University of California, Berkeley

U.C. Berkeley Division of Biostatistics Working Paper Series

Year 2002   *Paper* 119

Locally Efficient Estimation with Bivariate Right Censored Data

Christopher M. Quale[*]   Mark J. van der Laan[†]

James M. Robins[‡]

[*]Division of Biostatistics, School of Public Health, University of California, Berkeley

[†]Division of Biostatistics, School of Public Health, University of California, Berkeley, laan@berkeley.edu

[‡]Department of Epidemiology, Harvard School of Public Health

Locally Efficient Estimation with Bivariate Right Censored Data

Christopher M. Quale, Mark J. van der Laan, and James M. Robins

Abstract

Estimation for bivariate right censored data is a problem that has had much study over the past 15 years. In this paper we propose a new class of estimators for the bivariate survivor function based on locally efficient estimation. The locally efficient estimator takes bivariate estimators Fn and Gn of the distributions of the time variables T1,T2 and the censoring variables C1,C2, respectively, and maps them to the resulting estimator. If Fn and Gn are consistent estimators of F and G, respectively, then the resulting estimator will be nonparametrically efficient (thus the term "locally efficient"). However, if either Fn or Gn (but not both) is not a consistent estimator of F or G, respectively, then the estimator will still be consistent and asymptotically normally distributed. We propose a locally efficient estimator which uses a consistent, non-parametric estimator for G and allows the user to supply lower dimensional (semi-parameteric or parametric) model for F. Since the estimator we choose for G will be a consistent estimator of G, the resulting locally efficient estimator will always be consistent and asymptotically normal, and our simulation studies have indicated that using a lower dimensional model for F gives excellent small sample performance. In addition, our algorithm for calculation of the efficient influence curve at true distributions for F and G yields also the efficiency bound which can be used to calculate relative efficiencies for any bivariate estimator. In this paper we will introduce the locally efficient estimator for bivariate right censored data, present an asymptotic theorem, present the results of simulation studies and perform a brief data analysis illustrating the use of the locally efficient estimator.

1 Introduction

Bivariate right censored data arises when there are two time to event variables of interest (T_1, T_2) in which for some observations, a process (independent of the event of interest) prevents us from observing the full time to event of one or both time variables. This process is represented by the censoring variables (C_1, C_2). Thus in a bivariate right censored data set we are seeing n i.i.d copies of $Z_i = (Y_{1i}, Y_{2i}, \Delta_{1i}, \Delta_{2i})$ $(i = 1, 2, \ldots, n)$, where $Y_{1i} \equiv T_{1i} \wedge C_{1i}$, $Y_{2i} \equiv T_{2i} \wedge C_{2i}$ and $\Delta_{1i} \equiv I(T_{1i} \leq C_{1i})$ and $\Delta_{2i} \equiv I(T_{2i} \leq C_{2i}$. Here $(T_{1i}, T_{2i}) \sim F$ (with corresponding bivariate survival function S) and $(C_{1i}, C_{2i}) \sim G$, where F and G are unspecified and (T_1, T_2) is independent of (C_1, C_2). Note that $Z_i \sim P_{F,G}$ for a distribution indexed by F and G. Let $\mu = \mu(F)$ be the parameter of interest. A typical parameter of interest is the survival function $\mu = S(t_1, t_2)$ at a given point (t_1, t_2).

There are several existing non-parametric estimators of the bivariate survival function. Some prominent estimators include those of Dabrowska (1988), Prentice and Cai (1992a), Pruitt (1991b), and van der Laan (1996b), among others. It is known that the NPMLE for continuous data is not consistent (Tsai, Leurgans and Crowley, 1986). Thus many of the existing bivariate estimators including Dabrowska and Prentice-Cai are explicit esti-

1

mators based on respresentations of the bivariate survival function in terms of distribution functions of the data.

Pruitt (1991b) proposed an estimator which is the solution of an ad hoc modification of the self-consistency equation. Pruitt's estimator tackles the non-uniqueness of the original self-consistency equation of the NPMLE by estimating conditional densities over the half-lines implied by the singly-censored observations. van der Laan (1996b) proves uniform consistency, $\sqrt{n}$-weak convergence, and validity of the bootstrap of Pruitt's estimator. However this estimator is not asymptotically efficient and its practical performance is not as strong as that of Dabrowska, Prentice-Cai and van der Laan (van der Laan, 1997).

The Dabrowska and Prentice and Cai estimators have been shown to have good practical performance (Bakker, 1990, Prentice and Cai, 1992b, Pruitt, 1993, van der Laan, 1997), but are not, in general, non-parametrically efficient. Dabrowksa's estimator is based on a clever representation of a multivariate survival function in terms of its conditional multivariate hazard measure. The Prentice-Cai estimator is related to Dabrowska's except that it also uses the Volterra structure suggested by Bickel (see Dabrowska 1988). Also, as these estimators are based on smooth functionals of the data, results such as consistency, asymptotic normality,

2

correctness of the bootstrap, consistent estimation of the variance of the influence curve all hold by application of the functional delta method: see Gill (1992) and Gill, van der Laan and Wellner (1995). In fact, both Prentice-Cai and Dabrowska are "locally" efficient in the sense that both are efficient at complete independence between T_1, T_2, C_1 and C_2 as proved in Gill, van der Laan and Wellner (1995).

The "Sequence of Reductions NPMLE," or SOR-NPMLE of van der Laan (1996b) makes use of the observation of Pruitt (1991a) that the inconsistency of the NPMLE is due to the fact that singly-censored observations imply half-lines for T which do not contain any uncensored observations. To deal with this problem, van der Laan proposes to interval censor the singly censored observations in the sense that he replaces the uncensored component, say T_{1i} of the singly censored observations by the observation that T_{1i} lies in a small predetermined interval around T_{1i}. Van der Laan (1996) also proposes a further reduction based on the discretization of the C_i's to facilitate factorization of the joint likelihood into an F part and a G part to avoid having to estimate G. This estimator was shown to have good practical performance (van der Laan, 1997) in comparison with the Dabrowksa, Prentice Cai and Pruitt estimators for small intervals around the uncensored components of singly censored observations. In van der Laan

3

(1996b) it is shown that if the reduction of the data converges to zero slowly enough, then the SOR-NPMLE is asymptotically efficient.

As noted above, the Dabrowska, Prentice-Cai and Pruitt estimators are not, in general, efficient estimators. As the SOR-NPMLE of van der Laan is "globally" efficient, a larger sample size may be necessary before its asymptotic properties take effect (this need for a larger sample size becomes more obvious when generalizing the estimator to higher dimensions). In addition, the SOR-NPMLE requires a choice of bandwidth. We propose a new class of estimators that is guaranteed to be consistent and asymptotically normal, is efficient if user supplied estimators for F are consistent, and can overcome the "curse of dimensionality" by guessing a lower dimensional model for F and thus can realize good small sample performance.

We will assume

$$\overline{G}(T_1, T_2) > \delta > 0 \; F\text{-a.e.,} \tag{1}$$

which establishes the desired invertibility of the information operator. However, we note that our locally efficient doubly robust estimator, which uses user supplied estimators F_n and G_n of F and G, respectively, will still be CAN if F_n is consistent for the true F, even when assumption (1) is violated.

Artificial censoring: The condition (1) will be true if the distribution of (T_1, T_2) has compact support contained in a rectangle $[0, \tau_1] \times [0, \tau_2] \subset \mathbb{R}^2$

4

and $\bar{G}(\tau_1, \tau_2) > \delta > 0$. Consequently, as proposed in van der Laan (1996a,b), one can artificially censor the data so that this assumption holds in the following manner: given a (τ_1, τ_2) satisfying $\bar{G}(\tau_1, \tau_2) > 0$, if $\tilde{T}_j > \tau_j$, then set $\tilde{T}_j = \tau_j$ and $\Delta_j = 1$, $j = 1, 2$. The artificially censored data now follows a distribution $P_{F_\tau, G}$, where F_τ equals F on $[0, \tau_1) \times [0, \tau_2)$ and $F(\tau_1, \tau_2) = 1$ and $\bar{G}(\tau) > 0$, as required. This means that we can still estimate the bivariate distribution F on $[0, \tau_1) \times [0, \tau_2)$ with the artificially censored data structure. In practice, this means that we obtain more robust estimators of F on this rectangle.

The locally efficient (LE) estimator takes estimators F_n and G_n of the bivariate distributions F and G, respectively, and maps them to an estimator $\hat{S}_{LE}$ of the bivariate survivor function S. We propose an estimator which uses a consistent, non-parametric estimator (such as Dabrowska's) for G, and a lower dimensional (semi-parametric or parametric) model for F, which has two beneficial properties: one, using a consistent estimator for G guarantees that the resulting estimator $\hat{S}_{LE}$ will be consistent and asymptotically normal, and two, our simulation studies indicate that using a lower dimensional model for F (with Dabrowska for G) produces excellent practical performance (see simulations in our technical report and our simulation section). The resulting LE estimator will be a consistent, asymptotically

5

normally distributed estimator of S, and will be efficient if the user supplied estimator for F is consistent. In addition, our simulation studies indicate that if the model for F is misspecified (namely $F_n \to F_1 \neq F$), the estimator is still stable and consistent, and produces surprisingly good small sample performance; the practical finding that it is better to parametrize F than it is to parametrize G might be partly explained by the fact that the estimator at inconsistent F_n is more efficient at a nonparametric model for G than it is at a parametric or semiparametric submodel for G (see Theorem A). Finally, in many applications one always observes (C_1, C_2) so that G can be well estimated with the empirical distribution of (C_{1i}, C_{2i}), $i = 1, \ldots, n$.

In this paper we will first describe the LE estimator for bivariate right censored data. We will show how to obtain an estimate of the influence curve (IC) of $\hat{S}_{LE}$ which allows us to estimate the variance of $\hat{S}_{LE}$ and construct confidence intervals for our resulting estimate. In subsection 2.3 we present an asymptotic theorem establishing a formal local efficiency result for our proposed estimator. In section 4 we will present the some of our results of a simulation study examining the performance of the LE estimator relative to Dabrowska's estimator, evaluating the amount of "protection" against misspecification of F or G we get by using LE estimation, and assessing the performance of estimated $(1 - \alpha)$ confidence intervals. For a complete

6

presentation of our simulation study we refer to our technical report. Finally, in section 5 we will implement the LE estimator on a dataset from a twin study examining time to onset of appendicitis (Duffy, 1990).

2   Locally Efficient Estimator for Bivariate Right Censored Data

In this paper we are primarily concerned with the estimation of the bivariate survivor function S of (T_1, T_2). We will represent this parameter more generally as μ defined at a point (t_1, t_2): $\mu(t_1, t_2) \equiv S(t_1, t_2)$. The one-step locally efficient estimator is defined as:

$$\hat{\mu}_{LE}(t_1, t_2) = \hat{\mu}_0(t_1, t_2) + \frac{1}{n} \sum_{i=1}^{n} IC(Z_i \mid F_n, G_n, \hat{\mu}_0(t_1, t_2)) \tag{2}$$

where $\hat{\mu}_0(\cdot, \cdot)$ is a consistent initial estimator of $\mu(\cdot, \cdot)$ and F_n and G_n are estimators of F and G, respectively. $IC(Z_i \mid F_n, G_n, \hat{\mu}_0(t_1, t_2))$ is an estimate of the efficient influence curve $IC(Z_i \mid F, G, \mu(t_1, t_2))$ of the parameter $\mu(t_1, t_2)$. The IC has a well known representation as defined in Bickel et. al. (1993) as:

$$IC(Z \mid F, G, \mu) = A_F I_{F,G}^{-}(\kappa(\mu)) \tag{3}$$

Here $A_F(\cdot) : L_0^2(F) \rightarrow L_0^2(P_{F,G})$ is the score operator for F defined as $E(\cdot \mid Z)$, $I_{F,G} : L_0^2(F) \rightarrow L_0^2(F)$ the information operator $I_{F,G}$ defined

7

as $A_G^T(A_F)$, where $A_G^T(\cdot) : L_0^2(P_{F,G}) \to L_0^2(F)$ is the transpose of the score operator for G and is defined as $E(\cdot \mid T_1, T_2)$ and $\kappa(\mu)$ is the efficient influence curve for μ under the full data model. Thus the IC can be seen as a mapping from a full data estimating function κ of (T_1, T_2) to a function of the observed data. For the case where $\mu(t_1, t_2) = S(t_1, t_2)$ the estimating function κ is defined to be $I(T_1 > t_1, T_2 > t_2) - S(t_1, t_2)$. In order for $I_{F,G}$ to be invertible, we need to impose the condition (1) that $\overline{G}(T_1, T_2) > \delta > 0$, F a.e.

To understand what type of estimator F_n is needed, in the next paragraphs we will inspect the smoothness of $IC(Y \mid F, G, D)$ in F, G in more detail. We will conclude that one should use a discretized version of a *smooth* estimate $\tilde{F}_n$ of F with consistent densities, and one can use any discrete estimator G_n that consistently estimates G. This motivates us to consider parametric models for F and nonparametric models for G.

For a cdf F we let $F(dt) = F(t) - F(t-)$ if F is discrete and $\frac{\partial}{\partial t}F(t)$ if F is absolutely continuous with respect to Lebesgue measure. Let $F(dt_1, dt_2) = F(t_1, t_2) - F(t_1-, t_2) - F(t_1, t_2-) - F(t_1-, t_2-)$ if F is discrete and $\frac{\partial}{\partial t_1 \partial t_2}F(t_1, t_2)$ if F is absolutely continuous with respect to Lebesgue measure. Let $F(dt_1, t_2) = F(t_1, t_2) - F(t_1-, t_2)$ if F is discrete and $\frac{\partial}{\partial t_1}F(t_1, t_2)$ if F is absolutely continuous with respect to Lebesgue measure (define $F(t_1, dt_2)$ in an analagous fashion). Also let

8

$F^1(t_1, t_2) = F(t_1, t_2) - F(t_1-, t_2)$ if $t_1 \rightarrow F(t_1, t_2)$ is discrete, otherwise

let $F^1(t_1, t_2) = \frac{\partial}{\partial t_1} F(t_1, t_2)$ if $t_1 \rightarrow F(t_1, t_2)$ is absolutely continuous with

respect to Lebesgue measure (define $F^2(t_1, t_2)$ similarly).

Let $T = (T_1, T_2)$, $s = (s_1, s_2)$, $c = (c_1, c_2)$, $dc = (dc_1, dc_2)$, $ds = (ds_1, ds_2)$. A straightforward application of Fubini's theorem gives us:

$$
\begin{aligned}
I_{F,G}(T) \quad = \quad & \overline{G}(T)h(T) \\
&+ \frac{\int_0^\infty \left\{ \int_0^\infty I(c_2 \le T_2, s_2 > c_2) h(T_1, s_2) G(T_1, dc_2) \right\} F^1(T_1, ds_2)}{\int_{c_2}^\infty F^1(T_1, ds_2)} \\
&+ \frac{\int_0^\infty \left\{ \int_0^\infty I(c_1 \le T_1, s_1 > c_1) h(s_1, T_2) G(dc_1, T_2) \right\} F^2(ds_1, T_2)}{\int_{c_1}^\infty F^2(ds_1, T_2)} \\
&+ \frac{\int_0^\infty \left\{ \int_0^\infty I(c_1 \le T_1, s_1 > c_1) I(c_2 \le T_2, s_2 > c_2) h(s) G(dc) \right\} F(ds)}{\int_c^\infty F(ds)}
\end{aligned}
$$

$$(4)$$

In our estimate of $IC(Z \mid F, G, \mu)$ we substitute for F and G discretized

estimates (possibly on a fine grid) of F and G so that we only need to define

the information operator and its inverse at discrete F and G. For example,

the frailty estimator (Clayton and Cusick 1985) F_n of F puts mass on the

grid of uncensored values of Y_1 and Y_2. If there are m_1 unique uncensored

values of Y_1 and m_2 unique uncensored values of Y_2, then the support is

defined on an m_1 by m_2 grid. Thus the support of F_n can be represented

by the $m_1 m_2$ dimensional vector $\tilde{t} = \{(t_{11}, t_{21}), \ldots, (t_{1 m_1 m_2}, t_{2 m_1 m_2})\}$, and

thus our estimate of $I_{F,G}$ will be an $m_1 m_2$ by $m_1 m_2$ matrix. An estimate of

9

$F(dt_1, dt_2)$ can be obtained as follows

$$F_n(dt_1, dt_2) = F_n(t_1 + \delta_1, t_2 + \delta_1) - F_n(t_1, t_2 + \delta_1) - F_n(t_1, t_2 + \delta_1) + F_n(t_1, t_2)$$

Estimation of $G(dc_1, dc_2)$ is accomplished in a similar fashion.

For a discrete underlying distribution F, this discretization should provide a good estimate of the true $F(dt_1, dt_2)$, and thus the estimates of the information operator should approach the true $I_{F,G}$. If F is continuous and the grid defined by $\tilde{t}$ gets finer and finer with sample size and F_n is a discretized version of an estimate $\tilde{F}$ of F for which $d/dx\tilde{F}$ consistently estimates d/dxF, then the estimated $I_{F_n,G}$ will converge to the true $I_{F,G}$. We refer the reader to van der Laan (1996) in which a similar result is proved. Of course as $m_1 m_2$ gets larger, it becomes more computationally expensive to calculate and invert the information operator (see section 2.1). In practical use if the value $m_1 m_2$ becomes too large, it is also possible to discretize the data (e.g. round to fewer significant digits) so that the gridsize $m_1 m_2$ is manageable. Both the simulation result in our technical report for continuos data table and the data analysis verify that the estimator performs well for continuous data. In our technical report we describe the implementation of the locally efficient estimator. The software was written using the R language (which works also on S-Plus),

10

2.1 Inversion of the Information Operator

In order to calculate the Influence Curve of the estimator $\hat{S}_{LE}$, it is necessary find a solution to the system of equations $\gamma = I_{F,G}^{-}(\kappa)$. The dimension of $I_{F,G}$ is determined by the dimension m of the vector of points on which F_n puts mass, and can be quite large. In order to make this calculation computationally feasible, we used a result of van der Laan (1998), in which he developed an iterative algorithm for calculating γ. This algorithm requires km^2 steps for a constant k, and our implementation indicates that the constant k is quite small for reasonably large m.

van der Laan (1998) showed the following results for the information operator $I_{F,G} : L_0^2(F) \to L_0^2(F)$:

- If for all $\gamma \in L_0^2(F)$ with $\| \gamma \|_F > 0$ we have $\| A_F(h) \|_F > 0$, then $I_{F,G}$ is 1-1.

- If there exists a $\delta > 0$ so that for all $\gamma \in L_0^2(F)$ we have $\| A_F(h) \|_{P_{F,G}} \geq \delta \| \gamma \|_F$ for some $\delta > 0$, then $I_{F,G}$ is onto and has bounded inverse with operator norm smaller than or equal to $1/\delta^2$ and its inverse is given by:

$$I_{F,G}^{-1} = \sum_{i=0}^{\infty} (I - I_{F,G})^i$$

11

This suggests that the following iterative algorithm may be used in order to calculate $\gamma = I_{F,G}^-(\kappa)$:

$$\gamma^{k+1} = \kappa - (I - I_{F,G})(\gamma^k)$$

where $\gamma^0 = \kappa$ and iteration continued until $\| \gamma^{k+1} - \gamma^k \|_F < \epsilon$ for some $\epsilon > 0$.

In order to evaluate the practical performance of this algorithm in the context of the LE estimator for bivariate right censored data, we recorded values of $\| \gamma^{k+1} - \gamma^k \|_{F_n}$ for each value of k over 100 simulated datasets (the datasets were generated with "low" dependency as described in section 4). Recall that for $\mu(t_1, t_2) = S(t_1, t_2)$, we have that $\kappa(t_1, t_2) = I(T_1 > t_1, T_2 > t_2) - S(t_1, t_2)$. The value of ϵ was chosen conservatively to be 1×10^{-7} to ensure accuracy in the calculation of γ, and for these simulated datasets, $m = 225$. The results may be seen in figure 5 and indicate that the algorithm performs well. The values of $\| \gamma^{k+1} - \gamma^k \|_{F_n}$ appear to fall quickly after the first iteration and the algorithm converges within 15 iterations for the timepoints chosen.

2.2 Double robustness.

We now present a lemma which provides the basis of the protection of the asymptotic consistency of $\hat{\mu}_{LE}(t_1, t_2)$ against misspecification of the model

12

for either F or G (but not both).

Lemma 2.1 *For any pair of measures P, P_1, we write $P \equiv P_1$ if dP/dP_1 and dP_1/dP are well-defined and have finite supremum norm. Let $\mathcal{F}$ be the set of bivariate failure time distributions with support included in $[0, \tau] \subset \mathbb{R}^2_{\geq 0}$. Let $\mathcal{G}$ be the set of bivariate censoring distributions G satisfying $\bar{G}(t_1, t_2) > \delta$ for some $\delta > 0$ for all $(t_1, t_2) \in [0, \tau] \subset \mathbb{R}^2_{\geq 0}$. For any $F_1 \in \mathcal{F}$ and $G_1 \in \mathcal{G}$, let $IC(Z \mid F_1, G_1, D)$ be defined as in (3). Then, given a $G_1 \in \mathcal{G}$, $E_{P_{F,G}}(IC(Z \mid F_1, G_1, D)) = E_F D(X)$ if either $F_1 = F$ and $G \ll G_1$ or $G_1 = G$ and $F_1 \equiv F$. (Here $X = (T_1, T_2)$ denotes the full data) We also have that, given $G_1 \in \mathcal{G}$, $E_{P_{F,G}} IC(Z \mid F_1, G_1, D) = E_F D(X)$ if either $F_1 = F$ and $G \ll G_1$ or $G_1 = G$ and F_1 is discrete. Finally, $E_G(IC(Z \mid F_1, G, D) \mid X) = D(X)$ F_X-a.e. at any $F_1 \in \mathcal{F}$ and $G \in \mathcal{G}$.*

Proof. Let $(D[0, \tau], \|\cdot\|_\infty)$ be the Banach space consisting of real-valued functions defined on $[0, \tau]$ endowed with the supremum norm. Given a general F_1, we only know that, given a G_1 with $\bar{G}_1(T_1, T_2) > \delta > 0$, F-a.e., $I_{F_1, G_1} : L^2(F_1) \to L^2(F_1)$ is boundedly invertible as a Hilbert space operator, while if F_1 is discrete, then Gill, van der Laan, and Robins (2000) prove that $I_{F_1, G_1} : (D[0, \tau], \|\cdot\|_\infty) \to (D[0, \tau], \|\cdot\|_\infty)$ has a bounded inverse. Let us first consider the case where $G_1 = G$. By first taking the conditional

13

expectation, given $X = (T_1, T_2)$, it follows that

$$E_{P_{F,G}} A_{F_1} (A_G^\top A_{F_1})^{-1}(D)(Z) = E_F D'(X),$$

where for general F_1, $D'(X) \equiv A_G^\top A_{F_1}(A_G^\top A_{F_1})^{-1}(D)$ equals $D(X)$ only in $L^2(F_1)$, and if F_1 is discrete, then $D'(X) = D(X)$ in $(D[0, \tau], \|\cdot\|_\infty)$. Thus, if $F_1 \equiv F$ or F_1 is discrete, then $E_{P_{F,G}} A_{F_1}(A_G^\top A_{F_1})^{-1}(D)(Z) = E_F D(X)$. This proves the unbiasedness for the case where $G_1 = G$.

Let us now consider the case where $F_1 = F$. Firstly note that $I_{F,G_1} : L_0^2(F) \to L_0^2(F)$ is 1-1 and onto under the condition $G_1 \in \mathcal{G}$. Thus

$$E_{P_{F,G}} A_F I_{F,G_1}^{-1}(D)(Z)) = E_F I_{F,G_1}^{-1}(D)(X)) = 0,$$

which proves the lemma.

Also it should be noted that if you use the true distributions F and G and the true parameter $\mu(t_1, t_2)$ to calculate the influence curve in (3) then the variance of the resulting influence curve equals the efficiency bound. This result allows us to calculate relative efficiencies for any bivariate estimator: we report such tables in our technical report.

14

3 Asymptotic performance and confidence Intervals for $\hat{S}_{LE}$

Consider the situation in which G_n is a consistent and efficient estimator of G according to the model we have assumed on G. For example, G_n might be an efficient estimator of G under the assumption of a Frailty model, G_n might be the SOR-NPMLE of van der Laan (1996b) under the nonparametric independence model (i.e. (C_1, C_2) is independent of (T_1, T_2)) or if (C_1, C_2) are always observed one can estimate G with the empirical distribution of (C_{1i}, C_{2i}), $i = 1, \ldots, n$, in the nonparametric independence model. In addition, assume that F_n converges to some F_1 (in a strong sense so that f_n converges to f_1) not necessarily equal to the true F, which one expects to be the case if F_n is an estimate of F according to some guessed (semi-)parametric model.

Our general asymptotic theorem A in the appendix proves that, under regularity conditions, $\hat{S}_{LE}$ is asympotically linear with an influence curve

$$IC(Z) = IC(Z \mid F_1, G, \mu) - \Pi(IC(Z \mid F_1, G, \mu) \mid T_G), \qquad (5)$$

where $T_G \subset L_0^2(P_{F,G})$ is the tangent space of G generated by all scores of the G-part of the likelihood of Z, given by $L(G) = P(Z \mid (T_1, T_2))$, under the proposed model for G. If $F_1 = F$, then $IC(Z \mid F_1, G, \mu)$ equals the efficient

15

influence curve which is orthogonal to T_G so that $IC(Z) = IC(Z \mid F, G, \mu)$.

This shows that we can use $IC(Z \mid F_1, G, \mu)$ as a conservative influence curve of our locally efficient estimator $\hat{S}_{LE}(t_1, t_2)$, which is actually correct if our guessed model for F is correct (or if G is known so that T_G is empty). Thus a conservative bound for the asymptotic variance of $\sqrt{n}(\hat{S}_{LE})(t_1, t_2) - S(t_1, t_2))$ is given by, $var(IC(Z \mid F_1, G, S)$. Therefore we may use our estimate of the influence curve to obtain a $\geq 95\%$ asymptotic confidence interval for $\hat{S}_{LE}$ in the following way:

- Calculate $\hat{\sigma}^2_{\hat{S}_{LE}(t_1, t_2)} = \frac{1}{n^2} \sum_{i=1}^{n} (\widehat{IC}(Z_i) - \overline{IC})^2$

- CI $= (\hat{S}_{LE}(t_1, t_2) - 1.96\hat{\sigma}_{\hat{S}_{LE}(t_1, t_2)}, \hat{S}_{LE}(t_1, t_2) + 1.96\hat{\sigma}_{\hat{S}_{LE}(t_1, t_2)})$

where $\widehat{IC(Z_i)} = \widehat{IC}(Z_i \mid F_n, G_n, \hat{S}_0)$ and

$$\overline{IC} = \frac{1}{n} \sum_{i=1}^{n} \widehat{IC}(Z_i \mid F_n, G_n, \hat{S}_0)$$

If G_n is not an efficient estimator according to a model for G, but is a good estimator such as Dabrowska's estimator, then we believe that this estimate of the limit variance and the corresponding confidence interval will still be a good practical choice. Our simulation study shows indeed the good practical performance of these confidence intervals for the case that G is estimated with Dabrowska's estimator.

16

The practical performance of any estimator and, in particular, $\hat{S}_{LE}$, is a trade off between first and second order asymptotics. Though the first order asymptotics suggest to estimate F, G nonparametrically (or just G with a smoothed NPMLE) the fact is that such globally efficient estimators can suffer from large second order terms. In addition, the influence curve depends on partial densities of F, therefore using a non-smooth nonparametric estimator for F such as Dabrowska's may not be appropriate. On the other hand, we do want to be nonparametrically consistent in the model only assuming that (C_1, C_2) is independent of (T_1, T_2). To control the second order terms of the estimator $\hat{S}_{LE}$, i.e. to control the curse of dimensionality, and to still be nonparametrically consistent under the sole assumption that (C_1, C_2) is independent of (T_1, T_2), we believe that estimating G with Dabrowska's estimator or the empirical distribution if (C_1, C_2) is always observed, and guessing a (very) low-dimensional model for F is a good strategy.

Below we state a formal theorem for local efficiency of our estimator for the special case that (C_1, C_2) is always observed with directly verifiable conditions, which is proved in the appendix of our technical report. The condition that (C_1, C_2) is always observed allows us to use a simple efficient estimator of G in the independent censoring model and thereby simplifies

17

the proof, but it is certainly not a necessary condition. For some more comments regarding the proof we refer to the appendix.

Theorem 3.1 *Consider the case that $Z = (C_1, C_2, \min(T_1, C_1), \min(T_2, C_2))$ includes observing the censoring times. Assume that F has support contained in $[0, \tau] \subset \mathbb{R}^2_{\geq 0}$ and that $\bar{G}(\tau) > 0$. Let F_n be discrete with support contained in a fixed set of (say) M points in $[0, \tau]$ and assume $\|F_n - F_1\|_\infty \to 0$ in probability for some discrete F_1 with $dF_1 > 0$ on each of the M-support points. Let G_n be the empirical distribution based on (C_{1i}, C_{2i}), $i = 1, \ldots, n$. Then, μ_n^1 is a regular asymptotically linear estimator with influence curve*

$$IC \equiv \Pi(IC(\cdot \mid F_1, G, \mu) \mid T_G^\perp(P_{F,G})),$$

where $T_G(P_{F,G}) = \{h(C_1, C_2) - E_G(h(C_1, C_2)) : h\} \subset L_0^2(P_{F,G})$. In particular, if $F_1 = F$, then μ_n^1 is asymptotically efficient.

The projection operator is given by:

$$\Pi(IC(\cdot \mid F_1, G, \mu) \mid T_G) = E(IC(Z \mid F_1, G, \mu) \mid C_1, C_2).$$

4 Simulation Methods and Results

In these simulations we have $\bar{G}(\tau) = 0.15$. For computational feasibility and so that we could use the full grid of support points for F, for all but two sim-

18

ulations the generated Y_1 and Y_2 were discretized so the support of F was on $\{1, 2, \ldots 15\} \times \{1, 2, \ldots, 15\}$, and the support of G was $\{1, 2, \ldots\} X \{1, 2, \ldots\}$. Two simulations were run in which the support of F was on the integers $\{1, 2, \ldots 40\} \times \{1, 2, \ldots, 40\}$ in order to verify that the locally efficient estimator works well in cases where the underlying F was less discrete. The amount of correlation between T_1 and T_2 was controlled by adjusting the α_t parameter, and the amount of censoring was controlled by the λ parameters. Further details for the data generation method may be found in our technical report.

In the simulation studies we studied the small sample performance of the locally efficient estimator using a variety of models for F and G. Due to the data generation scheme, using a bivariate frailty model would correspond to choosing semiparametric efficient estimators for F and G. Using the true distributions for F and G should give performance close to the efficiency bound and provide us with the best performance. Using Dabrowska's estimator for both F and G corresponds to the "globally" efficient estimator, as both F and G will be estimated consistently. However, as mentioned in section 3, as Dabrowska's estimator is non-smooth and highly non-parametric, it may not be the optimal choice for the estimator for F. Using a misspecification model ("guessing" a Uniform distribution for F or G) should give us an

19

indication of whether or not the locally efficient estimator is still consistent even if we "guess" wrong. The simulations were run with simulated datasets of size $n = 300$ over 625 iterations (except for the two continuous data simulations, in which the sample size was 100, so that the "gridsize" defined by $m_1 m_2$ in section 2 remained computationally feasible). Each simulation, aside from the first (described below), was run at two dependency levels, low ($\alpha_t = 0.5$, corresponding to a correlation between T_1 and T_2 ofapproximately 0.31), and high ($\alpha_t = 2$, corresponding to a correlation between T_1 and T_2 of approximately 0.72), and moderate censoring ($P(T_1 > C_1) = 0.30$ and $P(T_2 > C_2) = 0.30$). For the complete set of simulations we refer to our technical report.

In the first simulation, we generated data with heavy censoring on T_1 ($P(T_1 > C_1) = 0.65$) and high correlation between T_1 and T_2 (0.72) and mild censoring on T_2 $P(T_2 > C_2) = 0.30$). The fraily estimator was used for F and Dabrowska's estimator was used for G. The results may be seen in table 1. Here we see that the locally efficient estimator greatly outperforms Dabrowska's estimator, at points by almost a factor of 3. We see that the region where the LE estimator performs best is for the marginal distribution of T_1 ($t_2 = 0$). The marginal distribution of Dabrowska's estimator equals the Kaplan-Meier estimator based on $(\min(T_1, C_1), \Delta_1)$ and suffers thus heavily

20

from the high amount of censoring. However, an efficient estimator tries to borrow information from T_2 when estimating the marginal distribution of T_1, which is in this simulation very beneficial since T_2 is almost always observed and T_1 and T_2 are strongly correlated.

Table 2 illustrates the "protection" property we have when G is mis-specified. Although we propose to always use a consistent estimator for G in practice, these simulations indicate that indeed we are protected against misspecification of G. Table 3 shows the estimated coverage probabilities of 95% confidence intervals constructed as described in section 3. Here we see that the confidence intervals provide estimates close to the ideal 0.95. For the complete set of simulations results we refer to our technical report.

5   Data Analysis

To demonstrate the use of the LE estimator, we looked at a dataset originally analyzed by Duffy et al (1990) of 1218 monozygotic female twins in which the outcome of interest was age (in years) at appendectomy. The data was obtained from a questionnaire sent in 1980 to twins over the age of 17 registered with the Australian National Health and Medical Research Council Twin Registry. Thus if T_1 is the time to appendectomy of the first twin (where assignment of a twin to T_1 or T_2 was determined by birth order,

where that information was available) and T_2 is the time to appendectomy of the second twin, then these time variables were censored by a common censoring variable C (i.e. $C_1 = C_2$). Specifically, the twin data can be represented by $Z_i = (Y_{1i}, Y_{2i}, \Delta_{1i}, \Delta_{2i})$ where $Y_{1i} = T_{1i} \wedge C_i$, $Y_{2i} = T_{2i} \wedge C_i$, $\Delta_{1i} = I(T_{1i} \leq C_i)$ and $\Delta_{2i} = I(T_{2i} \leq C_i)$, for $i = 1, \ldots, 1218$. To take advantage of this structure in the data, we estimated the distribution G of (C_1, C_2) using the well known Kaplan Meier estimator. Specifically, we created the variable Z', where $Z'_i = (Y'_i, \Delta'_i)$ where $Y'_i = \min(T_{1i}, T_{2i}, C_i)$, and $\Delta'_i = I(C \leq T_{1i} \wedge T_{2i})$, for $i = 1, \ldots, 1218$. Since $C_1 = C_2$, $P(C_1 > c_1, C_2 > c_2) = P(C > \max(c_1, c_2))$, and thus we can easily determine the estimate of the bivariate distribution of (C_1, C_2) from the univariate estimate of the distribution of C. To estimate the distribution of (T_1, T_2) we used the bivariate frailty estimator. There were 42 unique uncensored values of Y_1 and 41 unique uncensored values of Y_2, thus the set of support points $\tilde{t}$ (defined in section 2) had dimension 1722. The resulting estimate of the bivariate surface may be found in our technical report.

We assessed the amount of dependence between T_1 and T_2 by utilizing two tests of independence developed by Quale and van der Laan (2000). The first tests the null hypotheses that the events $T_1 > t_{1i}$ and $T_2 > t_{2i}$ are independent, for a given set of points $\mathbf{t} = \{(t_{1i}, t_{2i}), \ldots, (t_{1k}, t_{2k})\}$. This

22

test involves looking at the difference between the locally efficient estimate of the bivariate survival function (no assumption of independence) and the bivariate product of the estimated marginals (which is correct under the assumption of independence) at the points **t**. The distribution of this test statistic can be determined if one has an estimate of the influence curve. The second test tests the null hypothesis that all the events mentioned above are independent, thus giving us an idea of the overall dependence of T_1 and T_2. We refer the reader to Quale and van der Laan (2000) for details. The results for the first test are given in table 4 and indicate that the time to appendectomy appear to be more dependent the older the twins become. The second test resulted in a P-Value less than 0.001 ($= P(\chi_{49} > 275.6)$, for χ_{49} a Chi Square random variable with 49 degrees of freedom), indicating that the time to appendectomy was related among the two twins (a result which Duffy et al (1990) found among the monozygotic female twins).

Also of interest is the question of whether or not the frailty was the correct model for S. To determine this, we looked at the difference between the locally efficient estimator and the estimator of S using the frailty estimator ($\hat{S}_{LE}(t_1, t_2) - \hat{S}_{frail}(t_1, t_2)$). The results may be seen in table 5, which tabulates the differences. The frailty estimator appears to be very close to the locally efficient estimator at all of the selected points (t_1, t_2), and indeed

23

the frailty estimator lies within the 95% confidence interval for all of the selected points (t_1, t_2).

A   Asymptotics

Recall the one-step estimator $\mu_n^1 = \mu_n^0 + 1/n \sum_i IC(Z_i \mid F_n, G_n, \mu_n^0)$. Below we provide a theorem which provides conditions under which our one-step estimator is consistent and asymptotically linear in the model where the bivariate distributions F and G of (T_1, T_2) and (C_1, C_2), respectively, are unspecified. The conditions of this theorem require that the estimator G_n of G is consistent, but F_n can be an inconsistent estimator. Since we advertised the one-step estimator μ_n^1 for which G_n is globally consistent and F_n is locally consistent as an important contribution to the rich literature on nonparametric bivariate right-censored data, we decided not to state the theorem which only assumes that either F_n or G_n is consistent. The latter theorem can be found in van der Laan, Robins (2002). For empirical process theory we refer to van der Vaart, Wellner (1996) and for efficiency theory we refer to Bickel, Klaassen, Ritov, Wellner (1993).

Theorem A.1 *Let the full-data model $\mathcal{M}^F$ for the distribution of $X = (T_1, T_2)$ be unspecified and assume that the conditional distribution of (C_1, C_2), given X, satisfies CAR. Suppose we observe n i.i.d. obser-*

24

vations $Z_1, \ldots, Z_n$ of $Z = (\min(T_1, C_1), \Delta_1, \min(T_2, C_2), \Delta_2)$ or $Z = (C_1, C_2, \min(T_1, C_1), \min(T_2, C_2))$ (i.e. now the censoring times are observed as well). Recall the one-step estimator μ_n^1 of $S(t_1, t_2)$:

$$\mu_n^1 = \mu_n^0 + \frac{1}{n} \sum_{i=1}^{n} IC(Z_i \mid F_n, G_n, \mu_n^0).$$

Assume

(i) $IC(\cdot \mid F_n, G_n, \mu_n^0)$ falls in a $P_{F_X, G}$-Donsker class with probability tending to 1.

(ii) $\|IC(\cdot \mid F_n, G_n, \mu_n^0) - IC(\cdot \mid F_1, G, \mu)\|_{P_{F_X,G}} \to 0$ in probability.

(iii) $P_{F, G_n - G} \{IC(\cdot \mid F_n, G_n, \mu) - IC(\cdot \mid F_1, G_n, \mu)\} = o_P(1/\sqrt{n})$.

If $IC(\cdot \mid F_1, G, \mu) = IC(\cdot \mid F, G, \mu)$, then μ_n^1 is asymptotically efficient. In general, define for a G_1

$$\Phi(G_1) = P_{F_X, G}(\cdot \mid F_1, G_1, \mu).$$

If also

(iv) $\Phi(G_n)$ is an asymptotically efficient estimator of $\Phi(G)$ for a CAR-

25

model containing the true G with tangent space $T_G(P_{F_X,G}) \subset T_{CAR}(P_{F_X,G})$, then μ_n^1 is a regular asymptotically linear estimator with influence curve given by

$$IC \equiv \Pi(IC(\cdot \mid F_1, G, \mu) \mid T_G^\perp(P_{F_X,G})).$$

Conditions (i) and (ii) requires detailed understanding of the entropy of a class of functions containing $I_{F_n,G_n}^-(D)$ with probability tending to one. Since the inverse of the information operator $I_{F,G}$ is, in general, only understood in a $L^2(F)$-sense this is a hard task. However, in the proof of Theorem 3.1, as provided in our technical report, we show that this inverse is well understood in supremum norm and variation norm sense at a discrete F. This allows us to prove that our one-step estimator μ_n^1 using a discrete estimate F_n with support contained in a fixed set of (say) M points is consistent, asymptotically linear and if F_n happens to be consistent, then it is also efficient. Our simulations suggest that μ_n^1 is also locally efficient if F_n is a discrete estimate of F according to a arbitrarily fine discrete approximation (depending on n) of a parametric or semiparametric smooth model. In other words, we are not claiming that our condition on F_n is needed.

26

References

Bakker, D. M. (1990). Two nonparametric estimators of the survival function of bivariate right censored observations. Technical Report Report BS-R9035, Centre for mathematics and computer science, Amsterdam.

Bickel, P. J., Classen, C. A. J., Ritov, Y., and Wellner, J. A. (1993). *Efficient and adaptive inference in semiparametric models.* Hopkins University Press, Baltimore.

Clayton, D. 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. M. (1988). Kaplan meier estimate on the plane. *The Annals of Statistics*, 18:308–325.

Duffy, D. L., Martin, N. G., and et al, J. D. M. (1990). Appendectomy in australian twins. *American Journal of Human Genetics*, 47:590–592.

Gill, R. D. (1992). Multivariate survival analysis. *Theory Prob. Appl.*, 37:19–36.

Gill, R. D., van der Laan, M. J., and Robins, J. M. (1997). Coarsening at random; characterizations, conjectures and counter examples. In

27

Lin, D. Y. and Fleming, T., editors, *Survival Analysis*, pages 255–294. Springer Lecture Notes in Statistics 123.

Gill, R. D., van der Laan, M. J., and Wellner, J. A. (1995). Inneficient estimators of the bivariate survival function for three models. *Annalles de L'Institute Henri J. Poincare*, 31:545–597.

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

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

Prentice, R. L. and Cai, J. (1992b). Marginal and conditional models for the analysis of multivariate failiure time data. In Klein, J. P. and Goel, P. K., editors, *Survival Analysis State of the Art*, pages 495–512, Dordrecht. Kluwer.

Pruitt, R. C. (1991a). On negative mass assigned by the bivariate kaplan-meier estimator. *The Annals of Statistics*, 19:443–453.

Pruitt, R. C. (1991b). Strong consistency of self-consistent estimators: Gen-

28

eral theory and an application to bivariate survival analysis. Technical Report 543, University of Minnesota.

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

Quale, C. M. and van der Laan, M. J. (2000). Inference with bivariate truncated data. *Lifetime Data Analysis*, 6:391–408.

Tsai, W. Y., Leurgans, S., and Crowley, J. (1986). Nonparametric estimation of a bivariate survival function in the presence of censoring. *The Annals of Statistics*, 14:1351–1365.

van der Laan, M. J. (1996a). *Efficient and Inefficient Estimation in Semiparametric Models*. CWI Tract 114, Centre for Mathematics and Computer Science, Amsterdam, the Netherlands.

van der Laan, M. J. (1996b). Efficient estimation of the bivariate censoring model and repairing NPMLE. *The Annals of Statistics*, 24:596–627.

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

29

van der Laan, M. J. (1998). Identity for the NPMLE in censored data models. *Lifetime Data Analysis*, 4:83–102.

van der Laan, M. J. and Gill, R. (1999). Efficiency of the NPMLE in non-parametric missing data models. *Mathematical Methods of Statistics*, 8(2):251–276.

van der Laan, M. J., Hubbard, A., and Robins, J. M. (2000). Locally efficient estimation of a multivariate survival function in longitudinal studies. Technical report, University of California, Berkeley.

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

30

Table 1: Heavy censoring on T_1 and high correlation; Frailty F, Dabrowska G. MSE ratio for estimates of S between Locally Efficient Estimator and Dabrowska ($\frac{MSE_{loc}}{MSE_{dab}}$) for a correlation between T_1 and T_2 of approximately 0.72. Simulations are for 300 subjects over 625 iterations, with $\lambda_{t_{1,2}} = 0.1$, $\lambda_{c_1} = 0.3$ and $\lambda_{c_2} = 0.08$, corresponding to $P(T_1 > C_1) = 0.65$ and $P(T_2 > C_2) = 0.30$.

	$t_2 = 0$	$t_2 = 1$	$t_2 = 2$	$t_2 = 4$	$t_2 = 6$	$t_2 = 8$	$t_2 = 10$
$t_1 = 0$	1.01	1.03	1.02	1.00	0.98	0.96	0.95
$t_1 = 1$	1.13	0.99	0.95	0.93	0.93	0.92	0.93
$t_1 = 2$	1.10	0.94	0.92	0.90	0.91	0.89	0.91
$t_1 = 4$	0.93	0.88	0.88	0.89	0.89	0.88	0.89
$t_1 = 6$	0.76	0.77	0.80	0.87	0.87	0.87	0.86
$t_1 = 8$	0.58	0.60	0.65	0.74	0.75	0.75	0.73
$t_1 = 10$	0.34	0.36	0.38	0.44	0.49	0.50	0.52

31

Table 2: Moderate censoring; low, high, dependence; Frailty F, Misspecification (Uniform) G. MSE ratio for estimates of S between Locally Efficient Estimator and Dabrowska ($\frac{MSE_{loc}}{MSE_{dab}}$) for correlations between T_1 and T_2 of approximately 0.31 and 0.72. Simulations are for 300 subjects over 625 iterations, with $\lambda_{t_{1,2}} = 0.1$ and $\lambda_{c_{1,2}} = 0.08$, corresponding to $P(T_1 > C_1) = 0.30$ and $P(T_2 > C_2) = 0.30$.

	$t_2 = 1$	$t_2 = 2$	$t_2 = 4$	$t_2 = 6$	$t_2 = 8$	$t_2 = 10$
$t_1 = 1$	1.08,1.08	1.07,1.06	1.06,1.02	1.00,0.97	0.95,0.91	0.92,0.87
$t_1 = 2$	1.07,1.05	1.06,1.05	1.06,1.01	1.01,0.97	0.96,0.92	0.92,0.86
$t_1 = 4$	1.04,1.00	1.04,1.00	1.05,1.03	1.01,0.98	0.97,0.96	0.93,0.90
$t_1 = 6$	0.97,0.94	0.98,0.94	1.00,0.99	0.99,0.98	0.99,0.96	0.97,0.92
$t_1 = 8$	0.94,0.91	0.94,0.90	0.93,0.93	0.95,0.94	0.97,0.93	0.97,0.91
$t_1 = 10$	0.93,0.87	0.93,0.86	0.91,0.88	0.91,0.90	0.93,0.90	0.95,0.92

32

Table 3: Empirical Coverage Probabilities for estimated 95% Confidence Intervals for Locally Efficient Estimator; Moderate censoring; low, high, dependence; Frailty F, Dabrowska G. Results for correlations between T_1 and T_2 of approximately 0.31 and 0.72. Simulations are for 300 subjects over 625 iterations, with $\lambda_{t_{1,2}} = 0.1$ and $\lambda_{c_{1,2}} = 0.08$, corresponding to $P(T_1 > C_1) = 0.30$ and $P(T_2 > C_2) = 0.30$.

	$t_2 = 1$	$t_2 = 2$	$t_2 = 4$	$t_2 = 6$	$t_2 = 8$	$t_2 = 10$
$t_1 = 1$	0.92,0.91	0.92,0.92	0.92,0.94	0.92,0.92	0.90,0.91	0.92,0.92
$t_1 = 2$	0.90,0.90	0.90,0.92	0.90,0.91	0.90,0.90	0.90,0.91	0.90,0.90
$t_1 = 4$	0.91,0.91	0.91,0.92	0.91,0.91	0.92,0.93	0.91,0.91	0.92,0.90
$t_1 = 6$	0.89,0.91	0.91,0.90	0.89,0.92	0.90,0.90	0.90,0.92	0.89,0.92
$t_1 = 8$	0.91,0.93	0.91,0.94	0.91,0.93	0.91,0.95	0.92,0.94	0.92,0.93
$t_1 = 10$	0.91,0.94	0.92,0.93	0.90,0.94	0.91,0.92	0.91,0.92	0.91,0.94

33

Table 4: Pointwise tests of independence, twin data

	$t_1 = 6$	$t_1 = 12$	$t_1 = 20$	$t_1 = 28$	$t_1 = 36$	$t_1 = 44$	$t_1 = 52$
$t_2 = 6$	0.298	0.083	0.022	0.032	0.072	0.11	0.11
$t_2 = 12$	0.179	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
$t_2 = 20$	0.345	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
$t_2 = 28$	0.604	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
$t_2 = 36$	0.402	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
$t_2 = 44$	0.319	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
$t_2 = 52$	0.307	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001

Table 5: Difference between locally efficient estimate of the bivariate distribution and the bivariate frailty estimate, twin data ($\hat{S}_{LE} - \hat{S}_{frailty}$)

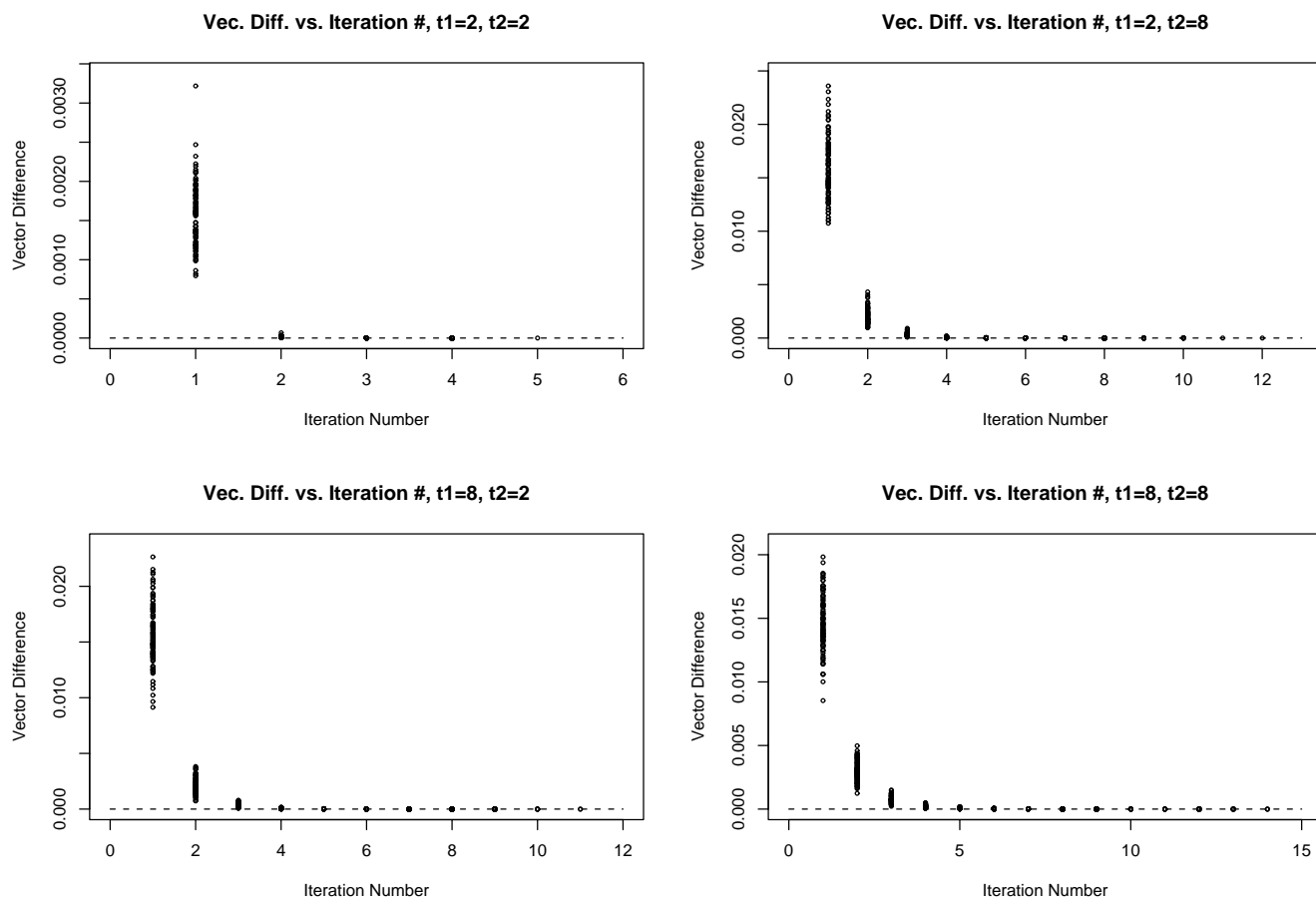|	$t_1 = 4$	$t_1 = 12$	$t_1 = 20$	$t_1 = 28$	$t_1 = 36$	$t_1 = 44$	$t_1 = 52$
$t_2 = 4$	-0.0003	-0.006	-0.015	-0.015	-0.026	-0.025	-0.021
$t_2 = 12$	0.004	0.002	-0.008	-0.007	-0.021	-0.020	-0.017
$t_2 = 20$	0.004	0.004	0.003	-0.004	-0.011	-0.011	-0.008
$t_2 = 28$	0.001	-0.002	-0.002	-0.006	-0.013	-0.015	-0.012
$t_2 = 36$	0.006	0.001	-0.008	-0.012	-0.017	-0.019	-0.017
$t_2 = 44$	0.001	-0.002	-0.009	-0.015	-0.020	-0.023	-0.021
$t_2 = 52$	-0.004	-0.007	-0.014	-0.021	-0.027	-0.030	-0.027

34

Figure 1: Inverse algorithm performance: Plots of $\| \gamma^{k+1} - \gamma^k \|_F$ ("Vector Difference") versus iteration number (k) for $I_{F,G}$ calculated at 100 simulated datasets, for $\kappa(t_1, t_2) = I(T_1 > t_1, T_2 > t_2) - S(t_1, t_2)$, at points $\mathbf{t} = \{(2,2), (2,8), (8,2), (8,8)\}$, where $\dim(I_{F,G}) = 225$. Dotted line indicates ϵ, which was set at 1×10^{-7}.

35