}(du,v)\bar{F}(u,dv)}{\bar{F}(u,v)^3}.$$

Then, the sum of the terms (32), (34), (36) equals

$$\int_0^{t_1\wedge T_1}\int_0^{t_2\wedge T_2}\left\{-\frac{\bar{F}(du,dv)}{\bar{F}(u,v)^2}+2\frac{\bar{F}(du,v)\bar{F}(u,dv)}{\bar{F}(u,v)^3}\right\}$$
$$= \int_0^{t_1\wedge T_1}\int_0^{t_2\wedge T_2}\frac{1}{du}\left(\frac{1}{dv}\left(\frac{1}{\bar{F}(u,v)}\right)\right)$$
$$= \int_0^{t_1\wedge T_1}\frac{1}{du}\left(\frac{1}{\bar{F}(u,t_2\wedge T_2)}-\frac{1}{\bar{F}(u,0)}\right\}$$
$$= \frac{1}{\bar{F}(t1\wedge T_1,t_2\wedge T_2)}-\frac{1}{\bar{F}(t1\wedge T_1,0)}-\frac{1}{\bar{F}(0,t_2\wedge T_2)}+\frac{1}{\bar{F}(0,0)}$$
$$= \frac{I(T_1\le t_1,T_2\le t_2)}{\bar{F}(T_1,T_2)}+\frac{I(T_1\ge t_1,T_2\le t_2)}{\bar{F}(t_1,T_2)}+\frac{I(T_1\le t_1,T_2\ge t_2)}{\bar{F}(T_1,t_2)}+\frac{I(T_1\ge t_1,T_2\ge t_2)}{\bar{F}(t_1,t_2)}$$
$$-\frac{I(T1\le t_1)}{\bar{F}(T_1,0)}-\frac{I(T1\ge t_1)}{\bar{F}(t_1,0)}-\frac{I(T2\le t_2)}{\bar{F}(0,T_2)}-\frac{I(T2\ge t_2)}{\bar{F}(0,t_2)}+1.$$

Bringing all the terms together we obtain

$$\begin{aligned}
E(IC(t_1,t_2)\mid X) &= \bar{F}(t_1,t_2)\left\{\frac{I(T_1\ge t_1)}{\bar{F}(t_1,0)}-1+\frac{I(T_2\ge t_2)}{\bar{F}(0,t_2)}-1+\frac{I(T_1\le t_1,T_2\le t_2)}{\bar{F}(T_1,T_2)}\right.\\
&- \frac{I(T_1\le t_1,T_2\le t_2)}{\bar{F}(T_1,T_2)}-\frac{I(T_1\le t_1,T_2\ge t_2)}{\bar{F}(T_1,t_2)}+\frac{I(T_1\le t_1)}{\bar{F}(T_1,0)}\\
&- \frac{I(T_1\le t_1,T_2\le t_2)}{\bar{F}(T_1,T_2)}-\frac{I(T_1\ge t_1,T_2\le t_2)}{\bar{F}(t_1,T_2)}+\frac{I(T_2\le t_2)}{\bar{F}(0,T_2)}\\
&+ \frac{I(T_1\le t_1,T_2\le t_2)}{\bar{F}(T_1,T_2)}+\frac{I(T_1\ge t_1,T_2\le t_2)}{\bar{F}(t_1,T_2)}+\frac{I(T_1\le t_1,T_2\ge t_2)}{\bar{F}(T_1,t_2)}+\frac{I(T_1\ge t_1,T_2\ge t_2)}{\bar{F}(t_1,t_2)}\\
&- \left.\frac{I(T1\le t_1)}{\bar{F}(T_1,0)}-\frac{I(T1\ge t_1)}{\bar{F}(t_1,0)}-\frac{I(T2\le t_2)}{\bar{F}(0,T_2)}-\frac{I(T2\ge t_2)}{\bar{F}(0,t_2)}+1\right\}\\
&= I(T_1\ge t_1,T_2\ge t_2)-\bar{F}(t_1,t_2).
\end{aligned}$$

This completes the proof. □

References

Andersen, P., Borgan, Ø., Gill, R. and Keiding, N. (1993). *Statistical models based on counting processes*, Springer-Verlag, New York.

Bickel, P., Klaassen, C. J., Ritov, Y. and Wellner, J. (1993). *Efficient and adaptive estimation for semiparametric models*, Johns Hopkins University Press, Baltimore, MD.

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

Clayton, D. G. and Cuzick, J. (1985). Multivariate generalizations of the proportional hazards model (with discussion), *Journal of the Royal Statistical Society, Series B* **148**: 82–117.

Dabrowska, D. (1988). Kaplan-meier estimate on the plane, *Annals of Statistics* **16**: 1475–1489.

Dabrowska, D. (1989). Kaplan-meier estimate on the plane: Weak convergence, LIL, and the bootstrap, *Journal of Multivariate Analysis* **29**: 308–325.

Genest, C. and MacKay, R. J. (1986). The joy of copulas: Bivariate distributions with given marginals, *The American Statistician* **40**: 280–283.

Genest, C. and Rivest, L. (1993). Statistical inference procedures for bivariate Arhimedean Copulas, *Journal of the American Statistical Association* **88**: 1034–1043.

Gill, R., van der Laan, M. and Robins, J. (1997). Coarsening at random, characterizations, conjectures and counter examples, *Proceedings of the First Seattle Symposium in Biostatistics 1995*, Springer Verlag, New York, pp. 255–294.

Gill, R., van der Laan, M. and Wellner, J. (1995). Inefficient estimators of the bivariate survival function for three models, *Ann. Inst. H. Poincaré Probab. Statist.* **31**(3): 545–597.

Heitjan, D. and Rubin, D. (1991). Ignorability and coarse data, *The Annals of Statistics* **19**(4): 2244–2253.

Hougaard, P. (1987). Modelling multivariate survival, *Scandinavian Journal of Statistics* **14**: 291–304.

Jacobsen, M. and Keiding, N. (1995). Coarsening at random in general sample spaces and random censoring in continuous time, *Annals of Statistics* **23**: 774–786.

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

Prentice, R. L. and Cai, J. (1992). Covariance and survivor function estimation using censored multivariate failure time data, *Biometrika* **79**(3): 495–512.

Pruitt, R. (1991). Strong consistency of self-consistent estimators: general theory and an application to bivariate survival analysis. Technical report nr. 543, University of Minnesota.

Pruitt, R. (1993). Small sample comparisons of six bivariate survival curve estimators, *Journal of Statistical Computing and Simulation* **45**: 147–167.

Quale, C., van der Laan, M. and Robins, J. (2001). Locally efficient estimation with bivariate right censored data. Technical report nr. 99, Division of Biostatistics.

Robins, J. (1989a). The analysis of randomized and non-randomized AIDS treatment trials using a new approach to causal inference in longitudinal studies, *in* L. Sechrest, H. Freeman and A. Mulley (eds), *Health Service Research Methdology: A Focus on AIDS*, NCHSR, U.S. Public Health Service, Dordrecht, pp. 113–159.

Robins, J. (1989b). Errata to: "Addendum to: 'A new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect'" [Comput. Math. Appl. **14** (1987), no. 9-12, 923–945; MR 89a:92048], *Computers and Mathematics with Applications* **18**(5): 477.

Robins, J. (1992). Estimation of the time-dependent accelerated failure time model in the presence of confounding factors, *Biometrika* **79**: 321–334.

Robins, J. (1998). Marginal structural models versus structural nested models as tools for causal inference, *Symposium on Prospects for a Common Sense Theory of Causation*, AAAI Technical Report, Stanford, CA.

Robins, J. (1999). Marginal structural models versus structural nested models as tools for causal inference, *in* M. Halloran and D. Berry (eds), *Statistical Models in Epidemiology: The Environment and Clinical Trials.*, Vol. 116, Springer-Verlag, NY, pp. 95–134.

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

Robins, J., Rotnitzky, A. and van der Laan, M. (2000). Comment on "On Profile Likelihood" by S.A. Murphy and A.W. van der Vaart, *Journal of the American Statistical Association – Theory and Methods* **450**: 431–435.

Robins, J., Rotnitzky, A. and Zhao, L. (1994). Estimation of regression coefficients when some regressors are not always observed, *Journal of the American Statistical Association* **89**(427): 846–866.

Sklar, A. (1959). Fonctions de répartition à n dimensionset leurs marges, *Publ. Inst. Statis. Univ* **8**: 229–231.

van der Laan, M. (1996). Efficient estimator of the bivariate survival function and repairing NPMLE, *Annals of Statistics* **24**: 596–627.

van der Laan, M. (1997). Nonparametric estimators of the bivariate survival function under random censoring, *Statist. Neerlandica* **51**(2): 178–200.

van der Laan, M. (2002). *Unified methods for censored longitudinal data and causality*, Springer, New York. To appear.

van der Laan, M., Hubbard, A. and Robins, J. (2000). Locally efficient estimation of multivariate survival function in longitudinal studies. Technical report nr. 82, Division of Biostatistics.

van der Laan, M. and Yu, Z. (2001). Comment on "Inference for semiparametric models: some questions and an answer" by P.J. Bickel and J. Kwon, *Statistica Sinica* **11**: 910–916.

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

Table 1: $MSE_{\mu_0}/MSE_{\mu_n^{Dab}}$ based on 200 simulated data sets of sample size 250. (T_1, T_2) and (C_1, C_2) are generated from frailty models with covariates (Z_1, Z_2) ~Bernoulli (0.5). $G(. \mid X)$ is estimated using a bivariate gamma frailty model with covariates. Correlations between T_1 and C_1 and T_2 and C_2 are approximately 0.4. $P(T1 < C1) = 0.65$ and $P(T_2 < C_2) = 0.65$.

	$t_1 = 0.1$	$t_1 = 1$	$t_1 = 4$	$t_1 = 10$
$t_2 = 0.1$	0.961543	0.8015517	0.2000056	0.2023185
$t_2 = 1$	0.9222071	0.6194325	0.2619613	0.2579991
$t_2 = 4$	0.1769921	0.3169093	0.1994758	0.2131806
$t_2 = 10$	0.2335622	0.3569389	0.2638717	0.2433467

Table 2: $MSE_{\mu_0}/MSE_{\mu_n^{Dab}}$ based on 200 simulated data sets of sample size 250. (T_1, T_2) are generated from frailty models with covariates (Z_1, Z_2) ~Bernoulli (0.5). $G(. \mid X)$ is from a bivariate gamma frailty model (no covariates). $G(. \mid X)$ is estimated using a bivariate gamma frailty with covariates Z. Correlations between T_1 and C_1 and T_2 and C_2 are approximately 0. $P(T1 < C1) = 0.70$ and $P(T_2 < C_2) = 0.70$.

	$t_1 = 0.05$	$t_1 = 0.2$	$t_1 = 3$	$t_1 = 8$
$t_2 = 0.05$	0.9990788	0.9985483	0.9702160	0.9813311
$t_2 = 0.2$	0.9924038	0.9946938	0.9650496	0.9741110
$t_2 = 3$	0.9814260	0.9789798	0.9504253	0.9650510
$t_2 = 8$	0.9806082	0.9821724	0.9665981	0.9789226

Table 3: μ_n^{Dab}, μ_n^0, μ_n^1 estimates of $P(T_1 \geq t_1, T_2 \geq t_2)$ with 95% confidence interval calculated for μ_n^1 on a data set simulated from simulation setup I.

(t_1, t_2)	$P(T_1 \geq t_1, T_2 \geq t_2)$	μ_n^{Dab}	μ_n^0	μ_n^1	95% CI of μ_n^1	$c_{nu,n}$
(0.1,0.1)	0.949409	0.974827	0.974531	0.9742604	(0.9381157,0.9981405)	2.510069
(0.1,1.0)	0.794018	0.826173	0.797576	0.7961396	(0.7290466,0.8632327)	2.300677
(0.1,4.0)	0.594459	0.720699	0.605632	0.6028978	(0.4939351,0.7118605)	2.361191
(0.1,10.0)	0.460084	0.658725	0.541393	0.5387704	(0.4232607,0.6542801)	2.404321
(1.0,1.0)	0.674216	0.696585	0.657125	0.6601110	(0.5843612,0.7298903)	2.460159
(1.0,4.0)	0.509527	0.620734	0.528965	0.5232135	(0.4216446,0.6247825)	3.649437
(1.0,10.0)	0.394196	0.571199	0.474861	0.46910810	(0.3595429,0.5786733)	3.441779
(4.0,4.0)	0.390691	0.539030	0.363114	0.3766390	(0.2439028,0.4823266)	2.525892
(4.0,10.0)	0.302516	0.482808	0.320330	0.3077152	(0.1829420,0.4324884)	2.539978
(10.0,10.0)	0.235763	0.453462	0.306258	0.2954581	(0.1740513,0.4168649)	3.399358