A class of non-parametric bivariate survival function estimators for randomly censored and truncated data

Hongsheng Dai[a*], Marialuisa Restaino[b] and Huan Wang[c]

[a]*Department of Mathematical Sciences, University of Essex, Wivenhoe Park, Colchester CO4 3SQ, United Kingdom*;

[b]*Department of Economics and Statistics, University of Salerno, Via Giovanni Paolo II 132, 84084 Fisciano (Salerno), Italy*

[c]*Dundee Epidemiology and Biostatistics Unit, University of Dundee, The Mackenzie Building, Kirsty Semple Way, Dundee DD2 4BF, United Kingdom*

This paper proposes a class of nonparametric estimators for the bivariate survival function estimation under both random truncation and random censoring. In practice, the pair of random variables under consideration may have certain parametric relationship. The proposed class of nonparametric estimators uses such parametric information via a data transformation approach and thus provides more accurate estimates than existing methods without using such information. The large sample properties of the new class of estimators and a general guidance of how to find a good data transformation are given. The proposed method is also justified via a simulation study and an application on an economic data set.

Keywords: bivariate survival function; random censoring; random truncation; correlated failure times; data transformation method

AMS Subject Classification: Primary: 62N01; Secondary: 62G05

[*]Corresponding author. Email: hdaia@essex.ac.uk

1. Introduction

The challenge in bivariate survival analysis usually comes from the incomplete information of the data, due to random censoring and random truncation (Wang 1991; van der Laan 1996a,b). Such kind of data occurs in many research areas, such as medicine, economics, insurance and social sciences. Consider that a business advisory team aims to study the failures for small and medium size businesses and then further provides advices to the businesses. In such a study, the time period T_1, from the establishment of a firm to the time of recruitment, and the time period T_2, from the recruitment time point to bankruptcy, are of interests. The times T_1 and T_2 are subject to random censoring by certain random variables C_1 and C_2, respectively (for example, the last follow-up). In practice, observations are also subject to random truncation. Suppose that data are accessible only from year 2004. Then only companies who were recruited after year 2004 will be available. Therefore there is a random time L_1, from the year of establishment to year 2004, such that only firms with $T_1 \geq L_1$ can be observed. This implied that T_1 is left-truncated. In summary, the observed data are $\{L_1, Y_1, \delta_1, Y_2, \delta_2\}$, where $Y_k = \min\{T_k, C_k\}$ and $\delta_k = I[X_k \leq C_k]$ and the aim of this study is to estimate the bivariate survival function of (T_1, T_2) under both censoring and truncation.

In this case only T_1 is subject to truncation L_1. In some situations both T_1 and T_2 are subject to truncation L_1 and L_2 respectively. An example is in (Huang et al. 2001), where the bivariate event times of interest are the parent's and child's ages of onset in genetic disease data and they are both right truncated at the parent's and child's ages at interview. For an affected parent-child pair to be included in the study, they have to be diagnosed with the disease before the time they are interviewed. No censoring is involved in their study.

Most existing research works (Woodroofe 1985; Keiding and Gill 1990; Wang 1991;

van der Laan 1996b) focused on bivariate survival analysis under either censoring or truncation. Some other existing methods dealt with bivariate survival function estimation under the scenario where one component is censored and truncated, but the other one is fully observed (Gürler 1997; Gijbels and Gürler 1998). Bivariate survival function estimation when both components are censored and truncated has received considerable attention recently (Shen 2006, 2007; Shen and Yan 2008). These methods, however, used an iterative computing method which is computationally heavy. Shen (2014) used the idea in Sankaran and Antony (2007) for competing risks set up, to propose two types of estimators ad generalizations of Dabrowska and Campbell and Foldes estimators. These estimators are easy to implement and do not require iteration. Dai and Fu (2012) proposed an estimator based on a polar coordinate transformation, which does not require iterative calculations and its large sample properties are established.

In this paper, we employ the idea in Dai and Fu (2012) and extend their methods to a class of estimators, based on different data transformations. The large sample properties of the class of estimators are also derived and a guidance of selecting good transformation functions is also provided.

The paper is organized as follows. In Section 2 the statistical models and the data transformation are discussed, the estimator is constructed and its large sample properties are provided. Then how to choose a good data transformation function is pointed out in Section 3. In Section 4, we present numerical studies for the performance of the estimator under different data transformation functions. A real data analysis is also provided in this section. Section 5 gives a discussion.

2. Statistical models and data transformation

Let (T_1, T_2) be the pair of non-negative random variables with bivariate cumulative distribution function and survival function $F(t_1, t_2) = P(T_1 \le t_1, T_2 \le t_2)$ and $S(t_1, t_2) = P(T_1 > t_1, T_2 > t_2)$, respectively. The pair of survival times (T_1, T_2) is subject to right censoring by a pair of censoring times (C_1, C_2), i.e. one can only observe $Y_k = \min\{T_k, C_k\}$ and $\delta_k = I[T_k \le C_k]$ for $k = 1, 2$. The pair (T_1, T_2) is also subject to random left truncation by a pair of truncation times (L_1, L_2), i.e. only subjects with $L_1 \le T_1$ and $L_2 \le T_2$ can be observed. Note that we focus on such type of truncation throughout this paper, which is called the type-I bivariate truncation in Dai and Fu (2012). In practice the data may also be type-II truncated (truncation with $L_1 \le T_1$ or $L_2 \le T_2$), for which the proposed method in this paper can be simply extended. We denote the observed data as $(Y_{1i}, Y_{2i}, \delta_{1i}, \delta_{2i}, L_{1i}, L_{2i})$ for $i = 1, \ldots, n$. We assume that (T_1, T_2) is independent of the censoring and truncation times, but the censoring and truncation times themselves can be mutually correlated, in the sense that we have the following joint probability function

$$G(t_1, t_2) = P(L_1 \le t_1 < C_1, L_2 \le t_2 < C_2)$$

We do not specify any parametric function for the above function G and it is estimated nonparametrically in the paper.

To develop a new estimator for the joint survival function S, we first consider a transformation for the time points (t_1, t_2) at which the survival function $S(t_1, t_2)$ is to be estimated. For any given arbitrary values (t_1, t_2), we define a transformation from (t_1, t_2)

to (z, α) as

$$t_2 = \zeta(t_1, \alpha), \qquad z = \int_0^{t_1} \sqrt{1 + \left[\frac{\partial \zeta(u, \alpha)}{\partial u}\right]^2} du \tag{1}$$

where $t_2 = \zeta(t_1, \alpha)$ means a function (curve) depending on a parameter α.

We assume that for each pair of (t_1, t_2) there is one and only one value of α such that $t_2 = \zeta(t_1, \alpha)$. Then the value α is uniquely determined by (t_1, t_2), once the function ζ is given. The function ζ needs to be continous and such that $\zeta(0, \alpha) = 0$. Then in equation (1), the value z is the distance from (t_1, t_2) to $(0, 0)$ along the curve $\zeta(\cdot, \alpha)$. Note that we can also write

$$z = \int_0^{t_2} \sqrt{1 + \left[\frac{\partial \zeta^{-1}(u, \alpha)}{\partial u}\right]^2} du, \tag{2}$$

where ζ^{-1} is the inverse function of ζ such that $t_1 = \zeta^{-1}(t_2, \alpha)$.

With the above definition, we can transform the target survival function from $S(t_1, t_2)$ to $S(z; \alpha)$, by the following formula

$$\begin{aligned} S(t_1, t_2) &= P(T_1 > t_1, T_2 > t_2) \\ &= P\left(\int_0^{T_1} \sqrt{1 + \left[\frac{\partial \zeta(u, \alpha)}{\partial u}\right]^2} du > z, \ \int_0^{T_2} \sqrt{1 + \left[\frac{\partial \zeta^{-1}(u, \alpha)}{\partial u}\right]^2} du > z\right) \\ &= P(Z(\alpha) > z) := S(z; \alpha), \end{aligned} \tag{3}$$

where

$$Z(\alpha) = \min\left\{ \int_0^{T_1} \sqrt{1 + \left[\frac{\partial \zeta(u, \alpha)}{\partial u}\right]^2} du, \ \int_0^{T_2} \sqrt{1 + \left[\frac{\partial \zeta^{-1}(u, \alpha)}{\partial u}\right]^2} du \right\}. \tag{4}$$

The above transformation implies that we only need to find a consistent estimate for $S(z;\alpha)$, which is the same as $S(t_1,t_2)$.

From the expression $S(z;\alpha) = P(Z(\alpha) > z)$, we can see that $S(z;\alpha)$ can be easily estimated if we have the transformed data $(Z_1(\alpha),\ldots,Z_n(\alpha))$, based on (T_{1i},T_{2i}) for $i=1,\ldots,n$. In practice, although the values of $Z(\alpha)$ in (4) may not be obtained due to censoring and truncation, we can still do similar transformation as follows. Define

$$\tilde{Y}_{1i} = \int_0^{Y_{1i}} \sqrt{1+\left[\frac{\partial\zeta(u,\alpha)}{\partial u}\right]^2}du, \quad \tilde{Y}_{2i} = \int_0^{Y_{2i}} \sqrt{1+\left[\frac{\partial\zeta^{-1}(u,\alpha)}{\partial u}\right]^2}du$$

and

$$\tilde{L}_{1i} = \int_0^{L_{1i}} \sqrt{1+\left[\frac{\partial\zeta(u,\alpha)}{\partial u}\right]^2}du, \quad \tilde{L}_{2i} = \int_0^{L_{2i}} \sqrt{1+\left[\frac{\partial\zeta^{-1}(u,\alpha)}{\partial u}\right]^2}du$$

Then the transformed data are given by

$$\begin{aligned}
&\tilde{Z}_i(\alpha) = \min\{\tilde{Y}_{1i},\tilde{Y}_{2i}\},\\
&\Delta_i(\alpha) = \delta_{1i}I[\tilde{Y}_{1i}\le\tilde{Y}_{2i}] + \delta_{2i}I[\tilde{Y}_{1i}\ge\tilde{Y}_{2i}] - \min(\delta_{1i},\delta_{2i})I[\tilde{Y}_{1i}=\tilde{Y}_{2i}],\\
&V_i(\alpha) = \max\{\tilde{L}_{1i},\tilde{L}_{2i}\}.
\end{aligned} \tag{5}$$

Based on the above transformation, the following lemma implies a product-limit estimator for $S(z;\alpha)$ (i.e. for $S(t_1,t_2)$).

LEMMA 2.1 *For fixed α, the hazard rate function of $Z(\alpha)$ is denoted by $\Lambda(dz;\alpha) = -\frac{S(dz;\alpha)}{S(z-;\alpha)}$. Then we have*

$$\Lambda(dz;\alpha) = \frac{P(\tilde{Z}_i(\alpha)\in dz, z>V_i(\alpha), \Delta_i(\alpha)=1)}{P(\tilde{Z}_i(\alpha)\ge z > V_i(\alpha))},$$

where $\tilde{Z}_i(\alpha) \in dz$ *denotes* $z \le \tilde{Z}_i(\alpha) < z + dz$. □

The proof of this lemma is given in Appendix. Note that the polar-coordinate transformation in Dai and Fu (2012) actually uses a specific function $\zeta(t_1, \alpha) := \alpha t_1$.

Based on the transformed observations $\{\tilde{Z}_i(\alpha), \Delta_i(\alpha), V_i(\alpha), i = 1, \ldots, n\}$ in (5), we define

$$\begin{aligned} N(ds; \alpha) &= \sum_{i=1}^{n} N_i(ds; \alpha)/n, \\ &:= \sum_{i=1}^{n} I[\tilde{Z}_i(\alpha) \in ds, s > V_i(\alpha), \Delta_i(\alpha) = 1]/n, \\ H_{(n)}(s; \alpha) &= \sum_{i=1}^{n} H_i(s; \alpha)/n, \\ &:= \sum_{i=1}^{n} I[\tilde{Z}_i(\alpha) > s \ge V_i(\alpha)]/n, \end{aligned} \tag{6}$$

and

$$\begin{aligned} H_{(n)}(t_1, t_2) &= \sum_{i=1}^{n} H_i(t_1, t_2)/n, \\ &:= \sum_{i=1}^{n} I[Y_{1i} > t_1 \ge L_{1i}, Y_{2i} > t_2 \ge L_{2i}]/n. \end{aligned} \tag{7}$$

Note that $H_{(n)}(t_1, t_2) = H_{(n)}(z; \alpha)$ and $H_i(t_1, t_2) = H_i(z; \alpha)$. An estimator for $\Lambda(dz; \alpha)$ is then given by $\hat{\Lambda}(dz; \alpha) = N(dz; \alpha)/H_{(n)}(z-; \alpha)$ and the product-limit estimator for $S(z; \alpha)$ is

$$\hat{S}(z; \alpha) = \prod_{s \le z} \left[1 - \frac{N\{s; \alpha\}}{H_{(n)}(s-; \alpha)} \right], \tag{8}$$

where $N\{s; \alpha\} = N(s; \alpha) - N(s-; \alpha)$. Since $S(z; \alpha) = S(t_1, t_2)$, $\hat{S}(z; \alpha)$ is also an estimator for $S(t_1, t_2)$.

The large sample properties of $\hat{S}$ are given by the following theorems, which follow from Dai and Fu (2012).

Assumption 2.2 *The function* $G(t_1, t_2) = P(C_1 > t_1 \geq L_1, C_2 > t_2 \geq L_2) > 0$, *almost surely with respect to* $S(t_1, t_2)$ *in* $\mathcal{A}$, *where* $\mathcal{A}$ *is the support of the distribution for* (T_1, T_2).

Theorem 2.3 *Under Assumption 2.2, for any* $(t_1, t_2) \in \mathcal{A}$ *such that* $S(z;\alpha) > 0$ *we have* $\hat{S}(z;\alpha) - S(z;\alpha) = r_n(z;\alpha)$, *where* r_n *is such that* $\sup_{\alpha,z} E[r_n(z;\alpha)]^2 = o(1)$. $\square$

Theorem 2.4 *Define* $M_i(ds;\alpha) = N_i(ds;\alpha) - H_i(s-;\alpha)\Lambda(ds;\alpha)$ *and* $M(ds;\alpha) = n^{-1}\sum_i M_i(ds;\alpha)$. *Under Assumption 2.2, for all* (z, α) *such that* $S(z;\alpha) > 0$, *we have that*

$$\hat{S}(z;\alpha) - S(z;\alpha) = -S(z;\alpha)\int_0^z \frac{\hat{S}(s-;\alpha)}{S(s;\alpha)} \frac{I[H_{(n)}(s-;\alpha) > 0]}{H_{(n)}(s-;\alpha)} M(ds;\alpha) + B(z;\alpha), \quad (9)$$

where

$$B(z;\alpha) = S(z;\alpha)\int_0^z \frac{\hat{S}(s-;\alpha)}{S(s;\alpha)} I[H_{(n)}(s-;\alpha) = 0]\Lambda(ds;\alpha). \quad (10)$$

We further have $\sqrt{n}(\hat{S}(z;\alpha) - S(z;\alpha)) \Rightarrow N(0, \sigma^2(z;\alpha))$, *where*

$$\sigma^2(z;\alpha) = S(z;\alpha)^2 \int_0^z \frac{1}{H(s-;\alpha)}\Lambda(ds;\alpha).$$

$\square$

A consistent estimator for $\sigma^2(z;\alpha)$ is

$$\hat{\sigma}^2(z;\alpha) = \hat{S}(z;\alpha)^2 \int_0^z \frac{I[H_{(n)}(s-;\alpha) > 0]}{H_{(n)}(s-;\alpha)} \hat{\Lambda}(ds;\alpha). \tag{11}$$

3. Interpretation of the transformation and selecting ζ

Dai and Bao (2009) consider the function $t_2 = \zeta(t_1, \alpha) = \alpha t_1$ as the data transformation function for censored data. The bivariate function $S(z;\alpha)$ is estimated by fixing α first and then find the estimates for all different values of α. The advantage of using such transformation is that we can estimate $S(z;\alpha)$ as a univariate survival function (viewed as a univariate function with fixed α) and then the univariate function estimate can be transformed back to the bivariate function estimate. The estimation based on the transformed data makes use of the bivariate data information, when $S(z;\alpha)$ is estimated with a given α. A naive approach of not using such transformation is to estimate $S(t_1, t_2)$ by fixing t_2 first and then find the estimates for all different values of t_2. Such a naive approach will ignore the censoring or truncation information of the second component, when estimating $S(t_1, t_2)$ with fixed t_2. Therefore the polar-coordinate transformation provides better estimate comparing to the naive approach (Dai and Bao 2009).

If we consider the transformed data in (5), the transformation is required not changing the partial order of the observations, in the sense that $Y_{1i} < Y_{1j}, Y_{2i} < Y_{2j}$ (or $\tilde{Z}_i(\alpha) < \tilde{Z}_j(\alpha)$) will be kept unchanged under different functions of ζ. Therefore if the data are fully observed, the estimate based on the transformed data is the same as the bivariate empirical estimator. However, if the data are censored or truncated, different transformation function ζ will give a different result. This is because the transformed indicator $\Delta_i(\alpha)$ may choose either δ_{1i} or δ_{2i} as its value, depending on the function ζ, and the relation of $\tilde{Z}_i(\alpha)$ and $V_i(\alpha)$ is also dependent of ζ. Therefore, we may transform the data in different ways. For example consider the following two sets of data, given by the scatter plots in Figure 1. It may be more appropriate to use a linear function (the linear line in the plot) as ζ, for the transformation of the first data set (left plot), and it may be more suitable to use a non-linear function ζ (the curve in the plot) for the transformation of the second data set (right plot).

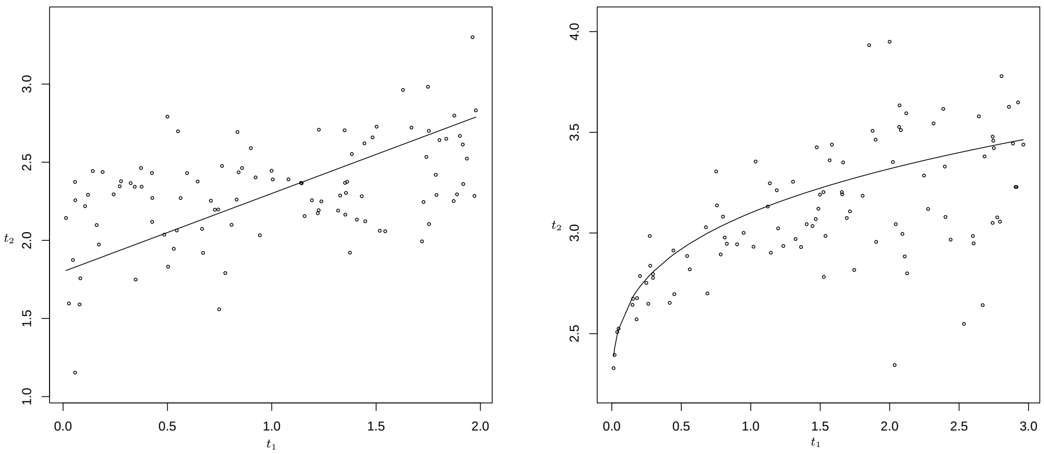

Figure 1. Using different data transformation function for different data sets.

In practice, we should select ζ which gives a small bias for $\hat{S}(z;\alpha) - S(z;\alpha)$, which is given in Theorem 2.4 by (9). Since the first term in (9) is a zero-mean martingale, we know that $E[\hat{S}(z;\alpha) - S(z;\alpha)] = E[B(z;\alpha)]$, where $B(z;\alpha)$ is given in (10). We should choose ζ to make $B(z;\alpha)$ as small as possible. Clearly we need to have a smaller value of $I[H_{(n)}(s-;\alpha) = 0]$ in order to have small $B(z;\alpha)$. This means that we should choose ζ to guarantee that the transformed data should have small possibilities of having $H_{(n)} = 0$.

Recall the definition of $H_n(s;\alpha)$ in (6). To make the bias smaller we have to assure that the number of observations, such that

$$\tilde{Z}_i(\alpha) > V_i(\alpha), \tag{12}$$

is as large as possible. Further recalling the definition of $\tilde{Z}_i(\alpha)$ and $V_i(\alpha)$ in (5), we need to choose ζ to guarantee that the number of observations with

$$\min\{\tilde{Y}_{1i}, \tilde{Y}_{2i}\} > \max\{L_{1i}, L_{2i}\} \tag{13}$$

is as large as possible. In other words we need to choose ζ which gives a large proportion of observations satisfying (13).

The above arguments may not be applied easily in practice. In practice, we may consider to use the following statistics

$$\mathbf{A} = \int\int \left[\hat{B}(s;\alpha)^2 + \hat{\sigma}^2(s;\alpha)\right] dsd\alpha \tag{14}$$

which can be viewed as the total mean square errors for all observed data points. Note that $\hat{B}$ is the estimated values for B, which can be obtained by replacing S and Λ by their consistent estimators in (10).In practice, when we compare several transformation functions ζ_k, which give total mse statistic $\mathbf{A}_k$ respectively, we should choose the function ζ_k which gives the smallest total mean square error $\mathbf{A}_k$.

The simulation studies in the following section confirm the above arguments.

4. Simulation studies and data analysis

4.1. *Simulation*

In this section we provide a simulation study to show the properties of the estimates based on different data-transformations and assess the performance of the proposed methods. We consider a scenario where data are generated from the model

$$T_1 \sim \text{Gamma}(2,1) \quad \text{and} \quad T_2 = a * \sqrt{T_1} + \varepsilon,$$

with $a = 0.5$ and $\varepsilon \sim \mathrm{N}(1, 0.5)$. We aim to estimate the joint survival functions of (T_1, T_2) via the proposed nonparametric approach, under censoring and truncation.

The censoring variables C_1 and C_2 are simulated via

$$C_1 = a_1\xi_1 + a_2\xi_2 \quad \text{and} \quad C_2 = b_1\xi_1 + b_2\xi_2, \tag{15}$$

where $\xi_1, \xi_2 \sim \exp(\beta)$ and $\beta = 0.02$. Moreover, we assume that truncation variables are given by

$$L_k = 0.05 \cdot (C_k - U[0, 1]), k = 1, 2.$$

Thus $L_k \leq C_k$ and they are correlated. Note that only observations with $L_k \leq T_k, k = 1, 2$ are recorded.

The values of a_1, a_2, b_1 and b_2 in (15) are chosen to achieve different censoring percentages and truncation probabilities. In our study the censoring percentages for both T_1 and T_2 are about 20% respectively and the truncation probability $P(L_1 < T_1, L_2 < T_2)$ is about 85%. We consider different sample sizes, $n = 100$, $n = 200$ and $n = 500$.

Tables 1, 2 and 3 summarize the simulation results, where the true $S(t_1, t_2)$, the mean estimates of our proposed estimator $\bar{S} = \sum_{k=1}^{m} \hat{S}_k/m$, the bias of $\hat{S}(t_1, t_2)$, the empirical standard errors of $\hat{S}(t_1, t_2)$ based on $m = 500$ simulations $\sqrt{\sum_{k=1}^{m}(\hat{S}_k - \bar{S})^2/(m-1)}$, the empirical means of standard errors $\sum_{k=1}^{m} \hat{\sigma}_k/m$, the mean squared error of $\hat{S}(t_1, t_2)$ and the proportion of $\tilde{Z}_i > V_i$ are respectively shown in rows (a), (b), (c), (d), (e), (f) and (g).

The distribution function estimators are evaluated at points (t_1, t_2) with values $(1, 1)$, $(2, 1)$, $(2, 2)$ and $(3, 2)$, respectively. Three data transformation functions have been considered: (i): $t_2 = a\sqrt{t_1}$; (ii): $t_2 = at_1$: (iii) $t_2 = at_1^2$.

When the sample size becomes smaller ($n = 500, n = 200, n = 100$) (Tables 1, 2 and 3), the choice of the ζ becomes more important. In particular $\widehat{S}(t_1, t_2)$ will have a larger bias, if ζ gives a smaller probability of $\{\tilde{Z}_i(\alpha) > V_i(\alpha)\}$. For example from Table 3, we can see that (iii) has probability of $\{\tilde{Z}_i(\alpha) > V_i(\alpha)\}$ uniformly smaller than (i) and (ii) and in the mean time the estimates of (iii) have larger bias and larger mean squared error. We also noticed that with $n = 200$, the standard error estimate of (iii) is much larger than (i) and (ii) and the standard error estimate of (iii) is not good enough (not close to the Monte Carlo standard error).

Table 1. Simulation study. Sample size: $n = 500$. (a): true $S(t_1, t_2)$; (b): empirical mean of $\widehat{S}(t_1, t_2)$; (c): the bias of $\widehat{S}(t_1, t_2)$; (d): empirical SE of $\widehat{S}(t_1, t_2)$; (e): empirical mean of estimated SE of $\widehat{S}(t_1, t_2)$; (f): the mean squared error of $\widehat{S}(t_1, t_2)$; (g): $P(\tilde{Z}_i > V_i)$ for selected time pairs (t_1, t_2) for three data transformations: (i): $T_2 = a\sqrt{T_1}$; (ii): $T_2 = aT_1$; (iii): $T_2 = aT_1^2$.

	(1,1) (i)	(1,1) (ii)	(1,1) (iii)	(2,1) (i)	(2,1) (ii)	(2,1) (iii)
(a)	0.6809	0.6809	0.6809	0.3908	0.3908	0.3908
(b)	0.6789	0.6783	0.6788	0.3906	0.3902	0.3936
(c)	-0.0021	-0.0027	-0.0021	-0.0002	-0.0006	0.0028
(d)	0.0248	0.0260	0.0395	0.0254	0.0280	0.0431
(e)	0.0251	0.0257	0.0331	0.0250	0.0269	0.0383
(f)	0.0006	0.0007	0.0016	0.0006	0.0008	0.0019
(g)	0.9800	0.9660	0.9140	0.9100	0.8560	0.7400
	(2,2) (i)	**(2,2) (ii)**	**(2,2) (iii)**	**(3,2) (i)**	**(3,2) (ii)**	**(3,2) (iii)**
(a)	0.1726	0.1726	0.1726	0.1025	0.1025	0.1025
(b)	0.1710	0.1713	0.1697	0.1022	0.1024	0.1027
(c)	-0.0017	-0.0014	-0.0029	-0.0004	-0.0001	0.0002
(d)	0.0198	0.0196	0.0234	0.0176	0.0180	0.0210
(e)	0.0196	0.0195	0.0214	0.0162	0.0164	0.0185
(f)	0.0004	0.0004	0.0006	0.0003	0.0003	0.0004
(g)	0.9700	0.9660	0.8660	0.9740	0.9340	0.7420

Moreover looking at the Table 3 for $n = 100$, we observed that although bias and mean squared error become worse for each transformation, they are always greater under the transformation (ii) and (iii) comparing to that under transformation (i). In other words, when we move from the "true relation" between T_1 and T_2, the results become unstable, especially for very small sample sizes. This is also confirmed by the statistic $\mathbf{A}$ in (14).

Table 2. Simulation study. Sample size: $n = 200$. (a): true $S(t_1, t_2)$; (b): empirical mean of $\widehat{S}(t_1, t_2)$; (c): the bias of $\widehat{S}(t_1, t_2)$; (d): empirical SE of $\widehat{S}(t_1, t_2)$; (e): empirical mean of estimated SE of $\widehat{S}(t_1, t_2)$; (f): the mean squared error of $\widehat{S}(t_1, t_2)$; (g): $P(\tilde{Z}_i > V_i)$ for selected time pairs (t_1, t_2) for three data transformations: (i): $T_2 = a\sqrt{T_1}$; (ii): $T_2 = aT_1$; (iii): $T_2 = aT_1^2$.

		(1,1)			(2,1)	
	(i)	(ii)	(iii)	(i)	(ii)	(iii)
(a)	0.6809	0.6809	0.6809	0.3908	0.3908	0.3908
(b)	0.6796	0.6791	0.6773	0.3869	0.3865	0.3819
(c)	-0.0014	-0.0018	-0.0036	-0.0039	-0.0043	-0.0089
(d)	0.0387	0.0400	0.0670	0.0367	0.0383	0.0786
(e)	0.0386	0.0398	0.0493	0.0394	0.0420	0.0563
(f)	0.0015	0.0016	0.0045	0.0014	0.0015	0.0063
(g)	0.9800	0.9450	0.9200	0.9400	0.8650	0.7500
		(2,2)			(3,2)	
	(i)	(ii)	(iii)	(i)	(ii)	(iii)
(a)	0.1726	0.1726	0.1726	0.1025	0.1025	0.1025
(b)	0.1728	0.1726	0.1735	0.1033	0.1031	0.1036
(c)	0.0002	0.0000	0.0008	0.0008	0.0005	0.0011
(d)	0.0318	0.0311	0.0381	0.0255	0.0257	0.0309
(e)	0.0306	0.0305	0.0331	0.0253	0.0255	0.0282
(f)	0.0010	0.0010	0.0015	0.0007	0.0007	0.0010
(g)	0.9700	0.9700	0.8650	0.9850	0.8800	0.7250

With 20% censoring and 80% truncation, the statistics **A** (total mean square errors) based on different tranformation function ζ are shown in Table 4. We can see that for all sample size under consideration, the square root transformation (the true transformation) gives the smallest total mse. These findings imply that although the choice of ζ is very arbitrary, for sample size which is not very large, we should choose a ζ which can give a small value of **A**.

In practice, we may choose ζ by inspecting the possible parametric relationship between T_1 and T_2. For example, we can check the scatter plots (see Figure 1) and a good ζ should be the one which gives a better fit for the relation of T_1 and T_2. Therefore, our method is actually a nonparametric method, but makes full use of certain parametric information about the relation between T_1 and T_2.

Table 3. Simulation study. Sample size: $n = 100$. (a): true $S(t_1, t_2)$; (b): empirical mean of $\widehat{S}(t_1, t_2)$; (c): the bias of $\widehat{S}(t_1, t_2)$; (d): empirical SE of $\widehat{S}(t_1, t_2)$; (e): empirical mean of estimated SE of $\widehat{S}(t_1, t_2)$; (f): the mean squared error of $\widehat{S}(t_1, t_2)$; (g): $P(\tilde{Z}_i > V_i)$ for selected time pairs (t_1, t_2) for three data transformations: (i): $T_2 = a\sqrt{T_1}$; (ii): $T_2 = aT_1$; (iii): $T_2 = aT_1^2$.

		(1,1)			(2,1)	
	(i)	(ii)	(iii)	(i)	(ii)	(iii)
(a)	0.6809	0.6809	0.6809	0.3908	0.3908	0.3908
(b)	0.6778	0.6768	0.6743	0.3891	0.3913	0.3921
(c)	-0.0031	-0.0042	-0.0067	-0.0017	0.0005	0.0013
(d)	0.0567	0.0574	0.0894	0.0577	0.0625	0.0887
(e)	0.0537	0.0555	0.0656	0.0550	0.0586	0.0754
(f)	0.0032	0.0033	0.0080	0.0033	0.0039	0.0079
(g)	0.9900	0.9800	0.9400	0.9200	0.8600	0.7500
		(2,2)			(3,2)	
	(i)	(ii)	(iii)	(i)	(ii)	(iii)
(a)	0.1726	0.1726	0.1726	0.1025	0.1025	0.1025
(b)	0.1704	0.1695	0.1699	0.1001	0.0994	0.0994
(c)	-0.0022	-0.0032	-0.0028	-0.0025	-0.0032	-0.0031
(d)	0.0438	0.0426	0.0506	0.0388	0.0389	0.0453
(e)	0.0423	0.0421	0.0447	0.0338	0.0340	0.0371
(f)	0.0019	0.0018	0.0026	0.0015	0.0015	0.0021
(g)	0.9800	0.9800	0.8900	0.9700	0.8900	0.7400

Table 4. Comparision of different ζ

	$n = 500$	$n = 200$	$n = 100$	$n = 50$
A with $t_2 = a\sqrt{t_1}$	0.973	0.934	0.899	0.867
A with $t_2 = at_1$	1.041	0.997	0.959	0.925
A with $t_2 = at_1^2$	1.826	1.600	1.410	1.218

4.2. *Simulation studies under different truncation probabilities*

In this section we show the effect of truncation percentage on the estimation of the bivariate survival function $\hat{S}(t_1, t_2)$. The scenario considered is the same as that illustrated in Section 4.1, except that truncation probability $P(L_1 < T_1, L_2 < T_2)$ is chosen to be about 50%. We fixed sample sizes at $n = 100$, $n = 200$ and $n = 500$.

Tables 5 to 7 summarize the simulation results for the proportion of truncation equal to about 50%, where the true $S(t_1, t_2)$, the mean estimates of our proposed estimator $\bar{S} = \sum_{k=1}^{m} \hat{S}_k/m$, the bias of $\hat{S}(t_1, t_2)$, the empirical standard errors of $\hat{S}(t_1, t_2)$ based on $m = 500$ simulations $\sqrt{\sum_{k=1}^{m}(\hat{S}_k - \bar{S})^2/(m-1)}$, the empirical means of standard errors $\sum_{k=1}^{m} \hat{\sigma}_k/m$, the mean squared error of $\hat{S}(t_1, t_2)$ and the proportion of $\tilde{Z}_i > V_i$ are respectively shown in rows (a), (b), (c), (d), (e), (f) and (g).

Compared the Table 7 and the Table 1 (large sample size $n = 500$), we have almost similar findings in terms of the mean square errors, i.e. using the true relation as the transformation function, the estimate will give smaller mean square errors. Note that when sample size becomes smaller $n = 200$, mean square errors in 6 do not confirm the "true" relation between T_1 and T_2. This is reasonable, as we need more samples when the data are severely biased. Also the results in Tables 5 - 7 have much larger bias and mean square errors, comparing to those in Tables 3-1, because the truncation probability is smaller (data are more biased).

Table 5. Simulation study. Sample size: $n = 100$. (a): true $S(t_1, t_2)$; (b): empirical mean of $\widehat{S}(t_1, t_2)$; (c): the bias of $\widehat{S}(t_1, t_2)$; (d): empirical SE of $\widehat{S}(t_1, t_2)$; (e): empirical mean of estimated SE of $\widehat{S}(t_1, t_2)$; (f): the mean squared error of $\widehat{S}(t_1, t_2)$; (g): $P(\tilde{Z}_i > V_i)$ for selected time pairs (t_1, t_2) for three data transformations: (i): $T_2 = a\sqrt{T_1}$; (ii): $T_2 = aT_1$; (iii): $T_2 = aT_1^2$.

		(1,1)			(2,1)	
	(i)	(ii)	(iii)	(i)	(ii)	(iii)
(a)	0.6809	0.6809	0.6809	0.3908	0.3908	0.3908
(b)	0.6845	0.6843	0.6875	0.3887	0.3923	0.4137
(c)	0.0036	0.0034	0.0066	-0.0021	0.0015	0.0229
(d)	0.0998	0.0869	0.1226	0.0743	0.0897	0.1459
(e)	0.0812	0.0779	0.0856	0.0677	0.0786	0.1024
(f)	0.0100	0.0076	0.0151	0.0055	0.0080	0.0218
(g)	0.9400	0.9400	0.8000	0.7400	0.7500	0.6600
		(2,2)			(3,2)	
	(i)	(ii)	(iii)	(i)	(ii)	(iii)
(a)	0.1726	0.1726	0.1726	0.1025	0.1025	0.1025
(b)	0.1710	0.1702	0.1710	0.1011	0.1012	0.1071
(c)	-0.0016	-0.0025	-0.0016	-0.0014	-0.0013	0.0046
(d)	0.0510	0.0425	0.0543	0.0333	0.0332	0.0514
(e)	0.0447	0.0398	0.0454	0.0301	0.0309	0.0402
(f)	0.0026	0.0018	0.0029	0.0011	0.0011	0.0027
(g)	0.7800	0.9000	0.8100	0.9700	0.8900	0.7100

4.3. *Data Analysis*

We apply our proposed method to analyze the probability of failure for a sample of 420 Italian firms, which was collected from the Amadeus Database, provided by Bureau

Table 6. Simulation study. Sample size: $n = 200$. (a): true $S(t_1, t_2)$; (b): empirical mean of $\widehat{S}(t_1, t_2)$; (c): the bias of $\widehat{S}(t_1, t_2)$; (d): empirical SE of $\widehat{S}(t_1, t_2)$; (e): empirical mean of estimated SE of $\widehat{S}(t_1, t_2)$; (f): the mean squared error of $\widehat{S}(t_1, t_2)$; (g): $P(\tilde{Z}_i > V_i)$ for selected time pairs (t_1, t_2) for three data transformations: (i): $T_2 = a\sqrt{T_1}$; (ii): $T_2 = aT_1$; (iii): $T_2 = aT_1^2$.

		(1,1)			(2,1)	
	(i)	(ii)	(iii)	(i)	(ii)	(iii)
(a)	0.6809	0.6809	0.6809	0.3908	0.3908	0.3908
(b)	0.6839	0.6809	0.6825	0.3898	0.3894	0.3985
(c)	0.0030	0.0000	0.0016	-0.0010	-0.0014	0.0077
(d)	0.0640	0.0623	0.1001	0.0516	0.0602	0.1133
(e)	0.0615	0.0583	0.0671	0.0496	0.0575	0.0823
(f)	0.0041	0.0039	0.0100	0.0027	0.0036	0.0129
(g)	0.9500	0.9600	0.8400	0.7400	0.7000	0.6750
		(2,2)			(3,2)	
	(i)	(ii)	(iii)	(i)	(ii)	(iii)
(a)	0.1726	0.1726	0.1726	0.1025	0.1025	0.1025
(b)	0.1711	0.1709	0.1721	0.1008	0.1002	0.1020
(c)	-0.0016	-0.0017	-0.0005	-0.0017	-0.0023	-0.0005
(d)	0.0349	0.0287	0.0398	0.0222	0.0233	0.0381
(e)	0.0328	0.0290	0.0337	0.0218	0.0222	0.0298
(f)	0.0012	0.0008	0.0016	0.0005	0.0005	0.0015
(g)	0.8150	0.9450	0.8600	0.9200	0.8900	0.6250

van Dijk. Following Altman (1968), the interest is in predicting the firms' potential ending up in financial distress and studying the relationship between the financial status of a company (acquisition, bankruptcy, liquidation, merger and so on) and its probability of failure. Many theoretical and empirical studies have been conducted to analyze and predict the occurrence of the business insolvency, by means of statistical techniques (i.e. discriminant analysis, logit and probit regressions, survival analysis) (for further details see Balcaen and Ooghe (2006), Gepp and Kumar (2012)). When survival analysis is applied in this context (see for example Gepp and Kumar (2008), Luoma and Laitinen (1991)), right censoring and truncation have been considered in only few papers (Bhattacharjee et al. 2009) in the univariate case.

It is expected that for all companies which entered this study, new businesses are more likely to be bankrupted during a crisis than well-established businesses. In other words the older is the firm, the smaller is the probability of bankruptcy and consequently bigger is the probability of being in activity. This motivates us to concentrate on two events:

Table 7. Simulation study. Sample size: $n = 500$. (a): true $S(t_1, t_2)$; (b): empirical mean of $\widehat{S}(t_1, t_2)$; (c): the bias of $\widehat{S}(t_1, t_2)$; (d): empirical SE of $\widehat{S}(t_1, t_2)$; (e): empirical mean of estimated SE of $\widehat{S}(t_1, t_2)$; (f): the mean squared error of $\widehat{S}(t_1, t_2)$; (g): $P(\tilde{Z}_i > V_i)$ for selected time pairs (t_1, t_2) for three data transformations: (i): $T_2 = a\sqrt{T_1}$; (ii): $T_2 = aT_1$; (iii): $T_2 = aT_1^2$.

		(1,1)			(2,1)	
	(i)	(ii)	(iii)	(i)	(ii)	(iii)
(a)	0.6809	0.6809	0.6809	0.3908	0.3908	0.3908
(b)	0.6820	0.6812	0.6847	0.3909	0.3885	0.3903
(c)	0.0011	0.0003	0.0038	0.0001	-0.0023	-0.0005
(d)	0.0441	0.0382	0.0596	0.0320	0.0425	0.0773
(e)	0.0408	0.0381	0.0481	0.0318	0.0368	0.0597
(f)	0.0019	0.0015	0.0036	0.0010	0.0018	0.0060
(g)	0.9260	0.9460	0.7700	0.7320	0.7100	0.6420
		(2,2)			(3,2)	
	(i)	(ii)	(iii)	(i)	(ii)	(iii)
(a)	0.1726	0.1726	0.1726	0.1025	0.1025	0.1025
(b)	0.1711	0.1712	0.1727	0.1044	0.1039	0.1059
(c)	-0.0015	-0.0014	0.0000	0.0019	0.0013	0.0033
(d)	0.0221	0.0191	0.0282	0.0148	0.0155	0.0240
(e)	0.0210	0.0184	0.0227	0.0143	0.0146	0.0211
(f)	0.0005	0.0004	0.0008	0.0002	0.0002	0.0006
(g)	0.8500	0.9420	0.8420	0.9180	0.8920	0.6460

the time period, T_1, from the establishment of a firm to recruitment time point, and T_2, from recruitment to bankruptcy. These two events are correlated. The time T_1 is usually truncated by a random variable L_1, the time period from the establishment of a firm to year 2004, such that only firms with $T_1 \geq L_1$ can be observed. In the mean time, T_1 and T_2 are subject to random censoring by certain random variables C_1 and C_2 respectively.

We estimate $S(t_1, t_2)$ via two different transformation functions $t_2 = at_1$ and $t_2 = a\sqrt{t_1}$, the results of which are given in Table 8 and Table 9, respectively. The estimates shown in Table 8 has smaller standard error estimates than that in Table 9. Also the total mse $\mathbf{A} = 39.8$ with linear tranformation is much smaller than the total mse $\mathbf{A} = 106$ with square root tranformation. Therefore, the transformation $t_2 = at_1$ is recommended. Table 8 and Table 9 present the time period for t_2 in the interval $[8.0, 8.5]$, which represents the period of four years after the financial crisis. From the result in Table 8 we find that a larger value of T_1 will give a smaller value of $P(T_2 \in [8.0, 8.5])$ (when T_1 increases from 12.57 to 16.57, $P(T_2 \in [8.0, 8.5])$ goes from 0.0237 down to 0.128), which indeed implies

Table 8. Business failure data. $\hat{S}(t_1, t_2)$ at the selected time pairs (t_1, t_2) their estimated standard error (in parentheses), when the data transformation is $t_2 = at_1$

		t_2							
		8.00	8.07	8.14	8.21	8.29	8.36	8.43	8.50
t_1	12.57	0.3845	0.3837	0.3825	0.3825	0.3825	0.3778	0.3758	0.3608
		(0.0538)	(0.0538)	(0.0539)	(0.0544)	(0.0523)	(0.0521)	(0.0522)	(0.0510)
		[0.4548]	[0.4548]	[0.4548]	[0.4476]	[0.4452]	[0.4405]	[0.4405]	[0.4405]
	13.71	0.3827	0.3727	0.3712	0.3712	0.3592	0.3476	0.3451	0.3394
		(0.0527)	(0.0523)	(0.0523)	(0.0523)	(0.0508)	(0.0500)	(0.0496)	(0.0492)
		[0.4714]	[0.4714]	[0.4714]	[0.4667]	[0.4619]	[0.4571]	[0.4571]	[0.4571]
	14.86	0.3438	0.3438	0.3438	0.3438	0.3376	0.3280	0.3244	0.3163
		(0.0486)	(0.0490)	(0.0494)	(0.0494)	(0.0485)	(0.0471)	(0.0468)	(0.0460)
		[0.5095]	[0.5071]	[0.5048]	[0.5024]	[0.5024]	[0.4952]	[0.4905]	[0.4857]
	15.43	0.3416	0.3416	0.3416	0.3416	0.3343	0.3223	0.3223	0.3163
		(0.0481)	(0.0487)	(0.0487)	(0.0487)	(0.0478)	(0.0467)	(0.0472)	(0.0467)
		[0.5333]	[0.5214]	[0.5190]	[0.5143]	[0.5095]	[0.5071]	[0.5048]	[0.5024]
	16.57	0.2896	0.2884	0.2884	0.2884	0.2834	0.2768	0.2768	0.2768
		(0.0429)	(0.0429)	(0.0433)	(0.0433)	(0.0423)	(0.0418)	(0.0421)	(0.0421)
		[0.5738]	[0.5690]	[0.5619]	[0.5595]	[0.5595]	[0.5548]	[0.5452]	[0.5381]

Table 9. Business failure data. $\hat{S}(t_1, t_2)$ at the selected time pairs (t_1, t_2), their estimated standard error (in parentheses), and $P(\tilde{Z}_i(\alpha) > V_i(\alpha))$ (in square brackets) when the data transformation is $t_2 = a\sqrt{t_1}$

		t_2							
		8.00	8.07	8.14	8.21	8.29	8.36	8.43	8.50
t_1	12.57	0.3331	0.3304	0.3304	0.3304	0.3150	0.2995	0.2995	0.2851
		(0.0987)	(0.0981)	(0.0986)	(0.0990)	(0.0953)	(0.0923)	(0.0928)	(0.0899)
		[0.4571]	[0.4452]	[0.4405]	[0.4333]	[0.4262]	[0.4167]	[0.4119]	[0.4095]
	13.71	0.3331	0.3304	0.3304	0.3150	0.3048	0.2938	0.2910	0.2851
		(0.081)	(0.0812)	(0.0811)	(0.0857)	(0.0831)	(0.0878)	(0.0872)	(0.0866)
		[0.4952]	[0.4810]	[0.4786]	[0.4667]	[0.4595]	[0.4571]	[0.4452]	[0.4405]
	14.86	0.3331	0.3304	0.3304	0.3150	0.3048	0.2938	0.2910	0.2715
		(0.0753)	(0.0745)	(0.0744)	(0.0743)	(0.0722)	(0.0715)	(0.0716)	(0.0747)
		[0.5333]	[0.5214]	[0.5167]	[0.5095]	[0.5048]	[0.4929]	[0.4810]	[0.4786]
	15.43	0.3331	0.3304	0.3304	0.3150	0.3048	0.2938	0.2910	0.2715
		(0.0775)	(0.0761)	(0.0753)	(0.0752)	(0.0726)	(0.0699)	(0.0697)	(0.0706)
		[0.5452]	[0.5381]	[0.5333]	[0.5262]	[0.5167]	[0.5119]	[0.5071]	[0.4952]
	16.57	0.3261	0.3261	0.3261	0.3150	0.3048	0.2863	0.2822	0.2715
		(0.0706)	(0.0725)	(0.0713)	(0.0699)	(0.0680)	(0.0648)	(0.0641)	(0.0640)
		[0.5738]	[0.5643]	[0.5548]	[0.5500]	[0.5452]	[0.5381]	[0.5333]	[0.5262]

that if a company is older, it has a smaller probability to fail during the crisis. Such a conclusion cannot be drawn if we use the transformation function $t_2 = a\sqrt{t_1}$, which is not appropriate. The plot of the estimated joint survival function is shown in Figure 2.

We can also estimate the truncation probability via the method in Shen (2006), or Dai and Fu (2012). Specifically, the truncation probability $\gamma = P(L_1 \leq T_1, L_2 \leq T_2)$ can be estimated by

$$\hat{\gamma} = \left[n^{-1} \sum_{i=1}^{n} \frac{1}{\hat{S}(L_{1i}-, L_{2i}-)} \right]^{-1},$$

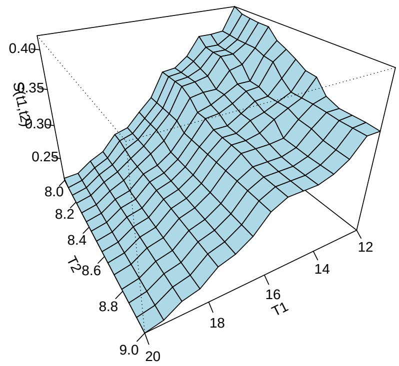

Figure 2. The bivariate survival function estimate for (T_1, T_2).

where $\hat{S}(L_{1i}-, L_{2i}-)$ is the left-continuous version of $\hat{S}(L_{1i}, L_{2i})$, which is our estimator evaluated at (L_{1i}, L_{2i}) for the ith observation. Here for our data of 420 Italian firms, the estimated truncation probability is $\hat{\gamma} = 0.21$.

5. Discussion

This paper proposes a class of nonparametric estimators based on a general data transformation for the bivariate survival function estimation. The large sample properties of the estimators have been provided. The performances of the estimators using different data transformation functions are compared via simulation studies. The proposed class of estimators is nonparametric, however it makes use of certain parametric information for the pair of random variables via a data transformation function. The estimator is equivalent to a univariate Kaplan-Meier estimator on the selected transformation function ζ. The non-uniqueness of ζ can also be interpreted by the fact that there is no unique partial order for the observations (under censoring and truncation) in the two-dimensional space. Such problems are equivalent to the challenges of martingales on the plane, where there is no unique order in $\mathbb{R}^{+2}$ (Merzbach and Nualart 1988). It is of interests to further study how to find the best transformation function ζ, if this is possible. We leave this as a future work.

Appendix A. Proof of Lemma 2.1

Proof. Given (t_1, t_2), the event $z \leq Z(\alpha) < z + dz$ is the same as $\{T_1 \in dt_1, T_2 > t_2\} \cup \{T_1 > t_1, T_2 \in dt_2\} \cup \{T_1 \in dt_1, T_2 \in dt_2\}$. Therefore

$$\begin{aligned}\Lambda(dz) &= \frac{P(z \leq Z(\alpha) < z + dz)}{P(Z(\alpha) \geq z)} \\ &= \frac{P(T_1 \in dt_1, T_2 > t_2) + P(T_1 > t_1, T_2 \in dt_2) + P(T_1 \in dt_1, T_2 \in dt_2)}{P(T_1 \geq t_1, T_2 \geq t_2)}. \qquad \text{(A1)}\end{aligned}$$

On the other hand Using the facts $\{z > V_i(\alpha)\} \Leftrightarrow \{t_1 > L_{1i}, t_2 > L_{2i}\}$ we have

$$
\begin{aligned}
&\frac{P(\tilde{Z}_i(\alpha) \in dz, z > V_i(\alpha), \Delta_i(\alpha) = 1)}{P(\tilde{Z}_i(\alpha) \geq z > V_i(\alpha))} \\
=& \frac{P(\tilde{Z}_i(\alpha) \in dz, z > V_i(\alpha), \Delta_i(\alpha) = 1 | L_{1i} \leq T_{1i}, L_{2i} \leq T_{2i})}{P(\tilde{Z}_i(\alpha) \geq z > V_i(\alpha) | L_{1i} \leq T_{1i}, L_{2i} \leq T_{2i})} \\
=& \frac{P(\tilde{Z}(\alpha) \in dz, z > V(\alpha), \Delta(\alpha) = 1)}{P(\tilde{Z}(\alpha) \geq z > V(\alpha)}
\end{aligned} \tag{A2}
$$

Now we consider the set $\{\tilde{Z}(\alpha) \in dz, z > V(\alpha), \Delta(\alpha) = 1\}$ in (A2). The definition in (5) indicates that

$$
\{\Delta(\alpha) = 1\} = \{\delta_1 = 1, \tilde{Y}_1 \leq \tilde{Y}_2\} \cup \{\delta_2 = 1, \tilde{Y}_1 \geq \tilde{Y}_2\}
$$

and because of $\{\delta_1 = 1, \tilde{Y}_1 \leq \tilde{Y}_2\} \cap \{\delta_2 = 1, \tilde{Y}_1 \geq \tilde{Y}_2\} = \{\min\{\delta_1, \delta_2\} = 1, \tilde{Y}_1 = \tilde{Y}_2\}$ we further have

$$
\begin{aligned}
& P(\tilde{Z}(\alpha) \in dz, z > V(\alpha), \Delta(\alpha) = 1) \\
=& P(\tilde{Z}(\alpha) \in dz, z > V(\alpha), \delta_1 = 1, \tilde{Y}_1 \leq \tilde{Y}_2) + P(\tilde{Z}(\alpha) \in dz, z > V(\alpha), \delta_2 = 1, \tilde{Y}_1 \geq \tilde{Y}_2) \\
& - P(\tilde{Z}(\alpha) \in dz, z > V(\alpha), \min\{\delta_1, \delta_2\} = 1, \tilde{Y}_1 = \tilde{Y}_2).
\end{aligned} \tag{A3}
$$

For the three sets in (A3), we have

$$
\begin{aligned}
& \{\tilde{Z}(\alpha) \in dz, z > V(\alpha), \delta_1 = 1, \tilde{Y}_1 \leq \tilde{Y}_2\} \Leftrightarrow \{T_1 \in dt_1, C_1 \geq t_1 > L_1, T_2 \geq t_2, C_2 \geq t_2 > L_2\}, \\
& \{\tilde{Z}(\alpha) \in dz, z > V(\alpha), \delta_2 = 1, \tilde{Y}_1 \geq \tilde{Y}_2\} \Leftrightarrow \{T_1 \geq t_1, C_1 \geq t_1 > L_1, T_2 \in dt_2, C_2 \geq t_2 > L_2\}, \\
& \{\tilde{Z}(\alpha) \in dz, z > V(\alpha), \min(\delta_1, \delta_2) = 1, \tilde{Y}_1 = \tilde{Y}_2\} \\
& \Leftrightarrow \{T_1 \in dt_1, T_2 \in dt_2, C_1 \geq t_1 > L_1, C_2 \geq t_2 > L_2\}.
\end{aligned} \tag{A4}
$$

Now (A3) and (A4) together imply that (A2) can be rewritten as

$$
\begin{aligned}
&\frac{P(\tilde{Z}_i(\alpha) \in dz, z > V_i(\alpha), \Delta_i(\alpha) = 1)}{P(\tilde{Z}_i(\alpha) \geq z > V_i(\alpha))} \\
= &\frac{P(T_1 \in dt_1, T_2 \geq t_2, C_1 \geq t_1 > L_1, C_2 \geq t_2 > L_2)}{P(T_1 \geq t_1, T_2 \geq t_2, C_1 \geq t_1 > L_1, C_2 \geq t_2 > L_2)} \\
&+ \frac{P(T_1 \geq t_1, T_2 \in dt_2, C_1 \geq t_1 > L_1, C_2 \geq t_2 > L_2)}{P(T_1 \geq t_1, T_2 \geq t_2, C_1 \geq t_1 > L_1, C_2 \geq t_2 > L_2)} \\
&- \frac{P(T_1 \in dt_1, T_2 \in dt_2, C_1 \geq t_1 > L_1, C_2 \geq t_2 > L_2)}{P(T_1 \geq t_1, T_2 \geq t_2, C_1 \geq t_1 > L_1, C_2 \geq t_2 > L_2)} \\
= &\frac{P(T_1 \in dt_1, T_2 \geq t_2) + P(T_1 \geq t_1, T_2 \in dt_2) - P(T_1 \in dt_1, T_2 \in dt_2)}{P(T_1 \geq t_1, T_2 \geq t_2)} \\
= &\frac{P(T_1 \in dt_1, T_2 > t_2) + P(T_1 > t_1, T_2 \in dt_2) + P(T_1 \in dt_1, T_2 \in dt_2)}{P(T_1 \geq t_1, T_2 \geq t_2)}
\end{aligned}
$$

which, from (A1), is $\Lambda(dz;\alpha)$. ■

References

Altam, E.I. (1968), "Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy", *Journal of Finance*, **23**: 589–609.

Balcaen, S., and Ooghe, H. (2006), "35 years of studies on business failure: an overview of the classic statistical methodologies and their related problems", *The British Accounting Review*, **38**, 63–93.

Bhattacharjee A., Higson C., Holly S., and Kattuman P. (2009), "Macroeconomic instability and business exit: determinants of failures and acquisitions of UK firms", *Economica*, **76**: 108–131.

Dai, H., and Bao, Y. (2009), "An inverse probability weighted estimator for the bivariate distribution function under right censoring", *Statistics and Probability Letters*, **79**:1789–1797.

Dai, H., and Fu, B. (2012), "A polar coordinate transformation for estimating bivariate survival functions with randomly censored and truncated data", *Journal of Statistical Planning and Inference*, **142**:248–262.

Gepp, A., and Kumar, K. (2008), "The Role of Survival Analysis in Financial Distress Prediction", *International Research Journal of Finance and Economics*, **16**, 13–34.

Gepp, A., and Kumar, K. (2012), *Business failure prediction using statistical techniques: A review*, in K. Kumar and A. Chaturvedi (Eds.). *Some Recent Developments in Statistical Theory and Applications*, pp 1-25. Boca Raton, Florida, USA: Brown Walker Press.

Gijbels, I., and Gürler, Ü. (1998), "Covariance function of a bivariate distribution function estimator for left truncated and right censored data", *Statistica Sinica*, **8**:1219–1232.

Gürler, Ü. (1997), "Bivariate distribution and hazard functions when a component is ramdoly truncated", *Journal of Multivariate Analysis*, **60**:20–47.

Huang, J., Vieland, V.J., and Wang, K. (2001), "Nonparametric estimation of marginal distributions under bivariate truncation with application to testing for age-of-onset anticipation", *Statistica Sinica*, **11**:1047–1068.

Keiding, N., and Gill, R. D. (1990), "Random Truncation Models and Markov Processes", *The Annals of Statistics*, **18**:582–602.

Luoma, M., and Laitinen, E. (1991), "Survival analysis as a tool for company failure prediction", *Omega*, **19**(6), 673–678.

Merzbach, E. and Nualart, D. (1988), "A Martingale Approach to Point Processes in The Plane", *The Annals of Probability*, **16**(1), 265–274.

Sankaran, P.G. and Antony A.A. (2007), "Bivariate Competing Risks Models Under Random Left Truncation and Right Censoring", *Sankhya: The Indian Journal of Statistics*, **69**(3):425-447.

Shen, P. (2006), "An inverse-probability-weighted approach to estimation of the bivariate survival function under left-truncation and right-censoring", *Journal of Statistical Planning and Inference*, **136**:4365-4384.

Shen, P.S. (2007), "A general semiparametric model for left truncated and right-censored data", *Journal of Nonparametric Statistics*, **19**:113-129.

Shen, P.S. (2014), "Simple nonparametri estimators of the bivariate survival function under random left truncation and right censoring", *Computational Statistics*, **29**:641-659.

Shen, P.S., and Yan, Y.F. (2008), "Nonparametric estimation of the bivariate survival function with left-truncated and right-censored data", *Journal of Statistical Planning and Inference*,

138:4041-4054.

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

van der Laan, M. J. (1996b), "Nonparametric estimation of the bivariate survival function with truncated data", *Journal of Multivariate Analysis*, **58**:107–131.

Wang, M.C. (1991), "Nonparametric Estimation from Cross-Sectional Survival Data", *Journal of the American Statistical Association*, **86**:130–143.

Woodroofe, M. (1985), "Estimating a distribution function with truncated data", *The Annals of Statistics*, **13**:163–177.