DOCTORAATSPROEFSCHRIFT

2008 | Faculteit Wetenschappen

Contributions to frailty and copula modelling with applications to clinical trials and dairy cows data

Proefschrift voorgelegd tot het behalen van de graad van
Doctor in de Wetenschappen, richting wiskunde, te verdedigen door:

Goele MASSONNET

Promotor: prof. dr. P. Janssen
Copromotor: prof. dr. T. Burzykowski
prof. dr. L. Duchateau

universiteit hasselt
CENTRUM
VOOR STATISTIEK

D/2008/2451/35

Dankwoord

Graag wil ik iedereen bedanken die gedurende de voorbije zes jaren heeft bijgedragen tot de realisatie van deze thesis.

In het bijzonder dank ik mijn promotor Prof. dr. Paul Janssen. Dankzij de vele gesprekken over dit onderwerp, zijn suggesties en adviezen heb ik veel mogen bijleren. Paul, bedankt voor de aangename samenwerking, de goede begeleiding en de steun.

Ik dank ook mijn copromotoren Prof. dr. Tomasz Burzykowski en Prof. dr. Luc Duchateau voor hun bijdrage aan deze thesis. Tomasz, thanks for your valuable advice and encouragement during the previous years. Luc, bedankt voor de fijne samenwerking en de leerrijke suggesties.

I would like to thank my colleagues at Censtat for the pleasant and stimulating environment. Een speciaal woordje van dank gaat naar Prof. dr. Noël Veraverbeke en Prof. dr. Roel Braekers voor de tijd die zij vrij maakten voor mijn vragen i.v.m. theoretische problemen. Het geven van een tip of nuttige referentie hielp me vaak weer verder. Mijn collega's assistenten ben ik dankbaar voor de aangename samenwerking.

Uiteraard wil ik ook mijn ouders en Ward bedanken voor hun steun tijdens de voorbije jaren. Ward, mama en papa, jullie belangstelling, motiverende woorden en begrip betekenden heel veel voor mij.

Goele Massonnet
Diepenbeek, 8 september 2008

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Introduction

In survival analysis the response of interest is the time from a well-defined time origin to the occurrence of a specific event. Some examples are the time from onset of a disease to death, the time from recovery to the time of recurrence of a disease, the time from purchase to breakdown of a machine. This response is called the survival time, the failure time or the event time. A special difficulty that often occurs in the analysis of survival data is the possibility that some responses are not observed for the full event time. Such incomplete observation of the event time is called (right) censoring. A classical model that is frequently used to model univariate survival data subject to right censoring, is the proportional hazards model (Cox, 1972).

In many studies there is a natural clustering in the data; event times within the same cluster may be correlated. Such data are known as clustered survival data. Clustered survival data are a particular example of multivariate survival data (see, e.g., Klein and Moeschberger, 2003, p.425). In recent years, extensive research on clustered survival data has been carried out. A lot of attention has been paid to frailty models and copula models. In frailty models the cluster effect is a random effect; therefore frailty models are conditional models. A frailty model is a multiplicative hazard model with three compo-

nents: a frailty factor that models the random cluster effect, a baseline hazard function (that can be modelled in a parametric way or that can be left unspecified) and a component that models (in a parametric way) the dependence of the hazard on the covariates.
Copula models are used to model multivariate survival data with small and equal cluster size. In this thesis we study copula models for four-dimensional survival data. In copula models the joint survival function of the four event times in a cluster is modelled as a function, called the copula, of the marginal survival functions of the four event times. The copula determines the type of dependence. The marginal survival functions can be modelled in a parametric, a semi-parametric or a nonparametric way, possibly taking into account the effect of covariates (Shih and Louis, 1995b; Glidden, 2000; Andersen, 2005).
In Section 1.2 we define the basic quantities that are used in survival analysis and we review classical models that can be used to analyse univariate survival data. In Section 1.3 we discuss multivariate survival data, which is the type of data considered in this thesis, in somewhat more detail. The examples, that will be used in future chapters to illustrate the developed methodology, are collected in Section 1.4. Section 1.5 concludes this chapter with a discussion of the thesis objectives.

1.2 Modelling univariate survival data

1.2.1 Survival data: notation and basic quantities

As explained in the introductory section, we consider event times that might be subject to right censoring. For a censored observation, the only information available is that the event time exceeds the censoring time. For instance, a patient is still disease-free at the end of the study so that the time to recurrence of the disease is unknown for this patient or a patient might decide to leave the study before the event of interest has occurred. For these subjects we only know that the unobserved event time is larger than the observed censoring time. We now give a formal description of what we mean by right censored survival data.

Let N denote the total number of subjects. For the jth subject we observe $X_j = \min(T_j, C_j)$, where T_j is the event time for this subject and C_j is the censoring time independent of T_j. Let $\delta_j = I(T_j \leq C_j)$ be the censoring indicator, which is equal to one if the event has been observed and is equal to zero otherwise.

Let f be the probability density function and F the corresponding cumulative distribution function of the (continuous) event time T. The basic quantities used to describe failure time data are the survival function

$$S(t) = 1 - F(t) = P(T > t),$$

which is the probability of surviving beyond time t, and the hazard function

$$\lambda(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t}. \tag{1.1}$$

The hazard function can be interpreted as the instantaneous failure rate, conditional on having survived up to time t. A related quantity is the cumulative hazard function, defined by

$$\Lambda(t) = \int_0^t \lambda(u) du.$$

For continuous event times, these important quantities are linked together by a series of relationships. First, it follows from (1.1) that

$$\lambda(t) = \frac{f(t)}{S(t)}$$

and hence,

$$\lambda(t) = -\frac{d}{dt} \log S(t).$$

After integrating and exponentiating the previous expression, we obtain that the survival function can be rewritten in terms of the cumulative hazard function

$$S(t) = \exp\left\{ -\int_0^t \lambda(u) du \right\} = \exp\left\{ -\Lambda(t) \right\}.$$

The survival function can be estimated in a nonparametric way using the product-limit estimator, proposed by Kaplan and Meier (1958). This estimator

is defined as follows for all values of t in the range of the observed times:

$$\hat{S}(t) = \begin{cases} 1 & \text{if } t < X_1, \\ \prod_{X_j \leq t} \left\{ \frac{r(X_j) - d(X_j)}{r(X_j)} \right\} & \text{if } X_1 \leq t, \end{cases} \tag{1.2}$$

where $r(v)$ is the number still at risk at time v and $d(v)$ is the number of events at time v. For values of t beyond the largest observed time, this estimator is not well defined (see, e.g., Klein and Moeschberger, 2003, Chapter 4, for suggestions of solutions of this problem). The Kaplan-Meier estimator (1.2) is a step function with jumps at the observed event times. The size of the jump at a certain event time $X_j = T_j$ depends on the number of events observed at T_j, as well as on the pattern of the censored event times before T_j.

1.2.2 Proportional hazards model

Assume that $\mathbf{z}_j = (z_{j1}, \ldots, z_{jp})'$ is a p-variate vector of covariates observed for the jth subject. To model the effect of the covariates on survival, we can model the hazard function as a function of the covariates. Two general classes of models that can be used to relate covariate effects to the hazard function, are the family of multiplicative hazards models and the family of additive risks models. The latter family will be discussed in the next section. For the family of the multiplicative hazards models, the hazard function of a subject with covariate vector $\mathbf{z}_j$ is the product of a baseline hazard function $\lambda_0(t)$, corresponding to the hazard function of a subject with covariate information $\mathbf{z}_j$ equal to $\mathbf{0}$, and a non-negative function of the covariates $c\left(\boldsymbol{\beta}'\mathbf{z}_j\right)$, i.e.,

$$\lambda_j(t) = \lambda_0(t) c\left(\boldsymbol{\beta}'\mathbf{z}_j\right).$$

The vector $\boldsymbol{\beta}$ contains the unknown regression coefficients associated with $\mathbf{z}_j$. For the classical choice $c\left(\boldsymbol{\beta}'\mathbf{z}_j\right) = \exp\left(\boldsymbol{\beta}'\mathbf{z}_j\right)$, we have

$$\lambda_j(t) = \lambda_0(t) \exp\left(\boldsymbol{\beta}'\mathbf{z}_j\right). \tag{1.3}$$

This implies that the difference between the logarithm of the hazard function and the logarithm of the baseline hazard function is linear in the risk

(regression) coefficients, i.e.,

$$\log\{\lambda_j(t)\} - \log\{\lambda_0(t)\} = \sum_{m=1}^{p} \beta_m z_{jm}.$$

Model (1.3) is known as the proportional hazards model. Its name comes from the fact that the ratio of the hazard functions of two subjects (say subject j and l), which is given by

$$\frac{\lambda_j(t)}{\lambda_l(t)} = \exp\left\{\sum_{m=1}^{p} \beta_m(z_{jm} - z_{lm})\right\},$$

is constant over time.

A popular choice, when assuming a parametric baseline hazard function, is $\lambda_0(t) = \lambda\rho t^{\rho-1}$, with $\lambda > 0$ and $\rho > 0$. This choice corresponds to a Weibull distribution for the event times. Alternatively, we can leave the baseline hazard function unspecified. Since the model then contains one parametric factor, $\exp\left(\boldsymbol{\beta}'\mathbf{z}_j\right)$, and one factor that is not specified in a parametric way, $\lambda_0(t)$, we call this a semi-parametric model. Assuming that there are no ties among the event times, the parameters in $\boldsymbol{\beta}$ can be estimated by maximising the partial likelihood, introduced by Cox (1972),

$$PL(\boldsymbol{\beta}) = \prod_{j=1}^{N} \left\{\frac{\exp\left(\boldsymbol{\beta}'\mathbf{z}_j\right)}{\sum_{l\in R(X_j)} \exp\left(\boldsymbol{\beta}'\mathbf{z}_l\right)}\right\}^{\delta_j},$$

where $R(X_j)$ is the set of subjects that are still at risk to experience the event at time X_j (risk set). Andersen and Gill (1982) and Fleming and Harrington (1991) show the consistency and the asymptotic normality of the partial likelihood estimator $\hat{\boldsymbol{\beta}}$.

1.2.3 Additive risks model

In the proportional hazards model, the covariates act multiplicatively on the baseline hazard function. In this section, we briefly describe an alternative to the proportional hazards model, namely the additive risks model. Under

the additive risks model, we assume that the hazard function at time t, for a subject with a p-dimensional vector of covariates $\mathbf{z}_j$, is a linear combination of the covariates, that is

$$\lambda_j(t) = \lambda_0(t) + \boldsymbol{\beta}'\mathbf{z}_j.$$

This model is a special case of the additive risks model proposed by Aalen (1980), which allows the regression coefficients to be functions whose values may change over time. More details on the additive risks model are given in Klein and Moeschberger (2003, Chapter 10) and Martinussen and Scheike (2006, Chapter 5).

1.2.4 Proportional odds model

The Cox proportional hazards model specifies that the covariate effect is a multiplicative factor on the baseline hazard function. When the covariates do not depend on time, this implies that the ratio of the hazard functions of subjects with different risk factors is constant. To deal with situations where the covariate effect diminishes over time, the proportional odds model may be useful (see Yang and Prentice, 1999, for more explanation about this). In the proportional odds model, the covariates act multiplicatively on the baseline odds function. Let F_j be the conditional distribution function of T_j given $\mathbf{z}_j$, for $j = 1, \ldots, N$, and let $F_j(t)\{1 - F_j(t)\}^{-1}$ be the odds of the event $\{T_j \leq t\}$. The proportional odds model specifies that the logodds of the event $\{T_j \leq t\}$ is

$$\log\left\{\frac{F_j(t)}{1 - F_j(t)}\right\} = h_0(t) + \boldsymbol{\beta}'\mathbf{z}_j,$$

where $h_0(t)$ is the baseline logodds function at time t corresponding to the logodds function at time t of a subject with covariate information $\mathbf{z}_j$ equal to $\mathbf{0}$. Bennett (1983) studies the estimation of the parameters $\boldsymbol{\beta}$ and of the baseline odds function using a maximum likelihood approach. Murphy *et al.* (1997) show that the maximum profile likelihood estimator for $\boldsymbol{\beta}$, proposed by Bennett (1983), is consistent, asymptotically normal and efficient. Yang and Prentice (1999) propose new inference techniques for the proportional

odds model that avoid the complex calculations of the profile likelihood while retaining good efficiency properties. They propose several classes of regression estimators, such as the pseudo maximum likelihood estimator, martingale residual-based estimators and minimum distance estimators.

1.3 Multivariate survival data

In the previous section we discussed some models that can be used for analysing univariate survival data. A typical feature of univariate survival data is the independence of the survival times. In this thesis we focus on clustered (or multivariate) survival data. Observational units within the same cluster are typically correlated.
In Section 1.4 we give examples of clustered survival data. These examples will be used in the future chapters to demonstrate the developed methodology. A first example is a data set on the time to infection in the four quarters of the udder of dairy cows. In such a study, where the clusters are the cows, we can expect association between the infection times within cows. This is an example where the cluster size is small and balanced; each cluster has four udder quarters, which are the observational units. In the following examples the cluster size differs from cluster to cluster and is larger than four. The second example is a data set on the time to culling for heifers, i.e., for cows which have experienced only one calving. Since the culling policy might differ between the herds, heifers are clustered within the herds. The third example is a multicenter clinical trial, which is a typical example of clustered survival data. The study is an early breast cancer clinical trial, where the event time is the time to death or recurrence. Since this trial is run in different cancer centers over the world, the center is considered as the cluster and the patient is the observational unit within the cluster. In this study, there are a few large centers and a few small centers. The fourth example consists of data from 27 randomised trials in advanced colorectal cancer data. This example is a meta-analysis, with the trial being the cluster and the patients being the observational units within the cluster. The event time is the survival time. In

the following section we give a more detailed description of the data sets.

1.4 Data sets used for illustration

1.4.1 Udder infection data

We consider a data set on mastitis, an infection of the udder of a dairy cow. Mastitis can be caused by many organisms, most of them bacteria, such as *Escherichia coli, Streptococcus uberis,* and *Staphylococcus aureus.* Since each udder quarter is separated from the three other udder quarters, one quarter might be infected while the other quarters are infection-free. In this study, 100 cows are followed up for infections. From each quarter, a milk sample is taken monthly and is screened for the presence of different bacteria. Due to the periodic follow up, the infection time is defined as the average of the time of the last milk sample that indicates that there is no infection and the time of the first milk sample that indicates an infection. Observations can be right censored if no infection occurs before the end of the lactation period, which is roughly 300 days but different for every cow, or if the cow is lost to follow up during the study, for example due to culling. Note that this implies that there is a common censoring time for the four udder quarters of a cow (i.e., for all units in the cluster). We model the time to infection with any bacteria, with cow being the cluster and udder quarter the observational unit within the cluster. The correlation between the infection times of the four udder quarters of a cow is an important parameter to take preventive measures. With high correlation, a lot of attention should be given to the uninfected udder quarters of a cow that has an infected quarter. Further, the difference between front and rear udder quarters has been put forward to explain the difference in infection status (Adkinson *et al.*, 1993). Therefore, we study the effect of a binary covariate indicating the location of the udder quarter (front or rear). This is a covariate at the udder quarter level since it takes different values within the cow. We further include parity as a covariate in the analysis. The parity of a cow is the number of calvings (and therefore the number of lacation periods) that the cow has already experienced. Several studies have

shown that prevalence as well as incidence of intramammary infections increase with parity (Weller *et al.*, 1992). We convert parity into a binary covariate, grouping all the cows with more than one calving in the group of multiparous cows (heifer=0) and grouping all the cows with only one calving in the group of primiparous cows or heifers (heifer=1). The covariate heifer is called a cow level covariate since every udder quarter of the cow takes the same value for this covariate. A subset of the data is presented in Table 1.1.
This data set has provided the motivation to develop the asymptotic properties of the semi-parametric and the nonparametric two-stage estimation approach for four-dimensional copulas, developed in Chapter 5. In Sections 6.3 and 6.5 of Chapter 6, we model the association between the infection times of the four udder quarters using a four-dimensional copula. In the marginal survival functions we model the effect of the location of the udder quarter (front or rear) and the parity. We use a pseudo likelihood ratio test to select an appropriate copula in the power variance copula family that describes the association between the infection times of the four udder quarters.

1.4.2 Time to culling data

The time to culling data are described in De Vliegher *et al.* (2005) and in Duchateau and Janssen (2008, p.10-12). The time to culling during the first lactation of dairy heifer cows is studied as a function of the somatic cell count (SCC) in early lactation. The somatic cell count is measured in the period 5 to 15 days after calving. We use the logarithm of the somatic cell count as a covariate. High somatic cell count in early lactation in heifers is associated with an increased probability of clinical mastitis during the first lactation (De Vliegher *et al.*, 2005). Heifers with infected udder quarters are quite expensive to keep due to the high costs for medical care and the loss in milk production. As a result, elevated somatic cell count in early lactation in heifers might be associated with an increased culling hazard during the first lactation. Dairy heifer cows are followed during the first lactation period (roughly 300-350 days). Cows that are still alive at the end of the lactation period are censored at that time. As already mentioned in Section 1.3, cows are clustered

Table 1.1: Udder infection data. The first column contains the cow identification number. The second column contains the time (in days) to infection. The third column gives the censoring indicator, taking value one (status = 1) if infection is observed and zero (status = 0) otherwise. The fourth column gives the position of the udder quarter (LR = left rear, LF = left front, RR = right rear, RF = right front). The last column gives the parity: multiparous cow (heifer = 0) or primiparous cow (heifer = 1).

Cowid	Time to infection	Status	Quarter	Heifer
1	278.5	0	LR	1
1	278.5	1	LF	1
1	62.5	1	RR	1
1	152.5	1	RF	1
2	317.5	0	LR	0
2	317.5	0	LF	0
2	317.5	0	RR	0
2	317.5	0	RF	0
...				
100	76.5	1	LR	1
100	76.5	1	LF	1
100	76.5	1	RR	1
100	76.5	1	RF	1

within herds since culling policy might differ substantially between the herds. The study includes 13835 heifer cows in total and consists of 3192 herds. The number of heifer cows per herd varies from 1 to 56 (the mean (median) number of heifers per herd is 4.33 (4)). The data for a few heifer cows from the first and the last herd are presented in Table 1.2.

In Chapter 3 we use this example to demonstrate a model-based bootstrap algorithm, proposed in Section 3.2, that can be used to estimate the standard

errors of the parameter estimates in a shared frailty model. In Section 3.4.1 we study the heterogeneity between herds and the effect of the logarithm of the somatic cell count on the culling hazard using a shared frailty model. We estimate the standard errors of the parameter estimates in the shared frailty model using model-based bootstrap. We compare the results of the model-based bootstrap algorithm with the results obtained by using the non-parametric resampling scheme proposed by Therneau and Grambsch (2000).

Table 1.2: Time to culling data. The first column contains the cow identification number. The second column gives the herd to which the cow belongs. The third column contains the time (in days) to culling. The fourth column gives the censoring indicator, taking value one (status = 1) if the cow is culled and zero (status = 0) otherwise. The last column gives the logarithm of the observed somatic cell count.

Cowid	Herd	Time to culling	Status	log(SCC)
1	1	331	0	4.09
2	1	312	0	4.83
3	1	96	1	3.93
...				
13833	3192	317	0	3.93
13834	3192	327	0	6.98
13835	3192	315	0	7.94

1.4.3 Early breast cancer data

This study is an early breast cancer phase III clinical trial from the European Organisation for Research and Treatment of Cancer (EORTC). One of the objectives of this trial is to study if perioperative chemotherapy, when compared with no further treatment, results in an increase in survival for women following potentially curative treatment of carcinoma of the breast. Details of this

trial are described in Clahsen *et al.* (1996). In practice, a multicenter clinical trial typically contains a few large centers and many rather small centers. This trial includes 2793 patients entered by 15 centers. The number of patients per center, sorted according to size, is 6, 19, 25, 40, 48, 53, 54, 60, 78, 184, 185, 206, 311, 622, 902. The endpoint that we consider is disease-free survival, defined as the time from randomisation to time to death or recurrence, whatever comes first. Patients that are still at risk at the end of the study are censored at that time. A subset of the data is presented in Table 1.3.

Table 1.3: Early breast cancer data. The first column contains the patient identification number. The second column gives the number of the center where the patient was treated. The third column contains the disease-free survival time (in days). The fourth column gives the censoring indicator, taking value one (status = 1) if death or recurrence is observed and zero (status = 0) otherwise. The last column indicates the treatment received (perioperative (periop=1) or not (periop=0)).

Patid	Center	Disease-free survival time	Status	Periop
1	1	6195	0	0
2	1	6215	0	0
3	1	6207	0	0
...				
2791	15	3257	0	1
2792	15	2569	0	0
2793	15	2900	0	1

In Chapter 3 we use this example to illustrate the use of bootstrap to estimate the standard errors of the parameter estimates in a shared frailty model. In Section 3.4.2 we use a shared frailty model to study the heterogeneity in the outcomes and to study the effect of perioperative chemotherapy compared

to no further treatment. We estimate the standard errors of the parameter estimates in the shared frailty model using a model-based resampling plan, proposed in Section 3.2, and using a nonparametric bootstrap algorithm proposed by Therneau and Grambsch (2000).

1.4.4 Advanced colorectal cancer data

This study is described in Burzykowski *et al.* (2004). The data come from 27 advanced colorectal cancer trials (Advanced Colorectal Cancer Meta-Analysis Project, 1992, 1994; Meta-Analysis Group in Cancer, 1996, 1998). In the four meta-analyses, the comparison was between an experimental treatment and a control treatment. In total there are 4007 patients, 1871 (46.7 %) in the control group and 2136 (53.3 %) in the experimental group. The number of patients per trial varies from 15 to 382 patients (the mean (median) number of patients per trial is 149 (148)). Our analysis is based on the survival time, defined as the time from randomisation to death from any cause. Most patients have died (3591 out of 4007 patients, i.e., 89.6%). The data for a few patients are given in Table 1.4.
In Chapter 4 we use this example to demonstrate a new approach to fit frailty models. In Section 4.5 we investigate the between-trial variation (heterogeneity) in both the baseline risk and the effectiveness of the experimental treatment. For this purpose, we use a frailty model including a fixed treatment effect, a random trial effect and a random treatment effect. We illustrate that this frailty model can be fitted by transforming the model to a linear mixed-effects model, as proposed in Section 4.2.2.
Note that the covariate "Treatment" is coded with values -1 and 1 (see Table 1.4). In this way we avoid convergence problems due to an ill-conditioned variance-covariance matrix in the SAS procedure PROC MIXED.

1.5 Thesis objectives

The main objective of this thesis is to develop estimation methods and resampling procedures for frailty models and copula models. The proposed methods

Table 1.4: Advanced colorectal cancer data. The first column contains the patient identification number. The second column gives the number of the trial in which the patient is involved. The third column contains the survival time (in days). The fourth column gives the censoring indicator, taking value one (status = 1) if death is observed and zero (status = 0) otherwise. The last column indicates the treatment received (experimental (treatment=1) or control treatment (treatment=-1)).

Patid	Trial	Survival time	Status	Treatment
1	1	341	1	-1
2	1	365	1	1
3	1	690	1	1
...				
32	1	388	0	-1
33	1	197	0	1
...				
4005	27	176	1	-1
4006	27	392	1	-1
4007	27	141	1	-1

are illustrated using examples from clinical trials and veterinary studies on dairy cows, as described in Section 1.4.

In Chapter 2 we give an overview of the models that will be studied in the future chapters and we briefly review the estimation methods that are proposed in the literature to fit these models.

Chapters 3 and 4 focus on frailty models. In Chapter 3 we consider the shared frailty model. This model extends the classical proportional hazards model by adding a multiplicative frailty term that accounts for the cluster effect. We propose two model-based bootstrap algorithms that can be used to estimate the standard errors of the parameter estimates in the shared frailty model

(Section 3.2). We compare the two model-based resampling schemes to a nonparametric resampling plan discussed in Therneau and Grambsch (2000). This comparison is based on a simulation study (Section 3.3). The results presented in Chapter 3 have been published in Massonnet *et al.* (2006).

In Chapter 4 we study a frailty model that extends the shared frailty model and that can be used in the clinical trials context: a frailty model with a random cluster effect and a random treatment effect. Classical estimation methods to fit such a model are likelihood-based. We propose an alternative estimation approach which is based on a model transformation (Section 4.2). Through the transformation, the parameters of interest in the frailty model become the parameters in a related mixed-effects model. We demonstrate that the idea of model transformation can also be used to fit other conditional survival models, such as the multivariate proportional odds model and the multivariate additive risks model (Section 4.3). Based on a simulation study, we evaluate the performance of the proposed estimation method for frailty models (Section 4.4). Massonnet *et al.* (2008a) contain the material discussed in Chapter 4.

In Chapters 5 and 6, we study copula models for four-dimensional survival data. The udder infection data set, described in Section 1.4.1, is the motivating example. In this example, we have correlated infection times in the four udder quarters of dairy cows and we use copulas to model the dependence between the four outcomes. In Chapter 5 we propose a semi-parametric and a nonparametric two-stage estimation procedure for four-dimensional copulas. In the first step of the estimation procedure, we estimate the marginal survival functions. We allow these survival functions to depend on a binary covariate at the cluster level and a binary covariate at the observational unit level. In a second step, we obtain likelihood estimates of the dependence parameter(s) after replacing the survival expressions in the likelihood by their estimated counterparts. We develop the asymptotic properties of the estimators obtained in the first and the second step of the estimation procedure (Sections 5.3 and 5.4). Chapter 5 contains the detailed proofs of the theorems that are given in Massonnet *et al.* (2008b).

In Chapter 6 we consider three copula models that are nested in the power variance copula family: Clayton, positive stable and inverse Gaussian copulas. Each of these three copulas model a different type of dependence. We use a pseudo likelihood ratio test to select a copula in the power variance copula family that provides a good description of the type of dependence between the outcomes in a cluster (Glidden, 2000; Andersen, 2005). We develop a bootstrap algorithm that can be used to obtain the p-value of this test (Section 6.2). This bootstrap algorithm also provides estimates for the standard errors of the estimated copula parameters. We perform simulations to study the type I error rate and the power of the pseudo likelihood ratio test. The simulation setting we use, mimics the parameter characteristics of the udder infection data. The material discussed in Chapter 6 is also in Massonnet *et al.* (2008b).
In Chapter 7 we collect the main conclusions from this thesis and discuss interesting topics for further research.

Chapter 2

Modelling multivariate survival data

2.1 Introduction

Frailty models and copula models are widely used to fit multivariate survival data. Both models provide a way to describe the within cluster dependence of the outcomes. Copula models are typically used to model the joint survival function of clustered data with small and equal cluster size (e.g., bivariate survival data). Frailty models can be used to fit clustered survival data having any balanced or unbalanced cluster size. In Section 2.2 we introduce the shared frailty model and briefly discuss the estimation methods for the shared frailty model. In Section 2.3 we consider a frailty model with a random cluster and a random treatment effect. This model extends the shared frailty model by adding a random treatment effect. We give a short review of classical likelihood-based estimation methods to fit this model. Frailty models are conditional hazards models. Other conditional survival models for clustered survival data are the multivariate additive risks model and the multivariate proportional odds model. In Section 2.4 we describe both models. In Section 2.5 we focus on copula models for quadruples. We summarize some general ideas on copulas (Section 2.5.1) and give the precise definitions

of some specific copulas that are nested in the power variance copula family (Section 2.5.2). In Section 2.5.3 we briefly introduce the two-stage estimation approach for copulas, which will be studied in more detail in Chapter 5.

2.2 The shared frailty model

2.2.1 Model formulation

Assume we have a total of N subjects that come from K different clusters, cluster i having n_i subjects ($N = \sum_{i=1}^{K} n_i$). Each subject is observed from time zero to a failure time T_{ij} or to a potential right censoring time C_{ij} independent of T_{ij}. Let $X_{ij} = \min(T_{ij}, C_{ij})$ be the observed time and δ_{ij} be the censoring indicator which is equal to 1 if $X_{ij} = T_{ij}$ and 0 otherwise. For each subject, we also have a p-variate vector of covariates $\mathbf{z}_{ij} = (z_{ij1}, \ldots, z_{ijp})$. The shared frailty model is given by

$$\lambda_{ij}(t) = \lambda_0(t) \exp(\boldsymbol{\beta}' \mathbf{z}_{ij} + w_i), \tag{2.1}$$

where $\lambda_{ij}(t)$ is the conditional hazard function at time t for the jth subject from the ith cluster (conditional on w_i), $\lambda_0(t)$ is the baseline hazard at time t, $\boldsymbol{\beta}$ is the fixed effects vector of dimension p and w_i is the random effect for cluster i. As in the Cox proportional hazards model for univariate survival data, $\lambda_0(t)$ can be left unspecified or it may be assumed to have some specific parametric form. The w_i's, $i = 1, \ldots, K$, are a sample (independent and identically distributed) from a density f_W.
Model (2.1) can be rewritten as:

$$\lambda_{ij}(t) = \lambda_0(t) u_i \exp(\boldsymbol{\beta}' \mathbf{z}_{ij}). \tag{2.2}$$

The factor $u_i = \exp(w_i)$ is termed the frailty corresponding to the ith cluster. Model (2.2) is called the shared frailty model because subjects in the same cluster all share the same value of the frailty factor. The u_i's, $i = 1, \ldots, K$, are a sample from a density f_U, which is called the frailty density. In this thesis, we consider the following choices for the frailty density:

(a) The one-parameter gamma density of the form

$$f_U(u) = \frac{u^{(1/\theta)-1}\exp(-u/\theta)}{\theta^{1/\theta}\Gamma(1/\theta)}, \quad \theta > 0. \tag{2.3}$$

The corresponding density for W is

$$f_W(w) = \frac{\{\exp(w)\}^{1/\theta}\exp\{-\exp(w)/\theta\}}{\theta^{1/\theta}\Gamma(1/\theta)}.$$

For the gamma density $\mathrm{E}(U) = 1$. Typically $\mathrm{Var}(U) = \theta$ is used to describe the heterogeneity.

(b) The normal density for W with $\mathrm{E}(W) = 0$ and $\mathrm{Var}(W) = \sigma^2$.
The corresponding density of U is the lognormal density:

$$f_U(u) = \frac{1}{u\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{(\log u)^2}{2\sigma^2}\right\}. \tag{2.4}$$

The mean and the variance of the frailty are then given by

$$\begin{aligned}\mathrm{E}(U) &= e^{\sigma^2/2}\\ \mathrm{Var}(U) &= e^{\sigma^2}\left(e^{\sigma^2}-1\right).\end{aligned}$$

Note that the mean of the frailty U is not one if we assume a zero-mean normal distribution for the random effect W. In the further discussion, σ^2 is chosen so that $\mathrm{Var}(U) = \theta$, i.e., $\sigma^2 = \log\left(\frac{1+\sqrt{1+4\theta}}{2}\right)$.

We use $\mathrm{Var}(U) = \theta$ to describe the heterogeneity induced by the two frailty distributions. The gamma and the lognormal distributions are often made choices for the frailty distribution in practice. Most of the software limits the choice of the frailty distributions to these cases. Other frailty distributions which have been used in the literature include the positive stable (Hougaard, 1986b), the inverse Gaussian (Hougaard, 1986a), the power variance function (Hougaard, 1986a; Aalen, 1988) and the compound Poisson (Aalen, 1992). A detailed discussion of these frailty distributions is given in Duchateau and Janssen (2008, Chapter 4).

2.2.2 Methods of estimation for the shared frailty model

The gamma frailty model is often used due to its mathematical convenience that results from the simple form of the Laplace transform that corresponds to (2.3):

$$L(s) = E\left(e^{-sU}\right) = (1+\theta s)^{-1/\theta}.$$

For the gamma frailty model, an explicit expression for the observable (marginal) likelihood can be obtained (Klein, 1992). For the lognormal frailty model, it is not possible to obtain an explicit expression for the Laplace transform. Therefore, estimation methods for this frailty distribution are often based on numerical integration (McGilchrist and Aisbett, 1991; Ripatti and Palmgren, 2000; Vaida and Xu, 2000).

For the gamma frailty model, Klein (1992) shows that the observable (marginal) loglikelihood is given by

$$\begin{aligned} l_{obs}\{\boldsymbol{\beta}, \theta, \lambda_0(.)\} \quad = \quad & \sum_{i=1}^{K} \Bigg[D_i \log\theta - \log\Gamma(\frac{1}{\theta}) + \log\Gamma(\frac{1}{\theta} + D_i) \\ & - \left(\frac{1}{\theta} + D_i\right) \log\{1 + \theta \sum_{j=1}^{n_i} \Lambda_0(x_{ij}) \exp(\boldsymbol{\beta}'\mathbf{z}_{ij})\} \\ & + \sum_{j=1}^{n_i} \delta_{ij}\{\boldsymbol{\beta}'\mathbf{z}_{ij} + \log\lambda_0(x_{ij})\} \Bigg], \end{aligned} \tag{2.5}$$

where $D_i = \sum_{j=1}^{n_i} \delta_{ij}$ is the number of observed events in cluster i.

As noted in the previous section, the baseline hazard $\lambda_0(t)$ in the frailty model can be specified explicitly or left unspecified. Under the parametric assumption, the parameters in the resulting model can be estimated using maximum likelihood estimation procedures. For example, for $\lambda_0(t) \equiv \lambda_0$ constant, the parameters $\boldsymbol{\beta}$, θ and λ_0 can be estimated by maximising the observable loglikelihood $l_{obs}(\boldsymbol{\beta}, \theta, \lambda_0)$. If $\lambda_0(t)$ is left unspecified, the EM algorithm (Klein, 1992) and the penalized partial likelihood approach (Therneau and Grambsch, 2000) can be used to estimate the unknown parameters in (2.5). The latter can also be used to estimate the parameters of the lognormal frailty model (McGilchrist and Aisbett, 1991; McGilchrist, 1993). Ducrocq and Casella (1996) develop a

Bayesian approach, using the Laplace approximation to derive the marginal posterior distribution of the variance of the random effects. We briefly discuss the EM algorithm for the gamma frailty model and the penalized partial likelihood approach for the gamma and the lognormal frailty model, since we will use these estimation methods in Chapter 3.

The EM algorithm for the gamma frailty model

To estimate $\boldsymbol{\xi} = (\theta, \boldsymbol{\beta})$, we would like to base the likelihood maximisation on the observable loglikelihood (2.5). However, this likelihood is difficult to maximise as it contains, apart from $\boldsymbol{\xi}$, also the unspecified baseline hazard. We therefore rely on the EM algorithm to estimate $\boldsymbol{\xi}$ (Klein, 1992); it has been described in detail by Duchateau *et al.* (2002).
It is worth noting that Therneau and Grambsch (2000, p.254) have shown that, for the gamma frailty case, the EM algorithm and the penalized partial likelihood approach lead to the same estimates. Since S-Plus contains a fast algorithm for the penalized partial likelihood approach, this property is very important from a practical point of view.

The penalized partial likelihood for shared frailty models

We consider the shared frailty model (2.1), where the model expression is in terms of the random effects w_i, for $i = 1, \ldots, K$. Define $\mathbf{w} = (w_1, \ldots, w_K)'$. For the estimation of $\boldsymbol{\xi} = (\theta, \boldsymbol{\beta})$, we use the penalized partial likelihood

$$l_{ppl}(\boldsymbol{\xi}, \mathbf{w}) = l_{part}(\boldsymbol{\xi}, \mathbf{w}) - l_{pen}(\boldsymbol{\xi}, \mathbf{w}),$$

where

$$l_{part}(\boldsymbol{\xi}, \mathbf{w}) = \sum_{l=1}^{r} \left[\sum_{t_{ij}=t_{(l)}} \eta_{ij} - d(t_{(l)}) \log \left\{ \sum_{t_{qs} \geq t_{(l)}} \exp(\eta_{qs}) \right\} \right],$$

with $\eta_{ij} = \mathbf{z}'_{ij}\boldsymbol{\beta} + w_i$, r denoting the number of different event times, $t_{(1)} < \ldots < t_{(r)}$ being the ordered event times, $d(t_{(l)})$ denoting the number of events

at time $t_{(l)}$, $l = 1, \ldots, r$, and

$$l_{pen}(\theta, \mathbf{w}) = -\sum_{i=1}^{K} \log f_W(w_i).$$

For random effects w_i, $i = 1, \ldots, K$, with corresponding one-parameter gamma density for the frailties, we have

$$l_{pen}(\theta, \mathbf{w}) = -\sum_{i=1}^{K} \left\{ \frac{w_i - \exp(w_i)}{\theta} \right\} - K \left\{ \frac{\log \theta}{\theta} - \log \Gamma \left(\frac{1}{\theta} \right) \right\}.$$

The maximisation of the penalized loglikelihood consists of an inner and an outer loop. In the inner loop the Newton-Raphson procedure is used to maximise, for a provisional value of θ, $l_{ppl}(\boldsymbol{\xi}, \mathbf{w})$ for $\boldsymbol{\beta}$ and $\mathbf{w}$. In the outer loop, a likelihood similar to (2.5) is maximised for θ as in the case of the EM algorithm. The process is iterated until convergence (for details see, e.g., Duchateau *et al.*, 2002).

For random effects w_i, $i = 1, \ldots, K$, having a normal density, we have

$$l_{pen}(\sigma^2, \mathbf{w}) = \frac{1}{2} \sum_{i=1}^{K} \left\{ \frac{w_i^2}{\sigma^2} + \log(2\pi\sigma^2) \right\}.$$

By reducing the penalized partial likelihood, this term penalizes random effects that are far away from the mean value. The maximisation of the penalized loglikelihood consists of an inner and an outer loop. The inner loop is identical to the one described for gamma frailty parameters. In the outer loop, the restricted maximum likelihood estimator for σ^2 is obtained using BLUP's. The process is iterated until convergence (for details see McGilchrist, 1993; Duchateau *et al.*, 2002).

2.3 Frailty model with random cluster and treatment effects

2.3.1 Model formulation

In this section we consider a frailty model with two random effects within a cluster. A typical example of such a model is a multicenter clinical trial with a

random effect describing the heterogeneity between centers and a second random effect describing the treatment heterogeneity between centers. We use this model to investigate the between-trial variation (heterogeneity) in both the baseline risk and the effectiveness of the experimental treatment in the advanced colorectal cancer data.

In this section we assume that we observe a binary covariate z_{ij}, representing the treatment to which the jth patient in the ith cluster has been randomised, with $z_{ij} = -1$ if the patient is in the control group and $z_{ij} = 1$ if the patient is in the experimental group. We consider a Cox proportional hazards model including a fixed treatment effect, a random cluster effect and a random treatment effect. The conditional hazard for the jth patient in the ith cluster is then given by

$$\lambda_{ij}(t) = \lambda_0(t) \exp\left\{b_{0i} + (\beta + b_{1i}) z_{ij}\right\}, \tag{2.6}$$

where $\lambda_0(t)$ represents the unspecified baseline hazard at time t, β is the fixed overall treatment effect, b_{0i} is the random cluster effect (contributing the factor $\exp(b_{0i})$ to the hazard) and b_{1i} is the random treatment effect providing information on how the treatment effect within cluster i deviates from the overall treatment effect captured by the regression coefficient β. The random effects b_{0i} and b_{1i} are assumed to follow zero-mean normal distributions. The variance-covariance matrix of the vector of random effects $\mathbf{b}^T = (b_{01}, b_{11}, \ldots, b_{0i}, b_{1i}, \ldots, b_{0K}, b_{1K})$ takes the form

$$\mathbf{G} = \mathbf{I}_K \otimes \begin{pmatrix} \sigma_0^2 & \sigma_{01} \\ \sigma_{01} & \sigma_1^2 \end{pmatrix}, \tag{2.7}$$

where $\otimes$ is the Kronecker product. The variance components σ_0^2 and σ_1^2 are a measure of the heterogeneity of the hazard due to the random cluster, resp. random treatment effect; σ_{01} is the covariance between the two random effects within a cluster.

Note that, in absence of a random treatment effect, model (2.6) reduces to the lognormal shared frailty model

$$\lambda_{ij}(t) = \lambda_0(t) \exp(b_{0i} + \beta z_{ij}) = \lambda_0(t) u_i \exp(\beta z_{ij}), \tag{2.8}$$

where $u_i = \exp(b_{0i})$ is the frailty for cluster i. The density of the frailty U is then given by (2.4). In absence of covariates this model further simplifies to

$$\lambda_i(t) = \lambda_0(t) \exp(b_{0i}) = \lambda_0(t) u_i. \tag{2.9}$$

In (2.8) and (2.9) b_{0i}, $i = 1, \ldots, K$, is a sample from a zero-mean normal density with variance σ_0^2, describing the heterogeneity between clusters.

2.3.2 Methods of estimation

The likelihood-based estimation methods for shared frailty models, described in Section 2.2.2, have been adapted to cover the extra complexity of the random treatment effect in model (2.6). Vaida and Xu (2000) propose the use of Markov Chain Monte Carlo methods to obtain the expected values of the frailties in the E-step of the EM-algorithm. Cortiñas Abrahantes and Burzykowski (2005) compute the expected values in the E-step using the Laplace approximation. Ripatti and Palmgren (2000) propose estimation using a penalized partial likelihood approach that is based on Laplace approximation of the marginal likelihood function. Legrand *et al.* (2005) extend the Bayesian approach of Ducrocq and Casella (1996) to a frailty model with a random cluster and a random treatment effect assuming that there is no correlation between the two random effects. For parametric frailty models with non-normal random effects, Liu and Yu (2008) propose a likelihood reformulation method. They reformulate the likelihood conditional on non-normal random effects to that conditional on normal random effects. Their method can be implemented using the SAS procedure PROC NLMIXED.

In Chapter 4 we propose an alternative way to fit frailty models. Using a model transformation, we reformulate the original problem of "fitting a frailty model" into a standard problem of "fitting a linear mixed-effects model". Note that Liu and Yu (2008) assume that the baseline hazard in model (2.6) has a specific parametric form, whereas the estimation methods proposed in Vaida and Xu (2000), Cortiñas Abrahantes and Burzykowski (2005), Ripatti and Palmgren (2000), Legrand *et al.* (2005) and in Chapter 4 leave the baseline hazard unspecified.

2.4 Other conditional survival models

In this section we briefly discuss the multivariate additive risks model and the multivariate proportional odds model. In Chapter 4 we show that the idea of a model transformation to obtain a linear mixed-effects model also works for these conditional survival models. This approach provides a new and alternative way to fit multivariate additive risks models and multivariate proportional odds models.

2.4.1 The multivariate additive risks model

The multivariate additive risks model extends the additive risks model, described in Section 1.2.3, by adding a random cluster effect b_{0i} and a random treatment effect b_{1i}:

$$\lambda_{ij}(t) = \lambda_0(t) + b_{0i} + (\beta + b_{1i})z_{ij}, \tag{2.10}$$

where $\lambda_{ij}(t)$ is the hazard function at time t for the jth subject in the ith cluster, $\lambda_0(t)$ is the baseline hazard function at time t, z_{ij} is a binary covariate as defined in Section 2.3.1 and β is the overall fixed treatment effect. As in Section 2.3.1, we assume that b_{0i} and b_{1i} follow zero-mean normal distributions. The variance-covariance matrix of the vector of random effects $\mathbf{b}^T = (b_{01}, b_{11}, \ldots, b_{0i}, b_{1i}, \ldots, b_{0K}, b_{1K})$ is given by (2.7).
Yin and Cai (2004) and Martinussen and Scheike (2006) consider marginal additive risks models for multivariate survival data. The study of conditional additive risks models for multivariate survival data, as given in (2.10), seems open.

2.4.2 The multivariate proportional odds model

Lam *et al.* (2002) and Lam and Lee (2004) extend the proportional odds model, described in Section 1.2.4, to multivariate survival data by incorporating random effects in the model. The multivariate proportional odds model is given

by

$$\log\left\{\frac{F_{ij}(t)}{1-F_{ij}(t)}\right\} = h_0(t) + b_{0i} + (\beta + b_{1i})z_{ij},$$

where $F_{ij}(t)$ is the conditional distribution function at time t for the jth subject in the ith cluster, $h_0(t)$ is the baseline logodds function at time t, b_{0i}, b_{1i}, β and z_{ij} are as defined in Section 2.3.1. Lam *et al.* (2002) and Lam and Lee (2004) study multivariate proportional odds models using a marginal likelihood approach.

2.5 Copula models

2.5.1 Definitions and properties

Copula models are typically used to model the joint survival function of clustered data with small and equal cluster size. The copula approach is often used for bivariate data (see, e.g., Shih and Louis, 1995b; Andersen, 2005; Duchateau and Janssen, 2008). We study copula models for four-dimensional failure time data. Our motivation for this is a data set on the correlated infection times in the four udder quarters of dairy cows, presented in Section 1.4.1. Let $(T_{i1}, T_{i2}, T_{i3}, T_{i4})$ be a quadruple of failure times of the observational units in cluster i and let $S_{ij}, j = 1, \ldots, 4$, be the marginal survival function of T_{ij}, where the index ij is used to indicate that the marginal survival function may depend on a covariate vector $\mathbf{z}_{ij}$. The survival copula is the function that links the marginal survival functions S_{ij} to generate the joint survival function, i.e.,

$$S_i(t_1, t_2, t_3, t_4; \boldsymbol{\zeta}) = C_{\boldsymbol{\zeta}}\left\{S_{i1}(t_1), S_{i2}(t_2), S_{i3}(t_3), S_{i4}(t_4)\right\} \tag{2.11}$$

for a four-dimensional distribution function $C_{\boldsymbol{\zeta}}$ defined on $(v_1, \ldots, v_4) \in [0,1]^4$, taking values in $[0,1]$ and having uniform marginals. $C_{\boldsymbol{\zeta}}$ is called a survival copula with parameter vector $\boldsymbol{\zeta}$. The existence (and uniqueness if the marginal survival functions are all continuous) follows from Sklar's theorem (Sklar, 1959). See Nelsen (2006) for an in-depth discussion on copulas. For clustered survival data the family of Archimedean copulas (Genest and MacKay,

1986) received considerable attention. Archimedean copulas are defined as

$$C_{\boldsymbol{\zeta}}(v_1, v_2, v_3, v_4) = \varphi^{-1}\{\varphi(v_1) + \varphi(v_2) + \varphi(v_3) + \varphi(v_4)\}, \tag{2.12}$$

where $v_j \in [0,1]$, $j = 1, \ldots, 4$, $\varphi : [0,1] \to [0,\infty]$ is a continuous strictly decreasing function such that $\varphi(0) = \infty$, $\varphi(1) = 0$; and φ^{-1} is completely monotonic on $[0,\infty)$, i.e. φ^{-1} is continuous on $[0,\infty)$ and $(-1)^k \frac{d^k}{ds^k}\varphi^{-1}(s) \geq 0$ for all $s \in (0,\infty)$ and $k = 0, 1, 2, \ldots$ (Nelsen, 2006, p.151-152). The function φ is called a generator of the copula.

There is a natural link between Archimedean copulas and shared frailty models, introduced in Section 2.2. To see this, take again a cluster of size four. Starting from the shared frailty model (2.2), the joint survival function can be obtained by integrating out the frailty in the conditional four-dimensional survival function. Given a frailty density f_U, having Laplace transform $L(s) = \mathrm{E}\{\exp(-sU)\}$, the joint survival function that corresponds to the shared frailty model takes the form

$$S_i(t_1, t_2, t_3, t_4) = L\left[\sum_{j=1}^{4} L^{-1}\left\{S_{ij,f}(t_j)\right\}\right],$$

with $S_{ij,f}, j = 1, \ldots, 4$, the marginal survival functions obtained by integrating out the frailty in the conditional survival distribution (obtained from the conditional hazard) corresponding to the jth observational unit. Note that the joint survival function takes the form of an Archimedean copula, where the generator φ is the inverse of the Laplace transform of the frailty density f_U. However, the marginal survival functions are modelled in a different way. Indeed, the marginal survival functions $S_{ij,f}$ depend on the parameter(s) of the frailty density. This implies that the modelling of the marginal survival functions and the modelling of the within cluster dependence cannot be separated when using frailty models. This is different for copulas. A nice feature of copulas is that the modelling of the marginal survival functions and the modelling of the within cluster dependence is separated. For a detailed discussion, see Goethals *et al.* (2008). What we do learn, however, from this short discussion, is that the inverse of Laplace transforms can be used to define generators for Archimedean copulas.

2.5.2 Power variance copula family

In this section we consider the large family of the power variance survival copulas. This family is generated by the inverse of the Laplace transform of the power variance function distributions. The latter family of distributions is a three parameter family (μ, θ, ν) that includes the gamma distribution, the inverse Gaussian distribution and the positive stable distribution (Hougaard, 1986b). These three families of distributions are obtained for specific choices of the parameters.

The Laplace transform of the power variance function family is given by (Aalen, 1992)

$$L(s) = \exp\left[\frac{\nu}{\theta(1-\nu)}\left\{1-\left(1+\frac{\theta\mu s}{\nu}\right)^{1-\nu}\right\}\right], \tag{2.13}$$

with $\mu > 0$, $\theta \geq 0$ and $0 \leq \nu \leq 1$. The parameter μ corresponds to the mean and $\mu^2\theta$ is the variance. The generator of the power variance survival copula is $\varphi(v_j) = L^{-1}(v_j), j = 1, \ldots, 4$ in (2.12). The power variance survival copula is then given by (see also Andersen, 2005)

$$\begin{aligned} & C_{\nu,\theta}(v_1, v_2, v_3, v_4) \\ = \quad & \exp\left[\frac{\nu}{\theta(1-\nu)}\left[1-\left\{\sum_{j=1}^{4}\left(1+\theta(1-\frac{1}{\nu})\log(v_j)\right)^{\frac{1}{1-\nu}}-3\right\}^{1-\nu}\right]\right]. \end{aligned} \tag{2.14}$$

Note that the parameter μ disappears in the functional form of the copula. This means that μ does not influence the dependence structure.

For $\nu = 0.5$ and $0 \leq \theta < \infty$, the Laplace transform (2.13) reduces to the Laplace transform of the inverse Gaussian distribution with mean μ and variance $\mu^2\theta$, i.e., $L(s) = \exp\left[\frac{1}{\theta}\left\{1-(1+2\theta\mu s)^{1/2}\right\}\right]$. If ν tends to one and $0 \leq \theta < \infty$, we obtain the Laplace transform of the gamma distribution with mean μ and variance $\mu^2\theta$, i.e., $L(s) = (1+\theta\mu s)^{-1/\theta}$. The corresponding copulas are the inverse Gaussian copula and the Clayton copula.

The inverse Gaussian survival copula has the following form:

$$C_\theta(v_1, v_2, v_3, v_4) = \exp\left[\frac{1}{\theta} - \left[\frac{1}{\theta^2} + \sum_{j=1}^{4}\log(v_j)\left\{\log(v_j) - \frac{2}{\theta}\right\}\right]^{1/2}\right]. \tag{2.15}$$

The Clayton survival copula is given by

$$C_\theta(v_1, v_2, v_3, v_4) = \left(v_1^{-\theta} + v_2^{-\theta} + v_3^{-\theta} + v_4^{-\theta} - 3\right)^{-1/\theta}.$$

To obtain the positive stable distribution from the family of power variance function distributions, we need that the mean μ and the variance $\mu^2\theta$ tend to infinity (Hougaard, 1986b). We set

$$\mu = \left(\frac{\theta}{\nu}\right)^{\nu/(1-\nu)} (1-\nu)^{1/(1-\nu)}. \tag{2.16}$$

For $0 \leq \nu < 1$, (2.16) implies that if $\theta \to \infty$, then also $\mu \to \infty$. Using asymptotic arguments it follows that, for θ tending to infinity, the Laplace transform of the power variance family (depending on parameters ν and θ) reduces to the Laplace transform of a positive stable distribution with parameter $1-\nu$, i.e., $L(s) = \exp\left(-s^{1-\nu}\right)$. The copula that is generated by the inverse of this Laplace transform is the positive stable copula:

$$C_\nu(v_1, v_2, v_3, v_4) = \exp\left[-\left[\sum_{j=1}^{4} \{-\log(v_j)\}^{1/(1-\nu)}\right]^{1-\nu}\right].$$

The Clayton, positive stable and inverse Gaussian copulas model different types of dependence. The Clayton copula models late dependence in time, the positive stable copula models early dependence in time, whereas the inverse Gaussian copula takes a position in between. A more detailed discussion that illustrates this based on contour lines can be found in Hougaard (2000) and in Duchateau and Janssen (2008, p. 188).

2.5.3 Estimation method

Since the marginal survival functions in the copula model (2.11) do not depend on the parameters of the copula, the marginal survival functions and the dependence parameter(s) $\boldsymbol{\zeta}$ can be estimated separately. This is the idea of the two-stage estimation approach (see, e.g., Genest *et al.*, 1995; Shih and Louis, 1995b; Glidden, 2000; Andersen, 2005). In the first stage, the marginal

survival functions are estimated. In the second stage, we estimate the copula parameter $\boldsymbol{\zeta}$ by maximising a loglikelihood function in which the marginal survival functions are replaced by their estimates obtained in the first stage. In Chapter 5, we give a detailed discussion on a semi-parametric and a non-parametric two-stage estimation approach to fit copula models for quadruples in the presence of fixed binary covariates.

Chapter 3

Resampling plans for shared frailty models

3.1 Introduction

As mentioned in Section 2.2.2, one of the estimation methods that can be used to fit shared frailty models is the EM algorithm (Klein, 1992). It has been described in detail by Duchateau and Janssen (2008, Chapter 5). The EM algorithm provides estimates for the fixed effects and for the variance of the frailty density, but does not automatically provide estimates for the variances of these estimates. Klein and Moeschberger (2003, p.433) show how the standard errors of the estimates for the shared gamma frailty model can be obtained from the inverse of the observed information matrix. This information matrix is a square matrix which has rank equal to the number of distinct event times plus the number of covariates plus one (for the heterogeneity parameter). For large data sets, this procedure is not appropriate because of the high dimensionality.

As mentioned in Section 2.2.2, Therneau and Grambsch (2000, p.254) prove that, for the shared gamma frailty model, the estimates obtained from the penalized partial likelihood maximisation coincide with the estimates obtained from the EM algorithm. Hence we can use the fast algorithm for the penalized

partial likelihood procedure available in S-Plus. However, the standard error estimates of $\hat{\boldsymbol{\beta}}$ reported by S-Plus are computed under the assumption of θ known (Therneau and Grambsch, 2000, p.249). Since θ needs to be estimated, the given standard errors are too small. Further, S-Plus does not provide an estimate for the standard error of the heterogeneity parameter estimate. Thus, the issue of estimating the standard errors of the parameter estimates in the shared frailty model requires further investigation. An alternative approach for finding variance estimates might be provided by the bootstrap. The results developed for resampling in linear mixed models show that resampling schemes need to be chosen in a careful way (Davison and Hinkley, 1997, p. 100-102; Morris, 2002). Therneau and Grambsch (2000, p.249) propose a nonparametric bootstrap algorithm to obtain standard error estimates. For parametric frailty models, model-based resampling schemes might be preferred above nonparametric resampling plans. In Section 3.2 we propose two model-based resampling plans that can be used to find standard errors of the estimated parameters. In Section 3.3 we compare the two model-based bootstrap algorithms to the nonparametric resampling algorithm of Therneau and Grambsch (2000) based on a simulation study. We illustrate the nonparametric and model-based resampling schemes using the time to culling data and the early breast cancer data in Section 3.4. In Section 3.5 we collect main conclusions and topics for further research.

3.2 Bootstrap: resampling schemes

In this section we use resampling techniques to obtain estimates for the standard errors of the parameter estimates in the shared frailty model. Therneau and Grambsch (2000, p.249) propose the following nonparametric bootstrap technique to obtain standard error estimates:

1. A bootstrap sample is obtained by choosing K clusters by sampling with replacement from the K clusters in the study.

2. Fit a gamma (or lognormal) frailty model with covariates to this boot-

strap sample.

This procedure is repeated a number of times. The estimates of the coefficients $\hat{\boldsymbol{\beta}}^*$ and the estimates of the heterogeneity parameter $\hat{\theta}^*$ are stored for each bootstrap sample. The standard errors of the estimated parameters $\hat{\boldsymbol{\beta}}$ and $\hat{\theta}$ are calculated based on the variability of the different values of $\hat{\boldsymbol{\beta}}^*$ and $\hat{\theta}^*$.
If a parametric model is appropriate, we might prefer model-based resampling techniques above the nonparametric resampling plan. We therefore propose two model-based resampling schemes.
We rely on a resampling plan for a simple random effects model with a balanced design, proposed by Davison and Hinkley (1997, p.102). A random effects model can be written as

$$y_{ij} = \mu_i + \epsilon_{ij}, \quad j = 1, \ldots, n_i = n, \quad i = 1, \ldots, K,$$

where K is the number of groups, $n_i = n$ is the number of subjects per group, the μ_i's are randomly sampled from F_μ and independent of the ϵ_{ij}'s, which are randomly sampled from F_ϵ with $E(\epsilon) = 0$ to force uniqueness of the model.
In the "naive" version of their algorithm, Davison and Hinkley (1997, p.102) define

$$\hat{\mu}_i = \bar{y}_{i\cdot} \quad \text{and} \quad \hat{\epsilon}_{ij} = y_{ij} - \bar{y}_{i\cdot},$$

where $\bar{y}_{i\cdot} = n^{-1} \sum_{j=1}^{n} y_{ij}$.
The resampled data set is then obtained in the following way

1. Choose $\mu_1^*, \ldots, \mu_K^*$ by randomly sampling with replacement from $\hat{\mu}_1, \ldots, \hat{\mu}_K$;

2. Choose $\epsilon_{i1}^*, \ldots, \epsilon_{in}^*$ randomly with replacement from one group of residuals $\hat{\epsilon}_{k1}, \ldots, \hat{\epsilon}_{kn}$, either from a randomly selected group or the group corresponding to μ_i^*;

3. Set $y_{ij}^* = \mu_i^* + \epsilon_{ij}^*, \quad j = 1, \ldots, n, \quad i = 1, \ldots, K.$

To construct a resampling plan for frailty models, we can argue that sampling from the means of the groups in the case of the random effects model is like

sampling from the frailty estimates in the case of the frailty model. However, in the situation of frailty models, we do not have any residuals to resample from. Hjort (1985) (see also Davison and Hinkley, 1997, p.351) proposes a bootstrap algorithm for the Cox proportional hazards model in the case of univariate survival data. We extend the resampling algorithm of Hjort (1985) to a resampling scheme for shared frailty models in the context of multivariate survival data.

3.2.1 Model-based bootstrap, algorithm 1

For $j = 1, \ldots, n_i$, $i = 1, \ldots, K$,

1. Fit the model; obtain the estimate $\hat{\beta}$ and the estimates of the frailties $\hat{u}_1, \ldots, \hat{u}_K$.

2. Choose $u_1^*, \ldots, u_K^*$ by sampling with replacement from $\hat{u}_1, \ldots, \hat{u}_K$.

3. Generate the true failure time T_{ij}^* from the estimated failure time survival function $\hat{S}_{ij}(t) = \{\hat{S}_0(t)\}^{u_i^* \exp(\hat{\boldsymbol{\beta}}' \mathbf{z}_{ij}^*)}$, where $\mathbf{z}_{ij}^*$ is the vector of covariates recorded for the jth individual from the cluster that corresponds to u_i^*.

4. Let $\tilde{C}_{ij}^*$, $\tilde{\delta}_{ij}^*$ and $\tilde{X}_{ij}^*$ be the censoring time, the censoring indicator and the observed time for the jth individual from the cluster that corresponds to u_i^*. If $\tilde{\delta}_{ij}^* = 0$, set $C_{ij}^* = \tilde{X}_{ij}^*$, and if $\tilde{\delta}_{ij}^* = 1$, generate C_{ij}^* from the conditional censoring distribution given that $\tilde{C}_{ij}^* > \tilde{X}_{ij}^*$, namely

$$\frac{\hat{G}(t) - \hat{G}(\tilde{X}_{ij}^*)}{1 - \hat{G}(\tilde{X}_{ij}^*)},$$

   where $\hat{G}$ is an estimate (e.g., Kaplan-Meier) of the common censoring distribution G. Assume that G is independent of the covariates.

5. Set $X_{ij}^* = \min(T_{ij}^*, C_{ij}^*)$, with $\delta_{ij}^* = 1$ if $X_{ij}^* = T_{ij}^*$ and zero otherwise.

Steps 3, 4 and 5 are the adaption of the algorithm proposed by Hjort (1985) (see also Davison and Hinkley, 1997, p.351).

For a semi-parametric model, the true failure times in step 3 are generated from the estimated failure time survival function

$$\hat{S}_{ij}(t) = \{\hat{S}_0(t)\}^{u_i^* \exp(\hat{\boldsymbol{\beta}}' \mathbf{z}_{ij}^*)},$$

where $\hat{S}_0(t) = \exp(-\hat{\Lambda}_0(t))$ is the estimated baseline survival function, with

$$\hat{\Lambda}_0(t) = \sum_{t_{(l)} \leq t} \hat{\lambda}_{l0},$$

where $\hat{\Lambda}_0(t)$ is the estimated baseline cumulative hazard at time t and

$$\hat{\lambda}_{l0} = \frac{d(t_{(l)})}{\sum_{t_{sq} \geq t_{(l)}} u_s^* \ \exp(\hat{\boldsymbol{\beta}}' \mathbf{z}_{sq}^*)}.$$

For a parametric model, the true failure times are generated under the parametric assumption.

For mixed models it has been demonstrated (Morris, 2002) that the variances of the BLUP's are biased downwards as estimators of the variance components. Due to this bias, bootstrapping BLUP's results in underestimation of the variation in the data, causing standard error estimates biased downwards. The above-mentioned model-based resampling algorithm suffers from this problem. This will be illustrated in Section 3.3.4. Therefore, we propose a second resampling scheme, where resampled frailty parameters are obtained by sampling from the appropriate frailty distribution with variance $\hat{\theta}$. We again assume that censoring is independent of the covariates.

3.2.2 Model-based bootstrap, algorithm 2

For $j = 1, \ldots, n_i$, $i = 1, \ldots, K$,

1. Fit the model; obtain the estimates $\hat{\boldsymbol{\beta}}$, $\hat{\theta}$.
2. Sample $u_1^*, \ldots, u_K^*$ from a gamma or lognormal distribution with variance $\hat{\theta}$.
3. Generate the true failure time T_{ij}^* from the estimated failure time survival function $\hat{S}_{ij}(t) = \{\hat{S}_0(t)\}^{u_i^* \exp(\hat{\boldsymbol{\beta}}' \mathbf{z}_{ij})}$.

4. If $\delta_{ij} = 0$, set $C^*_{ij} = X_{ij}$, and if $\delta_{ij} = 1$, generate C^*_{ij} from the conditional censoring distribution given that $C_{ij} > X_{ij}$, namely

$$\frac{\hat{G}(t) - \hat{G}(X_{ij})}{1 - \hat{G}(X_{ij})}.$$

5. Set $X^*_{ij} = \min(T^*_{ij}, C^*_{ij})$, with $\delta^*_{ij} = 1$ if $X^*_{ij} = T^*_{ij}$ and zero otherwise.

3.3 Simulations

3.3.1 Motivation

Based on simulations we compare the two model-based resampling plans and the nonparametric resampling plan. As simulation model we consider the setting of a multicenter clinical trial. The following issues will be discussed:

(i) The comparison of the nonparametric and the model-based resampling schemes assuming that the model is correct.

(ii) The effect of the size of the multicenter clinical trial on the precision of the variance estimation. Note that the size of a trial is determined by K, the number of centers, and by the number of patients per center (which we assume to be equal over the centers for simplicity).

(iii) The effect of the size of θ, the heterogeneity parameter, and $\lambda_0(t)$, the event rate (assumed to be constant in time for simplicity) on the precision of the variance estimation.

(iv) The robustness of the resampling plans to misspecification of the model.

3.3.2 The simulation setting

For each specific setting $(K, n, \lambda_0, \theta, \beta)$, with β the treatment effect parameter, 100 data sets are generated assuming a constant baseline hazard. Given a particular setting, the observations for each data set are generated in the following way. First, K frailties $u_1, \ldots, u_K$ are generated from a gamma (or

lognormal) frailty distribution with variance θ; both frailty distributions are introduced in Section 2.2.1. The time to event for the jth patient from center i is randomly generated from an exponential distribution with parameter $\lambda_{ij} = \lambda_0 u_i \exp(\beta z_{ij})$, where z_{ij} is generated from a Bernoulli distribution with success parameter 0.5. The censoring time for each patient is randomly generated from a uniform distribution so that approximately 30% censoring is obtained.

For each simulated data set, two model assumptions are considered to investigate the performance of the bootstrap algorithms under the correct and misspecified models.

First, we assume that the frailties are gamma distributed. For each simulated data set, $R = 100$ bootstrap samples are taken by using the nonparametric bootstrap and the two model-based resampling plans under the assumption of gamma distributed frailties. In the nonparametric resampling plan, we consider a semi-parametric gamma frailty model to estimate the treatment effect β and the heterogeneity parameter θ, as proposed by Therneau and Grambsch (2000). The penalized partial likelihood approach, discussed in Section 2.2.2, is used to obtain the parameter estimates (Therneau and Grambsch, 2000). In the two model-based resampling plans, we consider both a semi-parametric gamma frailty model and a parametric gamma frailty model with a constant baseline hazard. For the parametric gamma frailty model, the model-based resampling schemes assume that the time to event follows an exponential distribution with parameter λ_{ij}. Under this assumption, the parameters β, θ and λ_0 can be estimated by maximising the observable loglikelihood $l_{obs}(\beta, \theta, \lambda_0)$, given in (2.5), using the Newton-Raphson method.

Second, we assume that the frailties are lognormal distributed. For each simulated data set, $R = 100$ bootstrap samples are taken by using the nonparametric and the two model-based resampling plans under the assumption of lognormal distributed frailties. A semi-parametric lognormal frailty model is considered to estimate the treatment effect and the heterogeneity parameter in the nonparametric and in the two model-based resampling schemes. We again use the penalized partial likelihood approach to estimate the parame-

ters. Note that we do not consider a parametric lognormal frailty model in the model-based resampling plans since it is not possible to obtain an explicit expression of the observable likelihood, as mentioned in Section 2.2.2.

3.3.3 Choice of the parameters

For the concrete simulation, the number of centers is taken equal to 15 or 30 centers, with 20 or 40 patients per center. For "true" frailties that are gamma distributed, we additionally consider 15 or 30 centers, with 5 patients per center. The parameter values λ_0, β and θ are chosen in such a way that a different magnitude of spread in the median time to event from center to center is induced. The median time to event T_{M_0} (for $z_{ij} = 0$) and T_{M_1} (for $z_{ij} = 1$) is the solution of $\exp(-\lambda_0 U T_{M_0}) = 0.5$ and $\exp(-\lambda_0 U \exp(\beta) T_{M_1}) = 0.5$, with U one-parameter gamma distributed, i.e. $T_{M_0} = \frac{\log 2}{\lambda_0 U}$ and $T_{M_1} = \frac{\log 2}{\lambda_0 U \exp(\beta)}$. The magnitude of spread in the median time to event from center to center was determined by computing the density functions of T_{M_0} and T_{M_1} (Figure 3.1). It can be shown that, for $z_{ij} = 1$ and for a gamma frailty density, the density function $f_{T_{M_1}}(t)$ is given by

$$f_{T_{M_1}}(t) = \left\{ \frac{\log 2}{\theta \lambda_0 \exp(\beta)} \right\}^{\frac{1}{\theta}} \frac{1}{\Gamma(1/\theta)} \left(\frac{1}{t} \right)^{1+1/\theta} \exp\left\{ -\frac{\log 2}{\theta t \lambda_0 \exp(\beta)} \right\}.$$

For the treatment effect, we use $\beta = 0.25$. As true values for the event rate, we take $\lambda_0 = 0.1$ and $\lambda_0 = 0.5$. The heterogeneity parameter is set at $\theta = 0.1$ and $\theta = 0.6$.

For the settings $(\theta,\ \lambda_0) = (0.6,\ 0.5)$ and $(0.1,\ 0.5)$, there is little spread in the median time to event over the centers, with a bigger spread for $\theta = 0.6$. For the settings $(\theta,\ \lambda_0) = (0.6,\ 0.1)$ and $(0.1,\ 0.1)$, there is much spread in the median time to event over the centers. Furthermore, Figure 3.2 clearly explains our motivation for choosing $\theta = 0.1$ and $\theta = 0.6$. For $\theta = 0.1$ we have a situation where the gamma and the lognormal density functions are close, whereas for $\theta = 0.6$ these densities are more apart.

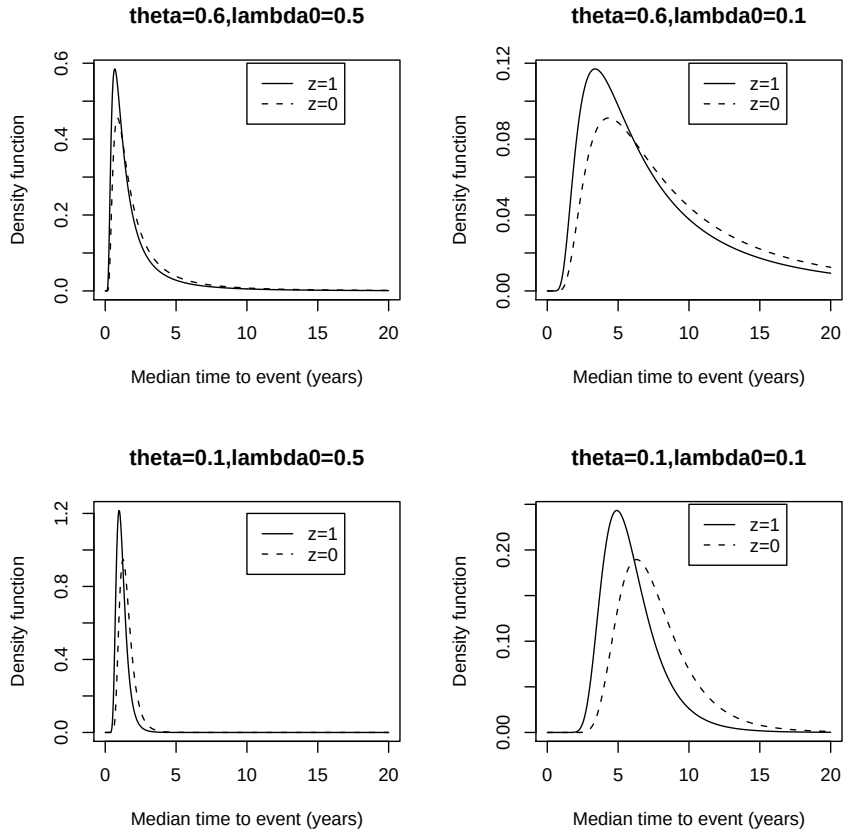

Figure 3.1: Density function of the median time to event over centers ($\beta = 0.25$).

3.3.4 Results

By performing the bootstrap, we obtain for each simulated data set a bootstrap estimate of the standard error of the treatment effect and the heterogeneity parameter. The mean of these 100 estimated standard errors is denoted by mean(SE^B). The values of mean(SE^B) for each resampling scheme are compared to the empirical standard error of $\hat{\beta}$ and $\hat{\theta}$, denoted by SE^E.
In the following discussion we focus on the standard error estimates of the heterogeneity. For completeness, the results for the treatment effect are given

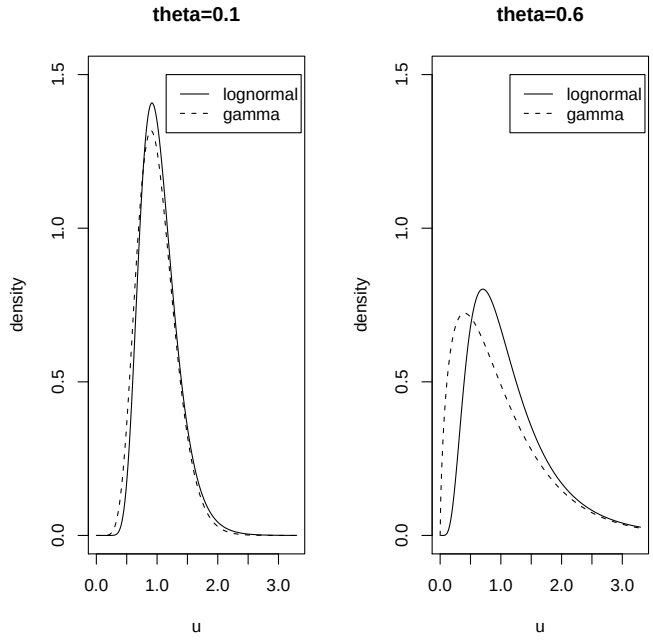

Figure 3.2: Density function for the lognormal and the gamma distribution.

in Tables 3.4 and 3.6 (end of this chapter). In almost all settings studied, the estimated standard error of the heterogeneity parameter obtained by the first model-based resampling plan underestimates the standard error, as compared to SE^E (see, e.g., Tables 3.3 and 3.5 at the end of this chapter). Since the estimates obtained by the second model-based resampling plan give in most cases a more precise assessment of the empirical variability of the parameter estimates than those obtained by the first model-based bootstrap algorithm, we only consider the results of the second model-based and the nonparametric resampling plan in the following discussion.

Nonparametric versus model-based resampling

Figures 3.3 and 3.4 are used to compare the nonparametric and the second model-based resampling plan assuming that the model is correct. In Figure 3.3 we consider "true" frailties that are gamma distributed. In the re-

sampling schemes we assume a semi-parametric gamma frailty model and use the penalized partial likelihood approach to estimate the parameters (gamma semi-par. in Table 3.3). Recall from Section 3.3.2 that we do not consider a parametric gamma frailty model in the nonparametric resampling scheme. Therefore, it is not possible to compare model-based and nonparametric resampling for a parametric gamma frailty model.

The resampling schemes are compared in terms of the absolute relative bias. Take, e.g., Figure 3.3 for the setting $(\theta, \lambda_0) = (0.6, 0.5)$. In that picture we plot for the settings $(K, n) = (15, 5), (15, 20), (15, 40), (30, 5), (30, 20)$ and $(30, 40)$ the points (RB_N, RB_{MB}) where RB_N, resp. RB_{MB}, is the absolute relative bias

$$|\text{mean}(SE^B) - SE^E|/SE^E$$

for the nonparametric, resp. the second model-based, resampling scheme. The actual value for, e.g., $(K, n) = (15, 40)$ and $(30, 40)$ can be obtained from Table 3.3. In the picture we add the bisector. Points (RB_N, RB_{MB}) that are below the bisector correspond to settings for which the absolute relative bias for the model-based resampling scheme is smaller than the absolute relative bias for the nonparametric resampling scheme. Points (RB_N, RB_{MB}) that are above the bisector indicate that nonparametric resampling performs better for the corresponding setting. Figure 3.4 is the equivalent of Figure 3.3 for "true" frailties that are lognormal (logn. semi-par. in Table 3.5).

There is no single consistent pattern for all settings in the results. We first compare the results of the nonparametric and the model-based bootstrap algorithms for "true" frailties that are gamma distributed (Figure 3.3). When $(\theta, \lambda_0) = (0.1, 0.1)$, model-based resampling has a smaller relative bias compared to the nonparametric resampling plan (i.e., most of the points (RB_N, RB_{MB}) are below the bisector), even if the cluster size is small $(n = 5)$. For $(\theta, \lambda_0) = (0.6, 0.5)$, $(0.1, 0.5)$ and $(0.6, 0.1)$, the general conclusion from Figure 3.3 is that, unless the cluster size is small $(n = 5)$, the performance of model-based resampling is often better than that of nonparametric resampling. In situations where the nonparametric resampling is better (points above the bisector) the performance of the model-based plan is almost as good as that of

the nonparametric resampling plan. Next, we consider "true" frailties that are lognormal (Figure 3.4). For $(\theta, \lambda_0) = (0.6, 0.5)$ and $(0.6, 0.1)$, the nonparametric resampling scheme has for most settings (K, n) a smaller absolute relative bias compared to the model-based resampling scheme. For $(\theta, \lambda_0) = (0.1, 0.5)$ and $(0.1, 0.1)$ model-based resampling often performs clearly better than nonparametric resampling.

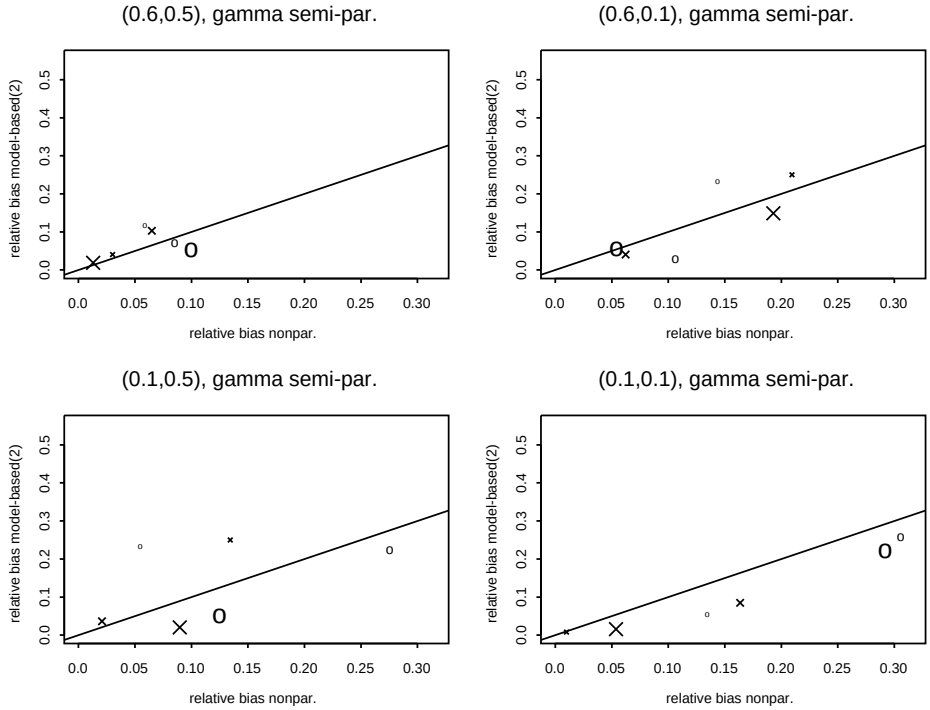

Figure 3.3: Absolute relative bias for the estimated standard error of the heterogeneity parameter (gamma frailties); $\circ = (15, 5)$, $\circ = (15, 20)$, $\bigcirc = (15, 40)$, x $= (30, 5)$, X $= (30, 20)$, X $= (30, 40)$.

Effect of the number of clusters and patients on the precision of the variance estimation

To study the effect of the number of clusters and the number of patients per cluster on the standard error we look at Figure 3.5 where, for the semi-

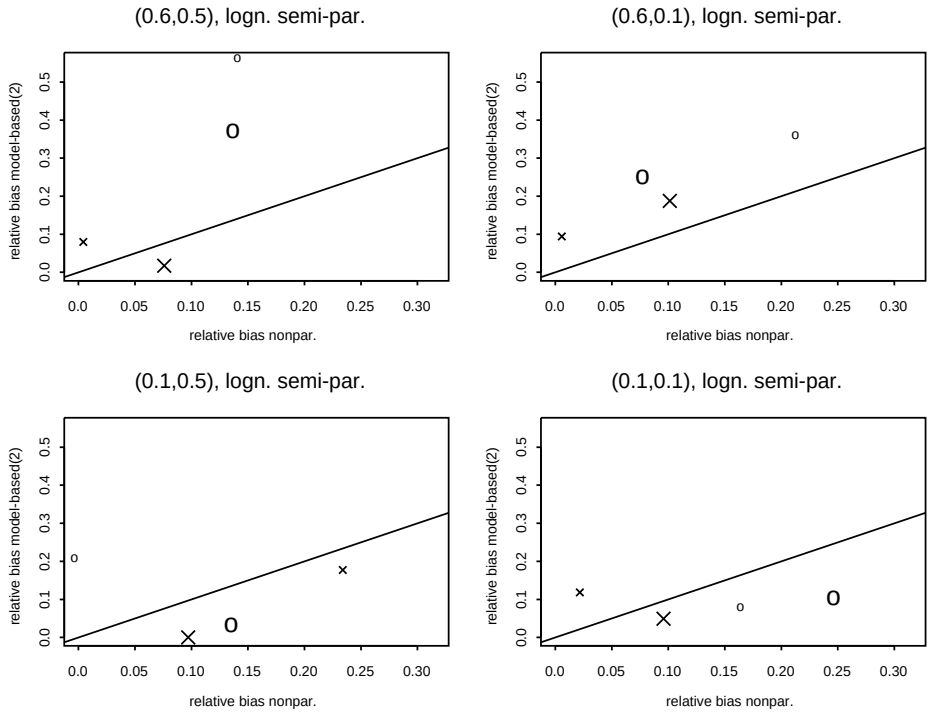

Figure 3.4: Absolute relative bias for the estimated standard error of the heterogeneity parameter (lognormal frailties); $\circ = (15, 20)$, $\bigcirc = (15, 40)$, x $= (30, 20)$, X $= (30, 40)$.

parametric gamma model, we plot for $(K, n) = (15, 5), (15, 20), (15, 40), (30, 5), (30, 20)$ and $(30, 40)$, SE^E, mean(SE^B) for nonparametric resampling and mean(SE^B) for the second model-based resampling scheme. The empirical standard error SE^E is considered as the reference point. The general conclusion, also based on pictures similar to Figure 3.5 for the semi-parametric lognormal model and for the parametric gamma model (pictures not shown), is that for both resampling plans the number of clusters is important to obtain accurate standard errors. We also see that, if the number of clusters is large enough (e.g., $K = 30$) we can only improve the accuracy of the standard errors in a moderate way by increasing the number of patients.

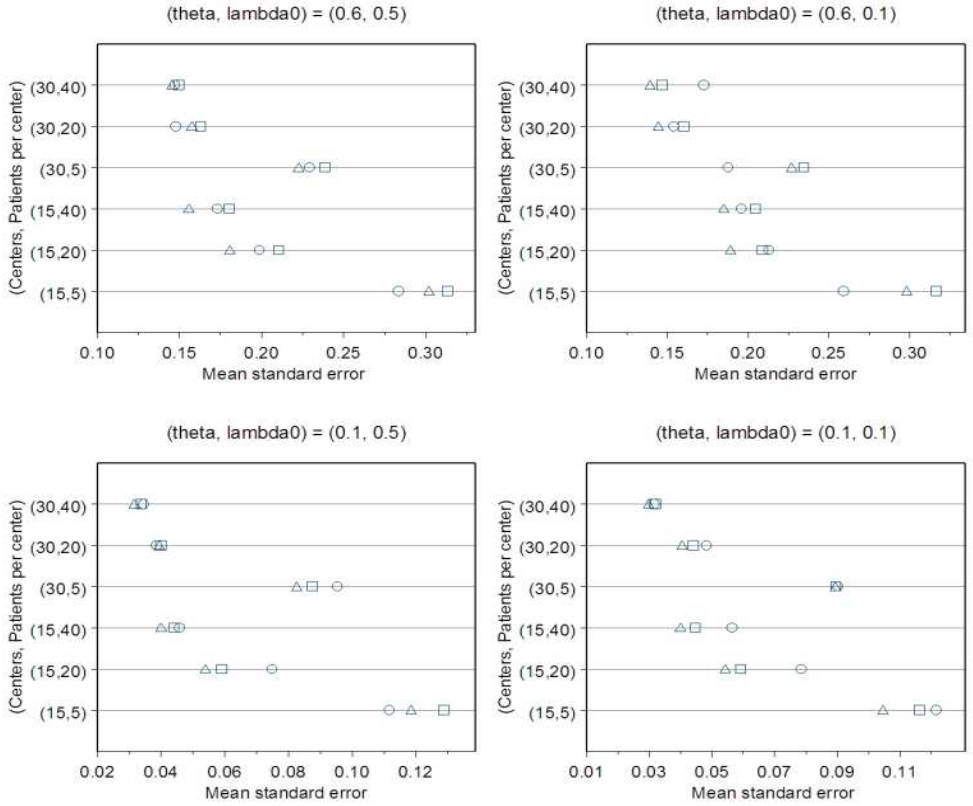

Figure 3.5: Effect of the number of clusters and patients on the mean estimated standard error of the heterogeneity parameter, semi-parametric gamma model; $\circ$ = empirical, $\triangle$ = nonparametric, $\square$ = model-based(2).

Effect of heterogeneity and event rate on the precision of the variance estimation

To study the effect of the heterogeneity and the event rate on the estimated standard error we look at Figure 3.6 where, for the semi-parametric gamma model, we plot for $(\theta, \lambda_0) = (0.6, 0.5), (0.6, 0.1), (0.1, 0.5)$ and $(0.1, 0.1)$, SE^E, mean(SE^B) for nonparametric resampling and mean(SE^B) for the second model-based resampling scheme. The empirical standard error SE^E is considered as the reference point. The general conclusion, also based on pictures similar to Figure 3.6 for the semi-parametric lognormal model and for the

parametric gamma model (pictures not shown), is that the bootstrap standard error obtained by both resampling plans are more accurate for small θ, i.e., $\theta = 0.1$. When λ_0 increases, the accuracy of the standard errors is improved in a moderate way, keeping θ constant.

Robustness

In all settings studied, the point estimates of the fixed effect in the correct and the misspecified model are close to each other (Tables 3.4 and 3.6). Also the estimated standard errors of the estimator of the fixed effect obtained by the nonparametric and the second model-based resampling scheme are similar, even if the model is misspecified. This means that there is robustness in terms of estimation of the fixed effects. This is in agreement with results in, e.g., Pickles and Crouchley (1995).
When $\theta = 0.6$, the point estimates of the heterogeneity parameter in the misspecified model are biased (Tables 3.3 and 3.5). The empirical variability is also quite different for the correct and the misspecified model. For $\theta = 0.1$, the bias of the point estimates is smaller and the difference in variability is less pronounced. This can be explained since there is only little difference in shape between the gamma distribution and the lognormal distribution when $\theta = 0.1$, whereas there is more difference when $\theta = 0.6$ (Figure 3.2). The relative bias, compared to the empirical standard error, indicates that in general the estimated standard error of the heterogeneity obtained by the nonparametric and the model-based resampling schemes are close to the corresponding empirical standard error, both for the correct and the misspecified model. So, bootstrap is useful to estimate the standard error of the heterogeneity parameter. However, since the empirical variability of the heterogeneity parameter is rather different for the correct and misspecified model, lack of robustness is an issue when fitting frailty models.

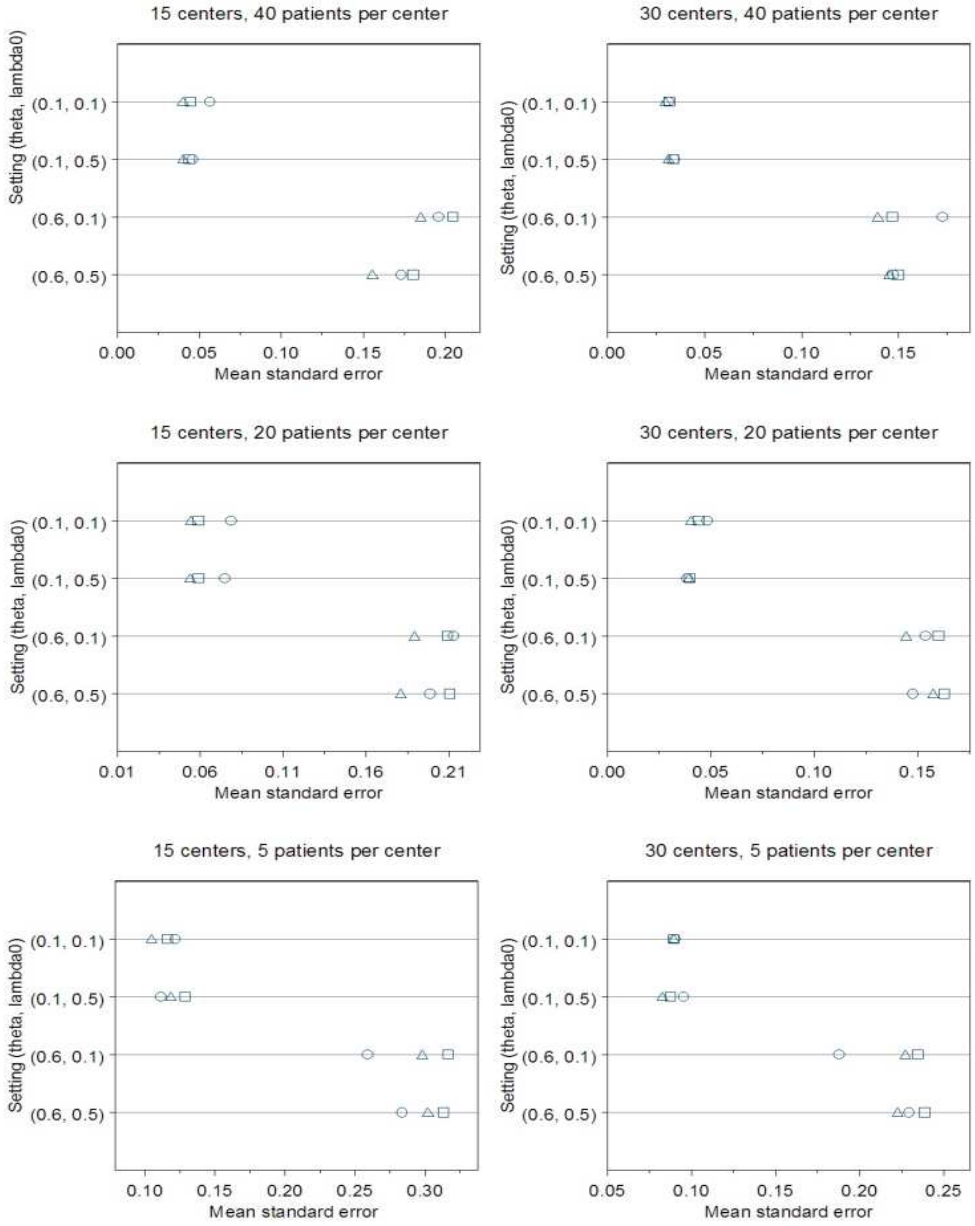

Figure 3.6: Effect of heterogeneity and event rate on the mean estimated standard error of the heterogeneity parameter, semi-parametric gamma model; $\circ$ = empirical, $\triangle$ = nonparametric, $\square$ = model-based(2).

3.4 Examples

We illustrate the nonparametric and the second model-based resampling scheme using the time to culling data and the early breast cancer data, introduced in Sections 1.4.2 and 1.4.3.

3.4.1 Analysis of the time to culling data

To study the heterogeneity between herds and the effect of the logarithm of the somatic cell count on the culling hazard, we fit both a semi-parametric gamma and a semi-parametric lognormal frailty model including the logarithm of the somatic cell count as a covariate and including a random herd effect. We estimate the fixed effect β and the heterogeneity parameter θ using the penalized partial likelihood approach (Therneau and Grambsch, 2000), available in S-Plus. To estimate the standard errors of the parameter estimates in the gamma, resp. lognormal, frailty model, we take $R = 500$ bootstrap samples by using the nonparametric and the second model-based resampling plan under the assumption of gamma, resp. lognormal, distributed frailties. For each bootstrap sample, we obtain an estimate for the fixed effect and the heterogeneity parameter by fitting a semi-parametric gamma, resp. lognormal, frailty model using the penalized partial likelihood approach. Based on the variability of these bootstrap parameter estimates, we obtain a bootstrap estimate for the standard errors of the fixed effect and the heterogeneity parameter. Both for the gamma and the lognormal frailty model, the parameter estimates, the estimate for the standard error of the estimated fixed effect reported by S-Plus and the bootstrap estimates for the standard errors of the parameter estimates are presented in Table 3.1. Recall that S-Plus does not provide an estimate for the standard error of the estimated heterogeneity parameter.

The fixed effect is estimated by $\hat{\beta} = 0.0683$, both in the gamma and the lognormal frailty model. As expected in Section 1.4.2, the culling hazard increases with increasing values of the logarithm of the somatic cell count. The bootstrap estimates for the standard error of $\hat{\beta}$ obtained using the nonpara-

metric and the second model-based resampling plan are comparable both for the gamma and the lognormal frailty model. Recall from Section 3.3.4 that there is robustness with respect to the estimation of the fixed effects. The estimate of the heterogeneity between herds is $\hat{\theta} = 0.1407$ for the gamma frailty model and $\hat{\theta} = 0.1555$ for the lognormal frailty model. Note that these values are rather small which implies that the gamma and lognormal frailty distributions are close (Figure 3.2). The bootstrap estimate for the standard error of $\hat{\theta}$ obtained using the second model-based resampling scheme is slightly larger than the estimate obtained using nonparametric bootstrap.

Table 3.1: Results for the time to culling data, parameter estimates for treatment effect and heterogeneity, estimated standard errors using nonparametric and model-based bootstrap.

Frailty distribution	Parameter	Est. S-Plus	SE S-Plus	nonpar. SE^B	model-based(2) SE^B
gamma	β	0.0683	0.0153	0.0152	0.0156
	θ	0.1407		0.0327	0.0355
lognormal	β	0.0683	0.0153	0.0152	0.0152
	θ	0.1555		0.0406	0.0421

3.4.2 Analysis of the early breast cancer data

We use a semi-parametric gamma and a semi-parametric lognormal frailty model to study the heterogeneity between centers and the effect of the perioperative chemotherapy compared to no further treatment. For both frailty models, the parameter estimates, the estimate for the standard error of the estimated fixed effect provided by S-Plus and the bootstrap estimates for the standard errors of the parameter estimates are presented in Table 3.2. These

estimates are obtained in the same way as explained in Section 3.4.1.
For the gamma frailty model, the treatment effect is estimated as -0.1417 ($SE = 0.0602$), meaning that the hazard corresponding to the disease-free survival is smaller for patients receiving perioperative chemotherapy compared to patients who receive no further treatment. The estimated standard error for $\hat{\beta}$ provided by S-Plus is comparable with the bootstrap estimate for the standard error obtained using the second model-based resampling scheme (0.0609), whereas the nonparametric bootstrap estimate for the standard error is smaller (0.0426). The variance of the frailties is estimated as 0.0629. The bootstrap estimate for the standard error of $\hat{\theta}$ obtained by using model-based bootstrap is slightly larger than the estimate obtained by using nonparametric bootstrap. Assuming a lognormal distribution for the frailty terms, the treatment effect is estimated as -0.1418 ($SE = 0.0602$), which is almost the same as the estimate obtained using a gamma frailty model. The model-based bootstrap estimate for the standard error of $\hat{\beta}$ (0.0629) is slightly larger than the estimate provided by the penalized partial likelihood approach, whereas the nonparametric resampling scheme leads to a smaller estimate (0.0425). The heterogeneity parameter is estimated as 0.0687. Note that the estimated variance for the frailty terms is small, which implies that the gamma and the lognormal frailty distribution are close (Figure 3.2). This can be seen from Table 3.2 since the results obtained for the gamma and the lognormal frailty model are comparable.

3.5 Conclusions

In this chapter, the use of bootstrap for the estimation of the standard errors of the parameter estimates in a shared frailty model is proposed. To complement the existing nonparametric resampling plan, we propose two model-based bootstrap algorithms. The comparison between the nonparametric and model-based resampling schemes and the robustness of the schemes to the model assumptions is studied by simulation. The results indicate that the first model-based resampling plan, based on resampling of the estimated frail-

Table 3.2: Results for the early breast cancer data, parameter estimates for treatment effect and heterogeneity, estimated standard errors using nonparametric and model-based bootstrap.

Bootstrap assumption	Parameter	Est. S-Plus	SE S-Plus	nonpar. SE^B	model-based(2) SE^B
gamma	β	-0.1417	0.0602	0.0426	0.0609
	θ	0.0629		0.0222	0.0318
lognormal	β	-0.1418	0.0602	0.0425	0.0626
	θ	0.0687		0.0244	0.0403

ties, underestimates the empirical variability of the parameter estimates for almost all settings studied. This corresponds to the conclusion drawn by Morris (2002) for linear mixed models. On the other hand, the second model-based algorithm, based on resampling from the estimated frailty distribution, provides in general precise assessment of the empirical variability of the parameter estimates, even if the model is misspecified. However, the empirical variability of the heterogeneity parameter can be rather different for the correct and misspecified models. This provides evidence that robustness in terms of the heterogeneity parameter is not guaranteed for the bootstrap algorithms (including the nonparametric bootstrap); but robustness holds for the fixed effects. This finding clearly illustrates the need for diagnostic tests for the choice of the frailty distribution. Oakes (1989) and Viswanathan and Manatunga (2001) propose a diagnostic test for bivariate survival data with a gamma frailty distribution which is based on the cross ratio function. Shih and Louis (1995a) propose a diagnostic test for the gamma frailty model for multivariate survival data with clusters of arbitrary size. This test is based on the evolution of the conditional posterior mean of the frailties over time. A detailed description of these diagnostic measures can be found in Duchateau

and Janssen (2008, Chapter 4). Further research on diagnostic techniques for assessing the frailty distribution assumption is needed.

The values of the parameters in the two examples, presented in Section 3.4, do not correspond to the simulation setting considered in Section 3.3. It would be interesting to perform simulations for parameter settings that correspond to the time to culling data and to the early breast cancer data to evaluate the performance of the resampling plans.

In the model-based resampling schemes we have made the assumption that censoring is independent of the covariates. In principle, it should be possible to extend the schemes to the more general situation where the censoring distribution depends on the covariates. Indeed, Davison and Hinkley (1997, p.351) give a resampling scheme for the Cox proportional hazards model in the case of univariate survival data where the censoring distribution depends on the covariates. This algorithm should be extended to the shared frailty model using ideas which are similar to the ideas proposed in Section 3.2. Furthermore, it also would be of interest to consider frailty densities other than gamma and lognormal, e.g., the positive stable or the inverse Gaussian frailty density. These are important topics for further research.

Table 3.3: Estimated standard error for heterogeneity parameter estimate; true gamma frailty distribution, for each setting: 40 patients per center, first line for 15 centers, second line for 30 centers.

True Setting	Bootstrap			nonpar.	model-based(1)	model-based(2)
(θ, λ_0)	assumption	mean $(\hat{\theta})$	SE^E	mean(SE^B)	mean(SE^B)	mean(SE^B)
(0.6, 0.5)	Gamma par.	0.4994	0.1722		0.1469	0.1760
		0.5850	0.1441		0.1302	0.1436
Gamma	Gamma semi-par.	0.4930	0.1732	0.1556	0.1494	0.1805
		0.5846	0.1472	0.1453	0.1356	0.1500
	Logn. semi-par.	2.0760	2.2220	4.4679	2.1952	8.8867
		2.8328	2.2885	6.0186	2.2860	2.6916
(0.6, 0.1)	Gamma par.	0.5678	0.1939		0.1736	0.1963
		0.5712	0.1687		0.1294	0.1411
Gamma	Gamma semi-par.	0.5688	0.1959	0.1849	0.1764	0.2048
		0.5703	0.1726	0.1393	0.1313	0.1469
	Logn. semi-par.	3.0123	2.7641	11.6203	6.1434	7.5224
		2.8285	2.7590	5.4391	1.3970	2.6621
(0.1, 0.5)	Gamma par.	0.0934	0.0375		0.0290	0.0392
		0.0983	0.0276		0.0217	0.0277
Gamma	Gamma semi-par.	0.0903	0.0458	0.0400	0.0338	0.0441
		0.0968	0.0346	0.0315	0.0263	0.0339
	Logn. semi-par.	0.1154	0.0673	0.0623	0.0487	0.0705
		0.1190	0.0518	0.0476	0.0367	0.0501
(0.1, 0.1)	Gamma par.	0.0969	0.0487		0.0310	0.0410
		0.0940	0.0245		0.0209	0.0275
Gam.	Gamma semi-par.	0.0924	0.0565	0.0399	0.0349	0.0447
		0.0910	0.0316	0.0299	0.0251	0.0321
	Logn. semi-par.	0.1229	0.0919	0.0710	0.0534	0.0761
		0.1119	0.0450	0.0438	0.0336	0.0474

Table 3.4: Estimated standard error for estimate of treatment effect; true gamma frailties; for each setting: 40 patients per center, first line for 15 centers, second line for 30 centers.

True Setting (θ, λ_0)	Bootstrap assumption	mean $(\hat{\beta})$	SE^E	nonpar. mean(SE^B)	model-based(1) mean(SE^B)	model-based(2) mean(SE^B)
(0.6, 0.5)	Gamma par.	0.2538	0.0922		0.0982	0.0985
		0.2464	0.0633		0.0698	0.0703
Gamma	Gamma semi-par.	0.2530	0.0915	0.0931	0.1002	0.1062
		0.2469	0.0640	0.0681	0.0716	0.0754
	Logn. semi-par.	0.2529	0.0915	0.0936	0.1001	0.1062
		0.2469	0.0639	0.0695	0.0707	0.0776
(0.6, 0.1)	Gamma par.	0.2456	0.1095		0.0992	0.1007
		0.2508	0.0707		0.0698	0.0708
Gamma	Gamma semi-par.	0.2451	0.1115	0.0988	0.1007	0.1078
		0.2507	0.0713	0.0675	0.0702	0.0760
	Logn. semi-par.	0.2454	0.1117	0.0986	0.1010	0.1098
		0.2508	0.0713	0.0674	0.0709	0.0773
(0.1, 0.5)	Gamma par.	0.2370	0.1008		0.0986	0.0983
		0.2469	0.0701		0.0688	0.0700
Gamma	Gamma semi-par.	0.2372	0.1015	0.0947	0.1010	0.1014
		0.2469	0.0710	0.0703	0.0694	0.0714
	Logn. semi-par.	0.2373	0.1013	0.0936	0.0980	0.0993
		0.2470	0.0710	0.0703	0.0710	0.0707
(0.1, 0.1)	Gamma par.	0.2541	0.1038		0.0982	0.0991
		0.2512	0.0710		0.0689	0.0699
Gamma	Gamma semi-par.	0.2533	0.1056	0.0921	0.0986	0.1002
		0.2513	0.0709	0.0680	0.0697	0.0701
	Logn. semi-par.	0.2531	0.1056	0.0929	0.0988	0.0922
		0.2513	0.0709	0.0690	0.0697	0.0724

Table 3.5: Estimated standard error for heterogeneity parameter estimate; true lognormal frailties; for each setting: 40 patients per center, first line for 15 centers, second line for 30 centers.

True Setting	Bootstrap			nonpar.	model-based(1)	model-based(2)
(θ, λ_0)	assumption	mean $(\hat{\theta})$	SE^E	mean(SE^B)	mean(SE^B)	mean(SE^B)
(0.6, 0.5)	Gamma par.	0.2976	0.1007		0.0871	0.1086
		0.3142	0.0805		0.0688	0.0821
Logn.	Gamma semi-par.	0.2915	0.1029	0.0972	0.0891	0.1110
		0.3085	0.0825	0.0779	0.0709	0.0838
	Logn. semi-par.	0.5605	0.3134	0.3568	0.2883	0.4258
		0.5836	0.2810	0.2597	0.2166	0.2858
(0.6, 0.1)	Gamma par.	0.3020	0.1066		0.0873	0.1099
		0.3300	0.0780		0.0729	0.0860
Logn.	Gamma semi-par.	0.2964	0.1072	0.0971	0.0898	0.1125
		0.3279	0.0801	0.0830	0.0749	0.0895
	Logn. semi-par.	0.5926	0.4060	0.3739	0.3083	0.5026
		0.6371	0.2704	0.2978	0.2450	0.3212
(0.1, 0.5)	Gamma par.	0.0818	0.0318		0.0255	0.0352
		0.0906	0.0265		0.0202	0.0263
Logn.	Gamma semi-par.	0.0756	0.0390	0.0347	0.0288	0.0394
		0.0851	0.0319	0.0289	0.0238	0.0308
	Logn. semi-par.	0.0942	0.0584	0.0504	0.0407	0.0596
		0.1030	0.0453	0.0409	0.0317	0.0453
(0.1, 0.1)	Gamma par.	0.0832	0.0377		0.0259	0.0352
		0.0821	0.0292		0.0184	0.0244
Logn.	Gamma semi-par.	0.0766	0.0451	0.0338	0.0288	0.0393
		0.0755	0.0296	0.0267	0.0215	0.0284
	Logn. semi-par.	0.0966	0.0661	0.0497	0.0403	0.0601
		0.0898	0.0407	0.0368	0.0278	0.0387

Table 3.6: Estimated standard error for estimate of treatment effect; true lognormal frailties; for each setting: 40 patients per center, first line for 15 centers, second line for 30 centers.

True Setting (θ, λ_0)	Bootstrap assumption	mean $(\hat{\beta})$	SE^E	nonpar. mean(SE^B)	model-based(1) mean(SE^B)	model-based(2) mean(SE^B)
(0.6, 0.5)	Gamma par.	0.2590	0.0937		0.0948	0.0955
		0.2512	0.0606		0.0672	0.0675
Logn.	Gamma semi-par.	0.2583	0.0935	0.0921	0.0954	0.1001
		0.2501	0.0619	0.0661	0.0674	0.0708
	Logn. semi-par.	0.2583	0.0935	0.0919	0.0946	0.1002
		0.2503	0.0617	0.0661	0.0676	0.0705
(0.6, 0.1)	Gamma par.	0.2567	0.1206		0.0968	0.0950
		0.2499	0.0754		0.0677	0.0680
Logn.	Gamma semi-par.	0.2562	0.1217	0.0941	0.0959	0.1007
		0.2503	0.0758	0.0654	0.0677	0.0703
	Logn. semi-par.	0.2564	0.1217	0.0941	0.0966	0.0995
		0.2501	0.0759	0.0653	0.0678	0.0701
(0.1, 0.5)	Gamma par.	0.2511	0.1032		0.0977	0.0990
		0.2553	0.0664		0.0693	0.0691
Logn.	Gamma semi-par.	0.2520	0.1045	0.0962	0.0984	0.0997
		0.2541	0.0662	0.0693	0.0691	0.0698
	Logn. semi-par.	0.2523	0.1044	0.0962	0.0983	0.1007
		0.2541	0.0662	0.0693	0.0700	0.0781
(0.1, 0.1)	Gamma par.	0.2462	0.0973		0.0985	0.0968
		0.2441	0.0720		0.0693	0.0698
Logn.	Gamma semi-par.	0.2455	0.0978	0.0968	0.0977	0.0984
		0.2435	0.0713	0.0682	0.0694	0.0701
	Logn. semi-par.	0.2456	0.0977	0.0967	0.0977	0.0990
		0.2437	0.0711	0.0682	0.0688	0.0698

Chapter 4

Fitting conditional survival models to meta-analytic data by using a transformation toward mixed-effects models

4.1 Introduction

Data from multicenter clinical trials are a typical example of multivariate survival data; data within the same center all share the same random cluster effect. The shared frailty model, introduced in Section 2.2.1, provides an appropriate way to describe the within cluster dependence of outcomes. In Section 2.2.2 we mentioned some likelihood-based estimation methods for the shared frailty model: the expectation-maximisation (EM) algorithm (Klein, 1992), the penalized partial likelihood approach (McGilchrist, 1993; Therneau and Grambsch, 2000), the Bayesian approach (Ducrocq and Casella, 1996). A more complex frailty model that can be used in the clinical trials context is introduced in Section 2.3.1: a frailty model with a random cluster effect and a random treatment effect. As discussed in Section 2.3.2, the likelihood-based estimation methods mentioned above have been adapted to fit this more complex

frailty model: the EM algorithm (Vaida and Xu, 2000; Cortiñas Abrahantes and Burzykowski, 2005), the penalized partial likelihood approach (Ripatti and Palmgren, 2000), the Bayesian approach (Legrand *et al.*, 2005).
In this chapter we propose an alternative way to fit frailty models. We start from the following observation: the integral of the weighted (over time) conditional cumulative loghazard depends in a linear way on the random effects describing the cluster and the treatment effect over clusters. Using the data within a cluster, we can estimate the integral using nonparametric estimation techniques. Considering the estimated integral as a response we can reformulate the original problem of "fitting a frailty model" into a problem of "fitting a linear mixed-effects model". We can summarize the idea as follows: based on the original data we obtain pseudo data (the estimated integrals) on which we can apply mixed models methodology. Model transformation also works for multivariate proportional odds models and multivariate additive risks models. Related references dealing with model transformation in the classical context of proportional hazards, additive risks and proportional odds models are Grigoletto and Akritas (1999) and Cao and Gonzalez-Manteiga (2008).
In Sections 4.2 and 4.3 we give, for right censored clustered survival data, the details on how multivariate proportional hazards models (frailty models), multivariate proportional odds models and multivariate additive risks models can be transformed into mixed-effects models. The finding that parameters of interest of a multivariate survival model become parameters of interest of a related mixed-effects model, provides an interesting link between two seemingly segregated fields. In Sections 4.4 and 4.5, we focus on frailty models to study the performance of the proposed method. The simulation study in Section 4.4 illustrates that, compared to the classical likelihood-based approaches, the transformation method provides a good and simple alternative for fitting frailty models for data sets with a sufficiently large number of clusters and moderate to large sample sizes within covariate level subgroups in the clusters. In Section 4.5 we discuss the performance of the method for the advanced colorectal cancer data, introduced in Section 1.4.4. We finally present some remarks and discuss possible further extensions in Section 4.6.

4.2 From frailty model to linear mixed-effects model

4.2.1 Model formulation

In this chapter we assume that z_{ij} is a binary covariate representing the treatment to which the jth patient in the ith cluster has been randomised, with $z_{ij} = -1$ if the patient is in the control group and $z_{ij} = 1$ if the patient is in the experimental group, for $j = 1, \ldots, n_i$ and $i = 1, \ldots, K$.
We consider a frailty model including a fixed overall treatment effect, a random cluster effect and a random treatment effect. Recall from (2.6) in Section 2.3.1 that the conditional hazard for the jth patient in the ith cluster is defined as

$$\lambda_{ij}(t) = \lambda_0(t) \exp\left\{b_{0i} + (\beta + b_{1i}) z_{ij}\right\}, \tag{4.1}$$

where the random effects b_{0i} and b_{1i} are assumed to follow zero-mean normal distributions. As discussed in Section 2.3.1, the variance-covariance matrix of the vector of random effects

$$\mathbf{b}' = (b_{01}, b_{11}, \ldots, b_{0i}, b_{1i}, \ldots, b_{0K}, b_{1K})$$

is given by

$$\mathbf{G} = \mathbf{I}_K \otimes \begin{pmatrix} \sigma_0^2 & \sigma_{01} \\ \sigma_{01} & \sigma_1^2 \end{pmatrix}, \tag{4.2}$$

where σ_0^2 and σ_1^2 provide information on the heterogeneity of the hazard due to the random cluster, resp. random treatment effect; σ_{01} is the covariance between the two random effects within a cluster.
Also recall that, in absence of a random treatment effect, model (4.1) reduces to the shared frailty model:

$$\lambda_{ij}(t) = \lambda_0(t) \exp(b_{0i} + \beta z_{ij}) = \lambda_0(t) u_i \exp(\beta z_{ij}), \tag{4.3}$$

where $u_i = \exp(b_{0i})$ is the frailty for cluster i. In absence of covariates this model further simplifies to

$$\lambda_i(t) = \lambda_0(t) \exp(b_{0i}) = \lambda_0(t) u_i. \tag{4.4}$$

In (4.3) and (4.4) b_{0i}, $i = 1, \ldots, K$, is a sample from a zero-mean normal density with variance σ_0^2, describing the heterogeneity between clusters.

4.2.2 The transformation

With $\Lambda_{ij}(t) = \int_0^t \lambda_{ij}(s)ds$ the cumulative hazard for the jth patient in cluster i, $j = 1, \ldots, n_i$ and $i = 1, \ldots, K$, and $\Lambda_0(t) = \int_0^t \lambda_0(s)ds$, we easily obtain from (4.1) that

$$\log \Lambda_{ij}(t) = \log \Lambda_0(t) + b_{0i} + (\beta + b_{1i})z_{ij}. \tag{4.5}$$

Let $w(.)$ be a weight function $\left(W(t) = \int_0^t w(s)ds\right)$ satisfying $w(s) \geq 0, s \in [0, \infty)$, $\int_0^\infty w(s)ds = 1$, and assigning zero weight to regions where the logarithm of the cumulative hazard function cannot be estimated due to censoring. The choice of the weight function is discussed in Section 4.2.3. Integrating both sides in (4.5) with respect to the weight function we obtain

$$\int_0^\infty \log \Lambda_{ij}(t)dW(t) \quad = \quad \alpha_F + b_{0i} + (\beta + b_{1i})z_{ij},$$

with $\alpha_F = \int_0^\infty \log \Lambda_0(t)dW(t)$. The patients in cluster i are divided, by the binary covariate z_{ij}, in a control and a treatment group. Let $\Lambda_i^{(0)}$, resp. $\Lambda_i^{(1)}$, be the cumulative hazard function shared by all control, resp. treated, patients in cluster i. Define, for $k = 0, 1$,

$$\Omega_{ik} = \int_0^\infty \log \Lambda_i^{(k)}(t)dW(t).$$

Then $\Omega_{i0} = \alpha_F + b_{0i} - (\beta + b_{1i})$ (control) and $\Omega_{i1} = \alpha_F + b_{0i} + (\beta + b_{1i})$ (treated). Following the ideas of Grigoletto and Akritas (1999), pseudo observations for the Ω_{ik}'s can be obtained as

$$\hat{\Omega}_{ik} = \int_0^\infty \log \hat{\Lambda}_i^{(k)}(t)dW(t),$$

where $\hat{\Lambda}_i^{(k)}$ is the estimated cumulative hazard based on the observations (X_{ij}, δ_{ij}) for all patients in cluster i with, for $k = 0$, $z_{ij} = -1$ and, for $k = 1$, $z_{ij} = 1$. As concrete estimator we use $\hat{\Lambda}_i^{(k)}(t) = -\log \hat{S}_i^{(k)}(t)$ with $\hat{S}_i^{(0)}(t)$ the

Kaplan-Meier estimator, introduced in (1.2) of Section 1.2.1, for the control group ($z_{ij} = -1$):

$$\hat{S}_i^{(0)}(t) = \prod_{j:X_{ij}\leq t, z_{ij}=-1} \left\{ \frac{r(X_{ij}) - d(X_{ij})}{r(X_{ij})} \right\},$$

with $r(v)$ the number still at risk at time v and $d(v)$ the number of events at time v and with $\hat{S}_i^{(1)}(t)$ the Kaplan-Meier estimator for the experimental group ($z_{ij} = 1$).
In terms of the pseudo observations we now can propose the model

$$\begin{aligned} \hat{\Omega}_{ik} &= \alpha_F + b_{0i} + (\beta + b_{1i}) z_{ik} + (\hat{\Omega}_{ik} - \Omega_{ik}) \\ &= \alpha_F + b_{0i} + (\beta + b_{1i}) z_{ik} + e_{ik} \end{aligned} \tag{4.6}$$

with $z_{i0} = -1$ and $z_{i1} = 1$. This is a linear mixed model with a random intercept and a random slope (treatment effect). Note that the error terms $e_{ik} = \hat{\Omega}_{ik} - \Omega_{ik}$ correct for the fact that the mixed model is applied to the pseudo data because the transformed cumulative hazard function cannot directly be observed. As $e_{ik} = \hat{\Omega}_{ik} - \Omega_{ik}$, it is clear that the random error terms do not satisfy the homogeneity assumption (because different subclusters have different sample sizes). In Section 4.2.3 we explain how to account for this heterogeneity when mixed models software is used to fit the model. A further remark is that for the special case (4.4) we obtain the following model after transformation:

$$\hat{\Omega}_i = \alpha_F + b_{0i} + \left(\hat{\Omega}_i - \Omega_i \right) = \alpha_F + b_{0i} + e_i. \tag{4.7}$$

For this one-way random effects model we only have one observation per cluster. At first glance this leads to identifiability problems. We, however, do have estimators of the variances of the error terms so that estimation of the variance components associated with the random cluster effect is possible. More details on this are given in Section 4.2.3.

4.2.3 The error variance

To apply the methods proposed in Section 4.2.2, we need estimates for the error variances $\sigma^2_{e,ik} = \text{Var}(e_{ik})$, resp. $\sigma^2_{e,i} = \text{Var}(e_i)$, of the random error

terms in model (4.6), resp. (4.7). We consider the general model (4.6). The patients of cluster i are divided in two groups: the control group ($k = 0$) and the treatment group ($k = 1$). Let n_{ik} be the number of patients in group k of cluster i. Define $F_i^{(k)}(t) = 1 - S_i^{(k)}(t)$. By using the relationship $\Lambda_i^{(k)}(t) = -\log\left\{1 - F_i^{(k)}(t)\right\}$ and a first order Taylor approximation, we obtain

$$\begin{aligned} e_{ik} &= \hat{\Omega}_{ik} - \Omega_{ik} = \int_0^\infty \left\{\log \hat{\Lambda}_i^{(k)}(t) - \log \Lambda_i^{(k)}(t)\right\} dW(t) \\ &\cong \int_0^\infty \frac{1}{\Lambda_i^{(k)}(t)} \frac{1}{S_i^{(k)}(t)} \left\{\hat{F}_i^{(k)}(t) - F_i^{(k)}(t)\right\} dW(t). \end{aligned} \tag{4.8}$$

Using the i.i.d. representation for $\hat{F}_i^{(k)}(t) - F_i^{(k)}(t)$ proposed by Lo and Singh (1986), we easily obtain an i.i.d. representation of e_{ik}. Based on this representation and given an appropriate weight function, an approximation for the (estimated) variance of e_{ik} is obtained through (4.8). In the sequel we have chosen a uniform weight function W on the interval (A, B), where A and B are chosen such that the logarithm of the cumulative hazard can be estimated for $t \in (A, B)$ for the control and the treatment group in each cluster. The variance of the error term $\hat{\Omega}_{ik} - \Omega_{ik}$ ($i = 1, \ldots, K$, $k = 0, 1$) can then be estimated by

$$\begin{aligned} \hat{\sigma}^2_{e,ik} &= \frac{1}{n_{ik}^2} \frac{1}{(B-A)^2} \int_A^B \frac{1}{\hat{\Lambda}_i^{(k)}(s)} \int_A^s \frac{1}{\hat{\Lambda}_i^{(k)}(t)} \\ &\quad \times \sum_{j:\ z_{ij}=k} \frac{I\left(0 \le x_{ij} \le t, \delta_{ij} = 1\right)}{\left\{1 - \frac{1}{n_{ik}} \sum_{j:\ z_{ij}=k} I\left(X_{ij} < x_{ij}\right)\right\}^2} dt\ ds \\ &\quad + \frac{1}{n_{ik}^2} \frac{1}{(B-A)^2} \int_A^B \frac{1}{\hat{\Lambda}_i^{(k)}(s)} \int_s^B \frac{1}{\hat{\Lambda}_i^{(k)}(t)} \\ &\quad \times \sum_{j:\ z_{ij}=k} \frac{I\left(0 \le x_{ij} \le s, \delta_{ij} = 1\right)}{\left\{1 - \frac{1}{n_{ik}} \sum_{j:\ z_{ij}=k} I\left(X_{ij} < x_{ij}\right)\right\}^2} dt\ ds. \end{aligned} \tag{4.9}$$

The technical details of the derivation of (4.9) are presented in Appendix 4.7.1. For model (4.7), we obtain in a similar way the estimated variance of the error

term $\hat{\Omega}_i - \Omega_i$ $(i = 1, \ldots, K)$:

$$\begin{aligned}\hat{\sigma}^2_{e,i} &= \frac{1}{n_i^2}\frac{1}{(B-A)^2}\int_A^B \frac{1}{\hat{\Lambda}_i(s)}\int_A^s \frac{1}{\hat{\Lambda}_i(t)} \\ &\quad \times \sum_{j=1}^{n_i} \frac{I\left(0 \le x_{ij} \le t, \delta_{ij} = 1\right)}{\left\{1 - \frac{1}{n_i}\sum_{j=1}^{n_i} I\left(X_{ij} < x_{ij}\right)\right\}^2} dt\ ds \\ &\quad + \frac{1}{n_i^2}\frac{1}{(B-A)^2}\int_A^B \frac{1}{\hat{\Lambda}_i(s)}\int_s^B \frac{1}{\hat{\Lambda}_i(t)} \\ &\quad \times \sum_{j=1}^{n_i} \frac{I\left(0 \le x_{ij} \le s, \delta_{ij} = 1\right)}{\left\{1 - \frac{1}{n_i}\sum_{j=1}^{n_i} I\left(X_{ij} < x_{ij}\right)\right\}^2} dt\ ds.\end{aligned}$$

4.3 Other conditional survival models

As mentioned in Section 4.1, model transformation also works for conditional survival models that are different from frailty models. We give two examples. First we consider the multivariate proportional odds model, introduced in Section 2.4.2:

$$\log\left\{\frac{F_{ij}(t)}{1 - F_{ij}(t)}\right\} = h_0(t) + b_{0i} + (\beta + b_{1i})z_{ij},$$

where F_{ij} is the conditional distribution function for the jth patient in the ith cluster and h_0 is the baseline logodds function. Integrating out with respect to the weight function w, we obtain

$$\int_0^\infty \log\left\{\frac{F_{ij}(t)}{1 - F_{ij}(t)}\right\} dW(t) = \alpha_{PO} + b_{0i} + (\beta + b_{1i})z_{ij}, \tag{4.10}$$

with $\alpha_{PO} = \int_0^\infty h_0(t)dW(t)$.
A second example is the multivariate additive risks model, introduced in Section 2.4.1:

$$\lambda_{ij}(t) = \lambda_0(t) + b_{0i} + (\beta + b_{1i})z_{ij}.$$

Integrating out the corresponding cumulative hazard function with respect to the weight function $\tilde{W}(t) = W(t)/\int_0^\infty sdW(s)$, we obtain

$$\int_0^\infty \Lambda_{ij}(t)d\tilde{W}(t) = \alpha_{AR} + b_{0i} + (\beta + b_{1i})z_{ij}, \tag{4.11}$$

with $\alpha_{AR} = \int_0^\infty \Lambda_0(t) d\tilde{W}(t)$.
Starting from (4.10) and (4.11), transformations to mixed-effects models are obtained along the lines of the discussion given for the frailty models. The technical details on the estimation of the error variance are given in Appendix 4.7.1. In the rest of this chapter we focus on frailty models, as described in Section 4.2.1 to 4.2.3, to compare the results obtained by the proposed method with the results obtained by a classical likelihood-based approach.

4.4 Simulations

We study the performance of the proposed method in the context of frailty models by using a simulation study. As simulation model we consider the setting of a multicenter clinical trial. First, we consider the special case of the shared frailty model (4.4) including only a random center (cluster) effect. We compare the results obtained from the proposed method with those obtained from the penalized partial likelihood approach (Therneau and Grambsch, 2000). We use "coxph" in S-Plus 7.0.6 for the penalized partial likelihood inference. The precision of the parameter estimates is investigated for a varying number of clusters and a varying number of observations per cluster. We further look at different percentages of censoring, we consider different sizes for σ_0^2 and different values for the baseline event rate $\lambda_0(t)$ (which we assume constant in time for simplicity). We also discuss the robustness of the proposed method against misspecification of the frailty density. Next, we consider the general frailty model (4.1), including a fixed overall treatment effect, a random center effect and a random treatment effect. For this model, we allow for correlation between b_{0i} and b_{1i}. Also here we compare the results obtained by the proposed method with those based on the penalized partial likelihood approach (Ripatti and Palmgren, 2000) using "coxme" in S-Plus 7.0.6 for the likelihood inference. We further study the effect of the size of σ_0^2 and σ_1^2 on the precision of the parameter estimates.

4.4.1 Description of the simulations

For simplicity, we assume a constant sample size per cluster: $n_i = n$, for $i = 1, \ldots, K$. For each specific setting $(K, n, \lambda_0, \sigma_0^2, \sigma_{01}, \sigma_1^2)$, 500 data sets are generated from model (4.1), assuming a constant baseline hazard. Given a particular setting, observations for a particular data set are generated in the following way. First, we generate K observations (b_{0i}, b_{1i}), $i = 1, \ldots, K$, from a bivariate normal distribution with zero mean vector and variance-covariance matrix

$$\begin{pmatrix} \sigma_0^2 & \sigma_{01} \\ \sigma_{01} & \sigma_1^2 \end{pmatrix}.$$

Recall from Section 4.2.1 that $b_{01}, \ldots, b_{0K}$ are the random center effects and $b_{11}, \ldots, b_{1K}$ are the random treatment effects. The time to event for each patient is randomly generated from an exponential distribution with parameter $\lambda_{ij} = \lambda_0 \exp(b_{0i} + (\beta + b_{1i}) z_{ij})$, where z_{ij} is generated from a Bernoulli distribution with success probability 0.5. The censoring time for each patient is randomly generated from a uniform distribution, so that approximately 30% censoring is obtained. For each data set, pseudo data $\hat{\Omega}_{ik}$ are generated through the model transformation described in Section 4.2 by using a uniform weight function w on the interval (A, B), chosen so that $0 < \hat{S}_i^{(k)}(t) < 1$ for $t \in (A, B)$. For each cluster i, the estimated variance of $\hat{\Omega}_{ik} - \Omega_{ik}$ is computed as explained in Section 4.2.3. To fit model (4.6), we use the SAS procedure PROC MIXED (see Appendix 4.7.2 for details on PROC MIXED). For each data set we obtain estimates for β, σ_0^2, σ_1^2 and σ_{01}.

For the special case of model (4.4), the data are generated as explained above with $\beta = 0$, $\sigma_1^2 = 0$ and $\sigma_{01} = 0$. Here, we consider moderate censoring (around 30%) and heavy censoring (around 60%). To study the robustness of the proposed method against frailty misspecification, the data are generated assuming that the frailties $u_1 = \exp(b_{01})$, ..., $u_K = \exp(b_{0K})$ are gamma distributed with mean $E(U_i) = e^{\sigma_0^2/2}$ and variance $Var(U_i) = \theta = e^{\sigma_0^2}\left(e^{\sigma_0^2} - 1\right)$. This corresponds to random effects b_{0i} with mean 0 and variance σ_0^2. For each data set, pseudo data $\hat{\Omega}_i$ are generated as explained above. We fit model (4.4)

assuming, incorrectly, that the random effects b_{0i} are normally distributed with mean 0 and variance σ_0^2.

4.4.2 Choice of the parameters

Frailty model with a random center effect

For the concrete simulation, we take 20, 50, and 100 centers with 50 or 100 patients per center. The parameter values λ_0 and σ_0^2 in both settings are chosen in such a way that a different magnitude of spread in the median time to event from center to center is induced. Recall from Section 3.3.3 that the median time to event T_M is the solution of $\exp\{-\lambda_0 \exp(b_0) T_M\} = 0.5$, i.e., $T_M = \frac{\log 2}{\lambda_0 \exp(b_0)}$, where we assume in this case that b_0 is zero-mean normally distributed. The magnitude of spread in the median time to event from center to center was determined by computing the density function of T_M (Figure 4.1). It is easy to show that the density function $f_{T_M}(t)$ is given by

$$f_{T_M}(t) = \frac{1}{t\sqrt{2\pi\sigma_0^2}} \exp\left[-\frac{\left\{\log\left(\frac{\log 2}{\lambda_0 t}\right)\right\}^2}{2\sigma_0^2} \right].$$

As true values for the event rate, we take $\lambda_0 = 0.1$ and 0.5. The heterogeneity parameter is set at $\sigma_0^2 = 0.08765$ and 0.1577. To understand the choice of these values, note that the relation between σ_0^2 and the frailty variance θ is given by: $Var(U_i) = \theta = e^{\sigma_0^2}\left(e^{\sigma_0^2} - 1\right)$. The values of σ_0^2 correspond to a frailty variance of $\theta = 0.1$, resp. 0.2.
For the settings (σ_0^2, λ_0) = (0.08765, 0.1) and (0.1577, 0.1), there is much spread in the median time to event over the centers. For the settings (σ_0^2, λ_0) = (0.08765, 0.5) and (0.1577, 0.5), there is little spread in the median time to event over the centers, with a bigger spread for $\sigma_0^2 = 0.1577$. To study the robustness of the proposed method, we take $\sigma_0^2 = 0.3520$ ($\theta = 0.6$) and $\lambda_0 = 0.1$. The motivation for choosing $\theta = 0.6$ is that the gamma and the lognormal density functions are close for $\theta = 0.1$, whereas for $\theta = 0.6$ these densities are more apart, as we already illustrated in Figure 3.2.

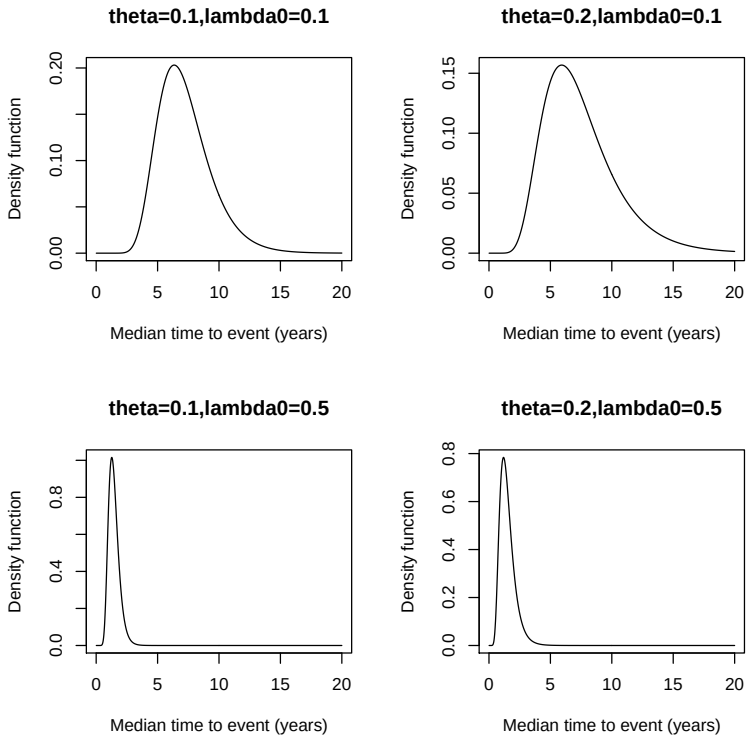

Figure 4.1: Density function of the median time to event over centers.

Frailty model with random center and treatment effects

We consider a situation with 50 centers that have 100 or 200 patients per center. The baseline hazard is assumed constant and equal to $\lambda_0 = 0.3$. For the treatment effect, we use $\beta = -0.2$. These parameter values are chosen so that the bladder cancer data considered in Legrand *et al.* (2005) can serve as a reference. This study investigates heterogeneity in disease-free interval due to center and treatment effect over centers in a large bladder cancer database including data from seven randomised clinical trials. We simulate data using different combinations of values of σ_0^2 and σ_1^2, varying from 0 to 0.08 ($\sigma_0^2, \sigma_1^2 = 0, 0.04$ or 0.08). The covariance parameter σ_{01} is chosen such that the correlation between b_0 and b_1 is equal to 0.5 (e.g., for $\sigma_0^2 = 0.08$, $\sigma_1^2 = 0.04$,

the covariance parameter $\sigma_{01} = 0.0283$ corresponds with a correlation of 0.5, see Table 4.3). This value mimics the correlation between the random effects observed in the bladder cancer data (Legrand *et al.*, 2005).

4.4.3 Simulation results

Frailty model with a random center effect

Table 4.1 presents, for the setting $(\sigma_0^2, \lambda_0) = (0.08765, 0.5)$, the relative bias, the mean and the empirical standard deviation computed for the 500 estimates of the variance of the random center effect. The results for the settings $(\sigma_0^2, \lambda_0) = (0.08765, 0.1), (0.1577, 0.5)$ and $(0.1577, 0.1)$ are not substantially different (see Tables 4.5, 4.6 and 4.7 at the end of this chapter).
The general conclusion for all parameter settings is that σ_0^2 is estimated well by the proposed method if the cluster size is large enough (i.e., $n = n_i = 100$). Both for the penalized partial likelihood approach (coxph in S-Plus 7.0.6) and the proposed method, the absolute relative bias decreases with the increasing cluster size, and is not substantially influenced by the number of clusters. In general, the estimates obtained by the proposed approach are on average closer to the true value σ_0^2 if the cluster size is large enough (i.e., $n = n_i = 100$). For a smaller cluster size ($n = n_i = 50$), the estimates obtained by the penalized partial likelihood are more precise. In general, the absolute relative bias increases if the amount of censoring increases. However, if the cluster size is large enough, σ_0^2 is estimated well by the proposed method.
Table 4.2 shows the results obtained by the penalized partial likelihood approach and the proposed method if the "true" frailties are gamma distributed with variance 0.6. The results illustrate that, for both methods, the point estimates of σ_0^2 are biased if the model is misspecified. This lack of robustness is also discussed in the bootstrap context in Section 3.3.4 (see also Massonnet *et al.*, 2006). It clearly shows the need for lack-of-fit measures for frailty models.

Table 4.1: Relative bias, mean and empirical standard deviation of the estimated values $\hat{\sigma}_0^2$ over the 500 simulations; true $\sigma_0^2 = 0.08765$ ($\theta = 0.1$), $\lambda_0 = 0.5$; first line for coxph, second line for PROC MIXED.

	30 % censoring			60 % censoring		
(K, n)	Rel. bias	Mean	Emp. std	Rel. bias	Mean	Emp. std
(100,100)	-0.0314	0.0849	0.0152	-0.0143	0.0864	0.0169
	-0.0257	0.0854	0.0148	-0.0177	0.0861	0.0162
(100,50)	-0.0382	0.0843	0.0174	-0.0280	0.0852	0.0198
	-0.0975	0.0791	0.0168	-0.1135	0.0777	0.0197
(50,100)	-0.0097	0.0868	0.0227	-0.0234	0.0856	0.0223
	0.0029	0.0879	0.0223	-0.0177	0.0861	0.0225
(50,50)	-0.0188	0.0860	0.0240	-0.0462	0.0836	0.0277
	-0.0667	0.0818	0.0248	-0.1204	0.0771	0.0274
(20,100)	-0.0439	0.0838	0.0327	-0.0747	0.0811	0.0339
	-0.0154	0.0863	0.0347	-0.0382	0.0843	0.0355
(20,50)	-0.0690	0.0816	0.0343	-0.0451	0.0837	0.0434
	-0.1010	0.0788	0.0365	-0.1067	0.0783	0.0436

Frailty model with random center and treatment effects

In Table 4.3 we report, for the parameter choice described in Section 4.4.2 and for 50 centers with 200 patients per center, the mean, the empirical standard deviation and the average of the model-based standard deviations computed over the 500 estimates of the fixed treatment effect and the variance-covariance components of the random effects. We compare the results obtained by the proposed method with those obtained by coxme in S-Plus 7.0.6. There is no

Table 4.2: Relative bias, mean and empirical standard deviation of the estimated values $\hat{\sigma}_0^2$ over the 500 simulations; true gamma frailties, $\sigma_0^2 = 0.3520$ ($\theta = 0.6$), $\lambda_0 = 0.1$; first line for coxph, second line for PROC MIXED; 30 % censoring.

(K, n)	Rel. bias	Mean	Emp. std
(100,100)	0.4102	0.4964	0.0862
	0.3226	0.4656	0.0740
(50,100)	0.3989	0.4925	0.1190
	0.3142	0.4626	0.1047
(20,100)	0.3563	0.4774	0.1825
	0.3422	0.4725	0.1714

reliable software available to compute the standard errors for the variance-covariance parameters for the penalized partial likelihood approach. The parameter β is in general estimated well by both methods. The bias of the fixed effect estimates obtained by coxme is in general a bit smaller than for the proposed method. The empirical variability of estimates of β is similar for both methods. The estimates of σ_0^2, σ_1^2 and σ_{01} for both methods are on average comparable. The estimates produced by coxme have in general the smallest empirical variability. The average of the model-based standard deviations for the fixed effect and for the variance components, give an adequate estimate of the empirical variability for the proposed method.

For situations with 50 centers that have 100 patients per center (50 in the control group and 50 in the treatment group), the proposed method still gives reasonable estimates for β, σ_0^2, σ_1^2 and σ_{01} (see Table 4.8 at the end of this chapter). However, compared to the situation with 50 centers and 200 patients per center, we obtain estimates of a somewhat lower quality. This illustrates

that the proposed method needs large enough samples sizes within the control group and the treatment group; this confirms our finding for the frailty model with a random center effect.

Table 4.3: Mean, empirical standard deviation and average of the model-based standard deviations of the estimated values over the 500 simulations; 50 centers, 200 patients per center (100 patients in control and treatment group); $\lambda_0 = 0.3$, $\beta = -0.2$.

		PROC MIXED			coxme	
	True	Mean	Emp. std.	Model std.	Mean	Emp. std.
β	-0.20	-0.2010	0.0124	0.0133	-0.2001	0.0120
σ_0^2	0	0.0005	0.0009	0.0008	0.0017	0.0011
σ_1^2	0	0.0006	0.0009	0.0008	0.0075	0.0062
σ_{01}	0	0.0002	0.0014	0.0014	0.0027	0.0024
β	-0.20	-0.1998	0.0423	0.0420	-0.2005	0.0422
σ_0^2	0	0.0006	0.0011	0.0009	0.0007	0.0010
σ_1^2	0.08	0.0801	0.0177	0.0182	0.0790	0.0127
σ_{01}	0	-0.0003	0.0047	0.0047	0.0011	0.0016
β	-0.20	-0.1997	0.0407	0.0418	-0.2001	0.0404
σ_0^2	0.04	0.0393	0.0102	0.0100	0.0394	0.0096
σ_1^2	0.08	0.0788	0.0179	0.0180	0.0777	0.0168
σ_{01}	0.0283	0.0277	0.0102	0.0104	0.0276	0.0097
β	-0.20	-0.1993	0.0135	0.0139	-0.1995	0.0123
σ_0^2	0.08	0.0794	0.0181	0.0181	0.0797	0.0124
σ_1^2	0	0.0005	0.0010	0.0008	0.0006	0.0009
σ_{01}	0	0.0000	0.0047	0.0046	0.0016	0.0014
β	-0.20	-0.2009	0.0331	0.0313	-0.2006	0.0321
σ_0^2	0.08	0.0805	0.0182	0.0184	0.0799	0.0179
σ_1^2	0.04	0.0399	0.0096	0.0101	0.0397	0.0092
σ_{01}	0.0283	0.0290	0.0105	0.0107	0.0287	0.0102
β	-0.20	-0.1986	0.0430	0.0421	-0.1996	0.0433
σ_0^2	0.08	0.0776	0.0173	0.0180	0.0776	0.0169
σ_1^2	0.08	0.0794	0.0174	0.0183	0.0790	0.0168
σ_{01}	0.04	0.0393	0.0142	0.0142	0.0394	0.0139

4.5 Analysis of the colorectal cancer data

To investigate the between-trial variation (heterogeneity) in both the baseline risk and the effectiveness of the therapy, we fit the frailty model (4.1) including a fixed treatment effect, a random trial effect and a random treatment effect. In this model, we also take into account a possible correlation between the two random effects within a trial. The parameter estimates and the corresponding standard errors, obtained by the proposed method and by the penalized partial likelihood approach (coxme in S-Plus 7.0.6), are presented in Table 4.4.
The point estimates for σ_1^2 and σ_{01} are very small (almost zero). For this reason, we fit the shared frailty model (4.3) including a fixed treatment effect and a random trial effect. The results are shown in Table 4.4.

Table 4.4: Results of the analysis of the survival time of the patients included in the colorectal cancer trials (standard error in parentheses).

Method	β	σ_0^2	σ_1^2	σ_{01}
PROC MIXED	-0.0458 (0.0219)	0.0476 (0.0187)	0.0000 (-)	-0.0084 (0.0056)
coxme	-0.0534 (0.0169)	0.0355	3.34×10^{-10}	1.78×10^{-11}
PROC MIXED	-0.0558 (0.0225)	0.0461 (0.0172)		
coxph	-0.0534 (0.0169)	0.0376		

The estimates obtained by the penalized partial likelihood and the transformation method are a bit different. However, the difference has only low impact on important medical quantities, e.g. on the density of the median time to event in the control group over trials (Figure 4.2). So both methods provide similar medical conclusions. A possible explanation for the difference between the estimates obtained by both methods, is that only 16 out of 27 trials have sample sizes of both the treatment and the control group larger than 50 patients. From the simulations, we know that the accuracy of the transformation method is comparable to the penalized partial likelihood if the cluster sizes are large enough.

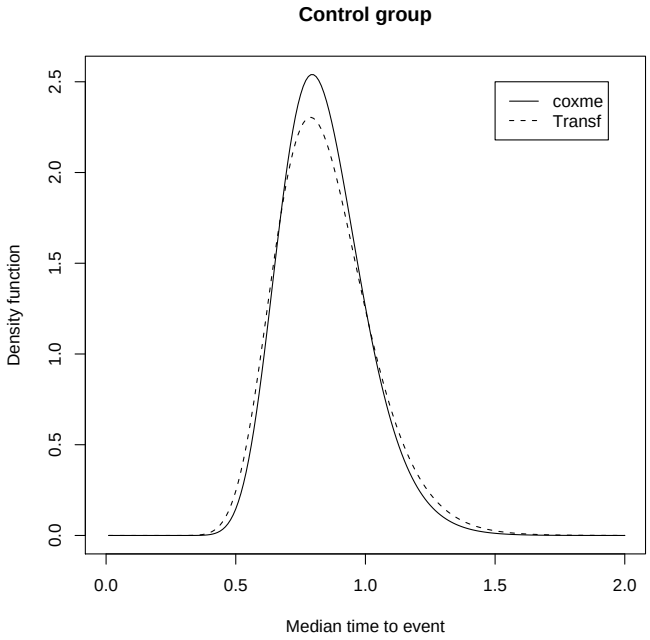

Figure 4.2: Density function of the median time to event over clusters for the colorectal cancer data.

The use of frailty models or linear mixed-effects models (for the pseudo data) raises questions on diagnostics. In the context of the method proposed here, we focus on model diagnostics for the linear mixed-effects model for the pseudo data (see, e.g., West *et al.*, 2007). If the diagnostic plots show that the linear mixed-effects model is not appropriate for the pseudo data, this indicates that the corresponding frailty model is not valid for the individual data. However, if the results of the mixed model diagnostics are good, there is no guarantee that the frailty model is the correct model for the individual data. It is indeed possible that other models lead to the same linear mixed-effects model using an appropriate model transformation. To check the fit of the linear mixed-effects model obtained from the shared frailty model using the model transformation, we consider a plot of the studentized conditional residuals versus the predicted values, a normal QQ plot of the studentized conditional residuals and a nor-

mal QQ plot of the empirical best linear unbiased predictions (i.e., EBLUP's) of the random trial effect (see Figure 4.3 and Figure 4.4). These diagnostic plots show that the pseudo values, obtained from the colorectal cancer data, can be analysed using the linear mixed-effects model that corresponds to the shared frailty model. However, there is still no guarantee that the shared frailty model is the correct model for the original data.

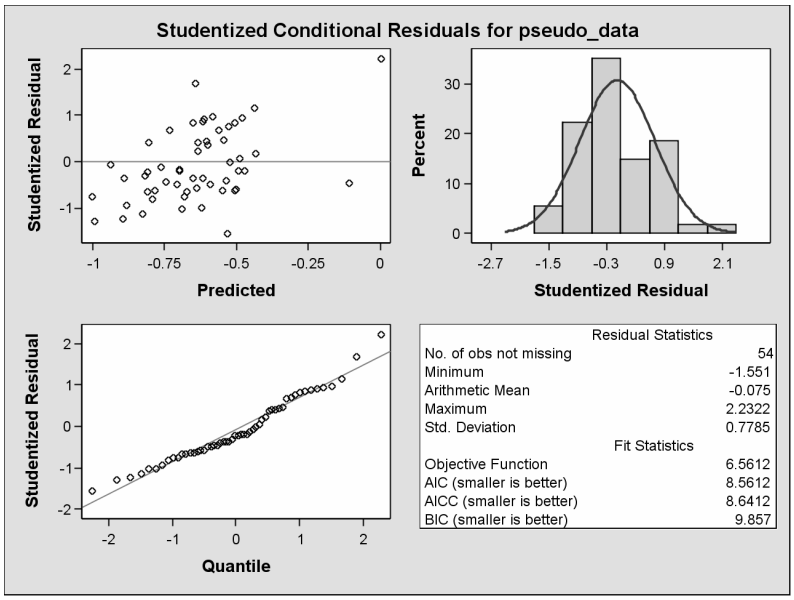

Figure 4.3: Plots of the studentized conditional residuals for the linear mixed-effects model that corresponds to the shared frailty model.

4.6 Conclusions

In this chapter, an alternative approach to fit frailty models is proposed. The original problem of "fitting a frailty model" is reformulated into a standard problem of "fitting a linear mixed-effects model". We show that the integral of the weighted (over time) conditional cumulative loghazard depends in a linear way on the random effects describing the cluster and/or the treatment effect over clusters. Using the data within a cluster, the integral can be esti-

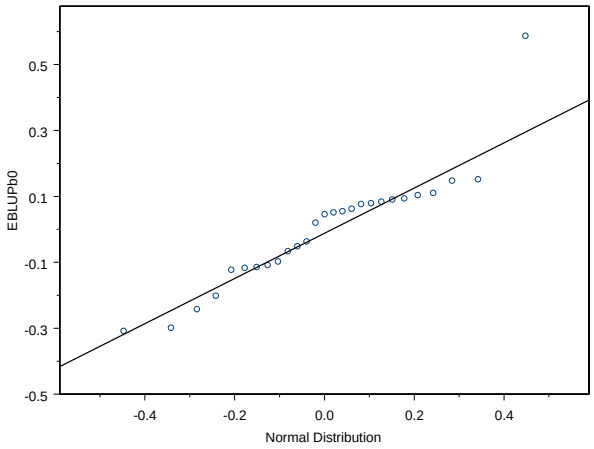

Figure 4.4: Normal QQ plot of the EBLUP's of the random trial effect in the linear mixed-effects model that corresponds to the shared frailty model.

mated using nonparametric estimation techniques. Considering the estimated integrals as a response, linear mixed models methodology can be applied. We illustrate that this transformation idea can also be used to fit multivariate proportional odds models and multivariate additive risks models. Most standard statistical packages contain procedures to fit complex linear mixed-effects models but offer only a limited number of procedures to fit conditional (random effects) survival models. The proposed model transformation is therefore a useful practical way to get insight in the heterogeneity in clustered data. The performance of the proposed method was studied by simulation in the context of frailty models. The results indicate a good performance of the proposed method for data sets with a sufficiently large number of clusters (i.e., $K = 20$) and moderate to large sample sizes within covariate level subgroups in the clusters (i.e., at least $n_{ik} = 50$). Given this finding, the proposed method is more suitable for a meta-analysis setting rather than for the setting of a multicenter clinical trial.

We considered a frailty model with a binary covariate and we therefore could use the Kaplan-Meier estimator for the survival function. It would be of in-

terest to extend the transformation idea to frailty models with a continuous covariate. We then need the Beran estimator to estimate the survival function (Beran, 1981). An asymptotic representation for the Beran estimator is proposed by Van Keilegom and Veraverbeke (1997). Such a representation is necessary to estimate the variance of the error terms in the mixed-effects model. This problem is currently under investigation (see Cao *et al.*, 2008). From the above discussion it is also clear that the transformation method is useful for censoring schemes that are different from the right censoring scheme discussed so far. Indeed, the transformation idea readily extends to any censoring scheme for which an i.i.d. representation for a nonparametric estimator for the cumulative hazard or the survival function is available (e.g., for interval-censored data, Lindsey and Ryan (1998); or for left truncated and right censored data, Gijbels and Wang (1993) and Zhou and Yip (1999)).
The performance of the transformation method for multivariate proportional odds models and multivariate additive risks models is a subject for further study.

4.7 Appendix

4.7.1 The error variance

Multivariate proportional hazards models

To apply the method proposed in Section 4.2.2 we need estimated values for the error variance. To obtain estimates we can rely on an asymptotic representation, proposed by Lo and Singh (1986), decomposing $\hat{F}_i^{(k)}(t) - F_i^{(k)}(t)$ as an average of i.i.d. terms and a lower order remainder term $r_{ik}(t)$, where $F_i^{(k)}$ is the continuous failure time distribution function for subjects in cluster i with $z_{ij} = k$.
Let G be the censoring distribution, $1 - H_{ik}(s) = \{1 - F_i^{(k)}(s)\}\{1 - G(s)\}$ and $H_{ik}^u(s) = P\left(X_{ij} \leq s, \delta_{ij} = 1 | z_{ij} = k\right) = \int_0^s 1 - G(y^-) dF_i^{(k)}(y)$. It follows from

Lo and Singh (1986) that

$$\hat{F}_i^{(k)}(t) - F_i^{(k)}(t) = \frac{1}{n_{ik}} \sum_{j:\ z_{ij}=k} \xi_{ik}(X_{ij}, \delta_{ij}, t) + r_{ik}(t),$$

where

$$\xi_{ik}(X_{ij}, \delta_{ij}, t) = \frac{I\left(X_{ij} \leq t, \delta_{ij} = 1\right)}{1 - H_{ik}(X_{ij})} - \int_0^t \frac{I\left(X_{ij} > s\right)}{\left\{1 - H_{ik}(s)\right\}^2}\, dH_{ik}^u(s),$$

for a subject in cluster i with observed information (X_{ij}, δ_{ij}) and $z_{ij} = k$. By using the relationship $\Lambda_i^{(k)}(t) = -\log\left\{1 - F_i^{(k)}(t)\right\}$ and first order Taylor expansions, we obtain

$$\begin{aligned}
\log \hat{\Lambda}_i^{(k)}(t) - \log \Lambda_i^{(k)}(t) &\cong \frac{1}{\Lambda_i^{(k)}(t)} \frac{1}{\left\{1 - F_i^{(k)}(t)\right\}} \left\{\hat{F}_i^{(k)}(t) - F_i^{(k)}(t)\right\} \\
&\cong \frac{1}{\Lambda_i^{(k)}(t)} \frac{1}{S_i^{(k)}(t)} \frac{1}{n_{ik}} \sum_{j:\ z_{ij}=k} \xi_{ik}(X_{ij}, \delta_{ij}, t).
\end{aligned}$$

Let w be a weight function, as defined in Section 4.2.2. Integrating both sides with respect to w gives

$$\begin{aligned}
\hat{\Omega}_{ik} - \Omega_{ik} &= \int_0^\infty \log \hat{\Lambda}_i^{(k)}(t) dW(t) - \int_0^\infty \log \Lambda_i^{(k)}(t) dW(t) \\
&\cong \frac{1}{n_{ik}} \sum_{j:\ z_{ij}=k} \int_0^\infty \frac{\xi_{ik}(X_{ij}, \delta_{ij}, t)}{\Lambda_i^{(k)}(t) S_i^{(k)}(t)} dW(t) \\
&= \frac{1}{n_{ik}} \sum_{j:\ z_{ij}=k} \eta_{ik}(X_{ij}, \delta_{ij}),
\end{aligned}$$

with $\eta_{ik}(X_{ij}, \delta_{ij}) = \int_0^\infty \frac{\xi_{ik}(X_{ij}, \delta_{ij}, t)}{\Lambda_i^{(k)}(t) S_i^{(k)}(t)} dW(t)$.

Noting that the function $\xi_{ik}(X_{ij}, \delta_{ij}, t)$ is a conditional version (conditioned on the cluster i and the subgroup with $z_{ij} = k$) of the function ξ in Lo and Singh (1986), it follows that for a subject with observed information (X_{ij}, δ_{ij}) in the subgroup with $z_{ij} = k$:

$$\begin{aligned}
E\left\{\xi_{ik}(X_{ij}, \delta_{ij}, t)\right\} &= 0, \\
\mathrm{Cov}\left\{\xi_{ik}(X_{ij}, \delta_{ij}, t), \xi_{ik}(X_{ij}, \delta_{ij}, s)\right\} &= \left\{1 - F_i^{(k)}(t)\right\}\left\{1 - F_i^{(k)}(s)\right\} \\
&\quad \times \int_0^{t \wedge s} \frac{dH_{ik}^u(y)}{\left\{1 - H_{ik}(y)\right\}^2}.
\end{aligned}$$

Assume that subject l in cluster i is in the subgroup with $z_{il} = k$. The asymptotic variance of the error terms $\hat{\Omega}_{ik} - \Omega_{ik}$ is given by

$$\begin{aligned}
\sigma^2_{e,ik} &= \mathrm{Var}(\hat{\Omega}_{ik} - \Omega_{ik}) \\
&= \mathrm{Var}\left\{ \frac{1}{n_{ik}} \sum_{j:\ z_{ij}=k} \eta_{ik}(X_{ij}, \delta_{ij}) \right\} \\
&= \frac{1}{n_{ik}} \mathrm{Var}\left\{ \eta(X_{il}, \delta_{il}) \right\} \\
&= \frac{1}{n_{ik}} \mathrm{E}\left\{ \int_0^\infty \frac{\xi_{ik}(X_{il}, \delta_{il}, t)}{\Lambda_i^{(k)}(t) S_i^{(k)}(t)} dW(t) \int_0^\infty \frac{\xi_{ik}(X_{il}, \delta_{il}, s)}{\Lambda_i^{(k)}(s) S_i^{(k)}(s)} dW(s) \right\} \\
&= \frac{1}{n_{ik}} \int_0^\infty \int_0^\infty \frac{1}{\Lambda_i^{(k)}(t) S_i^{(k)}(t) \Lambda_i^{(k)}(s) S_i^{(k)}(s)} \\
&\quad \times \mathrm{Cov}\left\{ \xi_{ik}(X_{il}, \delta_{il}, t), \xi_{ik}(X_{il}, \delta_{il}, s) \right\} dW(t) dW(s) \\
&= \frac{1}{n_{ik}} \int_0^\infty \int_0^\infty \frac{1}{\Lambda_i^{(k)}(t) \Lambda_i^{(k)}(s)} \int_0^{s \wedge t} \frac{dH^u_{ik}(y)}{\left\{1 - H_{ik}(y)\right\}^2} dW(t) dW(s).
\end{aligned}$$

Let W be a uniform weight function on the interval (A, B), where A and B are chosen so that $0 < \hat{S}_i^{(k)}(t) < 1$ for $t \in (A, B)$. Then

$$\begin{aligned}
\sigma^2_{e,ik} &= \frac{1}{n_{ik}} \frac{1}{(B-A)^2} \int_A^B \int_A^s \frac{1}{\Lambda_i^{(k)}(t) \Lambda_i^{(k)}(s)} \int_0^t \frac{dH^u_{ik}(y)}{\left\{1 - H_{ik}(y^-)\right\}^2} dt ds \\
&\quad + \frac{1}{n_{ik}} \frac{1}{(B-A)^2} \int_A^B \int_s^B \frac{1}{\Lambda_i^{(k)}(t) \Lambda_i^{(k)}(s)} \int_0^s \frac{dH^u_{ik}(y)}{\left\{1 - H_{ik}(y^-)\right\}^2} dt ds.
\end{aligned}$$

To obtain an estimate of the asymptotic error variance, we replace $H_{ik}(y^-)$ and $H^u_{ik}(y)$ by the following empirical estimators:

$$\begin{aligned}
\hat{H}^u_{ik}(y) &= \frac{1}{n_{ik}} \sum_{j:\ z_{ij}=k} I\left(X_{ij} \le y, \delta_{ij} = 1\right) \\
\hat{H}_{ik}(y^-) &= \frac{1}{n_{ik}} \sum_{j:\ z_{ij}=k} I\left(X_{ij} < y\right).
\end{aligned}$$

This gives the following estimated variances of the error terms:

$$\begin{aligned}\hat{\sigma}^2_{e,ik} &= \frac{1}{n_{ik}^2}\frac{1}{(B-A)^2}\int_A^B \frac{1}{\hat{\Lambda}_i^{(k)}(s)}\int_A^s \frac{1}{\hat{\Lambda}_i^{(k)}(t)} \\ &\quad \times \sum_{j:\ z_{ij}=k} \frac{I\left(0 \le x_{ij} \le t, \delta_{ij}=1\right)}{\left\{1-\frac{1}{n_{ik}}\ \sum_{j:\ z_{ij}=k} I\left(X_{ij} < x_{ij}\right)\right\}^2}\ dt\ ds \\ &\quad + \frac{1}{n_{ik}^2}\frac{1}{(B-A)^2}\int_A^B \frac{1}{\hat{\Lambda}_i^{(k)}(s)}\int_s^B \frac{1}{\hat{\Lambda}_i^{(k)}(t)} \\ &\quad \times \sum_{j:\ z_{ij}=k} \frac{I\left(0 \le x_{ij} \le s, \delta_{ij}=1\right)}{\left\{1-\frac{1}{n_{ik}}\sum_{j:\ z_{ij}=k} I\left(X_{ij} < x_{ij}\right)\right\}^2}\ dt\ ds.\end{aligned}$$

Multivariate proportional odds models

It follows from the first order Taylor expansion and the i.i.d. representation by Lo and Singh (1986) that

$$\begin{aligned}&\operatorname{logit}\left\{\hat{F}_i^{(k)}(t)\right\} - \operatorname{logit}\left\{F_i^{(k)}(t)\right\} \\ \cong\ & \frac{1}{F_i^{(k)}(t)\left\{1-F_i^{(k)}(t)\right\}}\left\{\hat{F}_i^{(k)}(t) - F_i^{(k)}(t)\right\} \\ \cong\ & \frac{1}{\left\{1-S_i^{(k)}(t)\right\}S_i^{(k)}(t)}\frac{1}{n_{ik}}\sum_{j:z_{ij}=k}\xi_{ik}\left(X_{ij},\delta_{ij},t\right).\end{aligned}$$

Integrating with respect to w gives

$$\begin{aligned}\hat{\Omega}_{PO;i}^{(k)} - \Omega_{PO;i}^{(k)} &= \int_0^\infty \operatorname{logit}\left\{\hat{F}_i^{(k)}(t)\right\} dW(t) - \int_0^\infty \operatorname{logit}\left\{F_i^{(k)}(t)\right\} dW(t) \\ &\cong \frac{1}{n_{ik}}\sum_{j:z_{ij}=k}\int_0^\infty \frac{\xi_{ik}\left(X_{ij},\delta_{ij},t\right)}{\left\{1-S_i^{(k)}(t)\right\}S_i^{(k)}(t)} dW(t).\end{aligned}$$

The estimated variance of the error terms $\hat{\Omega}_{PO;i}^{(k)} - \Omega_{PO;i}^{(k)}$ can be obtained using the same arguments as explained for the multivariate proportional hazards

model (frailty model):

$$\begin{aligned}\hat{\sigma}^2_{e\ PO,ik} &= \frac{1}{n_{ik}^2}\frac{1}{(B-A)^2}\int_A^B \frac{1}{\left\{1-\hat{S}_i^{(k)}(s)\right\}}\int_A^s \frac{1}{\left\{1-\hat{S}_i^{(k)}(t)\right\}} \\ &\quad\times \sum_{j:\ z_{ij}=k} \frac{I\left(0\le x_{ij}\le t, \delta_{ij}=1\right)}{\left\{1-\frac{1}{n_{ik}}\ \sum_{j:\ z_{ij}=k} I\left(X_{ij}<x_{ij}\right)\right\}^2}\ dt\ ds \\ &\quad+\frac{1}{n_{ik}^2}\frac{1}{(B-A)^2}\int_A^B \frac{1}{\left\{1-\hat{S}_i^{(k)}(s)\right\}}\int_s^B \frac{1}{\left\{1-\hat{S}_i^{(k)}(t)\right\}} \\ &\quad\times \sum_{j:\ z_{ij}=k} \frac{I\left(0\le x_{ij}\le s, \delta_{ij}=1\right)}{\left\{1-\frac{1}{n_{ik}}\sum_{j:\ z_{ij}=k} I\left(X_{ij}<x_{ij}\right)\right\}^2}\ dt\ ds.\end{aligned}$$

Multivariate additive risks models

Using $\Lambda_i^{(k)}(t) = -\log\left\{1-F_i^{(k)}(t)\right\}$ and the first order Taylor expansion gives

$$\begin{aligned}\hat{\Lambda}_i^{(k)}(t)-\Lambda_i^{(k)}(t) &\cong \frac{1}{1-F_i^{(k)}(t)}\left\{\hat{F}_i^{(k)}(t)-F_i^{(k)}(t)\right\} \\ &\cong \frac{1}{1-F_i^{(k)}(t)}\frac{1}{n_{ik}}\sum_{j:z_{ij}=k}\xi_{ik}\left(X_{ij},\delta_{ij},t\right).\end{aligned}$$

The last equation follows by Lo and Singh (1986). By integrating both sides with respect to the weight function $\tilde{W}(t) = W(t)/\int_0^\infty s dW(s)$, we obtain

$$\begin{aligned}\hat{\Omega}_{AR;i}^{(k)}-\Omega_{AR;i}^{(k)} &= \int_0^\infty \hat{\Lambda}_i^{(k)}(t)d\tilde{W}(t)-\int_0^\infty \Lambda_i^{(k)}(t)d\tilde{W}(t) \\ &\cong \frac{1}{n_{ik}}\sum_{j:z_{ij}=k}\int_0^\infty \frac{\xi_{ik}\left(X_{ij},\delta_{ij},t\right)}{S_i^{(k)}(t)}d\tilde{W}(t).\end{aligned}$$

As in the previous sections, we choose a uniform weight function W on the interval (A,B). The estimated variance of the error terms $\hat{\Omega}_{AR;i}^{(k)}-\Omega_{AR;i}^{(k)}$ can

be obtained in a similar way as the discussion given for the frailty model:

$$\begin{aligned}
&\hat{\sigma}^2_{e\ AR,ik} \\
= \ & \frac{1}{n_{ik}^2}\frac{4}{\left(B^2-A^2\right)^2}\int_A^B\int_A^s \sum_{j:\ z_{ij}=k} \frac{I\left(0\le x_{ij}\le t, \delta_{ij}=1\right)}{\left\{1-\frac{1}{n_{ik}}\ \sum_{j:\ z_{ij}=k} I\left(X_{ij}<x_{ij}\right)\right\}^2}\ dtds \\
& +\frac{1}{n_{ik}^2}\frac{4}{\left(B^2-A^2\right)^2}\int_A^B\int_s^B \sum_{j:\ z_{ij}=k} \frac{I\left(0\le x_{ij}\le s, \delta_{ij}=1\right)}{\left\{1-\frac{1}{n_{ik}}\sum_{j:\ z_{ij}=k} I\left(X_{ij}<x_{ij}\right)\right\}^2}\ dtds.
\end{aligned}$$

4.7.2 Fitting the linear mixed-effects model

To fit the transformed models (4.6) and (4.7) in Section 4.2.2, we use PROC MIXED in SAS. The mixed-effects model is written as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\gamma} + \mathbf{e}, \tag{4.12}$$

where $\mathbf{y}$ denotes the vector of dependent variable values, $\boldsymbol{\beta}$ is an unknown vector of fixed effects with known model matrix $\mathbf{X}$, $\boldsymbol{\gamma}$ is an unknown vector of random effects with known model matrix $\mathbf{Z}$, and $\mathbf{e}$ is the random error vector. A key assumption is that $\boldsymbol{\gamma}$ and $\mathbf{e}$ are normally distributed with

$$\mathrm{E}\begin{pmatrix}\boldsymbol{\gamma}\\ \mathbf{e}\end{pmatrix} = \begin{pmatrix}0\\0\end{pmatrix} \quad \text{and} \quad \mathcal{D}\begin{pmatrix}\boldsymbol{\gamma}\\ \mathbf{e}\end{pmatrix} = \begin{pmatrix}\mathbf{G} & 0\\ 0 & \mathbf{R}\end{pmatrix}.$$

The variance-covariance matrix of $\mathbf{y}$ is therefore $\mathbf{V} = \mathbf{ZGZ'} + \mathbf{R}$. To estimate the variance-covariance components in model (4.12), PROC MIXED implements two likelihood-based methods: maximum likelihood (ML) and restricted/residual likelihood (REML). We will consider the REML method. The corresponding loglikelihood function is:

$$l_R(\mathbf{G},\mathbf{R}) = -\frac{1}{2}\log|\mathbf{V}| - \frac{1}{2}\log|\mathbf{X'V^{-1}X}| - \frac{1}{2}\mathbf{r}'V^{-1}\mathbf{r} - \frac{n-p}{2}\log 2\pi, \tag{4.13}$$

where $\mathbf{r} = \mathbf{y} - \mathbf{X}(\mathbf{X'V}^{-1}\mathbf{X})^{-1}\mathbf{X'V}^{-1}\mathbf{y}$ and p is the rank of $\mathbf{X}$. PROC MIXED minimizes $-2\ l_R(\mathbf{G},\mathbf{R})$ over all unknown parameters using a ridge-stabilized Newton-Raphson algorithm.

For model (4.7) in Section 4.2.2, $\mathbf{y}' = \left(\hat{\Omega}_1, \hat{\Omega}_2, \ldots, \hat{\Omega}_K\right)$, $\mathbf{X} = \mathbf{1}_K$, $\mathbf{Z} = \mathbf{I}_K$,

$\boldsymbol{\gamma}' = (b_{01}, b_{02}, \ \ldots, b_{0K})$, $\mathbf{G} = \sigma_0^2 \mathbf{I}_K$ and $\mathbf{R} = \text{diag}(\sigma_{e,1}^2, \ldots, \sigma_{e,K}^2)$. As already mentioned, we only have one observation per level in model (4.7). To be able to estimate σ_0^2, we first estimate $\sigma_{e,1}^2, \ldots, \sigma_{e,K}^2$ as explained in Section 4.2.3. In the PARMS statement of PROC MIXED, initial values for the covariance parameters can be specified. We choose an arbitrary initial value for σ_0^2. The initial values for the error variances are chosen to be $\hat{\sigma}_{e,1}^2, \ldots, \hat{\sigma}_{e,K}^2$. By using the option EQCONS, the initial residual variances will be held constant during the estimation procedure. Maximisation of (4.13) over σ_0^2 gives an estimate for the heterogeneity σ_0^2. The following SAS code fits model (4.7) to the pseudo data for 20 clusters:

```
proc mixed data=pseudodata;
class cluster;
model omegaihat= ;
random cluster;
repeated/group=cluster;
parms /parmsdata=parmsdataset eqcons= 2 to 21;
run;
```

where *parmsdataset* is a SAS data set that contains the initial values for σ_0^2, $\sigma_{e,1}^2$, ..., $\sigma_{e,K}^2$.
For model (4.6), $\mathbf{y}' = \left(\hat{\Omega}_{10}, \hat{\Omega}_{11}, \hat{\Omega}_{20}, \hat{\Omega}_{21}, \ldots, \hat{\Omega}_{K0}, \hat{\Omega}_{K1}\right)$,

$$\mathbf{X} = \mathbf{I}_K \otimes \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}, \qquad \mathbf{Z} = \mathbf{I}_K \otimes \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix},$$

$\boldsymbol{\gamma}' = \mathbf{b}'$ and $\mathbf{G}$ is as defined in Section 4.2.1. Further, $\mathbf{R} = \text{diag}(\sigma_{e,10}^2, \sigma_{e,11}^2, \ldots, \sigma_{e,K0}^2, \sigma_{e,K1}^2)$. The error covariance matrix $\mathbf{R}$ can be estimated as explained in Section 4.2.3. By maximising (4.13) over $\mathbf{G}$ in PROC MIXED while fixing the error variances as described above, we obtain estimates for σ_0^2, σ_1^2 and σ_{01}. To obtain estimates of $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$, the mixed model equations are solved (Henderson, 1984). The solutions can be written as $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{X})^{-1}\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{y}$ and $\hat{\boldsymbol{\gamma}} = \hat{\mathbf{G}}\mathbf{Z}'\hat{\mathbf{V}}^{-1}(\mathbf{y} - \mathbf{X}\,\hat{\boldsymbol{\beta}})$.

Table 4.5: Relative bias, mean and empirical standard deviation of the estimated values $\hat{\sigma}_0^2$ over the 500 simulations; true $\sigma_0^2 = 0.08765$ ($\theta = 0.1$), $\lambda_0 = 0.1$; first line for coxph, second line for PROC MIXED.

	30 % censoring			60 % censoring		
(K, n)	Rel. bias	Mean	Emp. std	Rel. bias	Mean	Emp. std
(100,100)	-0.0222	0.0857	0.0146	-0.0200	0.0859	0.0170
	-0.0177	0.0861	0.0146	-0.0234	0.0856	0.0167
(100,50)	-0.0131	0.0865	0.0163	-0.0188	0.0860	0.0190
	-0.0793	0.0807	0.0164	-0.1010	0.0788	0.0189
(50,100)	-0.0051	0.0872	0.0204	-0.0280	0.0852	0.0232
	0.0017	0.0878	0.0210	-0.0234	0.0856	0.0238
(50,50)	-0.0200	0.0859	0.0245	-0.0211	0.0858	0.0281
	-0.0793	0.0807	0.0238	-0.0975	0.0791	0.0286
(20,100)	-0.0416	0.0840	0.0303	-0.0941	0.0794	0.0339
	-0.0086	0.0869	0.0319	-0.0645	0.0820	0.0354
(20,50)	-0.0690	0.0816	0.0355	-0.0747	0.0811	0.0397
	-0.0998	0.0789	0.0373	-0.1409	0.0753	0.0405

Table 4.6: Relative bias, mean and empirical standard deviation of the estimated values $\hat{\sigma}_0^2$ over the 500 simulations; true $\sigma_0^2 = 0.1577$ ($\theta = 0.2$), $\lambda_0 = 0.5$; first line for coxph, second line for PROC MIXED.

	30 % censoring			60 % censoring		
(K, n)	Rel. bias	Mean	Emp. std	Rel. bias	Mean	Emp. std
(100,100)	-0.0120	0.1558	0.0264	-0.0292	0.1531	0.0253
	-0.0120	0.1558	0.0261	-0.0323	0.1526	0.0253
(100,50)	-0.0355	0.1521	0.0272	-0.0406	0.1513	0.0282
	-0.0938	0.1429	0.0263	-0.1046	0.1412	0.0281
(50,100)	-0.0184	0.1548	0.0348	-0.0374	0.1518	0.0375
	-0.0101	0.1561	0.0351	-0.0317	0.1527	0.0382
(50,50)	-0.0146	0.1554	0.0388	-0.0609	0.1481	0.0400
	-0.0590	0.1484	0.0392	-0.1141	0.1397	0.0396
(20,100)	-0.0140	0.1555	0.0536	-0.0615	0.1480	0.0561
	0.0184	0.1606	0.0578	-0.0247	0.1538	0.0593
(20,50)	-0.0704	0.1466	0.0576	-0.0653	0.1474	0.0668
	-0.0900	0.1435	0.0610	-0.0964	0.1425	0.0672

Table 4.7: Relative bias, mean and empirical standard deviation of the estimated values $\hat{\sigma}_0^2$ over the 500 simulations; true $\sigma_0^2 = 0.1577$ ($\theta = 0.2$), $\lambda_0 = 0.1$; first line for coxph, second line for PROC MIXED.

	30 % censoring			60 % censoring		
(K, n)	Rel. bias	Mean	Emp. std	Rel. bias	Mean	Emp. std
(100,100)	-0.0127	0.1557	0.0265	-0.0120	0.1558	0.0276
	-0.0127	0.1557	0.0260	-0.0127	0.1557	0.0272
(100,50)	-0.0082	0.1564	0.0265	-0.0431	0.1509	0.0290
	-0.0634	0.1477	0.0267	-0.1053	0.1411	0.0296
(50,100)	-0.0152	0.1553	0.0345	-0.0520	0.1495	0.0357
	-0.0108	0.1560	0.0354	-0.0476	0.1502	0.0370
(50,50)	-0.0342	0.1523	0.0390	-0.0311	0.1528	0.0429
	-0.0755	0.1458	0.0398	-0.0881	0.1438	0.0436
(20,100)	-0.0317	0.1527	0.0538	-0.0653	0.1474	0.0571
	0.0038	0.1583	0.0573	-0.0317	0.1527	0.0599
(20,50)	-0.0330	0.1525	0.0568	-0.0977	0.1423	0.0635
	-0.0476	0.1502	0.0595	-0.1344	0.1365	0.0654

Table 4.8: Mean, empirical standard deviation and average of the model-based standard deviations of the estimated values over the 500 simulations; 50 centers, 100 patients per center (50 patients in control and treatment group); $\lambda_0 = 0.3$, $\beta = -0.2$.

		PROC MIXED			coxme	
	True	Mean	Emp. std.	Model std.	Mean	Emp. std.
β	-0.20	-0.2005	0.0184	0.0195	-0.2005	0.0167
σ_0^2	0	0.0009	0.0017	0.0013	0.0010	0.0011
σ_1^2	0	0.0009	0.0017	0.0014	0.0007	0.0014
σ_{01}	0	0.0006	0.0030	0.0030	0.0002	0.0003
β	-0.20	-0.2002	0.0429	0.0440	-0.2033	0.0425
σ_0^2	0	0.0006	0.0015	0.0012	0.0010	0.0016
σ_1^2	0.08	0.0773	0.0199	0.0200	0.0794	0.0127
σ_{01}	0	0.0006	0.0075	0.0073	0.0013	0.0026
β	-0.20	-0.1966	0.0438	0.0434	-0.1980	0.0440
σ_0^2	0.04	0.0371	0.0117	0.0120	0.0405	0.0110
σ_1^2	0.08	0.0738	0.0191	0.0195	0.0771	0.0185
σ_{01}	0.0283	0.0266	0.0111	0.0118	0.0282	0.0105
β	-0.20	-0.1981	0.0200	0.0204	-0.1999	0.0172
σ_0^2	0.08	0.0778	0.0201	0.0203	0.0798	0.0119
σ_1^2	0	0.0008	0.0016	0.0015	0.0011	0.0017
σ_{01}	0	0.0005	0.0075	0.0071	0.0013	0.0026
β	-0.20	-0.1945	0.0333	0.0336	-0.1973	0.0318
σ_0^2	0.08	0.0737	0.0193	0.0196	0.0773	0.0186
σ_1^2	0.04	0.0357	0.0114	0.0117	0.0384	0.0108
σ_{01}	0.0283	0.0255	0.0115	0.0116	0.0271	0.0108
β	-0.20	-0.1964	0.0442	0.0438	-0.1996	0.0436
σ_0^2	0.08	0.0715	0.0193	0.0194	0.0763	0.0186
σ_1^2	0.08	0.0743	0.0189	0.0199	0.0785	0.0174
σ_{01}	0.04	0.0358	0.0142	0.0152	0.0384	0.0139

Chapter 5

Two-stage estimation in copula models: methodology

5.1 Introduction

In Section 2.5 we introduced the copula model as a possible way to model the joint survival function of clustered data with small and equal cluster size. We described some copula models, nested in the power variance copula family, which can be used to model four-dimensional survival data. In this chapter we explain how copula models for quadruples can be fitted using the two-stage estimation approach. We estimate the copula parameter vector by maximising a likelihood function in which the marginal survival functions are replaced by their estimates. We study both semi-parametric and nonparametric estimation of the marginal survival functions. Our results extend results obtained by Glidden (2000) and Andersen (2005) on the asymptotic behaviour of the estimators of the marginal survival functions, the cumulative hazard functions and the copula parameter vector.

5.2 Estimation method

For observational unit j, $j = 1, \ldots, 4$, from cluster i, $i = 1, \ldots, K$, we observe a vector of covariates $\mathbf{z}_{ij}$. We introduce some additional notation to simplify the expression of the likelihood:

$$\begin{aligned}
\Delta_i &= \prod_{j=1}^{4} (1 - \delta_{ij}) \\
\Delta_i(j) &= \delta_{ij} \prod_{k=1;k\neq j}^{4} (1 - \delta_{ik}) \\
\Delta_i(j,k) &= \delta_{ij}\delta_{ik} \prod_{l=1;l\neq j,k}^{4} (1 - \delta_{il}) \quad , j \neq k \\
\Delta_i(j,k,l) &= \delta_{ij}\delta_{ik}\delta_{il}(1 - \delta_{im}) \quad , m \neq j,k,l;\ j \neq k \neq l \\
\Delta_i(1,2,3,4) &= \prod_{j=1}^{4} \delta_{ij}.
\end{aligned}$$

Let $\boldsymbol{\zeta}$ be the copula parameter vector, also called the dependence or association parameter vector. Let $S_i(.;\boldsymbol{\zeta})$ be the joint survival function for $(T_{i1}, \ldots, T_{i4})$. Denote the marginal survival functions by S_{ij}, where the index ij is used to indicate the dependence on a covariate vector $\mathbf{z}_{ij}$. The joint survival function is characterized by the copula $C_{\boldsymbol{\zeta}}$, that describes the dependence structure, and the marginal survival functions S_{ij}:

$$S_i(t_1, t_2, t_3, t_4; \boldsymbol{\zeta}) = C_{\boldsymbol{\zeta}}\left\{S_{i1}(t_1), S_{i2}(t_2), S_{i3}(t_3), S_{i4}(t_4)\right\}. \tag{5.1}$$

A two-stage estimation approach is often used to fit copula models. In the first stage, the marginal survival functions are estimated. This can be done in different ways. Shih and Louis (1995b) discuss parametric and nonparametric (Kaplan-Meier) estimation of the marginal survival functions. Their work is on bivariate survival data without covariates. Shih and Louis (1995b) use a different survival function for each member of the pair. For multivariate failure time data, Glidden (2000) considers semi-parametric two-stage estimation for the Clayton copula. The marginal survival functions are modelled through a marginal Cox model that can depend on covariates and all the data are used

to fit this Cox model. For bivariate failure times, Andersen (2005) extends the semi-parametric approach of Glidden (2000) to any copula family, including copula families with more than one parameter. She also studies parametric estimation of the marginal survival functions, where the two components have the same marginal survival function or where the two marginal survival functions depend on covariates. In this chapter, we focus on survival copulas for quadruples. We consider both a semi-parametric and a nonparametric approach to model the marginal survival functions. We show that the semi-parametric or nonparametric estimates of the marginal survival functions are consistent, and therefore can be used in the second step of a two-stage estimation approach. In that second step, we estimate the association parameter $\boldsymbol{\zeta}$ by maximising the loglikelihood, with the marginal survival functions replaced by their estimates obtained in the first step.
Using as notation $v_{ij} = S_{ij}\left(X_{ij}\right)$, the loglikelihood is given by

$$\begin{aligned}
\log L\left(\boldsymbol{\zeta}\right) \;=\; & \sum_{i=1}^{K}\Bigg[\Delta_i \log\left\{C_{\boldsymbol{\zeta}}\left(v_{i1},v_{i2},v_{i3},v_{i4}\right)\right\} \\
& + \sum_{j=1}^{4}\left[\Delta_i(j)\log\left\{\frac{\partial C_{\boldsymbol{\zeta}}\left(v_{i1},v_{i2},v_{i3},v_{i4}\right)}{\partial v_{ij}}\right\}\right] \\
& + \sum_{j\neq k}\left[\Delta_i(j,k)\log\left\{\frac{\partial^2 C_{\boldsymbol{\zeta}}\left(v_{i1},v_{i2},v_{i3},v_{i4}\right)}{\partial v_{ij}\;\partial v_{ik}}\right\}\right] \\
& + \sum_{j\neq k\neq l}\left[\Delta_i(j,k,l)\log\left\{\frac{\partial^3 C_{\boldsymbol{\zeta}}\left(v_{i1},v_{i2},v_{i3},v_{i4}\right)}{\partial v_{ij}\;\partial v_{ik}\;\partial v_{il}}\right\}\right] \\
& + \Delta_i(1,2,3,4)\log\left\{\frac{\partial^4 C_{\boldsymbol{\zeta}}\left(v_{i1},v_{i2},v_{i3},v_{i4}\right)}{\partial v_{i1}\;\partial v_{i2}\;\partial v_{i3}\;\partial v_{i4}}\right\}\Bigg]. \qquad (5.2)
\end{aligned}$$

In the second stage of the estimation, we replace in this loglikelihood expression $S_{ij}(X_{ij})$ by the estimated marginals $\hat{S}_{ij}(X_{ij})$ from the first step; the obtained expression is called the pseudo loglikelihood and is denoted by $\log L_P\left(\boldsymbol{\zeta}\right)$. To estimate $\boldsymbol{\zeta}$, the pseudo loglikelihood is maximised with respect to $\boldsymbol{\zeta}$, i.e., $\hat{\boldsymbol{\zeta}}$ is found by solving the following vector equation:

$$\mathbf{U}_{\boldsymbol{\zeta}}\left(\boldsymbol{\zeta}\right) = \frac{\partial}{\partial \boldsymbol{\zeta}}\log L_P\left(\boldsymbol{\zeta}\right) = \mathbf{0}. \qquad (5.3)$$

In Sections 5.3 and 5.4 we give consistency and asymptotic normality results for the estimators in the marginals and for the estimator of the dependence parameter vector $\boldsymbol{\zeta}$. These results provide support for the two-stage estimation method where the marginals are modelled in a semi-parametric or nonparametric way.

5.3 Semi-parametric approach

In the semi-parametric estimation approach, we model the marginal survival functions in the first step using a marginal Cox model with covariate vector $\mathbf{z}_{ij}$. In the second step, we replace the marginal survival functions in the loglikelihood by their estimates obtained in the first step. We then estimate the copula parameter $\boldsymbol{\zeta}$ by solving the score equation (5.3). Glidden (2000) studies this method for the Clayton copula in the case of multivariate failure time data. Andersen (2005) extends the results of Glidden (2000) to any copula family. However, she only considers bivariate survival data. We consider the use of copulas to model four-dimensional survival data and we study the semi-parametric approach for any copula family.

The asymptotic theory for the estimator of the copula parameter discussed in Glidden (2000) and Andersen (2005), builds on asymptotic theory, developed by Spiekerman and Lin (1998), for the estimators of the parameters in the marginal Cox model. For observational unit j in cluster i, Spiekerman and Lin (1998) consider a stochastic covariate vector $\mathbf{Z}_{ij}$. Let $\mathbf{T}_i = (T_{i1}, \ldots, T_{i4})'$ be the vector of failure times in cluster i and define $\mathbf{C}_i$ and $\mathbf{Z}_i$ similarly. Spiekerman and Lin (1998) then assume that $(\mathbf{T}_i, \mathbf{C}_i, \mathbf{Z}_i)$, for $i = 1, \ldots, K$, are independent and identically distributed. They use this assumption in their proof of the asymptotic normality of the parameter estimators in the marginal Cox model. We are interested in modelling the dependence of the infection times within a cluster in the udder infection data by using copula models. In the estimation of the marginal survival functions, we take into account the effect of the location of the udder quarter (front or rear) and the effect of the parity using covariates. Both covariates are fixed. Therefore, the situation we

consider is different from the situation in Spiekerman and Lin (1998).
In Section 5.3.1 we introduce definitions and assumptions that we need in the proofs of the theorems in Sections 5.3.2 and 5.3.3. In Section 5.3.2 we describe the marginal Cox model for a general cluster size n and explain how we can obtain the estimators for the fixed effects parameter and for the cumulative baseline hazard. We extend the asymptotic results of Spiekerman and Lin (1998) to specific cases of deterministic covariates. In Section 5.3.3 we give consistency and asymptotic normality results for the estimated copula parameter $\hat{\boldsymbol{\zeta}}$. We show how the results of Glidden (2000) and Andersen (2005) can be generalized to a general copula family for four-dimensional survival data.

5.3.1 Definitions and assumptions

Since the asymptotic results in Section 5.3.2 hold for a general cluster size n, we assume that we have K clusters with n observations per cluster. We observe a covariate vector $\mathbf{z}_{ij} = (z_{ij1}, \ldots, z_{ijp})'$. To derive the asymptotic properties of the parameter estimators in the marginal Cox model, we use counting process notation. Define the counting process $N_{ij}(t) = \delta_{ij} I(X_{ij} \leq t)$ and the at risk process $Y_{ij}(t) = I(X_{ij} \geq t)$, for the jth observation in the ith cluster, where $j = 1, \ldots, n$ and $i = 1, \ldots, K$. Note that $\{N_{ij}(t), Y_{ij}(t), \mathbf{z}_{ij} : t \geq 0\}$ carries the same information about the jth observation in cluster i as does $(X_{ij}, \delta_{ij}, \mathbf{z}_{ij})$. We assume that $\{N_{ij}(t), Y_{ij}(t)\}$, for $j = 1, \ldots, n$ and $i = 1, \ldots, K$, are observed in some time interval $[0, \tau], \tau < \infty$. Define

$$M_{ij}(t) = N_{ij}(t) - e^{\boldsymbol{\beta}' \mathbf{z}_{ij}} \int_0^t Y_{ij}(u) \lambda_0(u) du. \tag{5.4}$$

It is important to note that $M_{ij}(t)$ is a martingale with respect to the marginal filtration

$$\mathcal{F}_{t,ij} = \sigma \left\{ N_{ij}(u), Y_{ij}(u^+) : 0 \leq u \leq t \right\}.$$

This is natural because $\mathcal{F}_{t,ij}$ collects the history of the processes N_{ij}, Y_{ij} over the period $[0, t]$. However, due to the intra-cluster dependence, $M_{ij}(t)$ is not a martingale with respect to the joint filtration $\mathcal{F}_t = \bigvee_{i=1}^{K} \bigvee_{j=1}^{n} \mathcal{F}_{t,ij}$, where for σ-fields $\mathcal{A}_l$, $\bigvee_{l=1}^{L} \mathcal{A}_i$ denotes the smallest σ-field containing $\{\mathcal{A}_l : l = 1, \ldots, L\}$

(see, e.g., Spiekerman and Lin, 1998; Martinussen and Scheike, 2006, p.315). This limits the use of martingale theory in the proofs of the theorems in this section. However, using non-standard methods, it is still possible to perform statistical inference for the parameters in the marginal Cox model.
We use the following notation:

$$\begin{aligned}
\mathbf{S}^{(r)}(\boldsymbol{\beta}, u) &= K^{-1} \sum_{i=1}^{K} \sum_{j=1}^{n} Y_{ij}(u) \exp(\boldsymbol{\beta}' \mathbf{z}_{ij}) \mathbf{z}_{ij}^{\otimes r}, \text{ for } r = 0, 1, 2, \\
\mathbf{E}(\boldsymbol{\beta}, u) &= \mathbf{S}^{(1)}(\boldsymbol{\beta}, u) / S^{(0)}(\boldsymbol{\beta}, u), \\
\mathbf{V}(\boldsymbol{\beta}, u) &= \mathbf{S}^{(2)}(\boldsymbol{\beta}, u) / S^{(0)}(\boldsymbol{\beta}, u) - \mathbf{E}(\boldsymbol{\beta}, u)^{\otimes 2},
\end{aligned}$$

and

$$\begin{aligned}
\mathbf{s}^{(r)}(\boldsymbol{\beta}, u) &= \mathrm{E}\left\{ \mathbf{S}^{(r)}(\boldsymbol{\beta}, u) \right\}, \text{ for } r = 0, 1, 2, \\
\mathbf{e}(\boldsymbol{\beta}, u) &= \mathbf{s}^{(1)}(\boldsymbol{\beta}, u) / s^{(0)}(\boldsymbol{\beta}, u), \\
\mathbf{v}(\boldsymbol{\beta}, u) &= \mathbf{s}^{(2)}(\boldsymbol{\beta}, u) / s^{(0)}(\boldsymbol{\beta}, u) - \mathbf{e}(\boldsymbol{\beta}, u)^{\otimes 2},
\end{aligned}$$

where for a column vector $\mathbf{a}$, $\mathbf{a}^{\otimes 0} = 1$, $\mathbf{a}^{\otimes 1} = \mathbf{a}$ and $\mathbf{a}^{\otimes 2} = \mathbf{a}\mathbf{a}'$. For a function f with domain $[0, \tau]$, we define the norm

$$||f||_{[0,\tau]} = \sup_{t \in [0,\tau]} |f(t)|,$$

where, if f is vector-valued, $|f(t)|$ denotes the Euclidean norm.
We assume the following conditions:

(A1) $P\{Y_{ij}(t) = 1, \text{for all } t \in [0, \tau]\} > 0$, for all i, j.

(A2) $|z_{ijm}| < B$ for all i, j, for $m = 1, \ldots, p$ and for some constant $B < \infty$.

(A3) $\mathbf{I}_{\boldsymbol{\beta}} = \int_0^\tau \mathbf{v}(\boldsymbol{\beta}_0, t) s^{(0)}(\boldsymbol{\beta}_0, t) \lambda_0(t) dt$ is positive definite.

Conditions (A1) and (A2) entail some conditions that are useful in the proofs of Section 5.3.2:

Corollary 5.3.1. *Conditions (A1) and (A2) imply the following conditions:*

(C1) $\int_0^\tau \lambda_0(u) du \equiv B_L < \infty$.

(C2) There exists a neighbourhood $\mathcal{B}$ of $\boldsymbol{\beta}_0$ such that for $r = 0, 1, 2$:

$$\left|\left|\mathbf{S}^{(r)}(\boldsymbol{\beta}, t) - \mathbf{s}^{(r)}(\boldsymbol{\beta}, t)\right|\right|_{\mathcal{B}\times[0,\tau]} \xrightarrow{P} 0.$$

(C3) $\mathbf{s}^{(r)}(\boldsymbol{\beta}, t)$*, for* $r = 0, 1, 2$*, are continuous functions of* $\boldsymbol{\beta} \in \mathcal{B}$ *uniformly in* $t \in [0, \tau]$ *and are bounded on* $\mathcal{B} \times [0, \tau]$*,* $s^{(0)}(\boldsymbol{\beta}, t)$ *is bounded away from zero on* $\mathcal{B} \times [0, \tau]$*, i.e., there exist constants* L_0*,* U_0*,* U_1 *such that*

$$0 < L_0 \leq \left|\left|s^{(0)}(\boldsymbol{\beta}, t)\right|\right|_{\mathcal{B}\times[0,\tau]} \leq U_0,$$
$$\left|\left|\mathbf{s}^{(1)}(\boldsymbol{\beta}, t)\right|\right|_{\mathcal{B}\times[0,\tau]} \leq U_1.$$

Further,

$$\mathbf{s}^{(1)}(\boldsymbol{\beta}, t) = \frac{\partial}{\partial \boldsymbol{\beta}} s^{(0)}(\boldsymbol{\beta}, t) \tag{5.5}$$

and

$$\mathbf{s}^{(2)}(\boldsymbol{\beta}, t) = \frac{\partial^2}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} s^{(0)}(\boldsymbol{\beta}, t), \tag{5.6}$$

for $\boldsymbol{\beta} \in \mathcal{B}$ *and* $t \in [0, \tau]$*.*

Proof. Under the marginal Cox model we have

$$P\left(T_{ij} > \tau\right) = S_{ij}(\tau) = \exp\left\{-\exp\left(\boldsymbol{\beta}_0' \mathbf{z}_{ij}\right) \int_0^\tau \lambda_0(u) du\right\}.$$

Condition (A1) implies that $P\left(T_{ij} > \tau\right) > 0$. By condition (A2), we have that $\left|\exp\left(\boldsymbol{\beta}_0' \mathbf{z}_{ij}\right)\right|$ is bounded. Hence, condition (C1) follows from (A1) and (A2). Using similar arguments as discussed in Fleming and Harrington (1991, p.305-306), we obtain that condition (C2) follows from (A1) and (A2). We explain this for $S^{(0)}(\boldsymbol{\beta}, t)$. The proof for $\mathbf{S}^{(1)}(\boldsymbol{\beta}, t)$ and $\mathbf{S}^{(2)}(\boldsymbol{\beta}, t)$ is similar. Note that $S^{(0)}(\boldsymbol{\beta}, t)$ is a sum of independent random variables that are not identically distributed due to the presence of deterministic covariates. Therefore, we cannot use the strong law of large numbers which is used in the proof of Fleming and Harrington (1991). However, application of Theorem 5.6.1 in Appendix 5.6.1, with $X_i = \sum_{j=1}^n \left[Y_{ij}(u) - E\left\{Y_{ij}(u)\right\}\right] \exp\left(\boldsymbol{\beta}' \mathbf{z}_{ij}\right)$, provides the necessary result in our situation. We now check the conditions of Theorem 5.6.1. It is

easy to see that $E(X_i) = 0$. Further, we have that

$$\begin{aligned}
\sigma_i^2 &= \mathrm{Var}(X_i) \\
&= \sum_{j=1}^{n} \mathrm{Var}\left[Y_{ij}(u) - E\left\{Y_{ij}(u)\right\}\right] \left\{\exp\left(\boldsymbol{\beta}'\mathbf{z}_{ij}\right)\right\}^2 \\
&\quad + 2\sum\sum_{1\le j<k\le n} \Big[\mathrm{Cov}\left[Y_{ij}(u) - E\left\{Y_{ij}(u)\right\}, Y_{ik}(u) - E\left\{Y_{ik}(u)\right\}\right] \\
&\quad \times \left\{\exp\left(\boldsymbol{\beta}'\mathbf{z}_{ij}\right)\right\}\left\{\exp\left(\boldsymbol{\beta}'\mathbf{z}_{ik}\right)\right\}\Big] \\
&= \sum_{j=1}^{n} \mathrm{Var}\left\{Y_{ij}(u)\right\} \left\{\exp\left(\boldsymbol{\beta}'\mathbf{z}_{ij}\right)\right\}^2 \\
&\quad + 2\sum\sum_{1\le j<k\le n} \mathrm{Cov}\left\{Y_{ij}(u), Y_{ik}(u)\right\} \left\{\exp\left(\boldsymbol{\beta}'\mathbf{z}_{ij}\right)\right\}\left\{\exp\left(\boldsymbol{\beta}'\mathbf{z}_{ik}\right)\right\}.
\end{aligned}$$

Easy calculations show that an upper bound of the latter equation is given by $n(2n-1)\left\{\exp\left(|\boldsymbol{\beta}|pB\right)\right\}^2$, where B is as defined in condition (A2). It easily follows that

$$\sum_{i=1}^{\infty} i^{-2}\sigma_i^2 \;\le\; n(2n-1)\left\{\exp\left(|\boldsymbol{\beta}|pB\right)\right\}^2 \sum_{i=1}^{\infty} i^{-2} < \infty.$$

By Theorem 5.6.1 we then have that

$$S^{(0)}(\boldsymbol{\beta}, u) - s^{(0)}(\boldsymbol{\beta}, u) \xrightarrow{a.s.} 0 \text{ as } K \to \infty.$$

The proof that condition (C2) follows from (A1) and (A2) continues along the lines of Fleming and Harrington (1991, p.305-306).

We now show that (C3) follows from (A1) and (A2). Let "$\vee$" be the binary operator "maximum". For $t \in [0, \tau]$, and $\boldsymbol{\beta}, \boldsymbol{\delta} \in \mathbb{R}^p$, we have that

$$\begin{aligned}
&\left|s^{(0)}(\boldsymbol{\beta}+\boldsymbol{\delta}, t) - s^{(0)}(\boldsymbol{\beta}, t)\right| \\
= &\left|E\left[K^{-1}\sum_{i=1}^{K}\sum_{j=1}^{n} Y_{ij}(u)\left[\exp\left\{(\boldsymbol{\beta}+\boldsymbol{\delta})'\mathbf{z}_{ij}\right\} - \exp\left(\boldsymbol{\beta}'\mathbf{z}_{ij}\right)\right]\right]\right| \\
\le &\; K^{-1}\sum_{i=1}^{K}\sum_{j=1}^{n}\left|\exp\left\{(\boldsymbol{\beta}+\boldsymbol{\delta})'\mathbf{z}_{ij}\right\} - \exp\left(\boldsymbol{\beta}'\mathbf{z}_{ij}\right)\right|. \qquad (5.7)
\end{aligned}$$

It follows from condition (A2) and some easy calculations that an upperbound for (5.7) is given by

$$n\exp\left\{pB(|\boldsymbol{\beta}+\boldsymbol{\delta}| \vee |\boldsymbol{\beta}|)\right\} pB|\boldsymbol{\delta}|.$$

This shows that $s^{(0)}(\boldsymbol{\beta}, t)$ is continuous in $\boldsymbol{\beta}$, uniformly for $t \in [0, \tau]$. Similar inequalities can be obtained to prove the continuity of $\mathbf{s}^{(1)}(\boldsymbol{\beta}, t)$ and $\mathbf{s}^{(2)}(\boldsymbol{\beta}, t)$ in $\boldsymbol{\beta}$, uniformly in $t \in [0, \tau]$. Condition (A2) implies that $s^{(0)}(\boldsymbol{\beta}, t)$ and $\mathbf{s}^{(r)}(\boldsymbol{\beta}, t)$, for $r = 1, 2$, are bounded on $\mathcal{B} \times [0, \tau]$. Note that, for $t \in [0, \tau]$, $s^{(0)}(\boldsymbol{\beta}, t) \geq s^{(0)}(\boldsymbol{\beta}, \tau)$ and

$$s^{(0)}(\boldsymbol{\beta}, \tau) \geq \exp\left(-pB|\boldsymbol{\beta}|\right) K^{-1} \sum_{i=1}^{K} \sum_{j=1}^{n} P\left\{Y_{ij}(\tau) = 1\right\},$$

which is bounded away from zero for $\boldsymbol{\beta} \in \mathcal{B}$ by condition (A1). Condition (A2) and the dominated convergence theorem allow us to interchange the differentiation and the expectation to obtain equations (5.5) and (5.6). This completes the proof for (C3). □

To prove the asymptotic normality of the estimators for the fixed effects parameter and for the cumulative baseline hazard in Section 5.3.2, we need to restrict to specific cases of deterministic covariates. In condition (A4) we give one specific choice of deterministic covariates. This choice corresponds to the situation of the udder infection data, as we explain further in this section. In Theorem 5.3.2, Lemma 5.3.2 and Theorem 5.3.4 we replace condition (A2) by condition (A4). In Section 5.3.2 we explain how this condition can be relaxed.

(A4) We assume that $\mathbf{z}_{ij} = (z_{ij1}, z_{i2})'$, where z_{ij1} is a binary covariate at the observational unit level and where z_{i2} is a binary covariate at the cluster level. Assume that there are K_0 clusters with $z_{i2} = 0$. Further, we assume that each cluster i has n_0 observations with $z_{ij1} = 0$. Reorder the clusters so that the $K_0 < K$ clusters with $z_{i2} = 0$ have indices $i = 1, \ldots, K_0$ and the $K - K_0$ clusters with $z_{i2} = 1$ have indices $i = K_0 + 1, \ldots, K$.

Condition (A4) corresponds to the covariates in the udder infection data. The covariate z_{ij1} indicates the location (front or rear) of the jth udder quarter of cow i, for $j = 1, \ldots, 4$ and $i = 1, \ldots, K$. This is a binary covariate at the udder quarter level, such that $z_{ij1} = 0$ for the two front udder quarters (i.e.,

$n_0 = 2$) and $z_{ij1} = 1$ for the two rear udder quarters. The parity of cow i is a binary covariate z_{i2} at the cow level, such that $z_{i2} = 0$ if cow i is a multiparous cow and $z_{i2} = 1$ if cow i is a primiparous cow. The number of multiparous cows is K_0 and the number of primiparous cows or heifers is $K - K_0$.

5.3.2 First stage: estimation of the marginal survival functions

In the first stage of the semi-parametric estimation method, we estimate the marginal survival functions using the marginal Cox model. The marginal Cox model is given by

$$\lambda_{ij}(t) = \lambda_0(t) \exp(\boldsymbol{\beta}_0' \mathbf{z}_{ij}), \tag{5.8}$$

where $\lambda_0(t)$ is the baseline hazard at time t and $\boldsymbol{\beta}_0$ is the true p-dimensional fixed effect parameter. To estimate the fixed effect parameter $\boldsymbol{\beta}_0$ in (5.8), we ignore the cluster structure and we act as if the event times of the subjects are independent of each other. This is called the independence working assumption. Under the independence working assumption, we obtain the estimator $\hat{\boldsymbol{\beta}}$ for the true parameter $\boldsymbol{\beta}_0$ by solving the marginal score equation $\mathbf{U}_{\boldsymbol{\beta}}(\boldsymbol{\beta}) = 0$, where

$$\mathbf{U}_{\boldsymbol{\beta}}(\boldsymbol{\beta}) = \sum_{i=1}^{K} \sum_{j=1}^{n} \int_0^{\tau} \{\mathbf{z}_{ij} - \mathbf{E}(\boldsymbol{\beta}, u)\} \, dN_{ij}(u). \tag{5.9}$$

It follows from (5.4) and some easy calculations that, in the true value $\boldsymbol{\beta}_0$, we can write the score function (5.9) as (see also Spiekerman and Lin, 1998)

$$\mathbf{U}_{\boldsymbol{\beta}}(\boldsymbol{\beta}_0) = \sum_{i=1}^{K} \sum_{j=1}^{n} \int_0^{\tau} \{\mathbf{z}_{ij} - \mathbf{E}(\boldsymbol{\beta}_0, u)\} \, dM_{ij}(u). \tag{5.10}$$

The Aalen-Breslow type estimator for the cumulative baseline hazard Λ_0 is

$$\hat{\Lambda}_0(t; \hat{\boldsymbol{\beta}}) = \int_0^t \frac{dN_{\cdot\cdot}(u)}{KS^{(0)}(\hat{\boldsymbol{\beta}}, u)}.$$

The marginal survival functions are then estimated as follows:

$$\hat{S}_{ij}(t) = \exp\left\{-\hat{\Lambda}_0(t; \hat{\boldsymbol{\beta}}) \exp\left(\hat{\boldsymbol{\beta}}' \mathbf{z}_{ij}\right)\right\}. \tag{5.11}$$

In this section we give results on the consistency and the asymptotic distributional behaviour of the estimates $\hat{\boldsymbol{\beta}}$ and $\hat{\Lambda}_0(.;\hat{\boldsymbol{\beta}})$. We adapt results of Spiekerman and Lin (1998) to specific cases of deterministic covariates. As mentioned in Section 5.3.1, $M_{ij}(t)$, for $i = 1, \ldots, K$ and $j = 1, \ldots, n$, are not martingales with respect to the joint filtration $\mathcal{F}_t$. Therefore, the martingale convergence theorems cannot be applied to (5.10). However, the following lemma, presented in Spiekerman and Lin (1998), is a useful tool to prove the consistency and the asymptotic normality of $\hat{\boldsymbol{\beta}}$ and $\hat{\Lambda}_0(t, \hat{\boldsymbol{\beta}})$:

Lemma 5.3.1. *If f_K, for $K = 1, 2, \ldots$, is a sequence of random functions on $[0, \tau]$ that satisfies*

$$\int_0^\tau |df_K(u)| = O_p(1)$$

and

$$\sup_{t\in[0,\tau]} |f_K(t)| = o_p(1),$$

then, for $j = 1, \ldots, n$, $\sup_{t\in[0,\tau]} \left| K^{-1/2} \int_0^t f_K(u) dM_{.j}(u) \right| \xrightarrow{P} 0$.

The following theorems establish the consistency and the asymptotic normality of $\hat{\boldsymbol{\beta}}$.

Theorem 5.3.1. *If conditions (A1), (A2) and (A3) hold, then the estimator $\hat{\boldsymbol{\beta}}$ converges in probability to $\boldsymbol{\beta}_0$.*

Theorem 5.3.2. *If conditions (A1), (A3) and (A4) hold and if $K^{-1}K_0$ converges to p_0 for K tending to infinity, where $0 < p_0 < 1$, then $K^{1/2}\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right)$ converges weakly to a normal distribution with mean vector zero and variance-covariance matrix $\mathbf{I}_{\boldsymbol{\beta}}^{-1}\mathbf{B}\mathbf{I}_{\boldsymbol{\beta}}^{-1}$, where $\mathbf{I}_{\boldsymbol{\beta}}$ is as defined in (A3),*

$$\mathbf{B} = p_0 E\left(\mathbf{w}_{1.}^{\otimes 2}\right) + (1 - p_0) E\left(\mathbf{w}_{K_0+1,.}^{\otimes 2}\right), \tag{5.12}$$

and $\mathbf{w}_{ij} = \int_0^\tau \{\mathbf{z}_{ij} - \mathbf{e}(\boldsymbol{\beta}_0, u)\} dM_{ij}(u)$.

The proofs are adapted versions of the proofs in Spiekerman (1995) and in Spiekerman and Lin (1998). To demonstrate the methodology used in the proofs, we give as an example the proof of the following lemma, which is the

key lemma from which the asymptotic normality of $\hat{\boldsymbol{\beta}}$ follows. In the proof of Lemma 5.3.2 we show how the score function in (5.10) can be decomposed in two terms where each term is a sum of independent and identically distributed random variables. To obtain this, we assume the specific choice of deterministic covariates that is given in condition (A4). At this point the proof is different from the proof given in Spiekerman and Lin (1998) for the case of stochastic covariates, as described in the introduction of Section 5.3.

Lemma 5.3.2. *If conditions (A1) and (A4) hold and if $K^{-1}K_0$ converges to p_0 for K tending to infinity, where $0 < p_0 < 1$, then $K^{-1/2}\mathbf{U}_{\boldsymbol{\beta}}(\boldsymbol{\beta}_0)$ converges weakly to a normal distribution with mean vector zero and variance-covariance matrix $\mathbf{B}$, with $\mathbf{B}$ as defined in (5.12).*

Proof. Rewriting (5.10) gives

$$\begin{aligned}\mathbf{U}_{\boldsymbol{\beta}}(\boldsymbol{\beta}_0) &= \sum_{i=1}^{K}\sum_{j=1}^{n}\int_0^\tau \{\mathbf{z}_{ij} - \mathbf{e}(\boldsymbol{\beta}_0, u)\}\, dM_{ij}(u) \\ &\quad - \int_0^\tau \{\mathbf{E}(\boldsymbol{\beta}_0, u) - \mathbf{e}(\boldsymbol{\beta}_0, u)\}\, dM_{\cdot\cdot}(u). \qquad (5.13)\end{aligned}$$

We apply Lemma 5.3.1, with $f_K(u) = \mathbf{E}(\boldsymbol{\beta}_0, u) - \mathbf{e}(\boldsymbol{\beta}_0, u)$, to obtain

$$K^{-1/2}\int_0^\tau \{\mathbf{E}(\boldsymbol{\beta}_0, u) - \mathbf{e}(\boldsymbol{\beta}_0, u)\}\, dM_{\cdot\cdot}(u) \xrightarrow{P} 0. \qquad (5.14)$$

First, we check the required conditions of Lemma 5.3.1. By Corollary 2.1.6 in Spiekerman (1995), we have that

$$\int_0^\tau |d\{\mathbf{E}(\boldsymbol{\beta}_0, u) - \mathbf{e}(\boldsymbol{\beta}_0, u)\}| < \infty.$$

This implies that $\int_0^\tau |d\{\mathbf{E}(\boldsymbol{\beta}_0, u) - \mathbf{e}(\boldsymbol{\beta}_0, u)\}| = O_p(1)$, which is the first condition in Lemma 5.3.1. Further, for all $\epsilon > 0$, we have

$$\begin{aligned}& P\left(\sup_{u\in[0,\tau]} |\mathbf{E}(\boldsymbol{\beta}_0, u) - \mathbf{e}(\boldsymbol{\beta}_0, u)| > \epsilon\right) \\ = \; & P\left(\sup_{u\in[0,\tau]} \left|\frac{\mathbf{S}^{(1)}(\boldsymbol{\beta}_0, u)}{S^{(0)}(\boldsymbol{\beta}_0, u)} - \frac{\mathbf{s}^{(1)}(\boldsymbol{\beta}_0, u)}{s^{(0)}(\boldsymbol{\beta}_0, u)}\right| > \epsilon\right)\end{aligned}$$

$$= P\left(\sup_{u\in[0,\tau]}\left|\frac{\mathbf{S}^{(1)}(\boldsymbol{\beta}_0,u)s^{(0)}(\boldsymbol{\beta}_0,u)-\mathbf{s}^{(1)}(\boldsymbol{\beta}_0,u)S^{(0)}(\boldsymbol{\beta}_0,u)}{S^{(0)}(\boldsymbol{\beta}_0,u)s^{(0)}(\boldsymbol{\beta}_0,u)}\right|>\epsilon\right)$$

By adding and subtracting $s^{(0)}(\boldsymbol{\beta}_0,u)\mathbf{s}^{(1)}(\boldsymbol{\beta}_0,u)$ in the numerator, we obtain that the above expression equals

$$\begin{aligned}= P\Bigg(\sup_{u\in[0,\tau]}\Bigg|&\frac{\left\{\mathbf{S}^{(1)}(\boldsymbol{\beta}_0,u)-\mathbf{s}^{(1)}(\boldsymbol{\beta}_0,u)\right\}s^{(0)}(\boldsymbol{\beta}_0,u)}{S^{(0)}(\boldsymbol{\beta}_0,u)s^{(0)}(\boldsymbol{\beta}_0,u)}\\ &-\frac{\left\{S^{(0)}(\boldsymbol{\beta}_0,u)-s^{(0)}(\boldsymbol{\beta}_0,u)\right\}\mathbf{s}^{(1)}(\boldsymbol{\beta}_0,u)}{S^{(0)}(\boldsymbol{\beta}_0,u)s^{(0)}(\boldsymbol{\beta}_0,u)}\Bigg|>\epsilon\Bigg),\end{aligned}$$

which is smaller than

$$\begin{aligned}&P\left(\sup_{u\in[0,\tau]}\left|\frac{\left\{\mathbf{S}^{(1)}(\boldsymbol{\beta}_0,u)-\mathbf{s}^{(1)}(\boldsymbol{\beta}_0,u)\right\}s^{(0)}(\boldsymbol{\beta}_0,u)}{S^{(0)}(\boldsymbol{\beta}_0,u)s^{(0)}(\boldsymbol{\beta}_0,u)}\right|>\frac{\epsilon}{2}\right)\\ +&P\left(\sup_{u\in[0,\tau]}\left|\frac{\left\{S^{(0)}(\boldsymbol{\beta}_0,u)-s^{(0)}(\boldsymbol{\beta}_0,u)\right\}\mathbf{s}^{(1)}(\boldsymbol{\beta}_0,u)}{S^{(0)}(\boldsymbol{\beta}_0,u)s^{(0)}(\boldsymbol{\beta}_0,u)}\right|>\frac{\epsilon}{2}\right).\end{aligned}\tag{5.15}$$

Let δ be a constant such that $0<\delta<L_0$, where L_0 is as described in condition (C3).
It follows that (5.15) is smaller than

$$\begin{aligned}&P\left(\sup_{u\in[0,\tau]}\left|\frac{\left\{\mathbf{S}^{(1)}(\boldsymbol{\beta}_0,u)-\mathbf{s}^{(1)}(\boldsymbol{\beta}_0,u)\right\}s^{(0)}(\boldsymbol{\beta}_0,u)}{S^{(0)}(\boldsymbol{\beta}_0,u)s^{(0)}(\boldsymbol{\beta}_0,u)}\right|>\frac{\epsilon}{2},\right.\\ &\left.\sup_{u\in[0,\tau]}\left|S^{(0)}(\boldsymbol{\beta}_0,u)-s^{(0)}(\boldsymbol{\beta}_0,u)\right|\le\delta\right)\\ &+P\left(\sup_{u\in[0,\tau]}\left|\frac{\left\{S^{(0)}(\boldsymbol{\beta}_0,u)-s^{(0)}(\boldsymbol{\beta}_0,u)\right\}\mathbf{s}^{(1)}(\boldsymbol{\beta}_0,u)}{S^{(0)}(\boldsymbol{\beta}_0,u)s^{(0)}(\boldsymbol{\beta}_0,u)}\right|>\frac{\epsilon}{2},\right.\\ &\left.\sup_{u\in[0,\tau]}\left|S^{(0)}(\boldsymbol{\beta}_0,u)-s^{(0)}(\boldsymbol{\beta}_0,u)\right|\le\delta\right)\\ &+2P\left(\sup_{u\in[0,\tau]}\left|S^{(0)}(\boldsymbol{\beta}_0,u)-s^{(0)}(\boldsymbol{\beta}_0,u)\right|>\delta\right),\end{aligned}$$

which implies that (5.15) is smaller than

$$
\begin{aligned}
&P\left(\sup_{u\in[0,\tau]}\left|\frac{\left\{\mathbf{S}^{(1)}(\boldsymbol{\beta}_0,u)-\mathbf{s}^{(1)}(\boldsymbol{\beta}_0,u)\right\}s^{(0)}(\boldsymbol{\beta}_0,u)}{\left\{s^{(0)}(\boldsymbol{\beta}_0,u)-\delta\right\}s^{(0)}(\boldsymbol{\beta}_0,u)}\right|>\frac{\epsilon}{2}\right)\\
&+P\left(\sup_{u\in[0,\tau]}\left|\frac{\left\{S^{(0)}(\boldsymbol{\beta}_0,u)-s^{(0)}(\boldsymbol{\beta}_0,u)\right\}\mathbf{s}^{(1)}(\boldsymbol{\beta}_0,u)}{\left\{s^{(0)}(\boldsymbol{\beta}_0,u)-\delta\right\}s^{(0)}(\boldsymbol{\beta}_0,u)}\right|>\frac{\epsilon}{2}\right)\\
&+2P\left(\sup_{u\in[0,\tau]}\left|S^{(0)}(\boldsymbol{\beta}_0,u)-s^{(0)}(\boldsymbol{\beta}_0,u)\right|>\delta\right).
\end{aligned}
$$

By condition (C3), we have that $0 < L_0 \leq s^{(0)}(\boldsymbol{\beta}, u) \leq U_0$, for all $\boldsymbol{\beta} \in \mathcal{B}$ and for all $u \in [0, \tau]$, and $\mathbf{s}^{(1)}(\boldsymbol{\beta}, u) \leq U_1$, for all $\boldsymbol{\beta} \in \mathcal{B}$ and for all $u \in [0, \tau]$. This implies that

$$
\begin{aligned}
&P\left(\sup_{u\in[0,\tau]}|\mathbf{E}(\boldsymbol{\beta}_0,u)-\mathbf{e}(\boldsymbol{\beta}_0,u)|>\epsilon\right)\\
\leq\ &P\left(\sup_{u\in[0,\tau]}\left|\mathbf{S}^{(1)}(\boldsymbol{\beta}_0,u)-\mathbf{s}^{(1)}(\boldsymbol{\beta}_0,u)\right|>\frac{\epsilon}{2}\frac{(L_0-\delta)L_0}{U_0}\right)\\
&+P\left(\sup_{u\in[0,\tau]}\left|S^{(0)}(\boldsymbol{\beta}_0,u)-s^{(0)}(\boldsymbol{\beta}_0,u)\right|>\frac{\epsilon}{2}\frac{(L_0-\delta)L_0}{U_1}\right)\\
&+2P\left(\sup_{u\in[0,\tau]}\left|S^{(0)}(\boldsymbol{\beta}_0,u)-s^{(0)}(\boldsymbol{\beta}_0,u)\right|>\delta\right).
\end{aligned}
$$

It follows from condition (C2) that the three terms converge to zero, for K tending to infinity. We can conclude that the required conditions of Lemma 5.3.1 are satisfied. It then follows from (5.13) and (5.14) that

$$
\begin{aligned}
\mathbf{U}_{\boldsymbol{\beta}}(\boldsymbol{\beta}_0) &= \sum_{i=1}^{K}\sum_{j=1}^{n}\int_0^\tau\left\{\mathbf{z}_{ij}-\mathbf{e}(\boldsymbol{\beta}_0,u)\right\}dM_{ij}(u)+o_p(K^{1/2})\\
&= \sum_{i=1}^{K}\mathbf{w}_{i.}+o_p(K^{1/2}).
\end{aligned}
$$

By the choice of the covariates we have

$$\begin{aligned}
&\sum_{i=1}^{K} \mathbf{w}_{i.} \\
= &\sum_{i=1}^{K_0} \mathbf{w}_{i.} + \sum_{i=K_0+1}^{K} \mathbf{w}_{i.} \\
= &\sum_{i=1}^{K_0} \left(\sum_{j:z_{ij1}=0} \mathbf{w}_{ij} + \sum_{j:z_{ij1}=1} \mathbf{w}_{ij} \right) + \sum_{i=K_0+1}^{K} \left(\sum_{j:z_{ij1}=0} \mathbf{w}_{ij} + \sum_{j:z_{ij1}=1} \mathbf{w}_{ij} \right).
\end{aligned}$$

Note that observations that belong to different clusters are independent. Further, we assume in (A4) that the number of observational units in cluster i for which $z_{ij1} = 0$, resp. $z_{ij1} = 1$, is a fixed number n_0, resp. $n - n_0$, for $i = 1, \ldots, K$. Hence, it follows that $\sum_{i=1}^{K_0} \mathbf{w}_{i.}$, resp. $\sum_{i=K_0+1}^{K} \mathbf{w}_{i.}$, is a sum of independent and identically distributed random vectors.

Note that $e^{\boldsymbol{\beta}_0' \mathbf{z}_{ij}} \int_0^\tau Y_{ij}(t)\lambda_0(t)dt$ is bounded by conditions (A2) and (C1). Application of Theorem 5.6.2 in Appendix 5.6 then gives that M_{ij} is a local square integrable martingale with respect to the marginal filtration $\mathcal{F}_{t,ij}$. Further, by (5.4), M_{ij} is of finite variation since it is the difference of two increasing processes on $[0, \tau]$. We also have, by conditions (A2) and (C3), that $\mathbf{z}_{ij} - \mathbf{e}(\boldsymbol{\beta}_0, u)$ is a locally bounded function. It then follows from Theorem 5.6.3 in Appendix 5.6 that $\mathbf{w}_{ij} = \int_0^\tau \{\mathbf{z}_{ij} - \mathbf{e}(\boldsymbol{\beta}_0, u)\} \, dM_{ij}(u)$ is a local square integrable martingale with respect to the marginal filtration $\mathcal{F}_{t,ij}$. This implies that $\mathbf{w}_{ij}$ has mean vector zero for $i = 1, \ldots, K$, $j = 1, \ldots, n$ and hence, $\mathbf{w}_{i.}$ has mean vector zero.

Using the multivariate central limit theorem, we obtain that $K_0^{-1/2} \sum_{i=1}^{K_0} \mathbf{w}_{i.}$ converges to a normal distribution with mean vector zero and variance-covariance matrix $E\left(\mathbf{w}_{1.}^{\otimes 2}\right)$ and that $(K - K_0)^{-1/2} \sum_{i=K_0+1}^{K} \mathbf{w}_{i.}$ converges to a normal distribution with mean vector zero and variance-covariance matrix $E\left(\mathbf{w}_{K_0+1,.}^{\otimes 2}\right)$. Assume that $K^{-1}K_0$ converges to p_0 for K tending to infinity, where p_0 is a constant such that $0 < p_0 < 1$. Since observations that belong to different clusters are independent, we have that $K^{-1/2}\mathbf{U}(\boldsymbol{\beta}_0)$ converges to a normal distribution with mean vector zero and variance-covariance matrix

$$\mathbf{B} = p_0 E\left(\mathbf{w}_{1.}^{\otimes 2}\right) + (1 - p_0) E\left(\mathbf{w}_{K_0+1,.}^{\otimes 2}\right).$$

Straightforward calculations show that the components of $\mathbf{B}$ are finite since $\mathbf{w}_{ij}$ is a local square integrable martingale. □

We can easily generalize the proof of the previous lemma to a categorical covariate at the cluster level and a categorical covariate at the observational unit level, both with a finite number of categories. For covariates at the observational unit level, we then have to assume that the number of observational units within covariate level subgroups is the same for all clusters.

Theorems 5.3.3 and 5.3.4 establish the uniform consistency and the weak convergence of the cumulative baseline hazard estimator $\hat{\Lambda}_0(.;\hat{\boldsymbol{\beta}})$. The proofs are adaptations of the proofs in Spiekerman and Lin (1998) and are not included.

Theorem 5.3.3. *If conditions (A1), (A2) and (A3) hold, the estimator* $\hat{\Lambda}_0(t;\hat{\boldsymbol{\beta}})$ *converges in probability to* $\Lambda_0(t)$ *uniformly in* $t \in [0,\tau]$.

Define $W_K(t) = K^{1/2}\left\{\hat{\Lambda}_0(t;\hat{\boldsymbol{\beta}}) - \Lambda_0(t)\right\}$. Let $W(t)$ be a zero-mean Gaussian process with the covariance function between $W(t)$ and $W(s)$ being

$$p_0 E\left\{\Psi_1(s)\Psi_1(t)\right\} + (1-p_0)E\left\{\Psi_{K_0+1}(s)\Psi_{K_0+1}(t)\right\},$$

where

$$\Psi_i(t) = \int_0^t \frac{dM_{i.}(u)}{s^{(0)}(\boldsymbol{\beta}_0,u)} + \mathbf{h}(t)'\mathbf{I}_{\boldsymbol{\beta}}^{-1}\mathbf{w}_{i.}, \tag{5.16}$$

for $i = 1$, resp. $i = K_0+1$, $\mathbf{h}(t) = -\int_0^t \mathbf{e}(\boldsymbol{\beta}_0,u)\lambda_0(u)du$ and where p_0 is a constant such that $0 < p_0 < 1$.
Let $\mathcal{D}[0,\tau]$ be the space consisting of functions $f : [0,\tau] \to \mathbb{R}$, so that f is right continuous with left hand limits (cadlag). We make $\mathcal{D}[0,\tau]$ a metric space by equipping it with the supremum metric, i.e., the distance between two real-valued functions f, g in $\mathcal{D}[0,\tau]$ is the supremum distance $\|f-g\|_{[0,\tau]} = \sup_{t\in[0,\tau]}|f(t)-g(t)|$.

Theorem 5.3.4. *If conditions (A1), (A3) and (A4) hold and if* $K^{-1}K_0$ *converges to* p_0 *for* K *tending to infinity, the random process* W_K *converges weakly to* W *in* $\mathcal{D}[0,\tau]$.

Since $\hat{\boldsymbol{\beta}}$ and $\hat{\Lambda}_0(.;\hat{\boldsymbol{\beta}})$ are consistent estimators, we can use the estimated marginal survival functions (5.11) in the second stage of the estimation procedure. The asymptotic normality of $\hat{\boldsymbol{\beta}}$ and $\hat{\Lambda}_0(.;\hat{\boldsymbol{\beta}})$ is necessary to prove the asymptotic normality of the estimator for the copula parameter $\hat{\boldsymbol{\zeta}}$ in Section 5.3.3.

5.3.3 Second stage: estimation of the association parameter

Using the estimates $\hat{\boldsymbol{\beta}}$ and $\hat{\Lambda}_0(.;\hat{\boldsymbol{\beta}})$ we have that the pseudo loglikelihood is given by

$$\log L_P(\boldsymbol{\zeta}) = \sum_{i=1}^{K} l\left\{\boldsymbol{\zeta}, \hat{\boldsymbol{\beta}}, \hat{\Lambda}_0(X_{i1};\hat{\boldsymbol{\beta}}), \hat{\Lambda}_0(X_{i2};\hat{\boldsymbol{\beta}}), \hat{\Lambda}_0(X_{i3};\hat{\boldsymbol{\beta}}), \hat{\Lambda}_0(X_{i4};\hat{\boldsymbol{\beta}})\right\}, \tag{5.17}$$

where $l\left\{\boldsymbol{\zeta}, \hat{\boldsymbol{\beta}}, \hat{\Lambda}_0(X_{i1};\hat{\boldsymbol{\beta}}), \hat{\Lambda}_0(X_{i2};\hat{\boldsymbol{\beta}}), \hat{\Lambda}_0(X_{i3};\hat{\boldsymbol{\beta}}), \hat{\Lambda}_0(X_{i4};\hat{\boldsymbol{\beta}})\right\}$ is the contribution to the pseudo loglikelihood for cluster i. It can be seen from (5.2) that

$$\begin{aligned}
& l\left\{\boldsymbol{\zeta}, \hat{\boldsymbol{\beta}}, \hat{\Lambda}_0(X_{i1};\hat{\boldsymbol{\beta}}), \hat{\Lambda}_0(X_{i2};\hat{\boldsymbol{\beta}}), \hat{\Lambda}_0(X_{i3};\hat{\boldsymbol{\beta}}), \hat{\Lambda}_0(X_{i4};\hat{\boldsymbol{\beta}})\right\} \\
= & \left[\Delta_i \log\left\{C_{\boldsymbol{\zeta}}(v_{i1}, v_{i2}, v_{i3}, v_{i4})\right\}\right. \\
& + \sum_{j=1}^{4}\left[\Delta_i(j)\log\left\{\frac{\partial C_{\boldsymbol{\zeta}}(v_{i1}, v_{i2}, v_{i3}, v_{i4})}{\partial v_{ij}}\right\}\right] \\
& + \sum_{j\neq k}\left[\Delta_i(j,k)\log\left\{\frac{\partial^2 C_{\boldsymbol{\zeta}}(v_{i1}, v_{i2}, v_{i3}, v_{i4})}{\partial v_{ij}\,\partial v_{ik}}\right\}\right] \\
& + \sum_{j\neq k\neq l}\left[\Delta_i(j,k,l)\log\left\{\frac{\partial^3 C_{\boldsymbol{\zeta}}(v_{i1}, v_{i2}, v_{i3}, v_{i4})}{\partial v_{ij}\,\partial v_{ik}\,\partial v_{il}}\right\}\right] \\
& \left. + \Delta_i(1,2,3,4)\log\left\{\frac{\partial^4 C_{\boldsymbol{\zeta}}(v_{i1}, v_{i2}, v_{i3}, v_{i4})}{\partial v_{i1}\,\partial v_{i2}\,\partial v_{i3}\,\partial v_{i4}}\right\}\right]_{v_{ij}=\hat{S}_{ij}(X_{ij}) \text{ for } j=1,\ldots,4},
\end{aligned}$$

where $\hat{S}_{ij}(X_{ij}) = \exp\left\{-\hat{\Lambda}_0(X_{ij};\hat{\boldsymbol{\beta}})\exp\left(\hat{\boldsymbol{\beta}}'\mathbf{z}_{ij}\right)\right\}$, for $j = 1,\ldots,4$. We now estimate, in the second step of our estimation procedure, the association parameter $\boldsymbol{\zeta}$ by maximising $\log L_P(\boldsymbol{\zeta})$ in the previous expression with respect to

$\boldsymbol{\zeta}$.

Let $\boldsymbol{\zeta}$ be a q-dimensional vector. Define

$$\begin{aligned}
\mathbf{W}_{\boldsymbol{\zeta}}\{\boldsymbol{\zeta}, \mathbf{b}, c_1, c_2, c_3, c_4\} &= \frac{\partial}{\partial \boldsymbol{\zeta}} l(\boldsymbol{\zeta}, \mathbf{b}, c_1, c_2, c_3, c_4) \\
\mathbf{V}_{\boldsymbol{\zeta},j}(\boldsymbol{\zeta}, \mathbf{b}, c_1, c_2, c_3, c_4) &= \frac{\partial^2}{\partial \boldsymbol{\zeta} \partial c_j} l(\boldsymbol{\zeta}, \mathbf{b}, c_1, c_2, c_3, c_4), \quad j = 1, \ldots, 4 \\
\mathbf{V}_{\boldsymbol{\zeta}}\{\boldsymbol{\zeta}, \mathbf{b}, c_1, c_2, c_3, c_4\} &= \frac{\partial^2}{\partial \boldsymbol{\zeta} \partial \boldsymbol{\zeta}'} l(\boldsymbol{\zeta}, \mathbf{b}, c_1, c_2, c_3, c_4).
\end{aligned}$$

Note that $\mathbf{W}_{\boldsymbol{\zeta}}\{\boldsymbol{\zeta}, \mathbf{b}, c_1, c_2, c_3, c_4\}$ and $\mathbf{V}_{\boldsymbol{\zeta},j}(\boldsymbol{\zeta}, \mathbf{b}, c_1, c_2, c_3, c_4)$ are q-dimensional column vectors, whereas $\mathbf{V}_{\boldsymbol{\zeta}}\{\boldsymbol{\zeta}, \mathbf{b}, c_1, c_2, c_3, c_4\}$ is a $q \times q$-matrix. We further define

$$\begin{aligned}
\mathbf{I}_{1,\boldsymbol{\zeta}} &= p_0 \, \mathrm{E}\Big[-\mathbf{V}_{\boldsymbol{\zeta}}\{\boldsymbol{\zeta}_0, \boldsymbol{\beta}_0, \Lambda_0(X_{11}), \Lambda_0(X_{12}), \Lambda_0(X_{13}), \Lambda_0(X_{14})\}\Big] \\
&\quad + (1 - p_0) \, \mathrm{E}\big[- \mathbf{V}_{\boldsymbol{\zeta}}\{\boldsymbol{\zeta}_0, \boldsymbol{\beta}_0, \Lambda_0(X_{K_0+1,1}), \Lambda_0(X_{K_0+1,2}), \\
&\qquad \Lambda_0(X_{K_0+1,3}), \Lambda_0(X_{K_0+1,4})\}\big]
\end{aligned}$$

and

$$\begin{aligned}
\mathbf{I}_{\boldsymbol{\zeta}\boldsymbol{\beta}} &= \int_0^\tau \int_0^\tau \int_0^\tau \int_0^\tau \frac{\partial}{\partial \boldsymbol{\beta}'} \mathbf{W}_{\boldsymbol{\zeta}}\{\boldsymbol{\zeta}_0, \boldsymbol{\beta}_0, \Lambda_0(t_1), \Lambda_0(t_2), \Lambda_0(t_3), \Lambda_0(t_4)\} \\
&\qquad \times H_{\boldsymbol{\zeta}_0}(t_1, t_2, t_3, t_4),
\end{aligned}$$

where $H_{\boldsymbol{\zeta}_0}$ is the joint distribution function of $(X_{i1}, X_{i2}, X_{i3}, X_{i4})$, for $i = 1, \ldots, K$.

Note that $\mathbf{I}_{1,\boldsymbol{\zeta}}$ is a $q \times q$-matrix and $\mathbf{I}_{\boldsymbol{\zeta}\boldsymbol{\beta}}$ is a $q \times p$-matrix. We also have that $\mathbf{I}_{\boldsymbol{\beta}}$, as defined in condition (A3), is a $p \times p$-matrix.

To prove the consistency and the asymptotic normality of $\hat{\boldsymbol{\zeta}}$, we need the following assumption:

(A5) $\mathbf{W}_{\boldsymbol{\zeta}}\{\boldsymbol{\zeta}, \boldsymbol{\beta}, \Lambda_0(t_1), \Lambda_0(t_2), \Lambda_0(t_3), \Lambda_0(t_4)\}$,
$\mathbf{V}_{\boldsymbol{\zeta}}\{\boldsymbol{\zeta}, \boldsymbol{\beta}, \Lambda_0(t_1), \Lambda_0(t_2), \Lambda_0(t_3), \Lambda_0(t_4)\}$,
$\mathbf{V}_{\boldsymbol{\zeta},j}\{\boldsymbol{\zeta}, \boldsymbol{\beta}, \Lambda_0(t_1), \Lambda_0(t_2), \Lambda_0(t_3), \Lambda_0(t_4)\}$, for $j = 1, \ldots, 4$,
are continuous on $\mathcal{B}_{\boldsymbol{\zeta}} \times \mathcal{B}_{\boldsymbol{\beta}} \times [0, B_L + \delta_0] \times [0, B_L + \delta_0] \times [0, B_L + \delta_0] \times [0, B_L + \delta_0]$, where $\mathcal{B}_{\boldsymbol{\zeta}}$ is a compact neighbourhood of $\boldsymbol{\zeta}_0$ that contains $\hat{\boldsymbol{\zeta}}$, $\mathcal{B}_{\boldsymbol{\beta}}$ is a compact neighbourhood of $\boldsymbol{\beta}_0$ that contains $\hat{\boldsymbol{\beta}}$, where B_L is the

upper bound of $\Lambda_0(\tau)$, as defined in (C1), and where $\delta_0 > 0$ is a fixed constant that is used in the proof of Theorem 5.3.6.

The following theorems establish the consistency and the asymptotic normality of $\hat{\boldsymbol{\zeta}}$.

Theorem 5.3.5. *If the assumptions (A1), (A3), (A4) and (A5) hold, then the estimator $\hat{\boldsymbol{\zeta}}$ converges in probability to $\boldsymbol{\zeta}_0$.*

Theorem 5.3.6. *If the assumptions (A1), (A3), (A4) and (A5) hold and if $K^{-1}K_0$ converges to p_0 for K tending to infinity, then $K^{1/2}(\hat{\boldsymbol{\zeta}} - \boldsymbol{\zeta}_0)$ converges to a normal distribution with mean vector zero and variance-covariance*

$$\boldsymbol{\Sigma}_1 = \mathbf{I}_{1,\boldsymbol{\zeta}}^{-1} + \mathbf{I}_{1,\boldsymbol{\zeta}}^{-1}\mathbf{V}\left(\boldsymbol{\Phi}_1\right)\mathbf{I}_{1,\boldsymbol{\zeta}}^{-1},$$

where

$$\begin{aligned}
\mathbf{V}(\boldsymbol{\Phi}_1) &= p_0 E(\boldsymbol{\Phi}_{1,1}\boldsymbol{\Phi}_{1,1}') + (1-p_0)E(\boldsymbol{\Phi}_{1,K_0+1}\boldsymbol{\Phi}_{1,K_0+1}') \\
\boldsymbol{\Phi}_{1,i} &= \mathbf{I}_{\boldsymbol{\zeta}\boldsymbol{\beta}}\mathbf{I}_{\boldsymbol{\beta}}^{-1}\mathbf{w}_{i\cdot} + \int_0^\tau \mathbf{IC}_1(t_1)d\Psi_i(t_1) + \int_0^\tau \mathbf{IC}_2(t_2)d\Psi_i(t_2) \\
&\quad + \int_0^\tau \mathbf{IC}_3(t_3)d\Psi_i(t_3) + \int_0^\tau \mathbf{IC}_4(t_4)d\Psi_i(t_4),
\end{aligned}$$

for $i = 1$, resp. $i = K_0 + 1$, with Ψ_i as defined in (5.16) of Section 5.3.2, and

$$\begin{aligned}
\mathbf{IC}_1(t_1) &= \int_{t_1}^\tau \int_0^\tau \int_0^\tau \int_0^\tau \mathbf{V}_{\boldsymbol{\zeta},1}\{\boldsymbol{\zeta}_0, \boldsymbol{\beta}_0, \Lambda_0(u), \Lambda_0(t_2), \Lambda_0(t_3), \Lambda_0(t_4)\} \\
&\qquad \times dH_{\boldsymbol{\zeta}_0}(u, t_2, t_3, t_4).
\end{aligned}$$

Similar expressions hold for $\mathbf{IC}_2(t_2), \mathbf{IC}_3(t_3), \mathbf{IC}_4(t_4)$.

The proof of Theorem 5.3.5 is an adapted version of the proof of Theorem 1 in Glidden (2000), who proves a similar result for the estimator of the dependence parameter in a Clayton copula for the case where $(\mathbf{T}_i, \mathbf{C}_i, \mathbf{Z}_i)$ are i.i.d. The proof builds on the consistency of $\hat{\boldsymbol{\beta}}$ and on the uniform consistency of $\hat{\Lambda}_0(.; \hat{\boldsymbol{\beta}})$ on $[0, \tau]$. For our specific case of deterministic covariates, described in condition (C4), the proof of Glidden (2000) can be adapted using similar ideas as demonstrated in the proof of Lemma 5.3.2. Indeed, the loglikelihood

function, given in (5.2), can be decomposed in two terms where each term is a sum of independent and identically distributed random variables. The proof then follows along the lines of the proof presented in Glidden (2000).
The proof of Theorem 5.3.6 builds on the asymptotic results in Section 5.3.2 and can be given along the lines of the proof of Proposition 3.2 in Andersen (2005), who proves a similar result for bivariate failure time data for the situation where $(\mathbf{T}_i, \mathbf{C}_i, \mathbf{Z}_i)$ are i.i.d. In Section 5.4.3 we give a version of Theorem 5.3.6 following a nonparametric estimation approach. There we prove the asymptotic normality of $\hat{\boldsymbol{\zeta}}$ in a more rigorous way. Since the rigorous proof of Theorem 5.3.6 can be given by using similar ideas, the proof is not presented here.

5.4 Nonparametric approach

In the nonparametric approach, we assume that we observe a deterministic binary covariate z_{ij} at the observational unit level. In the first step, we estimate the marginal survival functions using

$$\hat{S}_{ij}(t) = \begin{cases} \exp\{-\hat{\Lambda}_1(t)\} & \text{if } z_{ij} = 0 \\ \exp\{-\hat{\Lambda}_2(t)\} & \text{if } z_{ij} = 1, \end{cases} \tag{5.18}$$

where $\hat{\Lambda}_1$, resp. $\hat{\Lambda}_2$, is the Nelson-Aalen estimator for the cumulative hazard function for the group with $z_{ij} = 0$, resp. the group with $z_{ij} = 1$. In the analysis of the udder infection data this corresponds to accounting for the effect of the location (front or rear) of the udder quarters in the estimation of the marginal survival functions without assuming a specific marginal model. In the second step, we replace the marginal survival functions in the loglikelihood expression (5.2) by their estimates (5.18) obtained in the first step and we estimate the copula parameter $\boldsymbol{\zeta}$ by solving the score equation (5.3). In Section 5.4.1 we give the definitions and assumptions that we need to develop the asymptotical results in Sections 5.4.2 and 5.4.3. In Section 5.4.2 we discuss the estimation of the marginal survival functions in the first stage. Note

that the observational units within the two covariate level subgroups (e.g., front and rear udder quarters) are correlated. We however prove that the estimators for the cumulative hazard functions $\hat{\Lambda}_1$ and $\hat{\Lambda}_2$ are consistent and asymptotically normal. Based on these results, we show in Section 5.4.3 that $\hat{\boldsymbol{\zeta}}$, the estimator for the copula parameter obtained in the second stage of the estimation, is consistent and asymptotically normal.

5.4.1 Definitions and assumptions

We first introduce the notation that will be used in this section. Recall from Section 5.3.1 that $Y_{ij}(t) = I\left(X_{ij} \geq t\right)$ denotes the at risk process and $N_{ij}(t) = \delta_{ij} I\left(X_{ij} \leq t\right)$ is the counting process for the jth observation in cluster i, for $j = 1, \ldots, n$ and $i = 1, \ldots, K$. Define

$$\begin{aligned} Y_{1,i.}(t) &= \sum_{j:z_{ij}=0} Y_{ij}(t) \\ Y_{2,i.}(t) &= \sum_{j:z_{ij}=1} Y_{ij}(t), \end{aligned}$$

and

$$\begin{aligned} N_{1,i.}(t) &= \sum_{j:z_{ij}=0} N_{ij}(t) \\ N_{2,i.}(t) &= \sum_{j:z_{ij}=1} N_{ij}(t). \end{aligned}$$

Reorder the observations so that the $n_0 < n$ observations with $z_{ij} = 0$ have indices $j = 1, \ldots, n_0$ and the $n - n_0$ observations with $z_{ij} = 1$ have indices $j = n_0 + 1, \ldots, n$. Then

$$\begin{aligned} Y_{1,i.}(t) &= \sum_{j=1}^{n_0} Y_{ij}(t) \\ Y_{2,i.}(t) &= \sum_{j=n_0+1}^{n} Y_{ij}(t), \end{aligned}$$

and

$$\begin{aligned} N_{1,i.}(t) &= \sum_{j=1}^{n_0} N_{ij}(t) \\ N_{2,i.}(t) &= \sum_{j=n_0+1}^{n} N_{ij}(t). \end{aligned}$$

We further define

$$\begin{aligned} Y_{1,..}(t) &= \sum_{i=1}^{K}\sum_{j=1}^{n_0} Y_{ij}(t) \\ Y_{2,..}(t) &= \sum_{i=1}^{K}\sum_{j=n_0+1}^{n} Y_{ij}(t), \end{aligned}$$

and

$$\begin{aligned} N_{1,..}(t) &= \sum_{i=1}^{K}\sum_{j=1}^{n_0} N_{ij}(t) \\ N_{2,..}(t) &= \sum_{i=1}^{K}\sum_{j=n_0+1}^{n} N_{ij}(t). \end{aligned}$$

We assume that $\{N_{ij}(t), Y_{ij}(t)\}$, for $i = 1, \ldots, K$ and $j = 1, \ldots, n_0$, are observed in some time interval $[0, \tau_1]$, where $\tau_1 < \infty$. For $i = 1, \ldots, K$ and $j = n_0 + 1, \ldots, n$, we assume that $\{N_{ij}(t), Y_{ij}(t)\}$ are observed in some time interval $[0, \tau_2]$, where $\tau_2 < \infty$. Further, define

$$M_{1,ij}(t) = N_{ij}(t) - \int_0^t Y_{ij}(u)d\Lambda_1(u) \qquad \text{for } j = 1, \ldots, n_0 \tag{5.19}$$

$$M_{2,ij}(t) = N_{ij}(t) - \int_0^t Y_{ij}(u)d\Lambda_2(u) \qquad \text{for } j = n_0 + 1, \ldots, n, \tag{5.20}$$

with Λ_1 and Λ_2 the cumulative hazard for the group with $z_{ij} = 0$, resp. $z_{ij} = 1$. Define

$$\begin{aligned} \Lambda_1(t) &= \int_0^t \lambda_1(u)du \\ \Lambda_2(t) &= \int_0^t \lambda_2(u)du, \end{aligned}$$

where λ_1, resp. λ_2, is the hazard function for the group with $z_{ij} = 0$, resp. $z_{ij} = 1$. We further denote, for $r = 1, 2$,

$$M_{r,.j}(t) = \sum_{i=1}^{K} M_{r,ij}(t),$$

and

$$M_{1,..}(t) = \sum_{i=1}^{K} \sum_{j=1}^{n_0} M_{1,ij}(t)$$

$$M_{2,..}(t) = \sum_{i=1}^{K} \sum_{j=n_0+1}^{n} M_{2,ij}(t).$$

To prove the consistency and the asymptotic normality of $\hat{\Lambda}_1$ and $\hat{\Lambda}_2$ in Section 5.4.2, we assume the following condition:

(A6) $P\{Y_{ij}(t) = 1, \text{for all } t \in [0, \tau_1]\} > 0$, for $i = 1, \ldots, K$, $j = 1, \ldots, n_0$,
$P\{Y_{ij}(t) = 1, \text{for all } t \in [0, \tau_2]\} > 0$, for $i = 1, \ldots, K$, $j = n_0 + 1, \ldots, n$.

Along the lines of the proof of Corollary 5.3.1, it follows that condition (A6) entails the following conditions, which are useful in the proofs of this section:

(C4) $\int_0^{\tau_r} \lambda_r(u)du \equiv B_{L_r} < \infty$, for $r = 1, 2$.

(C5)
$$\sup_{t \in [0, \tau_r]} \left| \frac{1}{K} \sum_{i=1}^{K} Y_{r,i.}(t) - E\{Y_{r,1.}(t)\} \right| \xrightarrow{P} 0.$$

(C6) There exist constants L_r and U_r such that
$0 < L_r \leq E\{Y_{r,1.}(t)\} \leq U_r$, for all $t \in [0, \tau_r]$.

5.4.2 First stage: estimation of the marginal survival functions

In the first step of the nonparametric approach, we estimate the marginal survival functions by using a Nelson-Aalen estimator for the cumulative hazard functions Λ_1 and Λ_2. The Nelson-Aalen estimator is given by

$$\hat{\Lambda}_r(t) = \int_0^t \frac{I\{Y_{r,..}(s) > 0\}}{Y_{r,..}(s)} dN_{r,..}(s) \quad \text{for } r = 1, 2,$$

with the convention that $0/0 = 0$. We then estimate the marginal survival functions as follows:

$$\begin{aligned} \hat{S}_{ij}(t) &= \exp\left\{-\hat{\Lambda}_1(t)\right\}, \text{ for } j = 1, \ldots, n_0, \\ \hat{S}_{ij}(t) &= \exp\left\{-\hat{\Lambda}_2(t)\right\}, \text{ for } j = n_0 + 1, \ldots, n. \end{aligned} \tag{5.21}$$

In this section we study the asymptotic properties of $\hat{\Lambda}_1$ and $\hat{\Lambda}_2$. Note that the observations in the group with $z_{ij} = 0$, resp. $z_{ij} = 1$, are not independent. This implies that $M_{1,ij}$ is not a martingale with respect to the joint filtration $\mathcal{F}_{1,t} = \bigvee_{i=1}^{K} \bigvee_{j=1}^{n_0} \mathcal{F}_{t,ij}$, and that $M_{2,ij}$ is not a martingale with respect to the joint filtration $\mathcal{F}_{2,t} = \bigvee_{i=1}^{K} \bigvee_{j=n_0+1}^{n} \mathcal{F}_{t,ij}$. Therefore, we cannot use martingale convergence theorems to prove the consistency and the asymptotic normality of $\hat{\Lambda}_1$ and $\hat{\Lambda}_2$. However, the following lemma, which is the version of Lemma 5.3.1 needed in the present nonparametric setting, is a useful tool to prove the asymptotic results for $\hat{\Lambda}_1$ and $\hat{\Lambda}_2$:

Lemma 5.4.1. *a) If f_K, for $K = 1, 2, \ldots$, is a sequence of random functions on $[0, \tau_1]$ that satisfies*

$$\int_0^{\tau_1} |df_K(u)| = O_p(1)$$

and

$$\sup_{t \in [0,\tau_1]} |f_K(t)| = o_p(1),$$

then, for $j = 1, \ldots, n_0$, $\sup_{t \in [0,\tau_1]} \left| K^{-1/2} \int_0^t f_K(u) dM_{1,\cdot j}(u) \right| \xrightarrow{P} 0$.

b) If f_K, for $K = 1, 2, \ldots$, is a sequence of random functions on $[0, \tau_2]$ that satisfies

$$\int_0^{\tau_2} |df_K(u)| = O_p(1)$$

and

$$\sup_{t \in [0,\tau_2]} |f_K(t)| = o_p(1),$$

then, for $j = n_0 + 1, \ldots, n$, $\sup_{t \in [0,\tau_2]} \left| K^{-1/2} \int_0^t f_K(u) dM_{2,\cdot j}(u) \right| \xrightarrow{P} 0$.

Proof. The proof can be given using similar arguments as used in the proof of Lemma 5.3.1 and is therefore not presented here. □

The proofs of the uniform consistency and the asymptotic normality of $\hat{\Lambda}_1$ and $\hat{\Lambda}_2$ rely on the following lemmas. Lemma 5.4.2, resp. Lemma 5.4.3, states that the first condition, resp. the second condition, of Lemma 5.4.1 holds, for $f_K(u) = K^{1/2} \{Y_{1,\cdot\cdot}(u)\}^{-1} I \{Y_{1,\cdot\cdot}(u) > 0\}$.

Lemma 5.4.2. *If conditions (C5) and (C6) hold, then*

$$\int_0^{\tau_1} \left| d \left[K^{1/2} \frac{I \{Y_{1,\cdot\cdot}(u) > 0\}}{Y_{1,\cdot\cdot}(u)} \right] \right| = O_p(1), \tag{5.22}$$

and

$$\int_0^{\tau_2} \left| d \left[K^{1/2} \frac{I \{Y_{2,\cdot\cdot}(u) > 0\}}{Y_{2,\cdot\cdot}(u)} \right] \right| = O_p(1). \tag{5.23}$$

Proof. We give the proof for (5.22). The proof for (5.23) is similar.
Let $\mathcal{D}$ be the set of all partitions of $[0, \tau_1]$: $0 = t_0 < t_1 < \ldots < t_L = \tau_1$. Note that

$$\begin{aligned} & \int_0^{\tau_1} \left| d \left[K^{1/2} \frac{I \{Y_{1,\cdot\cdot}(u) > 0\}}{Y_{1,\cdot\cdot}(u)} \right] \right| \\ = & \sup_{\mathcal{D}} \left[K^{-1/2} \sum_{l=1}^{L} \left| \frac{I \{Y_{1,\cdot\cdot}(t_l) > 0\}}{K^{-1} Y_{1,\cdot\cdot}(t_l)} - \frac{I \{Y_{1,\cdot\cdot}(t_{l-1}) > 0\}}{K^{-1} Y_{1,\cdot\cdot}(t_{l-1})} \right| \right]. \end{aligned}$$

For any constant M we have

$$\begin{aligned} & P \left(\sup_{\mathcal{D}} \left[K^{-1/2} \sum_{l=1}^{L} \left| \frac{I \{Y_{1,\cdot\cdot}(t_l) > 0\}}{K^{-1} Y_{1,\cdot\cdot}(t_l)} - \frac{I \{Y_{1,\cdot\cdot}(t_{l-1}) > 0\}}{K^{-1} Y_{1,\cdot\cdot}(t_{l-1})} \right| \right] > M \right) \\ = & P \left(\sup_{\mathcal{D}} \left[K^{-1/2} \sum_{l=1}^{L} \left| \frac{K^{-1} Y_{1,\cdot\cdot}(t_{l-1}) \left[I \{Y_{1,\cdot\cdot}(t_l) > 0\} - I \{Y_{1,\cdot\cdot}(t_{l-1}) > 0\} \right]}{K^{-1} Y_{1,\cdot\cdot}(t_l) K^{-1} Y_{1,\cdot\cdot}(t_{l-1})} \right. \right. \right. \\ & \left. \left. \left. + \frac{K^{-1} I \{Y_{1,\cdot\cdot}(t_{l-1}) > 0\} \{Y_{1,\cdot\cdot}(t_{l-1}) - Y_{1,\cdot\cdot}(t_l)\}}{K^{-1} Y_{1,\cdot\cdot}(t_l) K^{-1} Y_{1,\cdot\cdot}(t_{l-1})} \right| \right] > M \right). \end{aligned} \tag{5.24}$$

Let δ be a constant such that $0 < \delta < L_1$, with L_1 as in condition (C6). An upper bound for the probability (5.24) is given by

$$P \left(\sup_{\mathcal{D}} \left[K^{-1/2} \sum_{l=1}^{L} \left| \frac{K^{-1} Y_{1,\cdot\cdot}(t_{l-1}) \left[I \{Y_{1,\cdot\cdot}(t_l) > 0\} - I \{Y_{1,\cdot\cdot}(t_{l-1}) > 0\} \right]}{K^{-1} Y_{1,\cdot\cdot}(t_l) K^{-1} Y_{1,\cdot\cdot}(t_{l-1})} \right| \right] > \frac{M}{2}, \right.$$
$$\left. \sup_{t \in [0, \tau_1]} \left| K^{-1} Y_{1,\cdot\cdot}(t) - E \{Y_{1,1.}(t)\} \right| \leq \delta \right)$$

$$+P\left(\sup_{\mathcal{D}}\left[K^{-1/2}\sum_{l=1}^{L}\left|\frac{K^{-1}I\left\{Y_{1,..}(t_{l-1})>0\right\}\left\{Y_{1,..}(t_{l-1})-Y_{1,..}(t_l)\right\}}{K^{-1}Y_{1,..}(t_l)K^{-1}Y_{1,..}(t_{l-1})}\right|\right]>\frac{M}{2},\right.$$
$$\left.\sup_{t\in[0,\tau_1]}\left|K^{-1}Y_{1,..}(t)-E\left\{Y_{1,1.}(t)\right\}\right|\leq\delta\right)$$
$$+2P\left(\sup_{t\in[0,\tau_1]}\left|K^{-1}Y_{1,..}(t)-E\left\{Y_{1,1.}(t)\right\}\right|>\delta\right). \tag{5.25}$$

If $\sup_{t\in[0,\tau_1]}\left|K^{-1}Y_{1,..}(t)-E\left\{Y_{1,1.}(t)\right\}\right|\leq\delta$, it follows from condition (C6) that $\sup_{t\in[0,\tau_1]}\left\{K^{-1}Y_{1,..}(t)\right\}\leq U_1+\delta$ and $\sup_{t\in[0,\tau_1]}\left\{K^{-1}Y_{1,..}(t)\right\}^{-1}\leq(L_1-\delta)^{-1}$. This implies that (5.25) is smaller than

$$\begin{aligned}
& P\left(\sup_{\mathcal{D}}\left[K^{-1/2}\sum_{l=1}^{L}\left|I\left\{Y_{1,..}(t_l)>0\right\}-I\left\{Y_{1,..}(t_{l-1})\right\}\right|\right]>\frac{M(L_1-\delta)^2}{2\left(U_1+\delta\right)}\right)\\
+\quad & P\left(\sup_{\mathcal{D}}\left\{K^{-3/2}\sum_{l=1}^{L}\left|Y_{1,..}(t_l)-Y_{1,..}(t_{l-1})\right|\right\}>\frac{M\left(L_1-\delta\right)^2}{2}\right)\\
+\quad & 2P\left(\sup_{t\in[0,\tau_1]}\left|K^{-1}Y_{1,..}(t)-E\left\{Y_{1,1.}(t)\right\}\right|>\delta\right).
\end{aligned} \tag{5.26}$$

Note that

$$\sum_{l=1}^{L}\left|I\left\{Y_{1,..}(t_l)>0\right\}-I\left\{Y_{1,..}(t_{l-1})>0\right\}\right|\leq 1, \tag{5.27}$$

and

$$K^{-1}\sum_{l=1}^{L}\left|Y_{1,..}(t_l)-Y_{1,..}(t_{l-1})\right|\leq K^{-1}\sum_{i=1}^{K}\sum_{j=1}^{n_0}\sum_{l=1}^{L}\left|Y_{ij}(t_l)-Y_{ij}(t_{l-1})\right|=n_0, \tag{5.28}$$

for any partition of $[0,\tau_1]$ in $\mathcal{D}$. Since the upper bounds in (5.27) and (5.28) do not depend on the specific partition, it follows that the first and the second term of (5.26) become zero, if K tends to infinity. By condition (C5), the third term converges to zero, for K tending to infinity. This completes the proof. □

Lemma 5.4.3. *If conditions (C5) and (C6) hold, then*

$$\sup_{t\in[0,\tau_1]}\left|K^{1/2}\frac{I\left\{Y_{1,..}(t)>0\right\}}{Y_{1,..}(t)}\right|=o_p(1), \tag{5.29}$$

and

$$\sup_{t\in[0,\tau_2]}\left|K^{1/2}\frac{I\left\{Y_{2,\cdot\cdot}(t)>0\right\}}{Y_{2,\cdot\cdot}(t)}\right| = o_p(1). \tag{5.30}$$

Proof. For any $\epsilon > 0$ and $\delta > 0$, we have

$$\begin{aligned}
& P\left(\sup_{t\in[0,\tau_1]}\left|K^{1/2}\frac{I\left\{Y_{1,\cdot\cdot}(t)>0\right\}}{Y_{1,\cdot\cdot}(t)}\right|>\epsilon\right)\\
\leq\ & P\left(\sup_{t\in[0,\tau_1]}\left|K^{1/2}\left\{Y_{1,\cdot\cdot}(t)\right\}^{-1}\right|>\epsilon\right)\\
\leq\ & P\left(\sup_{t\in[0,\tau_1]}\left|K^{-1/2}\left\{K^{-1}Y_{1,\cdot\cdot}(t)\right\}^{-1}\right|>\epsilon,\right.\\
& \left.\sup_{t\in[0,\tau_1]}\left|K^{-1}Y_{1,\cdot\cdot}(t)-E\left\{Y_{1,1.}(t)\right\}\right|\leq\delta\right)\\
& + P\left(\sup_{t\in[0,\tau_1]}\left|K^{-1}Y_{1,\cdot\cdot}(t)-E\left\{Y_{1,1.}(t)\right\}\right|>\delta\right).
\end{aligned}$$

It follows from conditions (C5) and (C6) that the above expression converges to zero, for K tending to infinity. This concludes the proof of (5.29). The proof of (5.30) is similar. □

Lemma 5.4.4. *If condition (A6) holds, then*

$$\sup_{t\in[0,\tau_1]} I\left\{Y_{1,\cdot\cdot}(t)=0\right\} = o_p(1), \tag{5.31}$$

and

$$\sup_{t\in[0,\tau_2]} I\left\{Y_{2,\cdot\cdot}(t)=0\right\} = o_p(1). \tag{5.32}$$

Proof. We give the proof for (5.31). The proof for (5.32) is similar. For any $\epsilon > 0$, we have

$$\begin{aligned}
& P\left(\sup_{t\in[0,\tau_1]} I\left\{Y_{1,\cdot\cdot}(t)=0\right\}>\epsilon\right)\\
=\ & P\left(\sup_{t\in[0,\tau_1]} I\left\{Y_{1,\cdot\cdot}(t)=0\right\}=1\right).
\end{aligned} \tag{5.33}$$

Since $I\{Y_{1,\cdot\cdot}(t) = 0\}$ is a non-decreasing function in t, taking values zero or one, it follows that (5.33) equals

$$\begin{aligned} & P\left(\exists t^* \in [0,\tau_1], \forall t \in [t^*,\tau_1] : I\{Y_{1,\cdot\cdot}(t) = 0\} = 1\right) \\ = \; & P\left(\exists t^* \in [0,\tau_1], \forall t \in [t^*,\tau_1] : \sum_{i=1}^{K} Y_{1,i\cdot}(t) = 0\right) \\ = \; & P\left(\exists t^* \in [0,\tau_1], \forall t \in [t^*,\tau_1] : Y_{1,i\cdot}(t) = 0, \forall i = 1,\ldots,K\right). \quad (5.34) \end{aligned}$$

Since $Y_{1,1\cdot}(t), Y_{1,2\cdot}(t), \ldots, Y_{1,K\cdot}(t)$ are independent, we have that (5.34) is equal to

$$\begin{aligned} & \prod_{i=1}^{K} P\left(\exists t^* \in [0,\tau_1], \forall t \in [t^*,\tau_1] : Y_{1,i\cdot}(t) = 0\right) \\ = \; & \prod_{i=1}^{K} P\left(\exists t^* \in [0,\tau_1], \forall t \in [t^*,\tau_1] : Y_{1,ij}(t) = 0, \forall j = 1,\ldots,n_0\right). \quad (5.35) \end{aligned}$$

Hence, an upper bound for (5.35) is given by

$$\prod_{i=1}^{K} P\left(\exists t^* \in [0,\tau_1] : Y_{1,i1}(t^*) = 0\right). \quad (5.36)$$

Using condition (A6), we obtain that there exists a constant $0 < \delta \leq 1$, such that an upper bound for (5.36) is given by $(1-\delta)^K$. The latter upper bound converges to zero, for K tending to infinity. □

In the following theorem we establish the uniform consistency of the cumulative hazard estimators $\hat{\Lambda}_1$ and $\hat{\Lambda}_2$:

Theorem 5.4.1. *If condition (A6) holds, then the estimator $\hat{\Lambda}_1(t)$ converges in probability to $\Lambda_1(t)$, uniformly in $[0,\tau_1]$. Similary, $\hat{\Lambda}_2(t)$ converges in probability to $\Lambda_2(t)$, uniformly in $[0,\tau_2]$.*

Proof. We prove the theorem for $\hat{\Lambda}_1(t)$. The proof for $\hat{\Lambda}_2(t)$ is similar.
Note that

$$\begin{aligned} \hat{\Lambda}_1(t) - \Lambda_1(t) \; = \; & \int_0^t \frac{I\{Y_{1,\cdot\cdot}(u) > 0\}}{Y_{1,\cdot\cdot}(u)} dN_{1,\cdot\cdot}(u) - \int_0^t I\{Y_{1,\cdot\cdot}(u) > 0\}\lambda_1(u)du \\ & + \int_0^t I\{Y_{1,\cdot\cdot}(u) > 0\}\lambda_1(u)du - \int_0^t \lambda_1(u)du. \quad (5.37) \end{aligned}$$

It follows from (5.19) that

$$dM_{1,..}(u) = dN_{1,..}(u) - I\left\{Y_{1,..}(u) > 0\right\} Y_{1,..}(u)\lambda_1(u)du. \tag{5.38}$$

Substituting (5.38) in (5.37) leads to

$$\hat{\Lambda}_1(t) - \Lambda_1(t) = \int_0^t \frac{I\left\{Y_{1,..}(u) > 0\right\}}{Y_{1,..}(u)} dM_{1,..}(u) - \int_0^t I\left\{Y_{1,..}(u) = 0\right\} \lambda_1(u)du. \tag{5.39}$$

It easily follows from Lemma 5.4.4 and condition (C4) that the second term of the right hand side of the above equation converges in probability to zero. Hence,

$$\hat{\Lambda}_1(t) - \Lambda_1(t) = \int_0^t \frac{I\left\{Y_{1,..}(u) > 0\right\}}{Y_{1,..}(u)} dM_{1,..}(u) + o_p(1). \tag{5.40}$$

Note that $M_{1,ij}(t)$ are not martingales with respect to the joint filtration $\mathcal{F}_t = \bigvee_{i=1}^{K} \bigvee_{j=1}^{n_0} \mathcal{F}_{t,ij}$. Therefore, the martingale convergence theorems cannot be applied to (5.40). But we can apply Lemma 5.4.1 with $f_K(u) = K^{1/2}\left\{Y_{1,..}(u)\right\}^{-1} I\left\{Y_{1,..}(u) > 0\right\}$. For this choice the conditions required for Lemma 5.4.1 to hold are satisfied (Lemma 5.4.2 and Lemma 5.4.3). We therefore have

$$\sup_{t\in[0,\tau_1]} \left| \int_0^t \frac{I\left\{Y_{1,..}(u) > 0\right\}}{Y_{1,..}(u)} dM_{1,.j}(u) \right| \xrightarrow{P} 0,$$

for $j = 1, \ldots, n_0$. Since n_0 is fixed, it follows that

$$\sup_{t\in[0,\tau_1]} \left| \int_0^t \frac{I\left\{Y_{1,..}(u) > 0\right\}}{Y_{1,..}(u)} dM_{1,..}(u) \right| \xrightarrow{P} 0.$$

□

We use the following two lemmas in the proof of the asymptotic normality of $\hat{\Lambda}_1$ and $\hat{\Lambda}_2$.

Lemma 5.4.5. *If condition (C6) holds, then*

$$\int_0^{\tau_1} \left| d\left[E\left\{Y_{1,1.}(u)\right\}\right]^{-1} \right| = O_p(1), \tag{5.41}$$

and

$$\int_0^{\tau_2} \left| d\left[E\left\{Y_{2,1.}(u)\right\}\right]^{-1} \right| = O_p(1). \tag{5.42}$$

Proof. Let $\mathcal{D}$ be the set of all partitions of $[0, \tau_1]$: $0 = t_0 < t_1 < \ldots < t_L = \tau_1$. We have that

$$\begin{aligned} & \sup_{\mathcal{D}} \sum_{l=1}^{L} \left| \frac{1}{E\left\{Y_{1,1.}(t_l)\right\}} - \frac{1}{E\left\{Y_{1,1.}(t_{l-1})\right\}} \right| \\ = & \sup_{\mathcal{D}} \sum_{l=1}^{L} \left| \frac{E\left\{Y_{1,1.}(t_{l-1})\right\} - E\left\{Y_{1,1.}(t_l)\right\}}{E\left\{Y_{1,1.}(t_l)\right\} E\left\{Y_{1,1.}(t_{l-1})\right\}} \right|. \end{aligned} \tag{5.43}$$

Condition (C6) implies that $\sup_{t\in[0,\tau_1]} \left[E\left\{Y_{1,1.}(t)\right\}\right]^{-1} \leq L_1^{-1}$. Hence, an upper bound of (5.43) is given by

$$\frac{1}{L_1^2} \sup_{\mathcal{D}} \sum_{l=1}^{L} \left| E\left\{ \sum_{j=1}^{n_0} Y_{1,1j}(t_{l-1}) \right\} - E\left\{ \sum_{j=1}^{n_0} Y_{1,1j}(t_l) \right\} \right|. \tag{5.44}$$

Since $\sum_{l=1}^{L} |Y_{1,1j}(t_{l-1}) - Y_{1,1j}(t_l)|$ equals zero or one, straightforward calculations lead to the following upper bound for (5.44):

$$\frac{1}{L_1^2} \sum_{j=1}^{n_0} \left[\sup_{\mathcal{D}} P\left\{ \sum_{l=1}^{L} |Y_{1,1j}(t_{l-1}) - Y_{1,1j}(t_l)| = 1 \right\} \right].$$

An upper bound of the latter equation is given by n_0/L_1^{-2}. This concludes the proof of (5.41). Using similar arguments, we can obtain (5.42). □

Lemma 5.4.6. *If condition (A6) holds, then*

$$\sup_{t\in[0,\tau_1]} \left| \frac{I\left\{Y_{1,..}(t) > 0\right\}}{K^{-1}Y_{1,..}(t)} - \frac{1}{E\left\{Y_{1,1.}(t)\right\}} \right| = o_p(1), \tag{5.45}$$

and

$$\sup_{t\in[0,\tau_2]} \left| \frac{I\left\{Y_{2,..}(t) > 0\right\}}{K^{-1}Y_{2,..}(t)} - \frac{1}{E\left\{Y_{2,1.}(t)\right\}} \right| = o_p(1). \tag{5.46}$$

Proof. For $\epsilon > 0$, we have that

$$\begin{aligned} & P\left(\sup_{t\in[0,\tau_1]} \left| \frac{I\left\{Y_{1,..}(t) > 0\right\}}{K^{-1}Y_{1,..}(t)} - \frac{1}{E\left\{Y_{1,1.}(t)\right\}} \right| > \epsilon \right) \\ \leq & P\left(\sup_{t\in[0,\tau_1]} \left| \frac{I\left\{Y_{1,..}(t) > 0\right\}}{K^{-1}Y_{1,..}(t)} - \frac{1}{K^{-1}Y_{1,..}(t)} \right| > \frac{\epsilon}{2} \right) \\ & + P\left(\sup_{t\in[0,\tau_1]} \left| \frac{1}{K^{-1}Y_{1,..}(t)} - \frac{1}{E\left\{Y_{1,1.}(t)\right\}} \right| > \frac{\epsilon}{2} \right). \end{aligned} \tag{5.47}$$

Using the trivial equality $I\{Y_{1,..}(t) > 0\} + I\{Y_{1,..}(t) = 0\} = 1$, we easily obtain the following upper bound for (5.47):

$$
\begin{aligned}
&P\left(\sup_{t\in[0,\tau_1]}\left|\frac{I\{Y_{1,..}(t)=0\}}{K^{-1}Y_{1,..}(t)}\right| > \frac{\epsilon}{2}, \sup_{t\in[0,\tau_1]}\left|E\{Y_{1,1.}(t)\} - K^{-1}Y_{1,..}(t)\right| \le \delta\right)\\
&+P\left(\sup_{t\in[0,\tau_1]}\left|\frac{E\{Y_{1,1.}(t)\} - K^{-1}Y_{1,..}(t)}{K^{-1}Y_{1,..}(t)E\{Y_{1,1.}(t)\}}\right| > \frac{\epsilon}{2},\right.\\
&\left.\sup_{t\in[0,\tau_1]}\left|E\{Y_{1,1.}(t)\} - K^{-1}Y_{1,..}(t)\right| \le \delta\right)\\
&+2P\left(\sup_{t\in[0,\tau_1]}\left|E\{Y_{1,1.}(t)\} - K^{-1}Y_{1,..}(t)\right| > \delta\right), \qquad (5.48)
\end{aligned}
$$

with δ a fixed constant, such that $0 < \delta < L_1$. An upper bound for (5.48) is given by:

$$
\begin{aligned}
&P\left(\sup_{t\in[0,\tau_1]}\left|\frac{I\{Y_{1,..}(t)=0\}}{E\{Y_{1,1.}(t)\} - \delta}\right| > \frac{\epsilon}{2}\right)\\
&+P\left(\sup_{t\in[0,\tau_1]}\left|\frac{E\{Y_{1,1.}(t)\} - K^{-1}Y_{1,..}(t)}{[E\{Y_{1,1.}(t)\} - \delta]\,E\{Y_{1,1.}(t)\}}\right| > \frac{\epsilon}{2}\right)\\
&+2P\left(\sup_{t\in[0,\tau_1]}\left|E\{Y_{1,1.}(t)\} - K^{-1}Y_{1,..}(t)\right| > \delta\right).
\end{aligned}
$$

By condition (C6), the above expression is smaller than

$$
\begin{aligned}
&P\left(\sup_{t\in[0,\tau_1]}|I\{Y_{1,..}(t)=0\}| > \frac{\epsilon}{2}(L_1-\delta)\right)\\
&+P\left(\sup_{t\in[0,\tau_1]}\left|E\{Y_{1,1.}(t)\} - K^{-1}Y_{1,..}(t)\right| > \frac{\epsilon}{2}L_1(L_1-\delta)\right)\\
&+2P\left(\sup_{t\in[0,\tau_1]}\left|E\{Y_{1,1.}(t)\} - K^{-1}Y_{1,..}(t)\right| > \delta\right).
\end{aligned}
$$

It follows from condition (C5) that the second and the third term converge to zero, for K tending to infinity. The first term converges to zero, for K tending to infinity, by Lemma 5.4.4. □

Define $W_{1,K}(t) = K^{1/2}\left\{\hat{\Lambda}_1(t) - \Lambda_1(t)\right\}$. Let $W_1(t)$ be a zero-mean Gaussian random process with the covariance function between $W_1(t)$ and $W_1(s)$, for $0 \leq t, s, \leq \tau_1$, being $E\left\{\Psi_{1,1}(t)\Psi_{1,1}(s)\right\}$, where

$$\Psi_{1,i}(t) = \int_0^t \frac{1}{E\left\{Y_{1,1.}(u)\right\}} dM_{1,i.}(u). \tag{5.49}$$

Similar definitions can be given for $\hat{\Lambda}_2(.)$.
Recall from Section 5.3.2 that $\mathcal{D}[0, \tau_1]$, resp. $\mathcal{D}[0, \tau_2]$, is the metric space of all right continuous functions on $[0, \tau_1]$, resp. $[0, \tau_2]$, with left hand limits, equipped with the supremum metric.

Theorem 5.4.2. *If condition (A6) holds, the random process $W_{1,K}$ converges weakly to W_1 in $\mathcal{D}[0, \tau_1]$. Similarly, the random process $W_{2,K}$ converges weakly to W_2 in $\mathcal{D}[0, \tau_2]$.*

Proof. We prove the theorem for $W_{1,K}$. The proof for $W_{2,K}$ is similar.
The proof has two parts:

(a) the weak convergence of the finite dimensional distributions;

(b) the asymptotic tightness of $W_{1,K}$.

From (a) and (b) it follows, using Theorem 5.6.6 in Appendix 5.6.3, that the process $W_{1,K}$ converges weakly to W_1 in $\mathcal{D}[0, \tau_1]$. Some key results of the weak convergence theory for stochastic processes in $\mathcal{D}[a, b]$, where $[a, b] \subset \mathbb{R}$ is an arbitrary interval, are collected in Appendix 5.6.3.
In part (a) of the proof we write $W_{1,K}$ as a sum over the cluster index i so that we obtain a sum of independent and identically distributed random variables. The weak convergence of the finite dimensional distributions then easily follows. In part (b) we write $W_{1,K}$ as a sum over the observational unit index j. In this way we obtain that each term in this sum is a martingale with respect to the filtration $\bigvee_{i=1}^{K} \mathcal{F}_{t,ij}$. The proof of the asymptotic tightness of $W_{1,K}$ then builds on martingale theory.

(a) Weak convergence of the finite dimensional distributions:

To prove the weak convergence of the finite dimensional distributions, we start from (5.39). We then have

$$W_{1,K}(t) = K^{1/2}\left[\int_0^t \frac{I\{Y_{1,.}(u) > 0\}}{Y_{1,.}(u)} dM_{1,.}(u) - \int_0^t I\{Y_{1,.}(u) = 0\}\lambda_1(u)du\right].$$

Easy calculations using Lemma 5.4.4 and condition (C4) show that the second term of the right hand side of the above equation converges in probability to zero. It follows that

$$\begin{aligned} W_{1,K}(t) &= K^{-1/2}\int_0^t \left[\frac{I\{Y_{1,.}(u) > 0\}}{K^{-1}Y_{1,.}(u)} - \frac{1}{E\{Y_{1,1.}(u)\}}\right] dM_{1,.}(u) \\ &\quad + K^{-1/2}\int_0^t \frac{1}{E\{Y_{1,1.}(u)\}} dM_{1,.}(u) + o_p(1). \end{aligned} \tag{5.50}$$

By Lemma 5.4.1, with

$$f_K(u) = \left\{K^{-1}Y_{1,.}(u)\right\}^{-1} I\{Y_{1,.}(u) > 0\} - [E\{Y_{1,1.}(u)\}]^{-1},$$

we have that

$$\sup_{t\in[0,\tau_1]}\left|K^{-1/2}\int_0^t \left[\frac{I\{Y_{1,.}(u) > 0\}}{K^{-1}Y_{1,.}(u)} - \frac{1}{E\{Y_{1,1.}(u)\}}\right] dM_{1,.j}(u)\right| = o_p(1),$$

for $j = 1, \ldots, n_0$. Since n_0 is fixed, it follows that

$$\sup_{t\in[0,\tau_1]}\left|K^{-1/2}\int_0^t \left[\frac{I\{Y_{1,.}(u) > 0\}}{K^{-1}Y_{1,.}(u)} - \frac{1}{E\{Y_{1,1.}(u)\}}\right] dM_{1,.}(u)\right| = o_p(1). \tag{5.51}$$

For the present choice of f_K, Lemma 5.4.2, Lemma 5.4.5 and Lemma 5.4.6 imply that the assumptions of Lemma 5.4.1 hold. Substituting (5.51) in (5.50) then leads to

$$\begin{aligned} W_{1,K}(t) &= K^{-1/2}\int_0^t \frac{1}{E\{Y_{1,1.}(u)\}} dM_{1,.}(u) + o_p(1) \\ &= K^{-1/2}\sum_{i=1}^K \left[\int_0^t \frac{1}{E\{Y_{1,1.}(u)\}} dM_{1,i.}(u)\right] + o_p(1) \qquad (5.52) \\ &= K^{-1/2}\sum_{i=1}^K \Psi_{1,i}(t) + o_p(1), \qquad (5.53) \end{aligned}$$

which is a sum of independent and identically distributed random variables. From (5.53) the weak convergence of the finite dimensional distributions of $W_{1,K}$ to a zero-mean Gaussian process W_1 with covariance function

$$E\left\{\Psi_{1,1}(t)\Psi_{1,1}(s)\right\}$$

easily follows.

(b) Asymptotic tightness:
To prove the asymptotic tightness of $W_{1,K}$, we rewrite (5.52) in the following way:

$$\begin{aligned} W_{1,K}(t) &= K^{-1/2}\sum_{j=1}^{n_0}\left[\int_0^t \frac{1}{E\left\{Y_{1,1.}(u)\right\}} dM_{1,.j}(u)\right] + o_p(1) \\ &= \sum_{j=1}^{n_0} Q_j(t) + o_p(1), \end{aligned}$$

where

$$Q_j(t) = \int_0^t \frac{1}{E\left\{Y_{1,1.}(u)\right\}} d\left\{K^{-1/2}M_{1,.j}(u)\right\},$$

for $j = 1, \ldots, n_0$. Note that $Q_j(t)$ is a martingale with respect to the filtration $\bigvee_{i=1}^{K}\mathcal{F}_{t,ij}$. Hence, we can use martingale theory to obtain that Q_j converges weakly. The asymptotic tightness of Q_j then follows by Theorem 5.6.6 in Appendix 5.6.3, since asymptotic tightness is necessary for weak convergence. Using Theorem 5.6.7 in Appendix 5.6.3, with $X_K = W_{1,K}$, it is easy to show that the asymptotic tightness of $W_{1,K}$ follows from the asymptotic tightness of Q_j, for $j = 1, \ldots, n_0$, because $D[0, \tau_1]$ has been defined using the supremum metric.
To prove that Q_j converges weakly, we use Theorem 5.6.4 in Appendix 5.6.2 which gives sufficient conditions for the weak convergence of continuous time martingales in $D[0, \infty)$. We now check the conditions of Theorem 5.6.4. Condition (a) is trivially satisfied since $Q_j(0) = 0$.
By condition (C6), $E\left\{Y_{1,1.}(t)\right\}^{-1}$ is a bounded function. By applying Theorem 5.6.3 in Appendix 5.6.2, we obtain that the predictable variation process

for Q_j at t is given by

$$\begin{aligned} V_K(t) &= < Q_j, Q_j > (t) \\ &= K^{-1} \sum_{i=1}^{K} \left[\int_0^t E\left\{Y_{1,1.}(u)\right\}^{-2} d < M_{1,ij}, M_{1,ij} > (u) \right] \end{aligned}$$

Since $< M_{1,ij}, M_{1,ij} > (u) = \int_0^u Y_{ij}(s)\lambda_1(s)ds$, we have that

$$V_K(t) = \int_0^t \left\{ K^{-1} \sum_{i=1}^{K} Y_{ij}(u) \right\} E\left\{Y_{1,1.}(u)\right\}^{-2} \lambda_1(u)du.$$

For $i = 1, \ldots, K$ and for j fixed, $Y_{ij}(u)$ are independent and identically distributed random variables with finite mean. Application of the Kolmogorov strong law of large numbers gives

$$K^{-1} \sum_{i=1}^{K} Y_{ij}(u) \xrightarrow{a.s.} E\left\{Y_{1j}(u)\right\}.$$

It then follows from the dominated convergence theorem that $V_K(t)$ converges almost surely to $V(t)$, for each fixed t, where

$$V(t) = \int_0^t E\left\{Y_{1j}(u)\right\} E\left\{Y_{1,1.}(u)\right\}^{-2} \lambda_1(u)du.$$

Hence, condition (b) is also satisfied. For a function $f : \mathbb{R} \to \mathbb{R}$, define the map $J_T(f) = \sup_{t\in[0,T]} |f(t) - f(t^-)|$. To establish condition (c), note that

$$\begin{aligned} J_{\tau_1}(Q_j) &= \sup_{t\in[0,\tau_1]} \left| Q_j(t) - Q_j(t^-) \right| \\ &= \sup_{t\in[0,\tau_1]} \left| K^{-1/2} \sum_{i=1}^{K} \int_{t^-}^t E\left\{Y_{1,1.}(u)\right\}^{-1} dM_{1,ij}(u) \right|. \end{aligned}$$

It follows from (5.19) that

$$\begin{aligned} J_{\tau_1}(Q_j) &= \sup_{t\in[0,\tau_1]} \left| K^{-1/2} \sum_{i=1}^{K} \int_{t^-}^t E\left\{Y_{1,1.}(u)\right\}^{-1} dN_{ij}(u) \right. \\ &\qquad \left. -K^{-1/2} \sum_{i=1}^{K} \int_{t^-}^t E\left\{Y_{1,1.}(u)\right\}^{-1} Y_{ij}(u)\lambda_1(u)du \right|. \end{aligned}$$

The second term on the right hand side of the above equation is equal to zero since λ_1 is a continuous function and since Y_{ij} is left continuous. It then follows that

$$\begin{aligned} J_{\tau_1}(Q_j) &= \sup_{t\in[0,\tau_1]} \left| K^{-1/2} \sum_{i=1}^{K} E\left\{Y_{1,1.}(t)\right\}^{-1} \left\{N_{ij}(t) - N_{ij}(t^-)\right\} \right| \\ &= \sup_{t\in[0,\tau_1]} \left| K^{-1/2} \sum_{i=1}^{K} E\left\{Y_{1,1.}(t)\right\}^{-1} \delta_{ij} I\left(X_{ij} = t\right) \right|. \end{aligned}$$

Hence, we have that

$$J_{\tau_1}(Q_j)^2 = \sup_{t\in[0,\tau_1]} \left| K^{-1} \sum_{i=1}^{K} \sum_{i'=1}^{K} E\left\{Y_{1,1.}(t)\right\}^{-2} \delta_{ij}\delta_{i'j} I\left(X_{ij} = t\right) I\left(X_{i'j} = t\right) \right|.$$

Since $J_{\tau_1}(Q_j)^2$ is a non-negative random variable, we have that

$$E\left\{J_{\tau_1}(Q_j)^2\right\} = \int_0^{+\infty} P\left(J_{\tau_1}(Q_j)^2 > x\right) dx.$$

By using (C6) and the fact that the probability of having a tie between two observed event times is zero, we obtain that

$$E\left\{J_{\tau_1}(Q_j)^2\right\} = \int_0^{1/(L_1^2 K)} P\left(J_{\tau_1}(Q_j)^2 > x\right) dx \leq \frac{1}{L_1^2 K},$$

which converges to zero for K tending to infinity. Because, for each fixed k, $0 \leq J_k(Q_j)^2 \leq J_{\tau_1}(Q_j)^2$, we obtain that condition c) of Theorem 5.6.4 is satisfied. This completes the proof. □

Since $\hat{\Lambda}_1$, resp. $\hat{\Lambda}_2$, is a uniform consistent estimator of the cumulative hazard function Λ_1, resp. Λ_2, we can use the estimated marginal survival functions (5.21) in the second stage of the estimation. In Section 5.4.3, we use the weak convergence of $\hat{\Lambda}_1$ and $\hat{\Lambda}_2$ to prove the asymptotic normality of the estimator for the copula parameter $\hat{\boldsymbol{\zeta}}$.

5.4.3 Second stage: estimation of the association parameter

Assume that z_{ij} is a deterministic binary covariate at the observational unit level, such that $z_{ij} = 0$, for $j = 1, 2$, and $z_{ij} = 1$, for $j = 3, 4$, where

$i = 1, \ldots, K$. This choice corresponds to the binary covariate indicating the location (front or rear) of the udder quarters in the udder infection data, introduced in Section 1.4.1. Using the Nelson Aalen estimators $\hat{\Lambda}_1$, resp. $\hat{\Lambda}_2$, for the cumulative hazard function for the group with $z_{ij} = 0$, resp. $z_{ij} = 1$, we have that the pseudo loglikelihood is given by

$$\log L_P(\boldsymbol{\zeta}) = \sum_{i=1}^{K} l\left\{\hat{\Lambda}_1(X_{i1}), \hat{\Lambda}_1(X_{i2}), \hat{\Lambda}_2(X_{i3}), \hat{\Lambda}_2(X_{i4})\right\}, \tag{5.54}$$

where $l\left\{\hat{\Lambda}_1(X_{i1}), \hat{\Lambda}_1(X_{i2}), \hat{\Lambda}_2(X_{i3}), \hat{\Lambda}_2(X_{i4})\right\}$ is the contribution to the pseudo loglikelihood for cluster i. It can be seen from (5.2) that

$$\begin{aligned}
& l\left\{\hat{\Lambda}_1(X_{i1}), \hat{\Lambda}_1(X_{i2}), \hat{\Lambda}_2(X_{i3}), \hat{\Lambda}_2(X_{i4})\right\} \\
= & \left[\Delta_i \log\left\{C_{\boldsymbol{\zeta}}(v_{i1}, v_{i2}, v_{i3}, v_{i4})\right\}\right. \\
& + \sum_{j=1}^{4}\left[\Delta_i(j)\log\left\{\frac{\partial C_{\boldsymbol{\zeta}}(v_{i1}, v_{i2}, v_{i3}, v_{i4})}{\partial v_{ij}}\right\}\right] \\
& + \sum_{j\neq k}\left[\Delta_i(j,k)\log\left\{\frac{\partial^2 C_{\boldsymbol{\zeta}}(v_{i1}, v_{i2}, v_{i3}, v_{i4})}{\partial v_{ij}\,\partial v_{ik}}\right\}\right] \\
& + \sum_{j\neq k\neq l}\left[\Delta_i(j,k,l)\log\left\{\frac{\partial^3 C_{\boldsymbol{\zeta}}(v_{i1}, v_{i2}, v_{i3}, v_{i4})}{\partial v_{ij}\,\partial v_{ik}\,\partial v_{il}}\right\}\right] \\
& \left. + \Delta_i(1,2,3,4)\log\left\{\frac{\partial^4 C_{\boldsymbol{\zeta}}(v_{i1}, v_{i2}, v_{i3}, v_{i4})}{\partial v_{i1}\,\partial v_{i2}\,\partial v_{i3}\,\partial v_{i4}}\right\}\right]_{v_{ij}=\hat{S}_{ij}(X_{ij}) \text{ for } j=1,\ldots,4},
\end{aligned}$$

where

$$\hat{S}_{ij}(t) = \begin{cases} \exp\{-\hat{\Lambda}_1(t)\} & \text{for } j = 1, 2 \\ \exp\{-\hat{\Lambda}_2(t)\} & \text{for } j = 3, 4, \end{cases}$$

which corresponds to (5.18). In the second step of the estimation procedure we estimate the association parameter $\boldsymbol{\zeta}$ by maximising $\log L_P(\boldsymbol{\zeta})$ in the previous expression with respect to $\boldsymbol{\zeta}$.

Let $\boldsymbol{\zeta}$ be a q-variate association parameter vector. Define the following quan-

tities:

$$\begin{aligned}
\mathbf{W}_{\boldsymbol{\zeta}}(\boldsymbol{\zeta}, w_1, w_2, w_3, w_4) &= \frac{\partial}{\partial \boldsymbol{\zeta}} l(\boldsymbol{\zeta}, w_1, w_2, w_3, w_4), \\
\mathbf{V}_{\boldsymbol{\zeta},j}(\boldsymbol{\zeta}, w_1, w_2, w_3, w_4) &= \frac{\partial^2}{\partial \boldsymbol{\zeta} \partial w_j} l(\boldsymbol{\zeta}, w_1, w_2, w_3, w_4), \text{ for } j = 1, \ldots, 4, \\
\mathbf{V}_{\boldsymbol{\zeta}}(\boldsymbol{\zeta}, w_1, w_2, w_3, w_4) &= \frac{\partial^2}{\partial \boldsymbol{\zeta} \partial \boldsymbol{\zeta}'} l(\boldsymbol{\zeta}, w_1, w_2, w_3, w_4),
\end{aligned}$$

and

$$\mathbf{I}_{2,\boldsymbol{\zeta}} = E\left[-\mathbf{V}_{\boldsymbol{\zeta}}\{\boldsymbol{\zeta}_0, \Lambda_1(X_{11}), \Lambda_1(X_{12}), \Lambda_2(X_{13}), \Lambda_2(X_{14})\}\right].$$

Note that $\mathbf{W}_{\boldsymbol{\zeta}}(\boldsymbol{\zeta}, w_1, w_2, w_3, w_4)$ and $\mathbf{V}_{\boldsymbol{\zeta},j}(\boldsymbol{\zeta}, w_1, w_2, w_3, w_4)$ are q-dimensional column vectors; $\mathbf{V}_{\boldsymbol{\zeta}}(\boldsymbol{\zeta}, w_1, w_2, w_3, w_4)$ and $\mathbf{I}_{2,\boldsymbol{\zeta}}$ are $q \times q$-matrices.
We use the following assumption to prove the consistency and the asymptotic normality of $\hat{\boldsymbol{\zeta}}$:

(A7) $\mathbf{W}_{\boldsymbol{\zeta}}\{\boldsymbol{\zeta}, \Lambda_1(t_1), \Lambda_1(t_2), \Lambda_2(t_3), \Lambda_2(t_4)\}$,
$\mathbf{V}_{\boldsymbol{\zeta}}\{\boldsymbol{\zeta}, \Lambda_1(t_1), \Lambda_1(t_2), \Lambda_2(t_3), \Lambda_2(t_4)\}$,
$\mathbf{V}_{\boldsymbol{\zeta},j}\{\boldsymbol{\zeta}, \Lambda_1(t_1), \Lambda_1(t_2), \Lambda_2(t_3), \Lambda_2(t_4)\}$, for $j = 1, \ldots, 4$,
are continuous on $\mathcal{B}_{\boldsymbol{\zeta}} \times [0, B_{L_1} + \delta_1] \times [0, B_{L_1} + \delta_1] \times [0, B_{L_2} + \delta_2] \times [0, B_{L_2} + \delta_2]$, where $\mathcal{B}_{\boldsymbol{\zeta}}$ is a compact neighbourhood of $\boldsymbol{\zeta}_0$ that contains $\hat{\boldsymbol{\zeta}}$, where B_{L_1}, resp. B_{L_2}, is the upper bound of $\Lambda_1(\tau_1)$, resp. $\Lambda_2(\tau_2)$, as defined in (C4), and where $\delta_1, \delta_2 > 0$ are fixed constants that are used in the proof of (5.55).

The consistency of $\hat{\boldsymbol{\zeta}}$ follows from the following theorem:

Theorem 5.4.3. *If the assumptions (A6) and (A7) hold, then the estimator $\hat{\boldsymbol{\zeta}}$ converges in probability to $\boldsymbol{\zeta}_0$.*

The proof of Theorem 5.4.3 can be given along the lines of the proof of Theorem 1 in Glidden (2000). Recall from Section 5.3.3 that Glidden (2000) studies a semi-parametric two-stage estimation procedure for a Clayton copula in the case where $(\mathbf{T}_i, \mathbf{C}_i, \mathbf{Z}_i)$ are i.i.d. For this situation, he proves the consistency of the estimator of the association parameter. Since we use a nonparametric approach in the first step of the estimation, we adapt his proof by using the

uniform consistency of $\hat{\Lambda}_1$, resp. $\hat{\Lambda}_2$, on $[0, \tau_1]$, resp. $[0, \tau_2]$. Note that it follows from the choice of the covariate z_{ij} that the loglikelihood function, given in (5.2), is a sum of i.i.d. random variables. Hence, the proof can be given using similar arguments as presented in Glidden (2000).

The following theorem establishes the asymptotic normality of $\hat{\boldsymbol{\zeta}}$:

Theorem 5.4.4. *If the assumptions (A6) and (A7) hold, then* $K^{1/2}(\hat{\boldsymbol{\zeta}} - \boldsymbol{\zeta}_0)$ *converges to a normal distribution with mean vector zero and variance-covariance* $\boldsymbol{\Sigma}_2 = \mathbf{I}_{2,\boldsymbol{\zeta}}^{-1} + \mathbf{I}_{2,\boldsymbol{\zeta}}^{-1} \mathbf{V}\left(\boldsymbol{\Phi}_{2,1}\right) \mathbf{I}_{2,\boldsymbol{\zeta}}^{-1}$, *where*

$$\begin{aligned} \mathbf{V}\left(\boldsymbol{\Phi}_{2,1}\right) &= E\left(\boldsymbol{\Phi}_{2,1}\boldsymbol{\Phi}_{2,1}^{'}\right) \\ \boldsymbol{\Phi}_{2,1} &= \int_0^{\tau_1} \mathbf{IC}_1(t_1) d\Psi_{1,1}(t_1) + \int_0^{\tau_1} \mathbf{IC}_2(t_2) d\Psi_{1,1}(t_2) \\ &\quad + \int_0^{\tau_2} \mathbf{IC}_3(t_3) d\Psi_{2,1}(t_3) + \int_0^{\tau_2} \mathbf{IC}_4(t_4) d\Psi_{2,1}(t_4), \end{aligned}$$

with $\Psi_{1,1}$ *and* $\Psi_{2,1}$ *as defined in (5.49) in Section 5.4.2,*

$$\begin{aligned} \mathbf{IC}_1(t_1) &= \int_{t_1}^{\tau_1} \int_0^{\tau_1} \int_0^{\tau_2} \int_0^{\tau_2} \mathbf{V}_{\boldsymbol{\zeta},1}\left\{\boldsymbol{\zeta}_0, \Lambda_1(u), \Lambda_1(t_2), \Lambda_2(t_3), \Lambda_2(t_4)\right\} \\ &\qquad \times dH_{\boldsymbol{\zeta}_0}(u, t_2, t_3, t_4), \end{aligned}$$

and similar expressions for $\mathbf{IC}_2(t_2)$, $\mathbf{IC}_3(t_3)$ *and* $\mathbf{IC}_4(t_4)$.
Note that the expressions for $\mathbf{IC}_1(t_1)$, $\mathbf{IC}_2(t_2)$, $\mathbf{IC}_3(t_3)$ *and* $\mathbf{IC}_4(t_4)$ *differ from the expressions given in Theorem 5.3.6 since we use here a nonparametric approach in the first step of the estimation procedure.*

Proof. The first order Taylor series expansion for $\mathbf{U}_{\boldsymbol{\zeta}}\left(\boldsymbol{\zeta}, \hat{\Lambda}_1, \hat{\Lambda}_2\right)$ around $\boldsymbol{\zeta}_0$, evaluated at $\hat{\boldsymbol{\zeta}}$, is given by

$$\begin{aligned} &\mathbf{U}_{\boldsymbol{\zeta}}\left(\hat{\boldsymbol{\zeta}}, \hat{\Lambda}_1, \hat{\Lambda}_2\right) \\ = \; &\mathbf{U}_{\boldsymbol{\zeta}}\left(\boldsymbol{\zeta}_0, \hat{\Lambda}_1, \hat{\Lambda}_2\right) \\ &+ \sum_{i=1}^{K} \mathbf{V}_{\boldsymbol{\zeta}}\left\{\boldsymbol{\zeta}^*, \hat{\Lambda}_1\left(X_{i1}\right), \hat{\Lambda}_1\left(X_{i2}\right), \hat{\Lambda}_2\left(X_{i3}\right), \hat{\Lambda}_2\left(X_{i4}\right)\right\}\left(\hat{\boldsymbol{\zeta}} - \boldsymbol{\zeta}_0\right), \end{aligned}$$

where $\boldsymbol{\zeta}^*$ is on the line segment between $\hat{\boldsymbol{\zeta}}$ and $\boldsymbol{\zeta}_0$.
Since $\mathbf{U}_{\boldsymbol{\zeta}}\left(\hat{\boldsymbol{\zeta}}, \hat{\Lambda}_1, \hat{\Lambda}_2\right) = 0$, we obtain

$$\begin{aligned} & K^{1/2}\left(\hat{\boldsymbol{\zeta}} - \boldsymbol{\zeta}_0\right) \\ = & \left[-K^{-1}\sum_{i=1}^{K}\mathbf{V}_{\boldsymbol{\zeta}}\left\{\boldsymbol{\zeta}^*, \hat{\Lambda}_1\left(X_{i1}\right), \hat{\Lambda}_1\left(X_{i2}\right), \hat{\Lambda}_2\left(X_{i3}\right), \hat{\Lambda}_2\left(X_{i4}\right)\right\}\right]^{-1} \\ & \times K^{-1/2}\mathbf{U}_{\boldsymbol{\zeta}}\left(\boldsymbol{\zeta}_0, \hat{\Lambda}_1, \hat{\Lambda}_2\right). \end{aligned}$$

Straightforward calculations, using the consistency of $\hat{\boldsymbol{\zeta}}$, the uniform consistency of $\hat{\Lambda}_1$ and $\hat{\Lambda}_2$, condition (A7) and the weak law of large numbers, show that

$$-K^{-1}\sum_{i=1}^{K}\mathbf{V}_{\boldsymbol{\zeta}}\left\{\boldsymbol{\zeta}^*, \hat{\Lambda}_1\left(X_{i1}\right), \hat{\Lambda}_1\left(X_{i2}\right), \hat{\Lambda}_2\left(X_{i3}\right), \hat{\Lambda}_2\left(X_{i4}\right)\right\} \xrightarrow{P} \mathbf{I}_{2,\boldsymbol{\zeta}}. \quad (5.55)$$

Let $H_{\boldsymbol{\zeta}_0}$ be the joint distribution function of $(X_{i1}, X_{i2}, X_{i3}, X_{i4})$, for $i = 1, \ldots, K$ and let $h_{\boldsymbol{\zeta}_0}$ be the corresponding joint density function. Define H_K as the empirical distribution function of $(X_{i1}, X_{i2}, X_{i3}, X_{i4})$, i.e.

$$H_K(t_1, t_2, t_3, t_4) = \frac{1}{K}\sum_{i=1}^{K} I\left(X_{i1} \leq t_1, X_{i2} \leq t_2, X_{i3} \leq t_3, X_{i4} \leq t_4\right).$$

We then obtain that

$$\begin{aligned} & K^{-1/2}\mathbf{U}_{\boldsymbol{\zeta}}\left(\boldsymbol{\zeta}_0, \hat{\Lambda}_1, \hat{\Lambda}_2\right) \\ = & K^{-1/2}\sum_{i=1}^{K}\mathbf{W}_{\boldsymbol{\zeta}}\left\{\boldsymbol{\zeta}_0, \hat{\Lambda}_1(X_{i1}), \hat{\Lambda}_1(X_{i2}), \hat{\Lambda}_2(X_{i3}), \hat{\Lambda}_2(X_{i4})\right\} \\ = & K^{1/2}\int_0^{\tau_1}\int_0^{\tau_1}\int_0^{\tau_2}\int_0^{\tau_2}\mathbf{W}_{\boldsymbol{\zeta}}\left\{\boldsymbol{\zeta}_0, \hat{\Lambda}_1(t_1), \hat{\Lambda}_1(t_2), \hat{\Lambda}_2(t_3), \hat{\Lambda}_2(t_4)\right\} \\ & \qquad \times dH_K\left(t_1, t_2, t_3, t_4\right). \end{aligned}$$

Rewriting the above expression gives

$$\begin{aligned} & K^{1/2}\int_0^{\tau_1}\int_0^{\tau_1}\int_0^{\tau_2}\int_0^{\tau_2}\mathbf{W}_{\boldsymbol{\zeta}}\left\{\boldsymbol{\zeta}_0,\hat{\Lambda}_1(t_1),\hat{\Lambda}_1(t_2),\hat{\Lambda}_2(t_3),\hat{\Lambda}_2(t_4)\right\} \\ & \qquad\qquad \times dH_{\boldsymbol{\zeta}_0}(t_1,t_2,t_3,t_4) \\ & +K^{1/2}\int_0^{\tau_1}\int_0^{\tau_1}\int_0^{\tau_2}\int_0^{\tau_2}\mathbf{W}_{\boldsymbol{\zeta}}\left\{\boldsymbol{\zeta}_0,\hat{\Lambda}_1(t_1),\hat{\Lambda}_1(t_2),\hat{\Lambda}_2(t_3),\hat{\Lambda}_2(t_4)\right\} \\ & \qquad\qquad \times d\left(H_K-H_{\boldsymbol{\zeta}_0}\right)(t_1,t_2,t_3,t_4) \\ = \;& \mathbf{T}_K\left(\boldsymbol{\zeta}_0,\hat{\Lambda}_1,\hat{\Lambda}_2\right)+\mathbf{Z}_K\left(\boldsymbol{\zeta}_0,\hat{\Lambda}_1,\hat{\Lambda}_2\right). \end{aligned} \tag{5.56}$$

The last term, $\mathbf{Z}_K\left(\boldsymbol{\zeta}_0,\hat{\Lambda}_1,\hat{\Lambda}_2\right)$, can be written as

$$\begin{aligned} & K^{1/2}\int_0^{\tau_1}\int_0^{\tau_1}\int_0^{\tau_2}\int_0^{\tau_2}\mathbf{W}_{\boldsymbol{\zeta}}\left\{\boldsymbol{\zeta}_0,\Lambda_1(t_1),\Lambda_1(t_2),\Lambda_2(t_3),\Lambda_2(t_4)\right\} \\ & \qquad\qquad \times d\left(H_K-H_{\boldsymbol{\zeta}_0}\right)(t_1,t_2,t_3,t_4) \\ & + K^{1/2}\int_0^{\tau_1}\int_0^{\tau_1}\int_0^{\tau_2}\int_0^{\tau_2}\Big[\mathbf{W}_{\boldsymbol{\zeta}}\left\{\boldsymbol{\zeta}_0,\hat{\Lambda}_1(t_1),\hat{\Lambda}_1(t_2),\hat{\Lambda}_2(t_3),\hat{\Lambda}_2(t_4)\right\} \\ & \qquad\qquad -\mathbf{W}_{\boldsymbol{\zeta}}\left\{\boldsymbol{\zeta}_0,\Lambda_1(t_1),\Lambda_1(t_2),\Lambda_2(t_3),\Lambda_2(t_4)\right\}\Big] \\ & \qquad\qquad d\left(H_K-H_{\boldsymbol{\zeta}_0}\right)(t_1,t_2,t_3,t_4). \end{aligned}$$

By condition (A7) and the uniform consistency of $\hat{\Lambda}_1$, resp. $\hat{\Lambda}_2$, on $[0,\tau_1]$, resp. $[0,\tau_2]$, we obtain that

$$\begin{aligned} & \sup_{(t_1,t_2,t_3,t_4)\in[0,\tau_1]\times[0,\tau_1]\times[0,\tau_2]\times[0,\tau_2]}\Big|\mathbf{W}_{\boldsymbol{\zeta}}\left\{\boldsymbol{\zeta}_0,\hat{\Lambda}_1(t_1),\hat{\Lambda}_1(t_2),\hat{\Lambda}_2(t_3),\hat{\Lambda}_2(t_4)\right\} \\ & -\mathbf{W}_{\boldsymbol{\zeta}}\left\{\boldsymbol{\zeta}_0,\Lambda_1(t_1),\Lambda_1(t_2),\Lambda_2(t_3),\Lambda_2(t_4)\right\}\Big| = o_p(1). \end{aligned}$$

Further, we have that $K^{1/2}\left(H_K-H_{\boldsymbol{\zeta}_0}\right)=O_p(1)$. Hence, it follows that the second term of $\mathbf{Z}_K$ converges to zero in probability. The first term of Z_K is a sum of independent and identically distributed random vectors with mean vector zero and variance-covariance matrix $\mathbf{I}_{2,\boldsymbol{\zeta}}$.

We rewrite $\mathbf{T}_K\left(\boldsymbol{\zeta}_0,\hat{\Lambda}_1,\hat{\Lambda}_2\right)$ by using a Taylor expansion of $\mathbf{W}_{\boldsymbol{\zeta}}\{\boldsymbol{\zeta}_0,\Lambda_1(t_1),\Lambda_1(t_2),\Lambda_2(t_3),\Lambda_2(t_4)\}$ around $\{\Lambda_1(t_1),\Lambda_1(t_2),\Lambda_2(t_3),\Lambda_2(t_4)\}$, evaluated at

$\left\{\hat{\Lambda}_1(t_1), \hat{\Lambda}_1(t_2), \hat{\Lambda}_2(t_3), \hat{\Lambda}_2(t_4)\right\}$. This leads to

$$
\begin{aligned}
& \mathbf{T}_K\left(\boldsymbol{\zeta}_0, \hat{\Lambda}_1, \hat{\Lambda}_2\right) \\
= \; & K^{1/2} \int_0^{\tau_1}\int_0^{\tau_1}\int_0^{\tau_2}\int_0^{\tau_2} \Big[\mathbf{W}_{\boldsymbol{\zeta}}\left\{\boldsymbol{\zeta}_0, \Lambda_1(t_1), \Lambda_1(t_2), \Lambda_2(t_3), \Lambda_2(t_4)\right\} \\
& + \mathbf{V}_{\boldsymbol{\zeta},1}\left\{\boldsymbol{\zeta}_0, \gamma_1^*, \Lambda_1(t_2), \Lambda_2(t_3), \Lambda_2(t_4)\right\}\left\{\hat{\Lambda}_1(t_1) - \Lambda_1(t_1)\right\} \\
& + \mathbf{V}_{\boldsymbol{\zeta},2}\left\{\boldsymbol{\zeta}_0, \Lambda_1(t_1), \gamma_2^*, \Lambda_2(t_3), \Lambda_2(t_4)\right\}\left\{\hat{\Lambda}_1(t_2) - \Lambda_1(t_2)\right\} \\
& + \mathbf{V}_{\boldsymbol{\zeta},3}\left\{\boldsymbol{\zeta}_0, \Lambda_1(t_1), \Lambda_1(t_2), \gamma_3^*, \Lambda_2(t_4)\right\}\left\{\hat{\Lambda}_2(t_3) - \Lambda_2(t_3)\right\} \\
& + \mathbf{V}_{\boldsymbol{\zeta},4}\left\{\boldsymbol{\zeta}_0, \Lambda_1(t_1), \Lambda_1(t_2), \Lambda_2(t_3), \gamma_4^*\right\}\left\{\hat{\Lambda}_2(t_4) - \Lambda_2(t_4)\right\}\Big] \\
& \times dH_{\boldsymbol{\zeta}_0}(t_1, t_2, t_3, t_4),
\end{aligned}
$$

where γ_1^* is on the line segment between $\Lambda_1(t_1)$ and $\hat{\Lambda}_1(t_1)$, γ_2^* is on the line segment between $\Lambda_1(t_2)$ and $\hat{\Lambda}_1(t_2)$, γ_3^* is on the line segment between $\Lambda_2(t_3)$ and $\hat{\Lambda}_2(t_3)$, and γ_4^* is on the line segment between $\Lambda_2(t_4)$ and $\hat{\Lambda}_2(t_4)$.
Rewriting the above expression gives

$$
\begin{aligned}
& \mathbf{T}_K\left(\boldsymbol{\zeta}_0, \hat{\Lambda}_1, \hat{\Lambda}_2\right) \\
= \; & \mathbf{T}_K\left(\boldsymbol{\zeta}_0, \Lambda_1, \Lambda_2\right) \\
& + K^{1/2} \int_0^{\tau_1}\int_0^{\tau_1}\int_0^{\tau_2}\int_0^{\tau_2} \Big[\\
& \mathbf{V}_{\boldsymbol{\zeta},1}\left\{\boldsymbol{\zeta}_0, \Lambda_1(t_1), \Lambda_1(t_2), \Lambda_2(t_3), \Lambda_2(t_4)\right\}\left\{\hat{\Lambda}_1(t_1) - \Lambda_1(t_1)\right\} \\
& + \mathbf{V}_{\boldsymbol{\zeta},2}\left\{\boldsymbol{\zeta}_0, \Lambda_1(t_1), \Lambda_1(t_2), \Lambda_2(t_3), \Lambda_2(t_4)\right\}\left\{\hat{\Lambda}_1(t_2) - \Lambda_1(t_2)\right\} \\
& + \mathbf{V}_{\boldsymbol{\zeta},3}\left\{\boldsymbol{\zeta}_0, \Lambda_1(t_1), \Lambda_1(t_2), \Lambda_2(t_3), \Lambda_2(t_4)\right\}\left\{\hat{\Lambda}_2(t_3) - \Lambda_2(t_3)\right\} \\
& + \mathbf{V}_{\boldsymbol{\zeta},4}\left\{\boldsymbol{\zeta}_0, \Lambda_1(t_1), \Lambda_1(t_2), \Lambda_2(t_3), \Lambda_2(t_4)\right\}\left\{\hat{\Lambda}_2(t_4) - \Lambda_2(t_4)\right\}\Big] \\
& \times dH_{\boldsymbol{\zeta}_0}(t_1, t_2, t_3, t_4) \\
& + \mathbf{R}_K\left(\boldsymbol{\zeta}_0, \Lambda_1, \Lambda_2\right),
\end{aligned}
$$

where

$$
\begin{aligned}
& \mathbf{R}_K\left(\boldsymbol{\zeta}_0, \Lambda_1, \Lambda_2\right) \\
= \; & K^{1/2} \int_0^{\tau_1} \int_0^{\tau_1} \int_0^{\tau_2} \int_0^{\tau_2} \Big[\Big[\mathbf{V}_{\boldsymbol{\zeta},1} \left\{ \boldsymbol{\zeta}_0, \gamma_1^*, \Lambda_1(t_2), \Lambda_2(t_3), \Lambda_2(t_4) \right\} \\
& - \mathbf{V}_{\boldsymbol{\zeta},1} \left\{ \boldsymbol{\zeta}_0, \Lambda_1(t_1), \Lambda_1(t_2), \Lambda_2(t_3), \Lambda_2(t_4) \right\} \Big] \left\{ \hat{\Lambda}_1(t_1) - \Lambda_1(t_1) \right\} \\
& + \Big[\mathbf{V}_{\boldsymbol{\zeta},2} \left\{ \boldsymbol{\zeta}_0, \Lambda_1(t_1), \gamma_2^*, \Lambda_2(t_3), \Lambda_2(t_4) \right\} \\
& - \mathbf{V}_{\boldsymbol{\zeta},2} \left\{ \boldsymbol{\zeta}_0, \Lambda_1(t_1), \Lambda_1(t_2), \Lambda_2(t_3), \Lambda_2(t_4) \right\} \Big] \left\{ \hat{\Lambda}_1(t_2) - \Lambda_1(t_2) \right\} \\
& + \Big[\mathbf{V}_{\boldsymbol{\zeta},3} \left\{ \boldsymbol{\zeta}_0, \Lambda_1(t_1), \Lambda_1(t_2), \gamma_3^*, \Lambda_2(t_4) \right\} \\
& - \mathbf{V}_{\boldsymbol{\zeta},3} \left\{ \boldsymbol{\zeta}_0, \Lambda_1(t_1), \Lambda_1(t_2), \Lambda_2(t_3), \Lambda_2(t_4) \right\} \Big] \left\{ \hat{\Lambda}_2(t_3) - \Lambda_2(t_3) \right\} \\
& + \Big[\mathbf{V}_{\boldsymbol{\zeta},4} \left\{ \boldsymbol{\zeta}_0, \Lambda_1(t_1), \Lambda_1(t_2), \Lambda_2(t_3), \gamma_4^* \right\} \\
& - \mathbf{V}_{\boldsymbol{\zeta},4} \left\{ \boldsymbol{\zeta}_0, \Lambda_1(t_1), \Lambda_1(t_2), \Lambda_2(t_3), \Lambda_2(t_4) \right\} \Big] \left\{ \hat{\Lambda}_2(t_4) - \Lambda_2(t_4) \right\} \Big] \\
& \times dH_{\boldsymbol{\zeta}_0}(t_1, t_2, t_3, t_4).
\end{aligned}
$$

Straightforward arguments using condition (A7), Theorems 5.4.1 and 5.4.2, show that $R_K\left(\boldsymbol{\zeta}_0, \Lambda_1, \Lambda_2\right)$ converges to zero in probability.
Further,

$$
\begin{aligned}
& \mathbf{T}_K\left(\boldsymbol{\zeta}_0, \Lambda_1, \Lambda_2\right) \\
= \; & K^{1/2} \int_0^{\tau_1} \int_0^{\tau_1} \int_0^{\tau_2} \int_0^{\tau_2} \mathbf{W}_{\boldsymbol{\zeta}} \left\{ \boldsymbol{\zeta}_0, \Lambda_1(t_1), \Lambda_1(t_2), \Lambda_2(t_3), \Lambda_2(t_4) \right\} \\
& \qquad \times dH_{\boldsymbol{\zeta}_0}\left(t_1, t_2, t_3, t_4\right). \qquad (5.57)
\end{aligned}
$$

Since

$$
\mathbf{W}_{\boldsymbol{\zeta}} \left\{ \boldsymbol{\zeta}_0, \Lambda_1(t_1), \Lambda_1(t_2), \Lambda_2(t_3), \Lambda_2(t_4) \right\} = \frac{\partial \log h_{\boldsymbol{\zeta}_0}(t_1, t_2, t_3, t_4)}{\partial \boldsymbol{\zeta}},
$$

it follows that

$$
\begin{aligned}
& \mathbf{W}_{\boldsymbol{\zeta}} \left\{ \boldsymbol{\zeta}_0, \Lambda_1(t_1), \Lambda_1(t_2), \Lambda_2(t_3), \Lambda_2(t_4) \right\} h_{\boldsymbol{\zeta}_0}(t_1, t_2, t_3, t_4) \\
= \; & \frac{\partial h_{\boldsymbol{\zeta}_0}(t_1, t_2, t_3, t_4)}{\partial \boldsymbol{\zeta}}. \qquad (5.58)
\end{aligned}
$$

Substituting (5.58) in (5.57) and changing the order of the integral and the derivative, leads to

$$\begin{aligned} & \mathbf{T}_K\left(\boldsymbol{\zeta}_0, \Lambda_1, \Lambda_2\right) \\ = \; & K^{1/2} \int_0^{\tau_1} \frac{\partial}{\partial \boldsymbol{\zeta}} \left[\int_0^{\tau_1} \int_0^{\tau_2} \int_0^{\tau_2} h_{\boldsymbol{\zeta}_0}(t_1, t_2, t_3, t_4) dt_4 dt_3 dt_2 \right] dt_1. \end{aligned}$$

Since the marginal survival functions do not depend on the association parameter $\boldsymbol{\zeta}$, we obtain that

$$\mathbf{T}_K\left(\boldsymbol{\zeta}_0, \Lambda_1, \Lambda_2\right) = 0.$$

Using that $\hat{\Lambda}_1(t_j) - \Lambda_1(t_j) = \int_0^{t_j} d(\hat{\Lambda}_1 - \Lambda_1)(u)$, for $j = 1, 2$, and $\hat{\Lambda}_2(t_j) - \Lambda_2(t_j) = \int_0^{t_j} d(\hat{\Lambda}_2 - \Lambda_2)(u)$, for $j = 3, 4$, we obtain, after changing the order of the integrals, that

$$\begin{aligned} & \mathbf{T}_K\left(\boldsymbol{\zeta}_0, \hat{\Lambda}_1, \hat{\Lambda}_2\right) \\ = \; & K^{1/2} \int_0^{\tau_1} \mathbf{IC}_1(t_1) d\left(\hat{\Lambda}_1 - \Lambda_1\right)(t_1) + K^{1/2} \int_0^{\tau_1} \mathbf{IC}_2(t_2) d\left(\hat{\Lambda}_1 - \Lambda_1\right)(t_2) \\ & + K^{1/2} \int_0^{\tau_2} \mathbf{IC}_3(t_3) d\left(\hat{\Lambda}_2 - \Lambda_2\right)(t_3) + K^{1/2} \int_0^{\tau_2} \mathbf{IC}_4(t_4) d\left(\hat{\Lambda}_2 - \Lambda_2\right)(t_4), \end{aligned}$$

where

$$\begin{aligned} \mathbf{IC}_1(t_1) &= \int_{t_1}^{\tau_1} \int_0^{\tau_1} \int_0^{\tau_2} \int_0^{\tau_2} \mathbf{V}_{\boldsymbol{\zeta},1}\left\{\boldsymbol{\zeta}_0, \Lambda_1(u), \Lambda_1(t_2), \Lambda_2(t_3), \Lambda_2(t_4)\right\} \\ &\qquad \times dH_{\boldsymbol{\zeta}_0}(u, t_2, t_3, t_4) \\ \mathbf{IC}_2(t_2) &= \int_0^{\tau_1} \int_{t_2}^{\tau_1} \int_0^{\tau_2} \int_0^{\tau_2} \mathbf{V}_{\boldsymbol{\zeta},2}\left\{\boldsymbol{\zeta}_0, \Lambda_1(t_1), \Lambda_1(u), \Lambda_2(t_3), \Lambda_2(t_4)\right\} \\ &\qquad \times dH_{\boldsymbol{\zeta}_0}(t_1, u, t_3, t_4) \\ \mathbf{IC}_3(t_3) &= \int_0^{\tau_1} \int_0^{\tau_1} \int_{t_3}^{\tau_2} \int_0^{\tau_2} \mathbf{V}_{\boldsymbol{\zeta},3}\left\{\boldsymbol{\zeta}_0, \Lambda_1(t_1), \Lambda_1(t_2), \Lambda_2(u), \Lambda_2(t_4)\right\} \\ &\qquad \times dH_{\boldsymbol{\zeta}_0}(t_1, t_2, u, t_4) \\ \mathbf{IC}_4(t_4) &= \int_0^{\tau_1} \int_0^{\tau_1} \int_0^{\tau_2} \int_{t_4}^{\tau_2} \mathbf{V}_{\boldsymbol{\zeta},4}\left\{\boldsymbol{\zeta}_0, \Lambda_1(t_1), \Lambda_1(t_2), \Lambda_2(t_3), \Lambda_2(u)\right\} \\ &\qquad \times dH_{\boldsymbol{\zeta}_0}(t_1, t_2, t_3, u). \end{aligned}$$

It follows from (5.53) in the proof of Theorem 5.4.2 that $T_K\left(\boldsymbol{\zeta}_0, \hat{\Lambda}_1, \hat{\Lambda}_2\right)$ is asymptotically equivalent with

$$K^{-1/2}\sum_{i=1}^{K}\boldsymbol{\Phi}_{2,i} \equiv K^{-1/2}\sum_{i=1}^{K}\left[\int_0^{\tau_1}\mathbf{IC}_1(t_1)d\Psi_{1,i}(t_1)+\int_0^{\tau_1}\mathbf{IC}_2(t_2)d\Psi_{1,i}(t_2)\right.$$
$$\left.+\int_0^{\tau_2}\mathbf{IC}_3(t_3)d\Psi_{2,i}(t_3)+\int_0^{\tau_2}\mathbf{IC}_4(t_4)d\Psi_{2,i}(t_4)\right].$$

The above expression is a sum of independent and identically distributed random vectors with mean vector zero. By the central limit theorem, $\mathbf{T}_K\left(\boldsymbol{\zeta}_0, \hat{\Lambda}_1, \hat{\Lambda}_2\right)$ converges to a normal distribution with mean vector zero and variance-covariance matrix $\mathbf{V}\left(\boldsymbol{\Phi}_{2,1}\right) = E\left(\boldsymbol{\Phi}_{2,1}\boldsymbol{\Phi}_{2,1}^{'}\right)$. Since $\mathbf{T}_K$ and $\mathbf{Z}_K$ are asymptotically independent, it follows by (5.56) that $K^{-1/2}\mathbf{U}_{\boldsymbol{\zeta}}\left(\boldsymbol{\zeta}_0, \hat{\Lambda}_1, \hat{\Lambda}_2\right)$ converges to a normal distribution with mean vector zero and variance-covariance matrix $\mathbf{I}_{2,\boldsymbol{\zeta}} + \mathbf{V}\left(\boldsymbol{\Phi}_{2,1}\right)$. Hence, $K^{-1/2}\left(\hat{\boldsymbol{\zeta}} - \boldsymbol{\zeta}_0\right)$ converges to a normal distribution with mean vector zero and variance-covariance matrix $\mathbf{I}_{2,\boldsymbol{\zeta}}^{-1} + \mathbf{I}_{2,\boldsymbol{\zeta}}^{-1}\mathbf{V}\left(\boldsymbol{\Phi}_{2,1}\right)\mathbf{I}_{2,\boldsymbol{\zeta}}^{-1}$. □

5.5 Conclusions

In this chapter, we propose a semi-parametric and a nonparametric two-stage estimation approach for four-dimensional copulas. Our motivation for this is the udder infection data set, introduced in Section 1.4.1.

In the first stage of the estimation procedure, the marginal survival functions are estimated. In the semi-parametric approach, we estimate the marginal survival functions using a marginal Cox model with a deterministic binary covariate at the cluster level and a deterministic binary covariate at the observational unit level. This type of covariates corresponds to the situation we have in the udder infection data. Spiekerman and Lin (1998) prove the consistency and the asymptotic normality of the parameter estimators in the marginal Cox model in the case of stochastic covariates. Since we deal with deterministic covariates, the situation we consider is different from the situation in Spiekerman and Lin (1998). We extend the asymptotic results in

their paper to specific cases of deterministic covariates. In the nonparametric approach, we account for a binary covariate at the observational unit level. Observational units having the same covariate value are pooled. We then estimate the marginal survival functions by using a Nelson-Aalen estimator for the pooled data. It is important to note that the pooled observational units are correlated. The consistency and the asymptotic normality results that we obtain for the pooled data are new and extend results in Shih and Louis (1995b) and in Spiekerman and Lin (1998).
In the second step of the estimation, the association parameter is estimated by maximising the loglikelihood, with the marginal survival functions replaced by their estimates obtained in the first step. Since the proofs of the consistency and the asymptotic normality of the estimator of the association parameter build on the asymptotic properties of the estimators obtained in the first step and since we study four-dimensional copula models, our results extend results obtained by Shih and Louis (1995b), Glidden (2000) and Andersen (2005).
In the semi-parametric approach, we consider a binary covariate at the cluster level and a binary covariate at the observational unit level. The proofs of the theorems presented in Sections 5.3.2 and 5.3.3 can easily be generalized to a categorical covariate at the cluster level and a categorical covariate at the observational unit level, both with a finite number of categories. For covariates at the observational unit level, we then need that the number of observational units having the same covariate value is the same for all clusters.
We study both a semi-parametric and a nonparametric approach to estimate the marginal survival functions. It is also possible to consider a parametric approach in the estimation of the marginal survival functions, e.g., using a Weibull regression model. This is an extension of Andersen (2005) that is not discussed in this thesis.

5.6 Appendix

In this section, we collect a number of theorems and properties which were needed in the proofs of this chapter.

5.6.1 The Kolmogorov sufficient condition

Theorem 5.6.1. *Let X_i, for $i = 1, 2, \ldots$, be a sequence of independent random variables with mean zero and finite variances σ_i^2, and set $S_K = \sum_{i=1}^{K} X_i$, for $K = 1, 2, \ldots$. Then*

$$\sum_{i=1}^{\infty} i^{-2}\sigma_i^2 < \infty \quad \Rightarrow \quad K^{-1}S_K \xrightarrow{a.s.} 0 \quad \text{as } K \to \infty.$$

Proof. The proof is presented in Gut (2005, p.288). □

5.6.2 Properties of martingales

Theorem 5.6.2. *Let N be an arbitrary counting process.*

1. *Then there exists a unique right-continuous predictable increasing process A such that $A(0) = 0$ a.s., $A(t) < \infty$ a.s. for any t, and the process $M = N - A$ is a local martingale.*

2. *If A in (1) is locally bounded, M is a local square integrable martingale.*

Proof. The proof is presented in Fleming and Harrington (1991, p.61). □

The process A in the previous decomposition of an arbitrary counting process is called the compensator.

The following theorem is presented in Martinussen and Scheike (2006, p.22):

Theorem 5.6.3. *Let M and $\tilde{M}$ be finite variation local square integrable martingales, and let H and $\tilde{H}$ be locally bounded predictable processes. Then $\int H dM$ and $\int \tilde{H} d\tilde{M}$ are local square integrable martingales, and the predictable covariation processes are*

$$< \int H dM, \int \tilde{H} d\tilde{M} >= \int H\tilde{H} d < M, \tilde{M} > .$$

The following proposition gives sufficient conditions for the weak convergence of continuous time martingales in $D[0, \infty)$:

Theorem 5.6.4. *Let $\mathcal{M}_K$, for $K = 1, 2, \ldots$, be a sequence of zero-mean square-integrable martingales with predictable variation process V_K. Let V be a continuous, increasing function on $[0, \infty)$ with $V(0) = 0$. Sufficient conditions for weak convergence of $\mathcal{M}_K$ in $D[0, \infty)$ are:*

a) $\mathcal{M}_K(0) \xrightarrow{P} 0$;

b) $V_K(t) \xrightarrow{P} V(t)$ *for each fixed* t;

c) $E\left\{J_k \left(\mathcal{M}_K\right)^2\right\} \longrightarrow 0$ *for each fixed* k, *as* $K \to \infty$.

Proof. The proof is presented in Pollard (1984, p. 179-182). □

5.6.3 Weak convergence in $\mathcal{D}[a, b]$

Let $\mathbb{D}$ be an arbitrary metric space with metric d.
Let $(\Omega, \mathcal{A}, P)$ be an arbitrary probability space. We first define the inner probability and the outer probability of an arbitrary subset B of Ω (van der Vaart and Wellner, 1996, Chapter 1). The concept "inner probability" is used to define asymptotic tightness, whereas the "outer probability" will be used in Theorem 5.6.7.

Definition 5.6.1. *The inner probability of an arbitrary subset B of Ω is*

$$P_*(B) = \sup \{P(A) : A \subset B, A \in \mathcal{A}\}.$$

Definition 5.6.2. *The outer probability of an arbitrary subset B of Ω is*

$$P^*(B) = \inf \{P(A) : A \supset B, A \in \mathcal{A}\}.$$

Definition 5.6.3. *Let $(\Omega_K, \mathcal{A}_K, P_K)$ be a sequence of probability spaces. A sequence $X_K : \Omega_K \to \mathbb{D}$ is asymptotically tight if for every $\epsilon > 0$ there exists a compact set S such that*

$$\liminf_{K \to \infty} P_* \left(X_K \in S^\delta\right) \geq 1 - \epsilon,$$

for every $\delta > 0$, where $S^\delta = \{y \in \mathbb{D} : d(y, S) < \delta\}$.

Denote by $l^\infty[a,b]$ the metric space of uniformly bounded, real functions on $[a,b]$, equipped with the supremum metric. A function $f : [a,b] \to \mathbb{R}$ is uniformly bounded if

$$\|f\|_{[a,b]} \equiv \sup_{t\in[a,b]} |f(t)| < \infty.$$

Further, $\mathcal{D}[a,b]$ is the metric space of all right continuous functions on $[a,b]$ with left hand limits, equipped with the supremum metric.
Note that $\mathcal{D}[a,b] \subset l^\infty[a,b]$ (van der Vaart, 1998, p.261).
The following theorem establishes that we can consider the weak convergence of a sequence of maps with values in $\mathcal{D}[a,b]$ relative to $\mathcal{D}[a,b]$, but also relative to $l^\infty[a,b]$; this does not make a difference as long as we use the supremum metric for the two spaces (van der Vaart, 1998).

Theorem 5.6.5. *Let $\mathbb{D}_0 \subset \mathbb{D}$ be arbitrary metric spaces equipped with the same metric. If X and every X_K take their values in $\mathbb{D}_0$, then the sequence of maps X_K converges weakly to X as maps in $\mathbb{D}_0$ if and only if the sequence of maps X_K converges weakly to X as maps in $\mathbb{D}$.*

Proof. The proof is presented in van der Vaart (1998, p.261). □

Hence, we may focus on weak convergence in the space $l^\infty[a,b]$, and automatically obtain the weak convergence in $\mathcal{D}[a,b]$. One of the sufficient conditions of weak convergence in $l^\infty[a,b]$ is the weak convergence of the finite dimensional distributions:

Definition 5.6.4. *A sequence of elements $X_K \in l^\infty[a,b]$ converges finite dimensionally to X if for all finite subsets $\{t_1,\dots,t_l\} \subset [a,b]$ the sequence of random vectors $\{X_K(t_1),\dots,X_K(t_l)\}$ converges in distribution to $\{X(t_1),\dots, X(t_l)\}$. The random vectors $\{X_K(t_1),\dots,X_K(t_l)\}$ are also called the marginals.*

Weak convergence in $l^\infty[a,b]$ may be characterised as asymptotic tightness plus convergence of the finite dimensional distributions (i.e., the marginals):

Theorem 5.6.6. *Let $X_K : \Omega_K \to l^\infty[a,b]$ be arbitrary.*

a) X_K converges weakly to a tight limit if and only if X_K is asymptotically tight and the marginals $\{X_K(t_1), \ldots, X_K(t_l)\}$ converge weakly to a limit for every subset $t_1, \ldots, t_l$ of $[a,b]$.

b) If X_K is asymptotically tight and its marginals converge weakly to the marginals $\{X(t_1), \ldots, X(t_l)\}$ of a stochastic process X, then there is a version of X with uniformly bounded sample paths and X_K converges weakly to X.

Proof. The proof is presented in van der Vaart and Wellner (1996, p. 35). □

The following theorem can be used to prove asymptotic tightness:

Theorem 5.6.7. *A sequence $X_K : \Omega_K \to l^\infty[a,b]$ is asymptotically tight if and only if $X_K(t)$ is asymptotically tight in $\mathbb{R}$ for every t and, for all $\epsilon, \eta > 0$, there exists a finite partition $[a,b] = \bigcup_{l=1}^{L} T_l$, such that*

$$\limsup_{K\to\infty} P^* \left\{ \sup_{1\le l\le L} \sup_{s,t\in T_l} |X_K(s) - X_K(t)| > \epsilon \right\} < \eta.$$

Proof. The proof is presented in van der Vaart and Wellner (1996, p.36). □

Chapter 6

Two-stage estimation and goodness-of-fit for copulas: Udder infection example

6.1 Introduction

In Sections 5.3 and 5.4 we explained how four-dimensional copula models can be fitted using a semi-parametric and a nonparametric two-stage estimation approach. We developed the asymptotic properties of both estimation procedures. In this chapter, we consider Clayton, positive stable and inverse Gaussian copulas and we consider the power variance family of copulas to model the association between the outcomes in a cluster. As described in Section 2.5.2, the Clayton copula, the positive stable copula and the inverse Gaussian copula are nested in the power variance copula family. Recall that each of these three copulas models a different type of dependence. The Clayton copula models late dependence in time, the positive stable copula models early dependence in time and the inverse Gaussian copula takes a position in between. In Section 6.2, we explain how a pseudo likelihood ratio test can be used to select a copula, nested in the power variance copula family, that describes the appropriate type of dependence between the outcomes in a cluster. The concept of

such a pseudo likelihood ratio test was also proposed by Glidden (2000) and Andersen (2005). Power variance copulas become Clayton copulas for $\nu \to 1$ and positive stable copulas for $\theta \to \infty$. This means that the choice of ν, resp. θ, that corresponds with Clayton copulas, resp. positive stable copulas, is at the boundary of the parameter space of the power variance copula family. Therefore, as discussed in Andersen (2005), the classical likelihood ratio test asymptotics are not valid. Andersen (2005) gives the asymptotic distribution for the pseudo likelihood ratio test without a rigorous proof. We therefore propose in Section 6.2 a new bootstrap algorithm to obtain the p-value for this test. Using this bootstrap algorithm, we also obtain estimates for the standard errors of the estimated parameters in the copula.
In Sections 6.3 and 6.5 we illustrate the semi-parametric and the nonparametric two-stage estimation approach by analysing the udder infection data, introduced in Section 1.4.1. We apply the method proposed in Section 6.2 to test the null hypotheses: Clayton copula (late dependence in time), positive stable copula (early dependence in time) and inverse Gaussian copula for the infection times of the four udder quarters of a cow. This allows us to select a copula within the power variance copula family that describes well the dependence between the infection times of the four udder quarters of a cow. In Section 6.4 we study the type I error rate and the power of the pseudo likelihood ratio test based on a simulation study in a setting similar to the analysis of the udder infection data in Section 6.3. The results of the simulations give evidence that the pseudo likelihood ratio test provides a valid approach to select an appropriate copula within the power variance copula family.

6.2 Goodness-of-fit

To choose a copula model that describes the right type of dependence between the outcomes in a cluster, we start with the power variance copula family. We then use a pseudo likelihood ratio test to assess whether the observations show early dependence (positive stable copula), late dependence (Clayton copula) or whether the dependence is not explicitly early or late in time (inverse Gaus-

sian copula).
It is also important to check whether the semi-parametric and the nonparametric marginal survival functions, given by (5.11) and (5.18), can be used in the first step of the estimation. If the marginal survival functions are misspecified in the first step, the obtained estimators are not reliable to be used in the second step of the estimation (see Kim *et al.*, 2007, for a discussion in the case of complete data). In Section 6.3, we illustrate how model checking for the marginal survival functions can be done using a descriptive approach. A formal approach to check the adequacy of the marginal Cox model in the semi-parametric approach is based on the asymptotic distribution of the score process. This approach is described in Spiekerman and Lin (1996) and in Martinussen and Scheike (2006, p.318-319).

6.2.1 Pseudo likelihood ratio test

A pseudo likelihood ratio test is used to select a copula in the power variance copula family that models the type of dependence between the outcomes in a cluster. Recall from Section 2.5.2 that the parameter space of the power variance copula is given by

$$\mathcal{C}_{PVF} = \{(\nu, \theta) : 0 \leq \nu \leq 1, 0 \leq \theta < \infty \text{ or } 0 \leq \nu < 1, \theta \to \infty\},$$

where the notation $\theta \to \infty$ means that θ tends to infinity. For specific choices of the parameters ν and θ, we obtain the Clayton copula, the positive stable copula and the inverse Gaussian copula. These parameter choices are given by

$$\begin{aligned} \mathcal{C}_C &= \{(\nu, \theta) : \nu \to 1, 0 \leq \theta < \infty\}, \\ \mathcal{C}_{IG} &= \{(\nu, \theta) : \nu = 0.5, 0 \leq \theta < \infty\}, \\ \mathcal{C}_{PS} &= \{(\nu, \theta) : 0 \leq \nu < 1, \theta \to \infty\}. \end{aligned}$$

To test if the Clayton copula can be used to analyse the data, the following hypotheses testing problem is considered:

$$\begin{aligned} H_0 &: (\nu, \theta) \in \mathcal{C}_C \quad \text{(Clayton copula)} \\ H_1 &: (\nu, \theta) \in \mathcal{C}_{PVF} \setminus \mathcal{C}_C. \end{aligned} \tag{6.1}$$

To test if the inverse Gaussian copula describes the association between the outcomes in a cluster, we consider

$$\begin{aligned} H_0 &: (\nu, \theta) \in \mathcal{C}_{IG} \quad \text{(inverse Gaussian copula)} \\ H_1 &: (\nu, \theta) \in \mathcal{C}_{PVF} \setminus \mathcal{C}_{IG}. \end{aligned} \tag{6.2}$$

Let $\log L_{P,H_0}(\theta)$ be the pseudo loglikelihood, as defined in Section 5.2, under the null hypothesis where the estimates from the first stage are plugged in and let $\hat{\theta}_0$ be the maximiser of $\log L_{P,H_0}(\theta)$. Let $(\hat{\nu}, \hat{\theta})$ denote the maximiser of $\log L_P(\nu, \theta)$. The test statistic is then given by $W_{obs} = -2\{\log L_{P,H_0}(\hat{\theta}_0) - \log L_P(\hat{\nu}, \hat{\theta})\}$.
The following hypotheses testing problem is used to test if the positive stable copula is appropriate for the data:

$$\begin{aligned} H_0 &: (\nu, \theta) \in \mathcal{C}_{PS} \quad \text{(positive stable copula)} \\ H_1 &: (\nu, \theta) \in \mathcal{C}_{PVF} \setminus \mathcal{C}_{PS}. \end{aligned} \tag{6.3}$$

Denote the pseudo loglikelihood under the null hypothesis by $\log L_{P,H_0}(\nu)$ and let $\hat{\nu}_0$ be the maximiser of $\log L_{P,H_0}(\nu)$. The test statistic is given by $W_{obs} = -2\{\log L_{P,H_0}(\hat{\nu}_0) - \log L_P(\hat{\nu}, \hat{\theta})\}$.
This pseudo likelihood ratio test is proposed by Glidden (2000) and Andersen (2005). Proposition 3.4 in Andersen (2005) states that the asymptotic distribution of the test statistic W_{obs} that corresponds to (6.2), is a constant times a chi-square distribution with one degree of freedom. Since the Clayton copula and the positive stable copula are on the boundary of the parameter space, this asymptotic result does not hold for the test statistic corresponding to hypotheses (6.1), resp. (6.3). Andersen (2005) claims that, when the parameters are on the boundary of the parameter space, the asymptotic distribution of the test statistic W_{obs} is a constant times a $50 : 50$ mixture of a chi-square distribution with zero degrees of freedom and a chi-square distribution with

one degree of freedom. A rigid proof of this result is not yet available. Therefore, we propose a new bootstrap algorithm to obtain the p-value for this test. Using this bootstrap algorithm, we can also obtain estimates for the standard errors of the estimated parameters in the copula. In the bootstrap algorithm we assume that all observations in the same cluster have a common censoring time, i.e. $C_{ij} = C_i$ for $i = 1, \ldots, K$ and $j = 1, \ldots, 4$. This is the type of censoring in the udder infection data.

6.2.2 Bootstrap algorithm

1. Estimate $\boldsymbol{\zeta}$ (under H_0) using two-stage estimation.

2. Generate quadruples $(V_{i1}^*, V_{i2}^*, V_{i3}^*, V_{i4}^*)$, $i = 1, \ldots, K$, from the null model $C_{\hat{\boldsymbol{\zeta}}}$. The true failure times $(T_{i1}^*, T_{i2}^*, T_{i3}^*, T_{i4}^*)$ can be obtained from $(V_{i1}^*, V_{i2}^*, V_{i3}^*, V_{i4}^*)$ by using the expression for the survival function that corresponds to the marginal model that is used in the first stage.

3. Estimate the censoring distribution G with the Kaplan-Meier estimator based on the K observations of $(\tilde{X}_i = C_i \wedge \tilde{T}_i, I(C_i < \tilde{T}_i))$, where $\tilde{T}_i = \max(T_{i1}, T_{i2}, T_{i3}, T_{i4})$ now plays the role of the censoring variable for C_i. Generate C_i^*, $i = 1, \ldots, K$, from $\hat{G}$.

4. Set $X_{ij}^* = \min(T_{ij}^*, C_i^*)$ and $\delta_{ij}^* = I(X_{ij}^* = T_{ij}^*)$.

5. Estimate $\boldsymbol{\zeta}^*$ using the two-stage estimation procedure (under H_0 and H_1) and obtain the value of the test statistic W_r^* for the data $(X_{ij}^*, \delta_{ij}^*)$.

6. Step 2-5 is repeated R times to obtain an estimate of the null distribution of the test statistic and the p-value is

$$p = \sum_{r=1}^{R} I(W_r^* > W_{obs})/R,$$

where W_{obs} is the value of the test statistic for the original data (X_{ij}, δ_{ij}).

6.3 Analysis of the udder infection data

To investigate the dependence between the infection times of the four udder quarters of a cow, we fit a Clayton copula, a positive stable, an inverse Gaussian and a power variance copula using two-stage estimation. In the first stage, the marginal survival functions are estimated using the semi-parametric and the nonparametric approach discussed in Section 5.3.2 and Section 5.4.2. In the marginal survival functions we model the effect of the location of the udder quarters (front or rear). We assume that the marginal survival functions that correspond to the two front udder quarters, resp. the two rear udder quarters, are identical. The observations of the four udder quarters of each cow can be ordered in the following way: front left, front right, rear left, rear right. The values of the udder quarter specific covariate indicating the front or rear position of an udder quarter of cow i are then given by $z_{i1} = z_{i2} = 0$ (front) and $z_{i3} = z_{i4} = 1$ (rear).

6.3.1 Semi-parametric two-stage estimation approach

In the semi-parametric approach, the marginal survival functions are estimated as follows

$$\begin{aligned} \hat{S}_{i1}(t) &= \hat{S}_{i2}(t) = \exp\left\{-\hat{\Lambda}_0(t;\hat{\beta})\right\}, \\ \hat{S}_{i3}(t) &= \hat{S}_{i4}(t) = \exp\left\{-\hat{\Lambda}_0(t;\hat{\beta})\exp(\hat{\beta})\right\}. \end{aligned} \tag{6.4}$$

It follows from Theorem 5.3.1 and Theorem 5.3.3 that $\hat{\beta}$ and $\hat{\Lambda}_0(.;\hat{\beta})$ are consistent estimators, which therefore can be used in the second step of the estimation procedure. In the second step we estimate, for each copula model, the association parameter $\boldsymbol{\zeta}$ by maximising the pseudo loglikelihood in (5.17) with respect to $\boldsymbol{\zeta}$. Theorem 5.3.5 and Theorem 5.3.6 establish that $\hat{\boldsymbol{\zeta}}$, the estimator of the association parameter (vector), is consistent and that its limit distribution is normal.

Figure 6.1 shows the estimated marginal survival functions obtained by (6.4) and by using a Nelson-Aalen estimator for the cumulative hazard function of the pooled infection times of the front udder quarters, resp. the rear udder

quarters, as proposed in (5.18). This figure suggests that the marginal Cox model fits the udder infection data well.

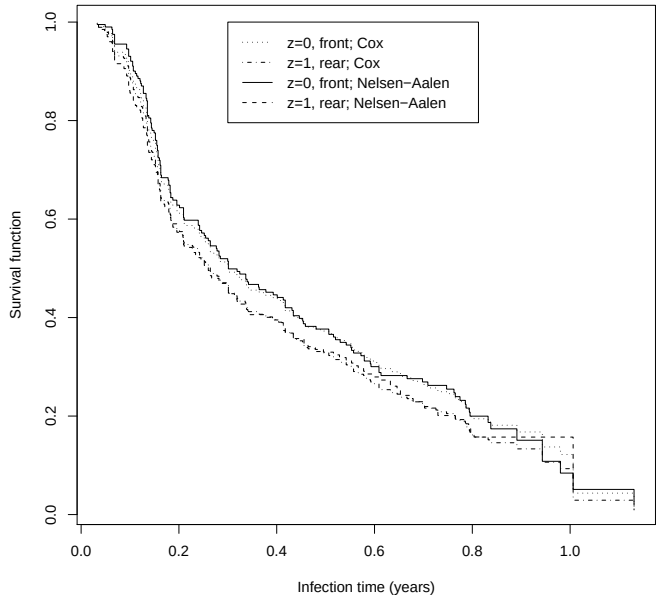

Figure 6.1: Udder infection data. Estimated marginal survival functions using a marginal Cox model with binary covariate "location" (front or rear) and using a Nelson-Aalen estimator for the pooled data with $z_{ij} = 0$, resp. $z_{ij} = 1$.

6.3.2 Nonparametric two-stage estimation approach

Recall from (5.18) that, in the nonparametric approach, the marginal survival functions are estimated by using a Nelson-Aalen estimator for the group with $z_{ij} = 0$ and $z_{ij} = 1$ separately:

$$\hat{S}_{ij}(t) = \begin{cases} \exp\{-\hat{\Lambda}_1(t)\} & \text{if } z_{ij} = 0 \\ \exp\{-\hat{\Lambda}_2(t)\} & \text{if } z_{ij} = 1. \end{cases} \tag{6.5}$$

By estimating the marginal survival functions in this way, we take into account that the teat end condition might be different for the front and the rear udder quarters, but we do not assume any functional form in the estimation of the marginal survival functions. Note that the udder quarters within the group of the front, resp. rear, udder quarters are not independent. However, we proved in Theorem 5.4.1 that $\hat{\Lambda}_1$, resp. $\hat{\Lambda}_2$, are uniformly consistent in $[0, \tau_1]$, resp. $[0, \tau_2]$. Therefore, the marginal survival functions given by (6.5) can be plugged into the pseudo loglikelihood (5.54). By maximising (5.54) with respect to $\boldsymbol{\zeta}$ for each copula model, we obtain an estimate for the association parameter (vector) in the copula model. Theorem 5.4.3 and Theorem 5.4.4 state that $\hat{\boldsymbol{\zeta}}$ is consistent and that its limit distribution is normal.

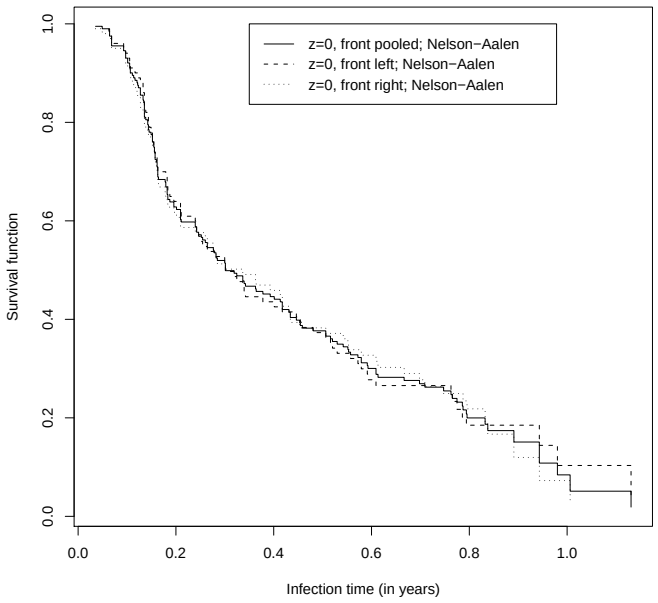

Figure 6.2: Udder infection data. Estimated marginal survival functions using a Nelson-Aalen estimator for the pooled front udder quarters data with $z_{i1} = 0$ and $z_{i2} = 0$ and using a Nelson-Aalen estimator for the data with $z_{i1} = 0$, resp. $z_{i2} = 0$.

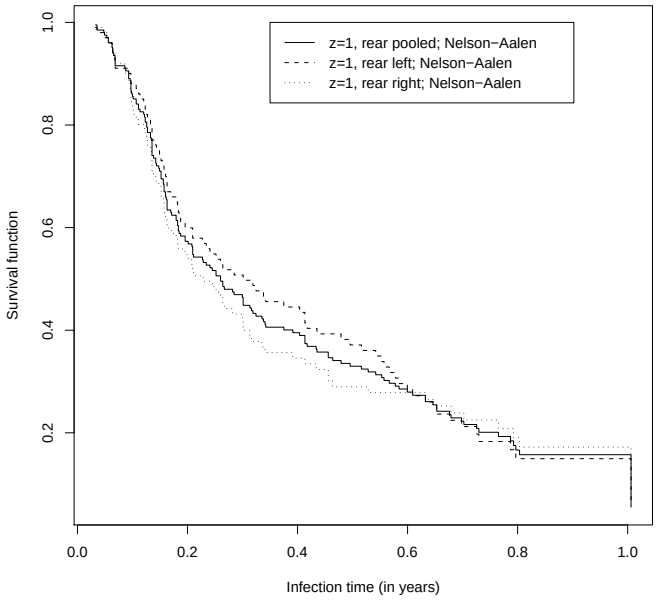

Figure 6.3: Udder infection data. Estimated marginal survival functions using a Nelson-Aalen estimator for the pooled rear udder quarters data with $z_{i3} = 1$ and $z_{i4} = 1$ and using a Nelson-Aalen estimator for the data with $z_{i3} = 1$, resp. $z_{i4} = 1$.

Figure 6.2 compares the estimated marginal survival function for the pooled infection times of the front udder quarters obtained by (6.5), with the two estimated marginal survival functions obtained by using a Nelson-Aalen estimator for the cumulative hazard functions of the infection times of the front left and the front right udder quarters separately. Since we assume in the nonparametric approach that the marginal survival functions for the front udder quarters are identical, we expect that the three survival curves are close. Figure 6.3 provides similar information for the rear udder quarters. The three estimated survival curves in Figure 6.2 almost coincide; the estimated survival curves in Figure 6.3 are a bit more apart for infection times between between 0.3 and 0.6 years. Overall, these figures suggest that the nonparametric approach can

be used to estimate the marginal survival functions in the first step.

6.3.3 Pseudo likelihood ratio test

To select which copula model, nested in the power variance family, is appropriate to model the dependence in the udder infection data, we perform pseudo likelihood ratio tests for the hypotheses (6.1), (6.2) and (6.3). The p-value for each test and the estimated standard errors of the estimated association parameters are obtained by taking $R = 1000$ bootstrap samples using the bootstrap algorithm in Section 6.2. To generate quadruples $(V_{i1}^*, V_{i2}^*, V_{i3}^*, V_{i4}^*)$ in the second step of the bootstrap algorithm, the algorithm provided by Embrechts *et al.* (2003) is used to generate from a Clayton or an inverse Gaussian copula, whereas the algorithm provided by Shih and Louis (1995a) is used to generate from a positive stable copula.
In the semi-parametric approach, the marginals are estimated through a Cox proportional hazards model with binary covariate "location", as explained in Section 6.3.1. The estimated parameters, the corresponding standard error estimates and the p-values for the pseudo likelihood ratio tests are presented in Table 6.1.

Table 6.1: Results for the udder infection data, marginal Cox model with udder location (front or rear) as binary covariate; 1000 bootstrap samples.

Copula	Parameter	Estimate	Std. error	P-value
	β	0.1207	0.0615	
Clayton	θ	1.4429	0.2503	0.048
Positive stable	ν	0.4049	0.0444	0.007
Inverse Gaussian	θ	8.4867	4.5790	0.346

In the nonparametric approach, the marginal survival functions are estimated as discussed in Section 6.3.2. Table 6.2 shows the estimated association parameters for the different copula models, the corresponding estimated standard

errors and the p-values for the pseudo likelihood ratio tests.
For both the semi-parametric and for the nonparametric approach, we conclude that, within the power variance copula family, the inverse Gaussian copula model is an appropriate model to describe the association between the infection times of the four udder quarters. This indicates that the dependence is not early or late but intermediate in time.

Table 6.2: Results for the udder infection data, marginal survival functions estimated by Nelson-Aalen estimator for pooled front, resp. rear, udder quarters data; 1000 bootstrap samples.

Copula	Parameter	Estimate	Std. error	P-value
Clayton	θ	1.4079	0.2520	0.012
Positive stable	ν	0.3976	0.0447	0.002
Inverse Gaussian	θ	7.9644	4.3509	0.402

6.4 Simulations

We perform simulations to evaluate the type I error rate and the power of the pseudo likelihood ratio test in a setting similar to the analysis of the udder infection data, discussed in Section 6.3. We generate 500 data sets with 100 clusters that contain four observations each. The observations for each data set are generated in the following way. First, 100 quadruples $(V_{i1}, V_{i2}, V_{i3}, V_{i4})$ are generated from an inverse Gaussian copula using the algorithm provided by Embrechts *et al.* (2003). The inverse Gaussian copula in (2.15) is often written in terms of a parameter $\alpha = 1/\theta$. For the udder infection data, we obtain $\hat{\alpha} = 0.1178$ ($\hat{\theta} = 8.4867$) using the semi-parametric approach and $\hat{\alpha} = 0.1256$ ($\hat{\theta} = 7.9644$) using the nonparametric approach, see Table 6.1 and Table 6.2. In the simulations, we generate 100 quadruples $(V_{i1}, V_{i2}, V_{i3}, V_{i4})$ from an inverse Gaussian copula with $\alpha = 0.12$; this corresponds to $\theta = 8.33$. To obtain the failure times T_{ij} we assume that the marginal survival functions

come from a Weibull regression model with a binary covariate ($z_{i1} = z_{i2} = 0$, $z_{i3} = z_{i4} = 1$), i.e.,

$$V_{ij} = S_{ij}(T_{ij}) = \exp\left\{-\Lambda_0(T_{ij})\exp(z_{ij}\beta)\right\},$$

with $\Lambda_0(T_{ij}) = \lambda T_{ij}^{\rho}$. The jth failure time in the ith cluster is then obtained as follows:

$$T_{ij} = \left\{\frac{-\log(V_{ij})}{\lambda\exp(z_{ij}\beta)}\right\}^{1/\rho},$$

for $j = 1, \ldots, 4$. The parameters λ, ρ and β are chosen to resemble the udder infection data (estimated values based on infection time in years): $\lambda = 2.3048$, $\rho = 1.1904$, $\beta = 0.1141$. A common censoring time C_i for the observations of each cluster is generated from a uniform distribution on $[0.1, 0.5]$ with probability 17% or from a normal distribution with mean 0.78 and variance 0.01 with probability 83%. We obtain approximately 20% censoring. We choose this censoring distribution since it mimics the distribution of the censoring times in the udder infection data. To illustrate this, Figure 6.4 shows an histogram of the censoring times based on 1000 values generated from the Kaplan-Meier estimator of the censoring distribution (obtained as explained in step 3 of the bootstrap algorithm proposed in Section 6.2.2), resp. from the specific mixture of a uniform and a normal distribution described in this section.

For each data set, a Clayton copula, a positive stable copula, an inverse Gaussian copula and a power variance copula is fitted using the semi-parametric and the nonparametric approach proposed in Sections 5.3 and 5.4. We perform a pseudo likelihood ratio test to test which copula model, nested in the power variance copula family, is appropriate for the simulated data. To obtain the p-value for this test, we take 500 bootstrap samples.

In the discussion that follows, we use a 5% significance level. To compute the type I error rate for the pseudo likelihood ratio test, we consider the hypotheses in (6.2). Based on 500 data sets, the probability of a type I error is 0.082, with a standard error of 0.012, if the marginal survival functions are estimated through a Cox model. The probability of a type I error is 0.056, with a standard error of 0.010, if the nonparametric approach is used in the

first stage of the estimation.

Since the generated data come from an inverse Gaussian copula, the hypotheses in (6.1) and (6.3) are considered for computing the power of the test. Based on 500 data sets, the power of the pseudo likelihood ratio test used to test (6.1), is 0.988, with a standard error of 0.005, using the semi-parametric approach in the first stage. Using the nonparametric approach, the power for this test is 0.992, with a standard eror of 0.004. For the test of (6.3), the power is 0.844, with a standard error of 0.016, using the semi-parametric approach. Using the nonparametric approach, the power is 0.802, with a standard error of 0.018. In terms of the type I error rate of the pseudo likelihood ratio test, the nonparametric approach is preferred in this setting, whereas both are comparable in terms of the power of the test.

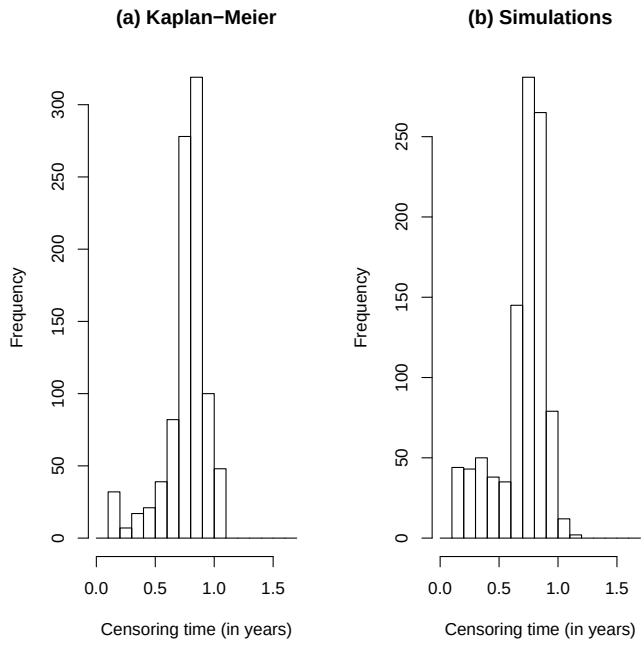

Figure 6.4: Histogram of censoring times based on 1000 values generated from (a) Kaplan-Meier estimator of censoring distribution for udder infection data (b) mixture of uniform and normal distribution.

6.5 Analysis of the udder infection data: continued

In Section 6.3 we accounted for the effect of the location of the udder quarters (front or rear) in the estimation of the marginal survival functions. In this section we denote this covariate by $z_{ij1} \equiv z_{ij}$, where z_{ij} is as defined in Section 6.3, i.e., $z_{i11} = z_{i21} = 0$ (front) and $z_{i31} = z_{i41} = 1$ (rear). Recall from Section 1.4.1 that also the parity of a cow might have an effect on the incidence of intramammary infections (Weller *et al.*, 1992). As already mentioned there, we convert parity into a binary covariate "heifer". Note that "heifer" is a covariate at the cluster level; therefore we use the notation $z_{ij2} = z_{i2}$, for $j = 1, \ldots, 4$ and $i = 1, \ldots, 100$. We define z_{i2} such that $z_{i2} = 0$ if cow i has experienced more than one calving (i.e., cow i is a multiparous cow) and $z_{i2} = 1$ if cow i has experienced only one calving (i.e., cow i is a primiparous cow or a heifer).
We fit a Clayton copula, a positive stable copula and an inverse Gaussian copula to study the dependence between the infection times of the four udder quarters of a cow. To fit these copula models, we use the semi-parametric two-stage estimation procedure, proposed in Section 5.3.

6.5.1 Semi-parametric two-stage estimation approach

To model the effect of the location of the udder quarter (front or rear) and the parity in the first step of the estimation procedure, we estimate the marginal survival functions as follows

$$\hat{S}_{ij}(t) = \exp\left\{-\hat{\Lambda}_0(t)\exp(\hat{\beta}_1 z_{ij1} + \hat{\beta}_2 z_{i2})\right\}, \tag{6.6}$$

for $j = 1, \ldots, 4$ and $i = 1, \ldots, 100$. It follows from Theorem 5.3.1 and Theorem 5.3.3 that $\hat{\boldsymbol{\beta}} = (\hat{\beta}_1, \hat{\beta}_2)'$ and $\Lambda_0(.; \hat{\boldsymbol{\beta}})$ are consistent estimators which therefore can be used in the second step of the estimation. In the second step, we estimate the association parameter (vector) $\boldsymbol{\zeta}$ by maximising the pseudo loglikelihood (5.17), where $\hat{S}_{ij}(t)$ is given by (6.6). The consistency and the asymptotic normality of the obtained estimator $\hat{\boldsymbol{\zeta}}$ follow from Theorem 5.3.5 and Theorem 5.3.6.

Figure 6.5 shows the estimated marginal survival function, obtained by (6.6), for the infection times of the front udder quarters of multiparous cows, resp. heifers. These two curves are compared with the two estimated marginal survival functions obtained using Nelson-Aalen estimators for the cumulative hazard functions of the front udder quarters data of multiparous cows and heifers. Note that we assume that the marginal survival functions of the infection times of the front udder quarters of cows with the same parity are identical. Figure 6.6 gives similar information for the rear udder quarters. Figures 6.5 and 6.6 suggest that the proportional hazards assumption of the Cox model used in (6.6) holds.

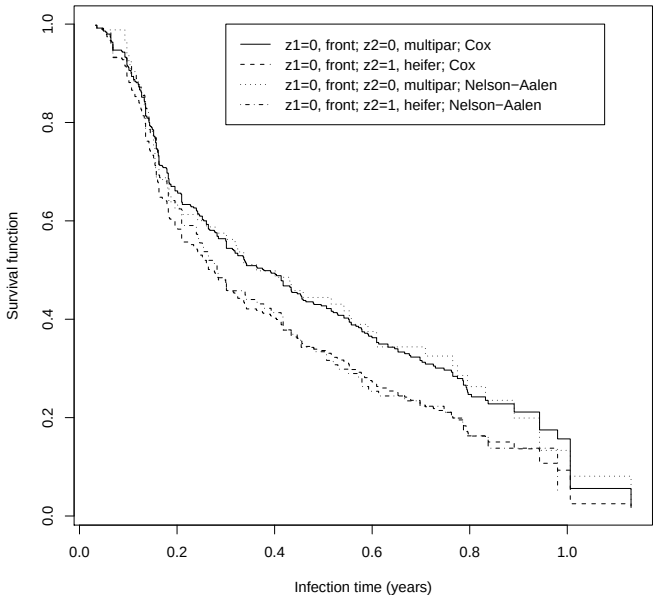

Figure 6.5: Udder infection data. Estimated marginal survival functions using a Cox model with binary covariates "location" (front or rear) and "heifer" and using a Nelson-Aalen estimator for the pooled data with $z_{ij1} = 0$ and $z_{i2} = 0$, resp. $z_{ij1} = 0$ and $z_{i2} = 1$.

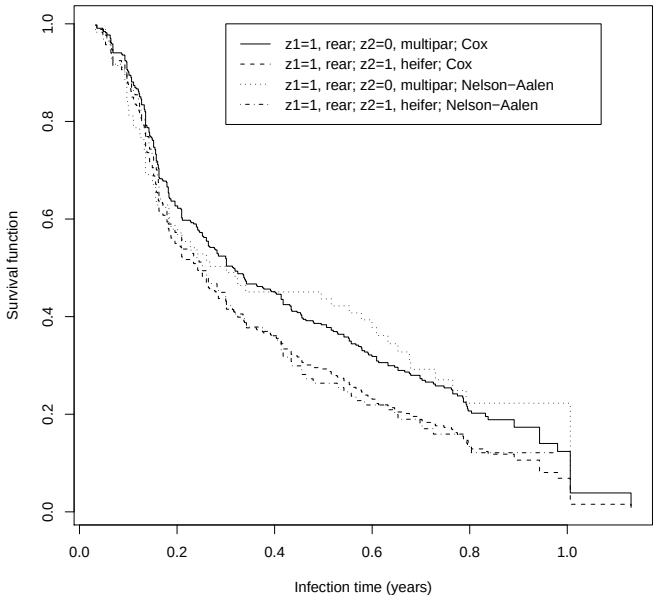

Figure 6.6: Udder infection data. Estimated marginal survival functions using a Cox model with binary covariates "location" (front or rear) and "heifer" and using a Nelson-Aalen estimator for the pooled data with $z_{ij1} = 1$ and $z_{i2} = 0$, resp. $z_{ij1} = 1$ and $z_{i2} = 1$.

6.5.2 Pseudo likelihood ratio test

In a similar way as explained in Section 6.3.3, we perform a pseudo likelihood ratio test to determine which copula, nested in the power variance family, describes the type of dependence between the infection times of the four udder quarters of a cow: the Clayton copula, the positive stable copula or the inverse Gaussian copula. Table 6.3 shows the estimate for the fixed effect $\boldsymbol{\beta}$, the estimated association parameters for the different copula models, the corresponding estimated standard errors and the p-values for the pseudo likelihood ratio tests. The results indicate that the inverse Gaussian copula model is an appropriate model within the power variance copula family to model the dependence between the infection times of the four udder quarters. This

indicates that the dependence is intermediate in time.

Table 6.3: Results for the udder infection data, marginal Cox model with udder location (front or rear) and heifer as binary covariates; 1000 bootstrap samples.

Copula	Parameter	Estimate	Std. error	P-value
	β_1	0.1199	0.0601	
	β_2	0.2476	0.1958	
Clayton	θ	1.4478	0.2424	0.0500
Positive stable	ν	0.4020	0.0441	0.0000
Inverse Gaussian	θ	8.0402	4.5269	0.2760

6.6 Conclusions

In this chapter, we illustrate the use of copulas to model the dependence between the infection times of the four udder quarters of cows. The copula model determines the type of association between the udder quarters. In the marginal survival functions we model the effect of the location of the udder quarter (front or rear) and/or the parity of the cow. Both a semi-parametric and a nonparametric approach are considered to estimate the marginal survival functions.

To select an appropriate copula from the power variance copula family, a pseudo likelihood ratio test is used. We propose a bootstrap algorithm that can be used to obtain a p-value for this test. In this bootstrap algorithm, we assume that all observations in the same cluster have a common censoring time. This is the type of censoring in the udder infection data. We can extend this algorithm to the more general situation where observations in the same cluster have different censoring times. For instance, if the censoring distribution is independent of the covariates, we can use step four of the second-model

based resampling plan, proposed in Section 3.2, to generate censoring times. Andersen *et al.* (2005) propose a class of goodness-of-fit tests for copula models based on bivariate right censored data. In these tests, a nonparametric estimate of the copula is compared to an estimate obtained by assuming a parametric family of copula models. To generalise these results to higher-dimensional failure time data, we need a nonparametric estimator for the copula that accounts for the presence of censoring in the data. This is not straightforward in general. If all observations in a cluster have a common censoring time, the nonparametric estimator proposed by van der Laan *et al.* (2002) can be used. These problems are subjects for further research.

Chapter 7

Conclusions and further research

Frailty models and copula models are often used to model clustered (or multivariate) survival data. In this thesis, we propose estimation methods and resampling procedures for frailty models and copula models. The developed methods are illustrated using examples from clinical trials and veterinary studies on dairy cows.

In Chapters 3 and 4 we consider frailty models for correlated survival data with a varying cluster size. In Chapters 5 and 6 we study copula models for four-dimensional survival data.

In Chapter 3 we propose two model-based resampling schemes that can be used to estimate the standard errors of the parameter estimates in a shared frailty model. Based on a simulation study, the performance of the proposed algorithms is compared with the performance of an existing nonparametric resampling plan (Therneau and Grambsch, 2000). The results indicate that the first model-based resampling plan, based on resampling of the estimated frailties, underestimates the empirical variability of the parameter estimates for almost all settings studied. The second model-based resampling plan, based on resampling from the estimated frailty distribution, provides in general precise assessment of the empirical variability of the parameter estimates, even if the

model is misspecified. However, the empirical variability of the heterogeneity parameter estimate can be rather different for the correct and the misspecified models. So, the results indicate that robustness in terms of the heterogeneity parameter is an issue of concern for the bootstrap algorithms (including the nonparametric bootstrap). In terms of the fixed effects, the bootstrap algorithms are robust against misspecification of the frailty distribution. The results clearly illustrate that further research on diagnostic tests for the choice of the frailty distribution is needed. More details on topics for further research are provided in Section 3.5.

In Chapter 4 we study a frailty model that extends the shared frailty model by accounting for a random treatment effect. Classical estimation methods to fit (shared) frailty models are likelihood-based. We propose an alternative estimation method which is based on a model transformation. We show that the integral of the weighted (over time) conditional cumulative loghazard depends in a linear way on the random effects describing the cluster and the treatment effect over clusters. Using the data within a cluster, we can estimate the integral using a nonparametric estimator for the cumulative hazard function. Considering this estimated integral as a response, we obtain a linear mixed-effects model. Through this model transformation, the parameters of interest in a frailty model become parameters in a linear mixed-effects model. Hence, linear mixed models methodology can be applied to estimate these parameters. We further demonstrate that model transformation also works for other conditional survival models, such as the multivariate proportional odds model and the multivariate additive risks model. Most software packages contain procedures to fit linear mixed-effects models but provide only a limited number of procedures to fit conditional (random effects) survival models. The proposed estimation method therefore is a useful approach to get insight in the heterogeneity in clustered data. To study the performance of the proposed method, we focus on frailty models. The results of a simulation study illustrate that the idea of model transformation provides a good and simple alternative for fitting frailty models for data sets with a sufficiently large number of clusters (i.e., $K = 20$) and moderate to large sample sizes within covariate

level subgroups in the clusters (i.e., at least $n_{ij} = 50$). We consider a frailty model with a binary covariate. Extending the proposed transformation idea to frailty models with a continuous covariate is an interesting topic that is currently under investigation by Cao *et al.* (2008). Further, it is also of interest to extend the proposed method to other censoring schemes than right censoring. More references related to these topics for further research are provided in Section 4.6.

In Chapters 5 and 6 we study copula models for four-dimensional survival data. The motivating example for this is the udder infection data; a data set on the correlated infection times in the four udder quarters of dairy cows, which is introduced in Section 1.4.1. In the copula model the joint survival function of the four outcomes is modelled as a function, called the copula, of the marginal survival functions of the four outcomes. The copula determines the type of dependence between the four outcomes.

In Chapter 5 we propose a semi-parametric and a nonparametric two-stage estimation approach for four-dimensional copulas. In the first step of the estimation procedure, we estimate the marginal survival functions. In the marginal survival functions we model the effect of a deterministic binary covariate at the cluster level and a deterministic binary covariate at the observational unit level. In the second step of the estimation procedure, we estimate the copula parameter by maximising a loglikelihood function in which the marginal survival functions are replaced by their estimates obtained in the first step. We prove the consistency and the asymptotic normality of the estimators obtained in the first and the second step of the estimation method. In Section 5.5 we explain how our results contribute to what is already available in the statistical literature.

In Chapter 6 we apply the methodology, developed in Chapter 5, to the udder infection data. To choose a copula model that describes the dependence of the outcomes within a cluster, we start with a large parametric copula family, i.e., the power variance copula family. We consider three copula models that are nested in the power variance copula family: the Clayton copula, the positive stable copula and the inverse Gaussian copula. Each of these three

copulas model a different type of dependence. To assess which copula model, nested in the power variance copula family, gives a good description of the dependence between the outcomes in a cluster, we use a pseudo likelihood ratio test. We propose a bootstrap algorithm that can be used to obtain the p-value for this test. We study the type I error rate and the power of the pseudo likelihood ratio test based on a simulation study in a setting similar to the udder infection data. The results indicate that the pseudo likelihood ratio test is a valid approach to select an appropriate copula model nested in the power variance copula family. For bivariate right censored data, Andersen *et al.* (2005) propose a goodness-of-fit test which compares a nonparametric estimate of the copula to an estimate obtained by assuming a parametric family of copula models. An interesting topic for further research is to extend this test to copulas for four-dimensional survival data. A more detailed discussion of topics for further research related to Chapter 6 is given in Section 6.6.

Bibliography

Aalen, O. O. (1980) A model for non-parametric regression analysis of counting processes. In: *Lecture Notes in Statistics 2: Mathematical Statistics and Probability Theory* (Eds. W. Klonecki, A. Kozek and J. Rosinski), 1–25. New York: Springer-Verlag.

Aalen, O. O. (1988) Heterogeneity in survival analysis. *Statistics in Medicine*, **7**, 1121–1137.

Aalen, O. O. (1992) Modelling heterogeneity in survival analysis by the compoind Poisson distribution. *Annals of Applied Probability*, **2**, 951–972.

Adkinson, R. W., Ingawa, K. H., Blouin, D. C. and Nickerson, S. C. (1993) Distribution of clinical mastitis among quarters of the bovine udder. *Journal of Dairy Science*, **76**, 3453–3459.

Advanced Colorectal Cancer Meta-Analysis Project (1992) Modulation of 5-fluorouracil by leucovorin in patients with advanced colorectal cancer: evidence in terms of response rate. *Journal of Clinical Oncology*, **10**, 896–903.

Advanced Colorectal Cancer Meta-Analysis Project (1994) Meta-analysis of randomized trials testing the bio-chemical modulation of 5-fluorouracil by methotrexate in metastatic colorectal cancer. *Journal of Clinical Oncology*, **12**, 960–969.

Andersen, E. W. (2005) Two-stage estimation in copula models used in family studies. *Lifetime Data Analysis*, **11**, 333–350.

Andersen, P. K., Ekstrom, C. T., Klein, J. P., Shu, Y. and Zhang, M.-J. (2005) A class of goodness of fit tests for a copula based on bivariate right-censored data. *Biometrical Journal*, **47**, 815–824.

Andersen, P. K. and Gill, R. D. (1982) Cox's regresssion model for counting processes: A large sample study. *Annals of Statistics*, **10**, 1100–1120.

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

Beran, R. (1981) Nonparametric regression with randomly censored survival data. *Technical report*, University of California, Berkeley.

Burzykowski, T., Molenberghs, G. and Buyse, M. (2004) The validation of surrogate end points by using data from randomized clinical trials: a case-study in advanced colorectal cancer. *Journal of the Royal Statistical Society A*, **167**, 103–124.

Cao, R. and Gonzalez-Manteiga, W. (2008) Goodness-of-fit tests for conditional models under censoring and truncation. *Journal of Econometrics*, **143**, 166–190.

Cao, R., Janssen, P. and Lopez de Ullibarri, I. (2008) Fitting frailty models via transformation: the continuous case. *Technical report (in preparation)*.

Clahsen, P. C., van de Velde, C. J., Julien, J. P., Floiras, J. L., Delozier, T., Mignolet, F. Y. and Sahmoud, T. M. (1996) Improved local control and disease-free survival after perioperative chemotherapy for early-stage breast cancer: a European Organization for Research and Treatment of Cancer breast cancer cooperative group study. *Journal of Clinical Oncology*, **14**, 745–753.

Cortiñas Abrahantes, J. and Burzykowski, T. (2005) A version of the EM algorithm for proportional hazard model with random effects. *Biometrical Journal*, **47**, 847–862.

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

Davison, A. C. and Hinkley, D. V. (1997) *Bootstrap methods and their application.* Cambridge: Cambridge University Press.

De Vliegher, S., Barkema, H. W., Opsomer, G., de Kruif, A. and Duchateau, L. (2005) Association between somatic cell count in early lactation and culling of dairy heifers using Cox frailty models. *Journal of Dairy Science*, **88**, 560–568.

Duchateau, L. and Janssen, P. (2008) *The frailty model.* New York: Springer.

Duchateau, L., Janssen, P., Lindsey, P., Legrand, C., Nguti, R. and Sylvester, R. (2002) The shared frailty model and the power for heterogeneity tests in multicenter clinical trials. *Computational Statistics and Data Analysis*, **40**, 603–620.

Ducrocq, V. and Casella, G. (1996) A Bayesian analysis of mixed survival models. *Genetics Selection Evolution*, **28**, 505–529.

Embrechts, P., Lindskog, F. and McNeil, A. J. (2003) *Handbook of Heavy Tailed Distributions in Finance*, chap. Modelling dependence with copulas and applications to risk management, 329–384. Amsterdam: Elsevier.

Fleming, T. R. and Harrington, D. P. (1991) *Counting processes and survival analysis.* New York: John Wiley & Sons.

Genest, C., Ghoudi, K. and Rivest, L.-P. (1995) A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. *Biometrika*, **82**, 543–552.

Genest, C. and MacKay, J. (1986) Copules Archimédiennes et familles de lois bidimensionnelles dont les marges sont données. *Canadian Journal of Statistics*, **14**, 145–159.

Gijbels, I. and Wang, J. L. (1993) Strong representations of the survival function estimator for truncated and censored data with applications. *Journal of Multivariate Analysis*, **47**, 210–229.

Glidden, D. (2000) A two-stage estimator of the dependence parameter for the Clayton-Oakes model. *Lifetime Data Analysis*, **6**, 141–156.

Goethals, K., Janssen, P. and Duchateau, L. (2008) Frailty models and copulas: similarities and differences. *Journal of Applied Statistics (in press)*.

Grigoletto, M. and Akritas, M. G. (1999) Analysis of covariance with incomplete data via semiparametric model transformations. *Biometrics*, **55**, 1177–1187.

Gut, A. (2005) *Probability: A graduate course*. New York: Springer.

Henderson, C. R. (1984) *Applications of linear models in animal breeding*. Guelph, Canada: University of Guelph Press.

Hjort, N. L. (1985) Bootstrapping Cox's regression model. *Tech. Rep. NSF-241*, Department of Statistics, Stanford University.

Hougaard, P. (1986a) A class of multivariate failure time distributions. *Biometrika*, **73**, 671–678.

Hougaard, P. (1986b) Survival models for heterogeneous populations derived from stable distributions. *Biometrika*, **73**, 387–396.

Hougaard, P. (2000) *Analysis of multivariate survival data*. New York: Springer.

Kaplan, E. L. and Meier, P. (1958) Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, **53**, 457–481.

Kim, G., Silvapulle, M. J. and Silvapulle, P. (2007) Comparison of semiparametric and parametric methods for estimating copulas. *Computational Statistics and Data Analysis*, **51**, 2836–2850.

Klein, J. P. (1992) Semiparametric estimation of random effects using the Cox model based on the EM algorithm. *Biometrics*, **48**, 795–806.

Klein, J. P. and Moeschberger, M. L. (2003) *Survival Analysis. Techniques for censored and truncated data.* New York: Springer-Verlag.

van der Laan, M. J., Hubbard, A. E. and Robins, J. M. (2002) Locally efficient estimation of a multivariate survival function in longitudinal studies. *Journal of the American Statistical Association*, **97**, 494–507.

Lam, K. F. and Lee, Y. W. (2004) Merits of modelling multivariate survival data using random effects proportional odds models. *Biometrical Journal*, **46**, 331–342.

Lam, K. F., Lee, Y. W. and Leung, T. L. (2002) Modeling multivariate survival data by a semiparametric random effects proportional odds model. *Biometrics*, **58**, 316–323.

Legrand, C., Ducrocq, V., Janssen, P., Sylvester, R. and Duchateau, L. (2005) A Bayesian approach to jointly estimate center and treatment by center heterogeneity in a proportional hazards model. *Statistics in Medicine*, **24**, 3789–3804.

Lindsey, J. C. and Ryan, L. M. (1998) Tutorial in biostatistics: Methods for interval-censored data. *Statistics in Medicine*, **17**, 219–238.

Liu, L. and Yu, Z. (2008) A likelihood reformulation method in non-normal random effects models. *Statistics in Medicine*, **27**, 3105–3124.

Lo, S. H. and Singh, K. (1986) The product-limit estimator and the bootstrap: Some asymptotic representations. *Probability Theory and Related Fields*, **71**, 455–465.

Martinussen, S. and Scheike, T. H. (2006) *Dynamic regression models for survival data.* New York: Springer.

Massonnet, G., Burzykowski, T. and Janssen, P. (2006) Resampling plans for frailty models. *Communications in Statistics - Simulation and Computation*, **35**, 497–514.

Massonnet, G., Janssen, P. and Burzykowski, T. (2008a) Fitting conditional survival models to meta-analytic data by using a transformation toward mixed-effects models. *Biometrics*, **64**, 834–842.

Massonnet, G., Janssen, P. and Duchateau, L. (2008b) Modelling udder infection data using copula models for quadruples. *Journal of Statistical Planning and Inference*, Submitted.

McGilchrist, C. (1993) REML estimation for survival models with frailty. *Biometrics*, **49**, 221–225.

McGilchrist, C. A. and Aisbett, C. W. (1991) Regression with frailty in survival analysis. *Biometrics*, **47**, 461–466.

Meta-Analysis Group in Cancer (1996) Reappraisal of hepatic arterial infusion in the treatment of nonresectable liver metastates from colorectal cancer. *Journal of the National Cancer Institute*, **88**, 252–258.

Meta-Analysis Group in Cancer (1998) Efficacy of intravenous continuous infusion of 5-fluorouracil compared with bolus administration in patients with advanced colorectal cancer. *Journal of Clinical Oncology*, **16**, 301–308.

Morris, J. S. (2002) The BLUPs are not "best" when it comes to bootstrapping. *Statistics and Probability Letters*, **56**, 425–430.

Murphy, S. A., Rossini, A. J. and van der Vaart, A. W. (1997) Maximum likelihood estimation in the proportional odds model. *Journal of the American Statistical Association*, **92**, 968–976.

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

Oakes, D. (1989) Bivariate survival models induced by frailties. *Journal of the American Statistical Society*, **84**, 487–493.

Pickles, A. and Crouchley, R. (1995) A comparison of frailty models for multivariate survival data. *Statistics in Medicine*, **14**, 1447–1461.

Pollard, D. (1984) *Convergence of stochastic processes.* New York: Springer-Verlag.

Ripatti, S. and Palmgren, J. (2000) Estimation of multivariate frailty models using penalized partial likelihood. *Biometrics*, **56**, 1016–1022.

Shih, J. H. and Louis, T. A. (1995a) Assessing gamma frailty models for clustered failure time data. *Lifetime Data Analysis*, **1**, 205–220.

Shih, J. H. and Louis, T. A. (1995b) 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. *Publications de l'Institut de Statistique de l' Université de Paris*, **8**, 229–231.

Spiekerman, C. F. (1995) *Regression methods for correlated survival data.* Ph.D. thesis, University of Washington.

Spiekerman, C. F. and Lin, D. Y. (1996) Checking the marginal Cox model for correlated failure time data. *Biometrika*, **83**, 143–156.

Spiekerman, C. F. and Lin, D. Y. (1998) Marginal regression models for multivariate failure time data. *Journal of the American Statistical Association*, **93**, 1164–1175.

Therneau, T. M. and Grambsch, P. M. (2000) *Modeling survival data. Extending the Cox model.* New York: Springer.

van der Vaart, A. W. (1998) *Asymptotic statistics.* Cambridge: Cambridge University Press.

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

Vaida, F. and Xu, R. (2000) Proportional hazards model with random effects. *Statistics in Medicine*, **19**, 3309–3324.

Van Keilegom, I. and Veraverbeke, N. (1997) Estimation and bootstrap with censored data in fixed design nonparametric regression. *Annals of the Institute of Statistical Mathematics*, **49**, 467–491.

Viswanathan, B. and Manatunga, A. K. (2001) Diagnostic plots for assessing the frailty distribution in multivariate survival data. *Lifetime Data Analysis*, **7**, 143–155.

Weller, J., Saran, A. and Zeliger, Y. (1992) Genetic and environmental relationships among somatic cell count, bacterial infection, and clinical mastitis. *Journal of Dairy Science*, **75**, 2532–2540.

West, B. T., Welch, K. B. and Galecki, A. T. (2007) *Linear mixed models: A practical guide using statistical software.* Boca Raton, Florida: Chapman & Hall/CRC.

Yang, S. and Prentice, R. L. (1999) Semiparametric inference in the proportional odds model. *Journal of the American Statistical Association*, **94**, 125–136.

Yin, G. and Cai, J. (2004) Additive hazards model with multivariate failure time data. *Biometrika*, **91**, 801–818.

Zhou, Y. and Yip, P. S. F. (1999) A strong representation of the product-limit estimator for left truncated and right censored data. *Journal of Multivariate Analysis*, **69**, 261–280.

Samenvatting

In de overlevingsanalyse is men geïnteresseerd in positieve stochastische veranderlijken die het verloop van de tijd aangeven tussen een beginmoment en het moment waarop een bepaalde gebeurtenis zich voordoet. Enkele voorbeelden zijn de tijd tussen de aanvang van een ziekte en het overlijden van de patiënt, de tijd tussen de genezing van een patiënt en de terugkeer van de ziekte, de tijd tussen de opstart van een machine en het falen van de machine. Deze positieve stochastische veranderlijke wordt de overlevingstijd of de faaltijd genoemd.
Een specifiek kenmerk, dat vaak voorkomt in de analyse van overlevingsgegevens, is dat de gegevens onderworpen zijn aan rechtse censurering. Voor rechts gecensureerde observaties is de enige beschikbare informatie dat de niet-geobserveerde overlevingstijd groter is dan de geobserveerde censureringstijd. In vele studies zijn de observaties gegroepeerd (geclusterde gegevens); overlevingstijden in dezelfde cluster zijn vaak gecorreleerd. Zulke gegevens worden gecorreleerde of multivariate overlevingsgegevens genoemd. In de literatuur zijn er gepaste modellen beschreven om multivariate overlevingsgegevens te modelleren en te analyseren: frailty modellen en copula modellen zijn typische voorbeelden (zie onder meer Therneau and Grambsch, 2000; Hougaard, 2000; Klein and Moeschberger, 2003; Duchateau and Janssen, 2008).
Frailty modellen zijn uitvalsmodellen waarin het effect van elke cluster wordt beschreven door een stochastische factor; daarom zijn frailty modellen conditionele modellen. Een shared frailty model is een multiplicatief uitvalsmodel dat bestaat uit drie componenten: een frailty factor die het toevallig effect van elke cluster modelleert, de gemeenschappelijke referentie uitvalsfunctie en een factor die parametrisch modelleert hoe de uitvalsfunctie afhangt van de

covariaten.

Copula modellen worden gebruikt om geclusterde overlevingsgegevens te modelleren waarbij de grootte van de cluster klein is en hetzelfde voor alle clusters. In dit proefschrift bestuderen we copula modellen voor vier-dimensionale overlevingsgegevens. In copula modellen wordt de gezamenlijke overlevingsfunctie van de vier overlevingstijden in een cluster gemodelleerd als een functie van de marginale overlevingsfuncties. Deze functie wordt de copula genoemd. De copula bepaalt het type van de afhankelijkheid. De marginale overlevingsfuncties kunnen parametrisch, semi-parametrisch of niet-parametrisch gemodelleerd worden (Shih and Louis, 1995b; Glidden, 2000; Andersen, 2005). Via de marginale overlevingsfuncties kan ook het effect van covariaten gemodelleerd worden.

In dit proefschrift ontwikkelen we nieuwe schattingsmethoden en bootstrap procedures voor frailty modellen en copula modellen. De ontwikkelde methoden worden geïllustreerd met voorbeelden uit klinische en veterinaire studies.

In hoofdstuk 1 definiëren we basisbegrippen uit de overlevingsanalyse en introduceren we de voorbeelden die gebruikt worden om de ontwikkelde methoden te illustreren. In hoofdstuk 2 geven we een overzicht van de modellen die zullen gebruikt worden in de volgende hoofdstukken voor het analyseren van multivariate overlevingsgegevens. We beschrijven ook de schattingsmethoden die in de literatuur reeds beschreven zijn voor deze modellen.

In hoofdstuk 3 en hoofdstuk 4 bestuderen we frailty modellen voor multivariate overlevingsgegevens waarbij de clusters een verschillende grootte mogen hebben. In hoofstuk 3 bestuderen we het shared frailty model. Dit model is een uitbreiding van het proportionele uitvalsmodel van Cox (Cox, 1972): in het Cox model wordt een multiplicatieve stochastische frailty factor toegevoegd die het effect van de verschillende clusters beschrijft. We ontwikkelen twee parametrische bootstrap algoritmen die gebruikt kunnen worden om de standaardafwijkingen van de schatters van de parameters in het shared frailty model te schatten. Gebaseerd op een simulatiestudie vergelijken we de twee parametrische bootstrap algoritmen met een bestaand niet-parametrisch bootstrap algoritme dat beschreven wordt in Therneau and Grambsch (2000).

De resultaten tonen aan dat het eerste parametrisch bootstrap algoritme, gebaseerd op het trekken uit de verzameling van geschatte frailties, de empirische variabiliteit van de geschatte parameters onderschat. Het tweede parametrisch bootstrap algoritme, gebaseerd op het genereren van frailty factoren uit de geschatte frailty verdeling, geeft een relatief nauwkeurige benadering voor de empirische standaardafwijkingen van de geschatte parameters, ook indien het model fout gespecifieerd is. De empirische variabiliteit van de heterogeniteitsparameter kan echter sterk verschillend zijn voor het correcte model en een fout gespecifieerd model. Uit de resultaten volgt dus dat de robuustheid wat betreft de heterogeniteitsparameter niet gegarandeerd kan worden voor de bootstrap algoritmen (ook niet voor het niet-parametrisch bootstrap algoritme). De robuustheid geldt wel voor de regressiecoëfficiënten.
In hoofdstuk 4 bestuderen we een frailty model dat een uitbreiding is van het shared frailty model en dat vaak nuttig is binnen de context van klinische studies: een frailty model met een stochastisch cluster effect en een stochastisch behandelingseffect. Bestaande schattingsmethoden voor (shared) frailty modellen zijn gebaseerd op de aannemelijkheidsfunctie (zie onder meer Vaida and Xu, 2000; Cortiñas Abrahantes and Burzykowski, 2005; Ripatti and Palmgren, 2000; Legrand *et al.*, 2005). Wij ontwikkelen een alternatieve schattingsmethode die gebaseerd is op een transformatie van het model. De parameters in het frailty model worden, door het toepassen van de modeltransformatie, parameters in een gemengd lineair model. We kunnen dan methodologie voor gemengde lineaire modellen toepassen om deze parameters te schatten. We tonen aan dat het idee van modeltransformatie ook gebruikt kan worden voor andere conditionele modellen in de overlevingsanalyse, zoals het multivariaat proportioneel odds model en het multivariaat additief risico model. De meeste statistische computerprogramma's bevatten procedures om de parameters in gemengde lineaire modellen te schatten maar bevatten slechts een beperkt aantal procedures voor conditionele overlevingsmodellen met stochastische effecten. Daarom is de ontwikkelde schattingsmethode nuttig om inzicht te verwerven in de heterogeniteit die aanwezig is in gecorreleerde overlevingsgegevens. We evalueren de voorgestelde methode voor het schatten van frailty

modellen gebaseerd op een simulatiestudie. Uit de resultaten blijkt dat de ontwikkelde methode een goed en eenvoudig alternatief is voor het schatten van de parameters in een frailty model indien er voldoende clusters zijn in de gegevens ($K = 20$) en indien er in elke cluster voldoende observaties zijn met dezelfde waarde voor de binaire covariaat (minstens $n_{ij} = 50$).

In hoofdstuk 5 en hoofdstuk 6 bestuderen we copula modellen voor vier-dimensionale overlevingsgegevens. De motivatie komt van een studie omtrent gecorreleerde infectietijden in de vier uierkwartieren van melkkoeien. We gebruiken copula modellen om de afhankelijkheid tussen de vier overlevingstijden te modelleren.

In hoofdstuk 5 ontwikkelen we een semi-parametrische en een niet-parametrische twee-stappen schattingsmethode voor vier-dimensionele copula modellen. In de eerste stap van de methode schatten we de marginale overlevingsfuncties. In de marginale survival functies modelleren we het effect van een deterministische binaire covariaat op het niveau van de cluster en een deterministische binaire covariaat op het niveau van de observationele eenheid in de cluster. Dit type van covariaten stemt overeen met de covariaten in het motiverend voorbeeld. In de tweede stap van de schattingsmethode, schatten we de copula parameter door het maximaliseren van de logaritme van de aannemelijkheidsfunctie waarin de marginale overlevingsfuncties vervangen zijn door hun schattingen die we in de eerste stap verkregen hebben. We tonen de consistentie en de asymptotische normaliteit aan van de schatters uit de eerste en de tweede stap van de schattingsprocedure.

In hoofdstuk 6 passen we de methodologie, die we ontwikkeld hebben in hoofdstuk 5, toe op het voorbeeld van de uierinfectie gegevens. We beschouwen drie copula modellen die genest zijn in de familie van de power variance copula modellen: Clayton, positief stabiele en invers Gaussische copula modellen. Elk van deze copula modellen beschrijft een ander type afhankelijkheid tussen de gegevens binnen een cluster. Als toets gebruiken we een pseudo aannemelijkheidsquotiënt (pseudo likelihood ratio test) om een geschikt copula model te selecteren in de power variance copula familie dat de afhankelijkheid tussen de overlevingstijden in een cluster goed beschrijft (Glidden, 2000; Andersen,

2005). We stellen een bootstrap algoritme voor dat kan gebruikt worden om de p-waarde van deze toetsingsgrootheid te bepalen. We bestuderen de kans op een type I fout en het onderscheidingsvermogen van de pseudo aannemelijkheidsquotiënt toets gebaseerd op een simulatiestudie die de situatie in de uierinfectie gegevens imiteert. De resultaten tonen aan dat de pseudo aannemelijkheidsquotiënt toets een waardevolle methode is om een geschikt copula model te selecteren in de power variance copula familie.