South African Statist. J. (2017) **51**, 1 – 18

LINEAR REGRESSION WITH RANDOMLY DOUBLE-TRUNCATED DATA

G. Frank [1]
Faculty for Economic and Social Sciences, University of Rostock
e-mail: *gordon.frank@uni-rostock.de*

A. Dörre
Faculty for Economic and Social Sciences, University of Rostock
e-mail: *achim.doerre@uni-rostock.de*

Key words: Insolvency risk, Linear regression, Non-parametric, Random double-truncation.

Abstract: Non-parametric estimation for a linear regression model under random double-truncation is investigated, i.e. the variables are observed if and only if the dependent variable lies in a random interval. The method requires only weak distribution assumptions to ensure identifiability, but does not require any specific distribution family for any variable, neither for the truncation variables nor for the error term. By using non-parametric estimators of several distribution functions, consistent and asymptotically normal estimators are established. A simulation study shows the tendency that the lower the probability of observation, the higher the mean squared error of the estimators, even for the same number of observations. Finally, the method is applied to a doubly truncated data set of German companies, where the age-at-insolvency is of interest.

1. Introduction

Truncation of data occurs if the event of interest is only recorded within a certain range and is otherwise not observed. For instance, cohort studies for which not all events are recorded due to a limited observation span of the study, constitute an important case of double-truncation. More generally, the variable of interest $\widetilde{Y}$ is observed if and only if it is in a random interval $[\widetilde{T}, \widetilde{T}+\widetilde{D}]$. Consider the following application.

Example: Insolvency of German companies

For German companies, the age-at-insolvency, $\widetilde{Y}$, is of interest, which possibly depends on several covariates $\widetilde{Z}$. Due to German law, an insolvent company is announced as such publicly only for a few days. Afterwards, the information is no longer publicly available. Let $\widetilde{T}$ be the age of a company at the starting date of observation. The observation span is denoted by $\widetilde{D}$ and is assumed be a known

[1] Corresponding author.

AMS: 62G05, 62G20, 62J05

constant. Thus, an insolvent company is observed if and only if $\widetilde{Y} \in [\widetilde{T}, \widetilde{T} + \widetilde{D}]$ (see Figure 1 for an illustration of the truncation mechanism).

In this article, it is assumed that the variable of interest is the outcome of a linear regression function and is subject to random double-truncation. Linear regression under left-truncation, which is a special case by setting $\widetilde{D} \equiv \infty$, has been studied by Bhattacharya, Chernoff and Yang (1983), Gross and Huber-Carol (1992), Gross and Lai (1996) and He and Yang (2003). Regarding random double-truncation, general regression problems have received more attention recently (see e.g. Shen, 2013; Shen, 2015; Moreira, Uña-Álvarez and Meira-Machado, 2016).

Here, a non-parametric estimation method for the regression coefficients is described, which employs the non-parametric maximum likelihood estimator (NPMLE) for distribution functions under random double-truncation. The NPMLE for the event-time distribution was introduced by Efron and Petrosian (1999). Shen (2010) added the NPMLE for the truncation distributions, which is needed for the proposed estimator, and derived asymptotic properties. Even though a closed-form variance estimator was not established then, this issue was investigated and solved for the event-time distribution by Emura, Konno and Michimae (2015). Furthermore, Zhang (2015) derived a variance estimator for both NPMLEs under the assumption that the lower and upper truncation variables are independent. However, there are many doubly truncated data sets where the upper truncation variable is the lower truncation variable plus a constant (see Section 5 or e.g. Moreira and Uña-Álvarez, 2010; Kalbfleisch and Lawless, 1989). Since there is no general closed-form variance estimator for the distribution of $\widetilde{T}$ and $\widetilde{T} + \widetilde{D}$, the bootstrap is applied.

The fundamental idea for the estimator goes back to Stute (1993b) where linear regression coefficients are estimated for right-censored data by considering weighted moments of covariates. He and Yang (2003) used this approach to establish an estimator for linear regression coefficients under left-truncation.

The general modelling approach towards random double-truncation is to introduce variables $\widetilde{T}, \widetilde{D}$ and $\widetilde{Y}$ with corresponding covariates $\widetilde{Z}$ and random errors $\widetilde{\varepsilon}$, and to let the observed sample (Y, Z, T, D) comprise all units that fulfil the observation criterion $\widetilde{T} \leq \widetilde{Y} \leq \widetilde{T} + \widetilde{D}$. By this construction, the size of the observed sample is random.

In Section 2, the regression model is specified and the modelling of the random sample size via random point measures is explained. In addition, the estimators are motivated and derived. The consistency and asymptotic normality are proved for the estimators of the regression coefficients in Section 3. In order to assess the finite sample performance, a simulation study is presented in Section 4. Finally, in Section 5, the estimation method is applied to a data set of German companies whose age-at-insolvency was recorded if and only if their date of insolvency fell into a particular observation span.

2. Regression Model

Let $(\widetilde{Y}_i, \widetilde{T}_i, \widetilde{D}_i, \widetilde{\varepsilon}_i, \widetilde{Z}_{i,1}, \dots, \widetilde{Z}_{i,k}) : \left(\bigtimes_{i=1}^{k+4} \Omega_i, \mathscr{A}, P \right) \to (\mathbb{R}^{k+4}, \mathscr{B}(\mathbb{R}^{k+4}))$ for $i = 1, \dots, n$ be i.i.d. random variables, where $\mathscr{B}(\mathbb{R}^{k+4})$ is the Borel σ-algebra of $\mathbb{R}^{k+4}$. For any random variable X, denote the associated c.d.f. by F^X. The omission of indices for random variables symbolizes the related random vector, e.g. $\widetilde{Y} := (\widetilde{Y}_1, \dots, \widetilde{Y}_n)^t$. In addition, $\widetilde{\mathbf{Z}}$ denotes the $n \times k$ matrix that contains for every

$j = 1,\dots,k$, $(\widetilde{Z}_{1,j},\dots,\widetilde{Z}_{n,j})^t$ as a column. Therefore, $\widetilde{\mathbf{Z}}_i$ is the ith row of $\widetilde{\mathbf{Z}}$. In the following, the regression model

$$\widetilde{Y} = \widetilde{\mathbf{Z}}\beta + \widetilde{\varepsilon}$$

is investigated, where $\beta := (\beta_1,\dots,\beta_k)^t$ is the parameter vector of interest. Denote by a_F the lower and by b_F the upper endpoint of a distribution function F, i.e. $a_F := \inf\{x|F(x) > 0\}$ and $b_F := \sup\{x|F(x) < 1\}$. The following assumptions are made for the model.

(A1) The distributions of the truncation variables $\widetilde{T}_i$ and $\widetilde{D}_i$ are continuous and $P(\widetilde{D}_i > 0) = 1$.

(A2) Every component of $(\widetilde{\varepsilon}_i, \widetilde{Z}_{i,1},\dots,\widetilde{Z}_{i,k})$ is square-integrable.

(A3) $\widetilde{Y}_i$ and $(\widetilde{T}_i, \widetilde{D}_i)$ are quasi-independent.

(A4) $\mathrm{E}(\widetilde{\varepsilon}_i) = 0$ and $\mathrm{Cov}(\widetilde{\mathbf{Z}}_i, \widetilde{\varepsilon}_i) = 0$.

(A5) $\alpha := P(\widetilde{T}_i \le \widetilde{Y}_i \le \widetilde{T}_i + \widetilde{D}_i) > 0$.

(A6) $a_{F^{\widetilde{T}}} \le a_{F^{\widetilde{Y}}} \le a_{F^{\widetilde{T}+\widetilde{D}}}$ and $b_{F^{\widetilde{T}}} \le b_{F^{\widetilde{Y}}} \le b_{F^{\widetilde{T}+\widetilde{D}}}$.

From a practical point of view, most of these assumptions are not critical. In many applications, (A1) is fulfilled because the truncation variables $\widetilde{T}_i$ and $\widetilde{D}_i$ represent time which is continuous. Regarding $\widetilde{D}_i$, the application in this paper and also many applications in the literature of random double-truncation (see e.g. Kalbfleisch and Lawless, 1989; Moreira and Uña-Álvarez, 2010) in fact deal with the case that $\widetilde{D}_i$ is a known constant. This even simplifies calculations and is discussed in Section 5. Square-integrability of covariates and error terms, as required by (A2), is a classical assumption and not an issue for real data sets. The same holds for the independence of the error term and covariates, i.e. (A4). In contrast, quasi-independence (definition in Tsai, 1990) of response and truncation variables is not obviously fulfilled. If the sampling mechanism does not imply this assumption, it is recommended to use appropriate tests (e.g. Martin and Betensky, 2005; Emura and Wang, 2010) in order to check (A3). Assumption (A6) is important to ensure that the whole support can be observed and hence the distribution of interest is identifiable. For non-parametric methods under random double-truncation, there is no workaround for this assumption. Note that (A5) is implied by (A6) and only stated for easier argumentation in the following sections.

2.1. Modelling of random number of observations

Due to random double-truncation, a vector $(\widetilde{Y}_i, \widetilde{T}_i, \widetilde{D}_i, \widetilde{\mathbf{Z}}_i)$ is observed if and only if $\widetilde{T}_i \le \widetilde{Y}_i \le \widetilde{T}_i + \widetilde{D}_i$. This truncation mechanism results in a random number of observations. In order to avoid technical difficulties and obtain independence of the observed quadruples, thinned binomial processes and the results of Reiss (1993) are used. For this purpose, random point measures are introduced first.

Let J be any subset of $\mathbb{N}$ and for each $l \in J$, let ε_{r_l} be the Dirac measure concentrated at $r_l \in \mathbb{R}^3$. Then, $\nu := \sum_{l\in J} \varepsilon_{r_l}$ is called a point measure. Denote the space of point measures on $\mathbb{R}^3$ by $\mathbb{M}(\mathbb{R}^3, \mathscr{B}(\mathbb{R}^3))$. The associated σ-algebra $\mathscr{M}(\mathbb{R}^3, \mathscr{B}(\mathbb{R}^3))$ for $\mathbb{M}(\mathbb{R}^3, \mathscr{B}(\mathbb{R}^3))$ contains subsets of $\mathbb{M}(\mathbb{R}^3, \mathscr{B}(\mathbb{R}^3))$ and is the smallest one such that the evaluation mappings $\pi_C : \mathbb{M}(\mathbb{R}^3, \mathscr{B}(\mathbb{R}^3)) \to \mathbb{N} \cup \{0\} \cup \{\infty\}$, $C \in \mathscr{B}(\mathbb{R}^3)$, with $\pi_C(\nu) = \nu(C)$ are measurable. The random measure $\widetilde{N} : \Omega_1 \times$

$\Omega_2 \times \Omega_3 \to \mathbb{M}(\mathbb{R}^3, \mathscr{B}(\mathbb{R}^3))$ is called a point process on $(\mathbb{R}^3, \mathscr{B}(\mathbb{R}^3))$ if it is measurable with respect to the σ-algebra of $\Omega_1 \times \Omega_2 \times \Omega_3$ and $\mathscr{M}(\mathbb{R}^3, \mathscr{B}(\mathbb{R}^3))$.
Specifically, for the considered model, $\widetilde{N}$ is given by

$$\widetilde{N}(\cdot) := \sum_{i=1}^{n} \varepsilon_{(\widetilde{Y}_i, \widetilde{T}_i, \widetilde{D}_i)}(\cdot \cap U),$$

where $U := \{(\widetilde{y}, \widetilde{t}, \widetilde{d}) | \widetilde{t} \leq \widetilde{y} \leq \widetilde{t} + \widetilde{d}\}$. Consequently, $\widetilde{N}(\mathbb{R}^3)$ is the random number of observations. In order to achieve a representation for which independence of the random sample size and the observations holds, consider a second point process

$$N(\cdot) := \sum_{i=1}^{\tau} \varepsilon_{(Y_i, T_i, D_i)}(\cdot),$$

where τ is binomially distributed with parameter vector (n, α) and independent of (Y_i, T_i, D_i). The distribution of the i.i.d. triples (Y_i, T_i, D_i) is the conditional distribution of $(\widetilde{Y}_i, \widetilde{T}_i, \widetilde{D}_i)$ given $\widetilde{T}_i \leq \widetilde{Y}_i \leq \widetilde{T}_i + \widetilde{D}_i$. It can be shown (see e.g. Reiss, 1993) that $\widetilde{N}$ and N are equal in distribution. Moreover, the point process N has always the argument $\mathbb{R}^3$. For this reason and the sake of simplicity, the argument of N is omitted throughout, i.e. $N \equiv N(\mathbb{R}^3)$ describes the random sample size.

2.2. Estimation

The ordinary least squares estimator for β can not be applied because $\mathrm{E}(\widetilde{Y}_i | \widetilde{\mathbf{Z}}_i, \widetilde{T}_i \leq \widetilde{Y}_i \leq \widetilde{T}_i + \widetilde{D}_i)$ is not linear in β. In particular,

$$\mathrm{E}(\varepsilon_i) = \mathrm{E}(\widetilde{\varepsilon}_i | \widetilde{\mathbf{Z}}_i, \widetilde{T}_i \leq \widetilde{Y}_i \leq \widetilde{T}_i + \widetilde{D}_i) \neq 0,$$

except in some special cases. However, it holds that

$$(\widetilde{\mathbf{Z}}^t \widetilde{\mathbf{Z}})\beta = \widetilde{\mathbf{Z}}^t \widetilde{Y} - \widetilde{\mathbf{Z}}^t \widetilde{\varepsilon} \implies n\mathrm{E}(\widetilde{\mathbf{Z}}_i^t \widetilde{\mathbf{Z}}_i)\beta = n\mathrm{E}(\widetilde{\mathbf{Z}}_i^t \widetilde{Y}_i).$$

In case of $\det(\mathrm{E}(\widetilde{\mathbf{Z}}_i^t \widetilde{\mathbf{Z}}_i)) \neq 0$, this implies

$$\beta = \mathrm{E}(\widetilde{\mathbf{Z}}_i^t \widetilde{\mathbf{Z}}_i)^{-1} \mathrm{E}(\widetilde{\mathbf{Z}}_i^t \widetilde{Y}_i).$$

The main idea for estimating β is thus to replace all true expectations by their estimated counterparts. Since $\widetilde{T}_i \leq \widetilde{Y}_i \leq \widetilde{T}_i + \widetilde{D}_i \iff \widetilde{Y}_i - \widetilde{D}_i \leq \widetilde{T}_i \leq \widetilde{Y}_i$, (A3) implies that

$$\begin{aligned}
F^{Y,\mathbf{Z}}(y, \mathbf{z}) &= P\left(\widetilde{Y}_i \leq y, \widetilde{\mathbf{Z}}_i \leq \mathbf{z} \middle| \widetilde{T}_i \leq \widetilde{Y}_i \leq \widetilde{T}_i + \widetilde{D}_i\right) \\
&= \frac{1}{\alpha} P\left(\left\{\widetilde{Y}_i \leq y, \widetilde{\mathbf{Z}}_i \leq \mathbf{z}\right\} \cap \left\{\widetilde{T}_i \leq \widetilde{Y}_i \leq \widetilde{T}_i + \widetilde{D}_i\right\}\right) \\
&= \frac{1}{\alpha} \int_0^y \int_0^{\mathbf{z}} \int_0^\infty F^{\widetilde{T}}(s) - F^{\widetilde{T}}(s-u)\, dF^{\widetilde{D}}(u)\, dF^{\widetilde{Y},\widetilde{\mathbf{Z}}}(s, \mathbf{w})
\end{aligned}$$

and hence

$$dF^{\widetilde{Y},\widetilde{\mathbf{Z}}}(y, \mathbf{z}) = \frac{\alpha}{\int_0^\infty F^{\widetilde{T}}(y) - F^{\widetilde{T}}(y-u)\, dF^{\widetilde{D}}(u)} dF^{Y,\mathbf{Z}}(y, \mathbf{z}), \tag{1}$$

where $\mathbf{Z}$ denotes the matrix of all observed covariates. Note that this relation is the key equation to deal with the unknown joint distribution of $(\widetilde{Y}_i, \widetilde{\mathbf{Z}}_i)$. It holds that

$$\begin{aligned} \mathrm{E}\left(\widetilde{Z}_{i,l}\widetilde{Z}_{i,m}\right) &= \int_0^\infty z_l z_m \, dF^{\widetilde{Y},\widetilde{\mathbf{Z}}}(y,\mathbf{z}) \\ &\overset{(1)}{=} \alpha \int_0^\infty \frac{z_l z_m}{\int_0^\infty F^{\widetilde{T}}(y) - F^{\widetilde{T}}(y-u)\, dF^{\widetilde{D}}(u)} \, dF^{Y,\mathbf{Z}}(y,\mathbf{z}) \end{aligned} \tag{2}$$

and

$$\begin{aligned} \mathrm{E}\left(\widetilde{Y}_i\widetilde{Z}_{i,m}\right) &= \int_0^\infty y z_m \, dF^{\widetilde{Y},\widetilde{Z}}(y,\mathbf{z}) \\ &\overset{(1)}{=} \alpha \int_0^\infty \frac{y z_m}{\int_0^\infty F^{\widetilde{T}}(y) - F^{\widetilde{T}}(y-u)\, dF^{\widetilde{D}}(u)} \, dF^{Y,\mathbf{Z}}(y,\mathbf{z}), \end{aligned} \tag{3}$$

for $1 \le l, m \le k$. To establish estimates for $\mathrm{E}\left(\widetilde{Z}_{i,l}\widetilde{Z}_{i,m}\right)$ and $\mathrm{E}\left(\widetilde{Y}_i\widetilde{Z}_{i,m}\right)$, several distribution estimates are necessary. Regarding $F^{Y,\mathbf{Z}}$, the empirical distribution function $\widehat{F}^{Y,\mathbf{Z}}$ is a reasonable choice. Discrete non-parametric estimators for $F^{\widetilde{T}}$ and $F^{\widetilde{T}+\widetilde{D}}$ are provided by Shen (2010). Let $I_{\{\cdot\}}$ denote the indicator function which is one if the related argument is true and zero otherwise. Define $\widehat{k}^{(0)} := (1/N, \ldots, 1/N)$ and $\widehat{K}_j^{(r)} := \sum_{l=1}^N \widehat{k}_l^{(r)} I_{\{T_l \le Y_j \le T_l + D_l\}}, r \in \mathbb{N}$. Then, the estimate is calculated by iterating

$$\widehat{k}_i^{(r+1)} = \left(\sum_{j=1}^N \frac{I_{\{T_i \le Y_j \le T_i + D_i\}}}{\widehat{K}_j^{(r)}} \right)^{-1}, \quad i = 1, \ldots, N,$$

until reaching r^* such that for a chosen $\delta > 0$ it holds $\left|\widehat{k}_i^{(r^*+1)} - \widehat{k}_i^{(r^*)}\right| < \delta$. The estimators for the lower, upper and the joint truncation distribution are respectively given by

$$\begin{aligned} \widehat{F}^{\widetilde{T}}(t) &:= \sum_{i=1}^N \widehat{k}_i^{(r^*)} I_{\{T_i \le t\}}, \\ \widehat{F}^{\widetilde{D}}(s) &:= \sum_{i=1}^N \widehat{k}_i^{(r^*)} I_{\{D_i \le s\}}, \text{ and} \\ \widehat{F}^{\widetilde{T},\widetilde{T}+\widetilde{D}}(t,s) &:= \sum_{i=1}^N \widehat{k}_i^{(r^*)} I_{\{T_i \le t, T_i + D_i \le s\}}. \end{aligned}$$

In the setting of Shen (2010), $F^{\widetilde{Y}}$ is also non-parametrically estimated, say by $\breve{F}^{\widetilde{Y}}$. Technically seen, the proof in Shen (2010) only covers the uniform consistency of $\breve{F}^{\widetilde{Y}}$ for fixed $[a_{F^{\widetilde{Y}}}, t] \subset [0,\infty], t \in (a_{F^{\widetilde{Y}}}, b_{F^{\widetilde{Y}}})$ under the following additional assumptions:

(B1) $[a_{F^{\widetilde{Y}}}, t]$ is such that $F^{\widetilde{Y}}(v) - F^{\widetilde{Y}}(u-) > \delta > 0$ for $[u,v] \subset [a_{F^{\widetilde{Y}}}, t]$.

(B2) $\displaystyle\int_{a_{F^{\widetilde{Y}}}}^t \frac{dF^{\widetilde{Y}}(y)}{F^{\widetilde{T},\widetilde{T}+\widetilde{D}}(y,\infty) - F^{\widetilde{T},\widetilde{T}+\widetilde{D}}(y,y)} < \infty.$

(B3) $\frac{dF^{\widetilde{T},\widetilde{T}+\widetilde{D}}(y,\infty) - dF^{\widetilde{T},\widetilde{T}+\widetilde{D}}(y,y)}{dF^{\widetilde{Y}}(y)}$ is uniformly bounded on $[a_{F^{\widetilde{Y}}}, t]$.

However, the uniform consistency of $\check{F}^{\widetilde{Y}}$ implies the uniform consistency of $\widehat{F}^{\widetilde{T},\widetilde{T}+\widetilde{D}}$. For further details, see Appendix. Considering the assumptions, (B1) ensures that $F^{\widetilde{Y}}(v) - F^{\widetilde{Y}}(u-)$ is uniformly bounded away from zero. Condition (B2) holds if $a_{F^{\widetilde{T}}} < a_{F^{\widetilde{Y}}}$ and $a_{F^{\widetilde{Y}}} \leq a_{F^{\widetilde{T}+\widetilde{D}}}$. Therefore (A6) and $a_{F^{\widetilde{T}}} \neq a_{F^{\widetilde{Y}}}$ imply (B2). The last assumption (B3) holds if the density of $(\widetilde{T}, \widetilde{T}+\widetilde{D})$ is bounded from above and the density of $\widetilde{Y}$ is positive on $[a_{F^{\widetilde{Y}}}, t]$. Considering real data sets, assumptions (B1) and (B3) are not critical. In many applications, for (B1) and (B3) it only has to be assumed that the distribution of $\widetilde{Y}$ has no inner intervals where the density is zero.

In the following, bracketed indices stand for an ascending order of the related vector. Using the established estimator for the truncation distribution, the observation probability α is estimated by

$$\begin{aligned}\widehat{\alpha} &= \left(\int_0^\infty \frac{1}{\int_0^\infty \widehat{F}^{\widetilde{T}}(y) - \widehat{F}^{\widetilde{T}}(y-u)\, d\widehat{F}^{\widetilde{D}}(u)}\, d\widehat{F}^{Y,\mathbf{Z}}(y) \right)^{-1} \\ &= N \left(\sum_{r=1}^N \frac{1}{\sum_{s=1}^N \left[\widehat{F}^{\widetilde{T}}(Y_r) - \widehat{F}^{\widetilde{T}}(Y_r - D_{(s)}) \right] \cdot \widehat{f}_s^{\widetilde{D}}} \right)^{-1},\end{aligned}$$

where $\widehat{f}_s^{\widetilde{D}} := \widehat{F}^{\widetilde{D}}(D_{(s)}) - \widehat{F}^{\widetilde{D}}(D_{(s-1)}), \widehat{F}^{\widetilde{D}}(D_{(0)}) := 0$. Plugging $\widehat{\alpha}$, $\widehat{F}^{Y,\mathbf{Z}}$, $\widehat{F}^{\widetilde{T}}$ and $\widehat{F}^{\widetilde{D}}$ into (2) yields

$$\begin{aligned}\widehat{\mathrm{E}}\left(\widetilde{Z}_{i,l}\widetilde{Z}_{i,m}\right) &:= \widehat{\alpha} \int_0^\infty \frac{z_l z_m}{\int_0^\infty \widehat{F}^{\widetilde{T}}(y) - \widehat{F}^{\widetilde{T}}(y-u)\, d\widehat{F}^{\widetilde{D}}(u)}\, d\widehat{F}^{Y,\mathbf{Z}}(y,\mathbf{z}) \\ &= \frac{\widehat{\alpha}}{N} \sum_{r=1}^N \frac{Z_{r,l} Z_{r,m}}{\sum_{s=1}^N \left[\widehat{F}^{\widetilde{T}}(Y_r) - \widehat{F}^{\widetilde{T}}(Y_r - D_{(s)}) \right] \cdot \widehat{f}_s^{\widetilde{D}}}.\end{aligned}$$

The estimated expectation $\widehat{\mathrm{E}}\left(\widetilde{Y}_i\widetilde{Z}_{i,m}\right)$ is defined analogously via (3). For $1 \leq l, m \leq k$, let

$$\begin{aligned}\widehat{\mathrm{E}}\left(\widetilde{\mathbf{Z}}_i^t\widetilde{\mathbf{Z}}_i\right)_{l,m} &:= \widehat{\mathrm{E}}\left(\widetilde{Z}_{i,l}\widetilde{Z}_{i,m}\right) \text{ and} \\ \widehat{\mathrm{E}}\left(\widetilde{\mathbf{Z}}_i^t\widetilde{Y}_i\right)_m &:= \widehat{\mathrm{E}}\left(\widetilde{Y}_i\widetilde{Z}_{i,m}\right),\end{aligned}$$

which is well-defined due to (A2). Finally, the estimator for β is given by

$$\widehat{\beta} := \widehat{\mathrm{E}}\left(\widetilde{\mathbf{Z}}_i^t\widetilde{\mathbf{Z}}_i\right)^{-1} \widehat{\mathrm{E}}\left(\widetilde{\mathbf{Z}}_i^t\widetilde{Y}_i\right), \tag{4}$$

if the inverse of $\widehat{\mathrm{E}}\left(\widetilde{\mathbf{Z}}_i^t\widetilde{\mathbf{Z}}_i\right)$ exists. Note that for the calculation of $\widehat{\beta}$, the factor $\widehat{\alpha}/N$ can be omitted because it gets cancelled out in (4). Regarding the variance of $\widetilde{\varepsilon}_i$, the estimator is defined by

$$\widehat{\mathrm{Var}}(\widetilde{\varepsilon}_i) := \widehat{\mathrm{E}}(\widetilde{Y}_i^2) - \left[\widehat{\mathrm{E}}\left(\widetilde{\mathbf{Z}}_i^t\widetilde{Y}_i\right)\right]^t \widehat{\beta},$$

where $\widehat{\mathrm{E}}(\widetilde{Y}_i^2)$, or any other estimated moment, is defined analogously to $\widehat{\mathrm{E}}\left(\widetilde{\mathbf{Z}}_i^t\widetilde{\mathbf{Z}}_i\right)$ and $\widehat{\mathrm{E}}\left(\widetilde{\mathbf{Z}}_i^t\widetilde{Y}_i\right)$ by interchanging z_m with y in (3). In addition, the distribution of $\widetilde{\varepsilon}$ can be estimated by

$$\widehat{F}^{\widetilde{\varepsilon}}(x) := \frac{\widehat{\alpha}}{N} \sum_{r=1}^N \frac{I_{\{Y_r - \mathbf{Z}_r\widehat{\beta} \leq x\}}(Y_r, \mathbf{Z}_r)}{\sum_{s=1}^N \left[\widehat{F}^{\widetilde{T}}(Y_r) - \widehat{F}^{\widetilde{T}}(Y_r - D_{(s)}) \right] \cdot \widehat{f}_s^{\widetilde{D}}},$$

as well as the coefficient of determination by

$$\widehat{R}^2 := 1 - \frac{\widehat{\mathrm{Var}}(\widetilde{\varepsilon}_i)}{\widehat{\mathrm{E}}(\widetilde{Y}_i^2) - \widehat{\mathrm{E}}(\widetilde{Y}_i)^2}.$$

3. Asymptotic Properties

In order to prove the consistency and asymptotic normality of $\widehat{\beta}$ under the assumptions (A1)-(A6) and (B1)-(B3), it will be shown first that $\widehat{\mathrm{E}}\left(\widetilde{\mathbf{Z}}_i^t\widetilde{\mathbf{Z}}_i\right)$ and $\widehat{\mathrm{E}}\left(\widetilde{\mathbf{Z}}_i^t\widetilde{Y}_i\right)$ are consistent and asymptotically normal. This is carried out with a mapping theorem for weak convergence and the central limit theorem. Afterwards, the multivariate delta method is applied to complete the proof. The consistency of the other estimators follows analogously.

For the sake of clarity, the generalised mapping theorem is stated first (Billingsley, 1968, p. 34).

Mapping Theorem Let h_n and h be measurable mappings from Ω to Ω', P_n,P probability measures on Ω and E be the set of x such that $h_n(x_n) \to h(x)$ fails to hold for some sequence $\{x_n\}$ approaching x. If $P_n \to P$ and $P(E) = 0$, then $P_n h_n^{-1} \to Ph^{-1}$.

To remove ambiguities, let f be a real function on Ω'. Then f is integrable w.r.t. Ph^{-1} if and only if fh is integrable w.r.t. P and, by definition, it holds that (Billingsley, 1968, p. 223)

$$\int f(h(x))dP(x) = \int f(x')dPh^{-1}(x').$$

Theorem 1 Under the model assumptions made in Section 2, it holds that

$$\widehat{\mathrm{E}}\left(\widetilde{\mathbf{Z}}_i^t\widetilde{\mathbf{Z}}_i\right) \xrightarrow{D} \mathrm{E}\left(\widetilde{\mathbf{Z}}_i^t\widetilde{\mathbf{Z}}_i\right) \text{ and}$$
$$\widehat{\mathrm{E}}\left(\widetilde{\mathbf{Z}}_i^t\widetilde{Y}_i\right) \xrightarrow{D} \mathrm{E}\left(\widetilde{\mathbf{Z}}_i^t\widetilde{Y}_i\right),$$

where $\xrightarrow{D}$ denotes convergence in distribution.

Proof. The proof consists of two steps, each using the above-mentioned mapping theorem. In both steps, $\Omega' := \mathbb{R}_+$ whereas h_n, h, Ω, E, P_n and P change. Let $y \in [a_{F^{\widetilde{Y}}}, b_{F^{\widetilde{Y}}}]$ be fixed, then

$$\begin{aligned}\int_0^\infty \widehat{F}^{\widetilde{T}}(y) - \widehat{F}^{\widetilde{T}}(y-x)d\widehat{F}^{\widetilde{D}}(x) &= \int_0^\infty h_n^{(1)}(x)dP_n^{(1)}(x)\\ &\xrightarrow{n\to\infty} \int_0^\infty h^{(1)}(x)dP^{(1)}(x)\\ &= \int_0^\infty F^{\widetilde{T}}(y) - F^{\widetilde{T}}(y-x)dF^{\widetilde{D}}(x),\end{aligned}$$

where $\Omega^{(1)} := \mathbb{R}_+$, $h^{(1)}(x) := F^{\widetilde{T}}(y) - F^{\widetilde{T}}(y-x)$, $h_n^{(1)}(x) := \widehat{F}^{\widetilde{T}}(y) - \widehat{F}^{\widetilde{T}}(y-x)$, $P^{(1)} := F^{\widetilde{D}}$ and $P_n^{(1)} := \widehat{F}^{\widetilde{D}}$. Because of the uniform consistency of $F^{\widetilde{T}}$, it holds that $E^{(1)} = \emptyset$. For $l, m \in \{1, \ldots, k\}$, let

$$h_n^{(2)}(y, \mathbf{z}) := \frac{z_l z_m}{\int_0^\infty \widehat{F}^{\widetilde{T}}(y) - \widehat{F}^{\widetilde{T}}(y-x)d\widehat{F}^{\widetilde{D}}(x)} \text{ and}$$
$$h^{(2)}(y, \mathbf{z}) := \frac{z_l z_m}{\int_0^\infty F^{\widetilde{T}}(y) - F^{\widetilde{T}}(y-x)dF^{\widetilde{D}}(x)},$$

where $\Omega^{(2)} := \mathbb{R}^1 \times \mathbb{R}^k$. Assumption (A5) is necessary to ensure that $\int_0^\infty F^{\widetilde{T}}(y) - F^{\widetilde{T}}(y-x) dF^{\widetilde{D}}(x) > 0$. The same inequality is true for the related estimator if there is at least one observation. Therefore, the continuous mapping theorem with the function $x \mapsto 1/x$ implies $h_n^{(2)}(y,\mathbf{z}) \xrightarrow{n\to\infty} h^{(2)}(y,\mathbf{z})$. In addition, $E^{(2)} = (-\infty, a_{F^{\widetilde{T}}}) \times \emptyset$ which is a null set as long as assumption (A6) holds. Choosing $P_n^{(2)} := \widehat{F}^{Y,\mathbf{Z}}$ and $P^{(2)} := F^{Y,\mathbf{Z}}$ allows for the second use of the mapping theorem. To complete the proof, the estimator $\widehat{\alpha}$ needs to be consistent. However, just consider the special case of $h_n^{(2)}(y,\mathbf{z})$ where $z_l z_m = 1$ to show that $1/\widehat{\alpha} \to 1/\alpha$. Then, the application of the continuous mapping theorem implies the consistency of $\widehat{\alpha}$. Again, $\widehat{\alpha}$ can not be zero if there is at least one observation. Finally, Slutsky's theorem completes the proof, i.e.

$$\begin{aligned}
\widehat{\mathrm{E}}\left(\widetilde{\mathbf{Z}}_i^t \widetilde{\mathbf{Z}}_i\right)_{l,m} &= \widehat{\alpha} \cdot \int_0^\infty \frac{z_l z_m}{\int_0^\infty \widehat{F}^{\widetilde{T}}(y) - \widehat{F}^{\widetilde{T}}(y-x) d\widehat{F}^{\widetilde{D}}(x)} d\widehat{F}^{Y,\mathbf{Z}}(y,\mathbf{z}) \\
&\xrightarrow{n\to\infty} \alpha \cdot \int_0^\infty \frac{z_l z_m}{\int_0^\infty F^{\widetilde{T}}(y) - F^{\widetilde{T}}(y-x) dF^{\widetilde{D}}(x)} dF^{Y,\mathbf{Z}}(y,\mathbf{z}) \\
&= \mathrm{E}\left(\widetilde{\mathbf{Z}}_i^t \widetilde{\mathbf{Z}}_i\right)_{l,m}.
\end{aligned}$$

The proof of

$$\widehat{\mathrm{E}}\left(\widetilde{\mathbf{Z}}_i^t \widetilde{Y}_i\right) \xrightarrow{D} \mathrm{E}\left(\widetilde{\mathbf{Z}}_i^t \widetilde{Y}_i\right)$$

follows analogously by interchanging z_m and y. ■

Corollary 1 It holds that

(i) $\widehat{\mathrm{Var}}(\widetilde{\varepsilon}_i) \xrightarrow{D} \mathrm{Var}(\widetilde{\varepsilon}_i)$,

(ii) $\widehat{F}^{\widetilde{\varepsilon}} \xrightarrow{D} F^{\widetilde{\varepsilon}}$,

(iii) $\widehat{R}^2 \xrightarrow{D} R^2$.

Proof.

(i) Consider first that

$$\begin{aligned}
\mathrm{Var}(\widetilde{\varepsilon}_i) &= \mathrm{E}(\widetilde{Y}_i - \widetilde{\mathbf{Z}}_i\beta)^2 = \mathrm{E}\left(\widetilde{Y}_i^2 - 2\widetilde{Y}_i\widetilde{\mathbf{Z}}_i\beta + \beta^t\widetilde{\mathbf{Z}}_i^t\widetilde{\mathbf{Z}}_i\beta\right) \\
&= \mathrm{E}(\widetilde{Y}_i^2) - \mathrm{E}(\widetilde{Y}_i\widetilde{\mathbf{Z}}_i)\beta.
\end{aligned}$$

Interchanging y^2 with $z_l z_m$ in the proof of Theorem 1 yields $\widehat{\mathrm{E}}(\widetilde{Y}_i^2) \xrightarrow{n\to\infty} \mathrm{E}(\widetilde{Y}_i^2)$. Furthermore, the continuous mapping theorem applied to the continuous function $(r,s,t) \to r - st$ for $r \in \mathbb{R}, s \in \mathbb{R}^{1\times k}, t \in \mathbb{R}^{k\times 1}$ ensures the consistency.

(ii) It holds that

$$P(\widetilde{\varepsilon} \le t) = \mathrm{E}\left(I_{\{\widetilde{\varepsilon}\le t\}}\right) = \int_0^\infty I_{\{y - \mathbf{z}\beta \le t\}}\, dF^{\widetilde{Y},\widetilde{\mathbf{Z}}}(y,\mathbf{z}).$$

In addition, consider the interchange of $I_{\{\widetilde{\varepsilon}\le t\}}$ with $z_l z_m$ for a fixed $t \in [a_{F^{\widetilde{\varepsilon}}}, b_{F^{\widetilde{\varepsilon}}}]$ in the proof of Theorem 1. Then the proof is completely analogous.

(iii) Since

$$\widehat{R}^2 = 1 - \frac{\widehat{\mathrm{Var}}(\widetilde{\varepsilon}_i)}{\widehat{\mathrm{E}}(\widetilde{Y}_i^2) - \widehat{\mathrm{E}}(\widetilde{Y}_i)^2}$$

and $\widehat{\mathrm{Var}}(\widetilde{\varepsilon}_i)$, $\widehat{\mathrm{E}}(\widetilde{Y}_i^2)$ and $\widehat{\mathrm{E}}(\widetilde{Y}_i)^2$ are consistent, the continuous mapping theorem applied to $(r,s,t) \to 1 - \frac{r}{s-t}$ for $r,s,t \in \mathbb{R}, s \neq t$ yields the consistency for $\widehat{R}^2$.

■

Theorem 2 The estimators $\widehat{\mathrm{E}}\left(\widetilde{\mathbf{Z}}_i^t \widetilde{\mathbf{Z}}_i\right)_{l,m}$ and $\widehat{\mathrm{E}}\left(\widetilde{\mathbf{Z}}_i^t \widetilde{Y}_i\right)_m$ are asymptotically normal for $1 \leq l,m \leq k$.

Proof. Let

$$\begin{aligned}
\widehat{c}(Y_r) &:= \int_0^\infty \widehat{F}^{\widetilde{T}}(Y_r) - \widehat{F}^{\widetilde{T}}(Y_r - u)\, d\widehat{F}^{\widetilde{D}}(u) \text{ and} \\
c(Y_r) &:= \int_0^\infty F^{\widetilde{T}}(Y_r) - F^{\widetilde{T}}(Y_r - u)\, dF^{\widetilde{D}}(u).
\end{aligned}$$

Note that assumption (A5) ensures that $c(Y_r)$ and $\widehat{c}(Y_r)$ are always positive. Additionally, since $\widetilde{T}$ and $\widetilde{D}$ are continuous it follows that

$$\sup_Y |\widehat{c}(Y) - c(Y)| \to 0 \implies \sup_Y \left| \frac{1}{\widehat{c}(Y)} - \frac{1}{c(Y)} \right| \to 0.$$

Then, for $l,m \in \{1,\ldots,k\}$, it holds that

$$\begin{aligned}
&\left| \frac{1}{N} \sum_{r=1}^N \frac{Z_{r,l} Z_{r,m}}{\widehat{c}(Y_r)} - \frac{1}{N} \sum_{r=1}^N \frac{Z_{r,l} Z_{r,m}}{c(Y_r)} \right| \\
&\qquad = \left| \frac{1}{N} \sum_{r=1}^N Z_{r,l} Z_{r,m} \left(\frac{1}{\widehat{c}(Y_r)} - \frac{1}{c(Y_r)} \right) \right| \\
&\qquad \leq \frac{1}{N} \sum_{r=1}^N |Z_{r,l} Z_{r,m}| \left| \frac{1}{\widehat{c}(Y_r)} - \frac{1}{c(Y_r)} \right| \\
&\qquad \leq \underbrace{\sup_Y \left| \frac{1}{\widehat{c}(Y)} - \frac{1}{c(Y)} \right|}_{\to 0} \cdot \underbrace{\left(\frac{1}{N} \sum_{r=1}^N |Z_{r,l} Z_{r,m}| \right)}_{\to \mathrm{E}(|Z_l Z_m|)} \to 0.
\end{aligned}$$

The last step follows from Slutsky's theorem. This shows that both sums have the same asymptotic distribution. Since the latter sum, i.e.

$$\frac{1}{N} \sum_{r=1}^N \frac{Z_{r,l} Z_{r,m}}{c(Y_r)},$$

is a mean of i.i.d. random variables, the central limit theorem implies the normal distribution. Analogous argumentation for $\widehat{\mathrm{E}}\left(\widetilde{\mathbf{Z}}_i^t \widetilde{Y}_i\right)_m$ proves the theorem. ■

Corollary 2 Assume that $\det\left(\widehat{\mathrm{E}}\left(\widetilde{\mathbf{Z}}_i^t\widetilde{\mathbf{Z}}_i\right)\right) \neq 0$. Then the estimator $\widehat{\beta}$ is asymptotically normal and consistent.

Proof. Let $\breve{\mathbf{Z}}$ be a $\frac{k(k-1)}{2}+k$-dimensional vector, containing all elements of the upper triangular matrix of $\widehat{\mathrm{E}}\left(\widetilde{\mathbf{Z}}_i^t\widetilde{\mathbf{Z}}_i\right)$ and all elements of $\widehat{\mathrm{E}}\left(\widetilde{\mathbf{Z}}_i^t\widetilde{Y}_i\right)$. Theorem 1 and Theorem 2 ensure that every component of $\breve{\mathbf{Z}}$ is consistent and asymptotically normal. Moreover, let W be the function such that

$$W(\breve{\mathbf{Z}}) := \widehat{\mathrm{E}}\left(\widetilde{\mathbf{Z}}_i^t\widetilde{\mathbf{Z}}_i\right)^{-1}\widehat{\mathrm{E}}\left(\widetilde{\mathbf{Z}}_i^t\widetilde{Y}_i\right).$$

The function W is differentiable with respect to its arguments and its gradient is never zero as long as $\det\left(\widehat{\mathrm{E}}\left(\widetilde{\mathbf{Z}}_i^t\widetilde{\mathbf{Z}}_i\right)\right) \neq 0$. This follows from the fact that for any matrix A with full rank, its inverse is given by $A^{-1} = \frac{1}{\det(A)}\mathrm{adj(A)}$, where adj(A) is the adjugate of matrix A. Applying the multivariate delta method proves the corollary (van der Vaart, 1998, p. 26).

■

Remark. In the previous corollary it was assumed that the determinant of the estimated expectation of $\widetilde{\mathbf{Z}}_i^t\widetilde{\mathbf{Z}}_i$ does not vanish. In order to weaken this assumption, it is interesting to know whether $\mathrm{Var}(\widetilde{Z}_{i,j}) > 0$ for all j already implies $\det\left(\widehat{\mathrm{E}}\left(\widetilde{\mathbf{Z}}_i^t\widetilde{\mathbf{Z}}_i\right)\right) \neq 0$. However, this implication does not even hold for the limiting determinant $\det\left(\mathrm{E}\left(\widetilde{\mathbf{Z}}_i^t\widetilde{\mathbf{Z}}_i\right)\right)$ and is therefore not true in general. In particular,

$$\begin{aligned}\det\left(\mathrm{E}\left(\widetilde{\mathbf{Z}}_i^t\widetilde{\mathbf{Z}}_i\right)\right) = &\sum_{v=0}^{k-2}(-1)^{k+1+v}(k-1-v)\\ &\cdot\sum_{\substack{S\subseteq\{1,\ldots,k\}\\|S|=v}}\ \prod_{m\in\{1,\ldots,k\}\setminus S}(\mathrm{E}\widetilde{Z}_{i,m})^2\prod_{l\in S}\mathrm{E}\widetilde{Z}_{i,l}^2\\ &+\prod_{h=1}^{k}\mathrm{E}\widetilde{Z}_{i,h}^2\end{aligned}$$

which can be derived by elementary calculations. Further consideration shows that this determinant is zero if

$$\frac{\mathrm{E}\widetilde{Z}_{i,j}^2}{(\mathrm{E}\widetilde{Z}_{i,j})^2} = \frac{\kappa_{1,j}}{\kappa_{2,j}+(-1)^k\prod_{h\in\{1,\ldots,k\}\setminus\{j\}}\mathrm{E}\widetilde{Z}_{i,h}^2},$$

for some j, where

$$\kappa_{r,j} = \sum_{v=0}^{k-r-1}(-1)^v(k-r-v)\sum_{\substack{S\subseteq\{1,\ldots,k\}\setminus\{j\}\\|S|=v}}\ \prod_{\substack{m\in\{1,\ldots,k\}\\m\notin(S\cup\{j\})}}(\mathrm{E}\widetilde{Z}_{i,m})^2\prod_{l\in S}\mathrm{E}\widetilde{Z}_{i,l}^2.$$

For instance, for $k = 3$ and $j = 1$, this expression simplifies to

$$\begin{aligned}\frac{\mathrm{E}\widetilde{Z}_{i,1}^2}{(\mathrm{E}\widetilde{Z}_{i,1})^2} &= \frac{(\mathrm{E}\widetilde{Z}_{i,2})^2\mathrm{Var}(\widetilde{Z}_{i,3}) + (\mathrm{E}\widetilde{Z}_{i,3})^2\mathrm{Var}(\widetilde{Z}_{i,2})}{\mathrm{Var}(\widetilde{Z}_{i,2}\widetilde{Z}_{i,3})} \\ &= \frac{(\mathrm{E}\widetilde{Z}_{i,2})^2\mathrm{Var}(\widetilde{Z}_{i,3}) + (\mathrm{E}\widetilde{Z}_{i,3})^2\mathrm{Var}(\widetilde{Z}_{i,2})}{(\mathrm{E}\widetilde{Z}_{i,2})^2\mathrm{Var}(\widetilde{Z}_{i,3}) + (\mathrm{E}\widetilde{Z}_{i,3})^2\mathrm{Var}(\widetilde{Z}_{i,2}) + \mathrm{Var}(\widetilde{Z}_{i,2})\mathrm{Var}(\widetilde{Z}_{i,3})} \\ &< 1\end{aligned}$$

which does not contradict $\mathrm{Var}(\widetilde{Z}_1) > 0$. It is therefore possible in some cases that $\det\left(\mathrm{E}\left(\widetilde{\mathbf{Z}}_i^t\widetilde{\mathbf{Z}}_i\right)\right) = 0$.

4. Finite Sample Properties

To analyse the finite sample properties of the estimators, two simulation studies are considered.

In the first study, $\widetilde{Y}_i$ has three covariates; $\widetilde{Z}_{i,1} \equiv 1$, $\widetilde{Z}_{i,2}$ following a (0.5)-Bernoulli distribution and $\widetilde{Z}_{i,3}$ is beta distributed with parameter vector $(1.5, 1.5)$. The corresponding regression coefficients are $\beta_1 = 0.4, \beta_2 = -0.2$ and $\beta_3 = 0.2$. The error variable follows a shifted $(0.3, 0.6)$-beta distribution with zero mean which is scaled to $[0, 0.2]$. This results in $\mathrm{Var}(\widetilde{\varepsilon}_i) \approx 0.0047$. The true R^2 is about 73%. Regarding the truncation variables, $\widetilde{T}_i$ is also beta distributed with 1 as the second shape parameter and scaled to $[0, 0.6]$, whereas $\widetilde{D}_i$ follows an exponential distribution. The first shape parameter of $\widetilde{T}_i$ and the parameter of $\widetilde{D}_i$ are adjusted to yield different values of α (see Table 1). Moreover, the sample size n was fitted to result in 200, 300 and 400 observations on average, namely EN. Every setting was repeated 10000 times. Table 2 displays the results.

Table 1: Configurations of parameters p and θ for truncation distributions $F^{\widetilde{T}}$ and $F^{\widetilde{D}}$ with associated observation probability α.

α	p for $\widetilde{T} \sim Beta(p, 1)$	θ for $\widetilde{D} \sim Exp(\theta)$
30%	2.5	2.6
40%	1.7	1.65
50%	1.15	1.13
60%	0.77	0.77
70%	0.5	0.5
80%	0.29	0.29

Due to consistency, the mean squared error (MSE) for every estimator decreases with increasing n. Clearly, the smaller α the higher the MSE, which was also observed in literature (see e.g. Shen, 2010; Moreira and Uña-Álvarez, 2010). This effect can not be seen for $\widehat{\alpha}$ because the change of α directly influences the calculation of the MSE of $\widehat{\alpha}$.

The second simulation study analyses the finite sample properties of the estimated standard error of $\widehat{\beta}$. Here, simple bootstrap is used, i.e. for a sample $(Y_i, T_i, D_i, \mathbf{Z}_i)$, $i = 1, \ldots, N$, there are $B \in \{200, 400, 1000\}$ resamples $(Y_{ib}, T_{ib}, D_{ib}, \mathbf{Z}_{ib})$, $b = 1, \ldots, B$, where each observation of the original sample has the same probability to get resampled. Then, the estimator is applied to all B resamples and the standard deviation of the B estimations is the estimate for the standard error. The

Table 2: Mean squared error for different observation probability α and mean number of observations EN.

α	EN	Mean squared error $\widehat{\beta}_1 \times 10^4$	$\widehat{\beta}_2 \times 10^4$	$\widehat{\beta}_3 \times 10^4$	$\widehat{\mathrm{Var}}(\widehat{\varepsilon}) \times 10^7$	$\widehat{\alpha} \times 10^3$	$\widehat{R}^2 \times 10^3$
30%	200	2.71	2.40	7.64	1.85	3.34	3.14
	300	1.83	1.74	5.17	1.21	2.45	2.24
	400	1.41	1.33	3.39	0.87	1.93	1.75
40%	200	1.81	1.48	5.00	1.35	3.89	1.42
	300	1.21	1.01	3.31	0.90	2.82	1.01
	400	0.93	0.75	2.58	0.67	2.16	0.79
50%	200	1.51	1.13	4.12	1.22	4.49	0.96
	300	1.10	0.76	2.76	0.77	3.01	0.65
	400	0.77	0.58	2.08	0.58	2.64	0.49
60%	200	1.44	1.02	4.00	1.14	4.30	0.78
	300	0.96	0.67	2.57	0.73	3.56	0.52
	400	0.72	0.51	1.95	0.56	2.58	0.39
70%	200	1.41	0.97	3.85	1.11	4.38	0.71
	300	0.93	0.64	2.51	0.73	3.27	0.47
	400	0.70	0.49	1.87	0.53	2.52	0.35
80%	200	1.37	0.93	3.77	1.10	3.83	0.65
	300	0.93	0.64	2.50	0.73	2.92	0.45
	400	0.69	0.46	1.85	0.54	2.36	0.33

simulation study considers three different observation probabilities, i.e. $\alpha \in \{0.25, 0.5, 0.75\}$. The lower truncation variable $\widetilde{T}_i$ follows a uniform distribution in contrast to $\widetilde{D}_i$ which is exponentially distributed. Table 3 displays the parameter settings for both distributions. The parameter of interest is $\beta' = (2, 0.5, 1)$ where $\widetilde{Z}_{i,1} \equiv 1$, $\widetilde{Z}_{i,2}$ again following a Bernoulli distribution with parameter 0.5 and $\widetilde{Z}_{i,3}$ is beta-distributed with parameter $(1.5, 1.5)$. The expected number of observations is 200 and every setting was repeated 1000 times. The error variable is uniformly distributed on $[-0.3, 0.3]$. Table 4 shows the simulation results. In this setting, the decrease of α has a similar effect on the MSE as in the first simulation study. Regarding the number of resamples B, even for $B = 200$ the estimates are quite good. The further increase in B reduces the MSE only slightly.

Table 3: Configurations of parameters r and θ for truncation distributions $F^{\widetilde{T}}$ and $F^{\widetilde{D}}$ with associated observation probability α.

α	r for $\widetilde{T} \sim Uni([0,r])$	θ for $\widetilde{D} \sim Exp(\theta)$
25%	4	0.91
50%	3	0.46
75%	2	0.165

The estimators were implemented in R. In order to achieve acceptable computational durations, it is strongly recommended to use C code in R which is possible by the use of the package `Rcpp`.

5. Application

For a data set of companies which were founded in the German federal state Hesse, the interest lies in the age-at-insolvency in days. The sample contains $N = 400$ companies which became insolvent

Table 4: Mean squared error of bootstrapped standard errors for β for different observation probabilities.

α	Number of resamples	Mean squared error $\widehat{s.e.}(\widehat{\beta}_1) \times 10^5$	$\widehat{s.e.}(\widehat{\beta}_2) \times 10^5$	$\widehat{s.e.}(\widehat{\beta}_3) \times 10^5$
25%	200	2.226	0.668	3.369
	400	1.960	0.497	2.856
	1000	1.468	0.456	2.317
50%	200	1.206	0.319	2.061
	400	1.058	0.255	1.794
	1000	1.011	0.195	1.459
75%	200	0.684	0.277	1.571
	400	0.573	0.189	1.249
	1000	0.489	0.146	1.040

during Aug-29 2013 to Mar-31 2014. This results in an observation period of $\widetilde{D} \equiv 214$ days. There were no observations with a foundation date after Aug-29 2013. Define $\widetilde{Y}_i$ as the age-at-insolvency of the ith company and $\widetilde{T}_i$ as the age at Aug-29 2013. Therefore, the ith company is observed if and only if $\widetilde{T}_i \leq \widetilde{Y}_i \leq \widetilde{T}_i + \widetilde{D}$. Figure 1 illustrates the truncation mechanism.

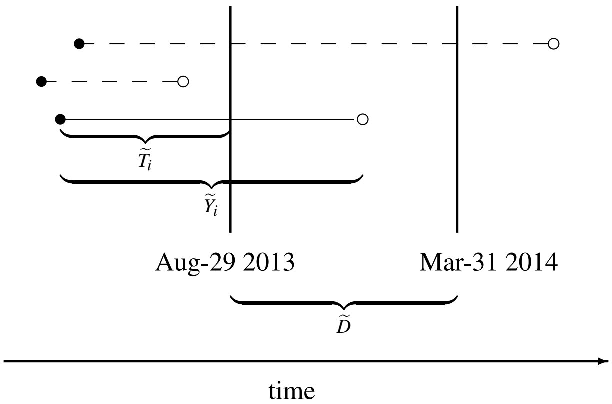

Figure 1: Three examples for foundation (black bullet) and insolvency (white bullet) of observed (solid) and truncated (dashed) companies.

The data set yields a special case for the established truncation model where $\widetilde{D}$ has only probability mass in 214. Shen (2010) did not explicitly cover this case, however, it was proven by Moreira and Uña-Álvarez (2010) that the asymptotic properties still hold. In this case, the calculation of the weights is slightly easier because

$$\int_0^\infty \widehat{F}^{\widetilde{T}}(y) - \widehat{F}^{\widetilde{T}}(y-u)\, d\widehat{F}^{\widetilde{D}}(u) = \widehat{F}^{\widetilde{T}}(y) - \widehat{F}^{\widetilde{T}}(y-\widetilde{D})$$

and hence the proofs are completely analogous. Due to the sampling mechanism, assumption (A3) is reasonable. The distribution assumption (A6) is fulfilled, because of

$$a_{F\tilde{T}} = 0 < a_{F\tilde{Y}} = 1 \ < a_{F\tilde{T}+D} = 214 \text{ and}$$
$$b_{F\tilde{T}} = \infty \leq b_{F\tilde{Y}} = \infty \leq b_{F\tilde{T}+D} = \infty.$$

Note that $a_{F\tilde{Y}} = 1$, because a company is assumed to survive at least for one day. This also implies the validity of assumption (B2). Every observation has the following 8 covariates:

$Z_{i,1}$ Always one for all observations.

$Z_{i,2}$ One if the company is a limited liability company, a limited partnership or a mixture. Otherwise zero.

$Z_{i,3}$ One if the observation is an entrepreneurial company with limited liability, otherwise zero.

$Z_{i,4}$ One if the company is part of the manufacturing sector, otherwise zero.

$Z_{i,5}$ One if the observation is a company of the building sector, otherwise zero.

$Z_{i,6}$ One if the observation is a company of the commerce sector, otherwise zero.

$Z_{i,7}$ One if the observation is a company of the maintenance sector, otherwise zero.

$Z_{i,8}$ One if the observation is a company of the car repair sector, otherwise zero.

Applying the proposed estimators and bootstrapping standard errors for confidence intervals yields Table 5. Note that the standard errors are bootstrapped in the same way as in the simulation study with 1000 resamples.

Table 5: Estimates with confidence intervals and standard errors.

	Value	90 % Confidence Intervals Lower Boundary	Upper Boundary
$\widehat{\beta}_1$	3923	2494	5351
$\widehat{\beta}_2$	96.1	-1273	1465
$\widehat{\beta}_3$	-3073	-4457	-1689
$\widehat{\beta}_4$	-717	-1667	232
$\widehat{\beta}_5$	1536	275	2796
$\widehat{\beta}_6$	218	-702	1140
$\widehat{\beta}_7$	-25.8	-968	944
$\widehat{\beta}_8$	636	-513	1785
		Estimated Standard Error	
$\widehat{\alpha}$	0.0101	0.0012	
$\widehat{n}$	39598	4192	
$\widehat{\mathrm{Var}}(\widetilde{\varepsilon})$	12381425	1476544	
$\widehat{R}^2$	0.054	0.040	

Even though the model cannot explain much of the variance ($\widehat{R}^2 \approx 5.54\%$), $\widehat{\beta}_1$, $\widehat{\beta}_3$ and $\widehat{\beta}_5$ are significantly (90%) different from zero. Therefore, entrepreneurial companies have a higher insolvency risk compared to other company forms. This is not surprising because an entrepreneurial company has insufficient share capital and may become a limited liability company after it has accumulated enough share capital. Regarding the different sectors, the building sector ($\widehat{\beta}_5$) has a lower insolvency risk compared to other sectors. The remaining sectors are clearly insignificant, and hence have no effect on the age-at-insolvency in this model.

As the simulation study indicates, one reason for the large confidence intervals lies in the low probability of observation ($\widehat{\alpha} \approx 1\%$). Looking at the estimated distribution function $\widehat{F^{\widetilde{\varepsilon}}}$, Figure 2 reveals the shape of a shifted exponential distribution. The confidence intervals were obtained by simple bootstrap. Denote $\widehat{F_b^{\widetilde{\varepsilon}}}$ as the estimated error distribution for the bth resample, $b = 1, \ldots, 1000$. Consequently, for every residual $\varepsilon_i, i = 1, \ldots, 400$ the 2.5% and 97.5% quantiles of all $\widehat{F_b^{\widetilde{\varepsilon}}}(\varepsilon_i)$ determine the confidence intervals.

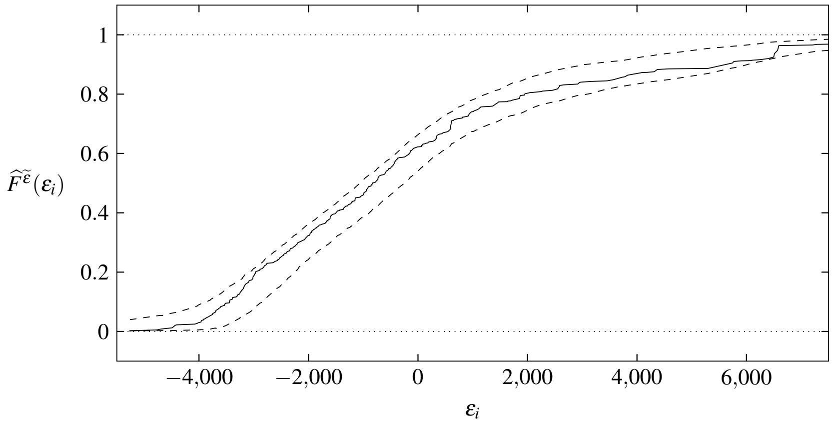

Figure 2: Estimated distribution function of $\widetilde{\varepsilon}$ with 95% bootstrap confidence intervals.

6. Discussion

Based on the idea of Stute (1993b) for censored data, consistent and asymptotically normal estimators under random double-truncation were derived. In order to obtain variance estimators, bootstrap estimators are recommended. He and Yang (2003) pursued the same idea of Stute (1993b) for randomly left-truncated data and established a closed-form variance estimator. They used the fact that the NPMLE for randomly left-truncated data (see e.g. Woodroofe, 1985), also known as Lynden-Bell or product-limit estimator, is a functional of empirical distribution functions for which useful approximations exist. In particular, Stute (1993a) developed error bounds which were employed by He and Yang (2003). This is only one reason why a product-limit estimator for randomly double-truncated data would be desirable. However, some technical aspects related to the cumulative hazard function reveal that the methods of Woodroofe (1985) can not be generalised to the case of random double-truncation. Nevertheless, Shen (2010) developed an NPMLE for randomly double-truncated data but there is no closed-form variance estimator for $\widehat{F}^{\widetilde{T},\widetilde{T}+\widetilde{D}}$. Therefore, bootstrapping the variance of $\widehat{\beta}$ seems to be a reasonable alternative.

Concerning the choice of covariates, forward selection and backward elimination or a combination of both are most commonly used in practice (see Hocking, 1976). Since the proposed method directly offers estimators of the underlying distributions, extending the standard selection methods to the investigated truncation setting seems natural. In particular, we suggest to use the derived estimators for R^2 and its standard error to perform forward selection. In this procedure, the default model contains no covariates, and in each successive step the covariate which contributes the largest significant improvement of R^2 is included, until no covariate carries a significant improvement. However, further research is needed to assess the exact performance of this selection method.

The simulation study and the application to German companies show that under a low probability of observation, reliable estimates are hard to obtain - even for a simple model like linear regression.

Acknowledgement

The authors thank the reviewer for the helpful and valuable comments and suggestions.

References

BHATTACHARYA, P. K., CHERNOFF, H., AND YANG, S. S. (1983). Nonparametric estimation of the slope of a truncated regression. *The Annals of Statistics*, **11** (2), 505–514.

BILLINGSLEY, P. (1968). *Convergence of Probability Measures*. Wiley Series in Probability and Statistics. John Wiley & Sons: New York.

CHUNG, K. L. (2001). *A Course in Probability Theory*. Academic Press: San Diego.

EFRON, B. AND PETROSIAN, V. (1999). Nonparametric methods for doubly truncated data. *Journal of the American Statistical Association*, **94**, 824–834.

EMURA, T., KONNO, Y., AND MICHIMAE, H. (2015). Statistical inference based on the nonparametric maximum likelihood estimator under double-truncation. *Lifetime Data Analysis*, **21**, 397–418.

EMURA, T. AND WANG, W. (2010). Testing quasi-independence for truncation data. *Journal of Multivariate Analysis*, **101**, 223–239.

GROSS, S. T. AND HUBER-CAROL, C. (1992). Regression models for truncated survival data. *Scandinavian Journal of Statistics*, **19** (3), 193–213.

GROSS, S. T. AND LAI, T. L. (1996). Nonparametric estimation and regression analysis with left-truncated and right-censored data. *Journal of the American Statistical Association*, **91** (435), 1166–1180.

HE, S. AND YANG, G. L. (2003). Estimation of regression parameters with left truncated data. *Journal of Statistical Planning and Inference*, **117** (1), 99–122.

HOCKING, R. R. (1976). The analysis and selection of variables in linear regression. *Biometrics*, **32** (1), 1–49.

KALBFLEISCH, J. AND LAWLESS, J. F. (1989). Inference based on retrospective ascertainment: An analysis of the data on transfusion-related aids. *Journal of the American Statistical Association*, **84** (406), 360–372.

MARTIN, E. C. AND BETENSKY, R. A. (2005). Testing quasi-independence of failure and truncation times via conditional kendall's tau. *Journal of The American Statistical Association*, **100**, 484–492.

MOREIRA, C. AND UÑA-ÁLVAREZ, J. (2010). Bootstrapping the NPMLE for doubly truncated data. *Journal of Nonparametric Statistics*, **22** (5), 567–583.

MOREIRA, C., UÑA-ÁLVAREZ, J., AND MEIRA-MACHADO, L. (2016). Nonparametric regression with doubly truncated data. *Computational Statistics and Data Analysis*, **93**, 294–307.

REISS, R.-D. (1993). *A Course on Point Processes*. Springer: New York.

SHEN, P.-S. (2010). Nonparametric analysis of doubly truncated data. *Annals of the Institute of Statistical Mathematics*, **62**, 835–853.

SHEN, P.-S. (2013). Regression analysis of interval censored and doubly truncated data with linear transformation models. *Computational Statistics*, **28**, 581–596.

SHEN, P.-S. (2015). Median regression model with doubly truncated data. *Journal of Applied Statistics*, **42**, 360–370.

STUTE, W. (1993a). Almost sure representations of the product-limit estimator for truncated data. *Annals of Statistics*, **21**, 146–156.

STUTE, W. (1993b). Consistent estimation under random censorship when covariables are present. *Journal of Multivariate Analysis*, **45**, 89–103.

TSAI, W.-Y. (1990). Testing the assumption of independence of truncation time and failure time. *Biometrika*, **77**, 169–177.

VAN DER VAART, A. (1998). *Asymptotic Statistics*. Cambridge University Press: United Kingdom.

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

ZHANG, X. (2015). Nonparametric inference for an inverse-probability-weighted estimator with doubly truncated data. *Communications in Statistics - Simulation and Computation*, **44**, 489–504.

Appendix

The consistency of $\widehat{F}^{\widetilde{T}+\widetilde{D},\widetilde{T}}$

For the sake of clarity, the setting in Shen (2010) will be introduced first. Note that Shen (2010) assumes a deterministic number of observations in contrast to this paper. However, it does not influence the asymptotic results. This follows from the monotonicity of N with respect to n (Chung, 2001, p. 143). As previously mentioned, the distribution of $\widetilde{Y}$ and the joint distribution of $(\widetilde{T},\widetilde{T}+\widetilde{D})$ are estimated non-parametrically. Consider Assumptions (A1), (A3), (A5), (A6) and (B1)-(B3). The full likelihood can be written as

$$L_N(f,k)=\prod_{i=1}^{N}\frac{f_ik_i}{\sum_{j=1}^{N}F_jk_j},$$

where $f=(f_1,\ldots,f_N)$ and $k=(k_1,\ldots,k_N)$ are discrete densities and $F_j:=\sum_{l=1}^{N}f_lI_{\{T_j\leq Y_l\leq T_j+D_j\}}$. The density f has only probability mass at the observations $(Y_1,\ldots,Y_N)$. The same is true for k with respect to $((T_1,T_1+D_1),\dots,(T_N,T_N+D_N))$. In order to achieve the inverse-probability-weighted estimator, the full likelihood can be written as

$$L_N(f,k)=\prod_{i=1}^{N}\frac{f_i}{F_i}\prod_{i=1}^{N}\frac{F_ik_i}{\sum_{j=1}^{N}F_jk_j}=L_1(f)\cdot L_2(f,k).$$

The NPMLE corresponding to $L_1(f)$ can be obtained by solving

$$\frac{1}{\widehat{f}_i}=\sum_{j=1}^{N}\frac{I_{\{T_j\le Y_i\le T_j+D_j\}}}{\widehat{F}_j},\quad i=1,\dots,N,$$

where $\widehat{F}_j=\sum_{l=1}^{N}\widehat{f}_lI_{\{T_j\le Y_l\le T_j+D_j\}}$. The estimation of k follows analogously, i.e. the full likelihood can also be written as

$$L_N(f,k)=\prod_{i=1}^{N}\frac{k_i}{K_i}\prod_{i=1}^{N}\frac{K_if_i}{\sum_{j=1}^{N}K_jf_j}=L_1(k)\cdot L_2(k,f),$$

where $K_j:=\sum_{l=1}^{N}k_lI_{\{T_l\le Y_j\le T_l+D_l\}}$. Similarly, solving the equation

$$\frac{1}{\widehat{k}_i}=\sum_{j=1}^{N}\frac{I_{\{T_i\le Y_j\le T_i+D_i\}}}{\widehat{K}_j},\quad i=1,\dots,N$$

leads to the NPMLE corresponding to $L_1(k)$. In addition, it was shown that the estimates of $\widehat{f}$ and $\widehat{k}$ can also be obtained by solving

$$\widehat{f}_i=\left(\sum_{j=1}^{N}\frac{1}{\widehat{K}_j}\right)^{-1}\frac{1}{\widehat{K}_i}\text{ and}$$
$$\widehat{k}_i=\left(\sum_{j=1}^{N}\frac{1}{\widehat{F}_j}\right)^{-1}\frac{1}{\widehat{F}_i}$$

simultaneously for $i=1,\dots,N$. This iterative procedure results in the same estimates but saves computational efforts. Furthermore, it was proven that $\widehat{f}$ and $\widehat{k}$ are also the NPMLE of the full likelihood. Let

$$\widehat{F}^{\widetilde{Y}}(y):=\sum_{i=1}^{N}\widehat{f}_iI_{\{Y_i\le y\}},$$
$$\widehat{F}^{\widetilde{T},\widetilde{T}+\widetilde{D}}(t,s):=\sum_{i=1}^{N}\widehat{k}_iI_{\{T_i\le t,T_i+D_i\le s\}},$$

and note that

$$F^{\widetilde{T},\widetilde{T}+\widetilde{D}}(t,s)=\left[\int_{a_{F^{\widetilde{T}+\widetilde{D}}}}^{\infty}\int_{a_{F^{\widetilde{T}}}}^{v}\frac{dF^{T,T+D}(u,v)}{F^{\widetilde{Y}}(v)-F^{\widetilde{Y}}(u-)}\right]^{-1}\int_{a_{F^{\widetilde{T}+\widetilde{D}}}}^{s}\int_{a_{F^{\widetilde{T}}}}^{\min(t,v)}\frac{dF^{T,T+D}(u,v)}{F^{\widetilde{Y}}(v)-F^{\widetilde{Y}}(u-)}\tag{5}$$

and that plugging $\widehat{F}^{\widetilde{Y}}$ into (5) results in $\widehat{F}^{\widetilde{T},\widetilde{T}+\widetilde{D}}$. Since $F^{T,T+D}$ can be estimated with the empirical distribution function and $\widehat{F}^{\widetilde{Y}}$ is consistent, the mapping theorem (Billingsley, 1968, p. 34) implies the consistency of $\widehat{F}^{\widetilde{T},\widetilde{T}+\widetilde{D}}$. Note that (A1) ensures that the consistency is uniform.

Manuscript received, 2016-05-10, revised, 2016-07-22, accepted, 2016-07-25.