Linear transformation models for censored data under truncation

Wang, Huan; Dai, Hongsheng; Restaino, Marialuisa; Bao, Yanchun

Published in:
Journal of Statistical Planning and Inference

DOI:
10.1016/j.jspi.2017.07.006

Publication date:
2018

Document Version
Peer reviewed version

Linear transformation models for censored data under truncation

Huan Wang[a,*], Hongsheng Dai[b], Marialuisa Restaino[c], Yanchun Bao[d]

[a] *Dundee Epidemiology and Biostatistics Unit, University of Dundee, U.K.*
[b] *Department of Mathematical Sciences, University of Essex, U.K.*
[c] *Department of Economic and Statistics, University of Salerno, Italy*
[d] *Institute for Social and Economic Research, Univeristy of Essex, U.K.*

Abstract

In many observational cohort studies, a pair of correlated event times are usually observed for each individual. This paper develops a new approach for the semiparametric linear transformation model to handle the bivariate survival data under both truncation and censoring. By incorporating truncation, the potential referral bias in practice is taken into account. A class of generalised estimating equations are proposed to obtain unbiased estimates of the regression parameters. Large sample properties of the proposed estimator are provided. Simulation studies under different scenarios and analyses of real-world datasets are conducted to assess the performance of the proposed estimator.

Keywords: Linear transformation model, bivariate survival function, truncation, censoring, survival analysis

1. Introduction

Bivariate survival data which contains pairs of correlated event times are often observed in many observational cohort studies. Incomplete information of the paired event times due to censoring and truncation leads to the challenge of analysing such bivariate survival data. Consider the following examples.

*Corresponding author
Email address: hwang@dundee.ac.uk (Huan Wang)

Example 1.1. In a hepatitis C cohort study (Fu et al., 2007) the epidemiological interest is to study progression to liver cirrhosis in patients with chronic hepatitis C. The paired event times (R, T) observed for each patient are the time from infection with hepatitis C virus (HCV) to recruitment to liver clinics and the time from HCV infection to development of cirrhosis, respectively. The cirrhosis time T is subject to right censoring by a random variable C, i.e. the last-follow up time. As only the data before the end of year 1999 are accessible, there is a random time L, from the time point of HCV infection to the end of year 1999, such that only patients who were referred to liver clinics before that $(R \leq L)$ can be included in the study cohort, i.e. R is right-truncated by L. This is an example of bivariate survival data where one component is right truncated and the other one is right censored. Table 1 summarises the two pairs of event times.

Table 1: Notations and descriptions of event times. (Only the patients with $R \leq L$ can be observed. For the observed patients, T is subject to right censoring by C.)

Notation	Description
R	Time from HCV infection to referral to liver clinic
L	Time from HCV infection to the end of study recruitment
T	Time from HCV infection to development of cirrhosis
C	Time from HCV infection to end of follow-up

Example 1.2. In a business failure data which includes 420 small and medium size Italian firms from Amadeus Database provided by Bureau van Dijk, a pair of event times (R, T) are collected for each firm, where R is the time period from establishment to the first financial statement available and T is the subsequent time period to bankruptcy (Figure 1). The database entry started from year

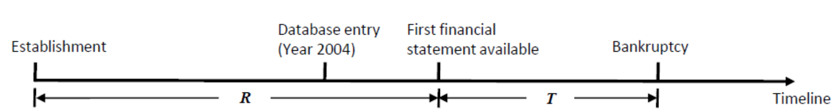

Figure 1: Structure of the business failure data in Example 1.2.

2004, therefore only firms who were recruited in the database after year 2004 could be observed. This introduces the left truncation for R, that is only the firms with $R \geq L$ can be observed, where L is the time period from establishment to database entry, which here is year 2004. The time T is subject to right censoring by the random variable C (e.g. the time from the first financial statement available to the last follow-up). This is an example of bivariate survival data under left truncation and right censoring.

In survival analysis, two widely used models to study the effect of covariates on the event time of interest are the proportional hazards (PH) model (Cox, 1972) and the accelerated failure time (AFT) model (Cox & Oakes, 1984). The AFT model has been well studied when the event time is subject to only right censoring. Its nonparametric version, $\log T = \boldsymbol{W}\boldsymbol{\beta} + \varepsilon$ with ε following an unknown distribution with mean 0, has also been studied in Wang et al. (2013) to handle bivariate survival data under both truncation and censoring. The PH model is even more widely used than the AFT model when assessing the effect of covariates. It can be generalised to the following semiparametric linear transformation model

$$h(T) = -\boldsymbol{W}\boldsymbol{\beta} + \varepsilon, \tag{1.1}$$

where $\boldsymbol{W}$ is the vector of covariates, $\boldsymbol{\beta}$ is the regression parameter to be estimated, and $h(\cdot)$ is a strictly increasing function which is completely unspecified. If the distribution function of ε is taken to be $F_\varepsilon(t) = 1 - \exp\{-\exp(t)\}$, (1.1) gives the famous PH model, while if F_ε is the standard logistic distribution function, (1.1) is the proportional odds model (Bennett, 1983; Dabrowska & Doksum, 1988; Pettitt, 1984; Murphy et al., 1997).

Under model (1.1), when T is subject to only right censoring, Cheng et al. (1995) proposed and justified a simple estimating equation for $\boldsymbol{\beta}$. Their estimator was further developed in Cheng et al. (1997); Fine et al. (1998); Cai et al. (2000). A key step in these approaches is estimating the survival function of the censoring variable by the Kaplan-Meier estimator. However, these univariate methods are not readily available when truncation is also incorporated, as the bivariate survival function of the truncation variable L and the censoring variable C needs to be estimated.

For bivariate survival function, Gürler (1996, 1997) proposed a nonparametric estimator when only a single component of the paired event times is subject to truncation. The case for doubly truncated data was discussed in van der Laan (1996) and Huang et al. (2001). Gijbels & Gürler (1998) considered the case where a single component of the paired event times is subject to both censoring and truncation but the other one can be fully observed. When both event times are under truncation and censoring, Shen (2006) proposed an inverse-probability-weighted (IPW) approach to estimate the bivariate survival function. Using similar idea, Shen & Yan (2008) generalised the approaches in Campbell & Földes (1982) and Dabrowska (1988, 1989) to estimate the bivariate survival function for left-truncated and right-censored data. However, their iteration algorithm is computationally heavy and relies on an assumption which may not be reasonable in practice. Dai & Fu (2012) proposed an nonparametric estimator for the bivariate survival function when both even times are subject to random truncation and censoring. Their method is based on a polar coordinate transformation which can transform the bivariate survival function to a univariate form without losing data information. The univariate survival function can be easily estimated by the product-limit estimator and can be transformed back to the bivariate form. Recently, their method was further extended to a class of bivariate survival function estimators based on different forms of data transformation (Dai et al., 2016).

In this paper, we develop a new approach for the semiparametric linear transformation model (1.1) to handle the bivariate survival data under both truncation and censoring. The bivariate survival function of (L, C) is estimated using the idea in Dai & Fu (2012). An unbiased estimating equation for $\boldsymbol{\beta}$ in model (1.1) is proposed. Our method is a new, flexible and important candidate for handling bivariate survival data with random truncation and censoring. It can also be extended to handle a general class of bivariate regression models with different types of censoring and truncation.

This paper is organised as follows. Section 2 describes the statistical models and introduces the estimating procedures. The large sample properties of the proposed estimator are established in Section 3. Simulation studies and analyses of real-world datasets are presented in Section 4 and 5 respectively to demonstrate the performance of the proposed estimator. Section 6 provides a brief discussion.

2. Statistical models and estimating procedure

2.1. Preliminaries

Denote (R, T) as the pair of event times, and $\boldsymbol{W}$ as the vector of covariates. For simplicity, here we focus on the case with right truncation and right censoring, i.e. R is right-truncated by L, and T is right-censored by C. For the case with left truncation we can simply replace R and L by $-R$ and $-L$ in practice, and our methodology still applies. In the presence of right truncation, since only individuals such that $R \leq L$ can be observed (opposite if under left truncation), we denote the observed data for the ith subject as $(R_i, L_i, X_i, \delta_i, \boldsymbol{W}_i)$, $i = 1, \ldots, n$, where $X_i = \min(T_i, C_i)$ and $\delta_i = I[T_i \leq C_i]$.

Throughout this paper, we assume that (L, C) are independent of the covariates vector $\boldsymbol{W}$ and are independent of the paired event times (R, T), similar to Wang et al. (2013). These assumptions are reasonable since in most retrospective studies the truncation time L (time to the end of recruitment) is determined independently before data collection, and also the censoring time C (last follow-up time) is usually a certain period of time after recruitment. Therefore (L, C) are not related to the individuals' information $\boldsymbol{W}$ and (R, T).

Let $G(t_1, t_2) = \mathbf{P}(L > t_1, C > t_2)$ be the continuous bivariate survival function for (L, C) and $\bar{F}(t_1, t_2) = \mathbf{P}(R \leq t_1, T \leq t_2)$ be the continuous joint distribution function for (R, T) with continuous support. We also assume the following conditions hold throughout this paper.

Condition 2.1. *The lower boundaries of support for $\bar{F}$ are coordinate axes of the first quadrant.*

Condition 2.2. *For* $\boldsymbol{t} = (t_1, t_2)$*, the function* $G(t_1, t_2) = \mathbf{P}(L > t_1, C > t_2) > 0$ *almost surely with respect to* $\bar{F}(dt_1, dt_2)$ *in* $\mathcal{A}$*, where* $\mathcal{A}$ *is the common support area of* $\bar{F}$ *and* G*.*

The two conditions are reasonable in practice. Condition 2.1 is a simple assumption which means $R \geq 0$ and $T \geq 0$, as we focus on event times with non-negative values. The explanations of Condition 2.2 for the hepatitis C data with right truncation (Example 1.1) can be found in Wang et al. (2013) or Dai & Wang (2016). Here we explain why Condition 2.2 is realistic for the case with left truncation (the business failure data described in Example 1.2).

Under left truncation, we have that the bivariate survival function $G(t_1, t_2) = \mathbf{P}(L < t_1, C > t_2)$. Condition 2.2 requires that: (i) the minimum and maximum values that R can take are greater than those of L, respectively. In the business failure data, it guarantees that the firms which are newly established (with very small R) or have been established for a long time (with very large L) can be collected. Condition 2.2 also requires that: (ii) the minimum and maximum values that T can take are smaller than those of C, respectively. This means that, in the business failure data, for a very short period of follow-up (very small C), it is always possible to have a firm which is bankrupt soon after submitting its first financial statement (i.e. observe even smaller $T \leq C$). In addition, for a firm with large value of T, it is possible to follow it long enough to observe its bankruptcy. In summary, Condition 2.2 guarantees that $\bar{F}$ and G can be identified in their common support region.

2.2. Estimating equation for β

Denote $\boldsymbol{\beta}^*$ as the true value of $\boldsymbol{\beta}$. When the indicators $\big\{I[T_i \geq T_j], i, j = 1, \ldots, n, i \neq j\big\}$ can be fully observed (i.e. no truncation or censoring), we have that

$$E\big\{I[T_i \geq T_j]\big|\boldsymbol{W}_i, \boldsymbol{W}_j\big\} = \mathbf{P}\big\{h(T_i) \geq h(T_j)]\big|\boldsymbol{W}_i, \boldsymbol{W}_j\big\} := \theta(\boldsymbol{W}_{ij}\boldsymbol{\beta}^*),$$

where $\boldsymbol{W}_{ij} = \boldsymbol{W}_i - \boldsymbol{W}_j$, $i, j = 1, \ldots, n$, $i \neq j$, and

$$\theta(t) = \int_{-\infty}^{\infty} [1 - F_\varepsilon(t + s)] F_\varepsilon(ds).$$

However in practice, the indicators $\{I[T_i \geq T_j], i, j = 1, \ldots, n, i \neq j\}$ may not be fully observed due to truncation and censoring. Let $\gamma = \mathbf{P}(R \leq L)$ be the truncation probability. Then we have that

$$\begin{aligned} & E\left\{\frac{\delta_j I[X_i \geq X_j]}{G(R_i-, X_j-)G(R_j-, X_j-)} - \frac{\theta(\boldsymbol{W}_{ij}\boldsymbol{\beta}^*)}{G(R_i-, 0)G(R_j-, 0)}\Big|\boldsymbol{W}_i, \boldsymbol{W}_j\right\} \\ &= \gamma^{-2} E\Big\{I[T_i \geq T_j] - \theta(\boldsymbol{W}_{ij}\boldsymbol{\beta}^*)\big|\boldsymbol{W}_i, \boldsymbol{W}_j\Big\} \\ &= 0. \end{aligned}$$

If the bivariate survival function G is known, an unbiased estimating equation for $\boldsymbol{\beta}$ is given by

$$\begin{aligned} \boldsymbol{U}(\boldsymbol{\beta}; G) = n^{-2} \sum_{i=1}^{n} \sum_{j=1}^{n} \boldsymbol{W}_{ij} \left\{\frac{\delta_j I[X_i \geq X_j]}{G(R_i-, X_j-)G(R_j-, X_j-)} - \frac{\theta(\boldsymbol{W}_{ij}\boldsymbol{\beta})}{G(R_i-, 0)G(R_j-, 0)}\right\} \\ = \mathbf{0}. \end{aligned}$$

If G is unknown, it can be replaced by a consistent estimator $\hat{G}$. Hence the estimating equation for $\boldsymbol{\beta}$ is

$$\boldsymbol{U}(\boldsymbol{\beta}; \hat{G}) = n^{-2} \sum_{i=1}^{n} \sum_{j=1}^{n} \boldsymbol{e}_{ij}(\boldsymbol{\beta}; \hat{G}) = \mathbf{0}, \tag{2.1}$$

where

$$\begin{aligned} \boldsymbol{e}_{ij}(\boldsymbol{\beta}; \hat{G}) = \boldsymbol{W}_{ij}\Bigg\{ & \frac{\delta_j I[X_i \geq X_j] \cdot I[\hat{G}(R_i-, X_j-) > 0] \cdot I[\hat{G}(R_j-, X_j-) > 0]}{\hat{G}(R_i-, X_j-)\hat{G}(R_j-, X_j-)} \\ & - \frac{\theta(\boldsymbol{W}_{ij}\boldsymbol{\beta}) \cdot I[\hat{G}(R_i-, 0) > 0] \cdot I[\hat{G}(R_j-, 0) > 0]}{\hat{G}(R_i-, 0)\hat{G}(R_j-, 0)}\Bigg\}. \end{aligned} \tag{2.2}$$

Solving the estimating equation (2.1) gives an unbiased estimator $\hat{\boldsymbol{\beta}}$.

2.3. *Estimating equation for* $h(t)$

Using a similar idea as above, we provide an unbiased estimating equation for $h(t)$ in this section. Denote $h^*(t)$ as the true value of $h(t)$. Without censoring and truncation, considering the indicators $\{I[T_i \geq t], i = 1, \cdots, n\}$, we have that

$$E\{I[T_i \geq t]\big|\boldsymbol{W}_i\} = \mathbf{P}\{h(T_i) \geq h^*(t)]\big|\boldsymbol{W}_i\} := g^{-1}\big(h^*(t) + \boldsymbol{W}_i\boldsymbol{\beta}^*\big),$$

where $g^{-1}(\cdot) = 1 - F_\varepsilon(\cdot)$ and F_ε is the specified distribution function of ε in the model (1.1).

In the presence of truncation and censoring, the indicators $\{I[T_i \geq t], i = 1, \cdots, n\}$ are not always observable. We can only observe $\{I[X_i \geq t], i = 1, \cdots, n\}$ in practice. Therefore we have that

$$\begin{aligned} & E\left\{\frac{I[X_i \geq t]}{G(R_i-, t-)} - \frac{g^{-1}\big(h^*(t) + \boldsymbol{W}_i \boldsymbol{\beta}^*\big)}{G(R_i-, 0)}\Big| \boldsymbol{W}_i\right\} \\ &= \gamma^{-1} E\Big\{I[T_i \geq t] - g^{-1}\big(h^*(t) + \boldsymbol{W}_i \boldsymbol{\beta}^*\big)\big| \boldsymbol{W}_i\Big\} = 0. \end{aligned}$$

Hence a reasonable and unbiased estimating equation for $h(t)$ is given by

$$n^{-1} \sum_{i=1}^{n} \left\{ \frac{I[X_i \geq t, \hat{G}(R_i-, t-) > 0]}{\hat{G}(R_i-, t-)} - \frac{g^{-1}\big(h(t) + \boldsymbol{W}_i \hat{\boldsymbol{\beta}}\big) I[\hat{G}(R_i-, 0) > 0]}{\hat{G}(R_i-, 0)} \right\} = 0, \tag{2.3}$$

where $\hat{G}$ is a consistent estimator of G and $\hat{\boldsymbol{\beta}}$ is the root of (2.1).

2.4. Estimation of G

The challenge of solving the estimating equations for $\boldsymbol{\beta}$ and $h(t)$ is getting a consistent estimate for the bivariate survival function G. Here in this paper, we use the idea in Dai & Fu (2012) and consider the polar coordinate transformation from (t_1, t_2) to $(z; \alpha)$ where $\alpha = t_2/t_1$, $z = \sqrt{t_1^2 + t_2^2}$. For fixed α, $G(t_1, t_2)$ can be transformed to a univariate function, $G(z; \alpha)$, by the following formula

$$G(t_1, t_2) = \mathbf{P}(L > t_1, C > t_2) = \mathbf{P}(Z(\alpha) > z) := G(z; \alpha),$$

where $Z(\alpha) = \min\{L\sqrt{1 + \alpha^2}, C\sqrt{1 + \alpha^{-2}}\}$.

In practice, due to truncation and censoring, the values of (L, C) may not be observed so that $Z(\alpha)$ may not be available. Denote the observed data as $(R_i, L_i, X_i, \delta_i, \boldsymbol{W}_i)$, $i = 1, \ldots, n$ and define $\tilde{L}_i = L_i\sqrt{1 + \alpha^2}$ and $\tilde{X}_i = X_i\sqrt{1 + \alpha^{-2}}$. We have the transformed observed data

$$\begin{aligned} \tilde{Z}_i(\alpha) &= \min\{\tilde{L}_i, \tilde{X}_i\}, \\ \Delta_i(\alpha) &= I[\tilde{L}_i \leq \tilde{X}_i] + (1 - \delta_i) I[\tilde{L}_i > \tilde{X}_i], \\ V_i(\alpha) &= R_i\sqrt{1 + \alpha^2}. \end{aligned}$$

Such a transformation introduces artificial censoring and truncation. Specifically, $\Delta_i(\alpha) = 1$ implies that $\tilde{Z}_i(\alpha)$ is the observed value of $Z_i(\alpha)$, $\Delta_i(\alpha) = 0$ implies censoring, and truncation information is given by $V_i(\alpha)$. Then based on the transformed observations $\big\{\tilde{Z}_i(\alpha), \Delta_i(\alpha), V_i(\alpha)\big\}$, $i = 1, \ldots, n$, we can estimate the univariate function $G(z; \alpha)$ using the following lemma.

Lemma 2.1. *For fixed α, the hazard function of $Z(\alpha)$ is denoted by $\Lambda(dz; \alpha) = -G(dz; \alpha)/G(z-; \alpha)$. Then we have that*

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

where $\cdot \in dz$ denotes $\cdot \in [z, z + dz)$. □

Define

$$\begin{aligned}
N(ds; \alpha) &= n^{-1} \sum_{i=1}^{n} N_i(ds; \alpha) = n^{-1} \sum_{i=1}^{n} I[\tilde{Z}_i(\alpha) \in ds, s > V_i(\alpha), \Delta_i(\alpha) = 1], \\
H_{(n)}(s; \alpha) &= n^{-1} \sum_{i=1}^{n} H_i(s; \alpha) = n^{-1} \sum_{i=1}^{n} I[\tilde{Z}_i(\alpha) > s \geq V_i(\alpha)], \\
H_{(n)}(t_1, t_2) &= n^{-1} \sum_{i=1}^{n} H_i(t_1, t_2) = n^{-1} \sum_{i=1}^{n} I[L_i > t_1 \geq R_i, X_i > t_2].
\end{aligned}$$

Note that $H_{(n)}(t_1, t_2) = H_{(n)}(z; \alpha)$ and $H_i(t_1, t_2) = H_i(z; \alpha)$. Hence Lemma 2.1 implies that an estimator for $\Lambda(dz; \alpha)$ is $\hat{\Lambda}(dz; \alpha) = N(dz; \alpha)/H_{(n)}(z-; \alpha)$. Then the product-limit estimator for $G(z; \alpha)$ is

$$\hat{G}(z; \alpha) = \prod_{s \leq z} \left\{1 - \frac{N(s; \alpha) - N(s-; \alpha)}{H_{(n)}(s-; \alpha)}\right\}.$$

Since $G(z; \alpha) = G(t_1, t_2)$, $\hat{G}(z; \alpha)$ is also an estimator for $G(t_1, t_2)$.

3. Large sample properties of $\hat{\boldsymbol{\beta}}$

3.1. Consistency of $\hat{\beta}$

Denote Φ as the distribution function of the covariates vector $\boldsymbol{W}$. We can show that with probability one, $\boldsymbol{U}(\boldsymbol{\beta}; G)(\boldsymbol{\beta}^* - \boldsymbol{\beta})$ converges to

$$\int_{\boldsymbol{w_1}, \boldsymbol{w_2}} \frac{\boldsymbol{w}_{12}\boldsymbol{\beta}^* - \boldsymbol{w}_{12}\boldsymbol{\beta}}{G(R_i-, 0)G(R_j-, 0)} \Big[\theta(\boldsymbol{w}_{12}\boldsymbol{\beta}^*) - \theta(\boldsymbol{w}_{12}\boldsymbol{\beta})\Big] d\Phi(\boldsymbol{w}_1) d\Phi(\boldsymbol{w}_2),$$

where $\boldsymbol{w}_{12} = \boldsymbol{w}_1 - \boldsymbol{w}_2$ and $\boldsymbol{\beta}^*$ is the true value of $\boldsymbol{\beta}$. Since $\theta(\cdot)$ is a decreasing function, the above limit is non-negative and is zero only if $\boldsymbol{\beta} = \boldsymbol{\beta}^*$. Following the idea in Cheng et al. (1995), together with $\boldsymbol{U}(\boldsymbol{\beta}; G) = \mathbf{0}$ at $\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}$, we have that $\hat{\boldsymbol{\beta}} \to \boldsymbol{\beta}^*$ in probability when $n \to \infty$. This implies that $\hat{\boldsymbol{\beta}}$ is a consistent estimator.

3.2. Asymptotic normality of $\hat{\beta}$

The following theorem provides the results of asymptotic normality of $\hat{\boldsymbol{\beta}}$. A heuristic proof of the theorem can be found in the Appendix.

Theorem 3.1. *Let $\boldsymbol{U}'(\boldsymbol{\beta}; G) = \partial \boldsymbol{U}(\boldsymbol{\beta}; G)/\partial \boldsymbol{\beta}$. Then we have that $n^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \to N(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{\beta}})$ as $n \to \infty$, where*

$$\boldsymbol{\Sigma}_{\boldsymbol{\beta}} = \left[\boldsymbol{U}'(\boldsymbol{\beta}; G)\right]^{-1} \boldsymbol{\Sigma}_{\boldsymbol{U}} \left\{ \left[\boldsymbol{U}'(\boldsymbol{\beta}; G)\right]^{-1} \right\}^{tr}, \tag{3.1}$$

$$\boldsymbol{\Sigma}_{\boldsymbol{U}} = \mathrm{Var}\left\{ n^{-5/2} \sum_{i,j,k=1}^{n} \left[\boldsymbol{e}_{ij}(\boldsymbol{\beta}; G) + \boldsymbol{\varsigma}_{ijk}(\boldsymbol{\beta}, G, M_k)\right] \right\}. \tag{3.2}$$

The term $\boldsymbol{\varsigma}_{ijk}(\boldsymbol{\beta}, G, M_k)$ is given by

$$\begin{aligned} \boldsymbol{\varsigma}_{ijk}(\boldsymbol{\beta}, G, M_k) = & \frac{\boldsymbol{W}_{ij} \delta_j I[X_i \geq X_j]}{G(R_i-, X_j-) G(R_j-, X_j)} \Big[M_k(Z_{ij}; \alpha_{ij}) + M_k(Z_{jj}; \alpha_{jj}) \Big] \\ & - \frac{\boldsymbol{W}_{ij} \theta(\boldsymbol{W}_{ij} \boldsymbol{\beta})}{G(R_i-, 0) G(R_j-, 0)} \Big[M_k(R_i, 0) + M_k(R_j, 0) \Big], \end{aligned} \tag{3.3}$$

where $Z_{ij} = \sqrt{R_i^2 + X_j^2}$, $Z_{jj} = \sqrt{R_j^2 + X_j^2}$, $\alpha_{ij} = X_j/R_i$, $\alpha_{jj} = X_j/R_j$. The term $M_k(z; \alpha)$ is defined as $M_k(z; \alpha) = \sum_{k=1}^{n} \int_{s \leq z} \frac{1}{H(s-; \alpha)} M_k(ds; \alpha)$, where $M_k(ds; \alpha) = N_k(ds; \alpha) - H_k(s-; \alpha) \Lambda(ds; \alpha)$ and $H(s; \alpha) = E[H_i(s; \alpha)]$.

Proof 3.1. *See Appendix A.*

Replacing $\boldsymbol{\beta}$, G and M_k by their estimates, then $\boldsymbol{\Sigma}_{\boldsymbol{\beta}}$ and $\boldsymbol{U}_{\boldsymbol{\beta}}$ given in (3.1) and (3.2) can be estimated by

$$\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\beta}} = \left[\boldsymbol{U}'(\hat{\boldsymbol{\beta}}; \hat{G})\right]^{-1} \hat{\boldsymbol{\Sigma}}_{\boldsymbol{U}} \left\{ \left[\boldsymbol{U}'(\hat{\boldsymbol{\beta}}; \hat{G})\right]^{-1} \right\}^{tr},$$

$$\hat{\boldsymbol{\Sigma}}_{\boldsymbol{U}} = n^{-5} \sum_{i,j,k=1}^{n} \left\{ \left[\boldsymbol{e}_{ij}(\hat{\boldsymbol{\beta}}; \hat{G}) + \boldsymbol{\varsigma}_{ijk}(\hat{\boldsymbol{\beta}}, \hat{G}, \hat{M}_k)\right]^{\otimes 2} \right\},$$

where the notation '*tr*' denotes the transpose of a matrix, and '$\otimes 2$' denotes the product of a matrix and its transpose.

4. Simulation studies

4.1. Estimation of β

In this subsection, simulation studies are conducted to show the properties of our proposed estimator for $\boldsymbol{\beta}$. We consider a three-dimensional covariates vector $\boldsymbol{W}$, where $W_1 \sim U[20, 30]$, $W_2 \sim$ Bernoulli(0.5) and $W_3 \sim$ Bernoulli(0.5), to mimic datasets in the real world which usually contain both continuous and discrete covariates (0/1 indicators). The event time of main interest, T, follows $h(T) = -\boldsymbol{W}\boldsymbol{\beta}^{*} + \varepsilon$, where the strictly increasing function $h(T) = 4\sqrt{T-5} + 8$, and the true value of $\boldsymbol{\beta}$ is $\boldsymbol{\beta}^{*} = (-0.2, -1.2, -1.5)^{tr}$. The distribution function of the error term ε is taken to be a standard extreme value distribution with $F_{\varepsilon}(t) = 1 - \exp\{-\exp(t)\}$. This makes the linear transformation model be its special case of PH model and allows us to compare the performance of our method with the conventional PH model that handles univariate survival data with only censoring. The event time R is generated by $R = T \times U[0.6, 0.8]$. The correlated truncation and censoring variables L and C are simulated via

$$L = a_1\nu_1 + a_2\nu_2 + U[3, 4] \quad \text{and} \quad C = b_1\nu_1 + b_2\nu_2 + U[4, 5], \qquad (4.1)$$

where $\nu_1, \nu_2 \sim Exp(5)$. Our scenario guarantees that the generated data satisfies the Conditions 2.1 and 2.2 mentioned in Section 2.

The values of a_1, a_2, b_1 and b_2 in Equation (4.1) are adjusted to achieve different truncation probabilities and censoring percentages. In this study the censoring percentage for T is considered to be around 10%, 30% and 50%, respectively. The truncation probability $\gamma = \mathbf{P}(R \le L)$ is considered to be around 0.9, 0.7, 0.5, 0.3 and 0.1, respectively. Note that only observations with $R \le L$ can be observed. We consider different observed sample size, $n = 200$ and $n = 500$. Number of simulations is taken to be 200 and 2000. The results are presented in Table 2 and 3, respectively.

The biases and standard errors of the estimates obtained from different number of simulations (200 or 2000) are similar. However, the difference between the empirical standard error ($\hat{s}_{\hat{\boldsymbol{\beta}}}$) and the mean of standard deviation estimates ($\hat{\sigma}_{\hat{\boldsymbol{\beta}}}$) becomes smaller when the iteration times increases from 200 to 2000.

For the same number of iterations and the same observed sample size, higher truncation probabilities and lower censoring percentages give better estimates with smaller biases and standard errors. Under the same truncation probability and censoring percentage, the biases are similar for different observed sample sizes. However the estimated standard errors are much smaller for large sample, which makes a difference when the observed sample is more biased from the population (lower truncation probability & heavier censoring). For example, when the observed sample size is relatively small ($n = 200$), the estimates are non-significant for all the three cases with truncation probability $\gamma = 0.1$, and for the case with $\gamma = 0.3$ and 50% censoring percentage. While for a larger observed sample ($n = 500$), the issue of non-significance only occurs when the observed data is severely biased (only 10% of the population can be observed and around half of the observations are censored).

4.2. Comparison with PH model

In this subsection, we compare our proposed method with the conventional PH model via 500 simulations. The data are generated using the same strategy as that in section 4.1, while the referral bias due to truncation is not considered in the PH model. The results are presented in Table 4 and 5.

Under 10% and 30% censoring, our method gives less biased estimates for all different truncation probabilities. However, in the presence of higher censoring percentage (50%), our proposal is not expected to be as efficient as the Cox procedure. This is due to the inverse probability weighted estimator of the bivariate survival function G used in our method does struggle for heavier censoring, as pointed out in Dai & Bao (2009); Dai & Fu (2012). Therefore the insignificance of improvement when using our method to analyse severely biased survival data (with censoring percentage around 50% or greater) is reasonable.

Table 2: Simulation results for 200 simulations. γ: truncation probability; $\hat{s}_{\hat{\boldsymbol{\beta}}}$: empirical standard errors for $\hat{\boldsymbol{\beta}}$ based on 200 simulations; $\hat{\sigma}_{\hat{\boldsymbol{\beta}}}$: means of standard deviation estimates obtained from a perturbation method (Wang & Zhu, 2006).

		Cens.% = 10%			Cens.% = 30%			Cens.% = 50%		
	$\boldsymbol{\beta}^*$	$\hat{\boldsymbol{\beta}}$ (bias)	$\hat{s}_{\hat{\boldsymbol{\beta}}}$	$\hat{\sigma}_{\hat{\boldsymbol{\beta}}}$	$\hat{\boldsymbol{\beta}}$ (bias)	$\hat{s}_{\hat{\boldsymbol{\beta}}}$	$\hat{\sigma}_{\hat{\boldsymbol{\beta}}}$	$\hat{\boldsymbol{\beta}}$ (bias)	$\hat{s}_{\hat{\boldsymbol{\beta}}}$	$\hat{\sigma}_{\hat{\boldsymbol{\beta}}}$
$n = 500$										
$\gamma = 0.9$	-0.2	-0.197 (0.003)	0.023	0.023	-0.190 (0.010)	0.026	0.024	-0.180 (0.020)	0.037	0.034
	-1.2	-1.183 (0.017)	0.116	0.125	-1.123 (0.077)	0.159	0.160	-1.066 (0.134)	0.199	0.184
	-1.5	-1.506 (0.006)	0.144	0.139	-1.423 (0.077)	0.201	0.164	-1.352 (0.148)	0.233	0.249
$\gamma = 0.7$	-0.2	-0.192 (0.008)	0.026	0.022	-0.178 (0.022)	0.028	0.031	-0.172 (0.028)	0.033	0.035
	-1.2	-1.125 (0.075)	0.180	0.159	-1.077 (0.123)	0.220	0.194	-1.028 (0.172)	0.153	0.179
	-1.5	-1.458 (0.042)	0.176	0.163	-1.405 (0.095)	0.197	0.190	-1.290 (0.210)	0.214	0.201
$\gamma = 0.5$	-0.2	-0.174 (0.026)	0.031	0.030	-0.168 (0.032)	0.030	0.032	-0.157 (0.043)	0.041	0.042
	-1.2	-1.021 (0.179)	0.163	0.169	-0.979 (0.221)	0.207	0.239	-0.971 (0.229)	0.259	0.254
	-1.5	-1.365 (0.135)	0.176	0.185	-1.271 (0.229)	0.243	0.264	-1.251 (0.249)	0.269	0.272
$\gamma = 0.3$	-0.2	-0.155 (0.045)	0.035	0.033	-0.147 (0.053)	0.041	0.037	-0.139 (0.061)	0.052	0.047
	-1.2	-0.930 (0.270)	0.200	0.196	-0.835 (0.365)	0.270	0.289	-0.835 (0.365)	0.333	0.267
	-1.5	-1.164 (0.336)	0.203	0.173	-1.165 (0.335)	0.272	0.274	-1.125 (0.375)	0.397	0.301
$\gamma = 0.1$	-0.2	-0.136 (0.064)	0.050	0.042	-0.122 (0.078)	0.052	0.061	-0.098 (0.102)	0.050	0.050
	-1.2	-0.852 (0.348)	0.351	0.277	-0.693 (0.507)	0.318	0.381	-0.603 (0.597)	0.338	0.431
	-1.5	-1.119 (0.381)	0.391	0.309	-0.963 (0.537)	0.429	0.395	-0.815 (0.685)	0.413	0.378
$n = 200$										
$\gamma = 0.9$	-0.2	-0.198 (0.002)	0.039	0.040	-0.183 (0.017)	0.049	0.048	-0.176 (0.024)	0.056	0.056
	-1.2	-1.184 (0.016)	0.241	0.243	-1.119 (0.081)	0.326	0.329	-1.049 (0.151)	0.339	0.371
	-1.5	-1.491 (0.009)	0.256	0.257	-1.410 (0.090)	0.337	0.336	-1.296 (0.204)	0.387	0.409
$\gamma = 0.7$	-0.2	-0.189 (0.011)	0.045	0.045	-0.175 (0.025)	0.048	0.053	-0.162 (0.038)	0.064	0.065
	-1.2	-1.125 (0.075)	0.321	0.316	-1.035 (0.165)	0.289	0.308	-0.960 (0.240)	0.400	0.447
	-1.5	-1.417 (0.083)	0.305	0.296	-1.354 (0.146)	0.325	0.328	-1.241 (0.259)	0.411	0.451
$\gamma = 0.5$	-0.2	-0.172 (0.028)	0.049	0.052	-0.163 (0.037)	0.056	0.049	-0.151 (0.049)	0.078	0.078
	-1.2	-1.026 (0.174)	0.335	0.428	-0.964 (0.236)	0.448	0.359	-0.906 (0.294)	0.440	0.442
	-1.5	-1.302 (0.198)	0.304	0.383	-1.295 (0.205)	0.437	0.402	-1.184 (0.316)	0.486	0.487
$\gamma = 0.3$	-0.2	-0.150 (0.050)	0.058	0.061	-0.148 (0.052)	0.078	0.071	-0.133 (0.067)	0.075	0.088
	-1.2	-0.941 (0.259)	0.367	0.376	-0.851 (0.349)	0.413	0.394	-0.824 (0.376)	0.501	0.593
	-1.5	-1.198 (0.302)	0.347	0.358	-1.115 (0.385)	0.487	0.487	-1.095 (0.405)	0.621	0.780
$\gamma = 0.1$	-0.2	-0.139 (0.061)	0.094	0.096	-0.124 (0.076)	0..092	0.099	-0.101 (0.099)	0.096	0.103
	-1.2	-0.804 (0.396)	0.629	0.486	-0.696 (0.504)	0.716	0.780	-0.610 (0.590)	0.721	0.670
	-1.5	-1.167 (0.333)	0.740	0.591	-0.975 (0.525)	0.650	0.718	-0.840 (0.660)	0.837	0.686

Table 3: Simulation results for 2000 simulations. γ: truncation probability; $\hat{s}_{\hat{\boldsymbol{\beta}}}$: empirical standard errors for $\hat{\boldsymbol{\beta}}$ based on 2000 simulations; $\hat{\sigma}_{\hat{\boldsymbol{\beta}}}$: means of standard deviation estimates obtained from a perturbation method (Wang & Zhu, 2006).

		Cens.% = 10%			Cens.% = 30%			Cens.% = 50%		
	$\boldsymbol{\beta}^*$	$\hat{\boldsymbol{\beta}}$ (bias)	$\hat{s}_{\hat{\boldsymbol{\beta}}}$	$\hat{\sigma}_{\hat{\boldsymbol{\beta}}}$	$\hat{\boldsymbol{\beta}}$ (bias)	$\hat{s}_{\hat{\boldsymbol{\beta}}}$	$\hat{\sigma}_{\hat{\boldsymbol{\beta}}}$	$\hat{\boldsymbol{\beta}}$ (bias)	$\hat{s}_{\hat{\boldsymbol{\beta}}}$	$\hat{\sigma}_{\hat{\boldsymbol{\beta}}}$
$n = 500$										
$\gamma = 0.9$	-0.2	-0.199 (0.001)	0.020	0.020	-0.190 (0.010)	0.025	0.025	-0.178 (0.022)	0.030	0.030
	-1.2	-1.194 (0.006)	0.114	0.116	-1.136 (0.064)	0.146	0.153	-1.072 (0.128)	0.186	0.194
	-1.5	-1.490 (0.010)	0.123	0.124	-1.424 (0.076)	0.163	0.168	-1.338 (0.162)	0.209	0.209
$\gamma = 0.7$	-0.2	-0.189 (0.011)	0.022	0.022	-0.178 (0.022)	0.026	0.027	-0.169 (0.031)	0.032	0.032
	-1.2	-1.143 (0.057)	0.133	0.135	-1.072 (0.128)	0.156	0.162	-1.027 (0.173)	0.205	0.204
	-1.5	-1.439 (0.061)	0.138	0.140	-1.356 (0.144)	0.175	0.181	-1.304 (0.196)	0.226	0.223
$\gamma = 0.5$	-0.2	-0.174 (0.026)	0.027	0.027	-0.165 (0.035)	0.030	0.031	-0.154 (0.046)	0.034	0.035
	-1.2	-1.055 (0.145)	0.171	0.173	-0.993 (0.207)	0.190	0.201	-0.934 (0.266)	0.233	0.236
	-1.5	-1.358 (0.142)	0.163	0.166	-1.280 (0.220)	0.206	0.203	-1.188 (0.312)	0.250	0.261
$\gamma = 0.3$	-0.2	-0.151 (0.049)	0.028	0.028	-0.149 (0.051)	0.036	0.036	-0.139 (0.061)	0.043	0.041
	-1.2	-0.904 (0.296)	0.169	0.170	-0.888 (0.312)	0.227	0.234	-0.844 (0.356)	0.275	0.312
	-1.5	-1.179 (0.321)	0.180	0.183	-1.158 (0.342)	0.251	0.255	-1.074 (0.426)	0.290	0.297
$\gamma = 0.1$	-0.2	-0.140 (0.060)	0.044	0.046	-0.114 (0.086)	0.048	0.046	-0.100 (0.100)	0.050	0.050
	-1.2	-0.861 (0.339)	0.311	0.318	-0.732 (0.468)	0.338	0.324	-0.605 (0.595)	0.338	0.324
	-1.5	-1.139 (0.361)	0.349	0.348	-0.958 (0.542)	0.375	0.366	-0.820 (0.680)	-0.377	0.358
$n = 200$										
$\gamma = 0.9$	-0.2	-0.196 (0.004)	0.040	0.040	-0.188 (0.012)	0.048	0.050	-0.176 (0.024)	0.064	0.066
	-1.2	-1.170 (0.030)	0.237	0.239	-1.121 (0.079)	0.306	0.312	-1.060 (0.140)	0.390	0.396
	-1.5	-1.472 (0.028)	0.244	0.245	-1.412 (0.088)	0.348	0.352	-1.323 (0.177)	0.452	0.446
$\gamma = 0.7$	-0.2	-0.188 (0.012)	0.048	0.047	-0.177 (0.023)	0.054	0.054	-0.161 (0.039)	0.061	0.062
	-1.2	-1.141 (0.059)	0.298	0.299	-1.065 (0.135)	0.334	0.338	-0.984 (0.216)	0.397	0.382
	-1.5	-1.437 (0.063)	0.299	0.296	-1.356 (0.144)	0.360	0.361	-1.253 (0.247)	0.448	0.434
$\gamma = 0.5$	-0.2	-0.171 (0.029)	0.058	0.058	-0.163 (0.037)	0.062	0.064	-0.152 (0.048)	0.070	0.071
	-1.2	-1.036 (0.164)	0.343	0.341	-0.973 (0.227)	0.364	0.375	-0.893 (0.307)	0.439	0.445
	-1.5	-1.350 (0.150)	0.352	0.361	-1.265 (0.235)	0.392	0.405	-1.176 (0.324)	0.497	0.505
$\gamma = 0.3$	-0.2	-0.149 (0.051)	0.057	0.057	-0.143 (0.057)	0.070	0.070	-0.134 (0.066)	0.079	0.081
	-1.2	-0.908 (0.292)	0.345	0.347	-0.890 (0.310)	0.457	0.468	-0.842 (0.358)	0.536	0.564
	-1.5	-1.169 (0.331)	0.355	0.367	-1.133 (0.367)	0.509	0.496	-1.074 (0.426)	0.625	0.642
$\gamma = 0.1$	-0.2	-0.142 (0.058)	0.090	0.087	-0.113 (0.087)	0.093	0.095	-0.101 (0.099)	0.100	0.096
	-1.2	-0.864 (0.336)	0.562	0.557	-0.738 (0.462)	0.655	0.629	-0.620 (0.580)	0.672	0.686
	-1.5	-1.141 (0.359)	0.647	0.653	-0.943 (0.557)	0.663	0.649	-0.839 (0.661)	0.750	0.731

Table 4: Simulation results for 500 simulations. γ: truncation probability; $\hat{s}_{\hat{\boldsymbol{\beta}}}$: empirical standard errors for $\hat{\boldsymbol{\beta}}$ based on 500 simulations; $\hat{\sigma}_{\hat{\boldsymbol{\beta}}}$: means of standard deviation estimates obtained from a perturbation method (Wang & Zhu, 2006).

		Cens.% = 10%			Cens.% = 30%			Cens.% = 50%		
	$\boldsymbol{\beta}^*$	$\hat{\boldsymbol{\beta}}$ (bias)	$\hat{s}_{\hat{\boldsymbol{\beta}}}$	$\hat{\sigma}_{\hat{\boldsymbol{\beta}}}$	$\hat{\boldsymbol{\beta}}$ (bias)	$\hat{s}_{\hat{\boldsymbol{\beta}}}$	$\hat{\sigma}_{\hat{\boldsymbol{\beta}}}$	$\hat{\boldsymbol{\beta}}$ (bias)	$\hat{s}_{\hat{\boldsymbol{\beta}}}$	$\hat{\sigma}_{\hat{\boldsymbol{\beta}}}$
$n = 500$										
$\gamma = 0.9$	-0.2	-0.199 (0.001)	0.021	0.019	-0.190 (0.010)	0.028	0.028	-0.178 (0.022)	0.032	0.033
	-1.2	-1.192 (0.008)	0.123	0.139	-1.150 (0.050)	0.180	0.177	-1.079 (0.121)	0.206	0.210
	-1.5	-1.492 (0.008)	0.125	0.134	-1.410 (0.090)	0.181	0.187	-1.351 (0.149)	0.237	0.241
$\gamma = 0.7$	-0.2	-0.187 (0.013)	0.023	0.024	-0.181 (0.019)	0.029	0.031	-0.173 (0.027)	0.039	0.033
	-1.2	-1.139 (0.061)	0.137	0.149	-1.064 (0.136)	0.177	0.190	-1.024 (0.176)	0.215	0.227
	-1.5	-1.440 (0.060)	0.155	0.152	-1.336 (0.164)	0.182	0.198	-1.293 (0.207)	0.255	0.250
$\gamma = 0.5$	-0.2	-0.173 (0.027)	0.026	0.027	-0.163 (0.037)	0.033	0.033	-0.157 (0.043)	0.041	0.039
	-1.2	-1.027 (0.173)	0.174	0.206	-0.979 (0.221)	0.221	0.193	-0.931 (0.269)	0.245	0.265
	-1.5	-1.351 (0.149)	0.168	0.198	-1.286 (0.214)	0.224	0.241	-1.168 (0.332)	0.259	0.280
$\gamma = 0.3$	-0.2	-0.150 (0.050)	0.030	0.032	-0.146 (0.054)	0.036	0.038	-0.135 (0.065)	0.045	0.042
	-1.2	-0.933 (0.267)	0.229	0.243	-0.888 (0.312)	0.241	0.258	-0.812 (0.388)	0.280	0.292
	-1.5	-1.185 (0.315)	0.199	0.203	-1.139 (0.361)	0.290	0.262	-1.102 (0.398)	0.341	0.351
$\gamma = 0.1$	-0.2	-0.138 (0.062)	0.046	0.041	-0.118 (0.082)	0.057	0.056	-0.093 (0.107)	0.051	0.057
	-1.2	-0.882 (0.318)	0.296	0.310	-0.747 (0.453)	0.335	0.364	-0.675 (0.525)	0.404	0.366
	-1.5	-1.177 (0.323)	0.478	0.319	-0.986 (0.514)	0.405	0.369	-0.824 (0.676)	0.399	0.374
$n = 200$										
$\gamma = 0.9$	-0.2	-0.198 (0.002)	0.041	0.041	-0.189 (0.011)	0.053	0.053	-0.177 (0.023)	0.066	0.066
	-1.2	-1.188 (0.012)	0.228	0.231	-1.125 (0.075)	0.308	0.304	-1.055 (0.145)	0.428	0.434
	-1.5	-1.489 (0.011)	0.254	0.256	-1.398 (0.102)	0.336	0.330	-1.327 (0.173)	0.471	0.484
$\gamma = 0.7$	-0.2	-0.186 (0.014)	0.046	0.047	-0.175 (0.025)	0.053	0.057	-0.164 (0.036)	0.062	0.064
	-1.2	-1.140 (0.060)	0.286	0.312	-1.066 (0.134)	0.318	0.328	-0.957 (0.243)	0.374	0.430
	-1.5	-1.434 (0.066)	0.290	0.301	-1.352 (0.148)	0.374	0.392	-1.259 (0.241)	0.405	0.457
$\gamma = 0.5$	-0.2	-0.171 (0.029)	0.056	0.056	-0.162 (0.038)	0.062	0.065	-0.152 (0.048)	0.072	0.071
	-1.2	-1.018 (0.182)	0.317	0.328	-0.988 (0.212)	0.437	0.377	-0.901 (0.299)	0.446	0.459
	-1.5	-1.350 (0.150)	0.326	0.329	-1.278 (0.222)	0.464	0.428	-1.193 (0.307)	0.517	0.540
$\gamma = 0.3$	-0.2	-0.152 (0.048)	0.056	0.055	-0.143 (0.057)	0.068	0.071	-0.133 (0.067)	0.079	0.080
	-1.2	-0.927 (0.273)	0.351	0.346	-0.871 (0.329)	0.417	0.435	-0.846 (0.354)	0.527	0.560
	-1.5	-1.181 (0.319)	0.386	0.381	-1.104 (0.396)	0.423	0.466	-1.070 (0.430)	0.546	0.593
$\gamma = 0.1$	-0.2	-0.143 (0.057)	0.091	0.088	-0.123 (0.077)	0.097	0.103	-0.094 (0.106)	0.083	0.082
	-1.2	-0.845 (0.355)	0.556	0.608	-0.762 (0.438)	0.696	0.783	-0.575 (0.625)	0.636	0.676
	-1.5	-1.098 (0.402)	0.641	0.723	-0.978 (0.522)	0.831	0.791	-0.801 (0.699)	0.709	0.729

Table 5: Simulation results from PH model for 500 simulations (univariate censoring only). γ: truncation probability; $\hat{s}_{\hat{\boldsymbol{\beta}}}$: empirical standard errors for $\hat{\boldsymbol{\beta}}$ based on 500 simulations.

		Cens.% = 10%		Cens.% = 30%		Cens.% = 50%	
	$\boldsymbol{\beta}^*$	$\hat{\boldsymbol{\beta}}$ (bias)	$\hat{s}_{\hat{\boldsymbol{\beta}}}$	$\hat{\boldsymbol{\beta}}$ (bias)	$\hat{s}_{\hat{\boldsymbol{\beta}}}$	$\hat{\boldsymbol{\beta}}$ (bias)	$\hat{s}_{\hat{\boldsymbol{\beta}}}$
$n = 500$							
$\gamma = 0.9$	-0.2	-0.190 (0.010)	0.018	-0.191 (0.009)	0.020	-0.192 (0.008)	0.023
	-1.2	-1.135 (0.065)	0.105	-1.129 (0.071)	0.117	-1.135 (0.065)	0.136
	-1.5	-1.428 (0.072)	0.110	-1.431 (0.069)	0.123	-1.430 (0.070)	0.143
$\gamma = 0.7$	-0.2	-0.165 (0.035)	0.018	-0.170 (0.030)	0.020	-0.172 (0.028)	0.023
	-1.2	-0.988 (0.212)	0.103	-1.018 (0.182)	0.116	-1.029 (0.171)	0.136
	-1.5	-1.267 (0.233)	0.108	-1.282 (0.218)	0.121	-1.293 (0.207)	0.141
$\gamma = 0.5$	-0.2	-0.152 (0.048)	0.018	-0.157 (0.044)	0.020	-0.158 (0.042)	0.024
	-1.2	-0.921 (0.279)	0.102	-0.914 (0.286)	0.114	-0.946 (0.254)	0.134
	-1.5	-1.141 (0.359)	0.106	-1.164 (0.336)	0.119	-1.188 (0.312)	0.140
$\gamma = 0.3$	-0.2	-0.135 (0.065)	0.017	-0.137 (0.063)	0.020	-0.142 (0.058)	0.023
	-1.2	-0.811 (0.389)	0.100	-0.803 (0.397)	0.114	-0.839 (0.361)	0.131
	-1.5	-1.019 (0.481)	0.103	-1.022 (0.478)	0.117	-1.063 (0.437)	0.134
$\gamma = 0.1$	-0.2	-0.123 (0.077)	0.018	-0.121 (0.079)	0.019	-0.126 (0.074)	0.038
	-1.2	-0.756 (0.444)	0.099	-0.746 (0.454)	0.112	-0.762 (0.438)	0.209
	-1.5	-0.950 (0.550)	0.102	-0.939 (0.561)	0.114	-0.986 (0.514)	0.212
$n = 200$							
$\gamma = 0.9$	-0.2	-0.189 (0.011)	0.029	-0.190 (0.010)	0.034	-0.190 (0.010)	0.037
	-1.2	-1.134 (0.066)	0.168	-1.138 (0.062)	0.186	-1.136 (0.064)	0.218
	-1.5	-1.421 (0.079)	0.176	-1.433 (0.067)	0.197	-1.415 (0.085)	0.229
$\gamma = 0.7$	-0.2	-0.167 (0.033)	0.029	-0.167 (0.033)	0.032	-0.170 (0.030)	0.039
	-1.2	-1.001 (0.199)	0.166	-1.002 (0.198)	0.187	-1.022 (0.178)	0.219
	-1.5	-1.270 (0.230)	0.173	-1.295 (0.205)	0.193	-1.298 (0.202)	0.224
$\gamma = 0.5$	-0.2	-0.155 (0.045)	0.027	-0.157 (0.043)	0.032	-0.153 (0.047)	0.040
	-1.2	-0.926 (0.274)	0.168	-0.920 (0.280)	0.183	-0.945 (0.255)	0.217
	-1.5	-1.150 (0.350)	0.173	-1.185 (0.315)	0.191	-1.144 (0.356)	0.229
$\gamma = 0.3$	-0.2	-0.136 (0.064)	0.027	-0.135 (0.065)	0.032	-0.142 (0.058)	0.039
	-1.2	-0.811 (0.389)	0.163	-0.839 (0.361)	0.181	-0.861 (0.339)	0.213
	-1.5	-1.063 (0.437)	0.166	-1.073 (0.427)	0.187	-1.055 (0.445)	0.217
$\gamma = 0.1$	-0.2	-0.125 (0.075)	0.030	-0.130 (0.070)	0.034	-0.127 (0.073)	0.037
	-1.2	-0.767 (0.433)	0.163	-0.778 (0.422)	0.177	-0.797 (0.403)	0.208
	-1.5	-0.953 (0.547)	0.169	-0.993 (0.507)	0.185	-0.993 (0.507)	0.213

4.3. Estimation of $h(t)$

In practice, the function $h(\cdot)$ in the semiparametric linear transformation model (1.1) is unknown. Therefore, unlike the AFT model studied in Wang et al. (2013), the model (1.1) discussed in this paper cannot be directly used to predict the event time of interest at individual level. Although the estimation of h is of lower interest, it can still be estimated using the estimating equation (2.3). For $i = 1, \ldots, n$, the estimated $h(t)$, evaluated at $t = X_i$, should be able to recover the rank of T_i, since h is a strictly increasing function. Here we introduce a simple method to assess the performance of the estimates for $h(t)$.

Given $F_\varepsilon(t) = 1 - \exp\{-\exp(t)\}$, for an observed sample size $n = 500$ and 500 simulations, we have that $\hat{\boldsymbol{\beta}} = (-0.181, -1.064, -1.336)^{tr}$ under around 0.7 truncation probability and 30% censoring percentage (Table 4). Then $h(T_i)$ can be estimated by substituting $\hat{\boldsymbol{\beta}}$ into the estimating equation (2.3). Since $g^{-1}(\cdot) = 1 - F_\varepsilon(\cdot)$, the distribution of $\{(\hat{h}(X_i) + \boldsymbol{W}_i\hat{\boldsymbol{\beta}}, \delta_i), i = 1, \ldots, n\}$ should be very close to the distribution of ε. As shown in Figure 2, the estimates are slightly biased at the tails because the bivariate survival function estimates are not very good when points are too sparse. But overall the distribution of $\{(\hat{h}(X_i) + \boldsymbol{W}_i\hat{\boldsymbol{\beta}}, \delta_i)\}$ are very close to the theoretical distribution of ε.

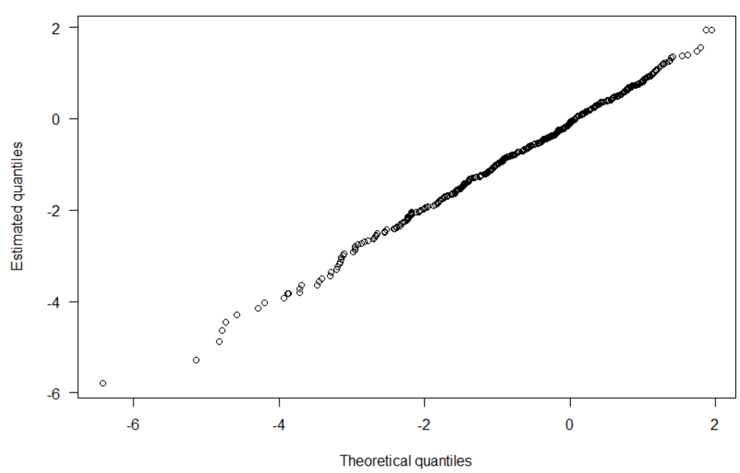

Figure 2: Q-Q plot for $\{(\hat{h}(X_i) + \boldsymbol{W}_i\hat{\boldsymbol{\beta}}, \delta_i), i = 1, \ldots, n\}$.

5. Data analyses

In this section we apply our proposed method on the two real-world datasets in Example 1.1 and 1.2 to illustrate its practicability in different research areas.

5.1. Hepatitis C data from Edinburgh Royal Infirmary Hospital

The dataset consists of 387 patients with chronic hepatitis C who had been recruited to the liver clinic in Edinburgh Royal Infirmary hospital by the end of year 1999. Patients were included to the study cohort with a referral bias (Dore et al., 2002). This is because the disease progression of chronic hepatitis C is often asymptomatic after initial infection. Most people with HCV infection do not seek medical advice until severe symptoms exhibit. Therefore, the patients with more rapid disease progression are preferentially referred to liver clinics or that referral is increasingly likely the closer a patient is to developing cirrhosis (Fu et al., 2007). To take such referral bias into account, we incorporate right truncation of the referral time R, that is, only the patients with $R \leq L$ can be recruited to the study cohort.

Among the 387 patients, no cirrhotic event occurred prior to referral and 63 (16%) developed cirrhosis during their follow-up time. The mean age at HCV infection is around 22 years old. The median duration from infection to referral is 17.1 years and the median follow-up time from referral to cirrhosis or censoring is 2.4 years. Our aim is to determine how the progression to cirrhosis is affected by the three covariates: age at infection, HIV co-infection (yes: 1 or no: 0), and alcohol excess (yes:1 or no: 0). An individual with excess alcohol intake is defined as one consuming more than 50 units per week for at least 5 years.

Given $F_\varepsilon(t) = 1 - \exp\{-\exp(t)\}$, the linear transformation model (1.1) gives the standard Cox model. Under this model we analyse the hepatitis C data using: 1) our proposed method (for bivariate data with both truncation and censoring); 2) the method proposed by Cheng et al. (1995) (for univariate data with only right censoring). Table 6 presents the estimates of regression parameters.

Table 6: Estimation results for hepatitis C data (standard errors in parenthesis).

	1) Truncation & censoring	2) Censoring only
	$\hat{\boldsymbol{\beta}}$ (SE)	$\hat{\boldsymbol{\beta}}$ (SE)
Age at HCV infection	0.008 (0.002)	0.050 (0.005)
HIV co-infection	0.135 (0.051)	0.857 (0.202)
Alcohol excess	0.117 (0.029)	-0.344 (0.101)

By incorporating truncation, the potential referral bias is considered in our proposed method. The results show that all the three covariates are identified as significant risk factors associated with more rapid progression to cirrhosis. The results from the standard Cox model under right censoring only (Cheng et al., 1995) show that ignorance of the potential referral bias leads to a nonsensical estimate of the impact of alcohol excess, i.e. heavy alcohol intake can slow down the progression to cirrhosis. In medical literatures, all these three covariates have been recognised as risk factors associated with more rapid disease progression of hepatitis C (Sharma & Sherker, 2010).

5.2. Business failure data

The dataset described in Example 1.2 consists of 420 small and medium size Italian firms having available information in Amadeus Database provided by Bureau van Dijk. As illustrated in Figure 1, we observe a paired event times (R, T) for each firm, where R is subject to left truncation by L and T is subject to right censoring by C. The potential referral bias exists due to the fact that newly established firms are more likely to be included in the study cohort. In the mean time they are more likely to be bankrupt (e.g. during a financial crisis) than those have been established for longer period of time.

Denote $C^{'}$ as the time period from establishment to the last follow-up. Then in theory, the time R may also be right censored by $C^{'}$, if the lost of follow-up happens in the very short time window between database entry and first financial statement available. However, we ignored this censoring here because it did not happen in our data. In the whole database, the right censoring of R is very rare.

We analyse this data to see how the progression to bankruptcy is affected by the three covariates: return on assets, medium size (yes:1 or no:0) and limited liability form (yes:1 or no:0). These covariates were identified in some existing literatures as risk factors associated with higher risk of bankruptcy (see for example Altman (1968); Bhattacharjee et al. (2009); Situm (2014)). Among the 420 firms, 381 (90 %) have medium size and 367 (87%) have limited liability form. The median duration of follow-up since the first financial statement available is 8.5 years with interquartile range of 6.1 to 9.5 years. Only 82 firms (around 19%) were still active after the last follow-up. Table 7 summarizes the estimates obtained from: 1) our proposed method; 2) the standard Cox model under right censoring only (Cheng et al., 1995).

Table 7: Estimation results for business failure data (standard errors in parenthesis).

	1) Truncation & censoring	2) Censoring only
	$\hat{\boldsymbol{\beta}}$ (SE)	$\hat{\boldsymbol{\beta}}$ (SE)
Return on assets	-0.186 (0.064)	0.009 (0.015)
Medium size	-2.572 (0.968)	-1.896 (0.420)
Limited liability	-1.534 (0.756)	-2.288 (0.471)

With truncation being considered, the results from our method show that lower values of return on assets, smaller firm size and non-limited liability form are associated with more rapid progression to bankruptcy. However, ignorance of truncation leads to a nonsensical result that higher return on assets gives higher risk of bankruptcy (Altman et al., 1977; Laitinen & Suvas, 2013).

5.3. Estimation of truncation probability

In practice, we can also estimate the truncation probability of a real-world data via the method in Shen (2006) or the one in Dai & Fu (2012). Specifically, if both of the bivariate event times are subject to left truncation by L_1 and L_2, the truncation probability γ can be estimated by

$$\hat{\gamma} = \left[n^{-1} \sum_{i=1}^{n} \frac{1}{\hat{S}(L_{1i}-, L_{2i}-)} \right]^{-1}, \tag{5.1}$$

where $S(t_1, t_2) = \mathbf{P}(R > t_1, T > t_2)$ is the bivariate survival function of (R, T), and $S(t_1-, t_2-)$ is its left-continuous version. The bivariate survival function $S(t_1, t_2)$ can be estimated by the methods in Dai & Fu (2012); Dai et al. (2016). In the case that only R is right-truncated by L, we let $R^{'} = const - R$ which is left-truncated at $(const - L)$, where '*const*' is a constant term such that $const - R \geq 0$ and $const - L$ may be negative. Then our method can be applied. Here the estimated truncation probability for the hepatitis C data and the business failure data is 0.08 and 0.21, respectively.

6. Conclusions

In this paper, we developed a new approach for a class of semiparametric linear transformation models $h(T) = -\boldsymbol{W}\boldsymbol{\beta} + \varepsilon$ to handle bivariate survival data under both censoring and truncation. The well-known Cox proportional hazards model can be seen as one special case of the linear transformation model, given the error term ε following a standard extreme value distribution. A new class of estimating equations for the parameter $\boldsymbol{\beta}$ were proposed to allow a flexible bivariate distribution structure between the two correlated event times R and T. By incorporating truncation, the potential referral bias in practice could be taken into account when estimating the regression parameters in the semiparametric linear transformation models. Simulation studies under different scenarios indicated that our method could effectively reduce the bias due to truncation and provide more precise estimates under moderate censoring percentages (around 30% or less). However in the presence of heavier censoring (around 50% or greater), our method might not perform as well as the Cox procedure. Analyses of two real-world dataset demonstrated the importance of applying our method on bivariate survival data with truncation, since our method provided more reliable estimates for the effects of covariates.

In summary, our proposed method is an important candidate of handling bivariate survival data with truncation and can be applied on many research areas. For future work, we may extend the method proposed in this paper to handle time-varying coefficients.

Appendix A. Proof of Theorem 3.1

First we show the asymptotic normality of $n^{1/2}\boldsymbol{U}(\boldsymbol{\beta}^*;\hat{G})$. For simplicity, let $G^{(1)} = G(R_i-, X_j-)$, $G^{(2)} = G(R_j-, X_j-)$, $G^{(3)} = G(R_i-, 0)$ and $G^{(4)} = G(R_j-, 0)$.

For $\boldsymbol{\beta} = \boldsymbol{\beta}^*$, we have

$$\begin{aligned} n^{1/2}\boldsymbol{U}(\boldsymbol{\beta}^*;\hat{G}) &= n^{1/2}\big[\boldsymbol{U}(\boldsymbol{\beta}^*;\hat{G}) - \boldsymbol{U}(\boldsymbol{\beta}^*;G)\big] + n^{1/2}\boldsymbol{U}(\boldsymbol{\beta}^*;G) \\ &= n^{-3/2}\sum_{i=1}^{n}\sum_{j=1}^{n}\big[\boldsymbol{e}_{ij}(\boldsymbol{\beta}^*;G) + \boldsymbol{\nu}_{ij}(\boldsymbol{\beta}^*)\big] + o_p(\mathbf{1}), \end{aligned}$$

where

$$\begin{aligned} \boldsymbol{e}_{ij}(\boldsymbol{\beta}^*;G) &= \boldsymbol{W}_{ij}\left\{\frac{\delta_j I[X_i \geq X_j] I[G^{(1)}G^{(2)} > 0]}{G^{(1)}G^{(2)}} - \frac{\theta(\boldsymbol{W}_{ij}\boldsymbol{\beta}^*) I[G^{(3)}G^{(4)} > 0]}{G^{(3)}G^{(4)}}\right\}, \\ \boldsymbol{\nu}_{ij}(\boldsymbol{\beta}^*) &= \frac{\boldsymbol{W}_{ij}\delta_j I[X_i \geq X_j]}{G^{(1)}G^{(2)}}\left[\frac{G^{(1)} - \hat{G}^{(1)}}{G^{(1)}} + \frac{G^{(2)} - \hat{G}^{(2)}}{G^{(2)}}\right] \\ &\quad - \frac{\boldsymbol{W}_{ij}\theta(\boldsymbol{W}_{ij}\boldsymbol{\beta}^*)}{G^{(3)}G^{(4)}}\left[\frac{G^{(3)} - \hat{G}^{(3)}}{G^{(3)}} + \frac{G^{(4)} - \hat{G}^{(4)}}{G^{(4)}}\right]. \end{aligned}$$

Following the results in Wang et al. (2013), we have

$$\frac{G(t_1,t_2) - \hat{G}(t_1,t_2)}{G(t_1,t_2)} = n^{-1}\sum_{k=1}^{n} M_k(z;\alpha) + o_p(1),$$

where $\alpha = t_2/t_1$, $z = \sqrt{t_1^2 + t_2^2}$ and $M_k(z;\alpha)$, given α, is a zero-mean martingale. Then we have

$$n^{1/2}\boldsymbol{U}(\boldsymbol{\beta}^*;\hat{G}) = n^{-5/2}\sum_{i,j,k=1}^{n}\big[\boldsymbol{e}_{ij}(\boldsymbol{\beta}^*;G) + \boldsymbol{\varsigma}_{ijk}(\boldsymbol{\beta}^*, G, M_k)\big] + o_p(\mathbf{1}),$$

where

$$\varsigma_{ijk}(\boldsymbol{\beta}^*, G, M_k) = \frac{\boldsymbol{W}_{ij}\delta_j I[X_i \geq X_j]}{G^{(1)}G^{(2)}}\Big[M_k(Z_{ij};\alpha_{ij}) + M_k(Z_{jj};\alpha_{jj})\Big] - \frac{\boldsymbol{W}_{ij}\theta(\boldsymbol{W}_{ij}\boldsymbol{\beta}^*)}{G^{(3)}G^{(4)}}\Big[M_k(R_i;0) + M_k(R_j;0)\Big].$$

Thus $n^{1/2}\boldsymbol{U}(\boldsymbol{\beta}^*;\hat{G})$ is a U-statistic and when $n \to \infty$, $n^{1/2}\boldsymbol{U}(\boldsymbol{\beta}^*;\hat{G}) \to N(\boldsymbol{0}, \boldsymbol{\Sigma}_{\boldsymbol{U}})$, where

$$\boldsymbol{\Sigma}_{\boldsymbol{U}} = \text{Var}\Big\{n^{-5/2}\sum_{i,j,k=1}^{n}\big[\boldsymbol{e}_{ij}(\boldsymbol{\beta}^*;G) + \varsigma_{ijk}(\boldsymbol{\beta}^*, G, M_k)\big]\Big\}.$$

Then the theorem follows from the first-order Taylor expansion for a vector field

$$\boldsymbol{U}(\boldsymbol{\beta}^*;\hat{G}) \approx \boldsymbol{U}(\hat{\boldsymbol{\beta}};\hat{G}) + \boldsymbol{U}^{'}(\hat{\boldsymbol{\beta}};\hat{G})(\boldsymbol{\beta}^* - \hat{\boldsymbol{\beta}}).$$

□

Acknowledgement

Yanchun Bao was supported by the Economic and Social Research Council (ESRC) through the Research Centre on Micro-Social Change (MiSoC) at the University of Essex, grant number ES/L009153/1.

References

Altman, E. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *J. Finance*, *23*, 589–609.

Altman, E., Haldeman, R., & Narayanan, P. (1977). Zeta$^{\text{TM}}$ analysis: A new model to identify bankruptcy risk of corporations. *J. Bank. Financ.*, *1*, 29–54.

Bennett, S. (1983). Analysis of survival data by the proportional odds model. *Stat. Med.*, *2*, 273–277.

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

Cai, T., Wei, L., & Wilcox, M. (2000). Semiparametric regression analysis for clusterd failure time data. *Biometrika*, *87*, 867–878.

Campbell, G., & Földes, A. (1982). Large-sample properties of nonparametric bivariate estimatiors with censored data. *Proceedings of the International Colloquium on Nonparametric Statistical Inference*, *32*, 103–121.

Cheng, S., Wei, L., & Ying, Z. (1995). Analysis of transformation models with censored data. *Biometrika*, *82*, 659–668.

Cheng, S., Wei, L., & Ying, Z. (1997). Predicting survival probabilities with semiparametric transformation models. *J. Amer. Statist. Assoc.*, *92*, 227–235.

Cox, D. R. (1972). Regression models and life tables. *J. Roy. Statist. Soc. Ser. B.*, *34*, 187–202.

Cox, D. R., & Oakes, D. (1984). *Analysis of Survival Data*. Chapman and Hall, London.

Dabrowska, D. (1988). Kaplan-meier estimate on the plane. *Ann. Stat.*, *16*, 1475–1489.

Dabrowska, D. (1989). Kaplan-meier estimate on the plane: weak convergence, lil, and the bootstrap. *J. Multivariate Anal.*, *29*, 308–325.

Dabrowska, D., & Doksum, K. (1988). Estimation and testing in the two-sample generalized odds-rate model. *J. Amer. Statist. Assoc.*, *83*, 744–749.

Dai, H., & Bao, Y. (2009). An inverse probability weighted estimator for the bivariate distribution function under right censoring. *Stat. Probabil. Lett.*, *79*, 1789–1797.

Dai, H., & Fu, B. (2012). A polar coordinate transformation for estimating bivariate survival functions with randomly censored and truncated data. *J. Statist. Plann. Inference*, *142*, 248–262.

Dai, H., Restaino, M., & Wang, H. (2016). A class of nonparametric bivariate survival function estimators for randomly censored and truncated data. *J. Nonparametric. Stats.*, *28*, 736–751.

Dai, H., & Wang, H. (2016). *Analysis for time-to-event data under censoring and truncation*. Academic Press, Elsevier.

Dore, G., Freeman, A., Law, M., & Kaldor, J. (2002). Is severe liver disease a common outcome for people with chronic hepatitis c? *J. Gastroenterol. Hepatol.*, *17*, 423–430.

Fine, J., Ying, Z., & Wei, L. (1998). On the linear transformation model with censored data. *Biometrika*, *85*, 980–986.

Fu, B., Tom, B. D., Delahooke, T., Alexander, G. J., & Bird, S. M. (2007). Event-biased referral can distort estimation of hepatitis c virus progression rate to cirrhosis and of prognostic influences. *J. Clin. Epidemiol*, *60*, 1140–1148.

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

Gürler, U. (1996). Bivariate estimation with right truncated data. *J. Amer. Statist. Assoc.*, *91*, 1152–1165.

Gürler, U. (1997). Bivariate distribution and hazard functions when a component is randomly truncated. *J. Multivariate Anal.*, *60*, 20–47.

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

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

Laitinen, E. K., & Suvas, A. (2013). International applicability of corporate failure risk models based on financial statement information: comparisons across european countries. *J. Financ. Econ.*, *1*, 01–26.

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

Pettitt, A. N. (1984). Proportional odds model for survival data and estimates using ranks. *Appl. Statist.*, *33*, 169–175.

Sharma, N. K., & Sherker, A. H. (2010). Epidemiology, risk factors, and natural history of chronic hepatitis c. In *Clinical Gastroenterology* (pp. 33–70). Human Press.

Shen, P. (2006). An inverse-probability-weighted approach to estimation of the bivariate survival function under left-truncation and right-censoring. *J. Statist. Plann. Inference*, *136*, 4365–4384.

Shen, P., & Yan, Y. (2008). Nonparametric estimation of the bivariate survival function with left-truncated and right-censored data. *J. Statist. Plann. Inference*, *138*, 4041–4054.

Situm, M. (2014). The age and size of the firm as relevant predictors for bankruptcy. *Journal of Applied Economics and Business*, *2*, 5–30.

Wang, H., Dai, H., & Fu, B. (2013). Accelerated failure time models for censored survival data under referral bias. *Biostatistics*, *14*, 313–326.

Wang, Y., & Zhu, M. (2006). Rank-based regression for analysis of repeated measures. *Biometrika*, *93*, 459–464.