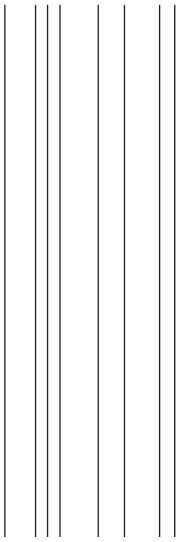


Maastricht University | universiteit hasselt

2015 | Faculty of Sciences

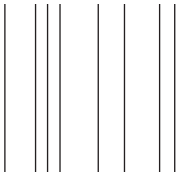
DOCTORAL DISSERTATION

Semi-parametric regression models for zero-inflated left-censored time to event data

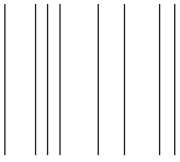
Doctoral dissertation submitted to obtain the degree of
Doctor of Science: Statistics, to be defended by

Yves Grouwels

Promoter: Prof. Dr Roel Braekers



D/2015/2451/15



Dankwoord

Ik ben aan het einde gekomen van zes leerrijke doctoraatsjaren. Tijd om even stil te staan bij iedereen die heeft bijgedragen tot de realisatie van mijn doctoraatsthesis.

In het bijzonder wil ik mijn promotor Prof. dr. Roel Braekers bedanken voor de begeleiding bij mijn onderzoek. De tips, die ik gekregen heb tijdens onze vele meetings, hebben me enorm geholpen.

Ik wil de Universiteit Hasselt bedanken om me de kans te geven om te doctoreren. Zonder de steun van de BOF-beurs BOF09D01 was deze dissertatie niet mogelijk geweest.

I would like to thank all the jury members for their willingness to read my thesis and for their useful comments on my work.

Tijdens de afgelopen jaren hebben uiteraard heel wat mensen mijn pad gekruist. Graag wil ik al mijn collega's bedanken voor de leuke werksfeer. Tijdens de pauzes kon ik mijn hoofd even leegmaken. In het wekelijkse badmintonuurtje kon ik volledig ontspannen. De quizavonden waren geestverruimend, maar vooral heel plezant. Furthermore, I want to thank all my former roommates. It was a pleasure for me to share an office with all of you. Ook wil ik mijn vrienden bedanken. Onze terrasjes en uitstapjes zorgen steeds voor een welgekomen ontspanning.

Mama en papa, ik wil mijn speciale dank uitspreken voor jullie onvoorwaardelijke steun en goede zorgen. Zonder jullie was het niet gelukt om mijn jarenlange studies

tot een goed einde te brengen. Jullie staan altijd voor mij klaar en weten me steeds op te peppen. Merci! Ook de rest van mijn familie en schoonfamilie wil ik bedanken voor alles.

Mijn laatste woorden van dank wil ik richten tot mijn echtgenote Ruth. Jij bent er altijd voor mij geweest en je helpt me steeds om mijn gedachten te verzetten. Ik wil je bedanken voor je liefde, steun, begrip, ... vanuit het diepste van mijn hart!

Yves Grouwels
Diepenbeek, 16 April 2015

Contents

Chapter 1

Introduction

1.1 Introduction and outline of the thesis

In some clinical, environmental or economical studies, researchers are interested in a semi-continuous outcome variable. Hereby the outcome variable attains on the one hand the value zero with a discrete probability mass, while on the other hand it has a continuous distribution for its non-zero response values. For example, in an environmental study where investigators want to know the amount of a certain toxic metal in an aquatic system, it is possible that this metal is not present in the system (Blackwood (1991)). Also, in an economical study where we want to obtain an insight in the amount of R & D which is done in different companies, we note that some companies do not perform any reported R & D. As a last example, we consider a biological study on ethanol-induced sleeping time in genetically selected mice (Markel et al. (1995)). Some mice did not fall asleep because their genetic metabolism was able to break down the alcohol in the blood very fast. In each of these studies, one can identify two groups of study subjects. On the one hand, there are individuals with a strictly positive value for the outcome variable which are called susceptibles, while on the other hand, there are subjects with a zero outcome value. These subjects are called non-susceptible because they do not show any response in the study.

In the literature, we can distinguish between two different ways to analyze a semi-continuous variable with an excess of zero values. A first way was introduced by Tobin (1958). He proposed to use an underlying continuous latent random variable

and to consider the zero outcomes as censored observations of this latent variable at zero. This Tobit model was later generalized by, for example, Cragg (1971), Amemiya (1973) and many others. A survey of the Tobit model and its generalizations can be found in Amemiya (1984). A second way to analyze semi-continuous variables is by a finite mixture model consisting of a degenerate component at zero and a continuous component for the non-zero values. This model was first introduced by Duan et al. (1983) and later extended to other settings by for example Olsen and Schafer (2001). In our work, we consider finite mixture structures for the response variables.

Due to technical limitations, it is in some studies not always possible to fully observe the semi-continuous outcome variable. For example by a fixed detection limit, we only observe an upper bound for the outcome variable in some individuals. We call these observations left-censored. Moulton and Halsey (1995) developed a finite mixture regression model for this type of left-censored semi-continuous data. Hereby they assumed a parametric logistic regression model for the discrete probability of a zero outcome value, and assumed a lognormal distribution for the strict positive values. Due to the left-censoring, we note that it is not possible to fully discriminate between the groups of susceptible and non-susceptible subjects. The uncensored observations are clearly susceptible subjects, but for the censored observations one cannot distinguish between unsusceptible subjects and susceptible subjects with a censored outcome variable. Recently, Yang and Simpson (2010) studied the computational issues in a more general class of parametric zero-inflated left-censored mixture models and extended the model of Moulton and Halsey (1995).

The zero-inflated left-censored mixture models are in a way related to the so called mixture cure models. Cure models are useful in modeling survival data with a cure fraction. For example, in cancer studies, a substantial fraction of the patients may be cured and will never experience relapse. An excellent book about this field of research is Maller and Zhou (1996). Typical data sets have heavy right censoring at the end of the follow-up period. Among the right censored observations, it is impossible to distinguish between cured individuals and the non-cured individuals who are censored due to incomplete follow-up. In the so called mixture cure models, one models separately the survival distribution for the susceptible individuals (who would eventually experience the event if there were no censoring) and the fraction of the cured individuals. Parametric mixture cure models are studied since many decades. Berkson and Gage (1952) used a model consisting of a mixture of the exponential distribution and a constant cure fraction. Farewell (1982, 1986) extended this

parametric modeling approach to Weibull regression for the susceptibles and logistic regression for the cure fraction. Kuk and Chen (1992) introduced semi-parametric mixture cure models. They modeled the survival times of susceptible subjects with proportional hazards regression models, while they used logistic regression models for the cure fraction. Lu (2008) used a joint parametric/nonparametric likelihood technique to estimate the parameters in this model from right-censored data and established the asymptotic properties of the estimators using the modern empirical process theory. We note for completeness that the so called bounded cumulative hazard cure models (see e.g. Yakovlev and Tsodikov (1996), Chen et al. (1999) and Tsodikov et al. (2003)) are an appealing alternative to the mixture cure models. The bounded cumulative hazard models assume that a latent biological process of propagation of latent clonogenic tumor cells (N latent factors) is generating the observed failure. In these models, an individual is at risk of failure if he/she is exposed to at least one of these latent factors. If not, the individual is considered cured. Failure occurs when one or more of these latent factors become activated. The random variable N can have any finite-mean integer-valued distribution (e.g. Poisson, geometric, etc.). In the most popular model, N is assumed to follow a Poisson distribution, with covariates introduced through the parameter of this distribution. The activation times are assumed to be independent and identically distributed. It can be shown that in these models the cumulative hazard function for the entire population of patients is bounded, which explains the name of this class of models. Cooner et al. (2007) generalized the framework of the bounded cumulative hazard models to a more flexible class of cure models under latent activation schemes.

In **Chapter 2**, we propose an extension of the mixture regression models by Moulton and Halsey (1995) and Yang and Simpson (2010) for left-censored semi-continuous data. Instead of a fixed detection limit, we assume a random censoring variable. Furthermore we consider a semi-parametric mixture regression model for this type of left-censored data. Hereby we take, as in the previous models, a parametric regression model to investigate the influence of covariates on the discrete probability of a zero outcome value. For the continuous positive part of the outcome variable, we consider a semi-parametric Cox's regression model instead of a parametric model to study the influence of the covariates. The different parameters in the mixture model are estimated using a likelihood method. Hereby the infinite dimensional baseline cumulative hazard function is estimated by a step function. In order to facilitate the maximum likelihood estimation procedure, we discuss some technical aspects about the optimization algorithm. As results, we show the

identifiability of the model and the consistency of the estimators for the different parameters in the model. Next, we prove the asymptotic normality of the maximum likelihood estimators applying theory about empirical processes (van der Vaart and Wellner (1996)). In order to study the finite sample behavior of the estimators, we set up a simulation study. Different simulation settings are considered. We compute an estimate of the bias and standard deviations of the estimates for all parameters in the zero-inflated semi-parametric regression model. For comparison, we also apply a standard left-censored semi-parametric Cox's regression model without zero-inflation (see Kim et al. (2010)) to the data sets. The model of Kim et al. (2010) is developed for doubly-censored data, but it can also be applied to left-censored data. Finally the model is illustrated on a practical data set of ethanol-induced sleep time in mice (Markel et al. (1995)).

In the simulation study of Chapter 2, we note that maximizing the empirical likelihood for univariate zero-inflated left-censored time to event data is rather time consuming and sometimes unstable. For example, when the probability of a zero response is low, there can be some optimization problems. Kim et al. (2013) became aware of similar problems in the computation of maximum likelihood estimators for the proportional hazards model with doubly-censored data. To overcome these problems, they proposed an approximated likelihood and developed an efficient EM-algorithm to obtain estimates for the different parameters. In **Chapter 3**, we investigate whether we can approximate the likelihood for the semi-parametric Cox's regression model for zero-inflated left-censored time to event data in a similar way. As result, the consistency of the maximum approximated likelihood estimators is proved. We further develop an efficient EM-algorithm to calculate the maximum approximated likelihood estimates. In the M-step of the EM-algorithm one obtains estimates for the logistic parameters and for the effect parameters of the proportional hazards model separately from each other. Moreover, one can use standard functions in the statistical software package **R** to compute these estimates. The estimates of the jump sizes of the baseline cumulative hazard function can be computed from the estimates for the effect parameters. By approximating the likelihood, we get a high dimension reduction, which makes the optimization procedure more stable and faster. Through a simulation study, the accuracies of the maximum approximated likelihood estimates and the maximum likelihood estimates are compared for finite data samples.

Furthermore, the biological study on ethanol-induced sleeping time in mice (Markel et al. (1995)) has a repeated measurement design, since mice are tested at

two different times. This is an example where researchers are interested in bivariate semi-continuous time to event data. Due to technical limitations, observations for both semi-continuous outcome variables may be left-censored. Berk and Lachenbruch (2002) extended the model of Moulton and Halsey (1995) to a multivariate setting with repeated measurements. By introducing a random effect for each subject, they allowed that the different measurements within a subject were not independent. Chu et al. (2005) considered a bivariate submodel of this extended model. They proposed the use of a bivariate Gaussian mixture model to model two measures of viral load with known lower limits of detection and investigated in detail the correlation coefficient between these measures.

In **Chapter 4**, we introduce bivariate parametric and semi-parametric regression models for left-censored observations where the underlying times until an event have a discrete probability of a zero value. We first model the marginal probabilities of a zero response and the marginal distributions of non-zero responses. Afterwards, we impose a dependence structure to model the joint probability of having zero responses in both measurements and the association between two non-zero responses. This kind of marginal approach was also used in the bivariate cure-mixture model of Chatterjee and Shih (2001), which was developed to analyze correlated survival data when there exists a cured proportion in the study. In our regression models, we assume in each measurement a parametric regression model for the marginal probability of a zero response. The non-zero parts of both outcome variables are modeled by parametric or semi-parametric proportional hazards models. For the joint probability of a zero-response in both measurements, we model the cross ratio of a contingency table containing all combinations of zero and non-zero responses for both measurements. The association between two non-zero responses is modeled by parametric families of copulas. Furthermore, we assume independent random censoring variables instead of fixed detection limits.

To estimate the different parameters in the bivariate regression models, we make use of maximum likelihood techniques. However, we note that the maximum likelihood estimators do not have a closed form and numerical optimization becomes more difficult as the number of parameters increases. The special structure of the model suggests that we can consider a two-stage estimation procedure. At the first stage we estimate the parameters in the margins, ignoring the dependence of the two measurements. The second stage involves maximum likelihood of the dependence parameters with the univariate parameters held fixed from the first stage. This

technique was also used by, among others Shih and Louis (1995), Joe (1997) and Chatterjee and Shih (2001). A partitioned form of the asymptotic variance-covariance matrix of the two-stage parametric estimators is deduced, together with a jackknife estimator for this matrix. We study the finite sample behavior of the parametric and semi-parametric estimators through a simulation study and illustrate the model on the practical data example of ethanol-induced sleeping time in mice.

Finally, in **Chapter 5**, general conclusions about the results are presented. In addition we give some indications about possible future research.

1.2 Practical data set of ethanol-induced sleep time in mice

We will illustrate the performance of our regression models and estimation methods on a practical study on ethanol-induced anesthesia (sleep time) in genetically-selected strains of mice. The original data set was created by Markel et al. (1995) and involved four different populations of mice, divided over three generations: two parental inbred strains, their isogenic F_1 population and a genetically-segregating F_2 population derived by crosses of F_1 mice. As primary goal, the populations of mice were used to study the genetic influence of ethanol on sleep time. The selection of the parental inbred strains was done in such a way that one of the parental strains had in general a very long ethanol-induced sleep time, while the other parental strain had a very short sleep time or did not sleep at all.

From Markel et al. (1995), we learn that the assessment of the ethanol-induced sleep time was done in the following way. Every mouse was injected intraperitoneally with a 4.1 g/kg dose of ethanol for a first time at 55-65 days of age (trial 1) and a second time 7-10 days later (trial 2). After the injection, the mouse was placed on its back and was considered anesthetized if it did not turn over more than three times within the first minute. Repeated attempts to observe this behavior were made within 15 min after the injection. Anesthetic recovery was indicated when an individual turned over three times within 1 min after being anesthetized. The sleeping time of a mouse was measured as the time interval between observed anesthesia and the final minute of recovery. We note that in the assessment of sleep time, the recording is left-censored by a fixed detection limit taken at 1 min. For a mouse in which repeated attempts to place it on its back failed, we know that the animal did not fall asleep or had a very short sleep time. In this case, the mouse was immune for the administered ethanol dose.

Next to the sleep time, several covariates were also recorded for each animal. We focus in this doctoral thesis on the covariates sex, weight at each trial and the coat color of the mouse. This variable coat color was dichotomized in the analysis and will be used to study whether an albino mouse reacts differently to alcohol than a non-albino mouse. Markel and Corley (1994) found that the gene coding for albinism (Tyr) had an effect on the ethanol-induced sleep time. They posed that either this gene or a gene closely linked to it, is important for sleep time.

In this doctoral thesis, we only consider the bivariate observations for the segregating F_2 population within this data set. In Table 1.1, we give an overview of the censoring status in both trials for the mice in the study.

		Second	
		Uncensored	Left-censored
First	Uncensored	978	51
	Left-censored	34	8

Table 1.1: Censoring status. The rows correspond to the censoring status at the first measurement, the columns to the censoring status at the second measurement.

Furthermore, we plot in Figure 1 the recorded sleep times for each mouse for the first and second measurement. Hereby we note that the mice, for which one or

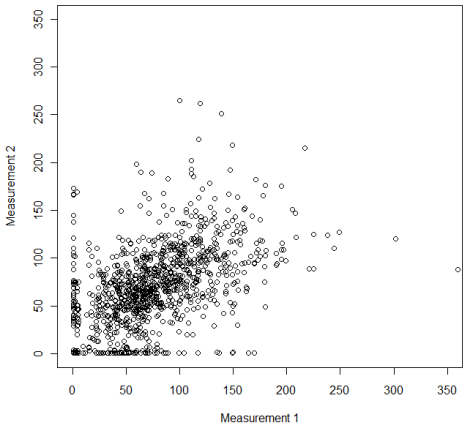


Figure 1.1: Scatter plot of the recorded sleep times for the first and second measurement.

both observations are censored (at the detection limit of 1 min), are given on lines parallel to the x- and y-axis.

Chapter 2

Zero-inflated semi-parametric Cox's regression model

The focus in this chapter is on the semi-parametric modeling of zero-inflated left-censored time to event data. In Section 2.1, we introduce mathematically the semi-parametric Cox's regression model for zero-inflated left-censored data. To estimate the different parameters in our model, we make use of maximum likelihood techniques. Hereby we note that the baseline cumulative hazard function is an infinite dimensional parameter which is estimated by a step function. In order to facilitate the maximum likelihood estimation procedure, we discuss some technical aspects about the optimization algorithm. After introducing some regularity conditions in Section 2.2, we show in Section 2.3 the identifiability of the model and the existence of the maximum likelihood estimators. We prove in Section 2.4 the consistency of the maximum likelihood estimators under the regularity conditions. The asymptotic normality of the maximum likelihood estimators is proved in Section 2.5. Next we present a simulation study in Section 2.6 to investigate the finite sample properties of these estimators. Afterwards, in Section 2.7, our model is illustrated on a practical data set of ethanol-induced sleep time in mice. Finally, in Section 2.8, we give some conclusions about our results.

2.1 Methodology

In this section, we introduce a semi-parametric Cox's regression model for zero-inflated time to event data. We denote by Y a semi-continuous outcome variable which attains the value zero with a discrete probability and has non-zero positive values with a continuous distribution. Furthermore we assume that there exist two vectors of covariates X and Z which may have components in common. We consider that the conditional distribution of the outcome Y is given by

$$F(y|x,z) = P(Y \leq y|X = x, Z = z) = \pi(x) + (1 - \pi(x))F_{Y>0}(y|z),$$

where $F_{Y>0}(y|z) = P(Y \leq y|Y > 0, Z = z)$ is a continuous conditional distribution for the non-zero values of the outcome Y and $\pi(x) = P(Y = 0|X = x)$ is the conditional probability of a zero outcome value. We assume a parametric model for the conditional probability $\pi(x)$ and denote it by $\pi(\gamma, x)$. The parameters of this model are expressed by a finite dimensional vector γ. For the conditional distribution $F_{Y>0}(y|z)$ of the non-zero outcome values, we consider a Cox's regression model (Cox (1972)). Hereto, we assume that the conditional hazard function has the following form

$$\lambda_{Y>0}(t|z) = \frac{f_{Y>0}(t|z)}{\bar{F}_{Y>0}(t|z)} = \lambda(t)g(\beta, z), \tag{2.1}$$

where $\bar{F}_{Y>0}(t|z) = 1 - F_{Y>0}(t|z)$ (resp. $f_{Y>0}(t|z) = \frac{d}{dt}F_{Y>0}(t|z)$) is the conditional survival (resp. density) function. The function $\lambda(t)$ in (2.1) is an unknown baseline hazard function and $g(\beta, z) > 0$ is a known parametric model which depends on a finite dimensional vector β.

In some studies, it is impossible to fully observe the outcome variable Y and we only see an upper bound. Therefore we assume that there exists a random variable C such that we only observe $T = \max(Y, C)$ and $\delta = I(Y \geq C)$. This type of data is called left-censored data. Conditionally on the covariate vectors X and Z, we assume that Y and C are independent. We denote the conditional distribution of C, given X and Z, by $F_{C|X,Z}(c|x,z) = P(C \leq c|X = x, Z = z)$.

To estimate the parameters γ and β and the baseline hazard function $\lambda(t)$ in this model, we use maximum likelihood techniques. Hereto we assume that $(T_1, \delta_1, X_1, Z_1)$, $\ldots$, $(T_n, \delta_n, X_n, Z_n)$ is an i.i.d. sample of the observed variables (T, δ, X, Z). The

contribution of an individual to the likelihood function is given by

$$\begin{aligned} p(t,\delta,x,z) &= \{f(t|x,z)F_{C|X,Z}(t|x,z)\}^{\delta}\{F(t|x,z)f_{C|X,Z}(t|x,z)\}^{1-\delta} \\ &= \left\{(1-\pi(\gamma,x))\,\lambda(t)g(\beta,z)\exp\left(-g(\beta,z)\Lambda(t)\right)F_{C|X,Z}(t|x,z)\right\}^{\delta} \\ &\quad \times\{[\pi(\gamma,x)+(1-\pi(\gamma,x)) \\ &\quad (1-\exp(-g(\beta,z)\Lambda(t)))]\,f_{C|X,Z}(t|x,z)\}^{1-\delta}, \end{aligned}$$

where $f(t|x,z)$ (resp. $f_{C|X,Z}(c|x,z)$) is the density function of the conditional outcome variable Y (resp. censoring variable C) and $\Lambda(t)=\int_0^t \lambda(s)ds$ is the baseline cumulative hazard function. Since the conditional distribution of the censoring variable does not depend on the unknown parameters, we get that the likelihood function is given by

$$\begin{aligned} L(\gamma,\beta,\lambda) &= \prod_{i=1}^{n}\{(1-\pi(\gamma,X_i))\,\lambda(T_i)g(\beta,Z_i)\exp\left(-g(\beta,Z_i)\Lambda(T_i)\right)\}^{\delta_i} \\ &\quad \{\pi(\gamma,X_i)+(1-\pi(\gamma,X_i))\,(1-\exp(-g(\beta,Z_i)\Lambda(T_i)))\}^{1-\delta_i}. \end{aligned}$$

However, it is impossible to maximize this expression over the space of all baseline cumulative hazard functions. We will replace the baseline hazard function $\lambda(t)$ by a difference in the cumulative hazard function $\Lambda(t)$ which we denote by $\Lambda\{t\}=\Lambda(t)-\Lambda(t^-)$. This leads to the following likelihood function:

$$\begin{aligned} L^e(\gamma,\beta,\Lambda) &= \prod_{i=1}^{n}\{(1-\pi(\gamma,X_i))\Lambda\{T_i\}g(\beta,Z_i)\exp\left(-g(\beta,Z_i)\Lambda(T_i)\right)\}^{\delta_i} \\ &\quad \{\pi(\gamma,X_i)+(1-\pi(\gamma,X_i))\,(1-\exp(-g(\beta,Z_i)\Lambda(T_i)))\}^{1-\delta_i}. \end{aligned}$$

In this way we note that the maximum likelihood estimator for the baseline cumulative hazard will be a step function. We consider the following non-parametric step function for this baseline cumulative hazard function:

$$\Lambda_n(t)=\sum_{k=1}^{q_n}\lambda_k \mathrm{I}(u_k\le t),$$

where $0<u_1<\ldots<u_{q_n}$ are the unique uncensored observations and $\lambda_1,\ldots,\lambda_{q_n}$ are the corresponding step sizes in these time points. For a given sample, we obtain maximum likelihood estimates for the different parameters γ, β and $\lambda_1,\ldots,\lambda_{q_n}$ by maximizing the likelihood function $L^e(\gamma,\beta,\Lambda_n)$. We denote these estimates by $\hat{\theta}=(\hat{\gamma},\hat{\beta},\hat{\Lambda}_n)$.

Some remarks:

1. In the maximization process of the likelihood function, we note that the score equations for the different step sizes in the baseline cumulative hazard function

$$\frac{\partial}{\partial \lambda_k} \log L^e(\gamma, \beta, \Lambda_n) = 0, \qquad k = 1, \ldots, q_n,$$

give the following expressions:

$$\hat{\lambda}_k = \frac{1}{\xi_n(u_k; \hat{\theta})}, \qquad k = 1, \ldots, q_n,$$

where

$$\begin{aligned} \xi_n(u;\theta) &= \sum_{i=1}^{n} g(\beta, Z_i)\{I(T_i \geq u, \delta_i = 1) + a_i(\theta) I(T_i \geq u, \delta_i = 0)\}, \\ a_i(\theta) &= -\frac{(1 - \pi(\gamma, X_i)) \exp\{-g(\beta, Z_i)\Lambda(T_i)\}}{\pi(\gamma, X_i) + (1 - \pi(\gamma, X_i))(1 - \exp\{-g(\beta, Z_i)\Lambda(T_i)\})}, \end{aligned}$$

with $\theta = (\gamma, \beta, \Lambda)$.

Hence, we see that we get a closed form solution for the step sizes of the non-parametric baseline cumulative hazard function in all uncensored observations which are larger than the largest censored observation. Suppose that all observations larger or equal than u_l are uncensored. Then

$$\hat{\lambda}_l = \frac{1}{\sum_{i=1}^{n} g(\hat{\beta}, Z_i) I(T_i \geq u_l)}.$$

Incorporating this into the likelihood function facilitates the maximum likelihood optimization procedure. The number of parameters reduces significantly, which makes the procedure much faster.

2. In the most extreme case, when all censored observations are smaller than the smallest uncensored observation, the likelihood looks as follows:

$$\begin{aligned} L^e(\gamma, \beta, \Lambda_n) &= \prod_{i=1}^{n} \{1 - \pi(\gamma, X_i)\}^{\delta_i} \{\pi(\gamma, X_i)\}^{1-\delta_i} \times \\ &\quad \prod_{i=1}^{n} \left\{ \left[\sum_{k=1}^{q_n} \lambda_k \mathrm{I}(u_k = T_i) \right] g(\beta, Z_i) \right. \\ &\quad \left. \exp\left(-g(\beta, Z_i) \left[\sum_{k=1}^{q_n} \lambda_k \mathrm{I}(u_k \leq T_i) \right] \right) \right\}^{\delta_i} \\ &= L^1(\gamma)\ L^2(\beta, \Lambda_n). \end{aligned}$$

We see that the first part of this likelihood is only concerned with the parametric model of the zero-probability, while the second part is the likelihood of a Cox's regression model based on the subgroup of the uncensored observations. Fitting the zero-inflated Cox's regression model simplifies in this case to fitting a logistic regression model on the censoring indicator variables and fitting a Cox's regression model on the subgroup of the uncensored observations. Hereto, one can use respectively the **glm**- and **coxph**-function in **R**.
This occurs for example in studies with a fixed detection limit, such as the biological study on ethanol-induced sleep time in mice (Markel et al. (1995)), which will be analyzed in Section 2.7.

2.2 Regularity conditions

We give in this section an overview of the assumptions that we impose for the following existence, identifiability, consistency and asymptotic normality results in the Cox's regression on zero-inflated left-censored data.
Hereby we first consider the support of the different random variables in this model. For the semi-continuous outcome variable Y, the support is given by the set $\{0\} \cup [\rho_Y, \tau_Y]$ where $\rho_Y = \sup\{y|F_{Y>0}(y) = 0\}$ and $\tau_Y = \inf\{y|F_{Y>0}(y) = 1\}$ are the lower- (resp. upper-) end of the support of the continuous non-zero part of the outcome Y. For the censoring variable C, we denote the support by the interval $[\rho_C, \tau_C]$. This leads to a support $[\rho_T, \tau_T]$ for the observed time T with $\rho_T = \max(0, \rho_C) = \rho_C$ and $\tau_T = \max(\tau_Y, \tau_C)$. Also, for the observed time of the uncensored observations, we see that the support is given by $[\max(\rho_Y, \rho_C), \tau_Y]$ while for the censored observations, this is $[\rho_C, \tau_C]$.
Assumptions:

A_1: The components in the covariate sets X and Z are bounded ($|X_j| < M_X$, $|Z_j| < M_Z$).

A_2: The parameter spaces of the finite dimensional parameters γ and β are compact and will be denoted by Θ_γ and Θ_β.

A_3: The function $\pi(\gamma, x)$ is a continuous function of γ and x with $0 < \pi(\gamma, x) < 1$ for $(\gamma, x) \in \Theta_\gamma \times [-M_X, M_X]^{d_X}$. We also assume that this function is identifiable (i.e. $\pi(\gamma, x) = \pi(\gamma^*, x) \Rightarrow \gamma = \gamma^*$).

A_4: The function $g(\beta, z) > 0$ is a continuous function of β and z, $((\beta, z) \in \Theta_\beta \times [-M_Z, M_Z]^{d_Z})$, which is identifiable (i.e. $g(\beta, z) = g(\beta^*, z) \Rightarrow \beta = \beta^*$).

A_5: There exist ρ and τ with $\rho_T < \rho < \tau < \tau_T$ such that:

a. $0 < \inf_{t\in[\rho,\tau]} \lambda_0(t) \leq \sup_{t\in[\rho,\tau]} \lambda_0(t) < \infty$, where λ_0 is the true baseline hazard function.

b. The conditional distribution function of C given X and Z, denoted with $F_{C|X,Z}$, has bounded derivatives on $[\rho, \tau]$, except at finitely many points. Furthermore, $F_{C|X,Z}(\rho) > 0$.

We will consider ρ and τ as the beginning and endpoint of any interval on which all uncensored observations are observed. ρ and τ can be arbitrarily close to resp. ρ_Y and τ_Y.

2.3 Identifiability and existence of the maximum likelihood estimators

In this section, we first show that the Cox's regression model for zero-inflated left-censored data is identifiable. Hereto we prove that the finite dimensional parameters γ and β, and the infinite dimensional baseline cumulative hazard function Λ are uniquely determined from the observed data. Next we show that the maximum likelihood estimators for these parameters exist.

Theorem 1. *Under assumptions $A_1 - A_5$, we get that the parameters (γ, β, Λ) are identifiable.*

Proof. Let us assume that we have two sets of parameters (γ, β, Λ) and $(\gamma^*, \beta^*, \Lambda^*)$ such that the observed quantities are the same,

$$p(t, \delta, x, z) = p^*(t, \delta, x, z).$$

This is equivalent to

$$(1 - \pi(\gamma, x)) \exp(-\Lambda(t)g(\beta, z)) = (1 - \pi(\gamma^*, x)) \exp(-\Lambda^*(t)g(\beta^*, z)).$$

Since this holds for all time points t in the support, we get that there exists a positive function $c(\tilde{x})$ such that

$$\frac{1 - \pi(\gamma, x)}{1 - \pi(\gamma^*, x)} = \frac{\exp(-\Lambda^*(t)g(\beta^*, z))}{\exp(-\Lambda(t)g(\beta, z))} = c(\tilde{x}),$$

where $\tilde{x}$ represents the set of covariates which is common for x and z. Hence we get that

$$\begin{aligned} \Lambda^*(t)g(\beta^*, z) &= -\log(c(\tilde{x})) + \Lambda(t)g(\beta, z), \\ 1 - \pi(\gamma^*, x) &= \frac{1 - \pi(\gamma, x)}{c(\tilde{x})}. \end{aligned}$$

For the Cox's regression model, we note that any multiplication of the function $g(\beta, z)$ by a constant c^* can be absorbed in the baseline cumulative hazard function. Therefore we can always find a value z_0 within the covariate space such that $g(\beta, z_0) = 1$ for all values of β. This leads to

$$\Lambda^*(t) = -\log(c(\tilde{x}_0)) + \Lambda(t).$$

By solving the previous relationships for $g(\beta^*, z)$, we get that

$$g(\beta^*, z) = \frac{-\log(c(\tilde{x})) + g(\beta, z)\Lambda(t)}{-\log(c(\tilde{x}_0)) + \Lambda(t)}.$$

In this expression, we need to show that the right-hand side does not depend on the time t. By using a first order Taylor expansion, we get that

$$g(\beta^*, z) = g(\beta, z) + \frac{-\log(c(\tilde{x}_0))(1 - g(\beta, z))}{-\log(c(\tilde{x}_0)) + \Lambda(t)} - \frac{c(\tilde{x}) - c(\tilde{x}_0)}{c^0(-\log(c(\tilde{x}_0)) + \Lambda(t))},$$

where c^0 is between $c(\tilde{x})$ and $c(\tilde{x}_0)$. If we want that the right-hand side of this equation does not depend on t, we need that $c(\tilde{x}) - c(\tilde{x}_0) = 0$ for all $\tilde{x}$ and $c(\tilde{x}_0) = 1$. This gives that $g(\beta^*, z) = g(\beta, z)$, $\Lambda^*(t) = \Lambda(t)$ and $\pi(\gamma^*, x) = \pi(\gamma, x)$. Under assumption A_3 and A_4, we get that this model is also identifiable. If the covariate sets x and z have no covariates in common, we get that the function $c(\tilde{x})$ is a constant and we can find in a similar way that it is equal to 1. □

Next, we prove the existence of the maximum likelihood estimators on a compact set.

Theorem 2. *Under the assumptions $A_1 - A_5$, there exists a vector $(\hat{\gamma}, \hat{\beta}, \hat{\Lambda}_n) = (\hat{\gamma}, \hat{\beta}, \hat{\lambda}_1, \ldots, \hat{\lambda}_{q_n})$ which maximizes the likelihood function $L^e(\gamma, \beta, \Lambda_n) = L^e(\gamma, \beta, \lambda_1, \ldots, \lambda_{q_n})$.*

Proof. Since $g(\beta, z)$ is a continuous function over the compact sets of β and z, we have that this function is bounded from below and above by positive constants, K_1 and K_2. We see that

$$0 \leq L^e(\gamma, \beta, \lambda_1, \lambda_2, \ldots, \lambda_{q_n}) \leq \prod_{i=1}^{n} \{\Lambda\{T_i\}K_2 \exp[-\Lambda(T_i)K_1]\}^{\delta_i}.$$

Now, let us consider the set

$$\Delta_M = \{(\lambda_1, \lambda_2, \ldots, \lambda_{q_n}) | \lambda_1 + \lambda_2 + \ldots + \lambda_{q_n} \leq M\} \subset \mathbb{R}_+^{q_n},$$

with $0 < M < +\infty$.
Since $L^e(\gamma, \beta, \lambda_1, \lambda_2, \ldots, \lambda_{q_n})$ is continuous in $(\gamma, \beta, \lambda_1, \lambda_2, \ldots, \lambda_{q_n})$, it will have a maximum on the compact subspace $\Theta_\gamma \times \Theta_\beta \times \Delta_M$, for any given value M. Denote by L_M the maximum value of $L^e(\gamma, \beta, \lambda_1, \lambda_2, \ldots, \lambda_{q_n})$ over $\Theta_\gamma \times \Theta_\beta \times \Delta_M$. However, since $Me^{-MK_1} \to 0$ as $M \to +\infty$, there exists an M_0 such that the maximum value L_M of $L^e(\gamma, \beta, \lambda_1, \lambda_2, \ldots, \lambda_{q_n})$ over $\Theta_\gamma \times \Theta_\beta \times \Delta_M$ does not get any larger than L_{M_0}. Therefore we can restrict the parameter space of $(\gamma, \beta, \lambda_1, \lambda_2, \ldots, \lambda_{q_n})$ to $\Theta_\gamma \times \Theta_\beta \times \Delta_{M_0}$. By the continuity of $L^e(\gamma, \beta, \lambda_1, \lambda_2, \ldots, \lambda_{q_n})$ on this space, we get the existence of the maximum likelihood estimator □

2.4 Consistency

In this section, we prove the consistency of the maximum likelihood estimators under some regularity conditions. We follow the ideas of Kim et al. (2010).
Let $(\gamma_0, \beta_0, \Lambda_0)$ be the true values of the parameters.

Theorem 3. *Under assumptions $A_1 - A_5$, the maximum likelihood estimators $(\hat{\gamma}, \hat{\beta}, \hat{\Lambda}_n)$ are consistent. This means that*

$$|\hat{\gamma} - \gamma_0| \to 0, \ |\hat{\beta} - \beta_0| \to 0 \ \textit{and} \ \sup_{t \in [\rho, \tau]} |\hat{\Lambda}_n(t) - \Lambda_0(t)| \to 0,$$

with probability 1.

Proof. The proof of the consistency is divided into several steps.

Before giving the full proof, we first give a short sketch. In the **first step**, we show that the maximum likelihood estimators for γ, β and Λ are bounded almost surely. In Section 2.1, we found an expression for the estimators of the step sizes of the nonparametric baseline cumulative hazard function:

$$\hat{\lambda}_r = \frac{1}{n\xi_n(u_r; \hat{\theta})}.$$

In **step 2**, we prove that there exists $M_\xi > 0$ such that $\inf_{t \in [\rho, \tau]} \xi_n(t; \hat{\theta}) > M_\xi$ for all sufficiently large n, with probability 1. Hereto we assume that such an M_ξ does not exist and derive a contradiction with the boundedness of $\hat{\Lambda}_n(\tau)$. In **step 3**, we define

an estimator $\tilde{\Lambda}_n(.)$. We show that this estimator converges to $\Lambda_0(.)$ uniformly on $[\rho, \tau]$. Next, we suppose that $\hat{\gamma}, \hat{\beta}$ and $\hat{\Lambda}_n$ converge to γ^+, β^+ and Λ^+ uniformly. In **step 4**, we show that

$$\frac{d\hat{\Lambda}_n}{d\tilde{\Lambda}_n}(t) \to \frac{d\Lambda^+}{d\Lambda_0}(t)$$

uniformly on $[\rho, \tau]$, with probability 1. In this part of the proof, we use the result from step 2. In **step 5**, we use Helly's selection theorem. We can find a convergent subsequence of $\hat{\theta} = (\hat{\gamma}, \hat{\beta}, \hat{\Lambda}_n)$, with convergence point $\theta^+ = (\gamma^+, \beta^+, \Lambda^+)$. The result from step 4 implies that

$$0 \leq \frac{1}{n} l^e(\hat{\gamma}, \hat{\beta}, \hat{\Lambda}_n) - \frac{1}{n} l^e(\gamma_0, \beta_0, \tilde{\Lambda}_n) \to KL(\theta^+, \theta_0),$$

where $KL(\theta^+, \theta_0)$ is the negative of the Kullback-Leibler divergence. We show that $KL(\theta^+, \theta_0) = 0$. Using the model identifiability, we get that that $\theta^+ = \theta_0$.

After this brief sketch, we describe every step of the proof in more detail.

Step 1: Since Θ_γ and Θ_β are compact sets, the MLE's of γ and β are bounded by some constants M_γ and M_β. We first prove that there exists a constant M_Λ such that $\hat{\Lambda}_n(\tau) \leq M_\Lambda$ for all sufficiently large n with probability 1. That is, the maximum likelihood estimator for Λ is bounded with probability 1. We have that, for any γ, β and Λ,

$$\begin{aligned}
&\frac{1}{n}\left(l^e(\hat{\gamma}, \hat{\beta}, \hat{\Lambda}_n) - l^e(\gamma, \beta, \Lambda)\right) = \\
&\frac{1}{n}\sum_{i=1}^{n} \delta_i \Bigg\{ \log \frac{1 - \pi(\hat{\gamma}, X_i)}{1 - \pi(\gamma, X_i)} + \log \frac{\hat{\Lambda}_n\{T_i\}}{\Lambda\{T_i\}} + \log \frac{g(\hat{\beta}, Z_i)}{g(\beta, Z_i)} \\
&\qquad\qquad -g(\hat{\beta}, Z_i)\hat{\Lambda}_n(T_i) + g(\beta, Z_i)\Lambda(T_i) \Bigg\} \\
&+\frac{1}{n}\sum_{i=1}^{n}(1 - \delta_i) \log \frac{\pi(\hat{\gamma}, X_i) + (1 - \pi(\hat{\gamma}, X_i))(1 - \exp[-g(\hat{\beta}, Z_i)\hat{\Lambda}_n(T_i)])}{\pi(\gamma, X_i) + (1 - \pi(\gamma, X_i))(1 - \exp[-g(\beta, Z_i)\Lambda(T_i)])}.
\end{aligned}$$

Since $g(\beta, z)$ is a continuous function over the compact sets of β and z, this function is bounded from below and above by positive constants K_1 and K_2. So there exists a positive constant K such that

$$\frac{1}{n}\sum_{i=1}^{n} \delta_i \log \frac{g(\hat{\beta}, Z_i)}{g(\beta, Z_i)} \leq K \frac{1}{n}\sum_{i=1}^{n} \delta_i \to KE(\delta),$$

as $n \to +\infty$, by the Glivenko-Cantelli theorem. Since also $\pi(\gamma, x)$ is a continuous function over the compact sets of γ and x, we get in a similar way that there exist constants K' and K'' such that

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\delta_i \log\frac{1-\pi(\hat{\gamma},X_i)}{1-\pi(\gamma,X_i)} \leq K'\frac{1}{n}\sum_{i=1}^{n}\delta_i \to K'E(\delta),\\
&\frac{1}{n}\sum_{i=1}^{n}(1-\delta_i)\log\frac{\pi(\hat{\gamma},X_i)+(1-\pi(\hat{\gamma},X_i))(1-\exp[-g(\hat{\beta},Z_i)\hat{\Lambda}_n(T_i)])}{\pi(\gamma,X_i)+(1-\pi(\gamma,X_i))(1-\exp[-g(\beta,Z_i)\Lambda(T_i)])}\\
&\leq K''\frac{1}{n}\sum_{i=1}^{n}(1-\delta_i) \to K''E(1-\delta).
\end{aligned}$$

If we take for $\Lambda(t)$, $\bar{\Lambda}_n(t) = \sum_{u_k \leq t}\frac{1}{n}$, we have that

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n}\delta_i \log\frac{\hat{\Lambda}_n\{T_i\}}{\bar{\Lambda}_n\{T_i\}} &= \frac{1}{n}\sum_{k=1}^{q_n}\log(n\hat{\lambda}_k)\\
&\leq \frac{q_n}{n}\log\left[\frac{1}{q_n}\sum_{k=1}^{q_n}n\hat{\lambda}_k\right] = \frac{q_n}{n}\log\hat{\Lambda}_n(\tau) + \frac{q_n}{n}\log\left(\frac{n}{q_n}\right)\\
&\leq \log\hat{\Lambda}_n(\tau) + O(1),
\end{aligned}$$

by an application of Jensen's inequality. Furthermore we note that

$$\frac{1}{n}\sum_{i=1}^{n}\delta_i g(\beta,Z_i)\bar{\Lambda}_n(T_i) \leq \frac{1}{n}\sum_{i=1}^{n}\delta_i g(\beta,Z_i) \leq \frac{1}{n}\sum_{i=1}^{n}g(\beta,Z_i) \to E[g(\beta,Z)],$$

for all β. We also find that

$$\begin{aligned}
-\frac{1}{n}\sum_{i=1}^{n}\delta_i g(\hat{\beta},Z_i)\hat{\Lambda}_n(T_i) &\leq -K_1\frac{1}{n}\sum_{k=1}^{q_n}\hat{\lambda}_k\sum_{i=1}^{n}\delta_i I(T_i \geq u_k)\\
&\leq K_1\hat{\Lambda}_n(\tau)\sup_{u\in[\rho,\tau]}\left|\frac{1}{n}\sum_{i=1}^{n}\delta_i I(T_i \geq u) - E[\delta I(T \geq u)]\right|\\
&\quad -K_1\min_{u\in[\rho,\tau]}E[\delta I(T \geq u)]\hat{\Lambda}_n(\tau).
\end{aligned}$$

By Glivenko-Cantelli, the first term in this last expression converges to 0 when n increases. We see that for sufficiently large n,

$$\frac{1}{n}\left(l^e(\hat{\gamma},\hat{\beta},\hat{\Lambda}_n) - l^e(\gamma,\beta,\bar{\Lambda}_n)\right) \leq \log\hat{\Lambda}_n(\tau) - K_1E[\delta I(T \geq \tau)]\hat{\Lambda}_n(\tau) + O(1).$$

Hence, if $\hat{\Lambda}_n(\tau) \to +\infty$, the right hand side of the inequality diverges to $-\infty$. But this would contradict with $\frac{1}{n}\left(l^e(\hat{\gamma},\hat{\beta},\hat{\Lambda}_n) - l^e(\gamma,\beta,\bar{\Lambda}_n)\right) \geq 0$. Therefore, this means that $\hat{\Lambda}_n(\tau)$ should be bounded for sufficiently large n, with probability 1.

Step 2: Let $\theta = (\gamma, \beta, \Lambda)$ and $\hat{\theta} = (\hat{\gamma}, \hat{\beta}, \hat{\Lambda}_n)$. We already mentioned that

$$\hat{\lambda}_r = \frac{1}{n\xi_n(u_r; \hat{\theta})},$$

where

$$\xi_n(t;\theta) = \frac{1}{n}\sum_{i=1}^{n} g(\beta, Z_i)\{I(T_i \geq t, \delta_i = 1) + a_i(\theta) I(T_i \geq t, \delta_i = 0)\},$$

with

$$a_i(\theta) = -\frac{(1-\pi(\gamma, X_i))\exp\{-g(\beta, Z_i)\Lambda(T_i)\}}{\pi(\gamma, X_i) + (1-\pi(\gamma, X_i))(1-\exp\{-g(\beta, Z_i)\Lambda(T_i)\})}.$$

We show that there exists $M_\xi > 0$ such that $\inf_{t\in[\rho,\tau]} \xi_n(t;\hat{\theta}) > M_\xi$ for all sufficiently large n, with probability 1.

Suppose that such an M_ξ does not exist. Let ϵ be fixed, whose value will be specified later. Then, we can find a sufficiently large n and $t^* \in [\rho, \tau]$ such that $\xi_n(t^*; \hat{\theta}) < \epsilon$. By assumption A_5, there exists $\kappa > 0$ such that T has a bounded density on either $(t-\kappa, t]$ or $(t, t+\kappa]$ for any $t \in [\rho, \tau]$. We will show that when T has a bounded density on $(t^* - \kappa, t^*]$, then $\hat{\Lambda}_n(t^*) - \hat{\Lambda}_n(t^* - \kappa) > M_\Lambda$ as $n \to \infty$, which is a contradiction (with step 1). If T has a bounded density on $(t^*, t^* + \kappa]$, a similar contradiction can be derived. The proof for this case is omitted.

From assumption A_3, there exists a constant $M_a > 0$ such that

$$\sup_i |a_i(\hat{\theta})| \leq \sup_i \frac{1-\pi(\gamma, X_i)}{\pi(\gamma, X_i)} \leq M_a$$

with probability 1.

Let $\nu = \sup_{z, |\beta|<M_\beta} g(\beta, z)(1 + M_a)$. For any $t \in (t^* - \kappa, t^*]$, we have

$$\begin{aligned} \xi_n(t;\hat{\theta}) &\leq |\xi_n(t;\hat{\theta}) - \xi_n(t^*;\hat{\theta})| + \xi_n(t^*;\hat{\theta}) \\ &\leq \frac{\nu}{n}\sum_{i=1}^{n} I(t \leq T_i < t^*) + \epsilon. \end{aligned}$$

Since T has a bounded density on $(t^* - \kappa, t^*]$, the Glivenko-Cantelli lemma implies that there exists a constant $\phi > 0$ such that

$$\frac{1}{n}\sum_{i=1}^{n} I(t < T_i \leq t^*) \leq \phi(t^* - t)$$

for all sufficiently large n.

Hence, we have that

$$\begin{aligned}
\hat{\Lambda}_n(t^*) - \hat{\Lambda}_n(t^* - \kappa) &= \frac{1}{n} \sum_{t^* - \kappa < u_r \leq t^*} \frac{1}{\xi_n(u_r, \hat{\theta})} \geq \frac{1}{n} \sum_{t^* - \kappa < u_r \leq t^*} \frac{1}{\nu\phi(t^* - u_r) + \epsilon} \\
&\to \int_{X,Z} \int_{t^*-\kappa}^{t^*} \frac{(1 - \pi(\gamma_0, x)) f_{Y>0}(t|z) F_{C|X,Z}(t|x, z)}{\nu\phi(t^* - t) + \epsilon} dt P(dxdz) \\
&\geq \frac{\zeta}{\nu\phi} \left[- \int_{t^*-\kappa}^{t^*} \frac{1}{\nu\phi(t^* - t) + \epsilon} d(\nu\phi(t^* - t) + \epsilon) \right] \\
&= \frac{\zeta}{\nu\phi} [\log(\nu\phi\kappa + \epsilon) - \log \epsilon] = \frac{\zeta}{\nu\phi} \log \left[\frac{\nu\phi\kappa}{\epsilon} + 1 \right], \qquad (2.2)
\end{aligned}$$

with

$$\zeta = \inf_{t \in [\rho, \tau]} E_{x,z} \left\{ (1 - \pi(\gamma_0, x)) f_{Y>0}(t|z) F_{C|X,Z}(t|x, z) \right\}.$$

In the expression for ζ, $f_{Y>0}(t|z)$ is the density of the non-zero part of the response, and $F_{C|X,Z}(t|x, z) = P(C \leq t|X = x, Z = z)$. $E_{x,z}$ means that the expectation is taken over X and Z. Due assumption A_5, we know that $\zeta > 0$. Finally, choose ϵ sufficiently small, so that (2.2) is larger than M_Λ, which is a contradiction with the boundedness of $\hat{\Lambda}_n(\tau)$.

Step 3: Define

$$\tilde{\lambda}_r = \frac{1}{n\xi_n(u_r; \theta_0)},$$

where $\theta_0 = (\gamma_0, \beta_0, \Lambda_0)$. We will prove that $\tilde{\Lambda}_n(.) = \sum_{k=1}^{q_n} \tilde{\lambda}_k I(u_k \leq .)$ converges to $\Lambda_0(.)$ uniformly on $[\rho, \tau]$. Note that

$$\tilde{\Lambda}_n(t) = \int_0^t \frac{1}{\xi_n(s; \theta_0)} \frac{1}{n} dN_n(s),$$

where $N_n(t) = \sum_{i=1}^{n} I(T_i \leq t, \delta_i = 1)$. Since classes of uniformly bounded and monotone functions are Glivenko-Cantelli (van der Vaart and Wellner (1996)), we have that

$$\frac{N_n(t)}{n} \to E_{x,z} \left\{ (1 - \pi(\gamma_0, x)) g(\beta_0, z) \int_{\rho_T}^{t} \lambda_0(s) \exp\{-g(\beta_0, z)\Lambda_0(s)\} F_{C|X,Z}(s|x, z) ds \right\}$$

uniformly on $[\rho_T, \tau_T]$ with probability 1. Furthermore we have that in

$$\xi_n(t; \theta_0) = \frac{1}{n} \sum_{i=1}^{n} g(\beta_0, Z_i) \{ I(T_i \geq t, \delta_i = 1) + a_i(\theta_0) I(T_i \geq t, \delta_i = 0) \},$$

$$\frac{1}{n}\sum_{i=1}^{n} g(\beta_0, Z_i) I(T_i \geq t, \delta_i = 1) \rightarrow$$
$$E_{x,z}\Big\{(1-\pi(\gamma_0,x))g(\beta_0,z)$$
$$\int_t^{\tau_T} g(\beta_0,z)\lambda_0(s)\exp\{-g(\beta_0,z)\Lambda_0(s)\}F_{C|X,Z}(s|x,z)ds\Big\}$$

and

$$\frac{1}{n}\sum_{i=1}^{n} g(\beta_0, Z_i) a_i(\theta_0) I(T_i \geq t, \delta_i = 0) \rightarrow$$
$$E_{x,z}\left\{-(1-\pi(\gamma_0,x))g(\beta_0,z)\int_t^{\tau_T}\exp\{-g(\beta_0,z)\Lambda_0(s)\}f_{C|X,Z}(s|x,z)ds\right\},$$

uniformly on $[\rho_T, \tau_T]$ with probability 1. Hence, after using integration by parts, we get that

$$\xi_n(t,\theta_0) \rightarrow E_{x,z}\{(1-\pi(\gamma_0,x))g(\beta_0,z)\exp\{-g(\beta_0,z)\Lambda_0(t)\}F_{C|X,Z}(t|x,z)\},$$

uniformly on $[\rho_T, \tau_T]$ with probability 1.
Since

$$\frac{d}{dt}E_{x,z}\left\{(1-\pi(\gamma_0,x))g(\beta_0,z)\int_{\rho_T}^{t}\lambda_0(s)\exp\{-g(\beta_0,z)\Lambda_0(s)\}F_{C|X,Z}(s|x,z)ds\right\}$$
$$= \lambda_0(t)E_{x,z}\{(1-\pi(\gamma_0,x))g(\beta_0,z)\exp\{-g(\beta_0,z)\Lambda_0(t)\}F_{C|X,Z}(t|x,z)\},$$

we get

$$\begin{aligned}\tilde{\Lambda}_n(t) &= \int_0^t \frac{1}{\xi_n(s,\theta_0)}\frac{1}{n}dN_n(s)\\ &\rightarrow \int_0^t \frac{\lambda_0(s)E_{x,z}\{(1-\pi(\gamma_0,x))g(\beta_0,z)\exp\{-g(\beta_0,z)\Lambda_0(s)\}F_{C|X,Z}(s|x,z)\}}{E_{x,z}\{(1-\pi(\gamma_0,x))g(\beta_0,z)\exp\{-g(\beta_0,z)\Lambda_0(s)\}F_{C|X,Z}(s|x,z)\}}ds\\ &= \Lambda_0(t)\end{aligned}$$

uniformly on $[\rho, \tau]$ with probability 1.
Step 4: Suppose that $\hat{\gamma}, \hat{\beta}$ and $\hat{\Lambda}_n$ converge to γ^+, β^+ and Λ^+ uniformly. We will show that

$$\frac{d\hat{\Lambda}_n}{d\tilde{\Lambda}_n}(t) \rightarrow \frac{d\Lambda^+}{d\Lambda_0}(t) \tag{2.3}$$

uniformly on $[\rho, \tau]$, with probability 1.

Since classes of uniformly bounded and monotone functions are Glivenko-Cantelli, we have that

$$\frac{1}{n}\sum_{i=1}^{n} g(\beta, Z_i) I(T_i \geq t, \delta_i = 1) \rightarrow$$
$$E_{x,z}\left\{ g(\beta, z)(1 - \pi(\gamma_0, x)) \int_t^{\tau_T} f_{Y>0}(s|z) F_{C|X,Z}(s|x, z) ds \right\} \tag{2.4}$$

and

$$\frac{1}{n}\sum_{i=1}^{n} g(\beta, Z_i) a_i(\theta) I(T_i \geq t, \delta_i = 0) \rightarrow \tag{2.5}$$
$$E_{x,z}\Bigg\{ g(\beta, z) \int_t^{\tau_T} - \frac{(1 - \pi(\gamma, x)) \exp\{-g(\beta, z)\Lambda(s)\}}{\pi(\gamma, x) + (1 - \pi(\gamma, x))(1 - \exp\{-g(\beta, z)\Lambda(s)\})} \times$$
$$[\pi(\gamma_0, x) + (1 - \pi(\gamma_0, x)) F_{Y>0}(s|z)] f_{C|X,Z}(s|x, z) ds \Bigg\},$$

uniformly on $[\rho_T, \tau_T]$, with probability 1. We have that $\xi_n(t; \hat{\theta}) \rightarrow \xi(t; \theta^+)$ uniformly on $[\rho_T, \tau_T]$ with probability 1, where $\xi(t, \theta)$ equals the sum of the right hand sides of (2.4) and (2.5) and $\theta^+ = (\gamma^+, \beta^+, \Lambda^+)$.
Since $\inf_{t\in[\rho,\tau]} \xi_n(t; \hat{\theta}) > M_\xi$, by Step 2, we have that $\inf_{t\in[\rho,\tau]} \xi(t; \theta^+) > M_\xi > 0$. In Step 3, we have proved that $\xi_n(t; \theta_0) \rightarrow \xi(t; \theta_0)$ uniformly on $[\rho_T, \tau_T]$, with probability 1. We get that,

$$\frac{d\hat{\Lambda}_n}{d\tilde{\Lambda}_n}(t) \rightarrow \frac{\xi(t; \theta_0)}{\xi(t; \theta^+)} \tag{2.6}$$

uniformly on $[\rho, \tau]$, with probability 1.
Since $d\tilde{\Lambda}_n(t) = \frac{1}{\xi_n(t;\theta_0)} \frac{1}{n} dN_n(t)$, we have

$$\begin{aligned} \Lambda^+(t) \leftarrow \hat{\Lambda}_n(t) &= \int_0^t \frac{1}{\xi_n(s; \hat{\theta})} \frac{1}{n} dN_n(s) \\ &= \int_0^t \frac{\xi_n(s; \theta_0)}{\xi_n(s; \hat{\theta})} d\tilde{\Lambda}_n(s) \rightarrow \int_0^t \frac{\xi(s; \theta_0)}{\xi(s; \theta^+)} d\Lambda_0(s). \end{aligned}$$

Hence,

$$\Lambda^+(t) = \int_0^t \frac{\xi(s; \theta_0)}{\xi(s; \theta^+)} d\Lambda_0(s).$$

Thus, $\Lambda^+(t)$ is absolutely continuous with respect to $\Lambda_0(s)$. We find

$$\frac{d\Lambda^+}{d\Lambda_0}(t) = \frac{\xi(t; \theta_0)}{\xi(t; \theta^+)}. \tag{2.7}$$

From (2.6) and (2.7), we conclude that

$$\frac{d\hat{\Lambda}_n}{d\tilde{\Lambda}_n}(t) \to \frac{d\Lambda^+}{d\Lambda_0}(t)$$

uniformly on $[\rho, \tau]$, with probability 1.

Step 5: By Helly's selection theorem, we can find a convergent subsequence of $\hat{\theta}$ (Parner (1998)). With abuse of notation, we let $\hat{\theta}$ be a convergent subsequence and θ^+ be the convergence point of $\hat{\theta}$. Then, (2.3) in Step 4 implies that

$$0 \leq \frac{1}{n} l^e(\hat{\gamma}, \hat{\beta}, \hat{\Lambda}_n) - \frac{1}{n} l^e(\gamma_0, \beta_0, \tilde{\Lambda}_n) \to KL(\theta^+, \theta_0),$$

where $KL(\theta^+, \theta_0)$ is the negative of the Kullback-Leibler divergence. It is defined by

$$\begin{aligned} KL(\theta^+, \theta_0) \quad = \quad & E[\delta\{\log(1 - \pi(\gamma^+, X)) + \log g(\beta^+, Z)\} - \delta g(\beta^+, Z)\Lambda^+(T) \\ & +(1-\delta)\log\{\pi(\gamma^+, X) + (1 - \pi(\gamma^+, X)) \\ & \quad (1 - \exp[-g(\beta^+, Z)\Lambda^+(T)])\}] \\ & -E[\delta\{\log(1 - \pi(\gamma_0, X)) + \log g(\beta_0, Z)\} - \delta g(\beta_0, Z)\Lambda_0(T) \\ & +(1-\delta)\log\{\pi(\gamma_0, X) + (1 - \pi(\gamma_0, X)) \\ & \quad (1 - \exp[-g(\beta_0, Z)\Lambda_0(T)])\}] \\ & +E\left[\delta \log \frac{\lambda^+(T)}{\lambda_0(T)}\right]. \end{aligned}$$

Since the Kullback-Leibler divergence is non-negative, $KL(\theta^+, \theta_0)$ is non-positive. So $KL(\theta^+, \theta_0) = 0$. Using the identifiability result in Theorem 1, this implies that $\theta^+ = \theta_0$, which terminates the proof of the consistency.

□

2.5 Asymptotic normality

In this section, we prove the asymptotic normality of the maximum likelihood estimators under some assumptions. We assume that the regularity conditions A_1, A_2 and A_5 of Section 2.2 still hold. We further assume a logistic regression model $\pi(\gamma, X) = \dfrac{\exp(X^\top\gamma)}{1 + \exp(X^\top\gamma)}$ for the probability of a zero response and we assume that $g(\beta, Z) = \exp(Z^\top\beta)$. These regression functions satisfy conditions A_3 and A_4. The first element of the X-vector is assumed to be one, such that the first element of γ represents the intercept of the logistic regression model. We also presume that $P(X^\top c = 0) = 1$ implies that $c = 0$. Let $\theta_0 = (\gamma_0, \beta_0, \Lambda_0)$ be the true values of the parameters.

Theorem 4. *Under the assumptions mentioned before, the scaled process* $\sqrt{n}(\hat{\gamma}-\gamma_0, \hat{\beta}-\beta_0, \hat{\Lambda}_n(\cdot)-\Lambda_0(\cdot))$ *converges weakly to a zero-mean Gaussian process in the metric space* $\Theta_\gamma \times \Theta_\beta \times l^\infty[\rho,\tau]$*, where* $l^\infty[\rho,\tau]$ *is the space containing all bounded valued cadlag (right continuous with left limits existing) functions on* $[\rho,\tau]$*, equipped with the supremum norm.*

Proof. To prove this theorem, we apply Theorem 3.3.1 of van der Vaart and Wellner (1996). We first need to define a random map Ψ_n and a fixed map Ψ from a set, say $\mathcal{A}$, containing the true values of the parameters and (at least asymptotically) the possible values of the estimators. Theorem 3.3.1 of van der Vaart and Wellner (1996) involves verifying four sufficient conditions: $\sqrt{n}(\Psi_n(\theta_0) - \Psi(\theta_0))$ converges in distribution to a tight Gaussian process $\mathcal{W}$ (weak convergence of the empirical process at the truth), an approximation condition of the score operator, Fréchet differentiability of the asymptotic score and continuous invertibility of the information operator.

We define a set

$$\begin{aligned}\mathcal{H}_p \quad = \quad & \{h=(h_1,h_2,h_3) : h_1 \in \mathbb{R}^{d_x+1},\ h_2 \in \mathbb{R}^{d_z} \text{ and } h_3 \text{ is a function} \\ & \text{of bounded variation on } [\rho,\tau];\ |h_1|+|h_2|+\|h_3\|_v \le p\},\end{aligned}$$

where $\|h_3\|_v$ is the absolute value of $h_3(\rho)$ plus the total variation of h_3 over the interval $[\rho,\tau]$, that is $\|h_3\|_v = |h_3(\rho)| + \int_\rho^\tau |dh_3|$. We shall restrict consideration to h_3 which is cadlag. We can consider the parameter $\theta=(\gamma,\beta,\Lambda)$ as a functional on $\mathcal{H}_p$ given by $\theta(h) = h_1^\top\gamma + h_2^\top\beta + \int_\rho^\tau h_3 d\Lambda$ and the parameter space Θ as a subset of $l^\infty(\mathcal{H}_p)$, the space of bounded functionals on $\mathcal{H}_p$, equipped with the supremum norm $\|\theta\|_p = \sup_{h\in\mathcal{H}_p} |\theta(h)|$.

The set $\mathcal{A}$ is defined as a small neighborhood containing the true parameters θ_0 as

$$A = \{\theta=(\gamma,\beta,\Lambda) : \|\theta-\theta_0\|_p < \epsilon\},$$

where ϵ is a small positive constant. Theorem 3 assures that, when the sample size is sufficiently large, the estimates $\hat{\theta}=(\hat{\gamma},\hat{\beta},\hat{\Lambda}_n)$ fall into the set A almost surely.

We now define Ψ_n and Ψ as random and fixed maps respectively from A to $l^\infty(\mathcal{H}_p)$. More specifically, we define

$$\Psi_n(\theta)[h_1,h_2,h_3] = \mathcal{P}_n(\phi(\theta,h)) \text{ and } \Psi(\theta)[h_1,h_2,h_3] = E(\phi(\theta,h)),$$

where $\mathcal{P}_n$ denotes the empirical measure, $\phi(\theta,h) = h_1^\top l_\gamma(\theta) + h_2^\top l_\beta(\theta) + l_\Lambda(\theta)[h_3]$, l_γ is the score vector for γ, l_β is the score vector for β and $l_\Lambda(\theta)[h_3]$ is the score for Λ along the submodel $\Lambda(\cdot) + \epsilon\int_\rho^\cdot h_3 d\Lambda$. The maximum likelihood estimators and the true parameters satisfy $\Psi_n(\hat{\theta}) = 0$ and $\Psi(\theta_0) = 0$.

Direct calculations give

$$\begin{aligned}
l_\gamma(\theta) &= -\delta \frac{\exp(X^\top\gamma)}{1+\exp(X^\top\gamma)} X \\
&\quad +(1-\delta)\frac{\dfrac{\exp(X^\top\gamma)}{1+\exp(X^\top\gamma)}\exp[-\exp(Z^\top\beta)\Lambda(T)]X}{\exp(X^\top\gamma)+1-\exp[-\exp(Z^\top\beta)\Lambda(T)]}, \\
l_\beta(\theta) &= \delta Z - \delta\Lambda(T)\exp(Z^\top\beta)Z \\
&\quad +(1-\delta)\frac{\exp[-\exp(Z^\top\beta)\Lambda(T)]\exp(Z^\top\beta)\Lambda(T)Z}{\exp(X^\top\gamma)+1-\exp[-\exp(Z^\top\beta)\Lambda(T)]}, \\
l_\Lambda(\theta)[h_3] &= \delta h_3(T) - \delta\exp(Z^\top\beta)\int_\rho^T h_3(s)d\Lambda(s) \\
&\quad +(1-\delta)\frac{\exp[-\exp(Z^\top\beta)\Lambda(T)]\exp(Z^\top\beta)\int_\rho^T h_3(s)d\Lambda(s)}{\exp(X^\top\gamma)+1-\exp[-\exp(Z^\top\beta)\Lambda(T)]}.
\end{aligned}$$

The classes A and $\mathcal{H}_p$ are Donsker classes. Since $\phi(\theta,h)$ is a bounded Lipschitz functional with respect to $A\times\mathcal{H}_p$, Theorem 2.10.6 of van der Vaart and Wellner (1996) yields that the class $\{\phi(\theta,h):(\theta,h)\in A\times\mathcal{H}_p\}$ is a Donsker class. Since

$$\{\phi(\theta_0,h):h\in\mathcal{H}_p\}\subset\{\phi(\theta,h):(\theta,h)\in A\times\mathcal{H}_p\},$$

$\{\phi(\theta_0,h):h\in\mathcal{H}_p\}$ is also a Donsker class. Hence $\sqrt{n}(\Psi_n(\theta_0)-\Psi(\theta_0))$ converges in distribution to a tight Gaussian process $\mathcal{W}$, so that the first sufficient condition of Theorem 3.3.1 in van der Vaart and Wellner (1996) is satisfied. We further have that the class $\{\phi(\theta,h)-\phi(\theta_0,h):(\theta,h)\in A\times\mathcal{H}_p\}$ is a Donsker class. One can also show that

$$E\sup_{h\in\mathcal{H}_p}(\phi(\theta,h)-\phi(\theta_0,h))^2<\infty,$$

which by the dominated convergence theorem implies that

$$\sup_{h\in\mathcal{H}_p}E(\phi(\theta,h)-\phi(\theta_0,h))^2\to 0,$$

when $\theta\to\theta_0$. Since $\hat\theta\to\theta_0$ uniformly and almost surely by Theorem 3, Lemma 3.3.5 of van der Vaart and Wellner (1996) implies that

$$\|\sqrt{n}\{\Psi_n(\hat\theta)-\Psi(\hat\theta)\}-\sqrt{n}\{\Psi_n(\theta_0)-\Psi(\theta_0)\}\|_p=o_p(1+\sqrt{n}\|\hat\theta-\theta_0\|). \tag{2.8}$$

The approximation of the score operator in (2.8) is the second sufficient condition of Theorem 3.3.1 of van der Vaart and Wellner (1996).

For the Fréchet differentiability of the asymptotic score Ψ at θ_0, first note that Ψ is Gâteaux differentiable in a neighborhood of θ_0 and that the derivative is continuous. Since X and Z are bounded, one can show that

$$\sup\left\{\left\|\frac{\partial}{\partial\lambda}\Psi(\theta_0+\lambda\theta)\right\|_p : \|\theta\|_p \leq 1, |\lambda| \leq \epsilon\right\} < \infty$$

for an $\epsilon > 0$. Proposition 1 in the Appendix of Bickel et al. (1993) assures that Ψ is Fréchet differentiable at θ_0, which is the third sufficient condition of Theorem 3.3.1 of van der Vaart and Wellner (1996).

It remains to show that the information operator $\dot{\Psi}_{(\gamma_0,\beta_0,\Lambda_0)}$ (i.e. the Fréchet derivative of Ψ at θ_0) is continuously invertible on its range. Let $\gamma_\epsilon = \gamma_0 + \epsilon\tilde{h}_1, \beta_\epsilon = \beta_0 + \epsilon\tilde{h}_2$ and $\Lambda_\epsilon(\cdot) = \Lambda_0(\cdot) + \epsilon\int_\rho^\cdot \tilde{h}_3 d\Lambda_0$. A calculation and algebraic manipulation shows that

$$\begin{aligned}
&\dot{\Psi}_{(\gamma_0,\beta_0,\Lambda_0)}(\tilde{h}_1,\tilde{h}_2,\int_\rho^\cdot \tilde{h}_3 d\Lambda_0)[h] \\
&= \frac{\partial}{\partial\epsilon}E\left\{h_1^\top l_\gamma(\gamma_\epsilon,\beta_\epsilon,\Lambda_\epsilon) + h_2^\top l_\beta(\gamma_\epsilon,\beta_\epsilon,\Lambda_\epsilon) + l_\Lambda(\gamma_\epsilon,\beta_\epsilon,\Lambda_\epsilon)[h_3]\right\}\Big|_{\epsilon=0} \qquad (2.9)\\
&= \tilde{h}_1^\top Q_1(h_1,h_2,h_3) + \tilde{h}_2^\top Q_2(h_1,h_2,h_3) + \int_\rho^\tau Q_3(h_1,h_2,h_3)(t)\tilde{h}_3(t)d\Lambda_0(t), \qquad (2.10)
\end{aligned}$$

where $Q_i, i = 1,2,3$ are linear operators satisfying

$$\begin{aligned}
Q_1(h_1,h_2,h_3) &= B_1h_1 + B_2h_2 + \int_\rho^\tau \left\{D_1(s)^\top\left(\int_\rho^s h_3(t)d\Lambda_0(t)\right)\right\}ds,\\
Q_2(h_1,h_2,h_3) &= B_3h_1 + B_4h_2 + \int_\rho^\tau \left\{D_2(s)^\top\left(\int_\rho^s h_3(t)d\Lambda_0(t)\right)\right\}ds,\\
Q_3(h_1,h_2,h_3)(t) &= D_3(t)^\top h_1 + D_4(t)^\top h_2 + D_5(t)h_3(t)\\
&\quad + \int_t^\tau \left\{D_6(s)\int_\rho^s h_3(t)d\Lambda_0(t)\right\}ds,
\end{aligned}$$

where $D_j(\cdot),\ j = 1,\ldots,6$ are continuously differentiable functions depending on the true distribution. They are obtained by doing some algebraic manipulations on the derivatives in (2.9) in the direction of the baseline cumulative hazard function. For example, we have that

$$\begin{aligned}
D_5(t) &= -\int_{X,Z}\frac{\exp(z^\top\beta_0)}{1+\exp(x^\top\gamma_0)}\left[\int_t^\tau f_{Y>0}(s|z)F_{C|X,Z}(s|x,z)ds\right]P(dxdz)\\
&\quad + \int_{X,Z}\frac{\exp(z^\top\beta_0)}{1+\exp(x^\top\gamma_0)}\left[\int_t^\tau \exp[-\exp(z^\top\beta_0)\Lambda_0(s)]f_{C|X,Z}(s|x,z)ds\right]P(dxdz).
\end{aligned}$$

The matrices $B_i,\ i = 1, \ldots, 4$ are given by

$$
\begin{aligned}
B_1 &= E\bigg(\frac{\partial^2}{\partial\gamma\partial\gamma^\top}\log L(\gamma_0,\beta_0,\Lambda_0)\bigg)\\
&= E\bigg(-\delta\frac{\exp(X^\top\gamma_0)}{[1+\exp(X^\top\gamma_0)]^2}XX^\top\\
&\quad -(1-\delta)\frac{\frac{\exp(X^\top\gamma_0)}{1+\exp(X^\top\gamma_0)}\exp[-\exp(Z^\top\beta_0)\Lambda_0(T)]\exp(X^\top\gamma_0)XX^\top}{\{\exp(X^\top\gamma_0)+1-\exp[-\exp(Z^\top\beta_0)\Lambda_0(T)]\}^2}\\
&\quad +(1-\delta)\frac{\frac{\exp(X^\top\gamma_0)}{[1+\exp(X^\top\gamma_0)]^2}\exp[-\exp(Z^\top\beta_0)\Lambda_0(T)]XX^\top}{\exp(X^\top\gamma_0)+1-\exp[-\exp(Z^\top\beta_0)\Lambda_0(T)]}\bigg),\\
B_2 &= E\bigg(\frac{\partial^2}{\partial\beta\partial\gamma^\top}\log L(\gamma_0,\beta_0,\Lambda_0)\bigg)\\
&= -E\bigg((1-\delta)\frac{\exp[-\exp(Z^\top\beta_0)\Lambda_0(T)]\exp(Z^\top\beta_0)\exp(X^\top\gamma_0)\Lambda_0(T)ZX^\top}{\{\exp(X^\top\gamma_0)+1-\exp[-\exp(Z^\top\beta_0)\Lambda_0(T)]\}^2}\bigg),\\
B_3 &= B_2^\top,\\
B_4 &= E\bigg(\frac{\partial^2}{\partial\beta\partial\beta^\top}\log L(\gamma_0,\beta_0,\Lambda_0)\bigg)\\
&= E\bigg(-\delta\Lambda_0(T)\exp(Z^\top\beta_0)ZZ^\top\\
&\quad -(1-\delta)\frac{\exp[-\exp(Z^\top\beta_0)\Lambda_0(T)][\exp(Z^\top\beta_0)]^2[\Lambda_0(T)]^2ZZ^\top}{\exp(X^\top\gamma_0)+1-\exp[-\exp(Z^\top\beta_0)\Lambda_0(T)]}\\
&\quad +(1-\delta)\frac{\exp[-\exp(Z^\top\beta_0)\Lambda_0(T)]\exp(Z^\top\beta_0)\Lambda_0(T)ZZ^\top}{\exp(X^\top\gamma_0)+1-\exp[-\exp(Z^\top\beta_0)\Lambda_0(T)]}\\
&\quad -(1-\delta)\frac{\{\exp[-\exp(Z^\top\beta_0)\Lambda_0(T)]\}^2[\exp(Z^\top\beta_0)]^2[\Lambda_0(T)]^2ZZ^\top}{\{\exp(X^\top\gamma_0)+1-\exp[-\exp(Z^\top\beta_0)\Lambda_0(T)]\}^2}\bigg).
\end{aligned}
$$

For the continuous invertibility, it suffices to show that $Q(h_1,h_2,h_3) = (Q_1(h_1,h_2,h_3), Q_2(h_1,h_2,h_3), Q_3(h_1,h_2,h_3))$, viewed as an operator from $\mathcal{H}_\infty$ to $\mathcal{H}_\infty$ is continuously invertible (see f.e. Parner (1998) and Kim et al. (2010)). We show that Q is one-to-one and can be written as a sum of a continuously invertible linear operator L and a compact operator C, which implies that Q is continuously invertible (see Rudin (1991)).

That Q is one-to-one means that if $\|h\| > 0$ then $\|Q(h)\| > 0$. Suppose that $Q_i(h_1,h_2,h_3) = 0$ for $i = 1,2,3$ for some $h = (h_1,h_2,h_3)$. We wish to show that $h_1 = 0$, $h_2 = 0$ and $h_3 = 0$. Considering (2.9)-(2.10) and choosing $\tilde{h}_1 = h_1$, $\tilde{h}_2 = h_2$,

$\tilde{h}_3 = h_3$, we have that

$$E\left\{h_1^\top l_\gamma(\gamma_0,\beta_0,\Lambda_0) + h_2^\top l_\beta(\gamma_0,\beta_0,\Lambda_0) + l_\Lambda(\gamma_0,\beta_0,\Lambda_0)[h_3]\right\}^2 = 0.$$

Hence,

$$h_1^\top l_\gamma(\gamma_0,\beta_0,\Lambda_0) + h_2^\top l_\beta(\gamma_0,\beta_0,\Lambda_0) + l_\Lambda(\gamma_0,\beta_0,\Lambda_0)[h_3] = 0 \tag{2.11}$$

almost surely. Note that the left hand side of (2.11) is the score function at the true value along a one-dimensional submodel $\{\gamma_0 + \epsilon h_1, \beta_0 + \epsilon h_2, \Lambda_0(\cdot) + \epsilon \int_\rho^\cdot h_3 d\Lambda_0\}$. Consider (2.11) and set $\delta = 0$ and $T = s \in [\rho, \tau]$. Then we obtain the following equation:

$$\frac{\exp(X^\top\gamma_0)}{1+\exp(X^\top\gamma_0)} X^\top h_1 + \exp(Z^\top\beta_0)\Lambda_0(s) Z^\top h_2 + \exp(Z^\top\beta_0)\int_\rho^s h_3(t) d\Lambda_0(t) = 0,$$

almost surely on $\{\delta = 0\}$. Suppose that not both of h_2 and $h_3(\cdot)$ are equal to zero. In that case, $\dfrac{\exp(X^\top\gamma_0)}{1+\exp(X^\top\gamma_0)} X^\top h_1$ would be constant on $\{\delta = 0\}$, which is impossible. Therefore, $h_2 = 0$ and $h_3(\cdot) = 0$. Since, by assumption, $P(X^\top h_1 = 0) = 1$ implies that $h_1 = 0$, we conclude that Q is a one-to-one map.

Finally, we prove that Q can be written as a sum of a continuously invertible linear operator L and a compact operator C. Let $L(h) = (L_{\gamma_0}(h_1), L_{\beta_0}(h_2), L_{\Lambda_0}(h_3))$, where

$$\begin{aligned} L_{\gamma_0}(h_1) &= E\left(\frac{\partial^2}{\partial\gamma\partial\gamma^\top}\log L(\gamma_0,\beta_0,\Lambda_0)\right) h_1 = B_1 h_1, \\ L_{\beta_0}(h_2) &= E\left(\frac{\partial^2}{\partial\beta\partial\beta^\top}\log L(\gamma_0,\beta_0,\Lambda_0)\right) h_2 = B_4 h_2, \\ L_{\Lambda_0}(h_3)(\cdot) &= D_5(\cdot) h_3(\cdot). \end{aligned}$$

One can show that the matrices B_1 and B_4 are invertible by a direct calculation. Hence L_{γ_0} and L_{β_0} are one-to-one. Since L_{γ_0} and L_{β_0} are finite dimensional operators and $D_5(\cdot) \neq 0$ on $]\rho, \tau[$, L is a continuously invertible linear operator. To show that $Q(h) - L(h)$ is compact, we show that for an arbitrary sequence $\{h_n\}_{n\geq 1}$ in $\mathcal{H}_p$, there exists a convergent subsequence of $\{Q(h_n) - L(h_n)\}$. Helly's selection theorem assures that there exists a subsequence $\{n_k\}_{k\geq 1}$, and a function, h, such that h_{n_k} converges pointwise to h since $\mathcal{H}_p$ is compact. Applying the dominated convergence theorem, one sees that the subsequence $Q(h_{n_k}) - L(h_{n_k})$ converges to $Q(h) - L(h)$.

We verified the four sufficient conditions of Theorem 3.3.1 of van der Vaart and Wellner (1996). Hence

$$\sqrt{n}\{\hat{\gamma} - \gamma_0, \hat{\beta} - \beta_0, \hat{\Lambda}_n(\cdot) - \Lambda_0(\cdot)\} \to -\dot{\Psi}^{-1}_{(\gamma_0,\beta_0,\Lambda_0)}\mathcal{W},$$

where we can calculate the asymptotic variance of $\mathcal{W}(h)$ as

$$\text{var}(\mathcal{W}(h)) = -\dot{\Psi}_{(\gamma_0,\beta_0,\Lambda_0)}(h_1, h_2, \int_\rho^\cdot h_3 d\Lambda_0)[h].$$

The asymptotic covariance of $\mathcal{W}(h)$ and $\mathcal{W}(g)$ can be computed as

$$\text{cov}(\mathcal{W}(h), \mathcal{W}(g)) = -\dot{\Psi}_{(\gamma_0,\beta_0,\Lambda_0)}(g_1, g_2, \int_\rho^\cdot g_3 d\Lambda_0)[h].$$

□

2.6 Simulation study

In this section, we show the performance of the zero-inflated semi-parametric Cox's regression model for finite data samples. Hereto we set up a simulation study and generate the data sets as follows:

- $X = Z \sim \text{Uniform}[0, 10]$; $C \sim \text{Uniform}[0, c]$.
- Probability of a zero response: $\pi(\gamma, X) = \frac{\exp(\gamma_0+\gamma_1 * X)}{1+\exp(\gamma_0+\gamma_1 * X)}$.
- $J \sim \text{Bernoulli}(1 - \pi(\gamma, X))$.
  If $J = 0$, then $T = C$ and $\delta = 0$.
  If $J = 1$, then $T = \max(Y, C)$ and $\delta = I(Y \geq C)$, where $Y \sim$ Cox's model ($g(\beta, z) = e^{\beta z}$; baseline hazard: Weibull(a_0,b_0)).

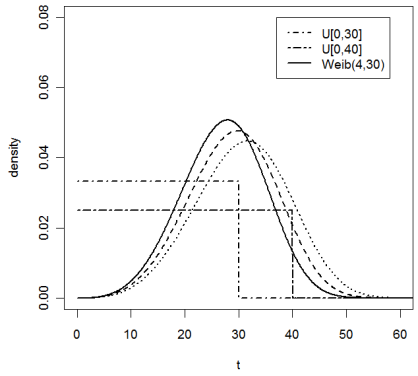
2.6.1 Settings

We consider in this simulation study five different settings, with various values of γ_0, γ_1, β, the baseline parameters a_0 and b_0, and c. The different settings are given in Table 2.1.

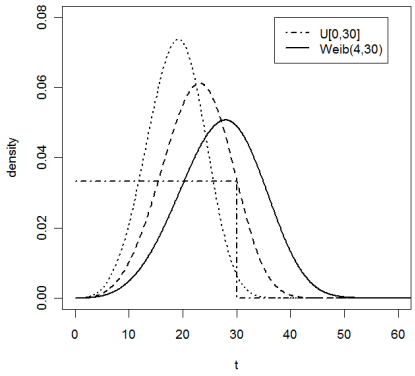
	γ_0	γ_1	β	a_0	b_0	c	$\Lambda_0(20)$	$\Lambda_0(30)$	$\Lambda_0(40)$
Setting 1	-0.3	0.15	-0.05	4	30	30	0.1975	1	3.1605
Setting 2	-2	0.10	-0.05	4	30	30	0.1975	1	3.1605
Setting 3	-0.3	0.15	-0.05	4	30	40	0.1975	1	3.1605
Setting 4	-0.3	0.15	0.15	4	30	30	0.1975	1	3.1605
Setting 5	0.5	-0.2	0.15	2	25	30	0.64	1.44	2.56

Table 2.1: Parameter values for γ_0, γ_1, β, a_0, b_0 and c in the different settings. The values for the true baseline cumulative hazard function at $t = 20$, $t = 30$ and $t = 40$ are also calculated.

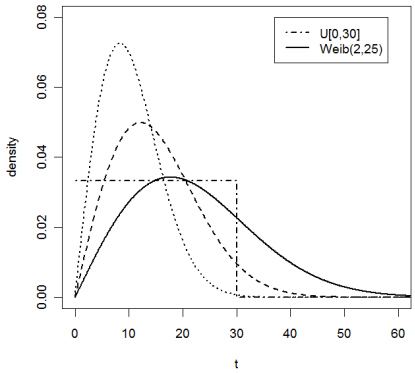
From the density plots in Figure 2.1, it is clear that in setting 3 we will have more censoring among the responders, compared to setting 1. Further, we note that in settings 1-3, the Weibull distribution of the responders has a larger scale parameter for individuals with larger values of X. In settings 4-5, the Weibull distribution of the responders has a smaller scale parameter for individuals with larger values of X.



(a) Setting 1-3



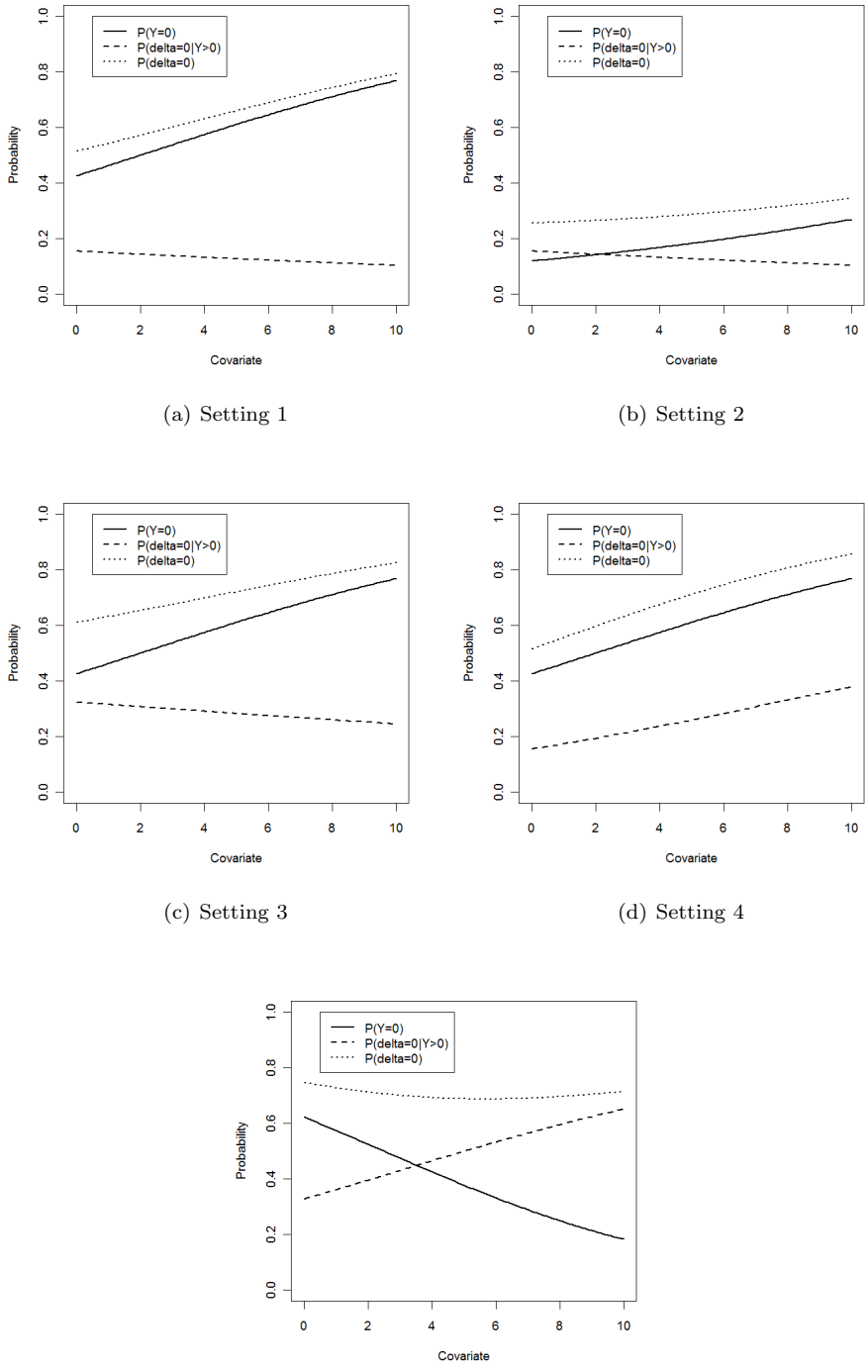
(b) Setting 4



(c) Setting 5

Figure 2.1: Densities of C (Uniform); $Y|Y > 0, X = 0$ (Weibull-solid); $Y|Y > 0, X = 5$ (Weibull-dashed) and $Y|Y > 0, X = 10$ (Weibull-dotted).

To get more insight in the censoring mechanism, we plot in Figure 2.2 for each (marginal) setting the conditional probability of a zero response and the conditional probability of a censored observation when we know that the response is not zero.



(a) Setting 1

(b) Setting 2

(c) Setting 3

(d) Setting 4

(e) Setting 5

Figure 2.2: The probability of a zero response and the conditional censoring probabilities.

These conditional probabilities are combined to also show the overall conditional probability of a censored observation. Note that 'conditional' means conditional on the covariates in the model.

For the first setting, we note that the overall conditional censoring probability is high. This is mainly due to a high conditional probability of a zero response. The second setting corresponds to a low conditional probability of a zero response, combined with a rather small conditional censoring probability for the responders. In the third and fourth setting, there is a high overall conditional censoring probability, induced by a high conditional probability of a zero response, together with a rather high conditional censoring probability for the responders. In setting 5, the conditional probability of a zero response is high and decreases with increasing covariate values, while the conditional censoring probability for the responders is increasing. This consequently gives us that for large values of the covariate, the overall conditional censoring probability is mostly determined by the high percentage of censoring for the responders.

In each simulation setting we generate 500 data sets, with $n = 100$, $n = 200$ and $n = 500$ observations. The maximization of the likelihood is conducted using the numerical optimizer **nlm** in the statistical software package **R**.

2.6.2 Results

For the five different settings, we calculate an estimate of the bias and the standard deviation of the 500 ML estimates. The results are shown in Table 2.2.

In the five different settings, we see in the table that the bias and the standard deviation of the maximum likelihood estimates decrease as the sample size increases. This is in line with what we expect based on the theoretical results. However we note that there are some differences between the settings. In the first setting, the bias and the standard deviation for all parameter estimates are low. This was expected since the parameters in this setting were created such that there was a clear conditional probability of a zero response value. The larger bias and standard deviation for the logistic parameters in setting 2, compared to setting 1, is due to the small rate of zero responses. The estimators for parameters in the Cox's regression model, on the other hand, produced smaller biases and standard errors. The higher percentage of censoring among the responders in settings 3 and 4 explains the higher biases and

		Setting 1	Setting 2	Setting 3	Setting 4	Setting 5
$n = 100$	$\hat{\gamma}_0$	-0.0235 (0.5113)	-0.3310 (1.8190)	-0.0148 (0.5980)	0.0017 (0.5203)	0.3075 (5.3915)
	$\hat{\gamma}_1$	0.0055 (0.0868)	-0.1398 (2.8894)	0.0071 (0.1045)	0.0090 (0.0959)	-0.3489 (7.3323)
	$\hat{\beta}$	-0.0001 (0.0721)	0.00003 (0.0397)	-0.0034 (0.0745)	0.0085 (0.0859)	0.0128 (0.0762)
	$\hat{\Lambda}(20)$	0.0031 (0.1288)	0.0009 (0.0825)	0.0098 (0.1473)	0.0029 (0.1141)	0.0073 (0.3153)
	$\hat{\Lambda}(30)$	0.0298 (0.4487)	0.0078 (0.2462)	0.0387 (0.4375)	0.0137 (0.3952)	0.0172 (0.6196)
	$\hat{\Lambda}(40)$	0.3001 (1.7984)	0.1126 (0.8649)	0.2564 (1.6413)	-0.5214 (0.6599)	-0.2221 (0.7988)
$n = 200$	$\hat{\gamma}_0$	0.0016 (0.3449)	-0.1056 (0.6292)	-0.0381 (0.4151)	-0.0121 (0.3713)	0.0542 (0.4934)
	$\hat{\gamma}_1$	0.0027 (0.0608)	0.0067 (0.0922)	0.0049 (0.0701)	0.0026 (0.0691)	-0.0084 (0.1083)
	$\hat{\beta}$	-0.0024 (0.0457)	-0.0008 (0.0305)	-0.0062 (0.0490)	0.0039 (0.0499)	0.0076 (0.0461)
	$\hat{\Lambda}(20)$	0.0003 (0.0913)	0.0045 (0.0591)	0.0084 (0.1027)	-0.0037 (0.0713)	-0.0116 (0.1945)
	$\hat{\Lambda}(30)$	0.0311 (0.2662)	0.0092 (0.1772)	0.0366 (0.2819)	0.0145 (0.2696)	-0.0064 (0.4012)
	$\hat{\Lambda}(40)$	0.1731 (0.9023)	0.0618 (0.6282)	0.1953 (0.9839)	-0.2055 (0.6089)	-0.0591 (0.6517)
$n = 500$	$\hat{\gamma}_0$	-0.0197 (0.2097)	0.0069 (0.3475)	-0.0195 (0.2479)	0.0070 (0.2330)	0.0063 (0.2989)
	$\hat{\gamma}_1$	0.0043 (0.0359)	-0.0020 (0.0524)	0.0042 (0.0439)	-0.0010 (0.0419)	-0.00004 (0.0687)
	$\hat{\beta}$	-0.0017 (0.0280)	-0.0018 (0.0179)	-0.0020 (0.0297)	0.0014 (0.0304)	0.0033 (0.0273)
	$\hat{\Lambda}(20)$	0.0036 (0.0524)	-0.0004 (0.0365)	0.0041 (0.0638)	-0.0005 (0.0437)	-0.0044 (0.1230)
	$\hat{\Lambda}(30)$	0.0198 (0.1505)	0.0097 (0.1076)	0.0110 (0.1605)	-0.0017 (0.1560)	-0.0160 (0.2354)
	$\hat{\Lambda}(40)$	0.0538 (0.4926)	0.0307 (0.3666)	0.0555 (0.5287)	-0.0377 (0.5590)	-0.0382 (0.4374)

Table 2.2: The bias and standard deviation (between brackets) of the simulated ML estimates.

standard deviations compared to setting 1. Furthermore we note that, for the data sets with 100 observations, the estimates of the bias and the standard deviations for the logistic parameters in the second and fifth settings are rather large. This can be explained by the fact that for some of these generated data sets our algorithm was not able to estimate the probability of a zero response properly. We see that the censored observations in the data sets generated in setting 2 are mostly censored non-zero response values for responders, such that a zero-inflated model may not be applicable in this case. Also in some data sets generated in setting 5 we see this behavior, but in general the censored observations in these data sets are a mix of censored positive response values and actually zero values for non-responders.

2.6.3 Comparison with a left-censored semi-parametric Cox's regression model

In Subsection 2.6.2, we noticed that in some simulation settings the zero-inflated semi-parametric model does not perform well for small data sets. In this aspect, we also apply a left-censored semi-parametric Cox's regression model (see Kim et al. (2010)) to the data sets of sample size 100 from settings 1-5 and compare the results with the previous ones. The model of Kim et al. (2010) is developed for doubly-censored data, but it can also be applied to left-censored data. The likelihood is given by

$$\begin{aligned} L^l(\beta,\Lambda) &= \prod_{i=1}^{n} \{\Lambda\{T_i\} g(\beta,Z_i) \exp\left(-g(\beta,Z_i)\Lambda(T_i)\right)\}^{\delta_i} \\ &\quad \{(1-\exp(-g(\beta,Z_i)\Lambda(T_i)))\}^{1-\delta_i}. \end{aligned}$$

Kim et al. (2010) consider a non-parametric step function for the baseline cumulative hazard function. The step function has jumps at all uncensored observations and at the first censored observation, if this observation is smaller than the smallest uncensored observation.

In Table 2.3, the bias and the standard deviation of the estimates for the hazard parameter β are shown.

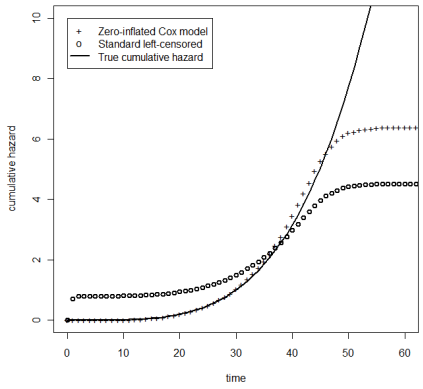
Setting 1	Setting 2	Setting 3	Setting 4	Setting 5
0.0870 (0.0402)	0.0257 (0.0366)	0.0810 (0.0413)	-0.0283 (0.0405)	-0.0976 (0.0416)

Table 2.3: The bias and standard deviation (between brackets) of the estimates for β in the left-censored semi-parametric Cox's regression model, based on the data sets with 100 observations.

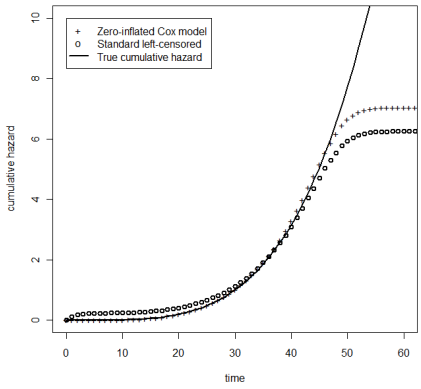
In Figure 2.3, we show the pointwise average of the estimates of the baseline cumulative hazard function at time points ranging from 0 to 60 for both the zero-inflated and the standard Cox's model for analyzing left-censored data. These averages are compared with the values of the true baseline cumulative hazard function at these time points.

From Table 2.3 and Figure 2.3, we learn that for setting 2 a standard left-censored semi-parametric Cox's regression model can be sufficient. For this setting, the average estimates of the baseline cumulative hazard function for the standard left-censored semi-parametric Cox's regression model and for the zero-inflated model are close to each other. For settings 1 and 3-5, we see in the beginning a clear and large jump in the graphs for the standard left-censored semi-parametric Cox's regression model, which indicates that it is very useful to account for the extra probability of a zero response. Furthermore we note that the average estimates of the baseline cumulative hazard functions for the zero-inflated model are very close to the true function values on a quite large interval of time points. This confirms that there is little bias in the estimates of the baseline cumulative hazard function at time points that are not too high.

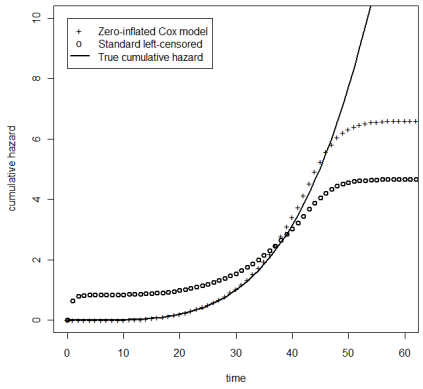
In general, we can conclude that the zero-inflated semi-parametric model performs better than the standard left-censored semi-parametric model when there is a clear conditional probability of a zero response value. So we suggest to use the zero-inflated model in practice for analyzing left-censored data for which it is known that there may be some underlying zeros present in the data set. As relevant examples, we can consider studies where investigators want to know the amount of a certain toxic product in ground water or in the food industry. There may be many samples in which the product is not present. However, by detection limits in the measuring mechanism, it is not possible to observe these zeros. In the next section, we consider another example, namely a biological study on ethanol-induced sleeping time in genetically selected mice (Markel et al. (1995)). Some mice did actually not fall asleep, but got the value of the detection limit for their response sleep time.



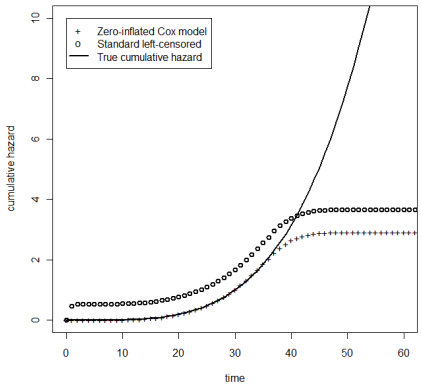
(a) Setting 1



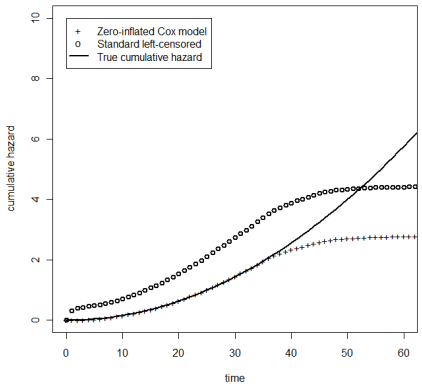
(b) Setting 2



(c) Setting 3



(d) Setting 4



(e) Setting 5

Figure 2.3: Pointwise average of the estimates of the cumulative hazard function at different time points, compared with the true baseline cumulative hazard function.

2.7 Example: Modeling ethanol-induced anesthesia (sleep time)

In this section, we illustrate the zero-inflated Cox's regression model with a practical study on ethanol-induced anesthesia (sleep time) in genetically-selected strains of mice, described in Section 1.2. In this example, we consider the same regression model for the sleep time as in the simulation study and take a logistic regression model for the probability of a zero outcome value and the regular Cox's regression model for the non-zero outcome value with $g(\beta, z) = e^{z^\top \beta}$. We investigate the influence of the following covariates on sleep time: sex, albinism (which is a binary variable indicating whether the mouse was albino), weight at the first test session, and an interaction between sex and albinism.

As we saw in the methodology, the zero-inflated Cox's regression model assumes that the baseline hazard is zero before the smallest uncensored observation. Consequentially, the probability of a value of sleep time between zero and the detection limit is also zero in this model. In Section 2.1, we noted that fitting the zero-inflated Cox's regression model simplifies in this case to fitting a logistic regression model on the censoring indicator variables and fitting a Cox's regression model on the subgroup of the uncensored observations. This can be done with standard statistical software. The parameter estimates and their standard errors are given in Table 2.4. In the same table we also give a parametric (zero-inflated) Logistic-Weibull model to compare with the zero-inflated Cox's regression model. In the parametric Logistic-Weibull model, we assume that the baseline hazard comes from a Weibull distribution. The probability of a non-zero censored value of the sleep time has, in this case, an expression which depends on the parameters of the Weibull baseline hazard and is non-zero. Furthermore we fit a standard left-censored semi-parametric Cox's regression model (see Kim et al. (2010)) to the data. The corresponding baseline hazard function has a mass at the detection limit and at all uncensored observations.

We notice in Table 2.4 that in both the zero-inflated Cox's model and the parametric Logistic-Weibull model, the same covariates have a significant effect in the logistic part of each model. An albino mouse has a significant higher probability of having a zero value for the sleep time than a non-albino mouse. Furthermore we note that the gender of a mouse also has a significant effect in the logistic part, through its

	Zero-inflated Cox's model	Parametric Logistic-Weibull model	Standard left-censored Cox's model
		Logistic part	
Intercept	-3.8111 (0.3540)***	-3.8233 (0.3580)***	
Sex	0.6626 (0.4929)	0.6692 (0.4975)	
Albinism	1.3294 (0.4457)***	1.3364 (0.4484)***	
Sex*Albinism	-1.2629 (0.6880)*	-1.2703 (0.6926)*	
Weight	0.0531 (0.0684)	0.0539 (0.0689)	
		Hazard part	
Sex	0.0560 (0.0895)	0.0475 (0.0911)	0.0748 (0.0880)
Albinism	0.1022 (0.1045)	0.0665 (0.1048)	0.1655 (0.1006)*
Sex*Albinism	-0.0169 (0.1484)	0.0273 (0.1484)	-0.0815 (0.1441)
Weight	-0.0193 (0.0130)	-0.0209 (0.0132)	-0.0166 (0.0129)

Table 2.4: The estimates for the different parameters in the zero-inflated Cox's model, in the parametric Logistic-Weibull model and in the standard left-censored Cox's model. Standard errors are given in brackets. * means significant at $p < 0.1$, ** significant at $p < 0.05$ and *** significant at $p < 0.01$.

interaction with albinism. Therefore we see that a female albino mouse has a lower probability of non-sleep than a male albino mouse. The other covariates do not have a significant effect in the logistic part of both models. For the hazard part of both zero-inflated models, we see that none of the covariates has a significant influence on the hazard. This means that as soon as the mice are sleeping, gender, albinism and weight don't have a significant effect on the duration of the sleep time. In Table 2.4, we also see that the estimates for the different parameters are almost the same in the zero-inflated Cox's model and in the parametric Logistic-Weibull model. Initially, we would not expect this because, as stated before, we assumed that the baseline cumulative hazard is zero before the first uncensored observation in the zero-inflated Cox's model, which is not the case for the parametric Logistic-Weibull model.
In the standard left-censored Cox's model, we notice that only the covariate albinism has a significant influence on the hazard. It indicates that an albino mouse has a higher hazard than a non-albino mouse, which means a shorter sleep time.

In Figure 2.4, we plotted the estimated baseline cumulative hazard functions for the three models. We also zoom in on the behavior of the graphs in the beginning of the time axis. The estimated baseline cumulative hazard functions in

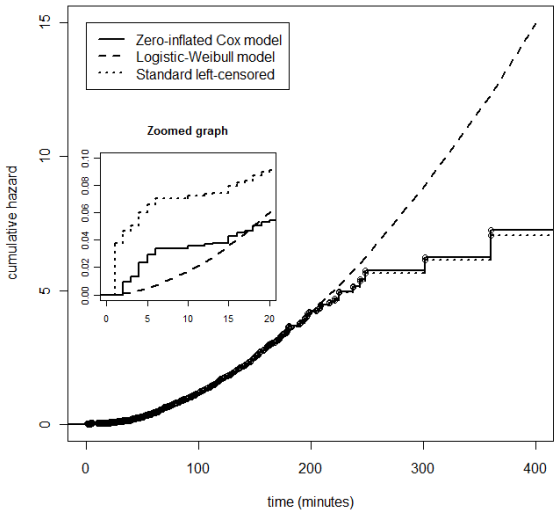


Figure 2.4: Estimates of the baseline cumulative hazard function.

the zero-inflated left-censored Cox's model and in the standard left-censored Cox's model are close to each other. However, the relatively large jump at the detection limit of the baseline cumulative hazard function in the standard left-censored Cox's model, reveals that a zero-inflated model may be better able to describe the data. The knowledge that some mice might be resistant to the alcohol dose fortifies the choice for a zero-inflated model.

The estimates of the baseline cumulative hazard functions in the parametric and in the semi-parametric zero-inflated model are almost the same for small values of sleep time. From the data we notice that the baseline cumulative hazard in the parametric Logistic-Weibull model is almost zero for small times which explains why there is not much difference here between the zero-inflated Cox's model and the parametric Logistic-Weibull model. This also explains why the estimates for the parameters do not differ much.

2.8 Conclusion

In several studies with left-censored data, the underlying time until an event can also become zero. To accommodate for this problem and to study the influence of covariates on the response variable, we introduced a zero-inflated Cox's regression model. In this model, we assumed that the probability of having a zero response is modeled through a logistic regression. Furthermore we assumed that the hazard of the non-zero part of the response follows a Cox's regression model. We estimate the baseline cumulative hazard function by a non-parametric step function. The different parameters in our model are estimated by maximum likelihood techniques. We proved the consistency and the asymptotic normality of the maximum likelihood estimators. The simulation results showed that our model performs well, especially when there is a clear conditional probability of a zero response. Finally, we have applied the regression model on a practical data set of ethanol-induced sleep time in mice.

Chapter 3

Approximated likelihood

In this chapter, we consider a new procedure for estimating the parameters in the semi-parametric Cox's regression model for univariate zero-inflated left-censored time to event data, introduced in Chapter 2. In the simulation study of Section 2.6, we noticed that the maximization of the likelihood was rather time consuming and sometimes unstable. For example, when the probability of a zero response is low, there can be some optimization problems. In Section 3.1, we propose an approximation of the likelihood and we develop an efficient EM-algorithm to calculate the maximum approximated likelihood estimates. A similar approximation was introduced by Kim et al. (2013), in order to solve stability problems in the computation of maximum likelihood estimates for the proportional hazards model with doubly-censored data. In Section 3.2, the consistency of the maximum approximated likelihood estimators is proved. Furthermore, the accuracies of the maximum approximated likelihood estimates and the maximum likelihood estimates for finite data samples are compared in Section 3.3. In Section 3.4, the new estimation technique is applied in the analysis of the data set of ethanol-induced sleep time in mice. Section 3.5 states some conclusions about the results.

3.1 Methodology

To estimate the parameters γ and β and the baseline hazard function $\lambda(t)$ in the model, introduced in Chapter 2, we derived the empirical likelihood:

$$\begin{aligned} L^e(\gamma,\beta,\Lambda) &= \prod_{i=1}^{n} \{(1-\pi(\gamma,X_i))\Lambda\{T_i\}g(\beta,Z_i)\exp\left(-g(\beta,Z_i)\Lambda(T_i)\right)\}^{\delta_i} \\ &\quad \{\pi(\gamma,X_i)+(1-\pi(\gamma,X_i))\left(1-\exp(-g(\beta,Z_i)\Lambda(T_i))\right)\}^{1-\delta_i}, \end{aligned}$$

where $\Lambda\{t\} = \Lambda(t) - \Lambda(t^-)$. We considered the following non-parametric step function for the baseline cumulative hazard function:

$$\Lambda_n(t) = \sum_{k=1}^{q_n} \lambda_k \mathrm{I}(u_k \leq t),$$

where $0 < u_1 < \ldots < u_{q_n}$ are the unique uncensored observations and $\lambda_1, \ldots, \lambda_{q_n}$ are the corresponding step sizes at these time points.

For a given sample, we can obtain maximum likelihood estimates for the different parameters γ, β and $\lambda_1, \ldots, \lambda_{q_n}$ by maximizing the likelihood function $L^e(\gamma,\beta,\Lambda_n)$. Note that the number of parameters to be optimized is proportional to the number of uncensored observations. This number can be very high when the sample size is large. Despite the technical aspects that facilitate the maximum likelihood estimation procedure, discussed in Section 2.1, the optimization procedure is sometimes very time consuming. Moreover, in the simulation study of Section 2.6 we noticed that in some cases the optimization procedure was not able to estimate the probability of a zero response well.

To obtain a more stable and faster algorithm, we propose an approximation of the likelihood and we develop an efficient EM-algorithm to obtain estimates for the different parameters. In the approximation of the likelihood function $L^e(\gamma,\beta,\Lambda_n)$, we use a first order Taylor expansion:

$$\exp\{g(\beta,z)\lambda(t)\} - 1 \approx g(\beta,z)\lambda(t).$$

Hence, we have that

$$\begin{aligned}
&1-\exp(-g(\beta,Z_i)\Lambda_n(T_i)) \\
&\quad=\sum_{l:u_l\le T_i}[\exp\{-g(\beta,Z_i)\Lambda_n(u_{l-1})\}-\exp\{-g(\beta,Z_i)\Lambda_n(u_l)\}] \\
&\quad=\sum_{l:u_l\le T_i}[(\exp\{g(\beta,Z_i)\Lambda_n\{u_l\}\}-1)\exp\{-g(\beta,Z_i)\Lambda_n(u_l)\}] \\
&\quad\approx\sum_{l:u_l\le T_i}[g(\beta,Z_i)\Lambda_n\{u_l\}\exp\{-g(\beta,Z_i)\Lambda_n(u_l)\}]\,.
\end{aligned}$$

We plug this approximating expression into the likelihood function $L^e(\gamma,\beta,\Lambda_n)$ and obtain the approximated empirical likelihood:

$$\begin{aligned}
\mathcal{L}^A(\gamma,\beta,\Lambda_n) \;=\; &\prod_{i=1}^{n}\Big\{(1-\pi(\gamma,X_i))\Lambda_n\{T_i\}g(\beta,Z_i)\exp\{g(\beta,Z_i)\Lambda_n(T_i)\}\Big\}^{\delta_i} \\
&\Big\{\pi(\gamma,X_i)+(1-\pi(\gamma,X_i))\times \\
&\sum_{l:u_l\le T_i}[g(\beta,Z_i)\Lambda_n\{u_l\}\exp\{-g(\beta,Z_i)\Lambda_n(u_l)\}]\Big\}^{1-\delta_i}.
\end{aligned}$$

The so called maximum approximated likelihood (MAL) estimates for γ, β and $\lambda_1,\ldots,\lambda_{q_n}$ are obtained by maximizing the approximated empirical likelihood $\mathcal{L}^A(\gamma,\beta,\Lambda_n)$. We can develop an efficient Expectation-Maximization algorithm to compute the MAL estimates, which is an important advantage. We will treat the left-censored observations as unobserved missing data.

The approximated empirical likelihood implies that the conditional distribution of $Y_i|Y_i>0$; given $T_i=t_i$, $\delta_i=0$ and $Z_i=z_i$ is discrete, having mass on $\{u_k:u_k\le t_i\}$ and

$$\begin{aligned}
&P_{Y_i>0}(Y_i=u_k|T_i=t_i,\delta_i=0,Z_i=z_i,\beta,\Lambda_n) \\
&\qquad=\frac{P_{Y_i>0}(Y_i=u_k,T_i=t_i,\delta_i=0|Z_i=z_i,\beta,\Lambda_n)}{P_{Y_i>0}(T_i=t_i,\delta_i=0|Z_i=z_i,\beta,\Lambda_n)} \\
&\qquad=\frac{\zeta_i(u_k;\beta,\Lambda_n)}{\sum_{l:u_l\le t_i}\zeta_i(u_l;\beta,\Lambda_n)},
\end{aligned}\tag{3.1}$$

where $\zeta_i(u;\beta,\Lambda_n)=g(\beta,z_i)\Lambda_n\{u\}\exp\{-g(\beta,z_i)\Lambda_n(u)\}$.
Define $\tilde{\delta}_i=1$ if $Y_i>0$ and $\tilde{\delta}_i=0$ if $Y_i=0$. For uncensored observations, we know Y_i and consequently $\tilde{\delta}_i$. For left-censored observations, Y_i and $\tilde{\delta}_i$ are unobserved. The

complete empirical likelihood (i.e. empirical likelihood of $(Y_i, \tilde{\delta}_i, X_i, Z_i)$) is given by

$$\begin{aligned} L^c(\gamma, \beta, \Lambda_n) &= \prod_{i=1}^{n} \Big[(1 - \pi(\gamma, X_i)) \Lambda_n\{Y_i\} g(\beta, Z_i) \exp\{-g(\beta, Z_i)\Lambda_n(Y_i)\} \Big]^{\tilde{\delta}_i} \\ &\quad \Big[\pi(\gamma, X_i) \Big]^{1-\tilde{\delta}_i} \\ &= \prod_{i=1}^{n} \Big[1 - \pi(\gamma, X_i) \Big]^{\tilde{\delta}_i} \Big[\pi(\gamma, X_i) \Big]^{1-\tilde{\delta}_i} \\ &\quad \prod_{i=1}^{n} \prod_{k=1}^{q_n} \Big[\{\lambda_k g(\beta, Z_i)\}^{I(\tilde{\delta}_i = 1, Y_i = u_k)} \\ &\qquad \exp\{-g(\beta, Z_i)\lambda_k I(\tilde{\delta}_i = 1, Y_i \geq u_k)\} \Big]. \end{aligned}$$

The corresponding log complete empirical likelihood becomes

$$\begin{aligned} l^c(\gamma, \beta, \Lambda_n) &= \sum_{i=1}^{n} \Big[\tilde{\delta}_i \log(1 - \pi(\gamma, X_i)) + (1 - \tilde{\delta}_i) \log \pi(\gamma, X_i) \Big] \\ &\quad + \sum_{i=1}^{n} \sum_{k=1}^{q_n} \Big[I(\tilde{\delta}_i = 1, Y_i = u_k)(\log \lambda_k + \log g(\beta, Z_i)) \\ &\qquad - g(\beta, Z_i)\lambda_k I(\tilde{\delta}_i = 1, Y_i \geq u_k) \Big]. \end{aligned}$$

Define $D = \{(T_1, \delta_1, X_1, Z_1), \ldots, (T_n, \delta_n, X_n, Z_n)\}$. In the E-step, we calculate

$$\begin{aligned} l^*(\gamma, \beta, \Lambda_n) &= E(l^c(\gamma, \beta, \Lambda_n) | D, \gamma^c, \beta^c, \Lambda_n^c) \\ &= \sum_{i=1}^{n} \Big[E(\tilde{\delta}_i | D, \gamma^c, \beta^c, \Lambda_n^c) \log(1 - \pi(\gamma, X_i)) \\ &\qquad + (1 - E(\tilde{\delta}_i | D, \gamma^c, \beta^c, \Lambda_n^c)) \log \pi(\gamma, X_i) \Big] \\ &\quad + \sum_{i=1}^{n} \sum_{k=1}^{q_n} \Big[E(I(\tilde{\delta}_i = 1, Y_i = u_k) | D, \gamma^c, \beta^c, \Lambda_n^c)(\log \lambda_k + \log g(\beta, Z_i)) \\ &\qquad - g(\beta, Z_i)\lambda_k E(I(\tilde{\delta}_i = 1, Y_i \geq u_k) | D, \gamma^c, \beta^c, \Lambda_n^c) \Big], \end{aligned}$$

where $E(\cdot | D, \gamma^c, \beta^c, \Lambda_n^c)$ is the conditional expectation of $\tilde{\delta}_i$, $I(\tilde{\delta}_i = 1, Y_i = u_k)$ and $I(\tilde{\delta}_i = 1, Y_i \geq u_k)$ given the data and the current parameter values $(\gamma^c, \beta^c, \Lambda_n^c)$. For $\delta_i = 1$, we know $\tilde{\delta}_i$, $I(\tilde{\delta}_i = 1, Y_i = u_k)$ and $I(\tilde{\delta}_i = 1, Y_i \geq u_k)$, so no expectation is needed. For $\delta_i = 0$, we can calculate all expected values using the conditional

distribution in (3.1). We have

$$\begin{aligned} m_i &= E(\tilde{\delta}_i|D,\gamma^c,\beta^c,\Lambda_n^c) \\ &= \frac{(1-\pi(\gamma^c,X_i))\sum\limits_{l:u_l\le T_i} p_{il}}{\pi(\gamma^c,X_i)+(1-\pi(\gamma^c,X_i))\sum\limits_{l:u_l\le T_i} p_{il}}, \end{aligned}$$

where $p_{il} = g(\beta^c, Z_i)\lambda_l^c \exp\{-g(\beta^c, Z_i)\Lambda_n^c(u_l)\}$. In a similar way, we have

$$\begin{aligned} e_{ik} &= E(I(\tilde{\delta}_i = 1, Y_i = u_k)|D,\gamma^c,\beta^c,\Lambda_n^c) \\ &= \frac{(1-\pi(\gamma^c,X_i))p_{ik}}{\pi(\gamma^c,X_i)+(1-\pi(\gamma^c,X_i))\sum\limits_{l:u_l\le T_i} p_{il}} \end{aligned} \tag{3.2}$$

for $u_k \le T_i$ and $e_{ik} = 0$ for $u_k > T_i$. Based on these expected values, we can also compute

$$\begin{aligned} r_{ik} &= E(I(\tilde{\delta}_i = 1, Y_i \ge u_k)|D,\gamma^c,\beta^c,\Lambda_n^c) \\ &= \sum_{l=k}^{q_n} e_{il} \end{aligned} \tag{3.3}$$

for $u_k \le T_i$ and $r_{ik} = 0$ for $u_k > T_i$.
For notation, we let

$$\begin{aligned} m_i &= \tilde{\delta}_i = 1, \\ e_{ik} &= I(\tilde{\delta}_i = 1, Y_i = u_k) = I(Y_i = u_k), \\ r_{ik} &= I(\tilde{\delta}_i = 1, Y_i \ge u_k) = I(Y_i \ge u_k), \end{aligned}$$

for $\delta_i = 1$.
We conclude with the result of the E-step:

$$\begin{aligned} l^*(\gamma,\beta,\Lambda_n) &= \sum_{i=1}^{n}[m_i\log(1-\pi(\gamma,X_i)) + (1-m_i)\log\pi(\gamma,X_i)] \\ &\quad \sum_{i=1}^{n}\sum_{k=1}^{q_n}[e_{ik}(\log\lambda_k + \log g(\beta,Z_i)) - r_{ik}g(\beta,Z_i)\lambda_k]. \end{aligned} \tag{3.4}$$

In the M-step, the parameter values are updated with $(\hat{\gamma},\hat{\beta},\hat{\Lambda}_n)$, which maximize $l^*(\gamma,\beta,\Lambda_n)$. From (3.4), we notice that the first part of $l^*(\gamma,\beta,\Lambda_n)$ is only concerned with the parametric model of the zero-probability, while the second part is only concerned with the parameters from the hazard part of the model. This yields that

the estimates of the logistic parameters and the hazard parameters can be found separately from each other.

$\hat{\gamma}$ maximizes

$$\sum_{i=1}^{n} \left[m_i \log(1 - \pi(\gamma, X_i)) + (1 - m_i) \log \pi(\gamma, X_i) \right]. \tag{3.5}$$

This optimization can be conducted using a numerical optimizer.

For the hazard part, we have that

$$\hat{\lambda}_k = \frac{\sum_{i=1}^{n} e_{ik}}{\sum_{i=1}^{n} r_{ik} g(\hat{\beta}, Z_i)}$$

and $\hat{\beta}$ is the solution of

$$\sum_{k=1}^{q_n} \sum_{i=1}^{n} e_{ik} \left[\frac{\frac{\partial}{\partial \beta} g(\beta, Z_i)}{g(\beta, Z_i)} - \bar{z}_k(\beta) \right] = 0, \tag{3.6}$$

where

$$\bar{z}_k(\beta) = \frac{\sum_{j=1}^{n} r_{jk} \frac{\partial}{\partial \beta} g(\beta, Z_i)}{\sum_{j=1}^{n} r_{jk} g(\beta, Z_j)}$$

is a weighted average. The equation in (3.6) is similar to the score equation of the partial likelihood for the Cox's model with right-censored observations.
Consider the case where $g(\beta, Z_i) = \exp(\beta^\top Z_i)$. The left-hand side of (3.6) can also be written as

$$\sum_{i=1}^{n} \sum_{k=1}^{q_n} e_{ik} \left[Z_i - \frac{\sum_{j=1}^{n} \sum_{l=1}^{q_n} e_{jl} I(u_l \geq u_k) \exp(\beta^\top Z_j) Z_j}{\sum_{j=1}^{n} \sum_{l=1}^{q_n} e_{jl} I(u_l \geq u_k) \exp(\beta^\top Z_j)} \right]. \tag{3.7}$$

We use standard functions in the statistical software package **R** to compute estimates for β and Λ. In order to use the **coxph**-function in **R**, we propose to use a data duplication method. The data set has to be constructed in the following way:

Row	Time	Indic	Z	Weights
1	u_1	1	Z_1	e_{11}
		$\vdots$		
n	u_1	1	Z_n	e_{n1}
$n+1$	u_2	1	Z_1	e_{12}
		$\vdots$		
$2n$	u_2	1	Z_n	e_{n2}
		$\vdots$		
$(q_n-1)n+1$	u_{q_n}	1	Z_1	e_{1q_n}
		$\vdots$		
$q_n n$	u_{q_n}	1	Z_n	e_{nq_n}

Table 3.1: Data duplication

Each individual i appears q_n times in this constructed data set, with (uncensored) time to events $u_1, \ldots, u_{q_n}$ and weights $e_{i1}, \ldots, e_{iq_n}$. We fit a weighted Cox's proportional hazards model with time variable Time, censoring indicator Indic and covariates Z. The weights are in the Weights-column. Ties have to be handled with the Breslow approximation method. The score equation of the partial likelihood of these observations corresponds to equation (3.7).

Some remarks:

1. In the most extreme case, all censored observations are smaller than the smallest uncensored observation. In this situation, $m_i = 0$, $e_{ik} = 0$, $r_{ik} = 0$ for $\delta_i = 0$ and $m_i = 1$, $e_{ik} = I(Y_i = u_k)$, $r_{ik} = I(Y_i \geq u_k)$ for $\delta_i = 1$. Expression (3.5) becomes the log-likelihood for fitting a logistic regression model on the censoring indicator random variables. Equation (3.6) collapses with the score equation of the partial likelihood for a Cox's regression model based on the subgroup of the uncensored observations. We conclude that the EM-algorithm simplifies to fitting a logistic regression model on the censoring indicator random variables and fitting a Cox's regression model on the uncensored observations. In the second remark of Section 2.1, we mentioned that the maximum likelihood estimates can be found similarly. The ML estimators and the MAL estimators are the same in this case. It occurs for example in studies with a fixed detection limit.

2. The covariance matrix of the MAL estimators can be estimated by the observed

information matrix with the Louis formula. Define $G(\theta) = \frac{\partial l^c(\theta)}{\partial \theta}$ and $H(\theta) = \frac{\partial^2 l^c(\theta)}{\partial \theta^2}$. The covariance matrix of the MAL estimators $\hat{\theta} = (\hat{\gamma}, \hat{\beta}, \hat{\Lambda}_n)$ can be estimated by $\frac{V^{-1}}{n}$, where

$$V = -E[H(\hat{\theta})|D, \hat{\theta}] - E[G(\hat{\theta})G(\hat{\theta})^\top|D, \hat{\theta}].$$

The calculation of $E[G(\hat{\theta})G(\hat{\theta})^\top|D, \hat{\theta}]$ requires the calculation of some extra conditional expectations, especially

$$E[\tilde{\delta}_i I(\tilde{\delta}_i = 1, Y_i = u_k)|D, \hat{\theta}] = E(I(\tilde{\delta}_i = 1, Y_i = u_k)|D, \hat{\theta}),$$

$$E[\tilde{\delta}_i I(\tilde{\delta}_i = 1, Y_i \geq u_k)|D, \hat{\theta}] = E(I(\tilde{\delta}_i = 1, Y_i \geq u_k)|D, \hat{\theta}),$$

$$E[I(\tilde{\delta}_i = 1, Y_i = u_k) I(\tilde{\delta}_i = 1, Y_i \geq u_l)|D, \hat{\theta}] = E(I(\tilde{\delta}_i = 1, Y_i = u_k)|D, \hat{\theta}) I(k \geq l),$$

$$E[I(\tilde{\delta}_i = 1, Y_i \geq u_k) I(\tilde{\delta}_i = 1, Y_i \geq u_l)|D, \hat{\theta}] = E(I(\tilde{\delta}_i = 1, Y_i \geq u_h)|D, \hat{\theta}),$$

with $h = \max\{l, k\}$. All these conditional expectations are in (3.2) and (3.3).

Note that the Louis formula is rather difficult to implement. We can also use a bootstrap approach to obtain estimates of the variances of the MAL estimators. In the bootstrap method, We construct a large number of resamples of the observed data set and of equal size to the observed data set. The resamples are obtained by random sampling with replacement from the original data set. For each of the resamples, one obtains MAL estimates of the model parameters and one computes their sample variances. In Section 3.3, we will show the performance of the bootstrap variance estimation method. The bootstrap approach is easy to program, but computer-intensive and rather time consuming.

3.2 Consistency and asymptotic normality

In this section we state and prove the consistency of the maximum approximated likelihood estimators. The assumptions that we impose for the consistency result are the same as in Section 2.2 of Chapter 2.
Let $(\gamma_0, \beta_0, \Lambda_0)$ be the true values of the parameters.

Theorem 5. *Under assumptions $A_1 - A_5$, given in Section 2.2 of Chapter 2, the maximum approximated likelihood estimators $(\hat{\gamma}, \hat{\beta}, \hat{\Lambda}_n)$ are consistent. This means*

that

$$|\hat{\gamma} - \gamma_0| \to 0,\ |\hat{\beta} - \beta_0| \to 0 \ \textit{and} \ \sup_{t\in[\rho,\tau]} |\hat{\Lambda}_n(t) - \Lambda_0(t)| \to 0$$

with probability 1.

Proof. To prove the consistency of the maximum approximated likelihood estimators, we follow the ideas stated in Chapter 2 and in Kim et al. (2013).

Step 1: Since Θ_γ and Θ_β are compact sets, the MAL estimators of γ and β are bounded by some constants M_γ and M_β. We first prove that there exists a constant M_Λ such that $\hat{\Lambda}_n(\tau) \leq M_\Lambda$ for all sufficiently large n with probability 1. That is, the maximum approximated likelihood estimator for Λ is bounded with probability 1. We have that, for any γ, β and Λ

$$\begin{aligned}
&\frac{1}{n}\left(l^A(\hat{\gamma},\hat{\beta},\hat{\Lambda}_n) - l^A(\gamma,\beta,\Lambda)\right)\\
&= \frac{1}{n}\sum_{i=1}^{n}\delta_i\Big\{\log\frac{1-\pi(\hat{\gamma},X_i)}{1-\pi(\gamma,X_i)} + \log\frac{\hat{\Lambda}_n\{T_i\}}{\Lambda\{T_i\}} + \log\frac{g(\hat{\beta},Z_i)}{g(\beta,Z_i)}\\
&\qquad -g(\hat{\beta},Z_i)\hat{\Lambda}_n(T_i) + g(\beta,Z_i)\Lambda(T_i)\Big\}\\
&+\frac{1}{n}\sum_{i=1}^{n}(1-\delta_i)\\
&\qquad \log\frac{\pi(\hat{\gamma},X_i) + (1-\pi(\hat{\gamma},X_i))\sum_{l:u_l\leq T_i}\left[g(\beta,Z_i)\hat{\lambda}_l\exp\{-g(\beta,Z_i)\hat{\Lambda}_n(u_l)\}\right]}{\pi(\gamma,X_i) + (1-\pi(\gamma,X_i))\sum_{l:u_l\leq T_i}[g(\beta,Z_i)\Lambda\{u_l\}\exp\{-g(\beta,Z_i)\Lambda(u_l)\}]}.
\end{aligned}$$

Since $g(\beta,z)$ is a continuous function over the compact sets of β and z, this function is bounded from below and above by positive constants K_1 and K_2. So there exists a positive constant K such that

$$\frac{1}{n}\sum_{i=1}^{n}\delta_i\log\frac{g(\hat{\beta},Z_i)}{g(\beta,Z_i)} \leq K\frac{1}{n}\sum_{i=1}^{n}\delta_i \to KE(\delta),$$

as $n\to+\infty$, by the Glivenko-Cantelli theorem. Since also $\pi(\gamma,x)$ is a continuous function over the compact sets of γ and x, we get in a similar way that there exists a constant K' such that

$$\frac{1}{n}\sum_{i=1}^{n}\delta_i\log\frac{1-\pi(\hat{\gamma},X_i)}{1-\pi(\gamma,X_i)} \leq K'\frac{1}{n}\sum_{i=1}^{n}\delta_i \to K'E(\delta).$$

Since $\sum_{l:u_l \leq T_i} \left[g(\hat{\beta}, Z_i)\hat{\lambda}_l \exp\{-g(\hat{\beta}, Z_i)\hat{\Lambda}_n(u_l)\} \right] \leq 1 - \exp[-g(\hat{\beta}, Z_i)\hat{\Lambda}_n(T_i)] \leq 1$, we have that

$$\frac{1}{n}\sum_{i=1}^{n}(1-\delta_i)\log\frac{\pi(\hat{\gamma},X_i)+(1-\pi(\hat{\gamma},X_i))\sum_{l:u_l\leq T_i}\left[g(\hat{\beta},Z_i)\hat{\lambda}_l\exp\{-g(\hat{\beta},Z_i)\hat{\Lambda}_n(u_l)\}\right]}{\pi(\gamma,X_i)+(1-\pi(\gamma,X_i))\sum_{l:u_l\leq T_i}\left[g(\beta,Z_i)\Lambda\{u_l\}\exp\{-g(\beta,Z_i)\Lambda(u_l)\}\right]}$$

$$\begin{aligned}
&\leq \frac{1}{n}\sum_{i=1}^{n}(1-\delta_i)\log\frac{\pi(\hat{\gamma},X_i)+(1-\pi(\hat{\gamma},X_i))(1-\exp[-g(\hat{\beta},Z_i)\hat{\Lambda}_n(T_i)])}{\pi(\gamma,X_i)}\\
&\leq \frac{1}{n}\sum_{i=1}^{n}(1-\delta_i)\log\frac{1}{\pi(\gamma,X_i)}\\
&\leq K''\frac{1}{n}\sum_{i=1}^{n}(1-\delta_i)\to K''E(1-\delta),
\end{aligned}$$

for some constant K''.
If we take for $\Lambda(t)$, $\bar{\Lambda}_n(t) = \sum_{u_k \leq t} \frac{1}{n}$, we have that

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n}\delta_i\log\frac{\hat{\Lambda}_n\{T_i\}}{\bar{\Lambda}_n\{T_i\}} &= \frac{1}{n}\sum_{k=1}^{q_n}\log(n\hat{\lambda}_k)\\
&\leq \frac{q_n}{n}\log\left[\frac{1}{q_n}\sum_{k=1}^{q_n}n\hat{\lambda}_k\right] = \frac{q_n}{n}\log\hat{\Lambda}_n(\tau)+\frac{q_n}{n}\log\left(\frac{n}{q_n}\right)\\
&\leq \log\hat{\Lambda}_n(\tau)+O(1),
\end{aligned}$$

by an application of Jensen's inequality.
Furthermore we note that

$$\frac{1}{n}\sum_{i=1}^{n}\delta_i g(\beta,Z_i)\bar{\Lambda}_n(T_i) \leq \frac{1}{n}\sum_{i=1}^{n}\delta_i g(\beta,Z_i) \leq \frac{1}{n}\sum_{i=1}^{n}g(\beta,Z_i)\to E[g(\beta,Z)],$$

for all β. We also find that

$$\begin{aligned}
-\frac{1}{n}\sum_{i=1}^{n}\delta_i g(\hat{\beta},Z_i)\hat{\Lambda}_n(T_i) &\leq -K_1\frac{1}{n}\sum_{k=1}^{q_n}\hat{\lambda}_k\sum_{i=1}^{n}\delta_i I(T_i\geq u_k)\\
&\leq K_1\hat{\Lambda}_n(\tau)\sup_{u\in[\rho,\tau]}\left|\frac{1}{n}\sum_{i=1}^{n}\delta_i I(T_i\geq u)-E[\delta I(T\geq u)]\right|\\
&\quad -K_1\min_{u\in[\rho,\tau]}E[\delta I(T\geq u)]\hat{\Lambda}_n(\tau).
\end{aligned}$$

By Glivenko-Cantelli, the first term in this last expression converges to 0 when n increases. We see that for sufficiently large n,

$$\frac{1}{n}\left(l^A(\hat{\gamma},\hat{\beta},\hat{\Lambda}_n) - l^A(\gamma,\beta,\bar{\Lambda}_n)\right) \leq \log\hat{\Lambda}_n(\tau) - K_1 E[\delta I(T\geq\tau)]\hat{\Lambda}_n(\tau) + O(1).$$

Hence, if $\hat{\Lambda}_n(\tau)\to+\infty$, the right hand side of the inequality diverges to $-\infty$. But this would contradict with $\frac{1}{n}\left(l^A(\hat{\gamma},\hat{\beta},\hat{\Lambda}_n) - l^A(\gamma,\beta,\bar{\Lambda}_n)\right)\geq 0$. Therefore, this means that $\hat{\Lambda}_n(\tau)$ should be bounded for sufficiently large n, with probability 1.

Step 2: Let $\theta=(\gamma,\beta,\Lambda)$ and $\hat{\theta}=(\hat{\gamma},\hat{\beta},\hat{\Lambda}_n)$.

Note that from the score equation

$$\frac{\partial}{\partial\lambda_r}\log L^A(\gamma,\beta,\Lambda_n)=0,$$

we find an expression for the different step sizes in the baseline cumulative hazard function:

$$\hat{\lambda}_r = \frac{1}{n\bar{\xi}_n(u_r;\hat{\theta})},$$

where

$$\bar{\xi}_n(t;\theta) = \frac{1}{n}\sum_{i=1}^{n} g(\beta,Z_i)\{I(T_i\geq t,\delta_i=1) + \bar{a}_i(t;\theta)I(T_i\geq t,\delta_i=0)\},$$

with

$$\bar{a}_i(t;\theta) = -\frac{(1-\pi(\gamma,X_i))\left[\sum_{l:t\leq u_l\leq T_i}\zeta_i(u_l;\theta) - \exp\{-g(\beta,Z_i)\Lambda(t)\}\right]}{\pi(\gamma,X_i) + (1-\pi(\gamma,X_i))\sum_{l:u_l\leq T_i}\zeta_i(u_l;\theta)}$$

and

$$\zeta_i(t;\theta) = g(\beta,Z_i)\Lambda\{t\}\exp\{-g(\beta,Z_i)\Lambda(u)\}.$$

We show that there exists $M_\xi>0$ such that $\inf_{t\in[\rho,\tau]}\bar{\xi}_n(t;\hat{\theta})>M_\xi$ for all sufficiently large n, with probability 1.

Suppose that such an M_ξ does not exist. Let ϵ be fixed, whose value will be specified later. Then, we can find a sufficiently large n and $t^*\in[\rho,\tau]$ such that $\bar{\xi}_n(t^*;\hat{\theta})<\epsilon$. By assumption A_5, there exists $\kappa>0$ such that T has a bounded density on either $(t-\kappa,t]$ or $(t,t+\kappa]$ for any $t\in[\rho,\tau]$. We will show that when T has a bounded density on $(t^*-\kappa,t^*]$, then $\hat{\Lambda}_n(t^*)-\hat{\Lambda}_n(t^*-\kappa)>M_\Lambda$ as $n\to\infty$, which is a contradiction

(with step 1). If T has a bounded density on $(t^*, t^*+\kappa]$, a similar contradiction can be derived. The proof for this case is omitted.

From assumption A_3 and the fact that $\sum_{l:u_l \le T_i} \zeta_i(u_l;\theta) \le 1 - \exp[-g(\hat{\beta}, Z_i)\Lambda(T_i)] \le 1$, there exists a constant $M_a > 0$ such that

$$\sup_{i,t} |\bar{a}_i(t;\theta)| \le \sup_i \frac{1-\pi(\gamma, X_i)}{\pi(\gamma, X_i)} \le M_a$$

with probability 1.

Let $\nu = \sup_{z, |\beta| < M_\beta} g(\beta, z)(1 + M_a)$. For any $t \in (t^* - \kappa, t^*]$, we have

$$\begin{aligned} \bar{\xi}_n(t;\hat{\theta}) &\le |\bar{\xi}_n(t;\hat{\theta}) - \bar{\xi}_n(t^*;\hat{\theta})| + \bar{\xi}_n(t^*;\hat{\theta}) \\ &\le \frac{\nu}{n}\sum_{i=1}^{n} I(t \le T_i < t^*) + \epsilon. \end{aligned}$$

Since T has a bounded density on $(t^* - \kappa, t^*]$, the Glivenko-Cantelli lemma implies that there exists a constant $\phi > 0$ such that

$$\frac{1}{n}\sum_{i=1}^{n} I(t < T_i \le t^*) \le \phi(t^* - t)$$

for all sufficiently large n.

Hence, we have that

$$\begin{aligned} \hat{\Lambda}_n(t^*) - \hat{\Lambda}_n(t^* - \kappa) &= \frac{1}{n}\sum_{t^*-\kappa < u_r \le t^*} \frac{1}{\xi_n(u_r, \hat{\theta})} \ge \frac{1}{n}\sum_{t^*-\kappa < u_r \le t^*} \frac{1}{\nu\phi(t^* - u_r) + \epsilon} \\ &\to \int_{X,Z}\int_{t^*-\kappa}^{t^*} \frac{(1-\pi(\gamma_0, x)) f_{Y>0}(t|z) F_{C|X,Z}(t|x,z)}{\nu\phi(t^* - t) + \epsilon} dt P(dxdz) \\ &\ge \frac{\zeta}{\nu\phi}\left[-\int_{t^*-\kappa}^{t^*} \frac{1}{\nu\phi(t^* - t) + \epsilon} d(\nu\phi(t^* - t) + \epsilon)\right] \\ &= \frac{\zeta}{\nu\phi}[\log(\nu\phi\kappa + \epsilon) - \log\epsilon] = \frac{\zeta}{\nu\phi}\log\left[\frac{\nu\phi\kappa}{\epsilon} + 1\right], \end{aligned} \tag{3.8}$$

with

$$\zeta = \inf_{t \in [\rho, \tau]} E_{x,z}\left\{(1 - \pi(\gamma_0, x)) g(\beta_0, z)\lambda_0(t)\exp\{-g(\beta_0, z)\Lambda_0(t)\} F_{C|X,Z}(t|x,z)\right\}.$$

In the expression for ζ, $F_{C|X,Z}(t|x,z) = P(C \le t | X = x, Z = z)$. $E_{x,z}$ means that the expectation is taken over X and Z. Due assumption A_5, we know that $\zeta > 0$. Finally,

choose ϵ sufficiently small, so that (3.8) is larger than M_Λ, which is a contradiction with the boundedness of $\hat{\Lambda}_n(\tau)$.

Step 3: Define

$$\tilde{\lambda}_r = \frac{1}{n\xi_n(u_r;\theta_0)}$$

where

$$\xi_n(t;\theta) = \frac{1}{n}\sum_{i=1}^{n} g(\beta, Z_i)\{I(T_i \geq t, \delta_i = 1) + a_i(\theta)I(T_i \geq t, \delta_i = 0)\}$$

with

$$a_i(\theta) = -\frac{(1-\pi(\gamma, X_i))\exp\{-g(\beta, Z_i)\Lambda(T_i)\}}{\pi(\gamma, X_i) + (1-\pi(\gamma, X_i))(1-\exp\{-g(\beta, Z_i)\Lambda(T_i)\})}$$

and $\theta_0 = (\gamma_0, \beta_0, \Lambda_0)$. In step 3 of the proof of the consistency of the ML estimators in Section 2.4, we proved that $\tilde{\Lambda}_n(.) = \sum_{k=1}^{q_n} \tilde{\lambda}_k I(u_k \leq .)$ converges to $\Lambda_0(.)$ uniformly on $[\rho, \tau]$.

Since $\tilde{\Lambda}_n$ converges uniformly to Λ_0 on $[\rho, \tau]$, $\tilde{\Lambda}_n$ is bounded on $[\rho, \tau]$ by assumption A_5. Through similar lines as in step 2, one can prove that $\tilde{\lambda}_r = O(\frac{1}{n})$. This also implies that

$$\sup_i |1 - \exp(-g(\beta_0, Z_i)\tilde{\Lambda}_n(T_i)) - \sum_{l:u_l \leq T_i} \zeta_i(u_l;\tilde{\theta})| = O\left(\frac{1}{n}\right), \tag{3.9}$$

where $\tilde{\theta} = (\gamma_0, \beta_0, \tilde{\Lambda}_n)$. We need (3.9) in step 5.

Step 4: Suppose that $\hat{\gamma}, \hat{\beta}$ and $\hat{\Lambda}_n$ converge to γ^+, β^+ and Λ^+ uniformly. We will show that

$$\frac{d\hat{\Lambda}_n}{d\tilde{\Lambda}_n}(t) \to \frac{d\Lambda^+}{d\Lambda_0}(t) \tag{3.10}$$

uniformly on $[\rho, \tau]$, with probability 1.

In step 4 of the proof of the consistency of the ML estimators in Section 2.4, we already showed that (3.10) holds for the maximum likelihood estimator. Since $\sup_l \hat{\lambda}_l = O(\frac{1}{n})$ and $\hat{\beta}$ is bounded, we have

$$\sup_i |1 - \exp(-g(\hat{\beta}, Z_i)\hat{\Lambda}_n(T_i)) - \sum_{l:u_l \leq T_i} \zeta_i(u_l;\hat{\theta})| = O\left(\frac{1}{n}\right). \tag{3.11}$$

Furthermore

$$\sup_{t\in[\rho,\tau]} |\bar{\xi_n}(t;\hat{\theta}) - \xi_n(t;\hat{\theta})| = O\left(\frac{1}{n}\right).$$

which implies that (3.10) also holds for the MAL estimator.

Step 5: By Helly's selection theorem, we can find a convergent subsequence of $\hat{\theta}$ (Parner (1998)). With abuse of notation, we let $\hat{\theta}$ be a convergent subsequence and θ^+ be the convergence point of $\hat{\theta}$. Then, (3.9) and (3.10) imply that

$$0 \leq \frac{1}{n} l^A(\hat{\gamma}, \hat{\beta}, \hat{\Lambda}_n) - \frac{1}{n} l^A(\gamma_0, \beta_0, \tilde{\Lambda}_n) \to KL(\theta^+, \theta_0)$$

where $KL(\theta^+, \theta_0)$ is the negative of the Kullback-Leibler divergence. It is defined by

$$\begin{aligned}
& KL(\theta^+, \theta_0) \\
= \; & E[\delta\{\log(1-\pi(\gamma^+, X)) + \log g(\beta^+, Z)\} - \delta g(\beta^+, Z)\Lambda^+(T) \\
& \quad +(1-\delta)\log\{\pi(\gamma^+, X) + (1-\pi(\gamma^+, X))(1-\exp[-g(\beta^+, Z)\Lambda^+(T)])\}] \\
& -E[\delta\{\log(1-\pi(\gamma_0, X)) + \log g(\beta_0, Z)\} - \delta g(\beta_0, Z)\Lambda_0(T) \\
& \quad +(1-\delta)\log\{\pi(\gamma_0, X) + (1-\pi(\gamma_0, X))(1-\exp[-g(\beta_0, Z)\Lambda_0(T)])\}] \\
& +E\left[\delta \log \frac{\lambda^+(T)}{\lambda_0(T)}\right].
\end{aligned}$$

Since the Kullback-Leibler divergence is non-negative, $KL(\theta^+, \theta_0)$ is non-positive. So $KL(\theta^+, \theta_0) = 0$. Using the identifiability result in Theorem 1 of Chapter 2, this implies that $\theta^+ = \theta_0$, which terminates the proof of the consistency.

□

Furthermore, we give some ideas to establish the asymptotic normality of the maximum approximated likelihood estimators. Expression (3.11) in the proof of the consistency of the MAL estimators implies that, for a large number of observations, the MAL estimators do not differ much from the ML estimators. In Section 2.5, we already proved the asymptotic normality of the ML estimators. We believe that the asymptotic normality also holds for the MAL estimators. In future research, we will try to give a formal proof. Hereto we will investigate the difference between the score functions for the maximum approximated likelihood estimators and the score functions for the maximum likelihood estimators. Statistical inference on the score functions for the ML estimators may lead to an asymptotic normality result for the MAL estimators.

3.3 Simulation study

In this section, we compare the accuracies of the maximum approximated likelihood estimates and the maximum likelihood estimates for finite data samples. In Section 2.6 of Chapter 2, we considered five different simulation settings. For each simulation setting we generated 500 data sets, with $n = 100$, $n = 200$ and $n = 500$ observations. For the five different settings, we calculated an estimate of the bias and the standard deviation of the 500 ML estimates. The results were shown in Table 2.2 of Section 2.6. From the same data sets we now calculate MAL estimates by applying the proposed EM-algorithm. Estimates for the bias and the standard deviations of the 500 MAL estimates, in all different settings, are shown in Table 3.2.

For the five different settings, we see in Table 3.2 that the bias and the standard deviation of the maximum aproximated likelihood estimates decrease as the sample size increases. This is in line with what we expect based on the theoretical results. However we notice some differences between the settings. In the first setting, the bias and the standard deviation for all parameter estimates are low. This was expected since the parameters in this setting were created such that there was a clear conditional probability of a zero response value. The larger bias and standard deviation for the logistic parameters in setting 2, compared to setting 1, is due to the small rate of zero responses. The estimators for parameters in the Cox's regression model, on the other hand, produced smaller biases and standard deviations. The higher percentage of censoring among the responders in setting 3 and 4 explains the higher biases and standard deviations compared to setting 1.

By comparing the biases and standard deviations in Table 2.2 and Table 3.2, we notice that the MAL estimators are always more accurate than the ML estimators. Especially the standard deviations of the MAL estimates are less than those of the ML estimates of Chapter 2.

Furthermore, in Table 2.2 we saw that the biases and standard deviations of the ML estimates for the logistic parameters in the second and fifth setting, based on data sets with 100 observations, were rather high. This is not the case for the biases and standard deviations of the MAL estimates. The MAL estimation approach is better able to handle with data sets from which the probability of a zero response is not easy to estimate.

In the ML estimation method of Chapter 2, we optimize for many parameters at once, while in the M-step of the EM-algorithm we obtain estimates of the γ- and β-parameters separately from each other and the estimates of the step sizes of the

		Setting 1	Setting 2	Setting 3	Setting 4	Setting 5
$n = 100$	$\hat{\gamma}_0$	-0.0049 (0.5065)	-0.2841 (1.9316)	0.0393 (0.5770)	0.0105 (0.5120)	0.0521 (0.7228)
	$\hat{\gamma}_1$	0.0042 (0.0862)	0.0256 (0.2246)	0.0032 (0.1017)	0.0136 (0.0942)	0.0314 (0.1541)
	$\hat{\beta}$	0.0001 (0.0717)	-0.00005 (0.0396)	-0.0026 (0.0732)	0.0056 (0.0837)	0.0075 (0.0721)
	$\hat{\Lambda}(20)$	-0.0058 (0.1225)	-0.0036 (0.0798)	-0.0071 (0.1341)	-0.0038 (0.1061)	-0.0227 (0.2874)
	$\hat{\Lambda}(30)$	0.0148 (0.4400)	0.0018 (0.2443)	0.0016 (0.4173)	0.0067 (0.3869)	-0.0148 (0.5886)
	$\hat{\Lambda}(40)$	0.2802 (1.7808)	0.1072 (0.8627)	0.1782 (1.5701)	-0.5198 (0.6493)	-0.2446 (0.7652)
$n = 200$	$\hat{\gamma}_0$	0.0121 (0.3428)	-0.0890 (0.6180)	-0.0110 (0.4088)	-0.0079 (0.3691)	0.0447 (0.4748)
	$\hat{\gamma}_1$	0.0020 (0.0605)	0.0054 (0.0912)	0.0030 (0.0693)	0.0052 (0.0687)	0.0197 (0.0976)
	$\hat{\beta}$	-0.0023 (0.0455)	-0.0008 (0.0305)	-0.0059 (0.0487)	0.0025 (0.0494)	0.0057 (0.0455)
	$\hat{\Lambda}(20)$	-0.0049 (0.0881)	0.0021 (0.0581)	-0.0009 (0.0966)	-0.0073 (0.0698)	-0.0299 (0.1863)
	$\hat{\Lambda}(30)$	0.0229 (0.2631)	0.0059 (0.1764)	0.0184 (0.2746)	0.0114 (0.2676)	-0.0257 (0.3932)
	$\hat{\Lambda}(40)$	0.1631 (0.8975)	0.0582 (0.6272)	0.1607 (0.9622)	-0.2033 (0.6070)	-0.0734 (0.6453)
$n = 500$	$\hat{\gamma}_0$	-0.0154 (0.2092)	0.0120 (0.3462)	-0.0082 (0.2462)	0.0089 (0.2324)	0.0019 (0.2925)
	$\hat{\gamma}_1$	0.0039 (0.0359)	-0.0024 (0.0523)	0.0034 (0.0438)	0.0001 (0.0417)	0.0146 (0.0628)
	$\hat{\beta}$	-0.0017 (0.0279)	-0.0018 (0.0179)	-0.0019 (0.0296)	0.0009 (0.0302)	0.0027 (0.0272)
	$\hat{\Lambda}(20)$	0.0014 (0.0517)	-0.0013 (0.0362)	-0.0001 (0.0617)	-0.0021 (0.0433)	-0.0143 (0.1199)
	$\hat{\Lambda}(30)$	0.0165 (0.1499)	0.0085 (0.1073)	0.0036 (0.1584)	-0.0030 (0.1556)	-0.0263 (0.2330)
	$\hat{\Lambda}(40)$	0.0501 (0.4917)	0.0293 (0.3663)	0.0427 (0.5251)	-0.0366 (0.5590)	-0.0466 (0.4361)

Table 3.2: The bias and standard deviation (between brackets) of the simulated MAL estimates.

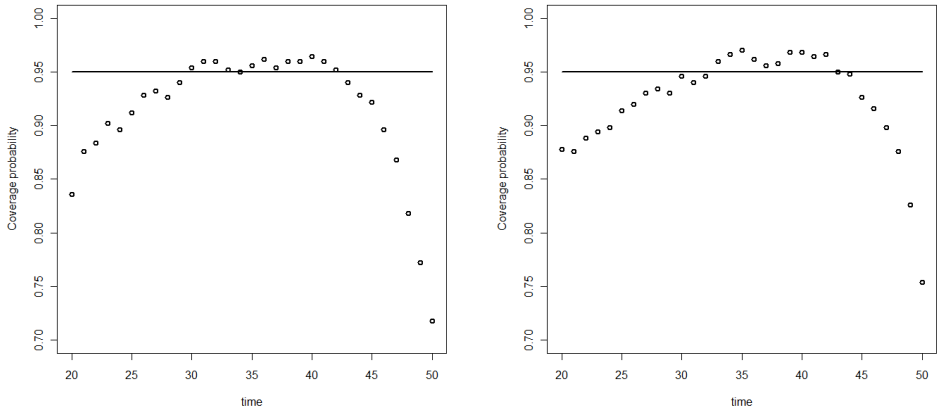
baseline cumulative hazard function can be calculated from the estimates for the β-parameters. This gives a high dimension reduction. It makes the optimization procedure more stable and it makes that the MAL estimates are obtained much faster than the ML estimates.
We conclude that the MAL estimation method is a competitive alternative to the ML estimation procedure, due to the computational stability and the speed at which the estimates are obtained.

We further investigate the performance of the bootstrap estimators for the variance of the MAL estimators for the first simulation setting. Hereto we calculate empirical coverage probabilities of the 95% confidence intervals of the regression coefficients and 95% pointwise confidence intervals of the cumulative hazard function at time points ranging from 20 to 50. The coverage probabilities are computed based on 500 data sets with resp. 100 and 200 observations. The bootstrap variance estimates for each of the 500 data sets are based on 1000 replications. For the finite dimensional parameters in the first simulation setting we show the results in Table 3.3.

	n=100	n=200
γ_0	0.940	0.950
γ_1	0.954	0.950
β	0.972	0.968

Table 3.3: Empirical coverage probabilities of the 95% confidence intervals for simulation setting 1.

The empirical coverage probabilities for the logistic regression coefficients γ_0 and γ_1 and for the hazard parameter β are rather close to the nominal level of 95%. Figure 3.1 shows empirical coverage probabilities of 95% pointwise confidence intervals of the cumulative hazard functions at several time points. The confidence intervals of the cumulative hazard functions at time points in the interval from 25 to 45 present no difficulties. For small and large time points, the empirical coverage probabilities are substantially smaller than 95%. For smaller time points, this is due to the higher percentage of censored observations. At larger time points there are not enough data to estimate the cumulative hazard function well. The estimated cumulative hazard functions remain constant at their estimates in the largest uncensored observations. However, we note that the coverage probabilities are getting closer to the nominal level as the sample size increases. In general, the bootstrap variance estimation method



(a) Setting 1 - n=100

(b) Setting 1 - n=200

Figure 3.1: Empirical coverage probabilities of 95% pointwise confidence intervals of the cumulative hazard functions.

performs reasonably well and is easy to implement in **R**.

3.4 Example: Modeling ethanol-induced anesthesia

In this section, we reconsider the study on ethanol-induced anesthesia (sleep time) in genetically-selected strains of mice, described in Section 1.2. We fit a zero-inflated regression model on the data. As in Section 2.7, we investigate the influence of the following covariates on sleep time: sex, albinism, weight at the first test session, and an interaction between sex and albinism. We apply the EM-algorithm to obtain MAL estimates for the different parameters. We use the bootstrap estimation method (based on 1000 replications) to obtain estimates of the standard errors of the MAL estimators. The results are shown in Table 3.4.

We note that in the assessment of sleep time, the recording is left-censored by a fixed detection limit at 1 min. We already mentioned that in case of a fixed detection limit, the EM-algorithm simplifies to fitting a logistic regression model on the censoring indicator random variables and fitting a Cox's regression model on the uncensored observations. In the second remark of Section 2.1, we noted that the maximum likelihood estimates can be found similarly in case of a fixed detection limit. So,

Logistic part	
Intercept	-3.8111 (0.4177)
Sex	0.6626 (0.5734)
Albinism	1.3295 (0.4503)
Sex*Albinism	-1.2629 (0.7251)
Weight	0.0531 (0.0757)
Hazard part	
Sex	0.0556 (0.0891)
Albinism	0.1026 (0.1086)
Sex*Albinism	-0.0179 (0.1498)
Weight	-0.0191 (0.0128)

Table 3.4: The MAL estimates for the different parameters in the zero-inflated Cox's model. Standard errors, obtained by the bootstrap estimation method, are given in brackets.

logically, the MAL estimates in Table 3.4 are similar to the ML estimates shown in Table 2.4 of Section 2.7. The bootstrap standard errors are only little larger than the standard errors shown in Table 2.4.

3.5 Conclusion

In this chapter, we introduced the maximum approximated likelihood estimators for the parameters in the semi-parametric Cox's regression model for zero-inflated left-censored time to event data. The estimates can be computed by an efficient EM-algorithm. In the M-step of the EM-algorithm one obtains estimates of the γ- and β-parameters separately using standard statistical software and the estimates of the step sizes can be calculated from the estimates for the β-parameters. This gives a high dimension reduction, which makes the optimization procedure more stable and faster. Our simulation results showed that the MAL estimator outperforms the ML estimator. As theoretical result, we proved the consistency of the MAL estimator.

Chapter 4

Bivariate parametric and semi-parametric regression models

The biological study on ethanol-induced sleeping time in mice, described in Section 1.2 of Chapter 1, has a repeated measurement design since mice are tested at two different times. This is an example where researchers are interested in bivariate semi-continuous time to event data. Due to technical limitations, observations for both semi-continuous outcome variables may be left-censored. In this chapter, the focus is on copula regression models for analyzing bivariate zero-inflated left-censored time to event data. In Section 4.1, we first describe mathematically the copula regression model. Next we describe a two-stage parametric and a two-stage semi-parametric estimation approach to obtain estimates of the different parameters in the model. The asymptotic variance-covariance matrix of the two-stage parametric estimators is derived by applying theory of inference functions. This matrix is estimated by a jackknife estimator. A simulation study is set up to illustrate the performance of the two-stage parametric and the two-stage semi-parametric estimation method for our regression model. The results are shown in Section 4.2. In Section 4.3, our model is illustrated on the practical data set of ethanol-induced sleep time in mice. Finally, Section 4.4 states some conclusions about the results.

4.1 Methodology

4.1.1 Model

We assume for each subject a bivariate vector of measurements (Y_1, Y_2) of times until an event. Y_1 and Y_2 are semi-continuous outcome variables which attain the value zero with a discrete probability and have non-zero positive values with a continuous distribution. We assume that conditionally on the covariate groups X and Z, the joint distribution of (Y_1, Y_2) is given by a bivariate mixture model:

$$\begin{aligned} F(y_1, y_2|X = x, Z = z) \\ &= P(Y_1 = 0, Y_2 = 0|X = x) \\ &+ \left[P(Y_2 = 0|X = x) - P(Y_1 = 0, Y_2 = 0|X = x)\right] F_{Y_1>0,Y_2=0}(y_1|Z = z) \\ &+ \left[P(Y_1 = 0|X = x) - P(Y_1 = 0, Y_2 = 0|X = x)\right] F_{Y_1=0,Y_2>0}(y_2|Z = z) \\ &+ \left[1 - P(Y_1 = 0|X = x) - P(Y_2 = 0|X = x) + P(Y_1 = 0, Y_2 = 0|X = x)\right] \\ &\quad \times F_{Y_1>0,Y_2>0}(y_1, y_2|Z = z). \end{aligned} \tag{4.1}$$

The notation X and Z for the two covariate groups is introduced to distinguish between the set of covariates which we will use in the marginal regression model for the zero responses and the set of covariates which will influence the non-zero response times. In a practical data analysis or as shown in the simulation section, a covariate can be part of both sets of covariates without any problem. In the previous expression, we denote by

$$\begin{aligned} F_{Y_1>0,Y_2=0}(y_1|Z = z) &= P(Y_1 \leq y_1|Y_1 > 0, Y_2 = 0, Z = z), \\ F_{Y_1=0,Y_2>0}(y_2|Z = z) &= P(Y_2 \leq y_2|Y_1 = 0, Y_2 > 0, Z = z), \\ F_{Y_1>0,Y_2>0}(y_1, y_2|Z = z) &= P(Y_1 \leq y_1, Y_2 \leq y_2|Y_1 > 0, Y_2 > 0, Z = z). \end{aligned}$$

We focus in this copula regression model on the marginal distribution of each measurement on the one hand and on the other hand on the association between the distributions of both measurements. In this aspect, we introduce a simplifying assumption that will allow us to set up later a two-stage estimation procedure where we first estimate the parameters in the marginal regression models under a working independence model, and afterwards in a second stage estimate the association parameters. Therefore we assume that the marginal distribution of a non-zero component does not depend on the susceptibility status of the other component, i.e.

$$\begin{aligned} F_{Y_1>0,Y_2>0}(y_1|Z = z) = F_{Y_1>0,Y_2=0}(y_1|Z = z) = F_{Y_1>0}(y_1|Z = z), \\ F_{Y_1>0,Y_2>0}(y_2|Z = z) = F_{Y_1=0,Y_2>0}(y_2|Z = z) = F_{Y_2>0}(y_2|Z = z). \end{aligned} \tag{4.2}$$

This means that the value of a non-zero response in one component, conditional on the covariates, does not depend on the status of the response of the other component. A similar assumption was made in the bivariate cure-mixture model of Chatterjee and Shih (2001) to reduce the number of terms in the model. Using assumption (4.2), we need to introduce the following part of the joint model in (4.1):

(i) The probability of a zero response for respectively measurement 1 and measurement 2: $P(Y_1 = 0|X = x)$ and $P(Y_2 = 0|X = x)$.

(ii) The distributions of Y_1 given $Y_1 > 0$ and Y_2 given $Y_2 > 0$: for the non-zero distributions $F_{Y_1>0}(y_1|Z = z)$ and $F_{Y_2>0}(y_2|Z = z)$.

(iii) The joint probability of having zero responses in both measurements: $P(Y_1 = 0, Y_2 = 0|X = x)$.

(iv) The joint distribution of Y_1, Y_2 given $Y_1 > 0, Y_2 > 0$: $F_{Y_1>0,Y_2>0}(y_1, y_2|Z = z)$.

(i) We assume a parametric regression model for the marginal probability of a zero response in each measurement. We denote these regression models with π_1 and π_2:

$$P(Y_1 = 0|X = x) = \pi_1(\gamma_1, x) \text{ and } P(Y_2 = 0|X = x) = \pi_2(\gamma_2, x).$$

(ii) For the conditional distribution of the non-zero outcome values, we consider proportional hazards models. We assume that the conditional hazard function for the first outcome has the following form:

$$\lambda_{Y_1>0}(y_1|Z = z) = \lambda_1(y_1) \exp(\beta_1^\top z),$$

where λ_1 is a baseline hazard function. The conditional distribution function is given by

$$F_{Y_1>0}(y_1|Z = z) = 1 - \exp[-\exp(\beta_1^\top z)\Lambda_1(y_1)],$$

with Λ_1 the baseline cumulative hazard function corresponding to λ_1. Similarly, we assume for the second measurement that

$$F_{Y_2>0}(y_2|Z = z) = 1 - \exp[-\exp(\beta_2^\top z)\Lambda_2(y_2)],$$

with Λ_2 a baseline cumulative hazard function.

Afterwards, a dependence structure between the two measurements is specified.

(iii) For the joint probability of having zero responses for both measurements, we model the cross ratio in a contingency table with $Y_1 = 0$ versus $Y_1 > 0$, and $Y_2 = 0$ versus $Y_2 > 0$. We assume that, conditionally on the covariates, this cross ratio is constant:

$$\psi = \frac{P(Y_1 = 0, Y_2 = 0|X = x)P(Y_1 > 0, Y_2 > 0|X = x)}{P(Y_1 = 0, Y_2 > 0|X = x)P(Y_1 > 0, Y_2 = 0|X = x)}.$$

(iv) To model the association between two non-zero responses, we use a family of copula functions. A two-dimensional copula is a function $C : [0,1]^2 \to [0,1]$ with properties

(a) $\forall u, v \in [0,1]$:

$$C(u,0) = 0 = C(0,v),\ C(u,1) = u \text{ and } C(1,v) = v;$$

(b) $\forall u_1, u_2, v_1, v_2 \in [0,1]$ with $u_1 \leq u_2$ en $v_1 \leq v_2$:

$$C(u_2, v_2) - C(u_2, v_1) - C(u_1, v_2) + C(u_1, v_1) \geq 0.$$

From this definition, we note that a two-dimensional copula function is a bivariate distribution function for which both marginal distribution functions are uniform distributions on the interval $[0,1]$. One of the major results of copula functions is given by Sklar's theorem and expresses that any bivariate distribution function can be written as a copula function which is evaluated in the marginal distribution functions. For an extended introduction into copula functions, we refer to Nelsen (2006).

In this doctoral thesis, we will consider copula families $C_\theta(u,v)$, where a parameter θ controls the strength of dependence. The copula links the univariate marginals to form a joint distribution. By assumption (4.2), we can express the joint distribution of the non-zero responses as (Sklar (1959)):

$$F_{Y_1>0,Y_2>0}(y_1, y_2|Z = z) = C_\theta(F_{Y_1>0}(y_1|Z = z), F_{Y_2>0}(y_2|Z = z)).$$

There are many parametric copula functions which could be used in this expression, like for example, the Gaussian copula family and the Plackett copula family. An overview of several copula families is found in Nelsen (2006). We consider in the simulation study and in the data example, some one-parameter families of Archimedean copula functions of which the expression is given by:

$$C_\theta(u,v) = \phi_\theta^{[-1]}\{\phi_\theta(u) + \phi_\theta(v)\} \quad \forall u, v \in [0,1],$$

where $\phi_\theta : [0,1] \to [0,\infty]$ is a continuous, strictly decreasing convex generating function such that $\phi_\theta(1) = 0$ and $\phi_\theta^{[-1]}$ is the pseudo-inverse of ϕ_θ, given by

$$\phi_\theta^{[-1]}(t) = \begin{cases} \phi_\theta^{-1}(t), & 0 \leq t \leq \phi_\theta(0), \\ 0, & \phi_\theta(0) \leq t \leq \infty. \end{cases}$$

We will consider some commonly used examples of Archimedean copula generators: the Clayton, Frank and Gumbel-Hougaard generator. They are shown in Table 4.1. Together with these generators, we also present the expression of Kendall's tau.

	$\phi_\theta(t)$	$\theta \in$	τ_θ
Clayton	$\frac{t^{-\theta}-1}{\theta}$	$[-1,\infty)\backslash\{0\}$	$\frac{\theta}{\theta+2}$
Frank	$-\ln\left(\frac{e^{-\theta t}-1}{e^{-\theta}-1}\right)$	$(-\infty,\infty)\backslash\{0\}$	$1-\frac{4}{\theta}[1-\frac{1}{\theta}\int_0^\theta \frac{t}{e^t-1}dt]$
Gumbel-Hougaard	$(-\ln t)^\theta$	$[1,\infty)$	$\frac{\theta-1}{\theta}$

Table 4.1: Some families of Archimedean copula functions with their generator and association expression for Kendall's tau.

Kendall's tau is often used as a global measure of the association. It can be expressed by a simple function of the generator ϕ_θ (Genest and MacKay (1986a,b)):

$$\tau_\theta = 1 + 4\int_0^1 \frac{\phi_\theta(t)}{\phi'_\theta(t)}dt.$$

Until this point, we only described the four different parts of the bivariate zero-inflated regression model for the outcome variables (Y_1, Y_2). Due to technical limitations, it is impossible in some studies to fully observe the outcome variables (Y_1, Y_2), for example by a detection limit. In that case, we only see an upper bound for one or both outcome variables. So, both observations may be left-censored. We assume that there exists a couple of random variables (C_1, C_2), independent of (Y_1, Y_2) conditionally on the covariate groups X and Z such that we observe $(T_1, T_2, \delta_1, \delta_2, X, Z)$, with

$$T_1 = \max(Y_1, C_1) \text{ and } \delta_1 = I(Y_1 \geq C_1),$$
$$T_2 = \max(Y_2, C_2) \text{ and } \delta_2 = I(Y_2 \geq C_2).$$

In this doctoral thesis, we limit ourselves to independent censoring, because when we would also assume an association structure between the lifetimes (Y_1, Y_2) and the censoring times (C_1, C_2) (dependent censoring), this would complicate the likelihood functions for the estimation of the parameters even more.

4.1.2 Estimation

To estimate the different parameters in our model, we make use of maximum likelihood techniques. The likelihood of a random sample $(T_{1i}, T_{2i}, \delta_{1i}, \delta_{2i}, X_i, Z_i)$; $i = 1, \ldots, n$ is given by

$$\begin{aligned}\mathcal{L} &= \prod_{i=1}^{n} \left[h^{1,1}(T_{1i}, T_{2i}|X = X_i, Z = Z_i)\right]^{\delta_{1i}\delta_{2i}} \left[h^{0,0}(T_{1i}, T_{2i}|X = X_i, Z = Z_i)\right]^{(1-\delta_{1i})(1-\delta_{2i})} \\ &\quad . \left[h^{1,0}(T_{1i}, T_{2i}|X = X_i, Z = Z_i)\right]^{\delta_{1i}(1-\delta_{2i})} \left[h^{0,1}(T_{1i}, T_{2i}|X = X_i, Z = Z_i)\right]^{(1-\delta_{1i})\delta_{2i}}\end{aligned}$$

where

$$h^{j,k}(t_1, t_2|X = x, Z = z) = \frac{\partial^2}{\partial t_1 \partial t_2} P(T_1 \leq t_1, T_2 \leq t_2, \delta_1 = j, \delta_2 = k|X = x, Z = z),$$

$j, k \in \{0, 1\}$. Based on expression (4.1) for (Y_1, Y_2), we note that the contribution of an individual is given by a mixture of one, two or four terms depending on whether none, one or both observations are left-censored. Since (Y_1, Y_2) and (C_1, C_2) are assumed to be independent, we can remove all factors concerning the distribution of (C_1, C_2). These factors don't yield any information about the unknown parameters. In the following subsections, we specify parametric forms for the baseline cumulative hazard functions $\Lambda_1(t)$ and $\Lambda_2(t)$ and afterwards we estimate them non-parametrically. Hereby we note that the maximum likelihood estimators do not have a closed form and numerical optimization becomes more difficult as the number of parameters increases. Due to the simplifying assumption (4.2) we are able to set up in the both cases a two-stage estimation procedure, similar to the one used in Shih and Louis (1995), Joe (1997) and Chatterjee and Shih (2001). In this way, we reduce the number of parameters that have to be estimated at the same time. First, we estimate the parameters of the marginal regression function under an independent working assumption and afterwards we estimate in a second stage the association parameters.

4.1.2.1 Parametric estimation

In the parametric estimation approach, we assume that $\Lambda_1(t)$ and $\Lambda_2(t)$ have known parametric forms $\Lambda_1(t, \zeta_1)$ and $\Lambda_2(t, \zeta_2)$, where ζ_1 and ζ_2 are vectors of parameters. Let $l(\alpha_1, \alpha_2, \rho) = \sum_{i=1}^{n} L(W_i, \alpha_1, \alpha_2, \rho)$ denote the bivariate log-likelihood, where $\alpha_1^\top = (\gamma_1^\top, \beta_1^\top, \zeta_1^\top)$, $\alpha_2^\top = (\gamma_2^\top, \beta_2^\top, \zeta_2^\top)$ and $\rho^\top = (\psi, \theta)$. The W_i, $i = 1, \ldots, n$ represent the observations $(T_{1i}, T_{2i}, \delta_{1i}, \delta_{2i}, X_i, Z_i)$. When we optimize this likelihood function over all parameters by looking for the solution $(\hat{\alpha}_1, \hat{\alpha}_2, \hat{\rho})$ of the set of equations,

$$(\partial l/\partial \alpha_1^\top, \partial l/\partial \alpha_2^\top, \partial l/\partial \rho^\top) = 0^\top,$$

we get the one-stage maximum likelihood estimates.

As already mentioned, we also consider a two-stage estimation procedure. At the first stage, we estimate the marginal parameters α_1 and α_2 using a univariate zero-inflated parametric Cox's regression approach that ignores the dependence of the two components. Hereto, we maximize the marginal log-likelihoods l_1 and l_2, where

$$\begin{aligned} l_j(\alpha_j) &= \sum_{i=1}^{n} L_j(W_{ji}, \alpha_j) \\ &= \sum_{i=1}^{n} \delta_{ji} \log\{[1-\pi_j(\gamma_j, X_i)]\lambda_j(T_{ji}, \zeta_j)\exp(\beta_j^\top Z_i)\exp[-\exp(\beta_j^\top Z_i)\Lambda_j(T_{ji}, \zeta_j)]\} \\ &\quad +(1-\delta_{ji})\log\{\pi_j(\gamma_j, X_i) + [1-\pi_j(\gamma_j, X_i)](1-\exp[-\exp(\beta_j^\top Z_i)\Lambda_j(T_{ji}, \zeta_j)])\}, \end{aligned}$$

with $j = 1, 2$ separately to obtain estimates $\tilde{\alpha}_1^\top = (\tilde{\gamma}_1^\top, \tilde{\beta}_1^\top, \tilde{\zeta}_1^\top)$ and $\tilde{\alpha}_2^\top = (\tilde{\gamma}_2^\top, \tilde{\beta}_2^\top, \tilde{\zeta}_2^\top)$. At the second stage, estimates of the association parameters ρ are obtained by fixing the parameters of the marginal distributions at their estimates and maximizing the bivariate log-likelihood $l(\tilde{\alpha}_1, \tilde{\alpha}_2, \rho)$ with respect to ρ. So, $(\tilde{\alpha}_1, \tilde{\alpha}_2, \tilde{\rho})$ is the solution of

$$(\partial l_1/\partial\alpha_1^\top, \partial l_2/\partial\alpha_2^\top, \partial l/\partial\rho^\top) = 0^\top. \tag{4.3}$$

For notation, we let $\eta^\top = (\alpha_1^\top, \alpha_2^\top, \rho^\top)$ be the vector of all parameters, $\hat{\eta}^\top = (\hat{\alpha}_1^\top, \hat{\alpha}_2^\top, \hat{\rho}^\top)$ be the one-stage maximum likelihood estimators and $\tilde{\eta}^\top = (\tilde{\alpha}_1^\top, \tilde{\alpha}_2^\top, \tilde{\rho}^\top)$ be the two-stage estimators. In a similar way as in Joe (2005), we can obtain a partitioned form for the asymptotic variance-covariance matrix of $\tilde{\eta}$. Hereto, we apply theory of inference functions, where the inference functions in the left-hand side of (4.3) are written as

$$\sum_{i=1}^{n} g(W_i; \eta),$$

where $g^\top = (g_1^\top, g_2^\top, g_3^\top)$, $g_j = \partial L_j/\partial\alpha_j$ for $j = 1, 2$; and $g_3 = \partial L/\partial\rho$.
Let $(\eta^0)^\top = (\alpha_1^0, \alpha_2^0, \rho^0)^\top$ be the true parameter vector. From maximum likelihood theory (for example in Lehmann (1998)), we know that $\sqrt{n}(\hat{\eta} - \eta^0)$ converges in distribution to a multivariate normal distribution with mean vector zero and variance-covariance matrix I^{-1}, where the Fisher information matrix I can be partitioned into blocks:

$$I = \begin{bmatrix} I_{11} & I_{12} & I_{13} \\ I_{21} & I_{22} & I_{23} \\ I_{31} & I_{32} & I_{33} \end{bmatrix},$$

where $I_{jk} = -E[\partial^2 L/\partial\alpha_j\partial\alpha_k^\top]$ for $1 \leq j,k \leq 2$, $I_{j3} = -E[\partial^2 L/\partial\alpha_j\partial\rho^\top]$, $I_{3j} = I_{j3}^\top$ for $j = 1,2$ and $I_{33} = -E[\partial^2 L/\partial\rho\partial\rho^\top]$.

In Theorem 6, we show the asymptotic normality of the two-stage estimators under regularity conditions $A_1 - A_4$:

A_1: There exists an open subset Θ of the parameter space containing the true parameter vector η^0, such that the third partial derivatives of $L_j(w_j,\alpha_j)$, $j = 1,2$, with respect to the components of α_j can be computed for almost all w_j and such that the third partial derivatives of $L(w,\alpha_1,\alpha_2,\rho)$, with respect to the components of α_1, α_2 and ρ can be computed for almost all w.

A_2: The first and second order partial derivatives satisfy the following equations:

$$E[\partial L_1(W_1,\alpha_1)/\partial\alpha_1] = 0,$$
$$E[\partial L_2(W_2,\alpha_2)/\partial\alpha_2] = 0,$$
$$E[\partial L(W,\alpha_1,\alpha_2,\rho)/\partial\eta] = 0,$$

and

$$E[-\partial^2 L_1(W_1,\alpha_1)/\partial\alpha_1\partial\alpha_1^\top] = E[\partial L_1(W_1,\alpha_1)/\partial\alpha_1 \cdot (\partial L_1(W_1,\alpha_1)/\partial\alpha_1)^\top],$$
$$E[-\partial^2 L_2(W_2,\alpha_2)/\partial\alpha_2\partial\alpha_2^\top] = E[\partial L_2(W_2,\alpha_2)/\partial\alpha_2 \cdot (\partial L_2(W_2,\alpha_2)/\partial\alpha_2)^\top],$$
$$E[-\partial^2 L(W,\alpha_1,\alpha_2,\rho)/\partial\eta\partial\eta^\top] = E[\partial L(W,\alpha_1,\alpha_2,\rho)/\partial\eta \cdot (\partial L(W,\alpha_1,\alpha_2,\rho)/\partial\eta)^\top].$$

Note that this condition is satisfied if the specified model allows the interchange of integrals and partial derivatives.

A_3: Let $J_{jk} = Cov(g_j,g_k) = E[g_j g_k^\top]$ for $1 \leq j,k \leq 2$, so that J_{jj} is the information matrix from the j-th marginal log-likelihood. The elements of J_{jk} and I have to be finite and J_{11}, J_{22} and I have to be positive definite for all η in Θ.

A_4: All third partial derivatives, mentioned in A_1, are bounded by functions, with a finite expected value w.r.t. the true parameter vector.

Theorem 6. *Under regularity conditions $A_1 - A_4$, $\sqrt{n}(\tilde{\eta} - \eta^0)$ converges in distribution to a multivariate normal distribution with mean vector zero and variance-covariance matrix V. The matrix V is given by*

$$V = D_g^{-1} M_g (D_g^{-1})^\top,$$

where $D_g = E[-\partial g(W;\eta)/\partial\eta^\top]$ and $M_g = Cov(g(W;\eta)) = E[gg^\top]$.

Proof. We will consider Taylor series expansions of the elements of the vectors $\sum_{i=1}^{n} g_1(W_{1i}, \alpha_1)$, $\sum_{i=1}^{n} g_2(W_{2i}, \alpha_2)$ and $\sum_{i=1}^{n} g_3(W_i, \eta)$. We introduce the following notation: $g_{1q} = \frac{\partial L_1}{\partial \alpha_{1q}}$, where α_{1q} is the q-th element of the vector α_1. Similarly, $g_{2r} = \frac{\partial L_2}{\partial \alpha_{2r}}$ and $g_{3s} = \frac{\partial L}{\partial \rho_s}$, where α_{2r} and ρ_s are the r-th and s-th element of α_2 and ρ respectively. A Taylor series expansion of $\sum_{i=1}^{n} g_{1q}(W_{1i}, \alpha_1)$ around α_1^0 and evaluated at $\tilde{\alpha}_1$ gives

$$\begin{aligned} \sum_{i=1}^{n} g_{1q}(W_{1i}, \tilde{\alpha}_1) = 0 &= \sum_{i=1}^{n} g_{1q}(W_{1i}, \alpha_1^0) - A_{q,\alpha_1}^{\top}(\alpha_1^0)(\tilde{\alpha}_1 - \alpha_1^0) \\ &\quad + \frac{1}{2}(\tilde{\alpha}_1 - \alpha_1^0)^{\top} H_{q,\alpha_1,\alpha_1}(\alpha_1^*)(\tilde{\alpha}_1 - \alpha_1^0), \end{aligned} \tag{4.4}$$

where

$$\begin{aligned} A_{q,\alpha_1}(\alpha_1) &= -\sum_{i=1}^{n} \frac{\partial g_{1q}(W_{1i}, \alpha_1)}{\partial \alpha_1} = -\sum_{i=1}^{n} \begin{bmatrix} \frac{\partial g_{1q}(W_{1i}, \alpha_1)}{\partial \alpha_{11}} \\ \frac{\partial g_{1q}(W_{1i}, \alpha_1)}{\partial \alpha_{12}} \\ \vdots \end{bmatrix}, \\ H_{q,\alpha_1,\alpha_1}(\alpha_1) &= \sum_{i=1}^{n} \frac{\partial^2 g_{1q}(W_{1i}, \alpha_1)}{\partial \alpha_1 \partial \alpha_1^{\top}} = \sum_{i=1}^{n} \begin{bmatrix} \frac{\partial^2 g_{1q}(W_{1i}, \alpha_1)}{\partial \alpha_{11}^2} & \frac{\partial^2 g_{1q}(W_{1i}, \alpha_1)}{\partial \alpha_{11} \partial \alpha_{12}} & \cdots \\ \frac{\partial^2 g_{1q}(W_{1i}, \alpha_1)}{\partial \alpha_{12} \partial \alpha_{11}} & \frac{\partial^2 g_{1q}(W_{1i}, \alpha_1)}{\partial \alpha_{12}^2} & \cdots \\ \vdots & \vdots & \vdots \end{bmatrix} \end{aligned}$$

and α_1^* is a point on the line segment connecting α_1 and α_1^0. Similarly,

$$\begin{aligned} \sum_{i=1}^{n} g_{2r}(W_{2i}, \tilde{\alpha}_2) = 0 &= \sum_{i=1}^{n} g_{2r}(W_{2i}, \alpha_2^0) - A_{r,\alpha_2}^{\top}(\alpha_2^0)(\tilde{\alpha}_2 - \alpha_2^0) \\ &\quad + \frac{1}{2}(\tilde{\alpha}_2 - \alpha_2^0)^{\top} H_{r,\alpha_2,\alpha_2}(\alpha_2^*)(\tilde{\alpha}_2 - \alpha_2^0), \end{aligned} \tag{4.5}$$

where

$$\begin{aligned} A_{r,\alpha_2}(\alpha_2) &= -\sum_{i=1}^{n} \frac{\partial g_{2r}(W_{2i}, \alpha_2)}{\partial \alpha_2}, \\ H_{r,\alpha_2,\alpha_2}(\alpha_2) &= \sum_{i=1}^{n} \frac{\partial^2 g_{2r}(W_{2i}, \alpha_2)}{\partial \alpha_2 \partial \alpha_2^{\top}} \end{aligned}$$

and α_2^* is a point on the line segment connecting α_2 and α_2^0.

We also have that

$$\begin{aligned}\sum_{i=1}^{n} g_{3s}(W_i,\tilde{\eta}) = 0 = \quad & \sum_{i=1}^{n} g_{3s}(W_i,\eta^0) - B_{s,\eta,1}^{\top}(\eta^0)(\tilde{\alpha}_1-\alpha_1^0) \\ & -B_{s,\eta,2}^{\top}(\eta^0)(\tilde{\alpha}_2-\alpha_2^0) - B_{s,\eta}^{\top}(\eta^0)(\tilde{\rho}-\rho^0) \\ & +\frac{1}{2}(\tilde{\alpha}_1-\alpha_1^0)^{\top}K_{s,\alpha_1,\alpha_1}(\eta^*)(\tilde{\alpha}_1-\alpha_1^0) \\ & +\frac{1}{2}(\tilde{\alpha}_1-\alpha_1^0)^{\top}K_{s,\alpha_1,\alpha_2}(\eta^*)(\tilde{\alpha}_2-\alpha_2^0) \\ & +\frac{1}{2}(\tilde{\alpha}_1-\alpha_1^0)^{\top}K_{s,\alpha_1,\rho}(\eta^*)(\tilde{\rho}-\rho^0) \\ & +\frac{1}{2}(\tilde{\alpha}_2-\alpha_2^0)^{\top}K_{s,\alpha_2,\alpha_1}(\eta^*)(\tilde{\alpha}_1-\alpha_1^0) \\ & +\frac{1}{2}(\tilde{\alpha}_2-\alpha_2^0)^{\top}K_{s,\alpha_2,\alpha_2}(\eta^*)(\tilde{\alpha}_2-\alpha_2^0) \\ & +\frac{1}{2}(\tilde{\alpha}_2-\alpha_2^0)^{\top}K_{s,\alpha_2,\rho}(\eta^*)(\tilde{\rho}-\rho^0) \\ & +\frac{1}{2}(\tilde{\rho}-\rho^0)^{\top}K_{s,\rho,\alpha_1}(\eta^*)(\tilde{\alpha}_1-\alpha_1^0) \\ & +\frac{1}{2}(\tilde{\rho}-\rho^0)^{\top}K_{s,\rho,\alpha_2}(\eta^*)(\tilde{\alpha}_2-\alpha_2^0) \\ & +\frac{1}{2}(\tilde{\rho}-\rho^0)^{\top}K_{s,\rho,\rho}(\eta^*)(\tilde{\rho}-\rho^0),\end{aligned} \tag{4.6}$$

where

$$B_{s,\eta,1}(\eta) = -\sum_{i=1}^{n}\frac{\partial g_{3s}(W_i,\eta)}{\partial\alpha_1},\ B_{s,\eta,2}(\eta) = -\sum_{i=1}^{n}\frac{\partial g_{3s}(W_i,\eta)}{\partial\alpha_2},$$

$$B_{s,\eta}(\eta) = -\sum_{i=1}^{n}\frac{\partial g_{3s}(W_i,\eta)}{\partial\rho}$$

and

$$K_{s,x,y}(\eta) = \quad = \quad \sum_{i=1}^{n}\frac{\partial^2 g_{3s}(W_i,\eta)}{\partial x\partial y^{\top}},$$

with x and y vectors of parameters. η^* is a point on the line segment connecting η and η^0. Rearranging the terms in (4.4), (4.5) and (4.6), we get

$$\begin{aligned}\frac{1}{\sqrt{n}}\sum_{i=1}^{n} g_{1q}(W_{1i},\alpha_1^0) &= \left[\frac{1}{n}A_{q,\alpha_1}^{\top}(\alpha_1^0) - \frac{1}{2n}(\tilde{\alpha}_1-\alpha_1^0)^{\top}H_{q,\alpha_1,\alpha_1}(\alpha_1^*)\right] \\ &\quad \sqrt{n}(\tilde{\alpha}_1-\alpha_1^0), \\ \frac{1}{\sqrt{n}}\sum_{i=1}^{n} g_{2r}(W_{2i},\alpha_2^0) &= \left[\frac{1}{n}A_{r,\alpha_2}^{\top}(\alpha_2^0) - \frac{1}{2n}(\tilde{\alpha}_2-\alpha_2^0)^{\top}H_{r,\alpha_2,\alpha_2}(\alpha_2^*)\right] \\ &\quad \sqrt{n}(\tilde{\alpha}_2-\alpha_2^0)\end{aligned}$$

and

$$\begin{aligned}\frac{1}{\sqrt{n}}\sum_{i=1}^{n} g_{3s}(W_i,\eta^0) &= \Bigg[\frac{1}{n}B_{s,\eta,1}^{\top}(\eta^0)-\frac{1}{2n}(\tilde{\alpha}_1-\alpha_1^0)^{\top}K_{s,\alpha_1,\alpha_1}(\eta^*)\\ &\quad -\frac{1}{2n}(\tilde{\alpha}_2-\alpha_2^0)^{\top}K_{s,\alpha_2,\alpha_1}(\eta^*)\\ &\quad -\frac{1}{2n}(\tilde{\rho}-\rho^0)^{\top}K_{s,\rho,\alpha_1}(\eta^*)\Bigg]\sqrt{n}(\tilde{\alpha}_1-\alpha_1^0)\\ &\quad +\Bigg[\frac{1}{n}B_{s,\eta,2}^{\top}(\eta^0)-\frac{1}{2n}(\tilde{\alpha}_1-\alpha_1^0)^{\top}K_{s,\alpha_1,\alpha_2}(\eta^*)\\ &\quad -\frac{1}{2n}(\tilde{\alpha}_2-\alpha_2^0)^{\top}K_{s,\alpha_2,\alpha_2}(\eta^*)\\ &\quad -\frac{1}{2n}(\tilde{\rho}-\rho^0)^{\top}K_{s,\rho,\alpha_2}(\eta^*)\Bigg]\sqrt{n}(\tilde{\alpha}_2-\alpha_2^0)\\ &\quad +\Bigg[\frac{1}{n}B_{s,\eta}^{\top}(\eta^0)-\frac{1}{2n}(\tilde{\alpha}_1-\alpha_1^0)^{\top}K_{s,\alpha_1,\rho}(\eta^*)\\ &\quad -\frac{1}{2n}(\tilde{\alpha}_2-\alpha_2^0)^{\top}K_{s,\alpha_2,\rho}(\eta^*)\\ &\quad -\frac{1}{2n}(\tilde{\rho}-\rho^0)^{\top}K_{s,\rho,\rho}(\eta^*)\Bigg]\sqrt{n}(\tilde{\rho}-\rho^0).\end{aligned}$$

Regularity condition A_4 ensures that the third derivatives of the marginal and bivariate log-likelihoods in the error terms are bounded in probability. In the assumptions, we introduced the following notation: $J_{jk} = Cov(g_j, g_k) = E[g_j g_k^{\top}]$ for $1 \leq j, k \leq 2$. By the law of large numbers, as $n \to +\infty$,

$$\frac{A_{\alpha_1}^{\top}(\alpha_1^0)}{n}, \frac{A_{\alpha_2}^{\top}(\alpha_2^0)}{n}, \frac{B_{\eta,1}^{\top}(\eta^0)}{n}, \frac{B_{\eta,2}^{\top}(\eta^0)}{n}, \frac{B_{\eta}^{\top}(\eta^0)}{n}$$

converge to J_{11}, J_{22}, I_{31}, I_{32} and I_{33} respectively. Remark that $A_{\alpha_1}^{\top}(\alpha_1^0)$ is the matrix with rows given by the vectors $A_{q,\alpha_1}^{\top}(\alpha_1^0)$. A similar statement holds for the other matrices. Hence, we have that

$$\left(\sum_{i=1}^{n} g_1(W_{1i},\alpha_1^0), \sum_{i=1}^{n} g_2(W_{2i},\alpha_2^0), \sum_{i=1}^{n} g_3(W_i,\eta^0)\right)^{\top}/\sqrt{n} \tag{4.7}$$

is approximately equivalent to $\sqrt{n}D_g(\tilde{\eta}-\eta^0)$, where

$$D_g = E[-\partial g(W;\eta)/\partial\eta^{\top}] = \begin{bmatrix} J_{11} & 0 & 0 \\ 0 & J_{22} & 0 \\ I_{31} & I_{32} & I_{33}\end{bmatrix}.$$

One can show that $Cov(g_j, g_3) = E[g_j g_3^{\top}] = 0$ for $j = 1, 2$. By the central limit

theorem,

$$\left(\sum_{i=1}^{n} g_1(W_{1i}, \alpha_1^0), \sum_{i=1}^{n} g_2(W_{2i}, \alpha_2^0), \sum_{i=1}^{n} g_3(W_i, \eta^0)\right)^\top / \sqrt{n}$$

converges to multivariate normal distribution with mean vector zero and variance-covariance matrix M_g, where

$$M_g = E[gg^\top] = \begin{bmatrix} J_{11} & J_{12} & 0 \\ J_{21} & J_{22} & 0 \\ 0 & 0 & I_{33} \end{bmatrix}.$$

Thus, $\sqrt{n}(\tilde{\eta} - \eta^0)$ converges to multivariate normal distribution with mean vector zero and variance-covariance matrix $V = D_g^{-1} M_g (D_g^{-1})^\top$. □

We have that

$$D_g^{-1} = \begin{bmatrix} J_{11}^{-1} & 0 & 0 \\ 0 & J_{22}^{-1} & 0 \\ a_1 & a_2 & I_{33}^{-1} \end{bmatrix},$$

where $a_j = -I_{33}^{-1} I_{3j} J_{jj}^{-1}$ for $j = 1, 2$. A matrix multiplication gives

$$D_g^{-1} M_g = \begin{bmatrix} J_{11}^{-1} J_{11} & J_{11}^{-1} J_{12} & 0 \\ J_{22}^{-1} J_{21} & J_{22}^{-1} J_{22} & 0 \\ a_1 J_{11} + a_2 J_{21} & a_1 J_{12} + a_2 J_{22} & I_3 \end{bmatrix},$$

where I_3 is the identity matrix with the same dimension as ρ. Carrying out a final matrix multiplication, we can conclude that V has (j,k) element $J_{jj}^{-1} J_{jk} J_{kk}^{-1}$ for $1 \leq j, k \leq 2$; $(j,3)$ element $J_{jj}^{-1} \sum_{k=1}^{2} J_{jk} a_k^\top$ for j=1,2; $(3,j)$ element $(\sum_{k=1}^{2} a_k J_{kj}) J_{jj}^{-1}$ for $j = 1, 2$; $(3,3)$ element $I_{33}^{-1} + \sum_{j=1}^{2} \sum_{k=1}^{2} a_j J_{jk} a_k^\top$. The diagonal elements J_{jj}^{-1}, for $j = 1, 2$, of V can be obtained directly from the Fisher information matrices corresponding to the marginal log-likelihoods.

In a practical data analysis, we need a consistent estimator $\tilde{V}$ for the variance-covariance matrix V of $\tilde{\eta}$. Instead of deriving analytic forms for the derivatives in V and computing empirical versions of the expectations, we avoid taking these derivatives of log-likelihoods by using the jackknife estimator for $n^{-1}V$, proposed

in Joe (1997). We only have to code the log-likelihoods (marginal and bivariate) and use a numerical optimizer to obtain estimates of the marginal and dependence parameters. Let $\tilde{\eta}^{(i)}$ be the two-stage estimator for η with the i-th observation W_i deleted, for $i = 1, \ldots, n$. The jackknife estimator for $n^{-1}V$ is

$$\sum_{i=1}^{n} (\tilde{\eta}^{(i)} - \tilde{\eta})(\tilde{\eta}^{(i)} - \tilde{\eta})^T.$$

We give a non-rigorous justification of this estimation method. From (4.7), we know that

$$(\tilde{\eta} - \eta^0) \approx \frac{1}{n} D_g^{-1} \sum_{k=1}^{n} g(W_k, \eta^0),$$

for n large enough. In the same way, we can show that

$$\begin{aligned}
(\tilde{\eta}^{(i)} - \eta^0) &\approx \frac{1}{n-1} D_g^{-1} \sum_{k \neq i} g(W_k, \eta^0) \\
&\approx \frac{1}{n-1} D_g^{-1} \\
&\quad [n D_g (\tilde{\eta} - \eta^0) - g(W_i, \eta^0)] \\
&= \frac{n}{n-1} (\tilde{\eta} - \eta^0) - \frac{1}{n-1} D_g^{-1} g(W_i, \eta^0).
\end{aligned}$$

So,

$$\begin{aligned}
(\tilde{\eta}^{(i)} - \tilde{\eta}) &= (\tilde{\eta}^{(i)} - \eta^0) - (\tilde{\eta} - \eta^0) \\
&\approx \frac{n}{n-1} (\tilde{\eta} - \eta^0) - \frac{1}{n-1} D_g^{-1} g(W_i, \eta^0) - (\tilde{\eta} - \eta^0) \\
&= \frac{1}{n-1} (\tilde{\eta} - \eta^0) - \frac{1}{n-1} D_g^{-1} g(W_i, \eta^0) \\
&\approx -\frac{1}{n} D_g^{-1} g(W_i, \eta^0),
\end{aligned}$$

for n large enough. We conclude that

$$\begin{aligned}
\sum_{i=1}^{n} (\tilde{\eta}^{(i)} - \tilde{\eta})(\tilde{\eta}^{(i)} - \tilde{\eta})^\top &\approx \frac{1}{n^2} D_g^{-1} \left[\sum_{i=1}^{n} g(W_i, \eta^0) g^\top (W_i, \eta^0) \right] (D_g^\top)^{-1} \\
&\approx \frac{1}{n} D_g^{-1} M_g (D_g^\top)^{-1} \\
&= \frac{1}{n} V,
\end{aligned}$$

for n large enough.

4.1.2.2 Two-stage semi-parametric estimation

In the semi-parametric estimation approach we estimate the baseline cumulative hazard functions $\Lambda_1(t)$ and $\Lambda_2(t)$ by non-parametric step functions:

$$\Lambda_{1n}(t) = \sum_{k=1}^{q_n} \lambda_{1k} \mathrm{I}(u_{1k} \leq t), \ \Lambda_{2n}(t) = \sum_{l=1}^{r_n} \lambda_{2l} \mathrm{I}(u_{2l} \leq t)$$

where $0 < u_{11} < \ldots < u_{1q_n}$, resp. $0 < u_{21} < \ldots < u_{2r_n}$ are the unique uncensored observations for the first, resp. second measurement and $\lambda_{11}, \ldots, \lambda_{1q_n}$, resp. $\lambda_{21}, \ldots, \lambda_{2r_n}$ are the corresponding step sizes in these time points. Since estimation of all marginal and dependence parameters simultaneously becomes computationally difficult if the number of uncensored observations increases, we only consider the two-stage estimation procedure. At the first stage, we estimate the marginal parameters $\gamma_1, \beta_1, \lambda_{1k}$ and $\gamma_2, \beta_2, \lambda_{2l}$ by fitting univariate zero-inflated semi-parametric Cox's regression models (introduced in Chapter 2) for each margin, ignoring the dependence of the two components. At the second stage, estimates of the association parameters ψ and θ are obtained by fixing the parameters of the marginal distributions at their estimates and maximizing the bivariate log-likelihood with respect to ψ and θ. We consider, as in the parametric setting, a jackknife approach to obtain standard errors of the semi-parametric estimators.

In the numerical optimization process to find the different parameter estimates, we note that the number of parameters over which we have to optimize, can be reduced in some studies. As in the univariate setting, described in Chapter 2, we see that if the largest observations in either of the measurements are uncensored, the corresponding step sizes of the baseline cumulative hazard functions have a closed form solution which only contains the parameters β. For example, suppose that all observations for the first outcome variable larger or equal than u_{1p} are uncensored, then

$$\hat{\lambda}_{1p} = \frac{1}{\sum_{i=1}^{n} \exp(Z_i^\top \hat{\beta}_1) I(T_{1i} \geq u_{1p})}.$$

In the most extreme case, where all censored observations are smaller than the smallest uncensored observation (f.e. in case of a fixed detection limit as in the example of Section 4.3), fitting the zero-inflated Cox's regression model simplifies to fitting a logistic regression model on the censoring indicator variables and fitting a Cox's regression model on the subgroup of the uncensored observations for the first outcome variable.

4.2 Simulation study

In this section, the performance of the one-stage parametric, the two-stage parametric and the two-stage semi-parametric estimation methods are compared. We set up a simulation study and generate data sets using the following simulation scheme:

(i) We generate a covariate related to the first measurement: $X_1 = Z_1 \sim$ Uniform$[0, 10]$ and a covariate related to the second measurement: $X_2 = Z_2 \sim$ Uniform$[0, 10]$. So $X = (X_1, X_2)^\top$ and $Z = (Z_1, Z_2)^\top$.

(ii) For the censoring times, we assume that $C_1 \sim$ Uniform$[0, c_1]$, $C_2 \sim$ Uniform$[0, c_2]$ with C_1 and C_2 independent.

(iii) We consider logistic regression models for the probability of zero values of the event times Y_1 and Y_2:

$$\pi_1(\gamma_1, X) = \frac{\exp(\gamma_1^0 + \gamma_1^1 X_1)}{1 + \exp(\gamma_1^0 + \gamma_1^1 X_1)}, \quad \pi_2(\gamma_2, X) = \frac{\exp(\gamma_2^0 + \gamma_2^1 X_2)}{1 + \exp(\gamma_2^0 + \gamma_2^1 X_2)}.$$

Hereby we take the cross ratio ψ to model the association.

(iv) The non-zero event times are drawn from an Archimedean copula model (parameter: θ) with weibull regression models as margins (parameters: β_1, β_2). The baseline hazards λ_1, λ_2 are from Weibull(a_1,b_1) and Weibull(a_2,b_2).

(v) For the observed measurements, we consider $T_1 = \max(Y_1, C_1)$ and $\delta_1 = I(Y_1 \geq C_1)$, $T_2 = \max(Y_2, C_2)$ and $\delta_2 = I(Y_2 \geq C_2)$.

In this simulation study, we investigate four different simulation settings. The marginal parameters for both measurements $\gamma_1, \beta_1, a_1, b_1$ and $\gamma_2, \beta_2, a_2, b_2$ are chosen to take the values in Table 4.2. The censoring variables C_1 and C_2 are assumed to have the same uniform distribution with support determined by the value of c. The choice of the four settings is made in such a way that we can compare the performance of the estimation methods for high and low marginal probabilities of a zero response, for different levels of censoring and for different strengths of association between two non-zero responses.

Note that the marginal parameters of setting 1 and 4 correspond to the parameter values of setting 1 in Section 2.6 of Chapter 2. From Figure 2.2, we notice

	$\gamma^\top$	β	a	b	c	ψ	θ
Setting 1	(-0.3,0.15)	-0.05	4	30	30	3	2
Setting 2	(-2,0.10)	-0.05	4	30	30	3	2
Setting 3	(-0.3,0.15)	-0.05	4	30	40	3	2
Setting 4	(-0.3,0.15)	-0.05	4	30	30	3	6

Table 4.2: Parameter values for γ, β, a, b, c, ψ and θ in the different settings.

that the overall conditional censoring probability is high. This is mainly due to a high conditional probability of a zero response. The marginal parameters of setting 2 correspond to the parameter values of setting 2 in Section 2.6. This marginal setting corresponds to a low conditional probability of a zero response, combined with a rather small conditional censoring probability for the responders. In the third setting, which corresponds to setting 3 of Section 2.6, there is a high overall conditional censoring probability, induced by a high conditional probability of a zero response, together with a rather high conditional censoring probability for the responders.

For the association between two non-zero responses, we consider the Clayton, Frank and Gumbel-Hougaard families of copula functions. Since the conclusions for the Frank and Gumbel-Hougaard copula functions are similar to those of Clayton, we only show results for the Clayton copula model. In Setting 1-3, we take $\theta = 2$ corresponding to a Kendall's $\tau = 0.50$ (association between two non-zero responses). In Setting 4, the values of the marginal parameters are equal to those in Setting 1, but the association between two non-zero responses is very high ($\tau = 0.75$). In all settings, we choose the cross ratio ψ fixed at three. Since the Clayton copula only has a strict generator when the event times have a positive association, we did not consider a negative association for the moment. Furthermore we kept the cross ratio fixed because we wanted to focus on the influence of the zero inflation probability in the four settings.
For each simulation setting we generate 500 data sets, with $n = 200$, or $n = 500$ observations. We use the **copula**-package (Yan (2007)) in **R** to draw samples from the different copula models. We conduct the one-stage and two-stage parametric estimation approach, assuming Weibull baseline cumulative hazard functions. We also estimate the marginal and association parameters by the two-stage semi-parametric approach. In each setting, we compute estimates of the bias and standard deviation of the one-stage and two-stage parametric estimates and the two-stage semi-parametric estimates. The results are shown in Table 4.3 and Table 4.4.

		$n = 200$						$n = 500$					
		Param				Semi-param		Param				Semi-param	
		1-stage		2-stage		2-stage		1-stage		2-stage		2-stage	
		Bias	Sd	Bias	Sd	Bias	Sd	Bias	Sd	Bias	Sd	Bias	Sd
Setting 1	$\hat{\gamma}_1^0$	-0.0084	0.3137	-0.0093	0.3234	-0.0066	0.3248	0.0060	0.2032	0.0067	0.2083	0.0069	0.2075
	$\hat{\gamma}_1^1$	0.0021	0.0566	0.0021	0.0578	0.0019	0.0578	-0.0010	0.0365	-0.0011	0.0369	-0.0010	0.0368
	$\hat{\beta}_1$	-0.0003	0.0404	0.0015	0.0461	0.0017	0.0460	-0.0014	0.0239	-0.0012	0.0275	-0.0016	0.0278
	$\hat{\Lambda}_1(20)$	-0.0027	0.0554	-0.0026	0.0645	-0.0052	0.0834	-0.0005	0.0351	-0.0004	0.0418	0.0006	0.0535
	$\hat{\Lambda}_1(30)$	0.0144	0.2114	0.0131	0.2411	0.0051	0.2483	0.0092	0.1265	0.0099	0.1486	0.0061	0.1579
	$\hat{\Lambda}_1(40)$	0.1625	0.7377	0.1676	0.8132	0.0306	0.8300	0.0762	0.4269	0.0853	0.4841	0.0605	0.5298
	$\hat{\gamma}_2^0$	-0.0007	0.3238	-0.0021	0.3340	-0.0005	0.3340	0.0204	0.1877	0.0249	0.1963	0.0259	0.1978
	$\hat{\gamma}_2^1$	0.0034	0.0595	0.0043	0.0612	0.0043	0.0612	-0.0038	0.0342	-0.0045	0.0360	-0.0045	0.0361
	$\hat{\beta}_2$	0.0002	0.0394	0.0004	0.0455	0.0008	0.0459	0.0003	0.0230	0.0008	0.0273	0.0009	0.0271
	$\hat{\Lambda}_2(20)$	-0.0049	0.0548	-0.0074	0.0624	-0.0052	0.0875	-0.0017	0.0357	-0.0032	0.0417	-0.0039	0.0558
	$\hat{\Lambda}_2(30)$	0.0065	0.2072	0.0045	0.2410	-0.0056	0.2452	0.0015	0.1297	-0.0020	0.1496	-0.0053	0.1557
	$\hat{\Lambda}_2(40)$	0.1445	0.7190	0.1755	0.8297	0.0796	0.8666	0.0464	0.4072	0.0502	0.4653	0.0075	0.5019
	$\hat{\psi}$	0.2810	1.3345	0.2291	1.2878	0.1865	1.2608	0.0982	0.7291	0.0789	0.7176	0.0650	0.7174
	$\hat{\theta}$	0.1711	0.7359	-0.1358	0.6065	-0.1351	0.6338	0.0731	0.4260	-0.0489	0.4049	-0.0730	0.4328
Setting 2	$\hat{\gamma}_1^0$	-0.0578	0.4984	-0.0659	0.5410	-0.0557	0.5364	-0.0193	0.3385	-0.0219	0.3636	-0.0278	0.3723
	$\hat{\gamma}_1^1$	0.0065	0.0758	0.0072	0.0815	0.0062	0.0806	0.0004	0.0508	0.0003	0.0544	0.0008	0.0551
	$\hat{\beta}_1$	-0.0001	0.0246	-0.000004	0.0290	0.00004	0.0293	-0.0008	0.0145	-0.0010	0.0176	-0.0010	0.0177
	$\hat{\Lambda}_1(20)$	-0.0022	0.0414	-0.0033	0.0458	-0.0040	0.0609	-0.0008	0.0232	-0.0007	0.0265	0.0014	0.0379
	$\hat{\Lambda}_1(30)$	0.0075	0.1540	0.0074	0.1754	0.0055	0.1839	0.0022	0.0878	0.0029	0.0982	0.0008	0.1026
	$\hat{\Lambda}_1(40)$	0.0926	0.5121	0.1083	0.5782	0.0689	0.6326	0.0299	0.2959	0.0352	0.3248	0.0060	0.3486
	$\hat{\gamma}_2^0$	-0.0620	0.5423	-0.0472	0.5657	-0.0447	0.5736	-0.0289	0.2940	-0.0303	0.3235	-0.0297	0.3279
	$\hat{\gamma}_2^1$	0.0040	0.0822	0.0015	0.0860	0.0015	0.0863	0.0031	0.0452	0.0034	0.0491	0.0034	0.0493
	$\hat{\beta}_2$	-0.0019	0.0221	-0.0005	0.0265	-0.0002	0.0264	-0.0003	0.0146	-0.0007	0.0175	-0.0006	0.0174
	$\hat{\Lambda}_2(20)$	-0.0007	0.0370	-0.0030	0.0422	-0.0046	0.0573	-0.0018	0.0254	-0.0019	0.0288	-0.0022	0.0365
	$\hat{\Lambda}_2(30)$	0.0148	0.1352	0.0073	0.1572	0.0026	0.1676	0.0008	0.0933	0.0020	0.1072	0.0010	0.1128
	$\hat{\Lambda}_2(40)$	0.1116	0.4587	0.1037	0.5321	0.0581	0.5699	0.0342	0.2974	0.0436	0.3345	0.0304	0.3579
	$\hat{\psi}$	0.8900	3.2902	0.7134	2.7538	0.7149	2.7478	0.3211	1.2525	0.2790	1.2238	0.2637	1.2210
	$\hat{\theta}$	0.0500	0.3688	-0.0653	0.3345	-0.0893	0.3688	0.0075	0.2173	-0.0341	0.2100	-0.0432	0.2266

Table 4.3: The bias and standard deviation of the one-stage parametric, two-stage parametric and two-stage semi-parametric estimators - Setting 1 and 2.

		$n = 200$						$n = 500$					
		Param				Semi-param		Param				Semi-param	
		1-stage		2-stage		2-stage		1-stage		2-stage		2-stage	
		Bias	Se	Bias	Sd	Bias	Sd	Bias	Sd	Bias	Sd	Bias	Sd
Setting 3	$\hat{\gamma}_1^0$	0.0291	0.4024	0.0368	0.4145	0.0414	0.4190	0.0272	0.2346	0.0291	0.2433	0.0312	0.2439
	$\hat{\gamma}_1^1$	-0.0044	0.0676	-0.0054	0.0689	-0.0057	0.0692	-0.0036	0.0403	-0.0039	0.0414	-0.0040	0.0413
	$\hat{\beta}_1$	-0.0038	0.0455	-0.0027	0.0491	-0.0017	0.0499	0.0010	0.0266	0.0015	0.0296	0.0016	0.0294
	$\hat{\Lambda}_1(20)$	0.0004	0.0624	-0.0028	0.0678	-0.0009	0.0975	-0.0015	0.0393	-0.0020	0.0459	-0.0036	0.0602
	$\hat{\Lambda}_1(30)$	0.0344	0.2414	0.0252	0.2523	0.0183	0.2687	0.0019	0.1417	-0.0001	0.1605	-0.0011	0.1745
	$\hat{\Lambda}_1(40)$	0.2484	0.8395	0.2534	0.8846	0.1565	0.9577	0.0537	0.4824	0.0570	0.5274	0.0194	0.5729
	$\hat{\gamma}_2^0$	0.0054	0.4006	0.0038	0.4135	0.0067	0.4202	0.0022	0.2225	0.0031	0.2300	0.0047	0.2357
	$\hat{\gamma}_2^1$	0.0007	0.0696	0.0007	0.0709	0.0009	0.0714	-0.0001	0.0395	0.0001	0.0404	0.00003	0.0408
	$\hat{\beta}_2$	-0.0013	0.0451	-0.0013	0.0483	-0.0015	0.0493	0.0001	0.0261	0.0004	0.0291	0.0007	0.0293
	$\hat{\Lambda}_2(20)$	-0.0001	0.0656	0.0009	0.0776	0.0033	0.1054	-0.0023	0.0376	-0.0036	0.0427	-0.0050	0.0609
	$\hat{\Lambda}_2(30)$	0.0251	0.2497	0.0282	0.2802	0.0254	0.3037	0.0074	0.1385	0.0037	0.1523	0.0037	0.1625
	$\hat{\Lambda}_2(40)$	0.2030	0.8474	0.2284	0.8955	0.1127	0.9590	0.0903	0.4589	0.0929	0.4927	0.0536	0.5316
	$\hat{\psi}$	0.5726	1.9700	0.4527	1.8052	0.4067	1.8243	0.2337	1.0021	0.2088	0.9796	0.1861	0.9728
	$\hat{\theta}$	0.3959	1.0455	-0.0615	0.7764	-0.0375	0.8303	0.1179	0.5453	-0.0424	0.4924	-0.0621	0.5110
Setting 4	$\hat{\gamma}_1^0$	0.0045	0.3426	0.0063	0.3614	0.0046	0.3635	-0.0060	0.1962	-0.0076	0.2039	-0.0074	0.2082
	$\hat{\gamma}_1^1$	0.0017	0.0584	0.0018	0.0610	0.0021	0.0611	0.0014	0.0345	0.0018	0.0357	0.0019	0.0358
	$\hat{\beta}_1$	-0.0006	0.0245	0.0001	0.0456	0.0003	0.0453	-0.00007	0.0153	0.0012	0.0278	0.0013	0.0282
	$\hat{\Lambda}_1(20)$	-0.0006	0.0477	-0.0032	0.0669	-0.0033	0.0914	-0.0017	0.0302	-0.0025	0.0420	-0.0016	0.0570
	$\hat{\Lambda}_1(30)$	0.0128	0.1627	0.0180	0.2673	0.0084	0.2708	0.0021	0.1041	0.0001	0.1544	-0.0055	0.1584
	$\hat{\Lambda}_1(40)$	0.1216	0.5745	0.2067	0.9341	0.0892	0.9649	0.0466	0.3427	0.0542	0.4952	0.0143	0.5168
	$\hat{\gamma}_2^0$	-0.0021	0.3254	-0.0013	0.3365	0.0016	0.3366	0.0035	0.1957	0.0042	0.2016	0.0048	0.2038
	$\hat{\gamma}_2^1$	0.0001	0.0585	0.0001	0.0601	-0.0002	0.0602	-0.0006	0.0337	-0.0005	0.0352	-0.0004	0.0353
	$\hat{\beta}_2$	-0.0010	0.0265	-0.0002	0.0444	0.00003	0.0457	-0.0004	0.0146	-0.0013	0.0252	-0.0017	0.0259
	$\hat{\Lambda}_2(20)$	-0.0027	0.0457	-0.0036	0.0644	-0.0042	0.0868	-0.0016	0.0292	-0.0017	0.0414	0.0001	0.0564
	$\hat{\Lambda}_2(30)$	0.0128	0.1555	0.0152	0.2341	0.0096	0.2496	0.0005	0.0952	0.0075	0.1452	0.0060	0.1540
	$\hat{\Lambda}_2(40)$	0.1507	0.5951	0.1964	0.7944	0.0867	0.8312	0.0369	0.3151	0.0848	0.4533	0.0519	0.4889
	$\hat{\psi}$	0.3846	1.4261	0.3262	1.3838	0.2857	1.3482	0.1680	0.7320	0.1444	0.7200	0.1288	0.7122
	$\hat{\theta}$	0.6684	1.7241	-1.2539	1.4073	-1.6529	1.2766	0.2513	0.8915	-0.6222	0.8999	-0.9433	0.8633

Table 4.4: The bias and standard deviation of the one-stage parametric, two-stage parametric and two-stage semi-parametric estimators - Setting 3 and 4.

In the four settings, we see that the biases and the standard deviations of the one-stage and two-stage estimates decrease as the sample size increases. This is in line with what we expect based on the theoretical results. Since the estimates of the marginal parameters in the two-stage procedure are obtained without using the information in the correlation, we expect the two-stage parametric estimates to be inefficient relative to the one-stage maximum likelihood estimates. However, we see that the estimators of the marginal parameters in the two-stage parametric method perform almost as well as the one-stage estimators. In simulation Settings 1, 2 and 3, we notice that the two-stage parametric approach performs even better than the one-stage approach in estimating the dependence parameters ψ and θ. This can be explained as follows: due to the two-stage procedure, the number of parameters over which, in each step, has to be optimized is lower and therefore the estimation can be done more accurately. When the association between two non-zero responses is very high, as in Setting 4, the bias of the two-stage estimator of θ is (in absolute value) higher than the bias of the one-stage estimator. Ignoring the high dependence of two non-zero responses in the estimation of the marginal parameters, results in a higher bias for the estimation of θ in the second stage.

The larger bias and standard deviation for the logistic parameters and the association parameter ψ in Setting 2, compared to Setting 1, is due to the small rate of zero responses for both measurements. The estimators for parameters in the Cox's regression models and the copula parameter, on the other hand, produced smaller biases and standard deviations. The higher percentage of censoring among the responders in Setting 3 explains the higher biases and standard deviations compared to Setting 1.

When we compare the performance of the two-stage parametric and the two-stage semi-parametric approach, we note that both approaches have about the same efficiency. Relaxing the parametric assumptions seems not to reduce the efficiency of the dependence parameters. We can conclude that the two-stage parametric and semi-parametric estimation methods work well, especially when the association between both measurements is not too high. We emphasize that two-stage estimates are obtained much faster than the one-stage estimates.

Furthermore we obtain for each data set with 200 observations in Setting 1 an estimate for the variance-covariance matrix of the one-stage parametric estimators (by inverting the Hessian matrix at the optimum) and for the variance-covariance matrix of the two-stage parametric estimators (estimate based on analytic derivatives and empirical versions of the expectations and a jackknife estimate). We show the means of the estimates of the standard errors in Table 4.5. By comparing the means of the empirical

	1-stage	2-stage	
	Inv Hes	Emp	Jack
$\hat{\gamma}_1^0$	0.3243	0.3371	0.3429
$\hat{\gamma}_1^1$	0.0573	0.0596	0.0606
$\hat{\beta}_1$	0.0380	0.0474	0.0472
$\hat{a}_1$	0.4007	0.4887	0.4753
$\hat{b}_1$	1.4938	1.7866	1.7573
$\hat{\gamma}_2^0$	0.3233	0.3361	0.3403
$\hat{\gamma}_2^1$	0.0573	0.0597	0.0603
$\hat{\beta}_2$	0.0381	0.0474	0.0469
$\hat{a}_2$	0.4046	0.4979	0.4773
$\hat{b}_2$	1.4901	1.7671	1.7272
$\hat{\psi}$	1.3026	1.3064	1.3285
$\hat{\theta}$	0.6650	0.6877	0.6781

Table 4.5: Estimation of the standard error of the one-stage and two-stage parametric estimators - Setting 1, n=200.

estimates (Emp) with the corresponding standard deviations in Table 4.3, we see that the sandwich estimator performs well in estimating the true standard errors. We also note that the means of the jackknife estimates of the standard errors of the two-stage estimators (Jack) are comparable to the empirical estimates and are only little larger than the estimates of the standard errors for the one-stage estimators (Inv Hes). The jackknife estimator performs well in estimating the standard errors of the estimators in the two-stage parametric approach.

4.3 Example: Modeling ethanol-induced anesthesia

In this section, we further analyze the data from the practical study of ethanol-induced anesthesia in genetically-selected strains of mice, described in Section 1.2. The data set has a repeated measurement design where mice are tested at two different times. In Sections 2.7 and 3.4, we only considered observations of the first test session for the segregating F_2 mouse population. In this section, we consider the observations of both test sessions for the F_2 population. In Figure 1.1, we saw that the range and the distribution of the observed lifetimes on the lines parallel to the x- and y-axis (with one or both observations censored at the detection limit of 1 min) is no much different from the range and the distribution of the lifetimes for the corresponding measurement within the plane. Except for the more extreme observed bivariate couples, we think that the simplifying assumption (4.2) of Section 4.1 is

not too much violated, such that we can assume that the marginal distribution of a non-zero component does not depend on the susceptibility status of the other component. In this way, we can use the zero-inflated regression model to analyze this data set.

We fit logistic regression models for the marginal probability of a zero response in each measurement:

$$\begin{aligned} \text{logit } \pi_1 &= \gamma_1^0 + \gamma_1^1 \text{Albinism}, \\ \text{logit } \pi_2 &= \gamma_2^0 + \gamma_2^1 \text{Albinism}. \end{aligned}$$

Hereby we considered after model selection only albinism a a binary variable indicating whether the mouse was albino. We further fit a proportional hazards model for the distribution of non-zero responses for both measurements:

$$\begin{aligned} \lambda_{Y_1>0}(t) &= \lambda_1(t) \exp[\beta_1^1 \text{Sex} + \beta_1^2 \text{Weight}_1], \\ \lambda_{Y_2>0}(t) &= \lambda_2(t) \exp[\beta_2^1 \text{Sex} + \beta_2^2 \text{Weight}_2]. \end{aligned}$$

In these models, we study the influence of the covariates sex and weight at respectively trial 1 and trial 2, on the hazard function of the non-zero response. Male mice with weight of 21 grams are used as reference in both trials. Since the sleep times in both trials are recorded on the same animal, they are correlated. The dependence structure of two non-zero responses is fit by a Clayton, Frank and Gumbel-Hougaard family of copula functions. Each of these copula functions represents a specific association structure in which resp. early event times, latter event times or, early and latter event times are more correlated. The cross ratio ψ induces the joint probability of having zero responses for both measurements.
We conduct both the two-stage parametric and the two-stage semi-parametric estimation approach. In the parametric approach, we fit Weibull hazard functions for the baseline hazard functions λ_1 and λ_2. In the semi-parametric approach, the baseline cumulative hazard functions are estimated by non-decreasing step functions, with jumps at the uncensored observations for respectively measurement 1 and measurement 2. As we saw in the methodology, the semi-parametric regression model assumes that the baseline hazards are zero before the smallest uncensored observations for respectively measurement 1 and measurement 2. Consequentially, the probabilities for a value of sleep time between zero and the detection limit are also zero. At the first stage, we estimate all marginal parameters ignoring the dependence of the two measurements. At the second stage, we hold the marginal parameters fixed from the

first stage and estimate the dependence parameters. The two-stage parametric and semi-parametric estimates for the marginal parameters are shown in Table 4.6. The standard errors are obtained by the jackknife estimation method. We note that, due to the two-stage estimation approach, estimates of the marginal parameters do not depend on the choice of the association models.

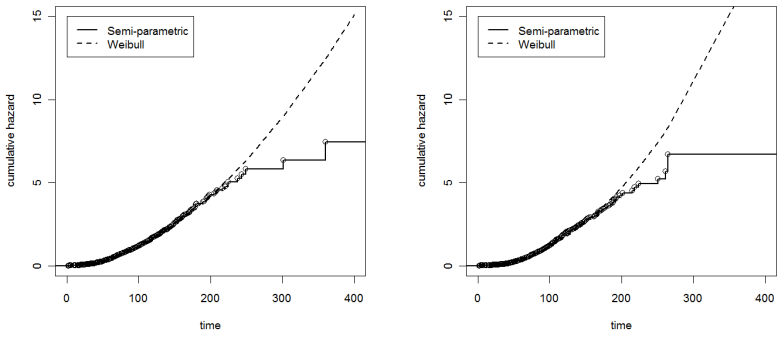
	Parametric	**Semi-parametric**
	Logistic part	
$\hat{\gamma}_1^0$ (Intercept)	-3.4616 (0.2089)	-3.4532 (0.2072)
$\hat{\gamma}_1^1$ (Albinism)	0.8147 (0.3328)	0.8099 (0.3311)
$\hat{\gamma}_2^0$ (Intercept)	-3.1373 (0.1782)	-3.1342 (0.1777)
$\hat{\gamma}_2^1$ (Albinism)	0.9124 (0.2791)	0.9107 (0.2785)
	Hazard part	
$\hat{\beta}_1^1$ (Sex)	0.0609 (0.0817)	0.0607 (0.0825)
$\hat{\beta}_1^2$ (Weight_1)	-0.0192 (0.0136)	-0.0171 (0.0131)
$\hat{\beta}_2^1$ (Sex)	0.3414 (0.0847)	0.3336 (0.0836)
$\hat{\beta}_2^2$ (Weight_2)	0.0517 (0.0130)	0.0495 (0.0131)
Shape_1	1.8407 (0.0565)	
Scale_1	91.5099 (2.6009)	
Shape_2	2.0174 (0.0657)	
Scale_2	92.7139 (2.5032)	

Table 4.6: Estimates for the marginal parameters. Standard errors are given in brackets.

We notice that in both the parametric and the semi-parametric model, the same parameters are significant. In the logistic part for both measurements, an albino mouse has a significant higher probability of having a zero value for the sleep time than a non-albino mouse. In the hazard part of the first measurement, the effects of both sex and weight are not significant. This means that as soon as the mice are sleeping, albinism and weight don't have a significant effect on the duration of the sleep time at the first measurement. The parameters β_2^1 (Sex) and β_2^2 (Weight_2) are highly significant. At the second measurement, a female mouse has a higher hazard than a male mouse with the same weight, which means a shorter sleep time for female mice. In contrast to the first trial, the hazard increases for heavier mice at the second measurement, resulting in a shorter sleep time compared to thinner mice of the same sex. We believe that heavier mice got used to the alcohol after the first

trial and that these mice are more resistant to the alcohol challenge at the second trial. We also notice that the estimates and standard errors of the parameters are almost the same in the parametric and semi-parametric model.

In Figure 4.1, we show the estimates of the baseline cumulative hazard functions for both the parametric and semi-parametric model. We see that, for both



(a) Measurement 1 (b) Measurement 2

Figure 4.1: Parametric and non-parametric estimates of the baseline cumulative hazard functions.

measurements, the estimated Weibull cumulative hazard function and the non-parametric step function are almost the same for a quite large interval of sleep times. The Weibull models fit the distributions of the non-zero response values in both measurements reasonably well. The Weibull cumulative baseline hazards are almost zero for small times, which explains why there is not much difference between the parametric and semi-parametric models.

The parameter estimates for the dependence parameters and their standard errors are given in Table 4.7.

The estimates of the hazard ratio ψ indicate a significant association between the susceptibilities of the mice at the two test sessions, regardless of the choice of parametric or semi-parametric marginal distributions and irrespective the choice of the copula model. Mice with a strict positive value for the sleep time at trial 1 are likely to have a strict positive value for sleep time at trial 2. In the semi-parametric models, the estimate of ψ does not depend on the choice of copula that models the association

Parametric			
	Clayton	Frank	Gumbel-Hougaard
$\hat{\psi}$	4.0651 (1.7941)	3.9756 (1.7397)	3.9927 (1.7492)
$\hat{\theta}$	0.3848 (0.0556)	4.2084 (0.2496)	1.4661 (0.0399)
$\hat{\tau}_\theta$	0.1613	0.4034	0.3179
Semi-Parametric			
	Clayton	Frank	Gumbel-Hougaard
$\hat{\psi}$	3.9416 (1.7174)	3.9416 (1.7174)	3.9416 (1.7174)
$\hat{\theta}$	0.7891 (0.0676)	3.8921 (0.2365)	1.4956 (0.0380)
$\hat{\tau}_\theta$	0.2829	0.3800	0.3314

Table 4.7: Estimates for the dependence parameters. Standard errors are given in brackets.

between two non-zero responses. This can be explained by the presence of a fixed detection limit for both measurements and both baseline hazards being zero before the smallest uncensored observations for respectively the first and second measurement. The full likelihood splits up so that the logistic paramaters and ψ can be estimated separately from the other parameters. The copula parameter can't be compared between models, because it has a different interpretation in different models. However, all copula parameter estimates correspond to a low to moderate positive global association between two non-zero responses. Kendall's tau values corresponding to the different copula parameter estimates $\hat{\theta}$ range between 0.16 and 0.40. Mice that have a longer sleep time at trial 1 tend to have a longer sleep time at trial 2. We conclude that there is a clear association between the sleep times in both trials.

4.4 Conclusion

In some studies with bivariate left-censored data, the underlying response variables also attain a zero value with a positive discrete probability. We introduced parametric and semi-parametric regression models for these bivariate zero-inflated left-censored survival data. The different parameters in the model are estimated using maximum likelihood techniques. The numerical optimization of the likelihood becomes more difficult as the number of parameters increases. Fortunately, the model structure suggests that a two-stage estimation procedure can be considered. Firstly we estimate

the parameters in the margins, ignoring the dependence of the two components. The second stage involves maximum likelihood of the dependence parameters with the univariate parameters held fixed from the first stage. We derived a partitioned form for the asymptotic variance-covariance matrix of the two-stage parametric estimators and discussed a jackknife estimator for this matrix. In the simulation study, we showed that the two-stage parametric and semi-parametric estimation methods perform well, especially when the association between two non-zero responses is low or moderate. Finally, we have applied our regression model on a practical data set of ethanol-induced sleep time in mice.

Chapter 5

Concluding remarks and possible future research

In Chapter 2, a new semi-parametric regression model for analyzing zero-inflated randomly left-censored time to event data was introduced. It was proposed to combine a parametric regression model for the probability of having no response with a semi-parametric model for the time to event for the responders. The different parameters in the mixture regression model were estimated by maximizing an empirical likelihood. Some nice properties such as the consistency and the asymptotic normality of the maximum likelihood estimators were proved.
For some data sets, the optimization procedure proposed in Chapter 2 was not able to estimate the probability of a zero response well. Therefore, in Chapter 3, an approximation of the likelihood was considered. We developed an efficient EM-algorithm to compute the maximum approximated likelihood estimates. The approximation of the likelihood led to a high dimension reduction, which made the optimization procedure more stable. The maximum approximated likelihood estimates were more accurate than the corresponding maximum likelihood estimates. Moreover, the MAL estimates were obtained much faster. As result, the consistency of the maximum approximated likelihood estimators was proved. In future research, the asymptotic normality of the MAL estimators may be studied.

The focus of Chapter 4 was on parametric and semi-parametric regression models for analyzing bivariate zero-inflated left-censored time to event data. A marginal mo-

deling approach was considered. First, the marginal probabilities of a zero response were modeled by a parametric regression model. The non-zero parts of both outcome variables were modeled by parametric or semi-parametric proportional hazards models. Afterwards, a dependence structure was imposed to model the association of the two measurements. For the joint probability of having zero-responses for both measurements, we modeled the cross ratio in a contingency table with combinations of zero and non-zero responses for both measurements. The association between two non-zero responses was modeled by a parametric family of copulas. In the proposed models, the cross ratio and the copula parameter where not allowed to vary with some covariates. As extension, we may model the influence of covariates on the copula parameter θ and on the cross ratio ψ with parametric regression models.

To estimate the different parameters in our model, maximum likelihood techniques were used. The special structure of the model suggested the consideration of a two-stage estimation procedure, instead of maximizing the full log-likelihood at once. At the first stage the parameters in the margins were estimated. Hereby, the dependence of the two measurements was ignored. In the semi-parametric estimation approach, we obtained estimates for the marginal parameters by fitting a univariate zero-inflated semi-parametric Cox's regression model, introduced in Chapter 2, for each margin. The second stage involved maximum likelihood of the dependence parameters with the univariate parameters held fixed from the first stage. The simulation study showed that the two-stage parametric and semi-parametric estimation methods perform reasonably well, especially when the association between both measurements is not too high.

In future research, it may be worth to investigate whether an approximation of the semi-parametric full likelihood, similar to the one used in the univariate case (Chapter 3), gives more accurate estimates for both the marginal and the dependence parameters. At the first stage of the two-stage estimation procedure, the efficient EM-algorithm may be used to obtain estimates for the marginal parameters.
Furthermore, in Chapter 4, the performance of the two-stage semi-parametric estimation approach was only investigated by simulations. We may try to prove the consistency and the asymptotic normality of the estimators under some regularity conditions. Finally, we may think about goodness-of-fit tests to verify whether a class of copula regression models complies with the data found in a practical study.

Finally, mention that the content of Chapter 2 is contained in Grouwels and Braekers (2011) and Braekers and Grouwels (2013), while Braekers and Grouwels (2015) deals with the bivariate parametric and semi-parametric regression models introduced in Chapter 4. In future, we will also try to publish the results of Chapter 3 in a scientific journal.

Bibliography

Amemiya, T. (1973). Regression analysis when the dependent variable is truncated normal, *Econometrica* **41**: 997 – 1016.

Amemiya, T. (1984). Tobit models: A survey, *Journal of Econometrics* **24**: 3–61.

Berk, K. N. and Lachenbruch, P. A. (2002). Repeated measures with zeros, *Statistical Methods in Medical Research* **11**: 303–316.

Berkson, J. and Gage, R. P. (1952). Survival curve for cancer patients following treatment, *Journal of the American Statistical Association* **47**: 501–515.

Bickel, P. J., Klaassen, C. A. J., Ritov, Y. and Wellner, J. A. (1993). *Efficient and adaptive estimation for semiparametric models*, The Johns Hopkins University Press, Baltimore and London.

Blackwood, L. G. (1991). Analyzing censored environmental data using survival analysis: single sample techniques, *Environmental Monitoring and Assessment* **18**: 25–40.

Braekers, R. and Grouwels, Y. (2013). A semi-parametric Cox's regression model for zero-inflated left-censored time to event data. Accepted for publication by Communications in Statistics - Theory and Methods.

Braekers, R. and Grouwels, Y. (2015). Modeling bivariate zero-inflated time to event data by a copula function. Revision sent to Biometrical Journal.

Chatterjee, N. and Shih, J. H. (2001). A bivariate cure-mixture approach for modeling familial association in diseases, *Biometrics* **57**: 779–786.

Chen, M.-H., Ibrahim, J. and Sinha, D. (1999). A new bayesian model for survival data with a surviving fraction, *Journal of the American Statistical Association* **94**: 909–919.

Chu, H., Moulton, L. H., Mack, W. J., Passaro, D. J., Barroso, P. F. and Munoz, A. (2005). Correlating two continuous variables subject to detection limits in the context of mixture distributions, *Applied Statistics* **54**: 831–845.

Cooner, F., Banerjee, S., Carlin, B. and Sinha, D. (2007). Flexible cure rate modeling under latent activation schemes, *Journal of the American Statistical Association* **102**: 560–572.

Cox, D. R. (1972). Regression models and life tables (with discussion), *Journal of the Royal Statistical Society, Series B (Methodological)* **34**: 187–220.

Cragg, J. G. (1971). Some statistical models for limited dependent variables with application to the demand for durable goods, *Econometrics* **39**: 829–844.

Duan, N., Manning, W. G. J., Morris, C. N. and Newhouse, J. (1983). A comparison of alternative models for the demand for medical care (corr: V2 p413), *Journal of Business and Economic Statistics* **1**: 115–126.

Farewell, V. T. (1982). The use of mixture models for the analysis of survival data with long-term survivors, *Biometrics* **38**: 1041–1046.

Farewell, V. T. (1986). Mixture models in survival analysis: are they worth the risk?, *Canadian Journal of Statistics* **14**: 257–262.

Genest, C. and MacKay, R. J. (1986a). Archimedean copulas and bivariate families with continuous marginals, *Canadian Journal of Statistics* **14**: 145–159.

Genest, C. and MacKay, R. J. (1986b). The joy of copulas: Bivariate distributions with uniform marginals, *American Statistician* **40**: 280–283.

Grouwels, Y. and Braekers, R. (2011). Zero-inflated semi-parametric Cox's regression model for left-censored survival data, *SCo 2011 Proceedings* .

Joe, H. (1997). *Multivariate models and dependence concepts*, Chapman & Hall.

Joe, H. (2005). Asymptotic efficiency of the two-stage estimation method for copula-based models, *Journal of Multivariate Analysis* **94**: 401–419.

Kim, Y., Kim, B. and Jang, W. (2010). Asymptotic properties of the maximum likelihood estimator for the proportional hazards model with doubly censored data, *Journal of Multivariate Analysis* **101**: 1339–1351.

Kim, Y., Kim, J. and Jang, W. (2013). An EM algorithm for the proportional hazards model with doubly censored data, *Computational Statistics and Data Analysis* **57**: 41–51.

Kuk, A. C. and Chen, C.-H. (1992). Amixturemodel combining logistic regression with proportional hazards regression, *Biometrika* **79**: 531–541.

Lehmann, E. L. (1998). *Theory of point estimation*, New York, N.Y. : Springer.

Lu, W. (2008). Maximum likelihood estimation in the proportional hazards cure model, *Annals of the Institute of Statistical Mathematics* **60**: 545–574.

Maller, R. A. and Zhou, S. (1996). *Survival analysis with long-term survivors*, John Wiley and Sons Ltd, United Kingdom.

Markel, P. D. and Corley, R. (1994). A multivariate analysis of repeated measures: linkage of the albinism gene (Tyr) to a QTL influencing ethanol-induced anasthesia in laboratory mice., *Psychiatric Genetics* **4**: 205–210.

Markel, P. D., DeFries, J. C. and Johnson, T. E. (1995). Ethanol-induced anesthesia in inbred strains of long-sleep and short-sleep mice: A genetic analysis of repeated measures using censored data, *Behavior Genetics* **25**: 67–73.

Moulton, L. H. and Halsey, N. A. (1995). A mixture model with detection limits for regression analysis of antibody response to vaccine, *Biometrics* **51**: 1570–1578.

Nelsen, R. B. (2006). *An introduction to copulas*, New York, N.Y. : Springer.

Olsen, M. K. and Schafer, J. L. (2001). A two-part random-effects model for semi-continuous longitudinal data, *Journal of the American Statistical Association* **96**: 730–745.

Parner, E. (1998). Asymptotic theory for the correlated gamma frailty model, *The Annals of Statistics* **26**: 183–214.

Rudin, W. (1991). *Functional Analysis, 2nd ed.*, McGraw-Hill, New York.

Shih, J. H. and Louis, T. A. (1995). Inferences on the association parameter in copula models for bivariate survival data, *Biometrics* **51**: 1384–1399.

Sklar, A. (1959). Fonctions de répartition à n dimensions et leurs marges, *Publ. Inst. Statist. Univ. Paris* **8**: 229–231.

Tobin, J. (1958). Estimation of relationships for limited dependent variables, *Econometrica* **26**: 24–36.

Tsodikov, A. D., Ibrahim, J. G. and Yakovlev, A. Y. (2003). Estimating cure rates from survival data: an alternative to two-component mixture models, *Journal of the American Statistical Association* **98**: 1063–1078.

van der Vaart, A. W. and Wellner, J. A. (1996). *Weak convergence and empirical processes*, New York, N.Y. : Springer.

Yakovlev, A. Y. and Tsodikov, A. D. (1996). *Stochastic models of tumor latency and their biostatistical applications.*, New Jersey: World Scientific.

Yan, J. (2007). Enjoy the joy of copulas: With a package copula, *Journal of Statistical Software* **21(4)**: 1–21.

Yang, Y. and Simpson, D. (2010). Unified computational methods for regression analysis of zero-inflated and bound-inflated data, *Computational Statistics and Data Analysis* **54**: 1525–1534.

Samenvatting

In sommige klinische, economische of milieustudies zijn onderzoekers geïnteresseerd in een semi-continue uitkomstvariabele Y. Deze uitkomstvariabele neemt de waarde nul aan met een discrete kansmassa en heeft een continue verdeling voor de niet-nul responswaarden. Bijvoorbeeld, in een milieustudie, waarbij onderzoekers geïnteresseerd zijn in de hoeveelheid van een bepaald toxisch metaal in een specifiek ecosysteem, is het mogelijk dat dit metaal niet aanwezig is. Als tweede voorbeeld kunnen we een biologische studie beschouwen over ethanol-geïnduceerde slaap in genetisch geselecteerde muizen. Tijdens deze studie vielen sommige muizen niet in slaap van de toegediende dosis ethanol omdat hun metabolisme de alcohol zeer snel kon afbreken (Markel et al. (1995)). In beide studies kunnen we twee groepen van subjecten onderscheiden. In een eerste groep hebben de subjecten een strikt positieve waarde voor de uitkomstvariabele. Deze subjecten worden de ontvankelijken genoemd. De tweede groep van subjecten reageert niet in de studie en heeft een waarde nul voor de uitkomstvariabele. We noemen hen de onontvankelijken.

In dit proefschrift modelleren we de verdeling van een semi-continue variabele met een mengselmodel:

$$F(y|x,z) = P(Y \leq y|X = x, Z = z) = \pi(\gamma, x) + (1 - \pi(\gamma, x))F_{Y>0}(y|z).$$

We beschouwen een parametrisch regressiemodel $\pi(\gamma, x)$ om de invloed van covariaten op de discrete kans op een nulwaarde voor de uitkomstvariabele te beschrijven. Voor de verdeling $F_{Y>0}(y|z)$ van de niet-nul responswaarden beschouwen we een semi-parametrisch Cox regressiemodel om de invloed van covariaten op deze verdeling te bestuderen. Hiervoor veronderstellen we dat de conditionele risicofunctie de volgende

vorm heeft:

$$\lambda_{Y>0}(t|z) = \lambda(t)g(\beta, z).$$

De functie $\lambda(t)$ is een ongespecifieerde referentie risicofunctie en $g(\beta, z) > 0$ is een parametrisch model.

In sommige studies is het onmogelijk om de uitkomstvariabele Y volledig te observeren en zien we slechts een bovengrens. Wiskundig drukken we dit uit door een tweede positieve, onafhankelijke stochastische variabele C te onderstellen die we de censureringstijd noemen. Voor elk studiesubject observeren we een stochastische variabele T die het maximum is van de tijd tot een gebeurtenis Y en de censureringstijd C. Voorts krijgen we een indicator die aangeeft welke van deze variabelen het grootst is. Subjecten waarvoor de tijd tot een gebeurtenis groter is dan de censureringstijd worden ongecensureerde observaties genoemd. Subjecten waarbij de censureringstijd groter is dan de tijd tot een gebeurtenis worden gecensureerde observaties genoemd. We noemen dit type van gegevens in het algemeen, linksgecensureerde gegevens. De ongecensureerde observaties zijn duidelijk ontvankelijken. Voor de gecensureerde observaties kan men echter geen onderscheid maken tussen onontvankelijke subjecten en ontvankelijke subjecten met een gecensureerde tijd tot een gebeurtenis.

Om de verschillende parameters in het mengselmodel te schatten maken we gebruik van de methode van de maximale aannemelijkheid. De oneindig dimensionale referentie cumulatieve risicofunctie wordt geschat met een een stap-functie, met sprongen op de ongecensureerde geobserveerde tijdstippen. In Hoofdstuk 2 worden enkele technische aspecten besproken die de berekening van de meest aannemelijke schattingen vergemakkelijken. Als resultaten worden de consistentie en de asymptotische normaliteit van de meest aannemelijke schatters bewezen. In een simulatiestudie wordt het gedrag van de schatters voor steekproeven uit verschillende onderliggende modellen en met verschillende steekproefgroottes bestudeerd. Daarnaast wordt het semi-parametrisch regressiemodel voor links-gecensureerde data met extra nullen geïllustreerd aan de hand van de data set over ethanol-geïnduceerde slaap in genetisch geselecteerde muizen.

In de simulatiestudie van Hoofdstuk 2 merken we op dat de optimalisatieprocedure voor het bepalen van de meest aannemelijke schattingen soms onstabiel is en enige tijd vergt. Wanneer bijvoorbeeld de kans op een nulwaarde voor de respons laag is, kunnen er zich optimalisatieproblemen voordoen. Om deze problemen op

te vangen, wordt in Hoofdstuk 3 onderzocht of de aannemelijkheidsfunctie kan benaderd worden, waardoor de optimalisatieprocedure op een efficiëntere manier kan verlopen. Als resultaat tonen we de consistentie van de maximale benaderende aannemelijkheidsschatters aan. We ontwikkelen een efficiënt EM-algoritme om de maximale benaderende aannemelijkheidsschattingen te berekenen. In de M-stap van het algoritme kan men schattingen voor de logistische parameters en voor de effectparameters van het Cox model afzonderlijk van elkaar bekomen. Daarenboven kan gebruik gemaakt worden van standaardfuncties in het statische softwarepakket **R** om deze schattingen te berekenen. De schattingen van de spronggroottes van de referentie cumulatieve risicofunctie kunnen berekend worden uit de schattingen van de effectparameters. In de simulatiestudie zien we dat de voorgestelde benaderende schattingsmethode een competitief alternatief vormt voor de standaard maximale aannemelijksmethode uit Hoofdstuk 2. Het benaderen van de aannemelijkheidsfunctie leidt immers tot een grote dimensieverlaging, hetgeen de optimalisatieprocedure stabieler en sneller maakt.

In de biologische studie over ethanol-geïnduceerde slaap in genetisch geselecteerde muizen worden de muizen op twee verschillende tijdstippen getest. Het is een voorbeeld waarin onderzoekers geïnteresseerd zijn in twee semi-continue uitkomstvariabelen Y_1 en Y_2, die mogelijk afhankelijk van elkaar zijn. In Hoofdstuk 4 wordt de gezamenlijke verdeling van (Y_1, Y_2) gemodelleerd door middel van een bivariaat mengselmodel. We modelleren eerst de marginale kansen op een nulwaarde voor iedere uitkomstvariabele en de marginale verdelingen van de niet-nul responswaarden. We veronderstellen parametrische regressiemodellen, bijvoorbeeld logistische regressiemodellen, voor de marginale kans op een nulwaarde voor iedere meting. De niet-nul gedeeltes van beide uitkomstvariabelen worden gemodelleerd met parametrische of semi-parametrische Cox regressiemodellen. Daarna wordt de afhankelijkheid van de uitkomstvariabelen Y_1 en Y_2 gemodelleerd. Voor de gezamenlijke kans op een nulwaarde voor beide uitkomstvariabelen, modelleren we de dubbelverhouding in een contingentietabel met combinaties van nul en niet-nul responswaarden voor beide metingen. De associatie tussen twee niet-nul responswaarden wordt gemodelleerd door middel van parametrische copulafamilies.

In sommige studies is het onmogelijk om de uitkomstvariabelen Y_1 en Y_2 volledig te observeren en zien we slechts een bovengrens. Beide observaties kunnen dus links-gecensureerd zijn. Wiskundig drukken we dit uit door een koppel stochastische veranderlijken (C_1, C_2), onafhankelijk van (Y_1, Y_2) conditioneel op de covariaten, te

veronderstellen. Voor elk studiesubject observeren we

$$T_1 = \max(Y_1, C_1) \text{ and } \delta_1 = I(Y_1 \geq C_1),$$
$$T_2 = \max(Y_2, C_2) \text{ and } \delta_2 = I(Y_2 \geq C_2).$$

Om de verschillende parameters in de bivariate parametrische en semi-parametrische regressiemodellen voor links-gecensureerde data met extra nullen te schatten, maken we gebruik van de methode van de maximale aannemelijkheid. Men heeft echter geen gesloten vorm voor de meest aannemelijke schatters en de numerieke optimalisatieprocedure wordt moeilijker naarmate het aantal parameters stijgt. De speciale structuur van het model suggereert dat we een schattingsprocedure in twee fasen kunnen beschouwen. In de eerste fase worden de marginale parameters geschat, waarbij de afhankelijkheid van de twee metingen genegeerd wordt. In de tweede fase worden de marginale parameters gefixeerd op hun geschatte waarden uit de eerste fase en wordt de bivariate aannemelijkheidsfunctie gemaximaliseerd om schattingen voor de afhankelijkheidsparameters te bekomen. Deze techniek werd ook gebruikt in onder andere Shih and Louis (1995), Joe (1997) en Chatterjee and Shih (2001). Als resultaat wordt een gepartitioneerde vorm voor de asymptotische variantie-covariantie matrix van de 2-fasen parametrische schatters afgeleid, samen met een jackknife schatter voor deze matrix. De prestaties van de 2-fasen parametrische en semi-parametrische schattingsmethodes worden onderzocht in een simulatiestudie. De resultaten van de simulaties tonen dat de 2-fasen schattingsmethodes goed werken, vooral wanneer de associatie tussen de twee uitkomstvariabelen niet extreem hoog is. Op het einde van Hoofdstuk 4 wordt de data set met de ethanol-geïnduceerde slaaptijden in genetisch geselecteerde muizen verder geanalyseerd door middel van parametrische en semi-parametrische regressiemodellen voor bivariate links-gecensureerde data met extra nullen.

In Hoofdstuk 5, ten slotte, worden algemene conclusies getrokken en worden indicaties gegeven voor mogelijk toekomstig onderzoek.