Open access • Journal Article • DOI:10.1002/BIMJ.201800375

Flexible parametric model for survival data subject to dependent censoring.

Negera Wakgari Deresa, Ingrid Van Keilegom

Institutions: Katholieke Universiteit Leuven

Published on: 01 Jan 2020 - Biometrical Journal (John Wiley & Sons, Ltd)

Topics: Censoring (clinical trials), Identifiability, Parametric model, Linear model and Conditional independence

Flexible parametric model for survival data subject to dependent censoring

Negera Wakgari Deresa *
KU Leuven

Ingrid Van Keilegom *
KU Leuven

August 8, 2019

Abstract

When modeling survival data, it is common to assume that the (log-transformed) survival time (T) is conditionally independent of the (log-transformed) censoring time (C) given a set of covariates. There are numerous situations in which this assumption is not realistic, and a number of correction procedures have been developed for different models. However, in most cases, either some prior knowledge about the association between T and C is required, or some auxiliary information or data is supposed to be available. When this is not the case, the application of many existing methods turns out to be limited. The goal of this paper is to overcome this problem by developing a flexible parametric model, that is a type of transformed linear model. We show that the association between T and C is identifiable in this model. The performance of the proposed method is investigated both in an asymptotic way and through finite sample simulations. We also develop a formal goodness-of-fit test approach to assess the quality of the fitted model. Finally, the approach is applied to data coming from a study on liver transplants.

Key Words: Association; dependent censoring; identifiability; parametric models; survival analysis.

*ORSTAT, KU Leuven, Naamsestraat 69, B-3000 Leuven, Belgium. Email: negerawakgari.deresa@kuleuven.be and ingrid.vankeilegom@kuleuven.be (corresponding author). Financial support from the European Research Council (2016-2021, Horizon 2020 / ERC grant agreement No. 694409) is gratefully acknowledged.

1 Introduction

In survival analysis, it is common to assume that the (log-transformed) survival time (T) and the (log-transformed) censoring time (C) are independent of each other. This assumption is reasonable in many contexts, in particular when censoring happens at the end of the study (the so-called administrative censoring). However, there are also numerous situations where this assumption is not realistic. For example, in medical studies, patients may withdraw from the study because their condition is deteriorating or because they are showing side effects that need alternative treatments. In this case, withdrawal from the study may indicate that death is likely sooner, and so we have a positive relation between the survival and the censoring time. On the other hand, patients could drop out because their health condition has improved and so they no longer follow the treatment (Etzioni et al., 1999). In this situation, the censoring time is negatively related to the survival time. Another situation where the assumption of independent censoring is questionable can be found in transplant studies. For this type of data, a patient is selected for a transplant based on his/her medical condition. Since the most sick patients are selected for a transplant, it is unconvincing that their expected survival on the waiting list is representative of those who have not been selected for the transplant (Staplin et al., 2015). In these examples, the issue of dependent censoring is induced.

A number of approaches have been proposed in the literature to deal with dependent censoring. According to Tsiatis (1975), the joint distribution of T and C is not identifiable from the joint distribution of the follow-up time and the censoring indicator when we have dependent censoring. Therefore, to identify the joint distribution of T and C, we need extra information about their dependence. In this regard, Zheng and Klein (1995) modeled the bivariate distribution of T and C by means of a known copula function, and estimated the marginal distribution of T non-parametrically under this copula model, which completely specifies the association between the two variables. Rivest and Wells (2001) further investigated the proposal of Zheng and Klein for the special case of an Archimedean copula. The copula approach has been extended to the context of regression by Braekers and Veraverbeke (2005), Huang and Zhang (2008) and Sujica and Van Keilegom (2018), who incorporated covariates in Zheng and Klein's model by assuming that the marginal relation between the survival time and the covariates is given by a fully nonparametric model, a Cox model and a location-scale model, respectively. The first textbook on dependent censoring under an assumed copula has been written by Emura and Chen (2018). The inverse probability of censoring weighted (IPCW) method (Collett, 2015), where the weight

is derived from a censoring time model, and the multiple imputation methods of Jackson et al. (2014), where the censored times are imputed under departures from independent censoring, are also useful to adjust for dependent censoring in the Cox model.

Although it is possible to adjust in some sense for possible dependent censoring using one of the existing approaches, we would like to come up with a novel way to adjust for dependent censoring, that is at the same time flexible enough to encompass many data structures, and easy to interpret thanks to its parametric nature. Under our proposed model the dependence between T and C will be identified. The key element in our model that ensures this identifiability is the bivariate normality of the errors in the model for T and the model for C. We will propose a method to estimate the model (and in particular the association between T and C) and to do inference. We will also develop a formal goodness-of-fit statistic to check the quality of the fit.

The proposed model is based on a certain parametric transformation of T and C, and we assume that the transformed variables have a bivariate normal distribution, after adjusting for possible covariate effects. Some authors have already used a bivariate model to induce an association. For instance, Emoto and Matthews (1990) suggested a bivariate Weibull model and Basu (1988) gave a review of existing bivariate exponential models, which could be used to model an association. However, these models do not allow for the adjustment of covariates. Here, we focus on bivariate normal models to induce an association, which are promising because of their mathematical tractability and because of the possibility of transforming non-normal continuous variables to approximately normal ones. For showing the identifiability of our model, we will use the papers of Nádas (1971) and Basu and Ghosh (1978), who showed that for a bivariate normal random vector, the distribution of the identified minimum (i.e. the observed minimum and the censoring indicator) determines the joint distribution of the bivariate normal pair. We will extend their result to our more flexible setting, in which we assume that the survival and censoring times behave like a bivariate normal random vector after a proper parametric transformation and after adjusting for covariates. This identification result is the crucial result of the paper.

The paper is organized as follows. In the next section we state the precise model and derive the distribution of the observed minimum. The identifiability and estimation of our proposed model is shown in Section 3. In Section 4, we give a formal goodness-of-fit test. A simulation study that investigates the finite sample behavior of the proposed estimator is given in Section 5. In Section 6, we apply our models and methods to real data coming from a study on liver transplants, and we conclude with a discussion in Section 7. All

proofs are deferred to Appendix A (for the identifiability result) and B (for the asymptotic results).

2 Model specification

Throughout the paper the variable T denotes the logarithm of the survival time, and C is the logarithm of the censoring time, both variables taking values in $(-\infty, +\infty)$. We assume that T and C are depending on each other, even after conditioning out the effect of covariates, and they are censoring each other. In this situation, the censoring time is worth to be modeled jointly with the survival time by including all relevant covariates. This enables us to identify the set of covariates having an influence on the survival time as well as on the censoring time. Denote the covariates having an influence on T by $X = (1, \tilde{X}^T)^T$ (of dimension p, say), and the covariates having an influence on C by $W = (1, \tilde{W}^T)^T$, which we suppose to be of dimension q. They may be identical, partially overlapping, or completely distinct. Then the proposed joint regression model has the following form:

$$\begin{cases} \Lambda_\theta(T) = X^T\beta + \epsilon_T \\ \Lambda_\theta(C) = W^T\eta + \epsilon_C, \end{cases} \tag{2.1}$$

where $\{\Lambda_\theta : \theta \in \Theta\}$ with $\Theta \subset \mathbb{R}$ is a parametric class of monotone increasing transformations, and β and η are the vectors of regression coefficients. The vector of error terms (ϵ_T, ϵ_C) has a bivariate normal distribution:

$$\begin{pmatrix} \epsilon_T \\ \epsilon_C \end{pmatrix} \sim N_2\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Sigma = \begin{pmatrix} \sigma_T^2 & \rho\sigma_T\sigma_C \\ \rho\sigma_T\sigma_C & \sigma_C^2 \end{pmatrix}\right), \tag{2.2}$$

where Σ is assumed to be a positive definite matrix, i.e. $\sigma_T > 0, \sigma_C > 0$ and $|\rho| < 1$. Here, we consider the same transformation parameter (θ) for T and C to make sure that $\min(\Lambda_\theta(T), \Lambda_\theta(C)) = \Lambda_\theta(\min(T, C))$, i.e. the censoring indicator does not change after the transformation. We shall assume that

(A1) (ϵ_T, ϵ_C) and (X, W) are independent.

(A2) The matrices $\mathrm{Var}(\tilde{X})$ and $\mathrm{Var}(\tilde{W})$ have full rank.

(A3) The family $\{\Lambda_\theta : \theta \in \Theta\}$ is a family of strictly increasing transformations that are defined on the whole real line and that satisfy $\lim_{t\to\pm\infty} \Lambda_\theta(t) = \pm\infty$ for all θ in Θ.

The latter assumption is required for identifiability reasons, and states that the family of transformations $\{\Lambda_\theta : \theta \in \Theta\}$ maps the whole real line to the whole real line. The Box-Cox transformation (depending on a parameter θ) does not satisfy this assumption, since it maps a positive random variable to a variable defined on $(-1/\theta, +\infty)$ for $\theta > 0$ and $(-\infty, +\infty)$ for $\theta = 0$. On the other hand, Yeo and Johnson (2000) proposed a family of power transformations that maps $(-\infty, +\infty)$ to $(-\infty, +\infty)$ provided $0 \leq \theta \leq 2$. It is an extension of the Box-Cox family to the whole real line, and is defined as follows:

$$\Lambda_\theta(t) = \begin{cases} \{(t+1)^\theta - 1\}/\theta & t \geq 0, \theta \neq 0 \\ \log(t+1) & t \geq 0, \theta = 0 \\ -\{(-t+1)^{2-\theta} - 1\}/(2-\theta) & t < 0, \theta \neq 2 \\ -\log(-t+1) & t < 0, \theta = 2. \end{cases} \tag{2.3}$$

The case $\theta = 1$ corresponds to the identity transformation, and hence $\Lambda_\theta(T)$ is the log of the survival time in that case. When $1 < \theta \leq 2$, the function $\Lambda_\theta(\cdot)$ lies entirely above the identity transformation and is convex, whereas when $0 \leq \theta < 1$, it lies completely below the identity function and it is concave. We refer to Yeo and Johnson (2000) for more details and properties of this flexible family of transformations. Assumption (A3) is also satisfied by a family of sinh-arcsinh transformations, which is defined by $\Lambda_\theta(t) = \sinh(\sinh^{-1}(t) + \theta)$, $\theta \in \mathbb{R}, t \in \mathbb{R}$ (see Jones and Pewsey (2009) for more details on this family).

In addition to condition (A3), we also need the following condition, which is also required in order to identify model (2.1)–(2.2):

(A4) For all $\theta_j, \mu_j, \sigma_j \in \mathbb{R}^3$, $j = 1, 2$: if $\theta_1 \neq \theta_2$, then $\lim_{t \to \pm\infty} K_{\theta_1,\mu_1,\sigma_1}(t)/K_{\theta_2,\mu_2,\sigma_2}(t) = 0$ or ∞, where

$$K_{\theta,\mu,\sigma}(t) = \exp\Big(-\frac{1}{2}\Big\{\frac{\Lambda_\theta(t) - \mu}{\sigma}\Big\}^2\Big)\Lambda'_\theta(t).$$

It is easily seen that this condition is satisfied for both the Yeo-Johnson family and the Jones-Pewsey family of transformations.

Note that model (2.1)–(2.2) is a flexible parametric model, depending on the parameter vector $(\theta, \beta, \eta, \sigma_T^2, \sigma_C^2, \rho) \in \mathbb{R}^{p+q+4}$. Due to censoring, the underlying random variables T and C are only observed through the follow-up time Z and the censoring indicator Δ, given by $Z = \min(T, C)$ and $\Delta = I(T \leq C)$, where $I(\cdot)$ is the indicator function. Finally, the data consist of n independent and identically distributed replications $(Z_i, \Delta_i, X_i, W_i)$, $i = 1, \ldots, n$ of (Z, Δ, X, W). We make the following additional assumption:

(A5) The probabilities $P(Z = T)$ and $P(Z = C)$ are strictly positive.

We will show in the next section that model (2.1)–(2.2) is identifiable. This might seem surprising, since the model determines the relation between T and C whereas for a given individual we observe either T or C, but never both of them.

It is known that when the response in a regression model is transformed by a power transformation, the regression function of the new model often has an additive structure and the new error term is often approximately normal and homoscedastic (Box and Cox, 1964). Hence, the assumption of bivariate normality of the error vector (ϵ_T, ϵ_C) is quite natural in this context.

Let $F_{T|X}(\cdot \mid x)$ and $F_{C|W}(\cdot \mid w)$ be the conditional distribution functions of T given $X = x$ and of C given $W = w$, respectively. Then, thanks to the independence between (ϵ_T, ϵ_C) and (X, W), we have that

$$\begin{aligned} F_{T|X}(t \mid x) &= F_{\epsilon_T}\big(\Lambda_\theta(t) - x^T\beta\big) = \Phi\Big(\tfrac{\Lambda_\theta(t)-x^T\beta}{\sigma_T}\Big) \\ F_{C|W}(c \mid w) &= F_{\epsilon_C}\big(\Lambda_\theta(c) - w^T\eta\big) = \Phi\Big(\tfrac{\Lambda_\theta(c)-w^T\eta}{\sigma_C}\Big), \end{aligned} \tag{2.4}$$

where F_{ϵ_T} and F_{ϵ_C} are the distribution functions of ϵ_T and ϵ_C, respectively, and where Φ is the distribution function of a standard normal variable. It follows that the conditional density of T given $X = x$ equals $f_{T|X}(t \mid x) = \sigma_T^{-1}\phi\big(\tfrac{\Lambda_\theta(t)-x^T\beta}{\sigma_T}\big)\Lambda'_\theta(t)$, where ϕ is the density function of a standard normal variable, and similarly for the conditional density $f_{C|W}(c \mid w)$ of C given $W = w$.

Let the parameter vector be denoted by $\alpha = (\theta, \beta, \eta, \sigma_T, \sigma_C, \rho) \in \mathbb{R}^{p+q+4}$. The sub-distribution function $F_{Z,\Delta|X,W}(\cdot, \cdot \mid x, w; \alpha)$ of (Z, Δ) given $(X, W) = (x, w)$ for a given α can now be derived as follows:

$$\begin{aligned} F_{Z,\Delta|X,W}(z, 1 \mid x, w; \alpha) &= P(Z \le z, \Delta = 1 \mid X = x, W = w) \\ &= P(\Lambda_\theta(T) \le \Lambda_\theta(z), \Lambda_\theta(T) \le \Lambda_\theta(C) \mid X = x, W = w) \\ &= \int_{-\infty}^{\Lambda_\theta(z)-x^T\beta} P(\epsilon_C \ge e + x^T\beta - w^T\eta \mid \epsilon_T = e) f_{\epsilon_T}(e)\, de \\ &= \frac{1}{\sigma_T}\int_{-\infty}^{\Lambda_\theta(z)-x^T\beta} \Big[1 - \Phi\Big(\frac{e + x^T\beta - w^T\eta - \rho\frac{\sigma_C}{\sigma_T}e}{\sigma_C(1-\rho^2)^{1/2}}\Big)\Big]\phi\Big(\frac{e}{\sigma_T}\Big)\, de, \end{aligned}$$

since $(\epsilon_C \mid \epsilon_T = e) \sim N\big(\rho\frac{\sigma_C}{\sigma_T}e, \sigma_C^2(1-\rho^2)\big)$, and hence the corresponding sub-density $f_{Z,\Delta|X,W}(\cdot, \cdot \mid x, w; \alpha)$ is given by

$$\begin{aligned} f_{Z,\Delta|X,W}(z, 1 \mid x, w; \alpha) &= \frac{1}{\sigma_T}\Big[1 - \Phi\Big(\frac{\Lambda_\theta(z) - w^T\eta - \rho\frac{\sigma_C}{\sigma_T}(\Lambda_\theta(z) - x^T\beta)}{\sigma_C(1-\rho^2)^{1/2}}\Big)\Big] \\ &\quad \times\phi\Big(\frac{\Lambda_\theta(z) - x^T\beta}{\sigma_T}\Big)\Lambda'_\theta(z). \end{aligned} \tag{2.5}$$

Similarly,

$$f_{Z,\Delta|X,W}(z,0 \mid x,w;\alpha) = \frac{1}{\sigma_C}\Big[1-\Phi\Big(\frac{\Lambda_\theta(z)-x^T\beta-\rho\frac{\sigma_T}{\sigma_C}(\Lambda_\theta(z)-w^T\eta)}{\sigma_T(1-\rho^2)^{1/2}}\Big)\Big] \times\phi\Big(\frac{\Lambda_\theta(z)-w^T\eta}{\sigma_C}\Big)\Lambda'_\theta(z). \quad (2.6)$$

In a similar way we can show that

$$F_{Z|X,W}(z \mid x,w;\alpha) = \Phi\Big(\frac{\Lambda_\theta(z)-x^T\beta}{\sigma_T}\Big)+\Phi\Big(\frac{\Lambda_\theta(z)-w^T\eta}{\sigma_C}\Big) -\Phi\Big(\frac{\Lambda_\theta(z)-x^T\beta}{\sigma_T},\frac{\Lambda_\theta(z)-w^T\eta}{\sigma_C};\rho\Big), \quad (2.7)$$

since $P(Z\le z)=1-P(Z>z)=1-P(T>z,C>z)=P(T\le z)+P(C\le z)-P(T\le z,C\le z)$, where $\Phi(\cdot,\cdot;\rho)$ is the cumulative distribution function of a standard bivariate normal distribution with correlation parameter ρ. These formulas will be useful in the sequel.

3 Model identification and estimation

We will show in this section that model (2.1)–(2.2) is identifiable. Note that we only observe the vector (Z,Δ,X,W) and on the basis of the joint distribution of this vector, we need to show that the model parameters $\theta,\beta,\eta,\sigma_T^2,\sigma_C^2$ and ρ are identifiable, in the sense that any two different sets of parameters yield different joint distributions of (Z,Δ,X,W). That the association parameter ρ is identifiable is quite surprising, since we only observe the minimum Z of T and C, but never both of them. Our proof relies on Basu and Ghosh (1978), who show the identifiability of model (2.1)–(2.2) when no covariates and no transformation are included. The proof is given in Appendix A.

Theorem 3.1. *Under assumptions* $(A1)-(A5)$*, suppose that the pair* (T_j,C_j) *satisfies model (2.1)–(2.2) with parameters* $\alpha_j=(\theta_j,\beta_j,\eta_j,\sigma_{T_j},\sigma_{C_j},\rho_j)$ *for* $j=1,2$*, and that* $Z_j=\min(T_j,C_j)$ *and* $\Delta_j=I(T_j\le C_j)$*. If* $f_{Z_1,\Delta_1|X,W}(\cdot,\ell \mid x,w;\alpha_1)\equiv f_{Z_2,\Delta_2|X,W}(\cdot,\ell \mid x,w;\alpha_2)$ *for* $\ell=0,1$ *and for almost every* (x,w)*, then*

$$\theta_1=\theta_2,\beta_1=\beta_2,\eta_1=\eta_2,\sigma_{T_1}=\sigma_{T_2},\sigma_{C_1}=\sigma_{C_2},\rho_1=\rho_2.$$

Remark 1: It is easily seen that the identification result given in Theorem 3.1 remains valid when the functional form of the covariates is non-linear, like e.g. polynomials, fractional polynomials or other fractional forms as long as assumption (A2) holds true.

Since our model is fully parametric, we will estimate the model parameters by maximizing the likelihood function. We write the joint density of (Z, Δ, X, W) as $f_{Z,\Delta,X,W} = f_{Z,\Delta|X,W} f_{X,W}$. Since the joint density $f_{X,W}$ of (X, W) does not depend on the model parameters, we can build the likelihood with the conditional density $f_{Z,\Delta|X,W}$. For each observation $(Z_i, \Delta_i, X_i, W_i)$, $i = 1, \ldots, n$, we obtain its contribution to the likelihood function from formulas (2.5) for the uncensored observations and (2.6) for the censored ones. This gives us the following likelihood function for the parameter vector $\alpha = (\theta, \beta, \eta, \sigma_T, \sigma_C, \rho)$:

$$\begin{aligned} L(\alpha) &= \prod_{i=1}^{n} f_{Z,\Delta|X,W}(Z_i, \Delta_i \mid X_i, W_i; \alpha) \\ &= \prod_{i=1}^{n} \Big\{ \frac{1}{\sigma_T} \Big[1 - \Phi\Big(\frac{\Lambda_\theta(Z_i) - W_i^T\eta - \rho\frac{\sigma_C}{\sigma_T}(\Lambda_\theta(Z_i) - X_i^T\beta)}{\sigma_C(1-\rho^2)^{1/2}} \Big) \Big] \phi\Big(\frac{\Lambda_\theta(Z_i) - X_i^T\beta}{\sigma_T} \Big) \Big\}^{\Delta_i} \\ &\quad \times \Big\{ \frac{1}{\sigma_C} \Big[1 - \Phi\Big(\frac{\Lambda_\theta(Z_i) - X_i^T\beta - \rho\frac{\sigma_T}{\sigma_C}(\Lambda_\theta(Z_i) - W_i^T\eta)}{\sigma_T(1-\rho^2)^{1/2}} \Big) \Big] \phi\Big(\frac{\Lambda_\theta(Z_i) - W_i^T\eta}{\sigma_C} \Big) \Big\}^{1-\Delta_i} \\ &\quad \times \Lambda'_\theta(Z_i). \end{aligned} \tag{3.1}$$

This likelihood will be maximized over the parameter space $A = \{(\theta, \beta, \eta, \sigma_T, \sigma_C, \rho) : \theta \in \Theta, \beta \in \mathbb{R}^p, \eta \in \mathbb{R}^q, \sigma_T > 0, \sigma_C > 0, -1 < \rho < 1\}$. Note that unlike the case where T and C are independent given (X, W), the above likelihood cannot be factorized into a factor only depending on the parameters of the model for T, and a second factor only depending on the parameters of the model for C. The only exception is when $\rho = 0$ and when θ would be known, in which case we find back the usual likelihood in the independent case.

We now define the maximum likelihood estimator (MLE) of α as follows:

$$\hat{\alpha} = (\hat{\theta}, \hat{\beta}, \hat{\eta}, \hat{\sigma}_T, \hat{\sigma}_C, \hat{\rho}) = \text{argmax}_{\alpha \in A} L(\alpha).$$

Note that it is not possible to maximize this likelihood function analytically, but instead it can be maximized numerically.

Let us now consider the asymptotic theory of our estimator. We will do this assuming that our model (2.1) is identified but potentially misspecified (see Remark 1 for identifiable classes of models that are more general than model (2.1)). In this context, the results of White (1982) on misspecified parametric models can be used to derive the consistency and asymptotic normality of our estimators. We give theorems related to consistency and asymptotic normality of our estimator in Appendix B.

We end this section with a discussion on how to compute standard errors (SE) and confidence intervals (CI) for the parameters in the model. We will do this by using the

asymptotic normality of the MLE given in Theorem 7.2 in Appendix B. For example, the SE for the regression parameter $\hat{\beta}_j$ is $SE(\hat{\beta}_j) = \sqrt{[V(\hat{\alpha})]_{\beta_j}}, j = 1, \ldots, p$, where $V(\hat{\alpha})$ is the variance–covariance matrix given in Theorem 7.2. The $(1-\omega)\times 100\%$ confidence interval for β_j is then $\hat{\beta}_j \pm z_{1-\omega/2}\widehat{SE}(\hat{\beta}_j)$, where $z_{1-\omega/2}$ is the $1-\omega/2$-quantile of the standard normal distribution, and $\widehat{SE}(\hat{\beta}_j)$ is an estimator of $SE(\hat{\beta}_j)$. Similarly, CI's for the regression coefficients $\eta_1, \ldots, \eta_q$ can be obtained. Note that for the variances σ_T and σ_C, and for the correlation ρ, the confidence interval will be based on the logarithm and on Fisher's Z transformation, respectively. Their corresponding standard errors will be obtained using the Delta method, and the confidence interval limits will then be transformed back to the original scale. For the logarithmic transformation we have $g(\hat{\sigma}_T) = \log(\hat{\sigma}_T)$, and then using the Delta method, $SE(g(\hat{\sigma}_T)) = \sqrt{[V(\hat{\alpha})]_{\sigma_T}}/\sigma_T$. The $(1-\omega)\times 100\%$ confidence interval is then given by

$$\exp\{g(\hat{\sigma}_T) - z_{1-\omega/2}\widehat{SE}(g(\hat{\sigma}_T))\} < \sigma_T < \exp\{g(\hat{\sigma}_T) + z_{1-\omega/2}\widehat{SE}(g(\hat{\sigma}_T))\}.$$

A similar result can be obtained for $\hat{\sigma}_C$. Similarly, for Fisher's transformation we have $g(\hat{\rho}) = 0.5\log\big\{(1+\hat{\rho})/(1-\hat{\rho})\big\}$ and $SE(g(\hat{\rho})) = \sqrt{[V(\hat{\alpha})]_{\rho}}/(1-\rho)$. Then the $(1-\omega)\times 100\%$ confidence interval is given by

$$\frac{\exp\{2[g(\hat{\rho}) - z_{1-\omega/2}\widehat{SE}(g(\hat{\rho}))]\} - 1}{\exp\{2[g(\hat{\rho}) - z_{1-\omega/2}\widehat{SE}(g(\hat{\rho}))]\} + 1} < \rho < \frac{\exp\{2[g(\hat{\rho}) + z_{1-\omega/2}\widehat{SE}(g(\hat{\rho}))]\} - 1}{\exp\{2[g(\hat{\rho}) + z_{1-\omega/2}\widehat{SE}(g(\hat{\rho}))]\} + 1}.$$

4 Goodness-of-fit test

Since the proposed estimation method relies on the model assumptions, we supplement the method with a formal goodness-of-fit test for testing whether the distribution of the observed survival time Z is equal to the distribution under our dependent censoring model. More precisely under the null hypothesis, we have

$$H_0 : P(Z \leq z) = F_Z(z; \alpha), \text{ for some } \alpha,$$

where the distribution of Z under H_0 is given by

$$\begin{aligned} F_Z(z;\alpha) &= \int\int P(Z \leq z \mid X = x, W = w) f_{X,W}(x,w)\, dx\, dw \\ &= \int \Phi\Big(\frac{\Lambda_\theta(z) - x^T\beta}{\sigma_T}\Big) f_X(x)\, dx + \int \Phi\Big(\frac{\Lambda_\theta(z) - w^T\eta}{\sigma_C}\Big) f_W(w)\, dw \\ &\quad - \int\int \Phi\Big(\frac{\Lambda_\theta(z) - x^T\beta}{\sigma_T}, \frac{\Lambda_\theta(z) - w^T\eta}{\sigma_C}; \rho\Big) f_{X,W}(x,w)\, dx\, dw, \end{aligned}$$

which can be approximated by

$$\begin{aligned} F_Z(z;\hat{\alpha}) &= \frac{1}{n}\sum_{i=1}^{n}\Phi\Big(\frac{\Lambda_{\hat{\theta}}(z)-X_i^T\hat{\beta}}{\hat{\sigma}_T}\Big)+\frac{1}{n}\sum_{i=1}^{n}\Phi\Big(\frac{\Lambda_{\hat{\theta}}(z)-W_i^T\hat{\eta}}{\hat{\sigma}_C}\Big) \\ &\quad -\frac{1}{n}\sum_{i=1}^{n}\Phi\Big(\frac{\Lambda_{\hat{\theta}}(z)-X_i^T\hat{\beta}}{\hat{\sigma}_T},\frac{\Lambda_{\hat{\theta}}(z)-W_i^T\hat{\eta}}{\hat{\sigma}_C};\hat{\rho}\Big), \end{aligned} \tag{4.1}$$

where the model parameters are replaced by their corresponding estimated values. It can be seen that both the survival and the censoring model parameters are involved in the estimation of $F_Z(\cdot;\alpha)$. For this reason, a deviation from H_0 suggests that either the survival or the censoring model is misspecified.

In order to test H_0, it is natural to assess the goodness-of-fit in terms of the distance between the empirical distribution function of Z, namely $F_n(z)=n^{-1}\sum_{i=1}^{n}I(Z_i\le z)$, and the proposed parametric estimator of $F_Z(z;\alpha)$. Specifically, the Cramér-von Mises type of statistic is given by

$$T_{CM}=\int_{\mathcal{R}} n\Big\{F_n(z)-F_Z(z;\hat{\alpha})\Big\}^2 dF_Z(z;\hat{\alpha}).$$

A large value of T_{CM} indicates a possible misspecification in the proposed model.

Now we will use a parametric bootstrap approach to determine the distribution of T_{CM} under the null hypothesis (Efron and Tibshirani, 1993; Emura and Michimae, 2017). Let B be the number of bootstrap samples. Then, we perform the goodness-of-fit test as follows:

1. Simulate bootstrap samples $(Z_i^b,\Delta_i^b,X_i^b,W_i^b)$, for $i=1,2,\cdots,n$, $b=1,2,\cdots,B$ under model (2.1) with $\alpha=\hat{\alpha}$.
2. Using the data generated in point 1, compute the bootstrap Cramér-von Mises statistic $T_{CM,b}^*$ for each bootstrap sample.
3. Reject H_0 with level ω if T_{CM} is greater than the $100\times(1-\omega)$ percent point of $\{T_{CM,b}^*, b=1,2,\cdots,B\}$.

5 Simulation study

In this section, we conduct a simulation study to show the effect of not taking into account the correlation between T and C in model (2.1). We will do this by comparing our estimation method with the method that assumes independent censoring. We will also evaluate the performance of the proposed method by comparing it with an estimator under a copula model. The copula model we will use in our simulations has the advantage

of being directly comparable with our model, whereas other competitors in the literature (like the weighted Cox model given in Collett (2015)) cannot be written in a form close to our model, and the parameters in these models have different interpretations and can therefore not be compared with the parameters in our model. Finally, we will assess the performance of the proposed goodness-of-fit test by means of simulations.

5.1 Comparison with the independent censoring model

The following data generating model is considered:

$$\begin{cases} \Lambda_\theta(T) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon_T \\ \Lambda_\theta(C) = \eta_0 + \eta_1 X_1 + \eta_2 X_2 + \epsilon_C, \end{cases} \tag{5.1}$$

where $\Lambda_\theta(\cdot)$ is the Yeo-Johnson transformation function defined in (2.3), $X_1 \sim \text{Bern}(0.5)$ and $X_2 \sim U[-1, 1]$. The vector of error terms (ϵ_T, ϵ_C) is drawn from a bivariate normal distribution with zero mean vector and $\sigma_T = 1$, $\sigma_C = 1.5$ and $\rho = 0.75$. We study two different settings. For setting 1, we set the regression parameters as follows: $\beta_0 = 2$, $\beta_1 = 1.2$, $\beta_2 = 1.5$ and $\eta_0 = 2.5$, $\eta_1 = 0.5$, $\eta_2 = 1$. Under this setting, the average proportion of censoring in the simulated data is approximately 45%. In setting 2, we set $\beta_0 = 3.25$, $\beta_1 = -0.45$, $\beta_2 = 1.2$ and $\eta_0 = 3.5$, $\eta_1 = 0.6$, $\eta_2 = 1$, so that the average proportion of censoring in the simulated data is approximately 25%. After simulating a set of normal transformed survival times and a second set of normal transformed censoring times, for each subject, we simply take the minimum to obtain the observed transformed survival time ($\Lambda_\theta(Z)$) and consequently the event indicator (Δ). Three different values of θ are considered ($\theta = 0$, 0.5 and 1.5), and we estimate θ along with the other model parameters to investigate their behavior in practice for two sample sizes, namely $n = 300$ and $n = 600$. The model estimation is done in *R* using the nonlinear optimization package *nloptr* for optimization and *numDeriv* for computing the Hessian matrix.

For each simulation setting, 2000 replicated datasets are created. For each of them, the model parameters are estimated under the dependent censoring model using the likelihood function given in (3.1), and the independent censoring model corresponding to $\rho = 0$ in (3.1). In Tables 1 and 2 we report the bias, the empirical standard deviation (ESD), the average of the model standard errors (ASE), the root mean squared error (RMSE), and the coverage rate of 95% confidence intervals (CR) to compare the fitted models.

The simulation results when the proportion of censoring is approximately 45% and $n = 300$ can be found in Table 1. From the table it is clear that the dependent censoring

model performs better than the model under independence, since it has much smaller bias and the coverage rate is close to the nominal level of 95% for all parameters. There is sometimes a slight increase in the variance with respect to the model under independence, but the RMSE is much smaller for the dependent censoring model. Comparing these results to those for $n = 600$ (see Table 7 in supplementary materials), as can be expected, the absolute bias becomes smaller as the sample size increases and/or the magnitude of θ decreases for the dependent censoring model. However, for the independent censoring model, we see that an increase in the sample size does not correspond to a decrease in the bias, which shows the inconsistency of the parameter estimators when non-zero association exists between T and C. Also, the coverage rates for the dependent censoring model improve for some parameters whereas the coverage rates for the independent censoring model decrease as we increase the sample size.

The simulation results for 25% censoring and $n = 300$ are shown in Table 2. The results in the table again show that the proposed dependent censoring model performs well: there is a much smaller bias in the parameter estimates compared to the independent censoring model, and the estimated coverage rates are reasonably close to the nominal level. Again we see that when we increase the sample size to $n = 600$ (Table 8 in supplementary materials), the biases and RMSE decrease. Comparing the coverage rates for 25% and 45% censoring for the dependent censoring model, we notice that the coverage rates are slightly worse for the case where we have 25% censoring. This is not surprising, since we need a reasonable number of uncensored and of censored observations in order to estimate well all model parameters. So, the model with 45% censoring is easier to estimate than the one with 25% censoring since we have enough observations on both T and C (in absolute terms) under the former censoring rate.

We also conduct a simulation study to assess the performance of the proposed method when the model for the censoring time is misspecified, since we like to know whether the estimated parameters are sensitive to the misspecification of the censoring model. To investigate this issue, we generate data similar to those in the previous setting under 45% censoring, except that the true censoring model is given by $\Lambda_\theta(C) = \eta_0 + \eta_1 X_1 + \eta_2 X_2 + \eta_3 X_2^2 + \epsilon_C$. Note that we fit a wrong censoring model $\Lambda_\theta(C) = \eta_0 + \eta_1 X_1 + \eta_2 X_2 + \epsilon_C$ without X_2^2. Then the effect of the misspecification of the censoring model on the survival model will be tested when X_2^2 has a moderate effect (with $\eta_3 = 1$) and a strong effect (with $\eta_3 = 2$). For each set of simulated data, the model parameters are estimated under the dependent censoring model. In Table 3 we present the bias, ESD, ASE, RMSE and

CR. The table shows that for $\eta_3 = 1$ the parameter estimators have a small bias and their coverage rates are close to the 95% nominal level. So the joint model does not suffer a lot from the misspecification of the censoring model. When the effect of X_2^2 is strong, we see an increase in bias and RMSE for the censoring model. However, the parameter estimates of the survival model and of the variance–covariance matrix are almost unaffected by the misspecification.

5.2 Comparison with the copula model

In this subsection, we will evaluate the performance of the proposed method by comparing it with an estimator constructed under a copula model allowing for dependent censoring. Unfortunately, none of the existing copula models is directly comparable to ours. A common assumption in the framework of copula models allowing for dependent censoring is the assumption that the copula is known and that the margins are completely unspecified or partially specified. Following this approach, we propose a Gaussian copula model, which has the same structure as our model except that one of the margins is unspecified. The proposed copula model is then given by

$$\begin{cases} \Lambda_\theta(T) = X^T\beta + \epsilon_T \\ \Lambda_\theta(C) = W^T\eta + \epsilon_C, \end{cases} \tag{5.2}$$

where

$$\begin{pmatrix} \Phi^{-1}(F_{\epsilon_T}(\epsilon_T)) \\ \epsilon_C \end{pmatrix} \sim N_2\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right), \tag{5.3}$$

and where F_{ϵ_T} is unknown, and ρ is known.

The estimation of this model is not straightforward when the error distribution is unknown, even under independent censoring. So, we will start by proposing a method for estimating this copula model. We follow a semi-parametric maximum likelihood approach. Note that the monotonicity of the transformations in the copula model allows the derivation of a likelihood function that can be factorized into the product of contributions from the survival and the censoring times. It follows that the likelihood function for the unknown

parameters $\xi = (\theta, \beta, \eta)$ and the unknown function F_{ϵ_T} can be written as

$$\begin{aligned} L(\xi, F_{\epsilon_T}) = \prod_{i=1}^{n} \Big\{ & f_{\epsilon_T}\big(\Lambda_\theta(Z_i) - X_i^T\beta\big) \\ & \times \Big[1 - \Phi\Big(\frac{\Lambda_\theta(Z_i) - W_i^T\eta - \rho\Phi^{-1}(F_{\epsilon_T}(\Lambda_\theta(Z_i) - X_i^T\beta))}{(1-\rho^2)^{1/2}}\Big)\Big]\Big\}^{\Delta_i} \\ & \times \Big\{\Big[1 - \Phi\Big(\frac{\Phi^{-1}(F_{\epsilon_T}(\Lambda_\theta(Z_i) - X_i^T\beta)) - \rho(\Lambda_\theta(Z_i) - W_i^T\eta)}{(1-\rho^2)^{1/2}}\Big)\Big] \\ & \times \phi\big(\Lambda_\theta(Z_i) - W_i^T\eta\big)\Big\}^{1-\Delta_i} \Lambda_\theta'(Z_i). \end{aligned} \tag{5.4}$$

Directly maximizing this likelihood is not feasible, as the likelihood contains the infinite dimensional parameter F_{ϵ_T} and its derivative. We will replace F_{ϵ_T} by a nonparametric estimator, and f_{ϵ_T} by a kernel estimator, which we will define below. We will also explain below how this nonparametric estimator can be used to estimate model (5.2)–(5.3). Note that this model has the extra difficulty that it imposes a copula model on the errors ϵ_T and ϵ_C and we do not know which of these two errors is the smallest, we only know the order between T and C, or equivalently between $\Lambda_\theta(T)$ and $\Lambda_\theta(C)$. This complicates the construction of the nonparametric estimator.

For a fixed value of $\xi = (\theta, \beta, \eta)$, the estimator of F_{ϵ_T} can be constructed as follows. Let $R_{(1)} < R_{(2)} < \ldots < R_{(m)} = \max\{R_i, i = 1, \ldots n\}$ be the distinct residual lifetimes $R_i = \Lambda_\theta(Z_i) - X_i^T\beta$, and let $\Delta_{(i)}, i = 1, \ldots, m$ be the corresponding censoring indicators. Here $m \leq n$ is the number of such distinct times. We suppose here that ties can occur among the uncensored residuals, or among the censored residuals, but that a censored residual cannot be equal to an uncensored residual. If this would happen, the censored residual is increased by a very small amount. Hence, $\Delta_{(i)}$ is well defined.

For fixed ξ, define now $F_{\epsilon_T,\xi}(t) = P(\Lambda_\theta(T) - X^T\beta \leq t)$. Note that

$$\begin{aligned} P(R \leq t, \Delta = 1) &= E(P(\Lambda_\theta(T) - X^T\beta \leq t, \Lambda_\theta(T) \leq \Lambda_\theta(C)|X, W)) \\ &= E\Big\{\int_{-\infty}^{t} \Big[1 - \Phi\Big(\frac{e^- - W^T\eta + X^T\beta - \rho\Phi^{-1}(F_{\epsilon_T,\xi}(e^-))}{(1-\rho^2)^{1/2}}\Big)\Big] dF_{\epsilon_T,\xi}(e)\Big\}. \end{aligned}$$

This suggests that in order to estimate $F_{\epsilon_T,\xi}(\cdot)$ for a fixed value of ξ, we solve the following equation (in $\hat{F}_{\epsilon_T,\xi}(\cdot)$) :

$$\begin{aligned} & n^{-1}\sum_{i=1}^{n} I(R_i \leq t, \Delta_i = 1) \\ & = \int_{-\infty}^{t} \Big\{1 - n^{-1}\sum_{i=1}^{n} \Phi\Big(\frac{e^- - W_i^T\eta + X_i^T\beta - \rho\Phi^{-1}(\hat{F}_{\epsilon_T,\xi}(e^-))}{(1-\rho^2)^{1/2}}\Big)\Big\} d\hat{F}_{\epsilon_T,\xi}(e). \end{aligned} \tag{5.5}$$

This shows that $\hat{F}_{\epsilon_T,\xi}$ is a nondecreasing step function with jumps only at the observed residual lifetimes. We can write (5.5) also as

$$\begin{aligned} & n^{-1}\sum_{i=1}^{n} I(R_i \le t, \Delta_i = 1) \\ & = \sum_{j=1}^{m} \big(\hat{F}_{\epsilon_T,\xi}(R_{(j)}) - \hat{F}_{\epsilon_T,\xi}(R_{(j-1)})\big) I(R_{(j)} \le t) \\ & \quad \times \Big\{1 - n^{-1}\sum_{i=1}^{n} \Phi\Big(\frac{R_{(j-1)} - W_i^T\eta + X_i^T\beta - \rho\Phi^{-1}(\hat{F}_{\epsilon_T,\xi}(R_{(j-1)}))}{(1-\rho^2)^{1/2}}\Big)\Big\}, \end{aligned} \tag{5.6}$$

where $R_{(0)} = -\infty$. This equation can be solved sequentially starting from $j = 1$ and going through the residuals $R_{(j)}$ from smallest to largest, using the following more direct formula that is equivalent to (5.6):

$$\hat{F}_{\epsilon_T,\xi}(R_{(j)}) \;=\; \hat{F}_{\epsilon_T,\xi}(R_{(j-1)}) + \frac{n^{-1}\sum_{i=1}^{n} I(R_i = R_{(j)}, \Delta_i = 1)}{1 - n^{-1}\sum_{i=1}^{n} \Phi\Big(\frac{R_{(j-1)} - W_i^T\eta + X_i^T\beta - \rho\Phi^{-1}(\hat{F}_{\epsilon_T,\xi}(R_{(j-1)}))}{(1-\rho^2)^{1/2}}\Big)}.$$

Since this only involves direct calculation, it is much easier than solving equation (5.6).

Next, we will use a kernel approach to obtain an estimator of the density $f_{\epsilon_T,\xi}$:

$$\hat{f}_{\epsilon_T,\xi}(t) = \frac{1}{h}\int K\Big(\frac{t-r}{h}\Big) d\hat{F}_{\epsilon_T,\xi}(r) = \frac{1}{h}\sum_{j=1}^{m} K\Big(\frac{t - R_{(j)}}{h}\Big) w_j,$$

where K is a smooth kernel function (we use a normal kernel in the simulations), h is a bandwidth parameter (set to $h = n^{-1/5}$), and $w_j = \hat{F}_{\epsilon_T,\xi}(R_{(j)}) - \hat{F}_{\epsilon_T,\xi}(R_{(j-1)})$. Finally, we plug in $\hat{F}_{\epsilon_T,\xi}$ and $\hat{f}_{\epsilon_T,\xi}$ in the likelihood function (5.4), and we maximize this likelihood that now only depends on ξ.

In order to compare our proposed estimator with the estimator under the copula model given above, we generate data from the following model:

$$\begin{cases} \Lambda_\theta(T) = \beta X + \epsilon_T \\ \Lambda_\theta(C) = \eta X + \epsilon_C, \end{cases}$$

where the transformation Λ_θ equals the Yeo-Johnson transformation with $\theta = 0.5$ and X follows a binomial distribution with equal probability. The other model parameters are set to $\beta = 0.75$ and $\eta = 1$. Since our method requires specification of the error distribution, we like to assess the performance of the proposed method when either the dependence structure or the marginal distribution is misspecified. We will also study the sensitivity of

the copula model when the dependence parameter ρ is misspecified. To investigate these issues, let $P(\epsilon_T \leq u, \epsilon_C \leq v) = \mathcal{C}\{F_{\epsilon_T}(u), F_{\epsilon_C}(v)\}$, where $\mathcal{C}$ is a parametric copula function. We consider the following scenarios:

Scenario 1: The copula $\mathcal{C}_\rho$ is a Gaussian copula or a t-copula with degrees of freedom $v = 5$, where $\rho = 0.75$, and the errors ϵ_T and ϵ_C follow a standard normal distribution. Here the assumption of linear relation holds true even though the dependence structure is misspecified under the t-copula.

Scenario 2: The copula $\mathcal{C}_\gamma$ is a Frank copula or a Gumbel copula, where the dependence parameter γ is chosen to yield Kendall's correlation $\tau = 0.54$, which corresponds to $\gamma = 6.5$ under the Frank copula and $\gamma = 2.2$ under the Gumbel copula. Note that for this value of τ, the correlation ρ is equal to 0.75. Again, both errors ϵ_T and ϵ_C follow a standard normal distribution. Here the assumption of linear dependence does not hold since both the Frank and the Gumbel copula introduce a non-linear dependence. We refer to the book of Nelsen (2006) for details about various copula models.

Scenario 3: The copula $\mathcal{C}_\rho$ is a Gaussian copula with $\rho = 0.75$. Under this scenario, the errors ϵ_T and ϵ_C follow a standard t-distribution with $v = 5$ degrees of freedom. Therefore, the assumption of normal marginals required by our method is not satisfied.

For each simulation scenario, a total of 1000 datasets with a sample size of $n = 500$ is considered. The average proportion of censored data is approximately 45%. For each dataset, the model parameters are estimated based on our dependent censoring model and based on the copula model using the true value of $\rho = 0.75$ and using a misspecified value of $\rho = 0.45$. Table 4 presents the bias, ESD and RMSE. The proposed method has a much smaller bias and RMSE in all scenarios except Scenario 3. For the considered scenarios, the misspecification in the dependence structure has little influence on the proposed model. Under Scenario 3, the proposed model gives inconsistent estimates for the model parameters, whereas the copula model provides a slightly smaller bias when ρ is correctly specified. So, our model seems to be sensitive to the misspecification of the marginal distributions. In order to decide whether the method can be used or not, we recommend the user of our method to carry out the goodness-of-fit test given in Section 4. On the other hand, when ρ is misspecified as $\rho = 0.45$, Table 4 shows an increase in absolute bias and RMSE for the copula model. This implies that prior knowledge of the dependence parameter is important to get good estimates. In all scenarios, the proposed method manifests advantages over the estimator under the copula model in terms of ESD and RMSE.

5.3 Performance of the goodness-of-fit test

We evaluate the performance of the goodness-of-fit test (Section 4) using simulations. The null hypothesis to be tested is the validity of the proposed dependent censoring model. The data are generated based on model (5.1), where $\Lambda_\theta(\cdot)$ is the Yeo-Johnson transformation function, $X_1 \sim \text{Bern}(0.5)$ and $X_2 \sim U[-1, 1]$. We draw error terms ϵ_T and ϵ_C from a bivariate normal distribution and from a bivariate t-distribution with either $v = 5$ or $v = 10$ degrees of freedom. The other model parameters are set to $\beta_0 = 2.25$, $\beta_1 = 1$, $\beta_2 = 1.7$, $\eta_0 = 2.6$, $\eta_1 = 0.5$, $\eta_2 = 1.2$, $\theta = 0.5$, $\sigma_T = 1.2$, $\sigma_C = 1.7$ and $\rho = 0.7$. For these parameter values, the average proportion of censored observations in the simulated data is approximately 42%. We create 500 data sets of size $n = 300$ and $n = 500$, and for each data set we compute the Cramér-von Mises statistic and test the goodness-of-fit hypothesis based on $B = 500$ bootstrap samples. So as to assess the Type I error rate and the power of the test, we count the number of rejections under levels 5% and 10% over 500 runs.

In Table 5 we show the performance of the goodness-of-fit test. When the data are generated from a bivariate normal distribution (so when the model is valid), we see that the percentage of times H_0 is rejected is close to the nominal levels. Moreover, the mean of the test statistic T_{CM} is close to the mean of the average bootstrap statistic $B^{-1}\sum_{b=1}^{B} T^*_{CM,b}$.

To get a better understanding of the power of the test, data are generated from a bivariate t with $v = 5$ degrees of freedom. When the data are generated from this distribution, we have seen in Section 5.2 that the proposed model does not fit well. Table 5 shows that the rejection rates are much higher than the specified levels. As expected, the powers become larger when the sample size increases from $n = 300$ to 500. It can be seen that the rejection rate is 0.488 (at level 0.05) under $n = 500$, showing about 48.8% power to reject the null hypothesis. Indeed, the bivariate normal and bivariate t (with small degrees of freedom) show different marginal behaviour as also shown in our simulation. As we increase the degrees of freedom to $v = 10$, the rejection rates are slightly higher than the specified levels, so the power is much smaller than the power under $v = 5$. For example, the rejection rate is 0.095 (at level 0.05) under $n = 500$, showing about 9.5% power to reject the null hypothesis. This is in agreement with our expectation that as the degrees of freedom increase, the t-distribution resembles a normal distribution. In summary, the proposed goodness-of-fit test shows a good control of the Type I error rate and exhibits a moderate power against a wrong null hypothesis.

6 Data application

We applied the proposed model and estimation method to two data examples. The analysis of the first data set is given below in detail, and concerns a liver transplant data set, given in the book of Collett (2015). For the second data set we refer to the online supplementary material.

In the UK, when a patient has been judged to require a liver transplant, he/she is added to the enrolment list. However, due to a national deficiency of livers some patients may die while waiting for their transplant. The aim of this data analysis is to identify the factors affecting time to death, which would inform policy makers of the selection of patients for transplantation.

The data given in Collett (2015) consist of 281 adults with primary biliary cirrhosis who were enrolled for a liver transplant in the UK in the five-year period starting from 1 January 2006, and the response variable of interest is the time from being registered for a liver transplant to death. Patients who got a transplant are considered to be censored at the time of the transplant, and individuals who were expelled from the study because of their poor health condition are categorized as dead at the time of the removal since their condition had deteriorated to the point where transplantation is no longer an alternative. Removal of a patient from the waiting list for transplantation is a form of dependent censoring since the livers were given on the basis of the patient's health condition. This means that the more the patient is seriously ill, the more likely it is that a liver will be allocated to that patient. Hence, patients who get a transplant tend to be those who are closer to death. The censoring time due to transplant is then dependent on the time to death without a transplant. About 27% of patients died while waiting for a transplant, whereas the remaining patients received a transplant. In addition to the observed survival time, the following characteristics of the patients were obtained: the age of the patient in years, their gender (1 = male, 0 = female), their body mass index in kg/m^2 (BMI), and their UKELD score, which is a UK end-stage liver disease score, where higher values correspond to a disease of greater severity. The UKELD score is calculated based on several clinical measurements such as international normalized ratio, serum bilirubin, sodium and serum creatinine (see details in Barber et al. (2011)). Also it seems natural to believe that the time until death while waiting for a liver transplant depends on the UKELD score, and on the time until transplant (for a given UKELD score). This relationship is shown in Figure 1, where we distinguish between non-censored (dead) and censored patients.

In a previous analysis of the data set (see Collett, 2015), a significant effect of the

UKELD score on both the survival and the censoring time was found. Hence, we fit our joint model by excluding the variables that have no significant effect on the censoring time, namely age, gender and BMI, but we will continue to use a survival time model that contains all four explanatory variables. The parameter estimates, model-based standard errors (SE), and p-values are shown in Table 6. Note that the p-values are computed based on a Wald test using the model-based standard errors. We found the estimated transformation parameter $\hat{\theta} = 1.76$ and the association parameter $\hat{\rho} = 0.73$ for the dependent censoring model. This is a strong correlation that will probably induce bias in the parameter estimates if this correlation would not be correctly acknowledged in the modeling process.

We therefore also provide in Table 6 the parameter estimates under the independent censoring model for purposes of comparison. The parameter estimates are somewhat different for the dependent and the independent censoring model, especially for the age, which shows that the association between the survival and censoring time affects the parameter estimates. In the independent censoring model, both the age and the UKELD score are highly significant (p-value < 0.014), whereas age ceases to be significant in the dependent censoring model at the 5% level.

We also compare our results for the dependent censoring model to the results obtained by using a weighted Cox model, where the IPCW method is used to adjust for dependent censoring as given in the book of Collett (2015) (Example 14.3 on page 467). The results are also given in Table 6. We see that the same variables are significant and insignificant in the two models, namely the UKELD score is in both models significant and the other variables are in both models insignificant. However, BMI negatively affects the survival of patients in our model, whereas it positively affects the survival of patients in the weighted Cox model. Other covariates such as age, gender, and UKELD score have similar impacts on survival in the two models.

Table 6 also shows the parameter estimates, bootstrap standard errors (BSE) and p-values under the copula model discussed in Section 5.2. Since the assumed copula needs a pre-specification of the dependence parameter, we select $\rho = 0.75$ which is very close to the estimated dependence parameter under the proposed model. The copula model is estimated based on the same set of covariates as in the dependent censoring model except that the continuous covariates are standardized so as to overcome convergence problems for this model. For computing the bootstrap standard errors, 500 bootstrap samples were drawn using a naive resampling of the original data. The table shows that age, gender, and BMI are not related to time to death, whereas the UKELD score is significant for time

to death. In conclusion, all models show that the UKELD score is an important covariate in predicting the risk of death.

In order to get more insight in the effect of dependent censoring, Figure 2 shows the estimated survival rates for a female patient aged 50 years with a UKELD score of 57 and a BMI of 25, under a dependent and independent censoring model for the time to death. From this figure, it is clear that failure to take the dependent censoring into account results in overestimated survival rates. In particular, if no account is made for dependent censoring, the survival rate at six months is estimated to be 88%, but after taking dependent censoring into account, the estimate is 66%. The 80% survival rate is overestimated by almost 2.8 months. Not accounting for dependent censoring can therefore result in misleading estimates of survival rates for patients awaiting a liver transplant. Figure 2 also displays the estimated survival rates under a weighted Cox model and under a copula model with $\rho = 0.75$ to compare them with our estimates. The results show a very good agreement between the estimates obtained from the dependent censoring model and the copula model. However, the estimates from the weighted Cox model deviate much from the two curves, particularly at middle time points. Given that we obtain a strong positive correlation under the dependent censoring model, it is expected that the survival curve from this model lies below the survival curve from the independent censoring model. The reason is that a positive correlation between the survival and censoring time indicates that censored patients will likely have their event soon after their censoring time, whereas under the independent censoring model it is believed that for a censored observation the (unobserved) survival time can be any value larger than the censoring time, not necessarily a value close to the censoring time. Hence, it is clear that under the independent censoring model, we obtain an estimator of the survival function that is too positive, i.e. an estimator that overestimates the true survival function. See also Etzioni et al. (1999) for a detailed discussion of this phenomenon. It should be noted that our model takes full potential of adjusting for dependent censoring by estimating the dependence parameter, whereas the usefulness of the IPCW method is conditional on the availability of important covariates that predict the censoring probability. On the other hand, as we decrease the dependence parameter of the copula model (not shown here), the survival estimates from this model deviate from our estimates. This is because the copula approach is sensitive to misspecification of the dependence parameter (Zheng and Klein, 1995).

One of the quantities of interest in survival analysis is the predicted median survival time. The median survival time is any time t satisfying $S_{T|X}(t) = 0.5$. Let $\tilde{T}$ be the original

survival time of interest, so $\tilde{T} = \exp(T)$. Using equation (2.4) we can estimate the median survival time by

$$\widehat{\text{med}(\tilde{T}|X)} = \exp\{\Lambda_{\hat{\theta}}^{-1}(\hat{\sigma}_T \Phi^{-1}(0.5) + X^T\hat{\beta})\} = \exp\left\{\Lambda_{\hat{\theta}}^{-1}(X^T\hat{\beta})\right\},$$

where θ and β are replaced by their corresponding estimates.

After we compute the predicted median survival time, we can use the Delta method to compute a confidence interval for the predicted median time. Treating $\widehat{\text{med}(\tilde{T}|X)}$ as a function of $\hat{\theta}$ and $\hat{\beta}$, the standard error can be calculated as

$$\widehat{SE}_m = \left\{ \begin{pmatrix} \frac{\partial \widehat{\text{med}(\tilde{T}|X)}}{\partial \hat{\beta}} \\ \frac{\partial \widehat{\text{med}(\tilde{T}|X)}}{\partial \hat{\theta}} \end{pmatrix}^T \hat{V}_{\hat{\beta},\hat{\theta}} \begin{pmatrix} \frac{\partial \widehat{\text{med}(\tilde{T}|X)}}{\partial \hat{\beta}} \\ \frac{\partial \widehat{\text{med}(\tilde{T}|X)}}{\partial \hat{\theta}} \end{pmatrix} \right\}^{1/2},$$

where $\hat{V}_{\hat{\beta},\hat{\theta}}$ is the estimated variance-covariance matrix of $\hat{\beta}$ and $\hat{\theta}$ obtained from Theorem 7.2. Then the $(1-\omega)\%$ confidence interval is given by

$$\widehat{\text{med}(\tilde{T}|X)} - z_{1-\omega/2}\,\widehat{SE}_m < \text{med}(\tilde{T}|X) < \widehat{\text{med}(\tilde{T}|X)} + z_{1-\omega/2}\,\widehat{SE}_m,$$

where $z_{1-\omega/2}$ is the $1-\omega/2$-quantile of the standard normal distribution.

For the real data example, we calculate the predicted median survival time for a female patient aged 58 years with a UKELD score of 58 and a BMI of 19.5. The estimated median time is 6.4 months with a 95% confidence interval ranging from 2 to 10.8 months.

Finally, we performed our goodness-of-fit test (see Section 4) on the dependent censoring model. The Cramér-von Mises type statistic T_{CM} produced a p-value of 0.314 based on 500 bootstrap replications. Therefore, there is no evidence against the fitted dependent censoring model at the 5% level.

7 Discussion and future research

When modeling survival data, it is common to assume that given a set of covariates, the survival time is independent of the censoring time. When this assumption is likely to be violated, the dependent censoring must be taken into account in order to obtain valid inferences.

In this paper, we proposed a flexible parametric model for the association between the survival and the censoring time. A bivariate normal distribution is assumed to induce this association. But since both the survival and the censoring time are typically skewed to

the right, these times are first transformed and the bivariate normality of the transformed times is assumed after adjusting for the possible effect of covariates. The model has the advantage that it is able to estimate a transformation parameter from the data rather than working with the classical logarithmic transformation. In addition, a remarkable advantage of our model over the well known copula approach is that the proposed model permits the estimation of the association parameter together with other model parameters instead of assuming the dependence parameter to be known.

The asymptotic normality of the parameter estimators is shown both when the model is correctly specified and when it is misspecified. Note that as a by-product of our method we obtain the parameter estimators of the model for the censoring time. Also note that our model can be used when the survival and censoring time are independent given the covariates. The estimated model parameters will then be close to those obtained under the assumption that the errors ϵ_T and ϵ_C are normal and independent, and hence there is only a small price to pay for this extra flexibility in the model.

A simulation study demonstrates the good performance of the dependent censoring model compared to the model under independence. Simulations (not shown here) also show that the method works well for a large range of censoring rates, as long as the sample size is reasonably large. Moreover, confidence intervals are obtained for the model parameters, whose coverage rate is close to the 95% nominal level, which is an indication that the asymptotic normality of our estimators is approximately satisfied even for small samples. Contrasting our model with the copula model, our model exhibits a major advantage in terms of reduction in bias and RMSE when the dependence structure is misspecified. However, when the marginal distributions are misspecified, our model gives biased estimates. To help the user of our model in checking the quality of the fitted model, we built up a formal goodness-of-fit test with the aid of parametric bootstrap. The proposed test is based on the distance between the model-based estimator and model-free estimator of the distribution function of the observed survival times. In the simulations, this test shows a good control of the Type I error rate and reveals a moderate power in rejecting a false null hypothesis.

The proposed approach can not only be followed for the analysis of medical survival data, but can also be used for analysing data on market disequilibria with minor modifications. Specifically, our approach is very useful if the quantity demanded depends on the quantity supplied, which is often the case in e.g. the housing market. Note that the data consist of the minimum of the demand and the supply, and on the information about which

of these two is the smallest, together with possible covariates that influence the market.

The limitation of the proposed model is that the model requires an adequate number of censored observations in order to estimate well all model parameters when the sample size is small. This is especially the case for the correlation parameter, whose estimator might be biased when the number of censored and/or uncensored observations is small. In the proposed approach, the censoring model is equally important as the survival model. Thus, it is worth noting that a decrease in the censoring rate does not necessarily correspond to a decrease in bias and an increase in the coverage rate as in the model under independence. However, this is not a problem when we have a large sample size as shown by our simulations (e.g., when $n = 600$ in the supplementary material, we observe small biases and coverage rates close to the nominal level for all parameters under both 25% and 45% censoring). So, the model parameters are well estimated irrespective of the proportion of censoring for a large sample size. Another limitation is that the proposed model assumes the same transformation parameter (θ) for both T and C. Without this condition, the derivation of the likelihood will be much more difficult, as it is then no longer clear whether the transformed T is smaller or larger than the transformed C, since the censoring indicator of the transformed variables depends on the unknown transformation parameters.

The proposed model is a first step towards a new stream of models that take the possible association between the survival and the censoring time into account, and it opens the door for many possible extensions and adaptations. A first possible extension that is currently under investigation is the extension to dependent competing risks, possibly including the case of administrative censoring as one of the risks, which can be assumed to be independent of all other risks. This will allow us to analyse data on dependent causes of death that are often encountered in practice. A second future line of research is the adaptation of the model to semi- or nonparametric regression functions, using splines, orthogonal series or kernel methods. This will also be studied in the near future. Finally, even more flexibility in the model can be obtained by allowing heteroscedasticity of the errors in the model, or by allowing the errors to follow other distributions than the bivariate normal distribution, although it is not clear for the moment under which other distributions the association parameter will be identifiable.

Supplementary materials

Tables 7 and 8 and software in the form of R code with complete documentation are available on Github (`https://github.com/N2143/Supplementarymaterials`), whereas the analysis of a second data example can be found in the online supplementary material.

Acknowledgments

The authors are very grateful to David Collett, Dan Jackson and Martin Law for providing the data on liver transplants that were used in this paper.

Conflict of interest

The authors have declared no conflict of interest.

Appendix A: Proof of the identifiability result

We start with a preliminary result that is needed in the proof of the main identifiability result, and which is interesting on its own.

Proposition 7.1. *Given assumptions* $(A1)-(A4)$*, suppose that the survival time* T_1 *satisfies* $\Lambda_{\theta_1}(T_1)-X^T\beta_1 \sim N(0,\sigma^2_{T_1})$ *and that the survival time* T_2 *satisfies* $\Lambda_{\theta_2}(T_2)-X^T\beta_2 \sim N(0,\sigma^2_{T_2})$*. Let* $f_{T_1|X}(\cdot)$ *and* $f_{T_2|X}(\cdot)$ *denote the probability density functions of* T_1 *and* T_2 *given* X*, respectively. Then,*

$$\lim_{t\to\pm\infty}\frac{f_{T_1|X}(t\mid x)}{f_{T_2|X}(t\mid x)}=1 \text{ for almost every } x \Longleftrightarrow \theta_1=\theta_2,\beta_1=\beta_2,\sigma_{T_1}=\sigma_{T_2}.$$

A similar result holds for the density of the censoring time C *given* W*.*

Proof. Write

$$1=\lim_{t\to\pm\infty}\frac{f_{T_1|X}(t\mid x)}{f_{T_2|X}(t\mid x)}=\frac{\sigma_{T_2}}{\sigma_{T_1}}\lim_{t\to\pm\infty}\frac{\phi\big(\frac{\Lambda_{\theta_1}(t)-x^T\beta_1}{\sigma_{T_1}}\big)}{\phi\big(\frac{\Lambda_{\theta_2}(t)-x^T\beta_2}{\sigma_{T_2}}\big)}\cdot\frac{\Lambda'_{\theta_1}(t)}{\Lambda'_{\theta_2}(t)}=\frac{\sigma_{T_2}}{\sigma_{T_1}}\lim_{t\to\pm\infty}\frac{K_{\theta_1,x^T\beta_1,\sigma_{T_1}}(t)}{K_{\theta_2,x^T\beta_2,\sigma_{T_2}}(t)}. \tag{7.1}$$

We know from condition (A4) that the right hand side of (7.1) equals either 0 or $+\infty$ when $\theta_1 \neq \theta_2$, which is impossible. Hence, $\theta_1 = \theta_2$. Straightforward calculus now shows that $(x^T\beta_1, \sigma_{T_1})$ has to be equal to $(x^T\beta_2, \sigma_{T_2})$ for almost every x. Writing $\beta_j^T = (\mu_j, \lambda_j^T)$, $j = 1, 2$, we have that $\tilde{x}^T(\lambda_1 - \lambda_2) = \mu_2 - \mu_1$ for almost every x. Hence, $\mathrm{Var}(\tilde{X}^T(\lambda_1 - \lambda_2)) = 0$. It now follows that $\gamma_1 = \gamma_2$ if $(A2)$ holds, and hence we also have that $\lambda_1 = \lambda_2$. □

We are now ready to show that the model parameters $(\theta, \beta, \eta, \sigma_T, \sigma_C, \rho)$ in model (2.1)–(2.2) are identifiable from the distribution of (Z, Δ, X, W).

Proof of Theorem 3.1. The proof is inspired by the proof of Theorem 2 in Basu and Ghosh (1978), which holds when no covariates and no transformation are included in the dependent censoring model. We know from (2.5) that

$$\begin{aligned} f_{Z_j,\Delta_j|X,W}(z, 1 \mid x, w; \alpha_j) \;=\; & \frac{1}{\sigma_{T_j}}\Big[1 - \Phi\Big(\frac{(1 - \rho_j\sigma_{C_j}/\sigma_{T_j})\Lambda_{\theta_j}(z) - w^T\eta_j + \rho_j(\sigma_{C_j}/\sigma_{T_j})x^T\beta_j}{\sigma_{C_j}(1-\rho_j^2)^{1/2}}\Big)\Big] \\ & \times\phi\Big(\frac{\Lambda_{\theta_j}(z) - x^T\beta_j}{\sigma_{T_j}}\Big)\Lambda'_{\theta_j}(z), \text{for } j = 1, 2. \end{aligned}$$

Define $\gamma_{11} = 1 - \rho_1\sigma_{C_1}/\sigma_{T_1}$, $\gamma_{12} = 1 - \rho_1\sigma_{T_1}/\sigma_{C_1}$, $\gamma_{21} = 1 - \rho_2\sigma_{C_2}/\sigma_{T_2}$, $\gamma_{22} = 1 - \rho_2\sigma_{T_2}/\sigma_{C_2}$, and define a variable ξ_{j1} $(j = 1, 2)$, whose distribution for given X and W is given by

$$(\xi_{j1} \mid X = x, W = w) \sim N\Big(\frac{w^T\eta_j - \rho_j(\sigma_{C_j}/\sigma_{T_j})x^T\beta_j}{\gamma_{j1}}, \frac{\sigma_{C_j}^2(1-\rho_j^2)}{\gamma_{j1}^2}\Big).$$

We consider a number of cases, depending on the signs of the γ coefficients.

Case 1: All $\gamma_{jk} > 0$, $j, k = 1, 2$. The positivity of γ_{j1} $(j = 1, 2)$ implies that

$$\begin{aligned} f_{Z_j,\Delta_j|X,W}(z, 1 \mid x, w; \alpha_j) \;&=\; P\big(\xi_{j1} > \Lambda_{\theta_j}(z) \mid X = x, W = w\big)\frac{1}{\sigma_{T_j}}\phi\Big(\frac{\Lambda_{\theta_j}(z) - x^T\beta_j}{\sigma_{T_j}}\Big)\Lambda'_{\theta_j}(z) \\ &=\; P(\xi_{j1} > \Lambda_{\theta_j}(z) \mid X = x, W = w) f_{T_j|X}(z \mid x). \end{aligned}$$

We know that $f_{Z_1,\Delta_1|X,W}(z, 1 \mid x, w; \alpha_1) = f_{Z_2,\Delta_2|X,W}(z, 1 \mid x, w; \alpha_2)$ for almost every (x, w) and z. We also know that $\lim_{z\to-\infty} P(\xi_{11} > \Lambda_{\theta_1}(z) \mid X = x, W = w) = \lim_{z\to-\infty} P(\xi_{21} > \Lambda_{\theta_2}(z) \mid X = x, W = w) = 1$. Hence, $\lim_{z\to-\infty} f_{T_1|X}(z \mid x) = \lim_{z\to-\infty} f_{T_2|X}(z \mid x)$ for almost every x, and application of Proposition 7.1 now implies that $\theta_1 = \theta_2, \beta_1 = \beta_2$ and $\sigma_{T_1} = \sigma_{T_2}$. Repeating the same arguments but with $\Delta_j = 1$ replaced by $\Delta_j = 0$, entails that $\eta_1 = \eta_2$ and $\sigma_{C_1} = \sigma_{C_2}$, using this time the fact that $\gamma_{12} > 0$ and $\gamma_{22} > 0$. Finally, to identify the correlation parameter, note that we have from (2.7) that

$$\Phi\Big(\frac{\Lambda_{\theta_1}(z) - x^T\beta_1}{\sigma_{T_1}}, \frac{\Lambda_{\theta_1}(z) - w^T\eta_1}{\sigma_{C_1}}; \rho_1\Big) = \Phi\Big(\frac{\Lambda_{\theta_1}(z) - x^T\beta_1}{\sigma_{T_1}}, \frac{\Lambda_{\theta_1}(z) - w^T\eta_1}{\sigma_{C_1}}; \rho_2\Big)$$

for almost every (x, w), from which it follows that $\rho_1 = \rho_2$.

Case 2: One of $(\gamma_{11}, \gamma_{12})$ is positive, and one of $(\gamma_{21}, \gamma_{22})$ is positive. Assume that $\gamma_{11} > 0$ and $\gamma_{12} < 0$. Now either $(\gamma_{21} > 0, \gamma_{22} < 0)$ or $(\gamma_{21} < 0, \gamma_{22} > 0)$. Let us first assume that $(\gamma_{21} > 0, \gamma_{22} < 0)$. Define this time the distribution of ξ_{j2} $(j = 1, 2)$ for given X and W by

$$(\xi_{j2} \mid X = x, W = w) \sim N\Big(- \frac{x^T\beta_j - \rho_j(\sigma_{T_j}/\sigma_{C_j})w^T\eta_j}{\gamma_{j2}}, \frac{\sigma_{T_j}^2(1-\rho_j^2)}{\gamma_{j2}^2}\Big).$$

Then the negativity of γ_{j2} $(j = 1, 2)$ implies that

$$\begin{aligned} f_{Z_j,\Delta_j|X,W}(z, 0 \mid x, w; \alpha_j) &= P\big(\xi_{j2} > -\Lambda_{\theta_j}(z) \mid X = x, W = w\big)\frac{1}{\sigma_{C_j}}\phi\Big(\frac{\Lambda_{\theta_j}(z) - w^T\eta_j}{\sigma_{C_j}}\Big)\Lambda'_{\theta_j}(z) \\ &= P(\xi_{j2} < \Lambda_{\theta_j}(z) \mid X = x, W = w) f_{C_j|W}(z|w). \end{aligned}$$

It is given that $f_{Z_1,\Delta_1|X,W}(z, 0 \mid x, w; \alpha_1) = f_{Z_2,\Delta_2|X,W}(z, 0 \mid x, w; \alpha_2)$ for almost every (x, w) and z. We know that $\lim_{z\to+\infty} P(\xi_{12} < \Lambda_{\theta_1}(z) \mid X = x, W = w) = \lim_{z\to+\infty} P(\xi_{22} < \Lambda_{\theta_2}(z) \mid X = x, W = w) = 1$, which implies $\lim_{z\to+\infty} f_{C_1|W}(z \mid w) = \lim_{z\to+\infty} f_{C_2|W}(z \mid w)$ for almost every w. Application of Proposition 7.1 to the latter equality now implies that $\theta_1 = \theta_2, \eta_1 = \eta_2$ and $\sigma_{C_1} = \sigma_{C_2}$. For the case $\Delta_j = 1$ and $(\gamma_{11} > 0, \gamma_{21} > 0)$, we have already shown the result under *Case 1.*

Similarly, let $(\gamma_{21} < 0, \gamma_{22} > 0)$. First consider $\gamma_{11} > 0$ and $\gamma_{21} < 0$, then we obtain

$$\begin{aligned} &P(\xi_{11} > \Lambda_{\theta_1}(z) \mid X = x, W = w) \\ &= P(\xi_{21} < \Lambda_{\theta_2}(z) \mid X = x, W = w)\big\{f_{T_1|X}(z \mid x)\big\}^{-1} f_{T_2|X}(z \mid x). \end{aligned}$$

Taking the limits on both sides as $z \to -\infty$ and using Proposition 7.1, the left hand side goes to 1, however, the right hand side does not go to 1, which is a contradiction. Following similar arguments when $\Delta_j = 0$, $\gamma_{12} < 0$ and $\gamma_{22} > 0$, one can show that the result is again a contradiction. Hence, these possibilities can not happen in practice.

Other cases, for example, when one (or more) of the $\gamma_{jk} = 0$ or one of γ_{jk} is negative can be considered, but we omit the details. □

Appendix B: Asymptotic results

Let $\alpha^* = (\theta^*, \beta^*, \eta^*, \sigma_T^*, \sigma_C^*, \rho^*)$ be the parameter vector that minimizes the Kullback-Leibler Information Criterion (KLIC), given by

$$E\Big[\log\Big\{\frac{f_{Z,\Delta|X,W}(Z, \Delta \mid X, W)}{f_{Z,\Delta|X,W}(Z, \Delta \mid X, W; \alpha)}\Big\}\Big],$$

where $f_{Z,\Delta|X,W}$ is the true density of (Z,Δ) given (X,W), and where the expectation is taken with respect to the true density $f_{Z,\Delta,X,W}$ (also in what follows). We then have the following results on consistency and asymptotic normality of our MLE estimator.

Theorem 7.1. *Under assumptions A*1 *to A*3 *in White (1982),*

$$(\hat{\theta},\hat{\beta},\hat{\eta},\hat{\sigma}_T,\hat{\sigma}_C,\hat{\rho}) \xrightarrow{P} (\theta^*,\beta^*,\eta^*,\sigma_T^*,\sigma_C^*,\rho^*) \quad as\ n\to\infty.$$

If the model is correctly specified the KLIC attains its unique minimum at $\alpha^* = \alpha$, which means that $\hat{\alpha}$ is a consistent estimator of the true parameter vector α in that case.

Theorem 7.2. *Under assumptions A*1 *to A*6 *in White (1982),*

$$n^{1/2}\Big((\hat{\theta},\hat{\beta},\hat{\eta},\hat{\sigma}_T,\hat{\sigma}_C,\hat{\rho}) - (\theta^*,\beta^*,\eta^*,\sigma_T^*,\sigma_C^*,\rho^*)\Big) \xrightarrow{d} N(0,V),$$

where $V = A(\alpha^*)^{-1}B(\alpha^*)A(\alpha^*)^{-1}$*, with*

$$A(\alpha) = \Big(E\Big\{\frac{\partial^2}{\partial\alpha_i\partial\alpha_j}\log f_{Z,\Delta|X,W}(Z,\Delta \mid X,W;\alpha)\Big\}\Big)_{i,j=1}^{p+q+4},$$

$$B(\alpha) = \Big(E\Big\{\frac{\partial}{\partial\alpha_i}\log f_{Z,\Delta|X,W}(Z,\Delta \mid X,W;\alpha)\cdot\frac{\partial}{\partial\alpha_j}\log f_{Z,\Delta|X,W}(Z,\Delta \mid X,W;\alpha)\Big\}\Big)_{i,j=1}^{p+q+4}.$$

Note that if the model is correctly specified, the variance-covariance matrix V in Theorem 7.2 is equal to $A(\alpha)^{-1}$, the inverse of Fisher's information matrix. The regularity conditions of White (1982) are assumptions regarding the true density $f_{Z,\Delta|X,W}$, the density $f_{Z,\Delta|X,W}(\cdot;\alpha)$ under our assumed model and its derivatives both with respect to α and z, and the parameter vector α.

References

Barber, K., Madden, S., Allen, J., Collett, D., Neuberger, J. and Gimson, A. (2011). Elective liver transplant list mortality: development of a united kingdom end-stage liver disease score, *Transplantation* **92**: 469–476.

Basu, A. P. (1988). Multivariate exponential distributions and their applications in reliability, *Handbook of Statistics* **7**: 467–477.

Basu, A. P. and Ghosh, J. K. (1978). Identifiability of the multinormal and other distributions under competing risks model, *Journal of Multivariate Analysis* **8**: 413–429.

Box, G. E. P. and Cox, D. R. (1964). An analysis of transformations, *Journal of the Royal Statistical Society – Series B* **26**: 211–243.

Braekers, R. and Veraverbeke, N. (2005). A copula-graphic estimator for the conditional survival function under dependent censoring, *Canadian Journal of Statistics* **33**: 429–447.

Collett, D. (2015). *Modelling Survival Data in Medical Research*, third edn, Taylor and Francis.

Efron, B. and Tibshirani, R. J. (1993). *An Introduction to the Bootstrap*, Chapman & Hall.

Emoto, S. E. and Matthews, P. C. (1990). A Weibull model for dependent censoring, *The Annals of Statistics* **18**: 1556–1577.

Emura, T. and Chen, Y. (2018). *Analysis of Survival Data with Dependent Censoring: Copula-Based Approaches*, Springer.

Emura, T. and Michimae, H. (2017). A copula-based inference to piecewise exponential models under dependent censoring, with application to time to metamorphosis of salamander larvae, *Environ Ecol Stat* **24**: 151–173.

Etzioni, R. D., Feuer, E. J., Sullivan, S. D., Lin, D., Hu, C. and Ramsey, S. D. (1999). On the use of survival analysis techniques to estimate medical care costs, *Journal of Health Economics* **18**: 365–380.

Huang, X. and Zhang, N. (2008). Regression survival analysis with an assumed copula for dependent censoring: A sensitivity analysis approach, *Biometrics* **64**: 1090–1099.

Jackson, D., White, I. R., Seaman, S., Evans, H., Baisley, K. and Carpenter, J. (2014). Relaxing the independent censoring assumption in the Cox proportional hazards model using multiple imputation, *Statistics in Medicine* **33**: 4681–4694.

Jones, M. C. and Pewsey, A. (2009). Sinh-arcsinh distributions, *Biometrika* **96**: 761–780.

Nádas, A. (1971). The distribution of the identified minimum of normal pair determines the distribution of the pair, *Technometrics* **13**: 201–202.

Nelsen, R. B. (2006). *An Introduction to Copulas*, second edn, Springer.

Rivest, L. P. and Wells, M. T. (2001). A martingale approach to the copula-graphic estimator for the survival function under dependent censoring, *Journal of Multivariate Analysis* **79**: 138–155.

Staplin, N., Kimber, A., Collett, D. and Roderick, P. (2015). Dependent censoring in piecewise exponential survival models, *Statistical Methods in Medical Research* **24**: 325–341.

Sujica, A. and Van Keilegom, I. (2018). The copula-graphic estimator in censored non-parametric location-scale regression models, *Econometrics and Statistics* **7**: 89–114.

Tsiatis, A. (1975). A nonidentifiability aspect of the problem of competing risks, *Proceedings of the National Academy of Sciences* **72**: 20–22.

White, H. (1982). Maximum likelihood estimation of misspecified models, *Econometrica* **50**: 1–25.

Yeo, I. and Johnson, R. A. (2000). A new family of power transformations to improve normality or symmetry, *Biometrika* **87**: 954–959.

Zheng, M. and Klein, J. P. (1995). Estimates of marginal survival for dependent competing risks based on an assumed copula, *Biometrika* **82**: 127–138.

Table 1: Simulation results for 45% censoring and $n = 300$. The true parameters are $(\beta_0, \beta_1, \beta_2) = (2, 1.2, 1.5)$, $(\eta_0, \eta_1, \eta_2) = (2.5, 0.5, 1)$, and $(\sigma_T, \sigma_C, \rho) = (1, 1.5, 0.75)$

θ	0					0.5					1.5				
Par.	Bias	ESD	ASE	RMSE	CR	Bias	ESD	ASE	RMSE	CR	Bias	ESD	ASE	RMSE	CR
	Dependent censoring model														
β_0	-0.001	0.144	0.144	0.143	0.948	-0.009	0.161	0.161	0.161	0.952	-0.019	0.168	0.169	0.169	0.948
β_1	-0.008	0.182	0.180	0.182	0.944	-0.014	0.188	0.186	0.189	0.941	-0.022	0.187	0.185	0.188	0.936
β_2	-0.004	0.169	0.165	0.169	0.939	-0.012	0.182	0.177	0.182	0.940	-0.022	0.182	0.177	0.183	0.930
η_0	-0.010	0.312	0.312	0.313	0.936	-0.020	0.322	0.322	0.323	0.934	-0.033	0.325	0.325	0.326	0.931
η_1	0.006	0.244	0.243	0.245	0.951	0.003	0.247	0.244	0.247	0.950	-0.001	0.246	0.243	0.246	0.949
η_2	0.006	0.214	0.211	0.214	0.945	0.001	0.220	0.216	0.220	0.950	-0.007	0.220	0.216	0.220	0.946
σ_T	0.004	0.097	0.096	0.097	0.948	-0.001	0.108	0.107	0.108	0.943	-0.008	0.108	0.107	0.108	0.944
σ_C	-0.014	0.167	0.165	0.167	0.938	-0.021	0.176	0.174	0.177	0.942	-0.031	0.175	0.172	0.178	0.938
ρ	-0.028	0.206	0.187	0.208	0.956	-0.030	0.210	0.188	0.212	0.956	-0.031	0.213	0.188	0.215	0.954
θ	-0.002	0.031	0.031	0.031	0.949	-0.007	0.061	0.059	0.061	0.942	-0.019	0.099	0.096	0.101	0.944
	Independent censoring model														
β_0	0.207	0.138	0.137	0.249	0.711	0.176	0.163	0.161	0.240	0.850	0.156	0.174	0.170	0.234	0.879
β_1	0.201	0.178	0.174	0.268	0.812	0.180	0.189	0.183	0.261	0.864	0.169	0.191	0.183	0.255	0.867
β_2	0.111	0.181	0.170	0.212	0.905	0.087	0.197	0.184	0.216	0.922	0.074	0.200	0.183	0.214	0.924
η_0	0.622	0.268	0.263	0.678	0.304	0.578	0.295	0.288	0.649	0.481	0.551	0.305	0.295	0.630	0.550
η_1	-0.380	0.243	0.246	0.451	0.654	-0.382	0.240	0.243	0.451	0.642	-0.383	0.238	0.241	0.451	0.630
η_2	-0.293	0.218	0.218	0.365	0.720	-0.304	0.218	0.218	0.374	0.694	-0.309	0.217	0.217	0.378	0.683
σ_T	-0.030	0.090	0.087	0.095	0.922	-0.044	0.101	0.096	0.110	0.904	-0.051	0.103	0.096	0.115	0.897
σ_C	0.257	0.162	0.164	0.303	0.616	0.232	0.177	0.177	0.292	0.704	0.221	0.180	0.175	0.285	0.711
θ	-0.021	0.031	0.031	0.038	0.881	-0.045	0.062	0.059	0.076	0.873	-0.080	0.102	0.095	0.130	0.858

Table 2: Simulation results for 25% censoring and $n = 300$. The true parameters are $(\beta_0, \beta_1, \beta_2) = (3.25, -0.45, 1.2)$, $(\eta_0, \eta_1, \eta_2) = (3.5, 0.6, 1)$, and $(\sigma_T, \sigma_C, \rho) = (1, 1.5, 0.75)$

θ	0					0.5					1.5				
Par.	Bias	ESD	ASE	RMSE	CR	Bias	ESD	ASE	RMSE	CR	Bias	ESD	ASE	RMSE	CR
	Dependent censoring model														
β_0	-0.009	0.240	0.245	0.240	0.947	-0.030	0.317	0.331	0.319	0.942	-0.065	0.378	0.401	0.384	0.938
$\beta 1$	0.005	0.158	0.160	0.158	0.957	0.009	0.160	0.161	0.160	0.954	0.014	0.160	0.162	0.160	0.949
$\beta 2$	-0.003	0.164	0.164	0.164	0.946	-0.013	0.196	0.199	0.197	0.940	-0.029	0.216	0.223	0.218	0.934
η_0	-0.010	0.330	0.339	0.330	0.943	-0.033	0.390	0.409	0.391	0.942	-0.070	0.442	0.470	0.447	0.938
η_1	0.005	0.441	0.425	0.441	0.926	0.001	0.440	0.421	0.439	0.922	-0.004	0.437	0.416	0.437	0.923
η_2	0.001	0.233	0.232	0.233	0.961	-0.008	0.252	0.254	0.252	0.951	-0.021	0.264	0.268	0.265	0.939
σ_T	-0.002	0.120	0.122	0.120	0.946	-0.011	0.152	0.157	0.152	0.952	-0.025	0.170	0.179	0.172	0.949
σ_C	-0.012	0.234	0.237	0.235	0.946	-0.024	0.254	0.261	0.256	0.953	-0.043	0.269	0.279	0.272	0.951
ρ	-0.036	0.225	0.203	0.228	0.928	-0.040	0.230	0.206	0.233	0.923	-0.044	0.235	0.208	0.239	0.924
θ	-0.003	0.037	0.037	0.037	0.950	-0.013	0.082	0.084	0.083	0.951	-0.037	0.156	0.161	0.160	0.951
	Independent censoring model														
β_0	0.167	0.249	0.252	0.300	0.933	0.060	0.340	0.337	0.345	0.947	-0.040	0.409	0.398	0.411	0.927
β_1	-0.209	0.137	0.139	0.250	0.725	-0.185	0.146	0.145	0.235	0.835	-0.164	0.153	0.150	0.224	0.881
β_2	-0.046	0.165	0.160	0.171	0.914	-0.089	0.194	0.185	0.213	0.872	-0.126	0.213	0.200	0.248	0.835
η_0	0.481	0.360	0.361	0.601	0.802	0.354	0.449	0.447	0.572	0.924	0.238	0.520	0.509	0.572	0.946
η_1	0.854	0.380	0.381	0.935	0.371	0.803	0.388	0.386	0.892	0.460	0.760	0.395	0.388	0.857	0.531
η_2	-0.224	0.281	0.282	0.359	0.847	-0.253	0.280	0.281	0.377	0.808	-0.277	0.278	0.279	0.393	0.768
σ_T	-0.071	0.109	0.107	0.130	0.891	-0.105	0.135	0.131	0.171	0.871	-0.135	0.152	0.145	0.203	0.847
σ_C	0.401	0.233	0.250	0.463	0.578	0.333	0.275	0.286	0.432	0.766	0.276	0.305	0.309	0.411	0.832
θ	-0.028	0.036	0.036	0.046	0.882	-0.072	0.082	0.080	0.109	0.863	-0.151	0.158	0.151	0.218	0.841

Table 3: Simulation results based on the proposed method, when the censoring model is misspecified. The true parameters are $(\beta_0, \beta_1, \beta_2) = (2, 1.2, 1.5)$, $(\eta_0, \eta_1, \eta_2) = (2.5, 0.5, 1)$, and $(\sigma_T, \sigma_C, \rho) = (1, 1.5, 0.75)$.

θ	0					0.5					1.5				
Par.	Bias	ESD	ASE	RMSE	CR	Bias	ESD	ASE	RMSE	CR	Bias	ESD	ASE	RMSE	CR
	η_3=1														
β_0	0.000	0.130	0.130	0.130	0.948	-0.007	0.153	0.152	0.153	0.946	-0.019	0.165	0.165	0.166	0.938
β_1	-0.010	0.168	0.168	0.169	0.944	-0.017	0.177	0.177	0.178	0.940	-0.026	0.178	0.178	0.180	0.932
β_2	0.008	0.162	0.158	0.162	0.936	-0.000	0.180	0.174	0.179	0.936	-0.013	0.183	0.177	0.183	0.933
η_0	0.284	0.424	0.422	0.510	0.914	0.273	0.434	0.432	0.513	0.919	0.257	0.438	0.436	0.508	0.926
η_1	0.024	0.280	0.280	0.281	0.935	0.021	0.282	0.282	0.283	0.937	0.016	0.282	0.281	0.283	0.937
η_2	0.087	0.232	0.236	0.248	0.923	0.081	0.241	0.244	0.254	0.935	0.071	0.242	0.245	0.252	0.940
σ_T	0.008	0.096	0.094	0.096	0.946	0.002	0.109	0.106	0.109	0.945	-0.006	0.112	0.109	0.112	0.948
σ_C	0.002	0.204	0.200	0.204	0.933	-0.005	0.213	0.209	0.213	0.934	-0.016	0.215	0.210	0.215	0.932
ρ	-0.027	0.212	0.193	0.213	0.952	-0.028	0.215	0.194	0.217	0.951	-0.031	0.220	0.195	0.222	0.950
θ	-0.000	0.031	0.030	0.031	0.946	-0.005	0.062	0.060	0.062	0.948	-0.018	0.104	0.101	0.106	0.945
	η_3=2														
β_0	0.003	0.125	0.124	0.125	0.950	-0.004	0.150	0.149	0.150	0.947	-0.016	0.165	0.164	0.165	0.940
β_1	-0.010	0.160	0.161	0.160	0.947	-0.016	0.170	0.171	0.170	0.942	-0.026	0.172	0.173	0.174	0.933
β_2	0.010	0.159	0.153	0.159	0.941	0.001	0.179	0.173	0.178	0.938	-0.012	0.184	0.178	0.184	0.936
η_0	0.552	0.549	0.555	0.779	0.863	0.543	0.560	0.563	0.779	0.873	0.526	0.564	0.565	0.771	0.881
η_1	0.037	0.316	0.320	0.318	0.927	0.033	0.319	0.322	0.320	0.928	0.026	0.319	0.320	0.320	0.931
η_2	0.194	0.232	0.251	0.302	0.849	0.187	0.243	0.261	0.306	0.874	0.174	0.246	0.263	0.301	0.886
σ_T	0.007	0.093	0.093	0.093	0.947	0.001	0.107	0.107	0.107	0.950	-0.008	0.111	0.110	0.111	0.946
σ_C	0.052	0.250	0.249	0.255	0.923	0.045	0.259	0.256	0.262	0.925	0.034	0.261	0.258	0.263	0.928
ρ	-0.047	0.221	0.212	0.226	0.936	-0.050	0.226	0.214	0.231	0.933	-0.053	0.231	0.215	0.237	0.932
θ	0.000	0.030	0.030	0.030	0.949	-0.005	0.061	0.061	0.062	0.951	-0.017	0.105	0.103	0.107	0.950

Table 4: Simulation results for the proposed model and the copula model when the data are generated under Scenarios 1–3. Note that the copula model is fitted using the true value of $\rho = 0.75$ and using a misspecified value of $\rho = 0.45$.

Type of copula	Par.	Proposed model			Copula model ($\rho = 0.75$)			Copula model ($\rho = 0.45$)		
		Bias	ESD	RMSE	Bias	ESD	RMSE	Bias	ESD	RMSE
		Scenario 1								
Normal	β	-0.001	0.073	0.073	-0.037	0.129	0.134	-0.186	0.179	0.258
	η	-0.003	0.104	0.104	-0.023	0.102	0.104	0.039	0.211	0.215
	θ	-0.001	0.054	0.054	0.014	0.077	0.078	-0.137	0.105	0.173
	ρ	-0.009	0.100	0.100						
t (df=5)	β	-0.016	0.072	0.073	-0.058	0.133	0.145	-0.223	0.168	0.279
	η	0.012	0.108	0.109	-0.012	0.103	0.104	0.027	0.212	0.214
	θ	-0.025	0.054	0.059	-0.017	0.080	0.081	-0.181	0.111	0.212
	ρ	-0.007	0.102	0.102						
		Scenario 2								
Frank	β	0.016	0.075	0.077	0.016	0.124	0.125	-0.176	0.199	0.266
(τ=0.54)	η	0.009	0.109	0.110	0.041	0.106	0.114	0.043	0.245	0.249
	θ	0.095	0.052	0.108	0.107	0.076	0.131	-0.078	0.121	0.144
	ρ	0.026	0.091	0.094						
Gumbel	β	-0.047	0.071	0.085	-0.120	0.141	0.186	-0.270	0.139	0.303
(τ=0.54)	η	-0.055	0.099	0.113	-0.065	0.111	0.129	0.061	0.149	0.189
	θ	-0.046	0.049	0.067	-0.107	0.081	0.134	-0.249	0.107	0.271
	ρ	0.052	0.081	0.096						
		Scenario 3								
Normal	β	-0.102	0.093	0.138	-0.119	0.158	0.198	-0.249	0.164	0.298
	η	-0.235	0.112	0.262	-0.088	0.120	0.149	-0.021	0.205	0.206
	θ	0.071	0.081	0.107	-0.066	0.093	0.114	-0.174	0.104	0.203
	ρ	0.171	0.040	0.176						

Table 5: Simulation results for the goodness-of-fit test for the proposed model when the data are generated from a bivariate normal and a bivariate t distribution.

Model	Sample size	Test statistics $E(T_{CM})$	Test statistics $E(T^*_{CM,b})$	Rejection rate at 5%	Rejection rate at 10%
Normal	$n = 300$	0.054	0.053	0.053	0.117
	$n = 500$	0.052	0.052	0.056	0.107
t with df = 5	$n = 300$	0.096	0.053	0.316	0.425
	$n = 500$	0.122	0.052	0.488	0.583
t with df = 10	$n = 300$	0.059	0.053	0.095	0.158
	$n = 500$	0.060	0.051	0.095	0.165

Table 6: Parameter estimates for the liver transplant data in a dependent censoring, an independent censoring model, a Weighted Cox model and a copula model. The standard errors (SE) (or the bootstrap standard errors (BSE)) and the p-values are also provided.

	Dependent model			Independent model		
Parameter	Estimate	SE	p-value	Estimate	SE	p-value
Age	-0.165	0.096	0.084	-0.267	0.109	0.014
Gender	0.915	0.895	0.307	0.988	1.318	0.456
BMI	-0.086	0.065	0.181	-0.121	0.085	0.155
UKELD	-0.610	0.214	0.005	-0.678	0.237	0.004
θ	1.764	0.196	0.000	1.680	0.195	0.000
ρ	0.730	0.250	0.004			
	Weighted Cox model			Copula model		
Parameter	Estimate	SE	p-value	Estimate	BSE	p-value
Age	0.012	0.032	0.708	-0.307	0.201	0.126
Gender	-0.990	0.655	0.131	0.234	1.092	0.830
BMI	-0.022	0.049	0.653	-0.265	0.164	0.107
UKELD	0.156	0.043	0.000	-0.859	0.192	0.000
θ				0.940	0.031	0.000

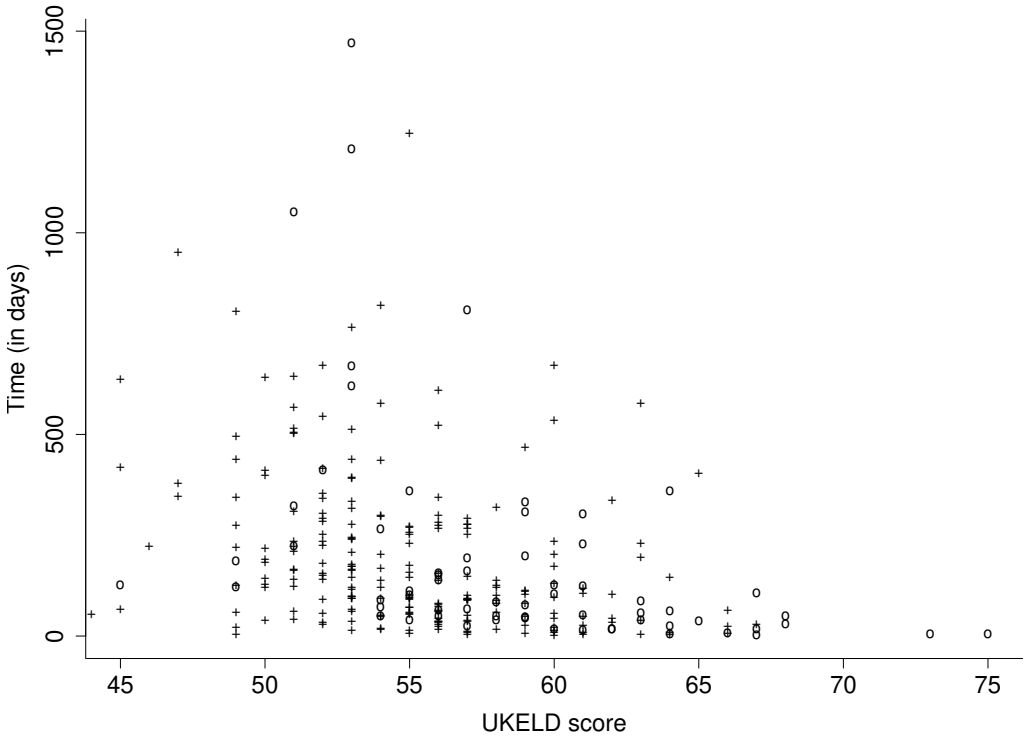

Figure 1: Scatter plot of the UKELD score versus the time to death while waiting for a liver transplant. Uncensored observations are indicated by a circle, censored ones by a plus.

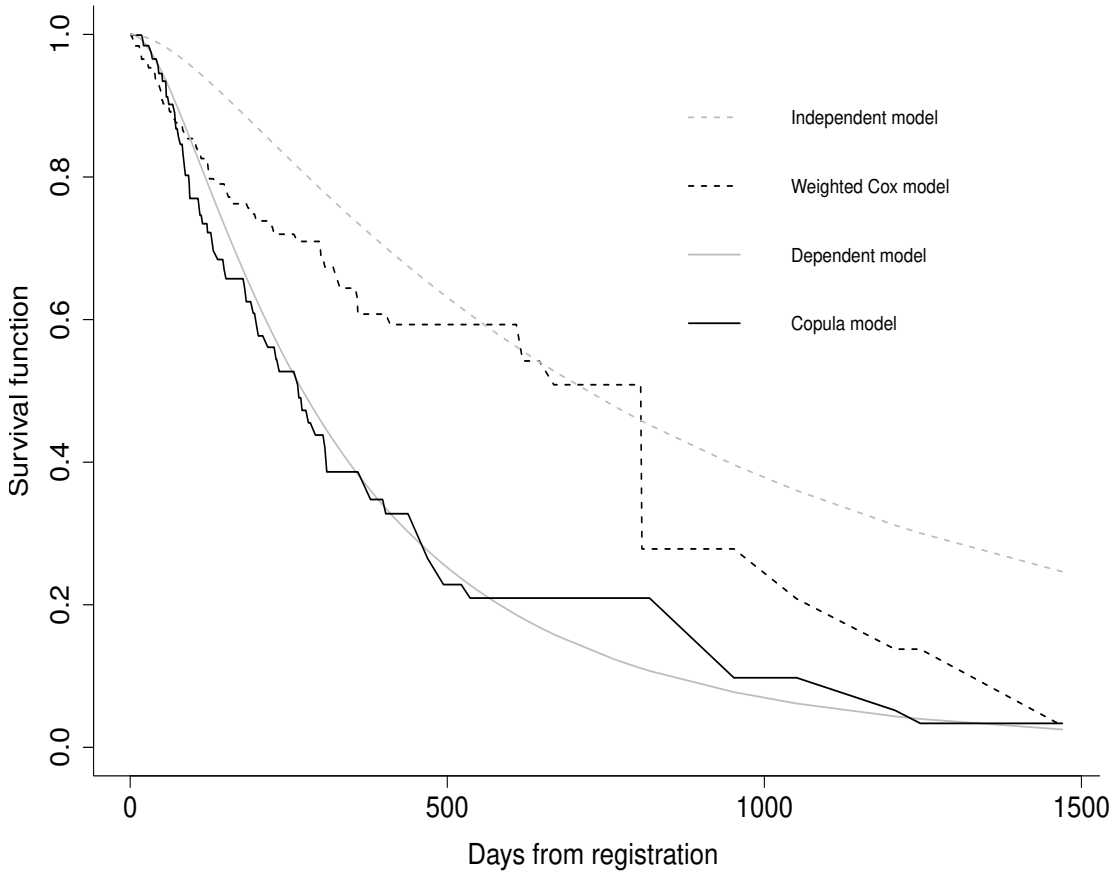

Figure 2: Estimated survival functions of T given X obtained under a dependent and independent censoring model and under a weighted Cox model and a copula model for a 50 year-old female with a UKELD score of 57 and a BMI of 25.