Random-Effects Modelling of Bivariate Survival Data

23-wöchige Abschlussarbeit im Rahmen der Prüfung im Studiengang Angewandte Statistik (M.Sc.) an der Universität Göttingen

Funding Acknowledgement: This work was supported by the Deutsche Forschungsgemeinschaft (DFG; German Research Foundation, grant number UN 400/2-1).

vorgelegt am: 10.10.2019

von: Maximilian Bardo

aus: Kassel

Matrikelnummer: 21214350

supervised by

Dr. Steffen Unkel and Dr. Andreas Leha

This master's thesis is dedicated to my mother who was in hospital at the UMG when this thesis was written as T finally realised to be 59 years and 1 month. Her kindness, positivity and inventiveness is an inspiration. Her help and advice is impossible to replace and impossible to forget ...

Contents

List of Figures

List of Tables

Abbreviations & Nomenclature	
Term	Meaning
$\boldsymbol{\beta}$	Vector of FE
c_i	Censoring time for cluster i
$d_{i,j}$	Censoring indicator
$\eta_{i,j}$	$\boldsymbol{x}_{i,j}^T\boldsymbol{\beta} + v_i$
$f_{i,j}(t)$	$\frac{\partial F_{i,j}(t)}{\partial t}$
$f_{i,.}(t_1, t_2)$	$\frac{\partial F_{i,.}(t_1,t_2)}{\partial t_1 \partial t_2}$
$f_{i,j\|Z}(t\|z)$	$\frac{\partial F_{i,j\|Z}(t\|z)}{\partial t}$
$f_{i,.\|Z}(t_1, t_2\|z)$	$\frac{\partial F_{i,.\|Z}(t_1,t_2\|z)}{\partial t_1 \partial t_2}$
$F_{i,j\|Z}(t\|z)$	$P(T_{i,j} \leq t\|Z = z)$
$F_{i,.\|Z}(t_1, t_2\|z)$	$P(T_{i,1} \leq t_1, T_{i,2} \leq t_2\|Z = z)$
FE	Fixed Effect
$g_Z(z)$/ $g_V(v)$	Density of frailties/ log-frailties respectively
$h_0(t)$	Baseline hazard
$h_{0,i}(t)$	Sub-population baseline hazard: $exp\{\boldsymbol{x}_{i,j}^T\boldsymbol{\beta}\}h_0(t)$
$h_{i,j\|Z}(t\|z)$	Conditional hazard of individual j from cluster i
$h_{i,j}(t)$	Sub-population hazard of individual j from cluster i
H-Likelihood	Hierarchical Likelihood
h-loglihood	Hierarchical log-likelihood
OR	Odds-Ratio
$\mathcal{I}$	Indicator function
$i; i^c$	i usually refers to cluster i. $i^c \neq i$
$j; j^c$	j usually refers to individual j. $j^c \neq j$
loglihood	Log-likelihood
MLE/s	Maximum Likelihood Estimate/s
n	Number of clusters
RE/REs	Random Effect/s
RV/RVs	Random Variable/s
$S_{i,j}(t)$	$1 - F_{i,j}(t)$
$S_{i,.}(t)$	$P(T_{i,1} > t_1, T_{i,2} > t_2)$
$S_{i,j\|Z}(t\|z)$	$1 - F_{i,j\|Z}(t\|z)$
$S_{i,.\|Z}(t_1, t_2\|z)$	$P(T_{i,1} > t_1, T_{i,2} > t_2\|Z = z)$
Sub-population	Any particular group of the entire population separated by some criterion

Abbreviations & Nomenclature	
Term	Meaning
t_j	Specific value for survival time. Chosen for analytical reasons (function input). $j = 1, 2$ refers to the specific individual, j might be dropped when unnecessary (univariate functions).
$t_{i,j}$	Observed survival time of individual j from cluster i
$T_{i,j}$	Random Variable: Survival time of individual j from cluster i
τ	Kendall's τ
θ	$V[Z_i]$
$\boldsymbol{x}_{i,j}$	Covariate vector of individual (i, j). i, j do not refer to matrix indices. Sometimes abbreviated as $\boldsymbol{x}_i$.
$\boldsymbol{X}$	Covariate matrix. Ordered by cluster.
v_i	Realisation of V_i, typically unobserved
$\hat{v}_i$	MLE of v_i
V_i	Random Variable: Log-Frailty for cluster i. Sometime V is used as generic RV
$V[]$	Variance operator
$y_{i,j}$	Observed value for $Y_{i,j}$
$Y_{i,j}$	Random Variable: $min(T_{i,j}, c_{i,j})$
z_i	Realisation of Z_i, typically unobserved
$\hat{z}_i$	MLE of z_i
Z_i	Random Variable: Frailty for cluster i. Sometime Z is used as generic RV
ζ	Cross Ratio Function

1 Introduction

Survival time analysis investigates systematic patterns of the time from origin to a certain event, for example, the time from birth to death (Collet, 2015, p.1). A key measure is the population hazard rate, which is the instantaneous risk to die in the very next moment if one is still alive, possibly dependent on covariates.

Variation that is left when accounted for covariates might in some cases still inhibit important information. This is the case if there is some structure between (certain) observational units. For example, twins might be more alike than total strangers due to unobserved factors like a common lifestyle or because of genetic disposition. Neglecting this kind of structure might disrupt inference: individuals with a higher risk will more likely experience the event in the early stages. This leaves those with a lower hazard at later stages. The interpretation of a model which neglects the structure between individuals could be that the hazard is steadily decreasing with time. However, due to the selection effect, this is entirely unclear and might even be upside down in a scenario where the increased risk is masked by the selection effect. (Aalen et al., 2008, p. 231-232) Additionally, the dependence in the data as described above destroys the iid assumption which is necessary for valid maximum Likelihood estimates (MLEs). Hence, an appropriate modelling of the cluster-specific hazard becomes compelling.

The structure within clusters can be accounted for by adding random effects (REs) to the analysis. In the context of survival time analysis, the term frailty is used for random effects. The modelling of unobserved influences via frailty models might simply be conducted to account for confounding influences or it might be of interest itself, as it will be in this thesis.

This thesis deals with bivariate clusters: the lifetime of twins is examined, where the unobserved factor, inducing a twin-specific hazard, is genetics and to a certain degree shared environmental influences. The dataset contains mono- and dizygotic twins, and the first, naturally, share more from the gene pool as the latter. Consequently, monozygotic twins should be more alike and the twin-specific hazard rates should be more pronounced for a lot of twins than for those of dizygotic twins. The variance (θ) of the frailty term indicates how pronounced the twin-specific hazard can get. Hence, the precision of θ is of major importance.

The H-Likelihood framework of Lee and Nelder (1996) is chosen for modelling. An implementation for survival analysis can be found in the frailtyHL package (version 2.2) (Ha et al., 2018) available for R (R Core Team, 2013) (version 3.6.0). The frailtyHL framework is very promising for two reasons: Firstly, it is very general and can, in general, be extended to any appropriate frailty distribution. Secondly, the order of approximation to the Likelihood function is higher and it includes further computational details as, for

example, coxph from the survival package. Hence, one expects improved estimates.

The frailtyHL package will be put in competition to coxph from the survival package (Therneau and Grambsch, 2015) (version 2.44-1.1). The differences in results will be discussed for the twin dataset. A simulation study finally evaluates the performance. This thesis considers only proportional hazard models with semi-parametric baseline hazard and gamma as well as log-normal distributed frailty distributions. The differences in the estimated frailty distributions will be discussed thoroughly for the twin dataset.

One problem with the frailtyHL package emerged early: it is utterly slow. The smallest sub-sample of the twin dataset (male monozygotic twins) converged after more than 5 days. All other sub-sets took much longer (eight days and more) and did not converge (with default settings). Hence, a sub-sample of the twin dataset was drawn for illustration.

The results of this thesis are that the differences in the two approaches are considerable for the twin data, the log-normal frailty distribution is a better fit to the twin data than the gamma and finally, the higher order of approximation in the frailtyHL package performs badly for lognormal frailty and small cluster size.

This thesis is structured as follows: Firstly, the twin data is introduced. A brief discussion of basic concepts of survival analysis follows. Afterwards, the dependence in the data will be measured before the dependency is modelled. The sixth chapter discusses the H-Likelihood concept and the optimisation scheme. Then, the results of the frailty models for the twin data will be compared for both estimation procedures. The eighth chapter evaluates the methods based on a simulation study. The last chapter concludes.

2 Data

Two types of datasets will be used in this thesis. The first one is a real-world example of Danish twins. The second type of data is simulated data, where all parameters are known. The latter will be used to evaluate the accuracy of the methods used in this thesis.

The real-world dataset used to illustrate the models in this thesis comes from the Danish twin register (Hauge et al., 1968). In the Danish twin register data of all same-sex twins born in Denmark from 1881 to 1930 were collected. Regarding the version of the dataset used in this thesis, those twins were followed up to 1980. If at least one twin remained untraced or emigrated within the time period the observations were expelled from the dataset. The dataset also only includes information on twins where both got at least 15 years old. This left a sample 8985 twin-pairing or 17970 individuals respectively. (Hougaard et al., 1992, p. 17)

The variates covered by this dataset are sex, the zygosity status, year of birth, information if the individual died within the time period of observation and a time variable in unit days that either gives information when the individual died or was known to be alive at least, depending on the former variable.

The zygosity status is also the reason why pairs, where an individual died before the age of 15, were not included in the dataset, as zygosity status is hard to obtain if an individual died in childhood (Hougaard, 2000, p.12). And so the information gathered in this dataset is conditional on the event not happening before that barrier. This is called left-truncation (Hougaard, 2000, p. 30). Of the 8985 twin pairings, the zygosity status of 927 pairs was unknown. This data was deleted from the dataset leaving a total sample of $n = 8085$ twin-pairings. Unfortunately, this dataset (and most useful subsets) is too big to be analysed by the frailtyHL package. The computation lasted either more than five days, failed to converge or broke up because matrices that had to be inverted were singular. Hence, a sub-sample was drawn by sampling 2400 twin ID's, leaving 4800 observations in total. The subscript i will be used to refer to an arbitrary twin-pair, i.e. $i \in \{1, 2, \ldots, 2400\}$. If an individual is to be addressed the second subscript j will be used, i.e. $j \in \{1, 2\}$. In the following, summary statistics will be discussed, with the values of the entire dataset in brackets.

A share of 0.34 (0.35) of the twins was monozygotic and the remaining share of 0.66 (0.65) was dizygotic. The corresponding variable of the zygosity status is defined as $zyg_{i,j} \in \{mono, di\}$. Further on, 48% (48%) were male and 52% (52%) female. The corresponding variable is defined as $sex_{i,j} \in \{m, f\}$. As already mentioned, the information of the birth-year covers the years from 1881 to 1930, but 1900 was deducted from the year of birth. The corresponding variable is defined as $birth_{i,j} \in \{-19, -18, \ldots, 30\}$. Those variables serve either directly as potential covariates in the models or indirectly, in order

to separate the dataset into more useful subsets. The covariate matrix will be called $\boldsymbol{X}$ and is of size $2n \times K$, with K being the number of chosen covariates. Note, that there will never be an intercept in the models used in this thesis. The value of n differs if different subsets are subject to examination, say monozygotic males or dizygotic females. The corresponding sample size will be mentioned when the subsets are analysed.

The random variable (RV) on which statistical inference focuses on, is the time to event $T_{i,j}$. It covers the time (in days) from birth to death. This means that $T_{i,j}$ has a different origin with respect to calender time as long as $birth_{i,j} \neq birth_{i^c,j^c}$, where $i \neq i^c$ and $j \neq j^c$ (throughout this thesis). The year of birth will serve as an important covariate as life expectation is expected to increase with ongoing time. Realisations of $T_{i,j}$ are called $t_{i,j}$. However, in some cases people survived the observation period and so $t_{i,j}$ cannot be observed. In such cases, however, it is known that $t_{i,j} >$ the number of days from birth up to the deadline in 1980 $= c_{i,j}$. Note, that the censoring time $c_{i,j}$ is not random because the data is collected from public registers on the deadline in 1980 and so the censoring time only depends on ones own birthday (Hougaard, 2000, p.12). This is what is called homogeneous right censoring (Hougaard, 2000, p.29), i.e. if both individuals are censored, then $c_{i,1} = c_{i,2} = c_i$. If only one of the twins is censored, say the second, it must be that $t_{i,2} > c_i > t_{i,1}$. In the dataset this information is recorded through the RVs $Y_{i,j} = min\{T_{i,j}, c_i\}$ and $D_{i,j} = \mathcal{I}\{T_{i,j} < c_i\}$, with $\mathcal{I}$ being the indicator function. The realisations of $Y_{i,j}$ and $D_{i,j}$ are $y_{i,j}$ and $d_{i,j}$ respectively. There is no further censoring, say left- or interval censoring, at play in this dataset. Further on, $(D_{i,1}, D_{i,2})$ and $(T_{i,1}, T_{i,2})$ are assumed to be independent, i.e. the knowledge that an observation is censored does not affect the distribution of $(T_{i,1}, T_{i,2})$, neither in the past nor in the future.

All birth cohorts together, there was a share of 0.52 (0.52) of the twins, where both survived the observation time period, another 0.24 (0.25) where at least one individual died and thus, there is a share of 0.24 (0.23) remaining where both individuals survived. In total the censoring rate equals 64% (64%). The earliest death occurred 15 years and 18 days (15 years and 15 days) after birth. The earliest censoring time is 49 years and 15 days. The latest death occurred 94 years and 222 days (94 years and 296 days) after birth, whereas there was an individual who got older than 98 years and 30 days (98 years and 348 days).

Figure 1 gives an impression of the survival times. The bulk of observations, being censored or not, can be found in the age from 50 to about 85 years. Considering that area only and only twins where at least one individual died (black and red dots), one might also suspect a positive dependency within the survival times of twins by a purely data-driven approach.

It should be noted that within a twin-pairing there is no information that distin-

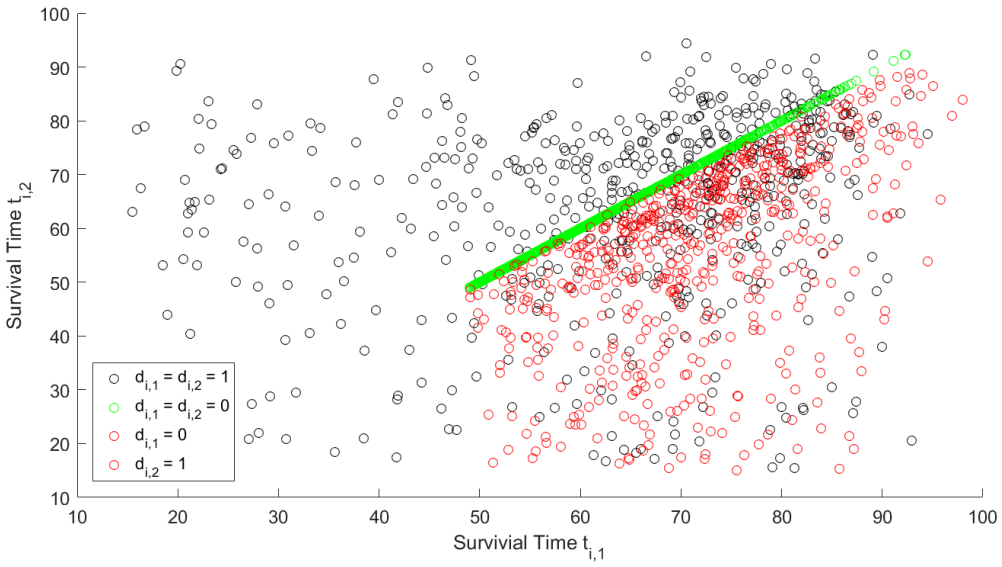

Figure 1: Survival times (in years) of the twins

guishes the twins that could be used as explanatory variable as $birth_{i,1} = birth_{i,2} = birth_i$, $sex_{i,1} = sex_{i,2} = sex_i$, $zyg_{i,1} = zyg_{i,2} = zyg_i$. There is also no natural ordering in the data. The first twin is simply the one that was named first in the public register. Thus, each twin pairing is an unordered cluster in which the twins are indistinguishable (apart from the dependent variable). Therefore, the sub-script j will often be dropped. In some cases, however, formulas are easier to understand with both sub-scripts. Hence, especially when sums and products are shown j will often be kept.

3 Survival Basics

Survival Time Analysis examines the time to a certain event. In the bivariate context of the twin dataset $(T_{i,1}, T_{i,2})$ denote the two RV of the duration from birth to death. Thus, $T_{i,j} \in \mathbb{R}_+$. The two pairings $(T_{i,1}, T_{i,2})$ and $(T_{i^c,1}, T_{i^c,1})$ are assumed to be independent. Within a cluster, however, the RVs $T_{i,1}$ and $T_{i,2}$ are assumed to be dependent.(Ha et al., 2017, p. 68) With respect to the twin dataset, this translates to the survival times of two twins being dependent as they share their genetic make-up but unrelated people being independent. The distribution of bivariate survival times might be different with respect to the chosen sub-sample, e.g. male monozygotic twins versus female dizygotic twins. To put it in a more general expression:

$$(T_{i,1}, T_{i,2}) \sim \pi_{sex_i, zyg_i}(\boldsymbol{\omega}_i), \tag{1}$$

where π_{sex_i, zyg_i} is some bivariate distribution with joint sub-population pdf $f_{i,\cdot}(t_1, t_2)$ and joint sub-population cdf $F_{i,\cdot}(t_1, t_2) = P(T_{i,1} \leq t_1, T_{i,2} \leq t_2)$ and parameter vector $\boldsymbol{\omega}_i$ possibly dependent on covariate $birth_i$. The "." refers to both of the twins, i.e. the joint distribution. A more simple model arises if, say, only monozygotic and dizygotic twins have different distributions, i.e.

$$(T_{i,1}, T_{i,2}) \sim \pi_{zyg_i}(\boldsymbol{\omega}_i). \tag{2}$$

This can be the case if sex_i and $birth_i$ are included as covariates. Model (1) and (2) will be put in competition to each other. Note, that there will be no notational difference for $f_{i,\cdot}$ and $F_{i,\cdot}$ whether the functions are from model (1) or (2). This will be clarified at the specific part of the thesis.

The individual sub-population pdf will be denoted as $f_{i,j}(t)$ and the individual sub-population cdf as $F_{i,j}(t) = P(T_{i,j} \leq t)$ respectively. The individual sub-population Survivor function is $S_{i,j}(t) = P(T_{i,j} > t) = 1 - P(T_{i,j} \leq t) = 1 - F_{i,j}(t)$. The joint sub-population Survivor function $S_{i,\cdot}$ is slightly more complicated and defined as

$$\begin{aligned}
S_{i,\cdot}(t_1, t_2) &= P(T_{i,1} > t_1, T_{i,2} > t_2) \\
&= 1 - P(T_{i,1} \leq t_1, T_{i,2} \leq t_2) - P(T_{i,1} > t_1, T_{i,2} \leq t_2) - \\
&\qquad P(T_{i,1} \leq t_1, T_{i,2} > t_2) \\
&= 1 + P(T_{i,1} \leq t_1, T_{i,2} \leq t_2) - \\
&\qquad \underbrace{P(T_{i,1} \leq t_1)}_{P(T_{i,1} \leq t_1, T_{i,2} > t_2) + P(T_{i,1} \leq t_1, T_{i,2} \leq t_2)} - \underbrace{P(T_{i,2} \leq t_2)}_{P(T_{i,1} > t_1, T_{i,2} \leq t_2) + P(T_{i,1} \leq t_1, T_{i,2} \leq t_2)} \\
&= 1 + F_{i,\cdot}(t_1, t_2) - F_{i,1}(t_1) - F_{i,2}(t_2).
\end{aligned} \tag{3}$$

In the context of bivariate Survival analysis, there are a couple of relevant hazard functions. The first, and with respect to univariate Survival analysis most straightforward one, is the sub-population hazard $h_{i,j}$ based on the univariate distribution of $T_{i,j}$. The second one is the sub-population hazard conditional on the survival of the other twin $h_{i,j|j^c}$ which is based on the joint sub-population distribution of $(T_{i,1}, T_{i,2})$. A third one, which will be introduced later in chapter 5, is the conditional hazard where conditionality refers to a RV Z_i that accounts for the dependence within twins. In the following, the first two concepts will be introduced and the interrelationships of Survival function, hazard function and pdf will be discussed.

The sub-population hazard $h_{i,j}$ is the risk to face the event in the very next moment given that the individual survived so far, i.e.

$$\begin{aligned} h_{i,j}(t) &= \lim_{\Delta \to 0} \frac{P(t < T_{i,j} \leq t + \Delta | T_{i,j} > t)}{\Delta} \\ &= \lim_{\Delta \to 0} \frac{P(t < T_{i,j} \leq t + \Delta)}{\Delta} \frac{1}{P(T_{i,j} > t)} \\ &= \frac{f_{i,j}(t)}{S_{i,j}(t)}. \end{aligned} \tag{4}$$

The sub-population hazard $h_{i,j}$ can also be expressed as

$$h_{i,j}(t) = -\frac{\partial ln\{S_{i,j}(t)\}}{\partial t}. \tag{5}$$

The cumulative sub-population hazard $H_{i,j} = \int_0^t h_{i,j}(u)du$ and using relationship (5)

$$\begin{aligned} H_{i,j}(t) &= -\int_0^t \frac{\partial ln\{S(u)\}}{\partial u} du \\ &= -[ln\{S(u)\}]_0^t \\ &= -ln\{S(t)\}, \end{aligned} \tag{6}$$

and consequently,

$$S_{i,j}(t) = exp\{-H_{ij}(t)\}. \tag{7}$$

These kind of relationships between the sub-population hazard, cumulative hazard and Survival function are needed for theoretical considerations as well as for the calculation of estimators of those quantities.

The second concept of hazard is that of the sub-population hazard conditional on the survival time of the other twin $h_{i,j|j^c}$. This measure will be needed to construct dependency measures in chapter 4. Here, two cases of conditionality are considered:

- $T_{i,j^c} = t_{j^c}$
- $T_{i,j^c} \geq t_{j^c}$

The measure $h_{i,j|j^c}$ has to be derived from the bivariate sub-population pdf $f_{i,\cdot}$. Let $T_{i,j} \in A$ denote one of the above two scenarios, then

$$\begin{aligned} h_{i,j|j^c}(t_1|T_{i,j^c} \in A) &= \lim_{\Delta \to 0} \frac{P(t_j < T_{i,j} \leq t_j + \Delta | T_{i,j} > t_j, T_{i,j^c} \in A)}{\Delta} \\ &= \lim_{\Delta \to 0} \frac{P(t_j < T_{i,j} \leq t_j + \Delta | T_{i,j^c} \in A)/\Delta}{P(T_{i,j} > t_j | T_{i,j^c} \in A)} \\ &= \frac{f_{i,j|j^c}(t_j|T_{i,j^c} \in A)}{S_{i,j|j^c}(t_j|T_{i,j^c} \in A)}, \end{aligned} \tag{8}$$

where $f_{i,j|j^c}$ and $S_{i,j|j^c}$ are the sub-population pdf and Survivor Function of $T_{i,j}$ conditional on $T_{i,j^c} \in A$.

Firstly, the case $A = t_{j^c}$ is considered. Then,

$$f_{i,j|j^c}(t_j|T_{i,j^c} = t_{j^c}) = \frac{f_{i,\cdot}(t_1, t_2)}{f_{i,j^c}(t_{j^c})}, \tag{9}$$

and

$$\begin{aligned} S_{i,j|j^c}(t_j|T_{i,j^c} = t_{j^c}) &= \lim_{\Delta \to 0} \frac{P(T_{i,j} > t_j, t_{j^c} \leq T_{i,j^c} < t_{j^c} + \Delta)/\Delta}{P(t_{j^c} \leq T_{i,j^c} < t_{j^c} + \Delta)/\Delta} \\ &= \frac{\partial S_{i,\cdot}(t_1, t_2) \ / \ \partial t_{j^c}}{-f_{i,j^c}(t_{j^c})}. \end{aligned} \tag{10}$$

Secondly, the case $A = \mathbb{R}_{>t_{j^c}}$ culminates in

$$\begin{aligned} f_{i,j|j^c}(t_j|T_{i,j^c} > t_{j^c}) &= \lim_{\Delta \to 0} \frac{P(t_j < T_{i,j} \leq t_j + \Delta, T_{i,j^c} > t_{j^c})/\Delta}{P(T_{i,j^c} > t_{j^c})} \\ &= \frac{-\partial S_{i,\cdot}(t_1, t_2) \ / \ \partial t_j}{S_{i,j^c}(t_{j^c})}, \end{aligned} \tag{11}$$

and

$$\begin{aligned} S_{i,j|j^c}(T_j > t_j|T_{i,j^c} > t_{j^c}) &= \frac{P(T_{i,j} > t_j, T_{i,j^c} > t_{j^c})}{P(T_{i,j^c} > t_{j^c})} \\ &= \frac{S_{i,\cdot}(t_1, t_2)}{S_{i,j^c}(t_{j^c})}. \end{aligned} \tag{12}$$

Hence, using (9) and (10) for (8) implies

$$h_{i,j|j^c}(t_j|T_{i,j^c} = t_{j^c}) = -\frac{f_{i,\cdot}(t_1, t_2)}{\partial S_{i,\cdot(t_1,t_2)} \ / \ \partial t_{j^c}}. \tag{13}$$

Using (11) and (12) for (8) yields

$$h_{i,j|j^c}(t_j|T_{i,j^c} > t_{j^c}) = -\frac{\partial S_{i,\cdot}(t_1,t_2) \ / \ \partial t_j}{S_{i,\cdot}(t_1,t_2)}. \tag{14}$$

Theses hazard functions will be needed for developing measures of dependence.

4 Measures of Dependence

Measures of Dependence are an important tool for assessing the degree and kind of dependence that is present in the data. So these measures are of significance when it comes to deciding for frailty or no-frailty modelling and to asses if a given frailty model is able to capture the characteristics of the data.

In the following, Kendall's τ and the cross-ratio function will be explained.

4.1 Kendall's τ

The measure τ is considered a global measure of dependence as the whole domain of the RVs is taken into account. (Hougaard, 2000, p. 129) Kendall's τ (Kendall, 1938, p. 82-86) in the survival context is defined as (Duchateau and Janssen, 2008, p. 123)

$$\tau = E[sign\{(T_{i,1} - T_{k,1})(T_{i,2} - T_{k,2})\}].$$

As

$$sign\{z\} = \begin{cases} -1 & \text{if } z < 0 \\ 0 & \text{if } z = 0 \\ 1 & \text{if } z > 0 \end{cases},$$

τ can be re-expressed as

$$\begin{aligned} \tau &= E[\mathcal{I}\{(T_{i,1} - T_{k,1})(T_{i,2} - T_{k,2}) > 0\}] - E[\mathcal{I}\{(T_{i,1} - T_{k,1})(T_{i,2} - T_{k,2}) < 0\}] \\ &= P[\mathcal{I}\{(T_{i,1} - T_{k,1})(T_{i,2} - T_{k,2}) > 0\}] - P[\mathcal{I}\{(T_{i,1} - T_{k,1})(T_{i,2} - T_{k,2}) < 0\}] \\ &= P[\mathcal{I}\{(T_{i,1} - T_{k,1})(T_{i,2} - T_{k,2}) > 0\}] - (1 - P[\mathcal{I}\{(T_{i,1} - T_{k,1})(T_{i,2} - T_{k,2}) > 0\}]) \\ &= 2P[\mathcal{I}\{(T_{i,1} - T_{k,1})(T_{i,2} - T_{k,2}) > 0\}] - 1, \end{aligned}$$

with $\mathcal{I}\{\}$ being the indicator function (Duchateau and Janssen, 2008, p. 124). The clusters i and k are (assumed to be) independent and identically distributed, i.e. $f_{i,\cdot} = f_{k,\cdot}$. The dependence might be within the pairing i and k respectively. The case $\tau > 0$ means that there is more probability mass when both values within a cluster are relatively big or small at once, compared to a combination of a relatively big and a small value within a cluster. This leads to a situation where there is relatively more probability mass for situations where $(T_{i,1} - T_{k,1})$ and $(T_{i,2} - T_{k,2})$ have the same sign, compared to a situation where $(T_{i,1} - T_{k,1})$ and $(T_{i,2} - T_{k,2})$ have opposite signs. Interpreted with respect to the life-times of twins, this kind of positive dependency means that we expect individual 1 of group i to live longer if individual 2 of twin pairing i lives longer, vice versa.

If $\tau < 0$, there is some negative dependency in the population: If $T_{i,1}$ tends to live

longer, $T_{i,2}$ tends to die earlier. The value $\tau = 0$ indicates no pattern in dependency at all: In case of independence between $T_{i,1}$ and $T_{i,2}$ it follows that $f_{i,\cdot} = f_{i,1}f_{i,2},\ f_{i,j} = f_{k,j}$, and, consequently,

$$\begin{aligned}
\tau &= 2 \times 2 \int_0^\infty \int_0^\infty f_{i,\cdot}(u_1, u_2) \left[\int_0^{u_1} \int_0^{u_2} f_{k,\cdot}(r_1, r_2)\ dr_1\ dr_2 \right] du_1\ du_2\ - 1 \\
&= 4 \int_0^\infty f_{i,1}(u_1) \int_0^{u_1} f_{i,1}(r_1)\ dr_1\ du_1 \int_0^\infty f_{i,2}(u_2) \int_0^{u_2} f_{i,2}(r_2)\ dr_2\ du_2\ - 1 \\
&= 4 \int_0^\infty F_{i,1}(u_1)\ dF_{i,1}(u_1) \int_0^\infty F_{i,2}(u_2)\ dF_{i,2}(u_2)\ - 1 \\
&= 4\ \frac{1}{2}\ \frac{1}{2} - 1 \\
&= 0.
\end{aligned}$$

Note, that $\tau = 0$ does not imply independence. This also implies that τ is not cohort adjusted in the case of the twin data and might soak up some of the cohort information.

The measure τ varies between -1 and 1, where an absolute value of 1 indicates perfect dependence but not necessarily $T_{i,1} = T_{i,2}$ as, say, $T_{i,2}$ might be shifted in location or on another scale as $T_{i,1}$. With respect to an estimation of τ, an absolute value of 1 refers to perfect dependence in rank (Kendall, 1938, p. 85).

A non-parametric estimate is obtained by

$$\hat{\tau} = \sum_{i=1}^{n} \sum_{k=1}^{n} \frac{a_{i,k} b_{i,k}}{n^2 - n},$$

with

$$a_{i,k} = \begin{cases} 1 \text{ if } t_{i,1} > t_{k,1} \\ 0 \text{ if } t_{i,1} = t_{k,1} \\ -1 \text{ if } t_{i,1} < t_{k,1} \end{cases}$$

and $b_{i,k}$ equivalently for the second individual from both clusters (Hougaard, 2000, p. 132). This, however, only works in a setting without censoring. To account for censoring, the empirical counterparts of $2P(T_{i,1} > T_{k,1} | Y_{i,1}, Y_{k,1}, D_{i,1}, D_{k,1}) - 1$ are calculated as values for $a_{i,k}$.(Wang and Wells, 2000, p. 1202) In a case where both are censored and $y_{i,1} > y_{k,1}$

(suppressed in following notation), this leads to

$$\begin{aligned}
2P(T_{i,1} > T_{k,1} | T_{i,1} > y_{i,1}, T_{k,1} > y_{k,1}) - 1 &= 2\frac{P(T_{i,1} > T_{k,1}, T_{i,1} > y_{i,1}, T_{k,1} > y_{k,1})}{S_{i,1}(y_{i,1})S_{k,1}(y_{k,1})} - 1 \\
&= 2\frac{\int_{y_{i,1}}^{\infty} f_{i,1}(u)\left[\int_{y_{k,1}}^{u} f_{k,1}(r)\ dr\right] du}{S_{i,1}(y_{i,1})S_{k,1}(y_{k,1})} - 1 \\
&= 2\frac{\int_{y_{i,1}}^{\infty} f_{i,1}(u)\left[S_{k,1}(y_{k,1}) - S_{k,1}(u)\right] du}{S_{i,1}(y_{i,1})S_{k,1}(y_{k,1})} - 1 \\
&= 2\frac{\int_{y_{i,1}}^{\infty} f_{i,1}(u)S_{k,1}(y_{k,1})du - \int_{y_{i,1}}^{\infty} f_{i,1}(u)S_{k,1}(u)du}{S_{i,1}(y_{i,1})S_{k,1}(y_{k,1})} - 1 \\
&= 2\frac{S_{i,1}(y_{i,1})S_{k,1}(y_{k,1}) - S_{i,1}(y_{i,1})^2/2}{S_{i,1}(y_{i,1})S_{k,1}(y_{k,1})} - 1 \\
&= 1 - \frac{S_{i,1}(y_{i,1})}{S_{k,1}(y_{k,1})}.
\end{aligned}$$

Note, that $S_{i,1}(t) = S_{k,1}(t),\ f_{i,1} = f_{k,1}$. The case where i is censored and $y_{i,1} < y_{k,1}$ (suppressed in following notation) leads to

$$\begin{aligned}
2P(T_{i,1} > T_{k,1} | T_{i,1} > y_{i,1}, T_{k,1} = y_{k,1}) - 1 &= 2\frac{P(T_{i,1} > T_{k,1}, T_{i,1} > y_{i,1} | T_{k,1} = y_{k,1})}{S_{i,1}(y_{i,1} | T_{k,1} = y_{k,1})} - 1 \\
&= 2\frac{\int_{y_{k,1}}^{\infty} f_{k,1}(y_{k,1})f_{i,1}(u)\ du\ / f_{k,1}(y_{k,1})}{S_{i,1}(y_{i,1})f_{k,1}(y_{k,1})/f_{k,1}(y_{k,1})} - 1 \\
&= 2\frac{S_{k,1}(y_{k,1})}{S_{i,1}(y_{i,1})} - 1
\end{aligned}$$

The cases where both are dead are either 1 or -1 and the for the other scenarios the values can be derived as the previous two. The corresponding estimator is obtained by putting the hats on the Survivor function. Ties are apart from the above definition defined to be zero. Table 1 gives an overview for all terms of $a_{i,k}$ (Hougaard, 2000, p. 133), where the indices for the individual sub-population Survivor functions $S_{i,1},\ S_{k,1}$ were dropped because of equivalence.

Table 1: $a_{i,k}$ for the Estimation of $\hat{\tau}$

$(d_{i,1}, d_{k,1})$	$y_{i,1} > y_{k,1}$	$y_{i,1} = y_{k,1}$	$y_{i,1} < y_{k,1}$
(1,1)	1	0	-1
(0,1)	1	1	$2\hat{S}(y_{k,1})/\hat{S}(y_{i,1}) - 1$
(1,0)	$1 - 2\hat{S}(y_{i,1})/\hat{S}(y_{k,1})$	-1	-1
(0,0)	$1 - \hat{S}(y_{i,1})/\hat{S}(y_{k,1})$	0	$\hat{S}(y_{k,1})/\hat{S}(y_{i,1}) - 1$

The values for $b_{i,k}$ are obtained by substituting the index 1 with 2. Finally, $\hat{\tau} = \frac{\sum_{i=1}^{n}\sum_{k=1}^{n} a_{i,k}b_{i,k}}{\sqrt{(\sum_{i=1}^{n} a_{i,k}^2)(\sum_{k=1}^{n} b_{i,k}^2)}}$ (Hougaard, 2000, p. 132). Note, that both approaches are equivalent in the absence of censoring. The corresponding univariate non-parametric Survivor functions are estimated by a separate estimation of the distribution of minimum lifetimes and of maximum lifetimes given the minimum. The comparison of the maximum and the minimum respectively also makes the estimated Survivor functions and Kendall's τ unaware of the arbitrary ordering in the clusters. (Hougaard et al., 1992, p. 20)

For the entire sample Hougaard et al. (1992) estimated $\hat{\tau}_{m,mono} = 0.173 > \hat{\tau}_{f,mono} = 0.147 > \hat{\tau}_{f,di} = 0.104 > \hat{\tau}_{m,di} = 0.091$.(p. 21) Thus, indicating stronger dependence within monozygotic twins versus dizygotitic twins in general. In particular the dependency among monozygotic male twins seems to be stronger than for monozygotic female twins. However, there is an indication of a stronger dependence for dizygotic female twins than for dizygotic male twins. Hougaard et al. (1992) also gives estimators for cohort adjusted Kendall's τ, which can be ordered as $\hat{\tau}_{m,mono} = 0.162 > \hat{\tau}_{f,mono} = 0.132 > \hat{\tau}_{f,di} = 0.090 > \hat{\tau}_{m,di} = 0.086$. (p. 24)

The dependence within the data makes inference based on univariate iid analysis poorly suited and theoretically wrong. Hence, ways to model the dependence within the data are needed to apply the iid principle and allocate all information in the data correctly.

4.2 Cross-Ratio Function

The cross-ratio function on the other hand is considered a local measure of dependence as only a subinterval of $T_{i,1}$ and $T_{i,2}$ is taken into account. (Hougaard, 2000, p. 136)

The cross-ratio function is the ratio of individual hazards conditioned on the survival status of the other twin at given points in time (t_1, t_2). More precisely (Duchateau and Janssen, 2008, p. 126)

$$\zeta(t_1, t_2) = \frac{h_{i,1|2}(t_1|T_{i,2} = t_2)}{h_{i,1|2}(t_1|T_{i,2} > t_2)}.$$

In words, $\zeta(t_1, t_2)$ is the factor by which the risk for individual 1 is increased if his or her twin dies at t_2 compared to the twin's (known) survival (longer than t_2).

Using (13) and (14), the cross ratio function can be expressed as

$$\begin{aligned}\zeta(t_1, t_2) &= \frac{f_{i,\cdot}(t_1, t_2) / \frac{\partial S_{i,\cdot}(t_1,t_2)}{\partial t_2}}{\frac{\partial S_{i,\cdot}(t_1,t_2)}{\partial t_1} / S_{i,,\cdot}(t_1, t_2)} \\ &= \frac{f_{i,\cdot}(t_1, t_2)}{\frac{\partial S_{i,\cdot}(t_1,t_2)}{\partial t_1}} \frac{S_{i,\cdot}(t_1, t_2)}{\frac{\partial S_{i,\cdot}(t_1,t_2)}{\partial t_2}}.\end{aligned} \tag{15}$$

From the last expression it can also be seen what happens with ζ if $T_{i,1}$ and $T_{i,2}$ are independent. In the case of independence, the factors of the denominator simplify to

$$\begin{aligned}
-\frac{\partial S_{i,\cdot}(t_1,t_2)}{\partial t_2} &= \lim_{\Delta\to 0}\frac{S_{i,\cdot}(t_1,t_2)-S_{i,\cdot}(t_1,t_2+\Delta)}{\Delta}\\
&= \lim_{\Delta\to 0}\frac{S_{i,1,\cdot}(t_1)S_{i,2}(t_2)-S_{i,1}(t_1)S_{i,2}(t_2+\Delta))}{\Delta}\\
&= S_{i,1}(t_1)\lim_{\Delta\to 0}\frac{S_{i,2}(t_2)-S_{i,2}(t_2+\Delta)}{\Delta}\\
&= S_{i,1,\cdot}(t_1)f_{i,2}(t_2)
\end{aligned}$$

and equivalently, $-\frac{\partial S_{i,\cdot}(t_1,t_2)}{\partial t_1} = S_{i,2}(t_2)f_{i,1}(t_1)$. Thus,

$$\begin{aligned}
\zeta(t_1,t_2) &\overset{T_{i,1}\perp T_{i,2}}{=} \frac{f_{i,\cdot}(t_1,t_2)\ S_{i,\cdot}(t_1,t_2)}{S_{i,2}(t_2)f_{i,1}(t_1)S_{i,1}(t_1)f_{i,2}(t_2)}\\
&= \frac{f_{i,1}(t_1)f_{i,2}(t_2)S_{i,1}(t_1)S_{i,2}(t_2)}{S_{i,2}(t_2)f_{i,1}(t_1)S_{i,1}(t_1)f_{i,2}(t_2)}\\
&= 1.
\end{aligned}$$

Note, that $\zeta(t_1,t_2)=1 \not\Longrightarrow T_{i,1}\perp T_{i,2}$ in general.

A non-parametric estimate of ζ can be justified from (15) as being the limiting case of an odds ratio (Anderson et al., 1992, p.642). Let $P_{11}^U = P(t_1 < T_{i,1} \le t_1+\delta, t_2 < T_{i,2} \le t_2+\delta)$, in $P_{01}^U = P(T_{i,1} > t_1+\delta, t_2 < T_{i,2} \le t_2+\delta)$, P_{10}^U is as P_{01}^U but with shifting the indices 1 and 2 and finally $P_{00}^U = P(T_{i,1} > t_1+\delta, T_{i,2} > t_2+\delta)$. Then,

$$\begin{aligned}
\zeta(t_1,t_2) &= \lim_{\delta\to 0}\frac{P_{11}^U/P_{01}^U}{P_{10}^U/P_{00}^U} \qquad (16)\\
&= \lim_{\delta\to 0}\frac{\frac{P_{11}/P(T_{i,1}>t_1,t_2<T_{i,2}\le t_2+\delta)}{P_{01}/P(T_{i,1}>t_1,t_2<T_{i,2}\le t_2+\delta)}}{\frac{P_{10}/P(T_{i,1}>t_1,T_{i,2}>t_2+\delta)}{P_{00}/P(T_{i,1}>t_1,T_{i,2}>t_2+\delta)}}\\
&= \lim_{\delta\to 0}\frac{\frac{P(t_1<T_{i,1}\le t_1+\delta|T_{i,1}>t_1,t_2<T_{i,2}\le t_2+\delta)}{1-P(t_1<T_{i,1}\le t_1+\delta|T_{i,1}>t_1,t_2<T_{i,2}\le t_2+\delta)}}{\frac{P(t_1<T_{i,1}\le t_1+\delta|T_{i,1}>t_1,T_{i,2}>t_2+\delta)}{1-P(t_1<T_{i,1}\le t_1+\delta|T_{i,1}>t_1,T_{i,2}>t_2+\delta)}}\\
&= \lim_{\delta\to 0}\underbrace{\frac{odds(t_1<t_{i,1}\le t_1+\delta|T_{i,1}>t_1,t_2<T_{i,2}\le t_2+\delta)}{odds(t_1<t_{i,1}\le t_1+\delta|T_{i,1}>t_1,T_{i,2}>t_2+\delta)}}_{\equiv OR(t_1,t_2,\delta)} \qquad (17)
\end{aligned}$$

(Duchateau and Janssen, 2008, p. 127), which is useful for interpretation.

A more useful expression for constructing an estimator, however, can be obtained if

fraction (16) is multiplied by $\frac{P(T_{i,1}>t_1,T_{i,2}>t_2)/P(T_{i,1}>t_1,T_{i,2}>t_2)}{P(T_{i,1}>t_1,T_{i,2}>t_2)/P(T_{i,1}>t_1,T_{i,2}>t_2)}$ which yields

$$OR(t_1, t_2, \delta) = \frac{P_{11}/P_{01}}{P_{10}/P_{00}}, \tag{18}$$

with $P_{11} = P(t_1 < T_{i,1} \leq t_1 + \delta, t_2 < T_{i,2} \leq t_2 + \delta | T_{i,1} > t_1, T_{i,2} > t_2)$, $P_{01} = P(T_{i,1} > t_1 + \delta, t_2 < T_{i,2} \leq t_2 + \delta | T_{i,1} > t_1, T_{i,2} > t_2)$, P_{10} is as P_{01} but, again, with shifting the indices 1 and 2 and lastly, $P_{00} = P(T_{i,1} > t_1 + \delta, T_{i,2} > t_2 + \delta | T_{i,1} > t_1, T_{i,2} > t_2)$. (Anderson et al., 1992, p.642)

There are two remaining issues to be clarified before an estimator can be calculated. Firstly, the estimator needs to be unaware of ordering in the cluster. Secondly, censoring has to be incorporated.

The first issue can be coped with by comparing the minimum of a cluster to the minimum of the function input and the maximum with the maximum. This is important to account for the conditionality in P_{11}, ..., i.e. for selecting the subset of the data which satisfies $t_{i,1} > t_1$, $t_{i,2} > t_2$ respectively. A simple selection based on the cluster indices only makes sense in a setting where the first and the second individual can be distinguished, say, by a treatment. This is not the case here. Hence, the corresponding subset of the data is selected by the condition $min\{y_{i,1}, y_{i,2}\} = y_{i,min} > min\{t_1, t_2\} = t_{min}$ and simultaneously $max\{y_{i,1}, y_{i,2}\} = y_{i,max} > max\{t_1, t_2\} = t_{max}$. This subset of the data is called $\tilde{\boldsymbol{y}}_{t_1,t_2}$ and the corresponding censoring indicators $\tilde{\boldsymbol{d}}_{t_1,t_2}$.

The second issue has to be dealt with by deleting observations that are censored within the interval of observation as it is impossible to know if they survived the period or not: The cluster i has to be removed from $\tilde{\boldsymbol{y}}_{t_1,t_2}$ in the case of $y_{i,min} \in (t_{min}, t_{min} + \delta)$ if $d_{i,min} = 0$ or $y_{i,max} \in (t_{max}, t_{max} + \delta)$ if $d_{i,max} = 0$. The remaining subset of the data is called $\boldsymbol{y}_{t_1,t_2}$ and $\boldsymbol{d}_{t_1,t_2}$, the number of clusters is n' and is dependent on (t_1, t_2).

Then, $\hat{P}_{11}(t_{min}, t_{max}) = \frac{1}{n'} \sum_{\boldsymbol{y}_{t_1,t_2}} \mathcal{I}\{y_{i,min} \leq t_{min} + \delta \wedge d_{i,min} = 1, y_{i,max} \leq t_{max} + \delta \wedge d_{i,max} = 1\}$, $\hat{P}_{10}(t_{min}, t_{max}) = \frac{1}{n'} \sum_{\boldsymbol{y}_{t_1,t_2}} \mathcal{I}\{y_{i,min} \leq t_{min} + \delta \wedge d_{i,min} = 1, y_{i,max} > t_{max} + \delta\}$, $\hat{P}_{01}(t_{min}, t_{max}) = \frac{1}{n'} \sum_{\boldsymbol{y}_{t_1,t_2}} \mathcal{I}\{y_{i,min} > t_{min} + \delta, y_{i,max} \leq t_{max} + \delta \wedge d_{i,max} = 1\}$, and P_{00} follows. This approach is problematic, however, if t_1 and t_2 are very close to each other as $\hat{P}_{01}(t_{min}, t_{max})$ is forced to be zero in the most extreme case of $t_{min} = t_{max}$. So by construction the the OR is forced to explode around the diagonal arguments. Adding 1 instead of 0.5 to the counter if $P_{0,1}$ is 0 did not solve the problem as the denominator of $OR(t_1, t_2, \delta) = \frac{P_{00}P_{11}}{P_{10}P_{01}}$ is forced to be relatively small because it is a product. This was solved by defining $P_{10/01} \equiv \frac{1}{2}(P_{01} + P_{10})$. This helps to inflate the the denominator of the OR. As there is no ordering in the data, this seems to be a more useful comparison as it yields much smoother results. The corresponding estimator of $P_{10/01}$ is $\hat{P}_{10/01}(t_1, t_2) = \frac{1}{2}(\hat{P}_{0,1}(t_1, t_2) + \hat{P}_{1,0}(t_1, t_2))$.

Finally, $\widehat{OR}(t_1, t_2, \delta) = \frac{\hat{P}_{11}(t_1,t_2)\hat{P}_{00}(t_1,t_2)}{\hat{P}_{01/10}(t_1,t_2)^2}$ will approximate an estimator for the cross-

ratio function. For this purpose, the survival-times of twins were transformed into yearly data. The distance δ was chosen to be ten years and the whole sample was taken. Figure 2 shows a plot of estimated ORs for a 55-year-old individual (left) and a 70-year-old individual (right) for any of the subsets. The time of his or her twin varies from 40 to 80 years. The ages of both individuals themselves are marked by the blue vertical lines for orientation. The black line represents an OR of 1 as reference.

One can see that the peak of one's risk is determined by one's own age. For the 55-year-old individual the risk typically starts to decline when his or her twin reaches at least the age of around 65. In the right panel of the older individual, however, the higher risk is shifted to later ages. This implies that the survival status of the other twin is especially informative if the twins are roughly equally old. This makes sense, as if one of the twins got substantially older than the other one, he or she might not be affected by, say, a genetic defect. On the other hand, this might also be because of deaths caused by different environmental influences or non-genetically reasons like accidents. The dependence in the data seems to differ locally.

Figure 3 shows the estimated OR in the area of $[40, 80] \times [40, 80]$ for monozygotic and dizygotic males. The risk factors are higher for monozygotic twins than for dizygotic twins, as already indicated by Kendall's τ. One can see that the risk peaks around the diagonal as could also be suspected from 2. This gets a little more narrow, the older the individuals get.

The low on the diagonal itself is somewhat hard to justify. This is caused by the adaptation too $P_{10/01}$ as explained above. If this would not be done, the values around the diagonal are inflated up to values of around 250 in regions where the bulk of observations are. This is entirely dominated by $\hat{P}_{00}(t_1, t_2)$ being very big and $\hat{P}_{01}(t, t) \to \frac{0.5}{n'}$ in most cases. This might also be exacerbated by the homogeneous censoring and the high-censoring rate. Because of those extremes, this did not seem to be a particularly good representation of reality. So the decision between two evils was made in favour of the depicted one. However, the diagonal itself should be regarded with caution and preferably only compared to other diagonal values. The general conclusions are not affected by choosing the adapted $\hat{P}_{10/01}$ apart from the numerical values and the diagonal itself.

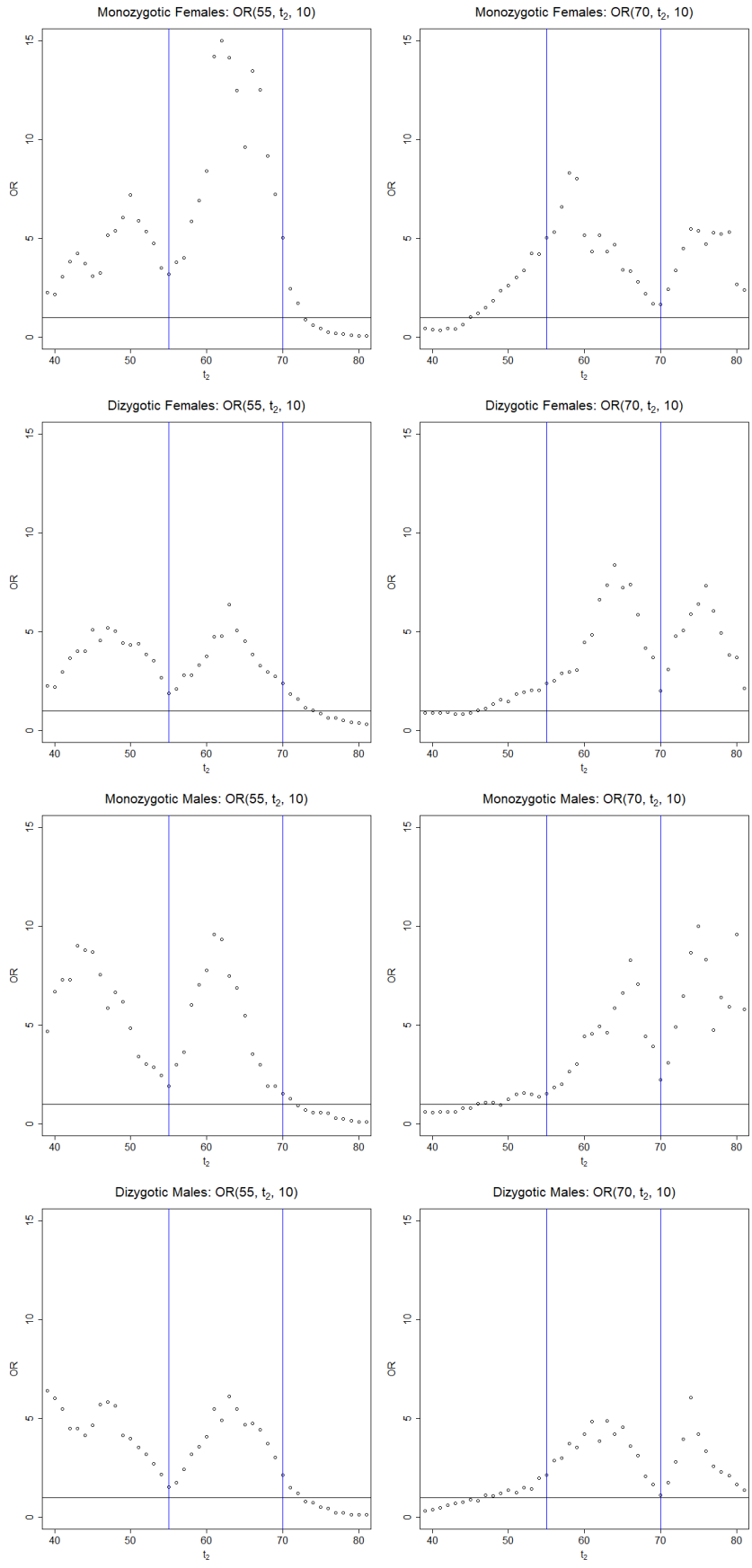

Figure 2: Odds Ratio of Survival Times

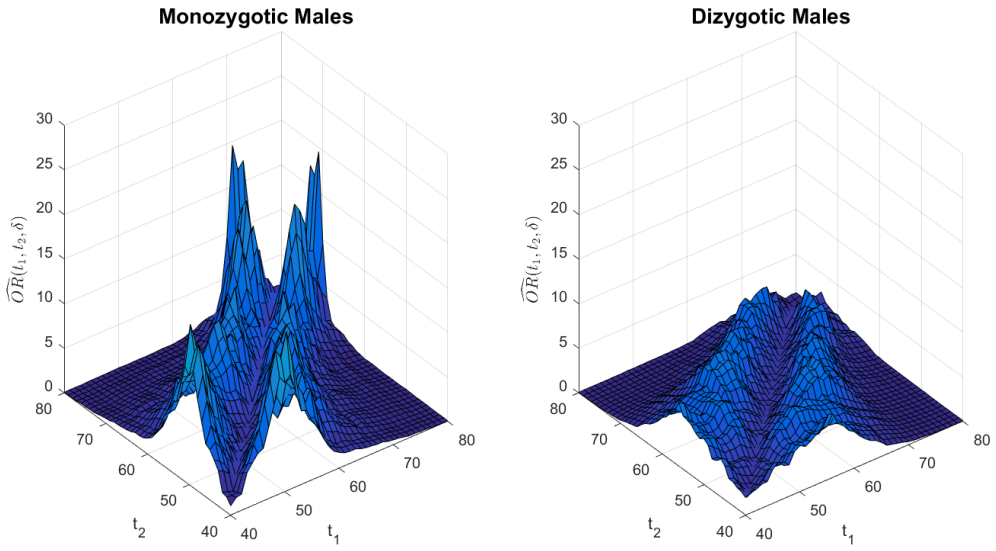

Figure 3: Odds Ratio of Survival Times of Male Twins

5 Frailty

The core idea of modelling dependence within the clusters is that they share a common feature in the risk to face the event of death in the very next moment. Say, a twin pair might inhibit a particular "unhealthy" set of genes, with respect to the susceptibility of getting cancer. Then, the risk of both twins might be particularly high of dying relatively early. This is modelled by introducing a random effect $Z_i > 0$, with realisation z_i in a multiplicative fashion on the hazard,(Aalen et al., 2008, p.275) i.e.

$$h_{i,j|Z}(t|Z = z_i) = z_i h_{0,i}(t),$$

where $h_{0,i}(t) = exp\{\boldsymbol{x}_i^T \boldsymbol{\beta}\} h_0(t)$ is the sub-population baseline hazard that might depend on time as well as on potential covariate information, indicating the subgroup of the population. A subscript j might also be needed if covariates on the individual level are present what is not the case in this thesis. Note that neither i nor (i, j) refers to the indices of the matrix $\boldsymbol{X}$ or any other matrix. The measure $h_{i,j|Z}$ will be called the conditional hazard rate from now on.

The random effect Z_i is assumed to follow a distribution π_Z $\forall i$, with parameter set $\boldsymbol{\theta}$ and density $g_Z(z)$. Further on, Z_i and Z_{i^c} are independent,(Ha et al., 2017, p. 69) that is, two different twin pairs do not share the same genetic make-up. The parameters of the frailty distribution are usually chosen such that $E[Z] = \int_0^\infty z g_Z(d) dz = const.$ for matters of identification (Ha et al., 2017, p. 70). The constant might chosen to be 1 but this is not always the case as will be seen later. The frailty distribution parameter $\boldsymbol{\theta}$ is a set of flexibility parameters. In the case of this thesis, $\boldsymbol{\theta} = \theta$ is a single parameter. Note

that the hazard will be relatively high if $z_i > const.$, and relatively low if $0 < z_i < const.$ relative to the expectation from the entire population. The sub-population baseline hazard $h_{0,i}$ should not be confused with the sub-population hazard rate even if $E[Z_i] = 1$ The reason is that with time running, people die and thus, the composition of the population changes. In particular, those with a relatively high conditional hazard will tend to die out much earlier and this leaves those with a relatively low conditional hazard. This will be investigated further below in more detail, but first, the conditional and unconditional Survival functions are required.

The conditional Survival function $S_{i,j|Z}$ can be derived via the conditional cumulative hazard function $H_{i,j|Z}$. The cumulative conditional hazard function can be obtained by integrating the conditional hazard up to time t:

$$\begin{aligned} H_{i,j|Z}(t|z_i) &= \int_0^t z_i h_{0,i}(u)du \\ &= z_i \int_0^t h_{0,i}(u)dt \\ &= z_i H_{0,i}(t). \end{aligned}$$

Using the relationship (7),

$$S_{i,j|Z}(t|z_i) = exp\{-z_i H_{0,i}(t)\}$$

yields the conditional univariate Survival function. The unconditional univariate Survival function can be obtained by integrating the frailty out of the joint density of survival time and frailty:

$$\begin{aligned} S_{i,j}(t) &= \int_t^\infty f_{i,j}(u)du \\ &= \int_t^\infty \int_0^\infty f_{i,j|Z}(u|z)g_Z(z)\ dz\ du \\ &= \int_0^\infty \int_t^\infty f_{i,j|Z}(u|z)du\ g_Z(z)\ dz \\ &= \int_0^\infty S_{i,j|Z}(t|z)g_Z(z)dz \\ &= \int_0^\infty exp\{-zH_{0,i}(t)\}g_Z(z)dz. \end{aligned} \tag{19}$$

The sub-population hazard rate can be obtained by using the relationship (4):

$$h_{i,j}(t) = \frac{\int_0^\infty h_{i,j|Z}(t|z)S_{i,j|Z}(t|z)g_Z(z)dz}{S_{i,j}(t)}. \tag{20}$$

In order to entirely understand the last expression it is worthwhile to derive the conditional distribution of Z_i given that we have a survivor at time t (Duchateau and Janssen, 2008, p. 112)

$$\begin{aligned} g_{Z|T}(z|T_{i,j} > t) &= \frac{P(T_{i,j} > t|Z_i = z)g_Z(z)}{P(T_{i,j} > T)} \\ &= \frac{S_{i,j|Z}(t|z)g_Z(z)}{S_{i,j}(t)}. \end{aligned} \tag{21}$$

Combining result (21) with equation (20), shows that

$$\begin{aligned} h_{i,j}(t) &= \int_0^\infty h_{i,j|Z}(t|z)g_{Z|T}(z|t)dz \\ &= h_{0,i}(t)\int_0^\infty z g_{Z|T}(z|t)dz \\ &= h_{0,i}(t)E[Z|T_{i,j} > t] \end{aligned}$$

the sub-population hazard rate is the expectation of the conditional hazard rate given the current distribution of frailty at time t. This feature also makes clear why frailty modelling is so important. If everything were to be modelled and interpreted based on the sub-population hazard rate one might only observe that the hazard decreases with increasing time. This might, however, only be a selection effect as only those with a low conditional hazard survive and thus, relatively high values for Z_i become rarer in the conditional distribution $g_{Z|T}(z|t)$. So it could still be the case that the baseline hazard $h_{0,i}$ increases with increasing time what would be entirely unobserved when modelling the sub-population hazard rate only. Frailty modelling is able to distinguish between a selection effect and the development of baseline hazard. Of course, this is not the only pattern of a selection effect and the development of baseline hazard but frailty modelling is able, if modelled adequately, to find those patterns. Pure population hazard modelling is not. (Aalen et al., 2008, p. 231-232)

In similar fashion to the univariate density, the bivariate density of the twins can be obtained by adding the fact, that conditional on z_i the event times $T_{i,1}$ and $T_{i,2}$ are independent (Ha et al., 2017, p. 69). Starting with the conditional Survivor function,

$$\begin{aligned} S_{i,.|Z}(t_1, t_2|z_i) &= S_{i,1|Z}(t_1|z_i)S_{i,2|Z}(t_2|z_i) \\ &= \int_0^\infty exp\{-z_i(H_{0,i}(t_1) + H_{0,i}(t_2))\} \end{aligned}$$

one can derive the unconditional Survivor function

$$\begin{aligned} S_{i,\cdot}(t_1, t_2) &= \int_{t_1}^{\infty} \int_{t_2}^{\infty} f_{i,\cdot}(u_1, u_2)\ dt_1 dt_2 \\ &= \int_{t_1}^{\infty} \int_{t_2}^{\infty} \int_{0}^{\infty} f_{i,\cdot|Z}(u_1, u_2|z) g_Z(z)\ dz dt_1 dt_2 \\ &= \int_{0}^{\infty} \int_{t_1}^{\infty} \int_{t_2}^{\infty} f_{i,\cdot|Z}(u_1, u_2|z) g_Z(z)\ dt_1 dt_2\ dz \\ &= \int_{0}^{\infty} S_{i,\cdot|Z}(u_1, u_2|z) g_Z(z)\ dz. \end{aligned}$$

The unconditional Survivor function $S_{i,j}$ and $S_{i,\cdot}$ are often expressed in a Laplaceian form. The Laplace function $\mathcal{L}$ is defined as

$$\begin{aligned} \mathcal{L}(c) &= \int_{0}^{\infty} exp\{-zc\} g_Z(z) dz \\ &= E[exp\{-Zc\}] \end{aligned}$$

ans so it is possible to re-express $S_{i,j}$ as

$$\begin{aligned} S_{i,j}(t) &= E[exp\{-ZH_{0,i}\}] \\ &= \mathcal{L}(H_{0,i}) \end{aligned}$$

(Aalen et al., 2008, p. 235) and $S_{i,\cdot}$ as

$$\begin{aligned} S_{i,\cdot}(t_1, t_2) &= E[exp\{-Z(H_{0,i}(t_1) + H_{0,i}(t_2))\}] \\ &= \mathcal{L}(H_{0,i}(t_1) + H_{0,i}(t_2)). \end{aligned}$$

Note, that $f_{i,j}(t) = -\frac{\partial \mathcal{L}(H_{0,i}(t))}{\partial t}$ and, because of (3), $f_{i,\cdot} = \frac{\partial^2 \mathcal{L}(H_{0,1}(t_1) + H_{0,2}(t_2))}{\partial t_1\ \partial t_2}$ what might be used to calculate hazard rates.

The advantage of the Laplace formulation is, that for some functions $g_Z(z)$ the results for $\mathcal{L}(c)$ are well established.

6 Likelihood

This chapter discusses the Maximum Likelihood Estimation of fixed effects (FE) $\boldsymbol{\beta}$, RE $\boldsymbol{z}_{n\times 1} = [z_i]$, baseline hazard $h_0(t)$ and flexibility parameter θ of the frailty distribution. This will be done with the Hierarchical-Likelihood (H-Likelihood) approach of Lee and Nelder (1996), which specifies the scale of the RE.

The chapter is structured as follows: The first sub-section discusses the H-Likelihood approach. In the second sub-section, the Likelihood for $\boldsymbol{\beta}$ and $\boldsymbol{z}$ is derived under the assumption that θ and $h_0(t)$ is known. Later in that sub-section a profile Likelihood approach is adopted which plugs the MLE of the baseline hazard $h_0(t)$ into the Likelihood. In the third sub-section, the restricted Likelihood for the flexibility parameter θ is discussed. In the fourth sub-section, the equations for maximum Likelihood estimations are gathered and the iterative algorithm to estimate all quantities of interest is introduced. This is the frailtyHL approach. Differences to the coxph command are then pointed out. The last sub-section discusses a Goodness of Fit measure and a hypothesis test.

6.1 H-Likelihood for Fixed and Random Effects

Care has to be taken when it comes to deriving MLEs for $\boldsymbol{\beta}$ and $\boldsymbol{z}$. Consider the extended likelihood

$$\begin{aligned} L(\boldsymbol{\beta}, \boldsymbol{z}; Y, Z) &= L(\boldsymbol{\beta}; Y|Z)\ L(\boldsymbol{z}; Z) \\ &= \prod_{i=1}^{n} \left[\prod_{j=1}^{2} f_{i,j|Z}(y_{i,j}|z_i)^{d_{i,j}} S_{i,j|Z}(y_{i,j}|z_i)^{1-d_{i,j}} \right] g_Z(z_i) \\ &= \prod_{i=1}^{n} \left[\prod_{j=1}^{2} h_{i,j|Z}(y_{i,j}|z_i)^{d_{i,j}} S_{i,j|Z}(y_{i,j}|z_i) \right] g_Z(z_i), \end{aligned}$$

where the RVs behind the semicolon in L specify the distribution on which the Likelihood is based:

- for $L(\boldsymbol{\beta}, \boldsymbol{z}; Y, Z)$ the Likelihood is based on the joint distribution of Y and Z
- for $L(\boldsymbol{\beta}; Y|Z)$ the Likelihood is based on the conditional distribution of Y given Z,
- for $L(\boldsymbol{z}; Z)$ the Likelihood is based on the distribution of Z.

For construction of the Likelihood $L(\boldsymbol{\beta}, \boldsymbol{z}; Y, Z)$ independence and non-informativeness between censoring and survival times is assumed (Ha et al., 2017, p. 69). Therefore, contributions to the Likelihood coming from the distribution of censoring times are ignored as they do not enrich the analysis of survival times at all. In case of the FEs $\boldsymbol{\beta}$ the MLEs

are invariant to transformations in the sense that $k(\hat{\boldsymbol{\beta}}_{MLE}) = \widehat{k(\boldsymbol{\beta})}_{MLE}$ for some element-wise function $k()$ due to its non-random nature (Ha et al., 2017, p. 42). This is not the case for Z, however, as there is a Jacobian term involved if the random variable is brought to another scale $V = k(Z)$ (Lee et al., 2017, p. 105)

$$g_V(v) = g_Z(k^{-1}(v)) \left| \frac{\partial k^{-1}(v)}{\partial v} \right|.$$

Then,

$$\begin{aligned} L(\boldsymbol{\beta}, \boldsymbol{v}; Y, V) &= L(\boldsymbol{\beta}; Y|V)\ L(\boldsymbol{v}; V) \\ &= \prod_{i=1}^{n} \left[\prod_{j=1}^{2} h_{i,j|V}(y_{i,j}|v_i)^{d_{i,j}} S_{i,j|V}(y_{i,j}|v_i) \right] g_V(v_i) \\ &= \prod_{i=1}^{n} \left[\prod_{j=1}^{2} h_{i,j|Z}(y_{i,j}|k^{-1}(v_i))^{d_{i,j}} S_{i,j|Z}(y_{i,j}|k^{-1}(v_i)) \right] \\ &\qquad \times g_Z(k^{-1}(v_i)) \left| \frac{\partial k^{-1}(v)}{\partial v}|_{v=v_i} \right|, \end{aligned}$$

with $h_{i,j|V}(y_{i,j}|v_i) = h_{i,j|Z}(y_{i,j}|k^{-1}(v_i))$ and $S_{i,j|V}(y_{i,j}|v_i) = S_{i,j|Z}(y_{i,j}|k^{-1}(v_i))$ as the RE is a fixed parameter in the conditional distribution of T.

Let $\boldsymbol{\beta}^{(1)}$ and $\boldsymbol{\beta}^{(2)}$ be some unequal vectors for $\boldsymbol{\beta}$ and $\hat{\boldsymbol{z}}^{(1)}$ and $\hat{\boldsymbol{z}}^{(2)}$ the corresponding MLEs of the REs. Further on, let $\tilde{\boldsymbol{v}}^{(1)} = k(\hat{\boldsymbol{z}}^{(1)})$ and $\tilde{\boldsymbol{v}}^{(2)} = k(\hat{\boldsymbol{z}}^{(2)})$ respectively, be the naive estimates of $\boldsymbol{v}$. In general,

$$\frac{L(\boldsymbol{\beta}^{(1)}, \tilde{\boldsymbol{v}}^{(1)}; Y, V)}{L(\boldsymbol{\beta}^{(2)}, \tilde{\boldsymbol{v}}^{(2)}; Y, V)} \neq \frac{L(\boldsymbol{\beta}^{(1)}, \hat{\boldsymbol{z}}^{(1)}; Y, Z)}{L(\boldsymbol{\beta}^{(2)}, \hat{\boldsymbol{z}}^{(2)}; Y, Z)}$$

because the Jacobian term $\left| \frac{\partial k^{-1}(v)}{\partial v}|_{v=\tilde{v}_i^{(q)}} \right|$, $q \in \{1, 2\}$, at any individual Likelihood contribution does not cancel out in the case of a non-linear function $k()$. The unequal Likelihood ratios show that inference on both, the RE and $\boldsymbol{\beta}$, is somewhat arbitrary, depending on the chosen parametrisation of the RE. Additionally, without further criteria no parametrisation of the random effects can be claimed to be correct or in some sense natural.

Criteria to find a natural or canonical scale of the random effects comes from the H-Likelihood approach. The H-Likelihood requires a parametrisation of the random effects $\boldsymbol{v} = k^*(\boldsymbol{z})$, with element-wise function $k^*()$, such that $\hat{\boldsymbol{\beta}}_{MLE}$ is identical in the maximisation of $L(\boldsymbol{\beta}, \boldsymbol{v}; Y, V)$ or $L(\boldsymbol{\beta}; Y) = \int_0^\infty L(\boldsymbol{\beta}, \boldsymbol{v}; Y, V) dv$. More precisely, with $\hat{\boldsymbol{v}}^{(q)}$ being

the MLE given $\hat{\boldsymbol{\beta}}^{(q)}$ the canonical scale of the random effect satisfies

$$\frac{L(\boldsymbol{\beta}^{(1)};Y)}{L(\boldsymbol{\beta}^{(2)};Y)} = \frac{L(\boldsymbol{\beta}^{(1)},\hat{\boldsymbol{v}}^{(1)};Y,V)}{L(\boldsymbol{\beta}^{(2)},\hat{\boldsymbol{v}}^{(2)};Y,V)},$$

i.e. the evidence for $\hat{\boldsymbol{\beta}}_{MLE}$ is the same in both Likelihood concepts. (Ha et al., 2017, p. 46) The extended Likelihood with canonical RE is called H-Likelihood. Once the H-Likelihood is found, the Likelihood is frozen and $\boldsymbol{v}$ is treated as a parameter, i.e. even if there is a transformation of V the Likelihood will not be changed, in particular, no Jacobian term will be multiplied (Ha et al., 2017, p. 71-72).

Unfortunately, there is not a canonical scale in every case and, in lack of an analytical expression for $\hat{\boldsymbol{v}}$, there is none here. There is a weak canonical scale, however. (Lee et al., 2017, p. 169)

Let $v_i = ln\{z_i\}$. Then, subject to estimation in $L(\boldsymbol{\beta},\boldsymbol{z};Y,Z)$ and $L(\boldsymbol{\beta},\boldsymbol{v};Y,V)$ is $\eta_i = \boldsymbol{x}_i^T\boldsymbol{\beta} + log\{z_i\}$ or $\eta_i = \boldsymbol{x}_i^T\boldsymbol{\beta} + v_i$ respectively, for all i. Obviously, both models are equivalent and should, therefore, lead to equivalent inference (Lee et al., 2017, p. 170). Thus, a scale of the RE must be chosen and kept fixed no matter if v_i or z_i is subject to estimation. The scale of the RE is called weakly canonical if inference is identical for trivial re-parametrisation of the form $v_i = m\ v_i^* + b$ or $ln\{z_i\} = m\ ln\{z_i^*\} + b$, with some $m \in \{\mathbb{R}\backslash 0\}$ and $b \in \mathbb{R}$ (Lee and Nelder, 2005, p. 146). In case of $ln\{z_i\} = m\ ln\{z_i^*\} + b \rightarrow z_i = z_i^{*m}\ exp\{b\}$, the Jacobian term in $g_{Z^*}(z^*) = g_Z(z(z^*))\ |m\ z^{*(m-1)} exp\{b\}|$ still depends on z^*. This affects the Likelihood ratio and therefore, inference is not invariant to the choice of the trivial reparameterisation. This is not the case for v, however. As $g_{V^*}(v^*) = g_V(v(v^*))\ |m|$, the multiplicative constant m simply cancels out in the Likelihood ratio and inference is unaffected by a trivial reparameterisation. In consequence, the parameter is (always) weakly canoncial if it combines additively with $\boldsymbol{x}_i^T\boldsymbol{\beta}$ (Lee et al., 2017, p. 170) and the H-Likelihood is set as

$$\begin{aligned}
L(\boldsymbol{\beta},\boldsymbol{v};Y,V) &= \prod_{i=1}^{n}\left[\prod_{j=1}^{2} h_{i,j|V}(y_{i,j}|v_i)^{d_{i,j}} S_{i,j|V}(y_{i,j}|v_i)\right] g_V(v_i) \\
&= \prod_{i=1}^{n}\left[\prod_{j=1}^{2} h_{i,j|V}(y_{i,j}|ln\{z_i\})^{d_{i,j}} S_{i,j|V}(y_{i,j}|ln\{z_i\})\right] g_V(ln\{z_i\}) \\
&= L(\boldsymbol{\beta},\boldsymbol{z};Y,V)
\end{aligned}$$

(Lee et al., 2017, p. 171).

Note, that in the above equations there was no Jacobian term involved and all functions remain functions of v, even though v is expressed as a function of z in the second and third line. Invariance with respect to parameter transformation is thereby (artificially) maintained by fixing the Likelihood on the scale of v.

From now on the H-Likelihood for the FE and RE $L(\boldsymbol{\beta}, \boldsymbol{v}; Y, V)$ is abbreviated as L^h. The hierarchical log-likelihood (h-loglihood)

$$\begin{aligned} ln\{L^h\} &= \sum_{i=1}^{n}\sum_{j=1}^{2}\Big\{d_{i,j}\ ln\{h_{i,j|V}(y_{i,j}|v_i)\} + ln\{S_{i,j|V}(y_{i,j}|v_i)\}\Big\} + ln\{g_V(v_i)\} \\ &= \sum_{i=1}^{n}\sum_{j=1}^{2}\Big\{d_{i,j}(\eta_i + ln\{h_0(y_{i,j})\}) - exp\{\eta_i\}H_0(y_{i,j})\Big\} + ln\{g_V(v_i)\}, \end{aligned}$$

using the interrelationship (7) between Survival and cumulative Hazard function. The h-loglihood is from now on abbreviated as l^h.

6.1.1 Profile H-Likelihood

The semi-parametric approach taken in this thesis leads to the assumption that the hazard is constant from one event in the dataset to the next. Let $y_{(0)} \coloneqq 0 < y_{(1)} < y_{(2)} < \cdots < y_{(D)}$ be the D distinct ordered event times, i.e. at least one person died at at every $y_{(m)}$. Further, let $dt_{(m)} = y_{(m)} - y_{(m-1)}$. Then, the assumption of a piecewise constant baseline hazard rate from event to event leads to $H_0(t) = \sum_{m>0:y_{(m)}<t, y_{(m+1)}<t} h_0^{(m)} dt_{(m)} + (t - y_{m^*})h(0)^{(m^*)}$, with $h(0)(y_m) = h(0)^{(m)}$ and $m^* = max(m : y_m < t)$ (Link, 1984, p. 602). Breslow (1974), however, simplified this estimator: "[...] all withdrawals, or censored observations, which occur in the interval (t_i, t_{i+1}) are adjusted to have occurred at t_i" (p. 93). Therefore,

$$H_0(t) = \sum_{m>0:y_m \leq t} dt_{(m)} h(0)^{(m)}.$$

Before proceeding to the l^h some definitions are necessary: $R_{(m)}$ denotes the risk set, i.e. it includes all individuals who are known to be alive just prior to $y_{(m)}$, $d_{(m)}$ is the number of people who died at $y_{(m)}$ and $\eta_{(m)} = \sum_{(i,j):y_{i,j}=y_{(m)}} \boldsymbol{x}_{i,j}^T \boldsymbol{\beta} + v_i$ and $g_V(\boldsymbol{v}) = \prod_i^n g_V(v_i)$. With all above definitions the h-loglihood can be expressed as

$$\begin{aligned} l^h &= \sum_{i=1}^{n}\sum_{j=1}^{2}\Big\{d_{i,j}(\eta_i + ln\{h_0(y_{i,j})\}) - exp\{\eta_i\}\sum_{m>0:y_{(m)}\leq y_{i,j}} h_0^{(m)} dt_{(m)}\Big\} + ln\{g_V(v_i)\} \\ &= \sum_{m=1}^{D}\Big\{\eta_{(m)} + d_{(m)} ln\{h_0^{(m)}\} - dt_{(m)} h_0^{(m)} \sum_{(i,j)\in R_{(m)}} exp\{\eta_i\}\Big\} + ln\{g_V(\boldsymbol{v})\}. \end{aligned}$$

The MLE for $h_0^{(m)}$ is derived by $\frac{\partial l^h}{\partial h_0^{(m)}} \stackrel{!}{=} 0$ and results in $\hat{h}_0^{(m)} = \frac{d_{(m)}}{dt_{(m)} \sum_{(i,j)\in R_{(m)}} exp\{\eta_i\}}$,[1] i.e.

[1] Note, that what here is defined as $h_0^{(m)} dt_{(m)}$ is simply $h_0^{(m)}$ in the frailtyHL package. This slight redefinition, however, does not affect inference as the loglihood, as defined in this thesis, differs only

the Breslow estimator. (Breslow, 1974, p. 93).

Using $\hat{h}_0^{(m)}$ as a plug-in estimator leads to the profile loglihood

$$\begin{aligned} l^h_{prof} &= \sum_{m=1}^{D} \Bigg\{ \eta_{(m)} + d_{(m)} ln \Bigg\{ \frac{d_{(m)}}{dt_{(m)} \sum_{(i,j)\in R_{(m)}} exp\{\eta_i\}} \Bigg\} - \\ &\qquad \frac{d_{(m)}}{\sum_{(i,j)\in R_{(m)}} exp\{\eta_i\}} \sum_{(i,j)\in R_{(m)}} exp\{\eta_i\} \Bigg\} + ln\{g_V(\boldsymbol{v})\} \\ &\propto \underbrace{\sum_{m=1}^{D} \Bigg\{ \eta_{(m)} - d_{(m)} ln\{ \sum_{(i,j)\in R_{(m)}} exp\{\eta_i\} \} \Bigg\}}_{l^p} + ln\{g_V(\boldsymbol{v})\}, \end{aligned} \tag{22}$$

where from (22) it can be seen that maximising l^h_{prof} is equivalent to maximising the partial h-loglihood l^p with Breslow approxmation to account for more than one death at a given point in time plus the penalty-term $ln\{g_V(\boldsymbol{v})\}$ (Ha et al., 2017, p. 73). From now on l^h is defined as (22).

6.2 Derivatives

This sub-section is purely technical and might be regarded as an integrated appendix. If technical details are not of interest one might skip this sub-section entirely and maybe come back to it if technical details or definitions are of interest in later chapters. The first and second-order derivatives of l^h with respect to $\boldsymbol{\beta}$ and $\boldsymbol{v}$ will be derived. They are be needed for estimation procedures and statistical testing. The estimation procedure itself will be discussed in one of the later sub-sections.

6.2.1 $\frac{\partial l^h}{\partial \boldsymbol{\beta}}$

The first-order derivatives will be needed for maximisation of the h-loglihood. The first set of derivatives is

$$\begin{aligned} \frac{\partial l^h}{\partial \boldsymbol{\beta}} &= \sum_{m=1}^{D} \Bigg\{ \boldsymbol{x}_{(m)} - \frac{d_{(m)}}{\sum_{(i,j)\in R_{(m)}} exp\{\eta_i\}} \sum_{(i,j)\in R_{(m)}} exp\{\eta_i\} \boldsymbol{x_{i,j}} \Bigg\} \\ &= \sum_{i,j} \Big\{ \boldsymbol{x}_{i,j} d_{i,j} - H_0(y_{i,j}) exp\{\eta_i\} \boldsymbol{x}_{i,j} \Big\}. \end{aligned}$$

by the constant term $-ln\{dt_{(m)}\}$. If values for the loglihood (via AIC) are reported there will be no adaptation to this redefinition.

This can be expressed in matrix notation as

$$\frac{\partial l^h}{\partial \boldsymbol{\beta}} = \boldsymbol{X}^T[\boldsymbol{d} - \boldsymbol{\mu}],$$

with

$$\boldsymbol{\mu} = \begin{bmatrix} H_0(y_{1,1})exp\{\eta_{1,1}\} \\ H_0(y_{1,2})exp\{\eta_{1,2}\} \\ H_0(y_{2,1})exp\{\eta_{2,1}\} \\ \vdots \\ H_0(y_{n,2})exp\{\eta_{n,2}\} \end{bmatrix}.$$

The vector $\boldsymbol{\mu} = \boldsymbol{W}_0\boldsymbol{M}\boldsymbol{A}\mathbf{1}$, with

$$\boldsymbol{W}_0 = diag\{exp\{\eta_{1,1}\}, exp\{\eta_{1,2}\}, exp\{\eta_{2,1}\}, \ldots, exp\{\eta_{n,2}\}\},$$

and M being an indicator matrix that shows if an individual (row) was still at risk at the distinct death times (column)

$$\boldsymbol{M} = \begin{bmatrix} \mathcal{I}\{y_{1,1} \geq y_{(1)}\} & \mathcal{I}\{y_{1,1} \geq y_{(2)}\} & \ldots & \mathcal{I}\{y_{1,1} \geq y_{(D)}\} \\ \mathcal{I}\{y_{1,2} \geq y_{(1)}\} & \mathcal{I}\{y_{1,2} \geq y_{(2)}\} & \ldots & \mathcal{I}\{y_{1,2} \geq y_{(D)}\} \\ \vdots & \vdots & \ddots & \vdots \\ \mathcal{I}\{y_{n,2} \geq y_{(1)}\} & \mathcal{I}\{y_{n,2} \geq y_{(2)}\} & \ldots & \mathcal{I}\{y_{n,2} \geq y_{(D)}\} \end{bmatrix},$$

and $\boldsymbol{A} = diag\{h_0^{(1)}dt_{(1)}, h_0^{(2)}dt_{(2)}, \ldots, h_0^{(D)}dt_{(D)}\}$, and $\mathbf{1}$ being a $D \times 1$ vector of ones.

The matrices $\boldsymbol{W}_0$ and $\boldsymbol{M}$ will also be needed for the following derivations.

6.2.2 $\frac{\partial l^h}{\partial \boldsymbol{v}}$

The second set of first-order derivatives is

$$\begin{aligned} \frac{\partial l^h}{\partial v_i} &= \sum_{j=1}^{2} d_{i,j} - \sum_{m=1}^{D} \frac{d_{(m)}}{\sum_{(i^*,j^*)\in R_{(m)}} exp\{\eta_{i^*,j^*}\}} \sum_{j=1}^{2} exp\{\eta_{i,j}\} + \frac{\partial ln\{g_V(v_i)\}}{\partial v_i} \\ &= \sum_{j=1}^{2} \Big\{ d_{i,j} - H_0(y_{i,j})exp\{\eta_{i,j}\} \Big\} + \frac{\partial ln\{g_V(v_i)\}}{\partial v_i}. \end{aligned}$$

Let $\boldsymbol{C}$ be a matrix, indicating cluster membership for each individual, i.e

$$\boldsymbol{C} = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \dots & 1 \\ 0 & 0 & \dots & 1 \end{bmatrix},$$

with units ordered by cluster index. Then,

$$\frac{\partial l^h}{\partial \boldsymbol{v}} = \boldsymbol{C}^T[\boldsymbol{d} - \boldsymbol{\mu}] + \frac{\partial ln\{g_V(\boldsymbol{v})\}}{\partial \boldsymbol{v}}.$$

6.2.3 $-\frac{\partial^2 l^h}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T}$

The negative Hessian will play an important role in the estimation process and its inverse is also the covariance estimate for the estimated fixed and random effects. The negative Hessian of $\boldsymbol{\beta}$ will be called $\boldsymbol{H}_\beta = -\frac{\partial^2 l^h}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T}$. The Hessian of $\boldsymbol{\beta}$ is as follows:

$$\begin{aligned} \boldsymbol{H}_\beta &= \frac{\partial}{\partial \boldsymbol{\beta}^T} \boldsymbol{X}^T \boldsymbol{\mu} \\ &= \boldsymbol{X}^T \begin{bmatrix} \frac{\partial}{\partial \beta_1} H_0(y_{1,1}) exp\{\eta_{1,1}\} & \dots & \frac{\partial}{\partial \beta_K} H_0(y_{1,1}) exp\{\eta_{1,1}\} \\ \frac{\partial}{\partial \beta_1} H_0(y_{1,2}) exp\{\eta_{1,2}\} & \dots & \frac{\partial}{\partial \beta_K} H_0(y_{1,2}) exp\{\eta_{1,2}\} \\ \vdots & \ddots & \vdots \\ \frac{\partial}{\partial \beta_1} H_0(y_{n,2}) exp\{\eta_{n,2}\} & \dots & \frac{\partial}{\partial \beta_K} H_0(y_{n,2}) exp\{\eta_{n,2}\} \end{bmatrix}, \end{aligned} \tag{23}$$

with

$$\begin{aligned} \frac{\partial}{\partial \beta_k} H_0(y_{i,j}) exp\{\eta_{i,j}\} = exp\{\eta_{i,j}\} \sum_{m>0: y_{(m)} \le y_{i,j}} \Bigg[& h_0^{(m)} dt_{(m)} x_{i,j:k} - \\ & \frac{(h_0^{(m)} dt_{(m)})^2}{d_{(m)}} \sum_{(i^*,j^*) \in R_{(m)}} exp\{\eta_{i^*,j^*}\} x_{i^*,j^*:k} \Bigg], \end{aligned} \tag{24}$$

where $x_{i,j:k}$ refers to covariate k of individual (i,j).

Letting $\boldsymbol{W}_1 = diag\{\boldsymbol{\mu}\}$, $\boldsymbol{U} = diag\{\frac{(h_0^{(1)} dt_{(1)})^2}{d_{(1)}}, \frac{(h_0^{(2)} dt_{(2)})^2}{d_{(2)}}, \dots, \frac{(h_0^{(D)} dt_{(D)})^2}{d_{(D)}}\}$. Then, com-

bining matrix notation in (23) and the individual derivative (24) leads to matrix expression

$$\begin{aligned}\frac{\partial^2 l^h}{\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}^T} &= \boldsymbol{X}^T[\boldsymbol{W}_1 - \boldsymbol{M}\boldsymbol{U}\boldsymbol{M}^{\boldsymbol{T}}\boldsymbol{W}_0]\boldsymbol{X} \\ &= \boldsymbol{X}^T\boldsymbol{W}^*\boldsymbol{X}, \end{aligned} \tag{25}$$

with $\boldsymbol{W}^* = \boldsymbol{W}_1 - \boldsymbol{M}\boldsymbol{U}\boldsymbol{M}^{\boldsymbol{T}}\boldsymbol{W}_0$.

6.2.4 $-\frac{\partial^2 l^h}{\partial \boldsymbol{v}\partial \boldsymbol{v}^T}$

The negative Hessian of $\boldsymbol{v}$ will be called $\boldsymbol{H}_v = -\frac{\partial^2 l^h}{\partial \boldsymbol{v}\partial \boldsymbol{v}^T}$. Using some results from the previous section,

$$\begin{aligned}\boldsymbol{H}_v &= \boldsymbol{C}^T\frac{\partial}{\partial \boldsymbol{v}}\boldsymbol{\mu} - \frac{\partial^2 ln\{g_V(\boldsymbol{v})\}}{\partial \boldsymbol{v}\partial \boldsymbol{v}^T} \\ &= \boldsymbol{C}^T\boldsymbol{W}^*\boldsymbol{C} + \boldsymbol{Q}, \end{aligned}$$

with $\boldsymbol{Q} = -diag\{\frac{\partial^2 ln\{g_V(v_1)\}}{\partial v_1^2}, \ldots, \frac{\partial^2 ln\{g_V(v_n)\}}{\partial v_n^2}\}$

6.2.5 $-\frac{\partial^2 l^h}{\partial \boldsymbol{v}\partial \boldsymbol{\beta}^T}$

The negative Hessian $-\frac{\partial^2 l^h}{\partial \boldsymbol{v}\partial \boldsymbol{\beta}^T}$ will be denoted as $\boldsymbol{H}_{v,\beta}$. Note that $\boldsymbol{H}_{v,\beta}^T = \boldsymbol{H}_{\beta,v} = -\frac{\partial^2 l^h}{\partial \boldsymbol{\beta}\partial \boldsymbol{v}^T}$. From the previous two sub-section one can see that

$$\begin{aligned}\boldsymbol{H}_{v,\beta} &= \frac{\partial}{\partial \boldsymbol{\beta}^T}\boldsymbol{C}^T\boldsymbol{\mu} \\ &= \boldsymbol{C}^T\boldsymbol{W}^*\boldsymbol{X}. \end{aligned}$$

6.2.6 All First and Second Order Derivatives

With letting $\boldsymbol{\tau} = \begin{bmatrix}\boldsymbol{\beta} \\ \boldsymbol{v}\end{bmatrix}$, $\boldsymbol{P} = \begin{bmatrix}\boldsymbol{X} & \boldsymbol{C} \\ \boldsymbol{0}_{n\times K} & \boldsymbol{I}_{n\times n}\end{bmatrix}$, $\boldsymbol{d}^* = \begin{bmatrix}\boldsymbol{d} \\ \boldsymbol{0}_{n\times 1}\end{bmatrix}$, $\boldsymbol{\mu}^* = \begin{bmatrix}\boldsymbol{\mu} \\ \boldsymbol{0}_{n\times 1}\end{bmatrix}$ and $\boldsymbol{b} = \begin{bmatrix}\boldsymbol{0}_{2n\times 1} \\ \frac{ln\{g_V(\boldsymbol{v})\}}{\partial \boldsymbol{v}}\end{bmatrix}$, the first order derivatives are (Ha et al., 2017, p. 102)

$$\frac{\partial l^h}{\partial \boldsymbol{\tau}} = \boldsymbol{P}^T[\boldsymbol{d^*} - \boldsymbol{\mu}^* + \boldsymbol{b}]. \tag{26}$$

With $\boldsymbol{V} = \begin{bmatrix}\boldsymbol{W}^* & \boldsymbol{0}_{2n\times n} \\ \boldsymbol{0}_{n\times 2n} & \boldsymbol{Q}\end{bmatrix}$, the negative Hessian is (Ha et al., 2017, p. 80)

$$\boldsymbol{H}_\tau = \boldsymbol{P}^T\boldsymbol{V}\boldsymbol{P}. \tag{27}$$

The inverse negative Hessian $\boldsymbol{H}_{\tau}^{-1}$ will serve as a covariance estimate of $\boldsymbol{\tau}$ (Ha et al., 2017, p. 47).

Note that in all those equations $h_0^{(m)}$, $m = 1, \ldots, D$, is a function of $\boldsymbol{\tau}$, even though this was not explicitly stated by notation.

6.3 Restricted Likelihood for the Variance of the Frailty

This sub-section deals with the restricted maximum Likelihood estimation of $\boldsymbol{\theta} = \theta$. Restricted maximum Likelihood estimation is necessary to obtain an unbiased estimator as a naive inference based on l^h would not account for the information lost through the estimation of $\boldsymbol{\tau}$. Hence, a restricted Likelihood is necessary that is free of both, $\boldsymbol{\beta}$ and $\boldsymbol{v}$.

The restricted Likelihood L^R is approximated by integrating $\boldsymbol{v}$ and $\boldsymbol{\beta}$ out of $L(\boldsymbol{\beta}, \boldsymbol{v}; Y, V)$ and treating it as a function of the formerly fixed θ.[2] (Lee et al., 2017, p. 181) As this function involves high dimensional integrals of dimension $n + K$ in this case, useful approximations need to be found. This is done by a second-order Laplace approximation in this case. The Laplace approximation will be discussed first, then the second-order Laplace approximation is applied to l^h to obtain the (approximated) restricted likelihood function.

6.3.1 Laplace Approximation

The notation used here is entirely unique to this sub-section. None of the previous notation is valid here and none of the notation used in this sub-section will be valid for any other. The results obtained here are general. Its use will be clarified at the end of this sub-section.

This sub-section is based on the univariate second-order Laplace approximation of the integral $I(x) = \int_{-\infty}^{\infty} exp\{x\phi(\tilde{t})\}d\tilde{t}$ as described in (Bender and Orszag, 1978, pp. 261-274) and extended to the multivariate integral $I(x) = \int_{\mathbb{R}^n} exp\{x\phi(\tilde{\boldsymbol{t}})\}d^n\tilde{\boldsymbol{t}}$, with $\tilde{\boldsymbol{t}} \in \mathbb{R}^n$ and $\phi(\tilde{\boldsymbol{t}}) \in \mathbb{R}^1$. At the end of this sub-section $x\phi()$ will be related to the loglihood and $\tilde{\boldsymbol{t}}$ to the random effects. Thus, $I(x)$ will be the basis of the REML.

Let $I(x) = \int_{-\infty}^{\infty} exp\{x\phi(\tilde{\boldsymbol{t}})\}d^n\tilde{t}$, with $x \to \infty$ and $-\infty < \tilde{\boldsymbol{t}}_m < \infty$ maximises $\phi(\tilde{\boldsymbol{t}})$, i.e. $\tilde{\boldsymbol{t}}_m$ is a stationary point. Then,

$$I(x) \approx \int_{\|\tilde{t}_m - \tilde{t}\| \leq \epsilon} exp\{x\phi(\tilde{\boldsymbol{t}})\}d^n t, \tag{28}$$

with ϵ being an arbitrary (small) positive real. This is because $exp\{x\phi(\tilde{\boldsymbol{t}}_m)\}$ will dominate the integral and the remainder is exponentially small. Approximating $\phi(\tilde{\boldsymbol{t}})$ by its fourth

[2] In this thesis there will be no notational difference made between the exact restricted Likelihood function and the approximated one, as the first is not available here.

order Taylor expansion leads to

$$\phi(\tilde{t}) \approx \phi^m + \sum_i t_i \phi_i^{(m)} + \frac{1}{2}\sum_{i,j} t_i t_j \phi_{i,j}^{(m)} + \frac{1}{6}\sum_{i,j,k} t_i t_j t_k \phi_{i,j,k}^{(m)} + \frac{1}{24}\sum_{i,j,k,l} t_i t_j t_k t_l \phi_{i,j,k,l}^{(m)}, \quad (29)$$

where $t_i = [\boldsymbol{t}]_i[\tilde{\boldsymbol{t}} - \tilde{\boldsymbol{t}}_m]_i$ and the number of indices in ϕ refer to the specific derivative evaluated at $\tilde{\boldsymbol{t}}_m$. For example, $\frac{\partial^4 \phi(\tilde{t})}{\partial \tilde{t}_i, \partial \tilde{t}_j, \partial \tilde{t}_k, \partial \tilde{t}_l}|_{\tilde{t}=\tilde{t}_m} = \phi_{i,j,k,l}^{(m)}$. Each of the indices in the sums in (29) runs from 1 to n, i.e. $i, j, k, l = 1, \dots n$.

The Taylor expansion is sharp around a very small ball around $\boldsymbol{t}_m$ and thus can be used to approximate $I(x)$ as in (28): by using (29) in (28) it follows that

$$I(x) \approx exp\{x\phi^{(m)}\} \int_{\|t\| \leq \epsilon} exp\Big\{ x\Big(\frac{1}{2}\sum_{i,j} t_i t_j \phi_{i,j}^{(m)} + \frac{1}{6}\sum_{i,j,k} t_i t_j t_k \phi_{i,j,k}^{(m)} + \frac{1}{24}\sum_{i,j,k,l} t_i t_j t_k t_l \phi_{i,j,k,l}^{(m)} \Big) \Big\} d^n t. \quad (30)$$

Let $y = x(\frac{1}{6}\sum_{i,j,k} t_i t_j t_k \phi_{i,j,k}^{(m)} + \frac{1}{24}\sum_{i,j,k,l} t_i t_j t_k t_l \phi_{i,j,k,l}^{(m)})$. Substituting $exp\{y\}$ by its power series $1+y+\frac{1}{2!}y^2+\frac{1}{3!}y^3+\frac{1}{4!}y^4+\cdots = 1+x(\frac{1}{6}\sum_{i,j,k} t_i t_j t_k \phi_{i,j,k}^{(m)}+\frac{1}{24}\sum_{i,j,k,l} t_i t_j t_k t_l \phi_{i,j,k,l}^{(m)})+ \frac{1}{72}x^2(\sum_{i,j,k} t_i t_j t_k \phi_{i,j,k}^{(m)})^2 + \dots$ and replacing the corresponding values in (30) leads to

$$\begin{aligned} I(x) \approx exp\{x\phi^{(m)}\} \int_{\|t\| \leq \epsilon} & exp\Big\{ x\frac{1}{2}\sum_{i,j} t_i t_j \phi_{i,j}^{(m)} \Big\} \\ & \times \big[1 + x\big(\frac{1}{6}\sum_{i,j,k} t_i t_j t_k \phi_{i,j,k}^{(m)} \\ & + \frac{1}{24}\sum_{i,j,k,l} t_i t_j t_k t_l \phi_{i,j,k,l}^{(m)}\big) + \frac{1}{72}x^2\big(\sum_{i,j,k} t_i t_j t_k \phi_{i,j,k}^{(m)}\big)^2 \\ & + \dots \big] d^n t. \end{aligned} \quad (31)$$

Considering, that all terms with an odd number of terms in t, for example $t_i t_j t_k$, vanish and neglecting all terms captured by the dots in (31), further simplifies the approximation to

$$\begin{aligned} I(x) \approx exp\{x\phi^{(m)}\} \int_{\mathbb{R}^n} & exp\Big\{ -x\frac{1}{2}\boldsymbol{t}^T(-\boldsymbol{\phi}_{\cdot\cdot}^{(m)})\boldsymbol{t} \Big\} \\ & \times \big[1 + \frac{1}{24}x\sum_{i,j,k,l} t_i t_j t_k t_l \phi_{i,j,k,l}^{(m)} \\ & + \frac{1}{72}x^2\big(\sum_{i,j,k} t_i t_j t_k \phi_{i,j,k}^{(m)}\big)^2\big] d^n t, \end{aligned} \quad (32)$$

where $\phi_{\cdot\cdot}^{(m)}$ denotes the matrix of second derivatives evaluated at $\tilde{\boldsymbol{t}}_m$. Note, that the

region of integration was changed to $\mathbb{R}^n$ what can be done because $\phi(\tilde{\boldsymbol{t}}_m)$ still dominates the integral.

Now, (32) is a sum of integrals where each can be integrated by its own and analytical expressions can be found of. The first integral $\int_{\mathbb{R}^n} exp\Big\{ -\frac{1}{2}\boldsymbol{t}^T(-x\boldsymbol{\phi}_{\cdot\cdot}^{(m)})\boldsymbol{t}\Big\}$ is known to be the Gaussian kernel and thus

$$\int_{\mathbb{R}^n} exp\Big\{ -\frac{1}{2}\boldsymbol{t}^T(-x\boldsymbol{\phi}_{\cdot\cdot}^{(m)})\boldsymbol{t}\Big\}d^nt = \frac{(2\pi)^{\frac{n}{2}}}{\sqrt{det\{-x\boldsymbol{\phi}_{\cdot\cdot}^{(m)}\}}}. \tag{33}$$

This result will be needed for the next two terms. What is also needed, is a generating function, with $\boldsymbol{J} = [j_1, \ldots, j_n]^T$,

$$Z(\boldsymbol{J}) = \int_{\mathbb{R}^n} exp\Big\{ -\frac{1}{2}\boldsymbol{t}^T(-x\boldsymbol{\phi}_{\cdot\cdot}^{(m)})\boldsymbol{t} + \boldsymbol{t}^T\boldsymbol{J}\Big\}d^nt \tag{34}$$

such that

$$\frac{\partial^4 Z(\boldsymbol{J})}{\partial J_i, \partial J_j, \partial J_k \partial J_l}|_{\boldsymbol{J}=\boldsymbol{0}_{n\times 1}} = \int_{\mathbb{R}^n} t_i t_j t_k t_l \; exp\Big\{ -\frac{1}{2}\boldsymbol{t}^T(-x\boldsymbol{\phi}_{\cdot\cdot}^{(m)})\boldsymbol{t}\Big\}d^nt. \tag{35}$$

Note, that this is not restricted to the case of the fourth derivative but can be used with any order. With (35), the second integral of (32) can be expressed as

$$\begin{aligned}\int_{\mathbb{R}^n} exp\Big\{ -\frac{1}{2}\boldsymbol{t}^T(-x\boldsymbol{\phi}_{\cdot\cdot}^{(m)})\boldsymbol{t}\Big\} \; \frac{1}{24}x\sum_{i,j,k,l} t_i t_j t_k t_l \phi_{i,j,k,l}^{(m)} d^nt =& \frac{x}{24}\sum_{i,j,k,l}\phi_{i,j,k,l}^{(m)} \\ &\times \frac{\partial^4 Z(\boldsymbol{J})}{\partial J_i, \partial J_j, \partial J_k \partial J_l}|_{\boldsymbol{J}=\boldsymbol{0}_{n\times 1}}. \end{aligned} \tag{36}$$

Hence, next a solution for (35) is derived in order to solve (36): Let $\boldsymbol{y} = \boldsymbol{t} - (-x\boldsymbol{\phi}_{\cdot\cdot}^{(m)})^{-1}\boldsymbol{J}$. Then (34) can be expressed as

$$\begin{aligned} Z(\boldsymbol{J}) &= \int_{\mathbb{R}^n} exp\Big\{ -\frac{1}{2}(\boldsymbol{t} - (-x\boldsymbol{\phi}_{\cdot\cdot}^{(m)})^{-1}\boldsymbol{J})^T(-x\boldsymbol{\phi}_{\cdot\cdot}^{(m)})(\boldsymbol{t} - (-x\boldsymbol{\phi}_{\cdot\cdot}^{(m)})^{-1}\boldsymbol{J}) \\ &\qquad + \frac{1}{2}\boldsymbol{J}^T(-x\boldsymbol{\phi}_{\cdot\cdot}^{(m)})^{-1}\boldsymbol{J}\Big\}d^nt \\ &= \int_{\mathbb{R}^n} exp\Big\{ -\frac{1}{2}\boldsymbol{y}^T(-x\boldsymbol{\phi}_{\cdot\cdot}^{(m)})\boldsymbol{y} + \frac{1}{2}\boldsymbol{J}^T(-x\boldsymbol{\phi}_{\cdot\cdot}^{(m)})^{-1}\boldsymbol{J}\Big\}d^nt \\ &= \frac{(2\pi)^{\frac{n}{2}}}{\sqrt{det\{-x\boldsymbol{\phi}_{\cdot\cdot}^{(m)}\}}} exp\Big\{\frac{1}{2}\boldsymbol{J}^T(-x\boldsymbol{\phi}_{\cdot\cdot}^{(m)})^{-1}\boldsymbol{J}\Big\}, \end{aligned} \tag{37}$$

and with this result (35) can be solved:

$$\frac{\partial^4 Z(\boldsymbol{J})}{\partial J_i, \partial J_j, \partial J_k \partial J_l}|_{\boldsymbol{J}=\mathbf{0}_{n\times 1}} = \frac{(2\pi)^{\frac{n}{2}}}{det\{-x\boldsymbol{\phi}_{\cdot\cdot}^{(m)}\}}\big[(x\boldsymbol{\phi}_{\cdot\cdot}^{(m)})_{i,j}^{-1}(x\boldsymbol{\phi}_{\cdot\cdot}^{(m)})_{k,l}^{-1} + (x\boldsymbol{\phi}_{\cdot\cdot}^{(m)})_{i,k}^{-1}(x\boldsymbol{\phi}_{\cdot\cdot}^{(m)})_{j,l}^{-1} + (x\boldsymbol{\phi}_{\cdot\cdot}^{(m)})_{i,l}^{-1}(x\boldsymbol{\phi}_{\cdot\cdot}^{(m)})_{j,k}^{-1}\big]. \tag{38}$$

Inserting (38) in (36) yields

$$\begin{aligned}\frac{x}{24}\sum_{i,j,k,l}\phi_{i,j,k,l}^{(m)}\frac{\partial^4 Z(\boldsymbol{J})}{\partial J_i, \partial J_j, \partial J_k \partial J_l}|_{\boldsymbol{J}=\mathbf{0}_{n\times 1}} &= \frac{x}{24}\sum_{i,j,k,l}\phi_{i,j,k,l}^{(m)}\frac{(2\pi)^{\frac{n}{2}}}{det\{-x\boldsymbol{\phi}_{\cdot\cdot}^{(m)}\}} \\ \times\big[(x\boldsymbol{\phi}_{\cdot\cdot}^{(m)})_{i,j}^{-1}(x\boldsymbol{\phi}_{\cdot\cdot}^{(m)})_{k,l}^{-1} + (x\boldsymbol{\phi}_{\cdot\cdot}^{(m)})_{i,k}^{-1}(x\boldsymbol{\phi}_{\cdot\cdot}^{(m)})_{j,l}^{-1} + (x\boldsymbol{\phi}_{\cdot\cdot}^{(m)})_{i,l}^{-1}(x\boldsymbol{\phi}_{\cdot\cdot}^{(m)})_{j,k}^{-1}\big] &= \frac{(2\pi)^{\frac{n}{2}}}{det\{-x\boldsymbol{\phi}_{\cdot\cdot}^{(m)}\}}\frac{x}{8}\sum_{i,j,k,l}\phi_{i,j,k,l}^{(m)} \\ &\quad\times(x\boldsymbol{\phi}_{\cdot\cdot}^{(m)})_{i,j}^{-1}(x\boldsymbol{\phi}_{\cdot\cdot}^{(m)})_{k,l}^{-1}.\end{aligned} \tag{39}$$

The third term in the bracket of (32) $\frac{x^2}{72}\big(\sum_{i,j,k} t_i t_j t_k \phi_{i,j,k}^{(m)}\big)^2$ can be derived in exactly the same fashion by applying the sixth order derivative of (37) and results in

$$-\sum_{i,j,k,l,o,p}\phi_{i,j,k}^{(m)}\phi_{l,o,p}^{(m)}\big[\frac{(x\boldsymbol{\phi}_{\cdot\cdot}^{(m)})_{i,j}^{-1}(x\boldsymbol{\phi}_{\cdot\cdot}^{(m)})_{k,l}^{-1}(x\boldsymbol{\phi}_{\cdot\cdot}^{(m)})_{o,p}^{-1}}{8} + \frac{(x\boldsymbol{\phi}_{\cdot\cdot}^{(m)})_{i,l}^{-1}(x\boldsymbol{\phi}_{\cdot\cdot}^{(m)})_{j,o}^{-1}(x\boldsymbol{\phi}_{\cdot\cdot}^{(m)})_{k,p}^{-1}}{12}\big]. \tag{40}$$

Using (33), (39) and (40) to get an expression for the approximation to the integral $I(x)$ as expressed in (32) yields

$$I(x) \approx exp\{x\phi(m)\}\frac{(2\pi)^{\frac{n}{2}}}{\sqrt{det\{-x\boldsymbol{\phi}_{\cdot\cdot}^{(m)}\}}}\big[1 + F^{(m)}\big], \tag{41}$$

with $F^{(m)}$ being equal to the sums of equations (39) and (40).

Considering that the interest lays in the loglihood l^R and the approximation of $I(x)$ will be used to get an approximation of L^R, the logarithm of (41) will be taken, where yet another approximation can be made: $ln\{1+\epsilon\} \approx \epsilon$, if $\epsilon \to 0$. As $x^{-1} \to 0$ in the inverse matrices in the expressions of (40) and (39),this can be applied to $ln\{1+F\}$, i.e

$$ln\{I(x)\} \approx \phi^{(m)} + \ln\left\{\frac{(2\pi)^{\frac{n}{2}}}{\sqrt{det\{-x\boldsymbol{\phi}_{\cdot\cdot}^{(m)}\}}}\right\} + F^{(m)}. \tag{42}$$

Now, (42) will be related to l^R after which the notation of this sub-sub-section is

invalid for the remainder of this thesis and the previous notation starts to kick in again:

- $x\phi(\tilde{\boldsymbol{t}}) = l^h$
- $\tilde{\boldsymbol{t}} = \boldsymbol{v}$
- $\phi^{(m)}_{i,j,k,l} = \frac{\partial^4 l^h}{\partial v_i \partial v_j \partial v_k \partial v_l}|_{\boldsymbol{v}=\hat{\boldsymbol{v}}}$, and so on.

Hence,

$$ln\Big\{ \int_{-\infty}^{\infty} L^h d^n v \Big\} \approx \Big[l^h + ln\Big\{ \frac{(2\pi)^{\frac{n}{2}}}{\sqrt{det\{\boldsymbol{H}_{\boldsymbol{v}}\}}} \Big\} + F \Big]_{\boldsymbol{v}=\hat{\boldsymbol{v}}}, \tag{43}$$

with

$$\begin{aligned} F =& \frac{1}{8} \sum_{i,j,k,l} \frac{\partial^4 l^h}{\partial v_i, \partial v_j \partial v_k \partial v_l} (\boldsymbol{H}_{\boldsymbol{v}})^{-1}_{i,j} (\boldsymbol{H}_{\boldsymbol{v}})^{-1}_{k,l} \\ & - \sum_{i,j,k,l,o,p} \frac{\partial^3 l^h}{\partial v_i \partial v_j \partial v_k} \frac{\partial^3 l^h}{\partial v_l, \partial v_o \partial v_p} \\ & \quad \Big[\frac{(\boldsymbol{H}_{\boldsymbol{v}})^{-1}_{i,j} (\boldsymbol{H}_{\boldsymbol{v}})^{-1}_{k,l} (\boldsymbol{H}_{\boldsymbol{v}})^{-1}_{o,p}}{8} + \frac{(\boldsymbol{H}_{\boldsymbol{v}})^{-1}_{i,l} (\boldsymbol{H}_{\boldsymbol{v}})^{-1}_{j,o} (\boldsymbol{H}_{\boldsymbol{v}})^{-1}_{k,p}}{12} \Big] \end{aligned}$$

(Lee et al., 2017, p.181).

For l^R the remaining issue is to condition on the sufficient statistics $\hat{\boldsymbol{\beta}}$ which is achieved by enhancing the approximation (43) to $\boldsymbol{\beta}$. (Ha et al., 2017, p. 52,76) This is only done for the first order Laplaceian approximation as the dimension of $\boldsymbol{\beta}$ does not increase with the data, i.e.

$$l^R = \Big[l^h + ln\Big\{ \frac{(2\pi)^{\frac{n}{2}}}{\sqrt{det\{\boldsymbol{H}_{\boldsymbol{\beta},\boldsymbol{v}}\}}} \Big\} + F \Big]_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}},\ \boldsymbol{v}=\hat{\boldsymbol{v}}}, \tag{44}$$

where $\hat{\boldsymbol{v}}$ and $\hat{\boldsymbol{\beta}}$ maximise l^h. Note, that $\boldsymbol{H}_{\boldsymbol{v}}$ in F is not changed to $\boldsymbol{H}_{\boldsymbol{\tau}}$.

6.3.2 Derivatives

This chapter discusses the first-order derivative of l^r (Ha et al., 2017, p. 103)

$$\frac{\partial l^R}{\partial \theta} = \Big[\frac{\partial l^h}{\partial \theta} - \frac{1}{2} trace\Big\{ \boldsymbol{H}^{-1}_{\boldsymbol{\tau}} \frac{\partial \boldsymbol{H}_{\boldsymbol{\tau}}}{\partial \theta} \Big\} - \frac{\partial F}{\partial \theta} \Big]_{\boldsymbol{\tau}=\hat{\boldsymbol{\tau}}},$$

where it must be considered that $l^h|_{\boldsymbol{\tau}=\hat{\boldsymbol{\tau}}} = l^h(\hat{\boldsymbol{\tau}}, \theta)$ and $\hat{\boldsymbol{\tau}} = \hat{\boldsymbol{\tau}}(\theta)$ are functions of θ. Hence,

$$\begin{aligned}\frac{\partial l^h}{\partial \theta}|_{\boldsymbol{\tau}=\hat{\boldsymbol{\tau}}} &= \frac{\partial l^h(\hat{\boldsymbol{\tau}}, \theta)}{\partial \theta} + \frac{\partial l^h(\boldsymbol{\tau}, \theta)}{\partial \boldsymbol{\tau}}|_{\boldsymbol{\tau}=\hat{\boldsymbol{\tau}}} \frac{\partial \hat{\boldsymbol{\tau}}(\theta)}{\partial \theta} \\ &= \frac{\partial l^h}{\partial \theta}|_{\boldsymbol{\tau}=\hat{\boldsymbol{\tau}}} \\ &= \frac{\partial ln\{g_V(\boldsymbol{v})\}}{\partial \theta}|_{\boldsymbol{v}=\hat{\boldsymbol{v}}}, \end{aligned} \tag{45}$$

because $\frac{\partial l^h}{\partial \boldsymbol{\tau}}|_{\boldsymbol{\tau}=\hat{\boldsymbol{\tau}}} = 0$ and in l^h, θ is only contained through g_V. The derivative $\frac{\partial ln\{g_V(\boldsymbol{v})\}}{\partial \theta}$ cannot be derived in general as it depends on the specific choice of distribution.

In the frailtyHL package the term $\frac{\partial \hat{\boldsymbol{v}}}{\partial \theta}$ is included but $\frac{\partial \hat{\boldsymbol{\beta}}}{\partial \theta}$ is ignored. Then, $\frac{\partial \boldsymbol{H}_{\boldsymbol{\tau}}}{\partial \theta}|_{\boldsymbol{\tau}=\hat{\boldsymbol{\tau}}}$ becomes

$$\frac{\partial \boldsymbol{H}_{\boldsymbol{\tau}}}{\partial \theta}|_{\boldsymbol{\tau}=\hat{\boldsymbol{\tau}}} = \boldsymbol{P}^T \begin{bmatrix} \frac{\partial \boldsymbol{W}^*}{\partial \theta}|_{\boldsymbol{\tau}=\hat{\boldsymbol{\tau}}} & \mathbf{0} \\ \mathbf{0} & \frac{\partial \boldsymbol{Q}}{\partial \theta} \end{bmatrix} \boldsymbol{P},$$

with

$$\begin{aligned}\frac{\partial \boldsymbol{W}^*}{\partial \theta}|_{\boldsymbol{\tau}=\hat{\boldsymbol{\tau}}} &= \left[\frac{\partial \boldsymbol{W}^*}{\partial \theta} + \frac{\boldsymbol{W}^*}{\partial \boldsymbol{v}}\frac{\partial \hat{\boldsymbol{v}}}{\partial \theta}\right]_{\boldsymbol{\tau}=\hat{\boldsymbol{\tau}}} \\ &= \left[\frac{\boldsymbol{W}^*}{\partial \boldsymbol{v}}\frac{\partial \hat{\boldsymbol{v}}}{\partial \theta}\right]_{\boldsymbol{\tau}=\hat{\boldsymbol{\tau}}}, \end{aligned}$$

and, according to (Lee and Nelder, 1996, pp. 105-106),

$$\frac{\partial \hat{\boldsymbol{v}}}{\partial \theta} = -\boldsymbol{H}_{\boldsymbol{v}}^{-1} \frac{\partial^2 ln\{g_V(\boldsymbol{v})\}}{\partial \boldsymbol{v} \partial \theta}|_{\boldsymbol{v}=\hat{\boldsymbol{v}}}.$$

The derivative of F and the second-order derivatives will not be discussed here.

6.4 Iterative Optimisation Procedure

The parameter estimates $\hat{\boldsymbol{\tau}}$ and $\hat{\theta}$ are obtained on the basis of the Newton-Raphson Procedure. (Ha et al., 2017, p. 78)

The Newton-Raphson procedure relies on a second order approximation of l^h, i.e.

$$\begin{aligned} l^h(\boldsymbol{\tau}; Y, V) \approx l^h(\boldsymbol{\tau}_0; Y, V) + [\boldsymbol{\tau} - \boldsymbol{\tau}_0]^T \boldsymbol{S}_{\boldsymbol{\tau}}(\boldsymbol{\tau}_0) - \frac{1}{2}[\boldsymbol{\tau} - \boldsymbol{\tau}_0]^T \boldsymbol{H}_{\boldsymbol{\tau}}(\boldsymbol{\tau}_0)[\boldsymbol{\tau} - \boldsymbol{\tau}_0] \\ = l_0^h, \end{aligned}$$

where $\boldsymbol{S}_{\boldsymbol{\tau}}(\boldsymbol{\tau}_0)$ and $\boldsymbol{H}_{\boldsymbol{\tau}}(\boldsymbol{\tau}_0)$ represent the first order derivatives and the negative Hessian evaluated at $\boldsymbol{\tau}_0$.

Then, $\boldsymbol{\tau}_+ = \underset{\boldsymbol{\tau}}{argmax}\ l_0^h$ leads to

$$\begin{aligned} & \frac{\partial l_0^h}{\partial \boldsymbol{\tau}} \overset{!}{=} 0 \\ \implies & \boldsymbol{S}_{\boldsymbol{\tau}}(\boldsymbol{\tau}_0) - \boldsymbol{H}_{\boldsymbol{\tau}}(\boldsymbol{\tau}_0)[\boldsymbol{\tau}_+ - \boldsymbol{\tau}_0] = 0 \\ \implies & \boldsymbol{H}_{\boldsymbol{\tau}}(\boldsymbol{\tau}_0)\boldsymbol{\tau}_+ = \boldsymbol{H}_{\boldsymbol{\tau}}(\boldsymbol{\tau}_0)\boldsymbol{\tau}_0 + \boldsymbol{S}_{\boldsymbol{\tau}}(\boldsymbol{\tau}_0) \end{aligned} \tag{46}$$

In this case, (46) is not pre-multiplied by $\boldsymbol{H}_{\boldsymbol{\tau}}^{-1}$ in order to solve for $\boldsymbol{\tau}_+$ to avoid the computation of the inverse. (Ha et al., 2017, p. 26) This is not done in the case of θ. (Ha et al., 2017, p. 103)

The equation (46) can be expressed as

$$\boldsymbol{P}^T\boldsymbol{V}\boldsymbol{P}\boldsymbol{\tau} = \boldsymbol{P}\boldsymbol{y}^*, \tag{47}$$

with $\boldsymbol{y}^* = \boldsymbol{V}\boldsymbol{P}\boldsymbol{\tau}_0 + \boldsymbol{d}^* + \boldsymbol{b}^* - \boldsymbol{\mu}^*$.

In similar fashion, the variance parameter θ will be calculated by the common Newton Raphson equation

$$\theta_+ = H_\theta^{-1}(\theta_0)S_\theta(\theta_0) + \theta_0. \tag{48}$$

Iteration takes place over the equations (47) and (48). The values $\boldsymbol{\tau}_0$ and θ_0 either indicate values from the previous iteration or initialized values: $\mathbf{0}_{K\times 1}$ or 0.1 respectively. The values $\boldsymbol{\tau}_+$ and θ_+ either indicate values for the the next iteration or $\hat{\boldsymbol{\beta}}$ and $\hat{\theta}$, if $\boldsymbol{\tau}_+ - \boldsymbol{\tau}_0 < 10^{-5}\mathbf{1}_{K\times 1}$ and $\theta_+ - \theta_0 < 10^{-5}$ in which case the procedure converged. (Ha et al., 2017, p. 80)

6.5 Differences to the coxph-Implementation

In estimating $\boldsymbol{\tau}$ both approaches use l^h. (Therneau et al., 2003, p. 158) One of the main differences is the estimation of θ, where the coxph package uses a marginal likelihood approach. That is, θ is estimated via $l^m = ln\left\{ \int_{-\infty}^{\infty} exp\{l^h(\boldsymbol{\beta}, \boldsymbol{v}; Y, V)\}dv \right\}$. In the case of a gamma distributed random term Z, this is done analytically. (Therneau et al., 2003, p. 160) In the case of a lognormal distributed random term Z this is done by a first order Laplace approximation with respect to V (Ripatti and Palmgren, 2000, p.1017), which is as in (44) but without F and using $\boldsymbol{H}_{\boldsymbol{v}}$ instead of $\boldsymbol{H}_{\boldsymbol{\beta},\boldsymbol{v}}$. The consequence is, that a maximum Likelihood estimator instead of a REML estiamtor for θ is obtained in both cases. (Therneau et al., 2003, p.161) The coxph implementation also ignores $\frac{\hat{\boldsymbol{v}}}{\theta}$ in l^r. (Ha et al., 2017, p. 77) According to Ha et al. (2017) this can lead to underestimation of θ especially when the cluster size is small.(p. 77)

A further simplification is made in the coxph package as $\boldsymbol{H_v}$ is simplified to be a diagonal matrix which speeds up computation especially if dimensions of the Hessian are large, i.e. if there are many clusters in the data at hand. (Therneau et al., 2003, p. 164)

Slight differences are also present in the iterative Newton-Raphson procedure. Given a value for θ, a couple of Newton-Raphson iterations are calculated to get an estimate for $\boldsymbol{\tau}$. Then it is returned to the outer Newton Raphson loop, where a new value for θ is calculated and so on. (Therneau et al., 2003, p. 159)

6.6 Model Selection

Model selection will be based on the Likelihood. For testing the presence of frailty the Likelihood Ratio Test (LRT) will be used. For comparisons across models, the Akaike Information Criterion (AIC) will be used. This subsection explains both, starting with the latter.

6.6.1 AIC

The AIC is standard tool for model selection in statistics. (Fahrmeier et al., 2013, p. 148) Its foundation lies in the Kullback Leibler Divergence. (Ha et al., 2017, p. 87) Consider the true density $\xi(y)$ of the data y and some density $\kappa(y)$. The $KLD = E_\xi[ln\{\xi(y)\} - ln\{\kappa(y)\}]$, where E_ξ denotes, that the expectation is taken over the real model. If κ is ξ, then there is no lost information by using κ and the $KLD = 0$. The model κ usually involves MLE's from the data $\hat{\psi} = \hat{\psi}(y)$. Hence, a useful measure of lost information is

$$\begin{aligned} E_\xi[KLD] &= E_{\xi(y)}[E_{\xi(y^*)}[ln\{\xi(y^*)\} - ln\{\kappa_{\hat{\psi}(y)}(y^*)\}]] \\ &= E_{\xi(y^*)}[ln\{\xi(y^*)\}] - E_{\xi(y)}[E_{\xi(y^*)}[ln\{\kappa_{\hat{\psi}(y)}(y^*)\}]], \end{aligned} \tag{49}$$

with y^* being another potential draw from the same RV. The first term of (49) is irrelevant to model comparison as it is merely a constant for any model. Considering only the second term multiplied by 2 is the Akaike Information. (Ha et al., 2017, p. 88) It measures the lost information by a chosen model. The Akaike Information is estimated by the AIC:

$$AIC = -2ln\{\kappa_{\hat{\psi}(y)}(y)\} + 2\tilde{K}, \tag{50}$$

where $\tilde{K}$ is the number of free parameters in the model. The AIC is an asymptotically unbiased estimator of the Akaike Information. (Ha et al., 2017, p. 88)

The AIC considered in this model is based on the conditional model l^p (partial h-loglihood), i.e.

$$AIC = -2l^p + 2df,$$

where $df = \boldsymbol{H_\tau^{-1}} \frac{-\partial^2 l^p}{\partial \tau \partial \tau^T}$ are the effective parameters. Note, that this excludes any kind of direct Likelihood contributions from θ. This measure is called conditional AIC by Ha et al.(2017, p.88) and simply AIC in Therneau (2003, p.163-164). The (conditional) AIC selects the model giving the best conditional predictions. A smaller AIC value indicates a better model with respect to conditional prediction.

6.6.2 LRT

Testing the hypothesis $H^0 : \theta = 0$ will be done with the LRT. The LRT test statistic (LR) equals $-2(l_0^r - l^r)$, where l_0^r is the restricted maximum likelihood under the null hypothesis, i.e. $\boldsymbol{v} = \mathbf{0}$ and $\theta = 0$. This essentially reduces to the first order Laplace approximation to the integral $ln\left\{ \int_{-\infty}^{\infty} exp\{l_0^p\} d\boldsymbol{\beta} \right\}$, which is given by $\left[l_0^p + ln\left\{ \frac{(2\pi)^{\frac{n}{2}}}{\sqrt{det\{\boldsymbol{H_\beta}\}}} \right\} \right]_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}}$, with l_0^p being the partial h-loglihood evaluated at $\boldsymbol{v} = \mathbf{0}$. Care has to be taken as the null hypothesis is on the boundary of the parameter space. Hence, a mixture distribution has to be used for p-values. The test statsitic is distributed as $\frac{1}{2}\chi_0^2 + \frac{1}{2}\chi_1^2$, where χ_0^2 has a point mass on 0. So the p-value can be calculated as $P(\frac{1}{2}\chi_0^2 + \frac{1}{2}\chi_1^2 > LR) = \frac{1}{2}P(\chi_0^2 > LR) + \frac{1}{2}P(\chi_1^2 > LR) = \frac{1}{2}P(\chi_1^2 > LR)$. (Ha et al., 2017, p. 80-81)

7 FrailtyModels

The models that were investigated are those of

$$Z_i \sim \pi(\theta_{sex_i,zyg_i}),$$

with Z_i being independent of Z_{i^c}. The conditional hazard is modelled as $h_{i,j|Z}(t|z_i) = z_i exp\{birth_i \beta_{sex_i,zyg_i}\} h_0(t)^{(sex_i,zyg_i)}$ for all i, j, where $h_0^{(sex_i,zyg_i)}$ indicates a separate baseline hazard depending on the combination of sex and zygosity. This, basically, led to four seperate models: female & monozygotic (fmono), female & dizygotic (fdi), male & monozygotic (mmono) and male & dizygotic (mdi). The distiribution π is either gamma or log-normal.

7.1 Gamma Frailty Model

The Gamma Frailty Model assumes

$$Z_i \overset{iid}{\sim} Gamma(\alpha, \theta), \text{ for } i = 1, \ldots, n.$$

The density of Z_i therefore is $g_Z(z) = \frac{z^{\alpha-1} exp\{-\frac{z}{\theta}\}}{\Gamma(\alpha)\theta^\alpha}$, where Γ is the Gamma function. The expectation of a Gamma distributed RV equals $\alpha\theta$ and its variance is $\alpha\theta^2$. As an identification constraint α is set to $\frac{1}{\theta}$ and thus $E[Z_i] = 1$ and $Var[Z_i] = \theta$. (Duchateau and Janssen, 2008, p. 44) The interpretation is that a twin-pair has a higher conditional hazard if $z_i > 1$ than is expected from the entire population at time origin and the expected deviation from $E(Z_i) = 1$ is $\sqrt{\theta}$. Hence, if θ is big the conditional hazard rate varies strongly and twins are more alike in their survival-times than unrelated people. The identification assumptions simplifies the density of Z_i to $g_Z(z) = \frac{z^{\frac{1}{\theta}-1} exp\{-\frac{z}{\theta}\}}{\Gamma(\frac{1}{\theta})\theta^{\frac{1}{\theta}}}$. With respect to the H-Likelihood approach, the density $g_V(v) = g_Z(exp\{v\})\frac{\partial z(v)}{\partial v} = g_Z(exp\{v\})exp\{v\} = \frac{exp\{v\}^{\frac{1}{\theta}} exp\{-\frac{exp\{v\}}{\theta}\}}{\Gamma(\frac{1}{\theta})\theta^{\frac{1}{\theta}}}$ is required to compute all necessary derivatives, like the Score function and the Hessian, for optimisation.

With respect to marginal distributions and sub-population hazard, however, it is easier to resort on well-known results of the gamma distribution. The univariate sub-population Survivor function $S_{i,j}$ can be found using well known results for the Laplace function when g_Z is the Gamma pdf, leading to $S_{i,j}(t) = \mathcal{L}(H_{0,i}(t)) = \frac{1}{1+\theta H_{0,i}(t)}^{\frac{1}{\theta}}$.(Aalen et al., 2008, p. 237) Equivalently, the bivariate Survivor function $S_{i,\cdot}$ can be established as $S_{i,\cdot}(t_1, t_2) = \mathcal{L}(H_{0,i}(t_1) + H_{0,i}(t_2)) = \frac{1}{[1+\theta(H_{0,i}(t_1)+H_{0,i}(t_2))]}^{\frac{1}{\theta}}$. The negative first-order deriva-

tive of $\mathcal{L}(H_{0,i}(t))$ then delivers the univariate sub-population density of $T_{i,j}$

$$\begin{aligned} f_{i,j}(t) &= -\frac{\partial \mathcal{L}(H_{0,i}(t))}{\partial t} \\ &= (1+\theta H_{0,i}(t))^{-\frac{1}{\theta}-1} h_{0,i}(t) \end{aligned}$$

and the bivariate sub-population density can be found by

$$\begin{aligned} f_{i,\cdot}(t_1,t_2) &= \frac{\partial^2 \mathcal{L}(H_{0,i}(t_1)+H_{0,i}(t_2))}{\partial t_1 \partial t_2} \\ &= (1+\theta)[1+\theta(H_{0,i}(t_1)+H_{0,i}(t_2))]^{-\frac{1}{\theta}-2} h_{0,i}(t_1) h_{0,i}(t_2) \end{aligned}$$

respectively.

In case of gamma distributed frailty the joint Survival and density function can be used to derive an analytic expression for Kendall's τ. Note, that τ is derived based on pairs with identical covariate information. (Duchateau and Janssen, 2008, p. 123) Hence, $f_{i,\cdot}(t_1,t_2) = f_{i^c,\cdot}(t_1,t_2)$ and - by applying the change of variable technique twice and integration by parts once - similar to (Duchateau and Janssen, 2008, p. 125,138-139),

$$\begin{aligned} \tau &= 2 \times 2 \int_0^\infty \int_0^\infty f_{i,\cdot}(t_1,t_2) \left[\int_0^{t_1} \int_0^{t_2} f_{i,\cdot}(r_1,r_2)\, dr_1\, dr_2 \right] dt_1\, dt_2 - 1 \\ &= 4 \int_0^\infty \int_0^\infty S_{i,\cdot}(t_1,t_2) f_{i,\cdot}(t_1,t_2)\, dt_1\, dt_2 - 1 \\ &= 4 \int_0^\infty \int_0^\infty \frac{(1+\theta) h_{0,i}(t_1) h_{0,i}(t_2)}{[1+\theta(H_{0,i}(t_1)+H_{0,i}(t_2))]^{\frac{2}{\theta}+2}}\, dt_1 dt_2 - 1 \\ &= 4 \int_0^\infty \int_0^\infty \frac{(1+\theta)}{[1+\theta(\underbrace{H_{0,i}(t_1)+H_{0,i}(t_2)}_{=\tilde{H}})]^{\frac{2}{\theta}+2}}\, dH(t_1) dH(t_2) - 1 \\ &= 4 \int_0^\infty \int_0^{\tilde{H}} \frac{(1+\theta)}{[1+\theta\tilde{H}]^{\frac{2}{\theta}+2}}\, dH(t_1) d\tilde{H} - 1 \\ &= 4 \int_0^\infty \tilde{H} \frac{(1+\theta)}{[1+\theta\tilde{H}]^{\frac{2}{\theta}+2}}\, d\tilde{H} - 1 \qquad \text{(applying integration by parts ...)} \\ &= 4 \int_0^\infty \frac{1+\theta}{[1+\theta\tilde{H}]^{\frac{2}{\theta}+1}} [\theta+2]^{-1}\, d\tilde{H} - 1 \\ &= -4 \left[\frac{1}{2} \frac{1+\theta}{[1+\theta\tilde{H}]^{\frac{2}{\theta}}}\, \frac{1}{2+\theta} \right]_0^\infty - 1 \\ &= \frac{2+2\theta}{2+\theta} - 1 \\ &= \frac{\theta}{2+\theta}. \end{aligned} \tag{51}$$

The cross-ratio function can also be derived analytically:

$$\begin{aligned}\zeta_i(t_1,t_2) &= \frac{\overbrace{(1+\theta)(1+\theta(H_{0,i}(t_1)+H_{0,i}(t_2))^{-\frac{1}{\theta}-2}h_{0,i}(t_1)h_{0,i}(t_2)}^{f_{i,\cdot}(t_1,t_2)}}{\underbrace{(1+\theta(H_{0,i}(t_1)+H_{0,i}(t_2)))^{-\frac{1}{\theta}-1}\ h_{0,i}(t_1)}_{-\frac{\partial\ S_{i,\cdot}(t_1,t_2)}{\partial\ t_1}}} \\ &\times \frac{\overbrace{(1+\theta(H_{0,i}(t_1)+H_{0,i}(t_2))^{-\frac{1}{\theta}}}^{S_{i,\cdot}(t_1,t_2)}}{\underbrace{(1+\theta(H_{0,i}(t_1)+H_{0,i}(t_2)))^{-\frac{1}{\theta}-1}\ h_{0,i}(t_2)}_{-\frac{\partial\ S_{i,\cdot}(t_1,t_2)}{\partial\ t_2}}} \\ &= 1+\theta.\end{aligned}$$

In the gamma frailty model, ζ_i is not a function of time and can easily be calculated once the variance is estimated.

7.1.1 Comparison of Estimation Approaches

Now, the different datasets will be examined with the gamma frailty model. Focus is first, on the differences between the frailtyHL and coxph estimates for each feature and second, on the interpretation of the estimated values. The measure $\hat{\gamma}$ is an estimate of $\frac{\sqrt{\theta}}{E[Z_i]}$ which is the standard deviation in the gamma case. This measure will not be discussed in this subsection but is needed for a comparison with the estimates from the lognormal model.

Table 2: Gamma Model: frailtyHL vs coxph

Variable	**fmono**	**fdi**	**mmono**	**mdi**
Birth^{HL}	-0.025***	-0.020***	-0.020***	-0.015***
	(0.0069)	(0.0046)	(0.0068)	(0.0044)
Birth^{coxph}	-0.025***	-0.020***	-0.020***	-0.015***
	(0.0065)	(0.0044)	(0.0066)	(0.0043)
$\hat{\theta}^{HL}$	0.613***	0.363***	0.489***	0.236***
	(0.2548)	(0.1578)	(0.1934)	(0.1274)
$\hat{\theta}^{coxph}$	0.534	0.196	0.440	0.201
$\hat{\gamma}^{frailtyHL}$	0.783	0.603	0.700	0.486
$\hat{\gamma}^{coxph}$	0.731	0.442	0.663	0.448
total observations	844	1638	808	1510
number of events	280	534	323	611

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

With respect to $\beta_{sex,zyg}$ it can be seen from Table 2 that the estimates are identically

(when rounded to three significant digits). Also, the estimated standard errors are pretty close. A difference can be seen, however, where the frailtyHL package claims to be more accurate: All estimated variances $\hat{\theta}_{sex,zyg}$ of the frailty RVs are bigger than their coxph counterparts. If the claim of frailtyHL is justified, this could hint on underestimation of the importance of genetics in the case of the twin's lifetimes when using procedures with less accuracy. According to intuition, in both estimation procedures, the variances in the fmono and mmono models are far bigger than those of fdi and mdi. However, there is a big difference between the variance in the fdi model in the two estimation procedures (0.363 vs 0.196). The difference in mdi is much smaller (0.236 vs 0.201). Also the difference between the variance of fdi and mdi under the frailtyHL regime is high (0.363 vs 0.236). Such a jump in the variance cannot be seen under the coxph regime (0.196 vs 0.201). In fact, the variance on the fdi model is even slightly smaller than in the mdi model.

In the following, interpretation relies on the estimates of the frailtyHL package unless otherwise mentioned.

With respect to the impact of the year of birth on expected survival times, it can be seen that $\beta_{f,mono} > \beta_{m,mono}$ and $\beta_{f,di} > \beta_{m,di}$ indicating that female mortality declined more heavily over the years than that of males. However, all confidence intervals of the slope parameters overlap on 0.95 level (not shown). Ceteris paribus, the conditional hazard $h_{i,j|Z}$ is estimated to decline by the factor 0.975 if a monozygotic female is one year younger. In contrast, this is only 0.985 for dizygotic males. The cross-ratio function is 1.613 for monozygotic females for every combination of the two time variables. Given the highly varying estimated ORs in chapter 4, where the risk was especially high around the diagonal and small far away from it, this does not seem to be a good representation of reality. This also holds for the other sub-samples.

Kendall's τ, as estimated by (51), is 0.235 for monozygotic females, lower for monozygotic males (0.197) and smallest in both dizygotic sub-populations (0.154 for females and 0.106 for males). In the coxph approach the order switches only for diyzgotic males and females. This structure cannot be found when using the entire sample where $\hat{\tau}_{m,mono} > \hat{\tau}_{f,mono} > \hat{\tau}_{f,di} > \hat{\tau}_{m,di}$ (or $\hat{\theta}_{m,mono} > \hat{\theta}_{f,mono} > \hat{\theta}_{f,di} > \hat{\theta}_{m,di}$) as can be seen in Hougaard (2000, p. 307) or from Table 3. The estimates of τ are all bigger than the (cohort adjusted) estimates of τ in chapter 4. This is also the case when using the entire sample except for dizygotic males.

In general, relatively high variance of the frailties means that the twins are highly dependent on each other and that there is a lot of difference in the population between different twins as there are relatively many high and low valued frailties that a pair shares. The frailty variance (at a given point in time) also reflects variability in mortality patterns across clusters. A high variance of the frailty distribution, however, does not indicate low or high mortality on the population level in general, as mortality is also governed by the

Table 3: Gamma Model: Entire Sample with coxph

Variable	**fmono**	**fdi**	**mmono**	**mdi**
Birth^{coxph}	-0.026***	-0.027***	-0.017***	-0.017***
	(0.0037)	(0.0025)	(0.0035)	(0.0023)
$\hat{\theta}^{coxph}$	0.389	0.226	0.464	0.120
$\hat{\gamma}^{coxph}$	0.624	0.475	0.681	0.346
total observations	2896	5512	2732	4976
number of events	904	1761	1069	1999

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

baseline hazard.

The impact of the frailty variance will further be analysed by comparing the first quartile and third quartile of the frailties. All comparison are certeris paribus comparisons, meaning, that everything is identical between the two hypothetical individuals, except the frailty value. As a standardised measure of importance, the ratio $\frac{h_{i,j|Z}(t|z^{3rd})}{h_{i,j|Z}(t|z^{1st})} = \frac{z^{3rd}}{z^{1st}}$, where z^{3rd} and z^{1st} indicate the third and the first quartile of the frailty distribution, will be used. This will be called quartiles ratio (QR) from now on. Interpretation is as follows: from the perspective of twins at $t = 0$ on the first quartile of the frailty distribution, there is a share of 0.25 in the population who have at least $\frac{z^{3rd}}{z^{1st}}$ times the risk at any t to die in the very next moment, given that they have survived up to t. The expression $\hat{z}^{1st}$, for example, will refer to the estimated first quartile of the frailty distribution by using $\hat{\theta}$ as the parameter for the gamma distribution.

This first quartile and third quartile for monozygotic females is $\hat{z}^{1st} = 0.427$ and $\hat{z}^{3rd} = 1.363$ respectively. At $t = 0$, there is an estimated share of 0.25 in the population who have at least $\frac{h_{i,j|Z}(t|\hat{z}^{3rd})}{h_{i,j|Z}(t|\hat{z}^{1st})} = 1.363/0.427 = 3.19$ times the conditional hazard than twins on the first quartile.

This is much more than in the estimate of the coxph procedure, where the quartiles ratio reduces to $1.352/0.463 = 2.92$, from the perspective of the same lucky twins as above. This means, that in the frailtyHL estimation the mortality patterns are more pronounced by individual frailties than in the coxph estimation, where the impact of the frailties on the baseline hazard is less variable in the gamma model.

This can be seen in figure 4 which shows the conditional hazard rate for twins born in 1900 ($birth_{i,j} = 0$) on the first quartile, the median and the third quartile of the frailty distribution. The individual on the first quartile is more off from the median in the case of frailtyHL what can best be seen around the value of 0.8.

Looking at the difference $[S^{HL}_{i,j|Z}(t|z^{1st}) - S^{HL}_{i,j|Z}(t|z^{3rd})] - [S^{cox}_{i,j|Z}(t|z^{1st}) - S^{cox}_{i,j|Z}(t|z^{3rd})]$ reveals that the spread in the conditional survivor function between the first and third

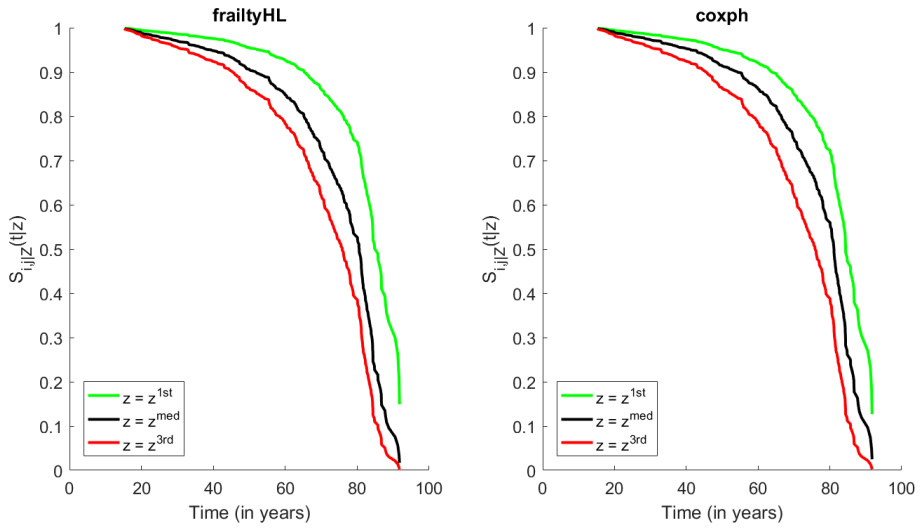

Figure 4: Conditional Survivor functions for an individual born in 1900, female, monozygotic and with the (gamma) frailites taking the value of the first quartile, median, and third quartile

quartile of the frailty distribution, is bigger in the frailtyHL approach than in the coxph approach. There are differences up to the second significant digit as seen in figure 5.

Differences are not visible in the sub-population Survivor function $S_{i,j}(t)$ (sub-population: female, monozygotic, twins born in 1900). The difference between the two Survivor functions is negligible with differences only up to lower values in the third significant digit (figure 6).

The QR of dizygotic females is much smaller (2.36) than that of monozygotic females (3.19), indicating, that the female dizygotic twins are much less depended, as they do not share that much from the gene pool as monozygotic females do. This is also mirrored by the males where the quartiles ratio is 2.77 for monozygotic twins and 1.97 for dizygotic twins.

Figure 7 shows the estimated frailties for monozygotic females for both estimation approaches. For the gamma density, the corresponding point estimate is used as "true" parameter. None of the theoretical distributions seems to be a particularly good fit. In particular, the empirical distribution is much more narrow around the reference value of $z = 1$ and the theoretical distribution is particularly bad in representing very small values of the frailties. This might be an indication that the frailties are not generated from a gamma distribution. Further on, the empirical $\hat{\boldsymbol{z}}$ look quite similar, with slightly more variation in the frailtyHL approach ($\hat{var}(\hat{z}) = 0.140$) than in the coxph approach ($\hat{var}(\hat{z}) = 0.114$). Slightly different patterns can be seen after the value of 1.

If the frailties are robust against the choice of distribution, as this seems to be the case here, this opens up the question of why one should use the theoretical quantiles and point estimates of the frailty variance parameter for analytic purposes as happened here rather

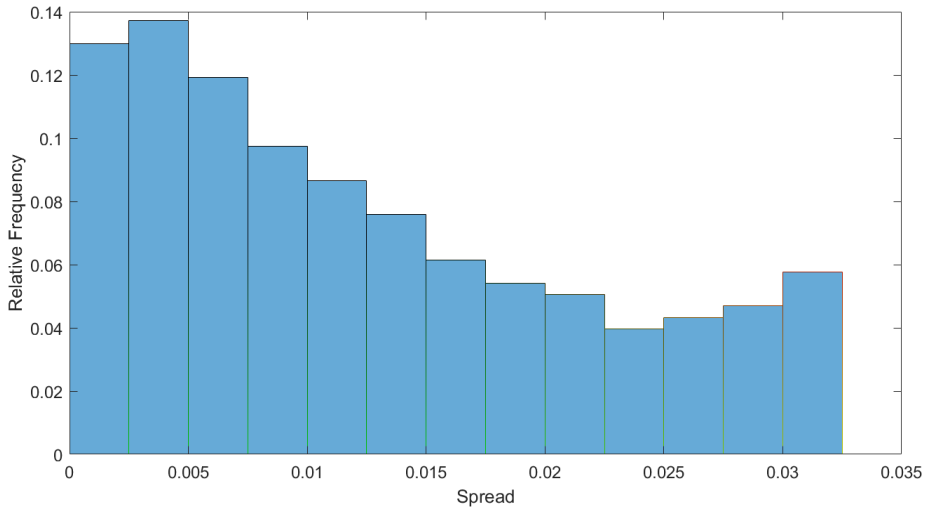

Figure 5: $Spread = [S^{HL}_{i,j|Z}(t|z^{1st}) - S^{HL}_{i,j|Z}(t|z^{3rd})] - [S^{cox}_{i,j|Z}(t|z^{1st}) - S^{cox}_{i,j|Z}(t|z^{3rd})]$

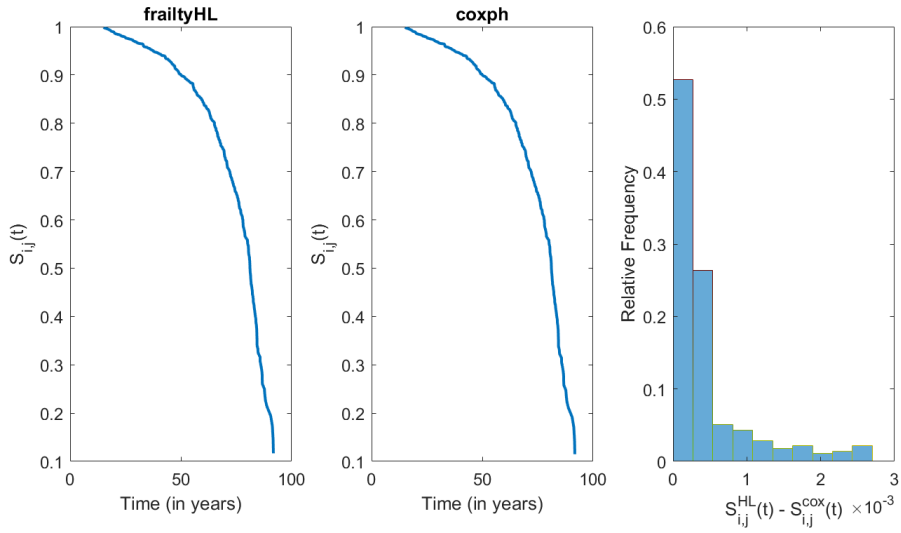

Figure 6: Sub-population Survivor function for monzygotic females born in 1900. Historgram shows $S^{HL}_{i,j}(t) - S^{cox}_{i,j}(t)$.

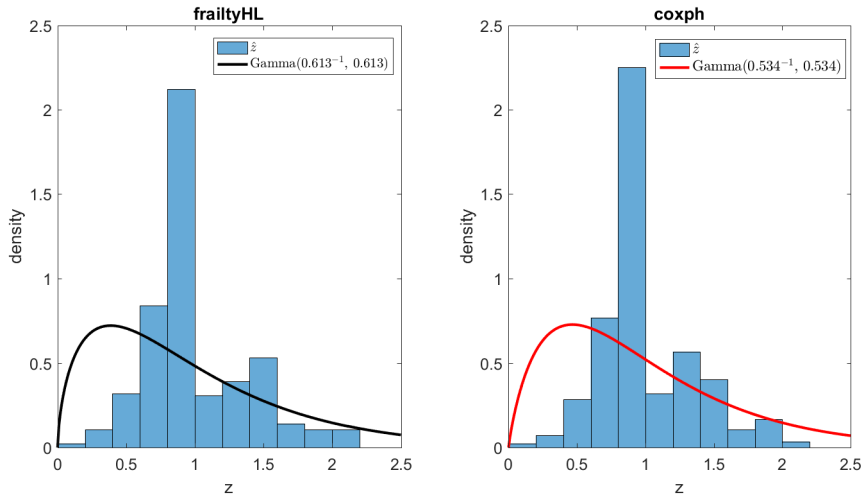

Figure 7: Density of frailites of monozygotic females for both estimation approaches

than using the empirical quantiles and variance estimate. A profound counterargument of using empirical quartiles and the variance of $\hat{\boldsymbol{z}}$ is that those measures do not consider the insecurity induced by $E[(\hat{\boldsymbol{v}} - \boldsymbol{v})(\hat{\boldsymbol{v}} - \boldsymbol{v})^T]$.

Estimates were also calculated for the models

$$Z_i \sim \pi(\theta_{zyg_i}),$$

and the correspondingly $h_{i,j|Z}(t|z_i) = z_i exp\{birth_i \beta_{zyg_i}^{(birth)} + sex_i \beta_{zyg_i}^{(sex)}\} h_0^{(zyg_i)}(t)$, i.e. separating only monzygotic (model: mono) and dizygotic (di) twins from each other. As mentioned previously, the performance of the models will be evaluated in terms of the AIC, with the estimates from coxph in brackets. The fmono and mmono model combine for an AIC of $3119.882 + 3631.994 = 6751.876$ ($3119.367 + 3632.09 = 6751.457$) the mono model had an AIC of 7582.59 (7587.39). This indicates a better conditional prediction if the sexes are estimated separately.

The same holds for the dizygotic twins: The joint model had an AIC of 15905.29 (15922.94) whereas the two separate models combine for only $6736.407 + 7591.075 = 14327.48$ ($6735.383 + 7591.039 = 14326.42$). This is probably caused by different mortality patterns between the sexes and hence, a joint baseline hazard and proportional hazards are not a good fit to reality. Additionally, the truncation as occurred in this dataset affects the frailty distribution at the beginning of the study and hence, a joint frailty distribution and again, a joint baseline hazard is not appropriate. (Hougaard, 2000, p. 300)

As the degree in dependence differs substantially between monozygotic and dizygotic twins a joint model for all individuals does not make sense, and was thus, not calculated.(Hougaard, 2000, p. 300)

A significant test for the frailty variance resulted in a rejection of the hypothesis

that $\theta = 0$ for any of the models: the LR test statistic for fmono was 14.71 (p-value: 6.25×10^{-5}), 12.11 for fdi (p-value: 2.50×10^{-4}), 13.85 for mmono (p-value: 9.88×10^{-5}) and 7.30 for mdi (p-value: 3.45×10^{-3}).

7.2 Log-normal Frailty Model

The Log-normal frailty model assumes

$$Z_i \overset{iid}{\sim} \mathcal{LN}(\mu, \sigma^2), \text{ for } i = 1, \ldots, n$$

(Duchateau and Janssen, 2008, p. 195). It follows that

$$V_i \overset{iid}{\sim} \mathcal{N}(\mu, \sigma^2), \text{ for } i = 1, \ldots, n,$$

and $g_v(\boldsymbol{v})$, as used in the likelihood function, is the product of individual normal p.d.f.'s with identical μ and σ. As an identification constraint μ is set to zero. The estimated parameter σ^2 is the variance of the frailties on the log-level. If $\theta = Var[Z_i]$, then $\theta = exp\{\sigma^2\}(exp\{\sigma^2\} - 1)$. Note, that from the identification restriction follows that $E[Z_i] = exp\{\frac{\sigma^2}{2}\} \neq 1$, unless $\sigma = 0$. Unfortunately, there is no analytical solution to calculate the unconditional Survivor or hazard functions and numerical (integration) procedures must be applied(Duchateau and Janssen, 2008, p. 196). This was not done here.

7.2.1 Comparison of Estimation Procedures & Distribution Assumptions

Now, the different datasets will be examined with the log-normal frailty model. Focus is, again, first, on the differences between the frailtyHL and coxph estimates and then, on the interpretation of the estimated values. The results are constantly compared to those of the gamma models.

Again, $\beta_{sex,zyg}$ is practically identical between the two methods, with only very slight differences between the estimators of mmono and mdi in the third significant digit (table 4). The same holds for estimated standard errors. Also, the size of the impact of the year of birth on one's hazard is similar to the gamma frailty model.

With respect to the estimators of σ^2, however, the picture is turned completely upside down. Now the estimators of coxph are all bigger than their frailtyHL counterparts. Also, the jump in the variance between the fdi and the mdi model, that could previously be seen in the frailtyHL approach but not in the coxph estimation, is now present in the coxph estimation, while less intense for frailtyHL.

As expected, the variances of both monozygotic models are bigger than the variances in the dizygotic models for both estimation procedures, as has been the case in the gamma model. In contrast to the gamma model, a pattern can also be seen across the sexes in

Table 4: Log-normal Model: frailtyHL vs coxph

Variable	**fmono**	**fdi**	**mmono**	**mdi**
Birth^{HL}	-0.023***	-0.019***	-0.019***	-0.015***
	(0.0065)	(0.0045)	(0.0065)	(0.0043)
Birth^{coxph}	-0.023***	-0.019***	-0.018***	-0.014***
	(0.0066)	(0.0046)	(0.0066)	(0.0044)
$\hat{\sigma}^{2\ HL}$	0.403	0.236	0.354	0.171
	(0.1774)	(0.1079)	(0.1473)	(0.0961)
$\hat{\sigma}^{2\ coxph}$	0.578	0.383	0.481	0.285
$\hat{\theta}^{HL}$	0.742	0.337	0.605	0.222
$\hat{\theta}^{coxph}$	1.395	0.683	1.000	0.437
$\hat{\gamma}^{frailtyHL}$	0.704	0.516	0.652	0.432
$\hat{\gamma}^{coxph}$	0.885	0.683	0.786	0.574
total observations	844	1638	808	1510
number of events	280	534	323	611

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

both estimation approaches: The variance of the female monozygotic twins is bigger than that of monozygotic males and so is the variance of dizygotic females bigger than the variance of dizygotic males. A feature that was only present in the frailtyHL estimators in the gamma model but not in the coxph approach.

Table 5: Log-normal Model: Entire Sample with coxph

Variable	**fmono**	**fdi**	**mmono**	**mdi**
Birth^{coxph}	-0.025***	-0.026***	-0.015***	-0.016***
	(0.0037)	(0.0025)	(0.0035)	(0.0023)
$\hat{\sigma^2}^{coxph}$	0.407	0.254	0.478	0.187
$\hat{\theta}^{coxph}$	0.755	0.373	0.988	0.248
$\hat{\gamma}^{coxph}$	0.709	0.538	0.782	0.453
total observations	2896	5512	2732	4976
number of events	904	1761	1069	1999

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Table 5 shows the estimates from the entire sample using coxph. As in the gamma model, the variance of monozygotic males is bigger than that of monozygotic females. The variances are all more moderate than in the coxph estimation with the sub-sample but still bigger as the estimates from frailtyHL.

The cross-distribution comparison shows, that the estimated variances in the lognormal model are bigger than their counterparts in the gamma frailty model for any subset

of the data and for both estimation procedures. However, it is not straightforward to compare these estimators between different distributions and distribution assumptions. This is especially important in the comparison of the lognormal frailty model and the gamma frailty model, as the former does not restrict $E[Z_i]$ to be unity but the latter does. Hence, a standard deviation (or variance) of Z_i has a different meaning relative to $E[Z_i]$. And this is especially important in the case of hazard modelling, as any restriction in $E[Z_i]$ merely redefines the scale of the baseline hazard but does not impact inference. Therefore, table 4 also shows an estimator of the relative standard deviation $\gamma = \frac{\sqrt{\theta}}{E[Z_i]}$ for each model.

For the relative standard deviation, the picture is again completely upside down. While the frailtyHL approach estimates less relative variation in the lognormal model than in the gamma model, the coxph approach shows an increase of relative variation: For any of the subsets $\gamma^{frailtyHL}_{sex,zyg}$ is smaller in the lognormal model than in the gamma model. However, for any of the subsets $\gamma^{coxph}_{sex,zyg}$ is bigger in the lognormal model than in the gamma model. But this is still only one measure of different distributions which, in particular, does not make the tail behaviour of the estimated frailty distributions visible. If one enriches the analysis of the relative standard deviation with the QR measure, examination is more thoroughly: For monozygotic females, the QR reduces to 2.35 in the lognormal model compared to 3.19 in the gamma model for the frailtyHL approach. This is contrasted by a QR of 2.79 in the coxph lognormal model versus 2.92 in the gamma model. This illustrates the shortcoming of analysing the relative standard deviation for itself. Even though the relative standard deviation is bigger in the lognormal frailty model than in the gamma model, the QR gets smaller. This is caused by the tail behaviour of the two corresponding distributions, where the lognormal density incorporates more mass on its right tail, as can be seen in figure 8. A fair comparison between the two distributions requires a "standardisation" of Z_i in the lognormal case. This was done by dividing Z_i with $exp\{\frac{\sigma^2}{2}\}$, leading to the parameters $\mu^* = -\frac{\sigma^2}{2}$ and $\sigma^2 = \sqrt{ln\{\gamma_{f,mono} + 1\}}$, such that $E[Z_i] = 1$ and $Var[Z_i] = \gamma^2_{f,mono}$. The corresponding parameter estimates of the two different approaches were plugged in. The figure also shows the first and third quartiles of the corresponding estimations. Note, that the QR is independent of μ as the quantile function of the lognormal distribution equals $exp\{\mu + \sqrt{2\sigma^2} erf^{-1}(2F - 1)\}$ and, consequently, μ cancels out in the ratio.

In the coxph estimation the Gamma and the Lognormal distribution cross at the value $z = 2.94$ and from then on, the Lognormal distribution is always above the Gamma distribution. From that point on, there is 0.014 more probability mass in the lognormal distribution than in the gamma distribution. That means, the increase in variation in the coxph estimation comes from "outliers" with a particular bad genetic make-up, while the majority of the female monozygotic twins are estimated to be closer to each other as

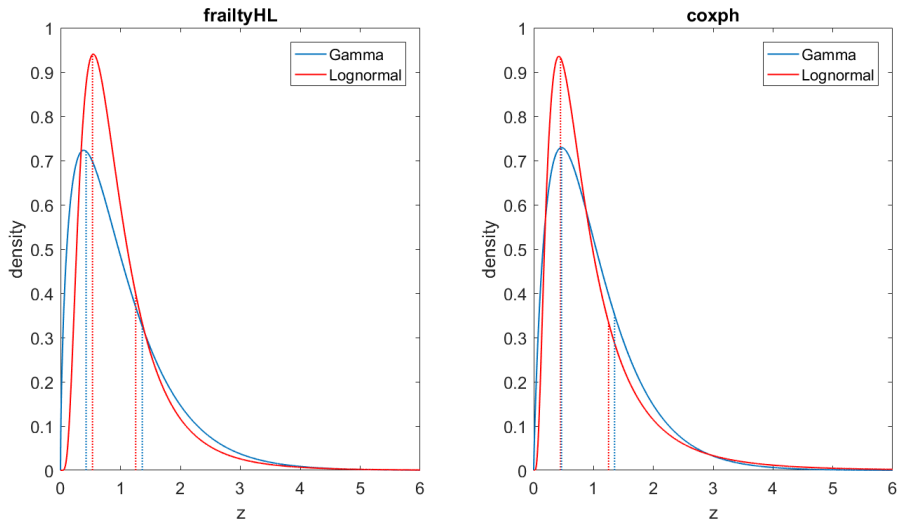

Figure 8: Estimated & "standardised" Lognormal distribution (left) vs the Gamma distribution of the frailtyHL approach. Coxph counterpart (right). Dotted lines are the first and the third quartile.

indicated by the QR. This is a notable feature as it is known that some individuals with specific genes like the breast cancer genes BRCA1 and BRCA2 have a much larger risk (Antoniou and Easton, 2006) and might, therefore, be one of the outliers. This effect is less pronounced in the frailtyHL approach. The distributions cross at $z = 4.77$ and from then on the density of the lognormal estimate is slightly bigger: From that point on there is only 0.0009 more probability mass in the lognormal distribution. So the effect of more outliers in the lognormal distribution starts to kick in later and is less pronounced as in the coxph estimation. Summed up a higher relative variance in the lognormal model leads more mass for outliers but not necessarily to a higher QR.

Figure 9 shows again the individual hazards of monozygotic female twins born in 1900. It can be seen that the difference to the median is now more pronounced in the coxph approach.

Looking at the difference $[S^{HL}_{i,j|Z}(t|z^{1st}) - S^{HL}_{i,j|Z}(t|z^{3rd})] - [S^{cox}_{i,j|Z}(t|z^{1st}) - S^{cox}_{i,j|Z}(t|z^{3rd})]$ reveals that the spread in the conditional survivor function between the first and third quartile of the frailty distribution, is bigger in the coxph approach than in the frailtyHL approach (figure 10). The differences are also much bigger with up to 0.08 in absolute value compared to the spread in the gamma model. The estimates of the lognormal model of the two approaches are considerably different.

Figure 11 shows the density of the estimated frailties versus the theoretical distribution with $\hat{\sigma}^2$ as value for the log-frailty variance. Though the theoretical densities in both approaches are still far off especially around the value $z = 0$. This seems to be less intense as for the gamma model. Further on, the theoretical density is a much better fit in the tail of the coxph approach. This might be a valuable feature if this captures the

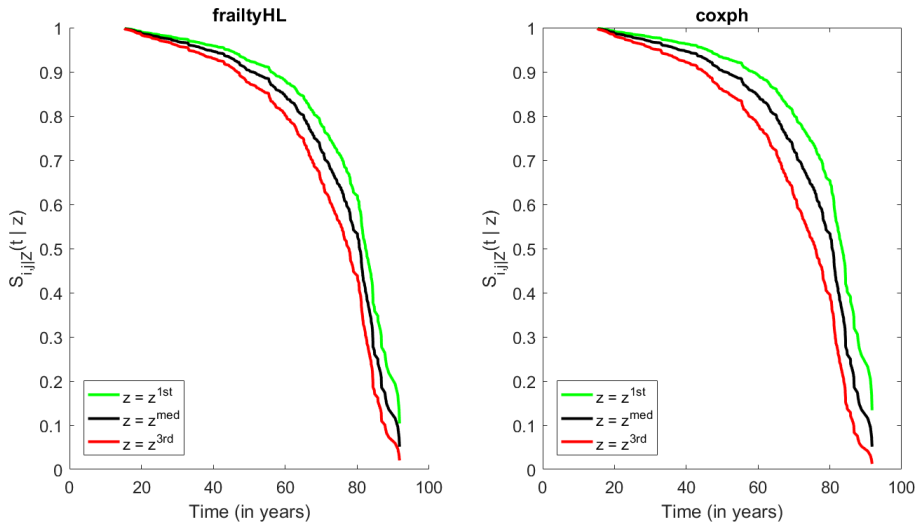

Figure 9: Conditional Survivor functions for an individual born in 1900, female, monozygotic and (log-normal) frailites taking the value of the first quartile, median, and third quartile

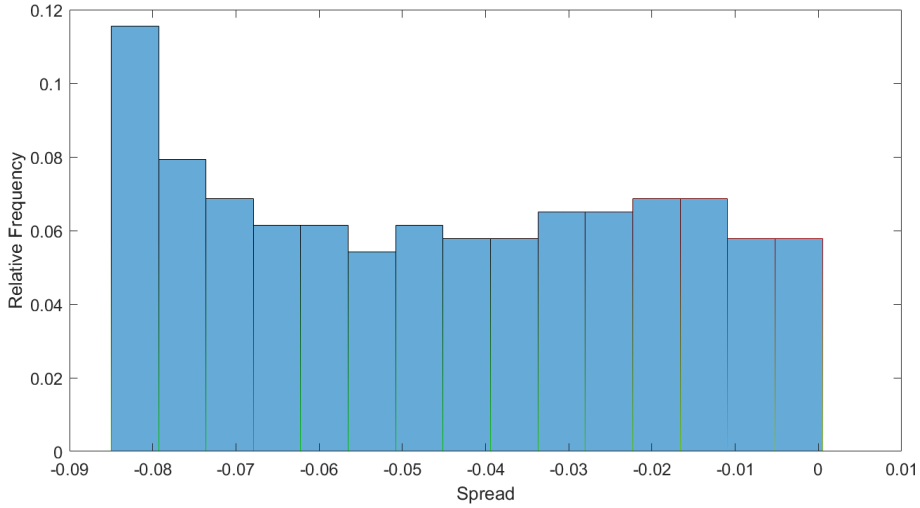

Figure 10: $Spread = [S^{HL}_{i,j|Z}(t|z^{1st}) - S^{HL}_{i,j|Z}(t|z^{3rd})] - [S^{cox}_{i,j|Z}(t|z^{1st}) - S^{cox}_{i,j|Z}(t|z^{3rd})]$

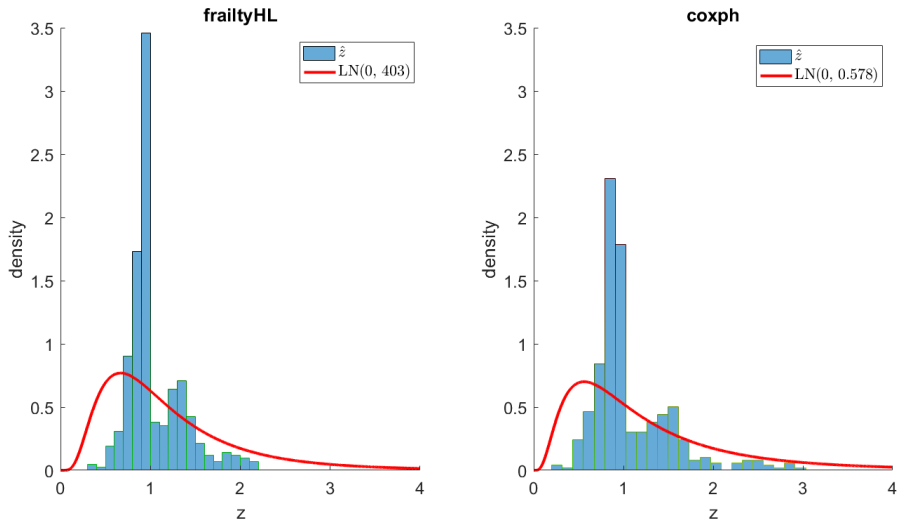

Figure 11: Density of the (log-normal) frailites of monozygotic females for both estimation approaches

true "outliers".

A rough visual inspection supports this idea. Figure 12 shows the estimated frailties against the survival times[3]: a trend is clearly visible.

And it is further supported by figure 13 which shows the estimated (log-centred) frailties [4] against the survival times for the gamma model (coxph estimation). The biggest value is $\hat{z}^{max} = 2.156$. In the lognormal model, some of the low survival times have much higher frailties. The highest value is $\hat{z}^{max} = 3.027$.

The remaining analysis of the QR of the other subsets do not differ substantially from that of the gamma model, in the sense that the QR of dizygotic females (frailtyHL: 1.92, coxph: 2.30) is smaller than that of monozygotic females and this also holds for monozygotic males (frailtyHL: 2.23, coxph: 2.54) and dizygotic males (frailtyHL: 1.75, coxph: 2.05).

As for the gamma model, estimates were also calculated for the models

$$Z_i \sim \pi(\theta_{zyg_i}),$$

and the correspondingly $h_{i,j|Z}(t|z_i) = z_i exp\{birth_i\beta_{zyg_i} + sex_i\beta_{zyg_i}^{(sex)}\}h_0^{(zyg_i)}(t)$, i.e. separating only monzygotic (model: mono) and dizygotic (di) twins from each other. The results of coxph can again be found in brackets. The fmono and mmono model combine for an

[3]Note that both twins are part of those plots. That means that every frailty value is present twice. In some cases, the points overlap, especially when both are censored. There is a rare case where both of the twins died at the same time ($\hat{z} = 2.91$ (lognormal) and $t_i = 27.59$ years). That is why on $\hat{z} = 2.91$ three dots can be seen in the vertical (and not four).

[4]Here the "standardisation" was taken in the opposite direction. The mean of $\boldsymbol{v}$ was deducted from $\boldsymbol{v}$ from the gamma model to have the same restriction as in the lognormal model. Then the centred frailties were exponentiated.

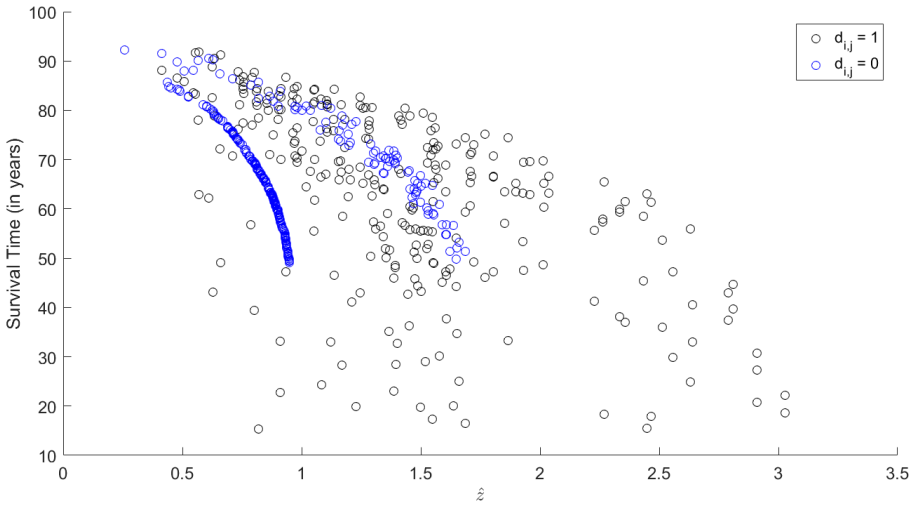

Figure 12: Estimated frailties of the coxph approach from the lognormal model against survival time

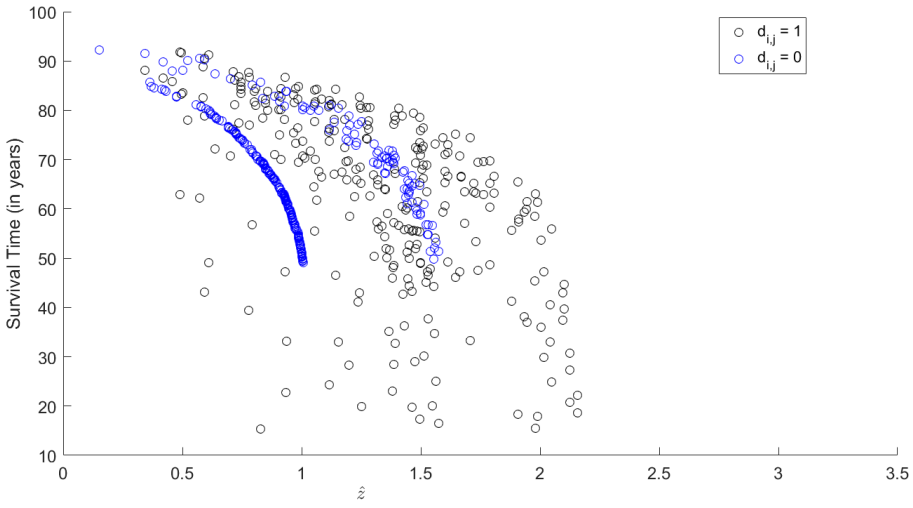

Figure 13: Estimated (log-centered) frailties of the coxph approach from the gamma model against survival time

AIC of 3108.897+3624.382 = 6733.279 (3102.423+3619.311 = 6721.734) the mono model had an AIC of 7565.171 (7556.311). This indicates a better conditional prediction if the sexes are estimated separately.

The same holds for the dizygotic twins: The joint model had an AIC of 15887.33 (15881.86) whereas the two separate models combine for only 6723.461 + 7584.363 = 14307.82 (6713.859 + 7575.725 = 14289.58).

Other models were again not considered.

A cross frailty distribution comparison shows that the lognormal frailty distribution led to a better fit in any case (estimates from the coxph approach are again in brackets):

- $AIC^{LN}_{f,mono} - AIC^{gamma}_{f,mono} = -10.985(-16.944)$
- $AIC^{LN}_{f,di} - AIC^{gamma}_{f,di} = -12.946(-21.524)$
- $AIC^{LN}_{m,mono} - AIC^{gamma}_{m,mono} = -7.612(-12.779)$
- $AIC^{LN}_{m,di} - AIC^{gamma}_{m,di} = -6.712(-15.314)$

Burnham and Anderson (2002) argue that a model which has a bigger AIC by the value of 10 or more has essentially no support as it fails to explain an essential part of the explainable variation in the model (p.71). The only models that do not cross this threshold are mmono and mdi in frailtyHL approach (but the lognormal model still being the better one). Particularly bad is the gamma fdi model of coxph estimation. This is the only estimation of dizygotic females where a jump in variation compared to dizygotic males cannot be seen. In all remaining models, the dependency of females was bigger than that of males with respect to their counterpart in zygosity status (and estimation approach).

The better AIC might be caused by a better fit of the lognormal model for extreme cases, i.e. twins with particularly high frailty.

The significance tests (frailtyHL) for the frailty variance resulted again in a rejection of the hypothesis that $\theta = 0$ for any of the models: the LR test statistic for fmono was 8.74 (p-value: 1.56×10^{-3}), 2.71 for fdi (p-value 3.05×10^{-3}), 9.09 mmono (p-value: 1.28×10^{-3}), 4.934 mdi (p-value: 1.32×10^{-3}).

Lastly, an overview of the findings in this chapter is given: The cross-estimation approach comparison showed practically no difference in the fixed effect. Also, the structure of the order of frailty variances across the different models were the same with one exception, namely the ordering of the variance between the fdi and the mdi model were different in the gamma model. The frailty variances were bigger in the frailtyHL approach than in the coxph estimation for the gamma model but smaller in the lognormal model. The cross-distribution comparison showed that the lognormal frailty distribution gives a better. This is most likely caused by the tail behaviour as the QR's were more closely

together in the lognormal model but the lognormal model allowed for more extreme values (big frailties). From the content-related perspective, all models gave the same interpretation: (unsurprisingly) genetics matter. That can be seen from the frailty variances. The monozygotic variances were bigger than the dizygotic variances, i.e. monozygotic twins are more similar than dizygotic twins. That does not mean that genetics matter less for dizygotic twins but as they share fewer genes they are less similar in survival times.

8 Simulation Study

To evaluate the performance of the applied estimation methods 1000 datasets were simulated. Each dataset has $n = 350$ clusters with a cluster size of 2. The simulation scheme is as follows:

Firstly, a covariate matrix $\boldsymbol{X}$ with size of 700×2 was simulated. Both covariates were drawn from a uniform distribution with a lower bound of -20 and an upper bound of $+20$. The first covariate $x_i^{(ind)}$ is individual-specific. In the models that will be estimated later on, this covariate will be left out. This was done to introduce some reality in the analysis of the simulated data, as it seems implausible that the lifetime of the twins in the real world example is entirely governed by the given cluster-specific covariates. The second covariate $x_i^{(cluster)}$, however, is a cluster-specific covariate and mimics the year of birth, i.e. for each cluster of artificial twins a single value was drawn. The values of $x^{(cluster)}$ were also rounded to whole numbers. The survival times (in unit days) $T_{i,j}$ of individuals were then drawn from a Weibull distribution with individual scale parameter $\lambda_{i,j} = \lambda_0 exp\{-(\boldsymbol{x}_{i,j}^T\boldsymbol{\beta} + ln\{z_i\})/\alpha_T\}$, with $\lambda_0 = 365 \times 80$, and shape parameter $\alpha_T = 5$. This parametrisation was chosen to have an easy partition of the conditional hazard into a baseline hazard and an individual part that fits the style of modelling in the semi-parametric approach. This will be discussed further below. The vector of slope parameters $\boldsymbol{\beta} = \begin{bmatrix} 0.0125 & -0.015 \end{bmatrix}^T$. The variable z_i is an iid realisation from a lognormal distribution with parameters $\mu = 0,\ \sigma = 0.5584849$, such that the median of Z_i is 1, $E[Z_i] = 1.169$ and $Var[Z_i] = 0.5$. Note, that $ln\{Z_i\}$ is normally distributed. For each of the 1000 datasets a single set of $\boldsymbol{z}$ was drawn.

For the first 1000 estimations, the censoring time was set to 72 (72×365.25 days) years for all individuals. This is in contrast to the real world twin example where a pair that were born one year earlier was observed one year longer (if they did not die previously). This is the high-censoring setting. For the second set of (1000) estimations, the censoring time was set to 90 years. This is the low-censoring setting. The aim is to examine whether there are systematic differences in performance between the approaches across high-censored and low-censored datasets.

Figure 14 shows the true conditional Survivor function for an individual with $\eta_i = 0$. Median survival time is about 74 years for this individual (1^{st}quartile $\approx$ 62 years, 3^{rd} quartile $\approx$ 85 years). Given the distribution of x^{ind}, $x^{cluster}$ and v_i and the values for $\boldsymbol{\beta}$, i is almost an "average" pair. It can be seen that the first censoring time is slightly before the median and consequently, the censoring rates vary around 53%. In contrast, the censoring rate for the second simulation run is slightly behind the third quartile and the censoring rates for those datasets vary around 20%. The censoring rates of each 1000 datasets are depicted in figure 15.

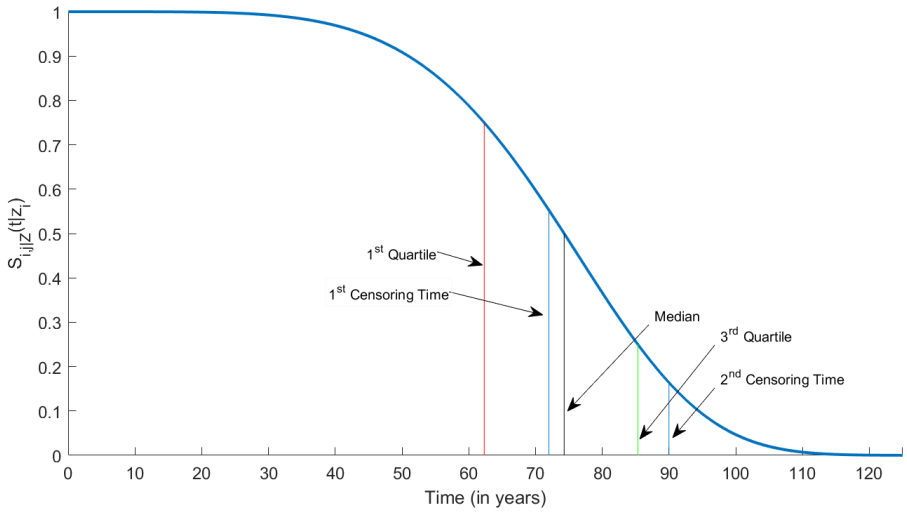

Figure 14: Survivor function for an individual with $\eta_i = 0$ with first quartile, median, third quartile survival-time and both censoring rates.

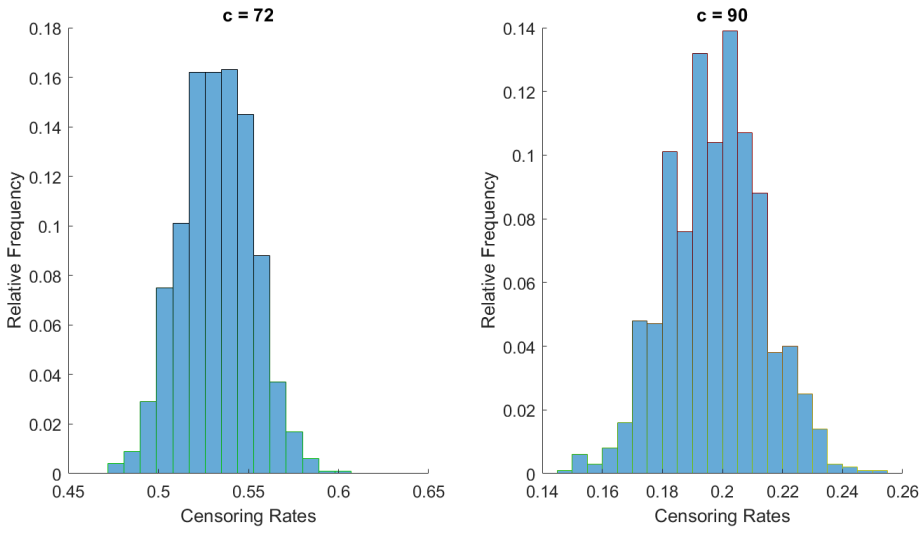

Figure 15: Censoring rates for high- (left) and low-censoring (right) setting.

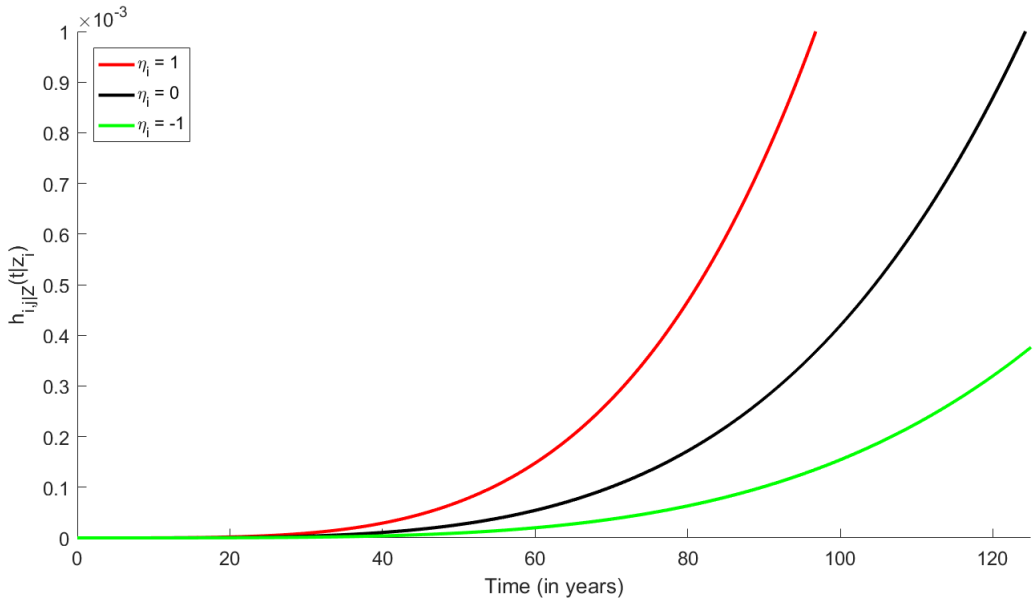

Figure 16: Conditional hazard rate

The conditional hazard of the simulated data fits in the proportional hazard framework, as

$$\begin{aligned} h_{i,j|V}(t|v_i) &= exp\{\boldsymbol{x}_{i,j}^T + v_i\}\frac{\alpha_T t^{\alpha_T - 1}}{\lambda_0^{\alpha_T}} \\ &= exp\{\boldsymbol{x}_{i,j}^T + v_i\}h_0(t) \\ &= z_i h_{i,0}(t). \end{aligned}$$

The true conditional hazard is identical to that of proportional hazards modelling as conducted here.

Hence, the true model of the simulated data is favourable to the estimated model. This is because, first, the proportional hazards assumption with respect to the FE and RE is correct and second, the distribution assumption of RE is also the correct choice. On the other hand, the true model is unfavourable to the estimated model because the true model includes $\boldsymbol{x}^{(ind)}$ and $\boldsymbol{x}^{(cluster)}$ whereas the estimated model only include $\boldsymbol{x}^{(cluster)}$. Also, the true baseline hazard is not modelled in the estimation but its non-parametric estimate is chosen.

As an example, figure 16 shows the true conditional hazard for an individual with $\eta_i = -1, 0, 1$ (values are chosen arbitrarily). The hazard is steadily increasing with time running.

8.1 Bias-Variance Analysis

The quality of an estimator, for example the frailty variance $\hat{\theta}$, is twofold: Firstly, it is desirable that $\hat{\theta}$ is unbiased, i.e. $E[\hat{\theta}] = \theta$. The bias of an estimator is $B(\hat{\theta}) = E[\hat{\theta}] - \theta$. Secondly, the variability $V[\hat{\theta}]$ around its expected value should be small. Both features have to be considered simultaneously: The variance of $\hat{\theta}$ gives an indication of how much $\hat{\theta}$ would change if it would be estimated from another dataset of the same random variable. Too much fidelity to the data, i.e. a high variance, increases the risk that one is confronted with an estimator that captures too much of unstructured variation for a given dataset. One can think of the $V[\hat{\theta}]$ as a measure of how good the estimator is brought into a certain shape and $B[\hat{\theta}]$ specifies how good that shape represents reality.

Ideally, one has an estimator that minimizes both features. However, that is often not the case. This should not be confused with the bias-variance trade-off, that arises if one increases the complexity of the model, by adding parameters. This typically causes a trade-off in that sense, that the bias is reduced but the model is more strongly fitted to the specific dataset what increases the variance.(James et al., 2015, p. 33-36) This is not the case here as there is no competition between models. The model stays the same but the equations to estimate it alter. Nevertheless, the different estimation procedures might lead to a different bias-variance structure. It is especially the bias perspective where frailtyHL claims to be more accurate. The increasing order of approximations might increase the variance by introducing numerical instabilities on the one hand and through a higher fidelity to the data by a higher precision on the other hand.

The mean squared error (MSE) serves as an overall measure to discriminate between estimators, those of frailtyHL and coxph in this case. This is because the MSE unifies both features, bias and variance, into a single measure, as

$$\begin{aligned} MSE(\hat{\theta}) &= E[(\hat{\theta} - \theta)^2] \\ &= E[(\hat{\theta} + E[\hat{\theta}] - E[\hat{\theta}] - \theta)^2] \\ &= E[(\hat{\theta} - E[\hat{\theta}])^2] + 2(E[\hat{\theta}] - \theta)(E[\hat{\theta}] - E[\hat{\theta}]) + (E[\hat{\theta}] - \theta)^2 \\ &= V[\hat{\theta}] + B(\hat{\theta})^2 \end{aligned}$$

By relying on asymptotics the expectation will be calculated as

$$E(\hat{\theta}) = \frac{1}{1000} \sum_{l=1}^{1000} \hat{\theta}_l, \tag{52}$$

with $\hat{\theta}_l = (exp\{\hat{\sigma}_l^2\} - 1)exp\{\hat{\sigma_l}^2\}$ as the estimator of the l^{th} dataset. And correspondingly,

the variance is calculated by

$$V(\hat{\theta}) = \frac{1}{1000} \sum_{l=1}^{1000} (\hat{\theta}_l - E(\hat{\theta}))^2.$$

As parameters might be very different in absolute size it is desirable to get the quality of goodness measure on a more readable and comparable scale. This will be done by dividing all criteria by its true value, the absolute value in the case of β. The following three criteria will be reported

- $SB(\hat{\theta}) = B(\hat{\theta})/\theta$
- $SSE(\hat{\theta}) = \sqrt{V(\hat{\theta})}/\theta$
- $SRMSE(\hat{\theta}) = \sqrt{MSE(\hat{\theta})}/\theta$,

where the additional S stands for scaled, the scaled standard error (SSE) for example.

8.2 Comparison of Estimation Approaches

The results from the simulation study mirror observations from the real world example: β is identical in essence, remarkable differences can be found for θ. The results of the simulation study can be found in table 6 for the high censoring setting.

The relative bias for β is small for both approaches, though bigger for frailtyHL. The relative standard error, however, is slightly smaller for frailtyHL. When taking the scaled root mean squared error as an ultimate criterion, frailtyHL is the better performing approach, though the difference is very small: the expected relative deviation from the real β is 0.357 for frailtyHL and 0.361 for coxph.

The scaled bias and scaled standard error are both bigger for θ for both approaches, indicating the difficulty to estimate parameters, that are further down the model hierarchy. The differences between the approaches are extreme, however. The frailty variance is enormously downward biased for frailtyHL with a scaled bias of -0.4305. This is much smaller in the coxph estimation (-0.079). Figure 17 gives an impression of how the estimators are distributed. The first (0.139), third quartile (0.353) and the median (0.228) are all smaller in the frailtyHL approach than in the coxph estimation. The true value is not even within the range of the first and the third quartile, demonstrating that the estimation is strongly biased. The coxph approach is also downward biased but the range from the first (0.258) and third quartile (0.583) include the real parameter and the median (0.408) is, accordingly, much closer to the real value.

The scaled standard error is smaller for the frailtyHL approach, as can also be seen from the boxplot. However, this means only that there is relatively low variation around a

"bad" value. Finally, the scaled root mean squared error is smaller for the coxph approach. Given the tremendous bias for θ in the frailtyHL estimation, the coxph procedure should be preferred.

Table 6: Goodness Measures in High-Censoring Setting

Parameter	**SB**	**SSE**	**SRMSE**
$\hat{\beta}^{HL}$	0.042	0.355^*	0.357^*
$\hat{\beta}^{coxph}$	0.028^*	0.360	0.361
$\hat{\theta}^{HL}$	-0.4305	0.475^*	0.641
$\hat{\theta}^{coxph}$	-0.079^*	0.597	0.603^*

* marks better performing approach for corresponding criterion

There is no systematic difference in the low-censoring setting. The results can be found in table 7. The SSE and SRMSE declined for both parameters and both approaches. Noteworthy - and alarmingly - is the increase in bias for β despite the increased amount of information in the dataset for both approaches. That is hard to explain and increases concerns that there was an error made in the simulation or estimation process or elsewhere. However, if there is one, it could not be found.[5] Contra a possible error is, that the results confirm what has been found in the real world example, especially in the lognormal model: rather big differences in the frailty variances but small or even negligible differences in the fixed effect. And: all other Q-criteria declined.

The scaled bias of θ declined heavily for both approaches. Relative to the results from the high censoring setting, even more for the coxph approach than for the frailtyHL approach. The bias in the frailtyHL approach is still intolerably high.

Table 7: Goodness Measures in Low-Censoring Setting

Parameter	**SB**	**SSE**	**SRMSE**
$\hat{\beta}^{HL}$	0.052	0.291^*	0.296^*
$\hat{\beta}^{coxph}$	0.034^*	0.296	0.298
$\hat{\theta}^{HL}$	-0.255	0.494^*	0.556
$\hat{\theta}^{coxph}$	-0.010^*	0.550	0.550^*

* marks better performing approach for corresponding criterion

Figure 18 shows again a boxplot for the estimated θ. Both got closer and the distance

[5]The following steps were extensively checked after this finding: were the same datasets (apart from censoring) used to calculate the models across the different censoring schemes? Were the same datasets used for the frailtyHL and the coxph estimation? Were model formulas correct? Are the correct values for the estimators retrieved? Is the bias calculated correctly? The answer is yes, to the best knowledge of the author, to all of those questions.

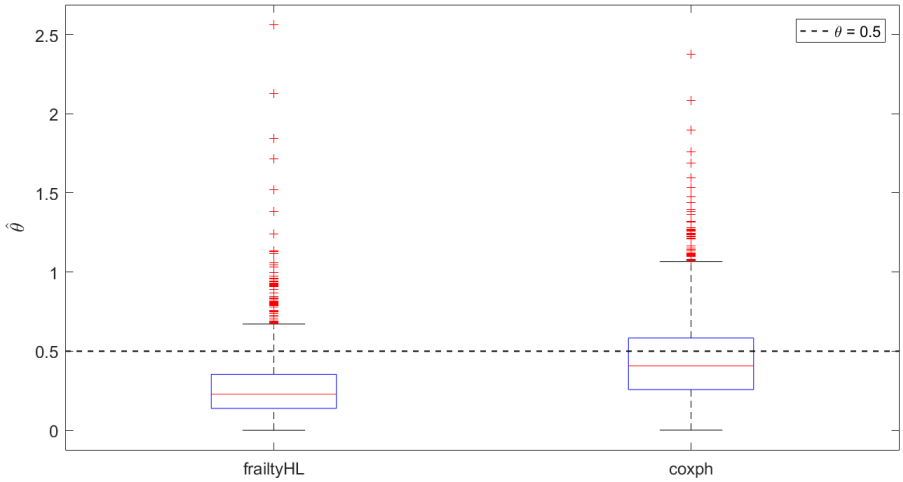

Figure 17: $\hat{\theta}$ from both approaches for all datasets of the high-censoring setting

between the third and the first quartile decreased for both but the true value is still not covered by the first and third quartile of estimated θ in the frailtyHL approach. The median (0.448) of the coxph estimations comes quite close to the real value of 0.5. The median of the frailtyHL estimators was 0.321.

Table 8 shows the performance of both approaches for both settings. The winners are constant through the different settings.

It is surprising that frailtyHL showed better performance in variance but worse in bias. It was expected to be the other way around as the higher order of approximation was assumed to be more sharp in getting all information from the given dataset.

Table 8: Better Estimator in Low- and High Censoring Setting

	Parameter	**SB**	**SSE**	**SRMSE**
$\hat{\beta}$:	Low Censoring	coxph	frailtyHL	frailtyHL
	High Censoring	coxph	frailtyHL	coxph
$\hat{\theta}$:	Low Censoring	coxph	frailtyHL	frailtyHL
	High Censoring	coxph	frailtyHL	coxph

The difference in SRMSE for β is rather small. Given that the differences are much stronger for θ one should opt for the coxph package again.

There are essentially three candidates who might be responsible for bad performance:

- $\boldsymbol{H_v}$ which is assumed to be a diagonal matrix in coxph but not in frailtyHL (no matter the order of approximation),
- $\frac{\partial \hat{\boldsymbol{v}}}{\partial \theta}$ which is ignored in coxph but not in frailtyHL (no matter the order of approximation),

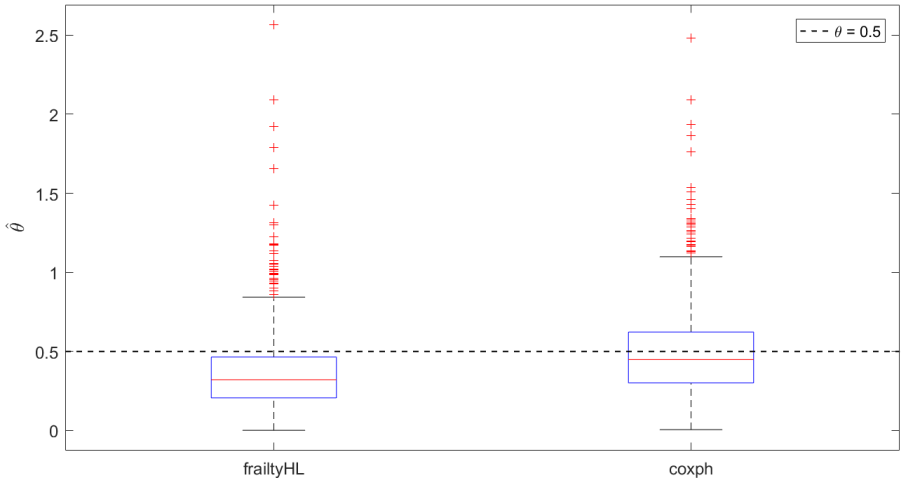

Figure 18: $\hat{\theta}$ from both approaches for all datasets of the low-censoring setting

- F, the higher order of approximation in frailtyHL (which is also a function of $\boldsymbol{H_v}$).

It is tempting to blame $\boldsymbol{H_v}$ in settings with low cluster size: little information on the clusters might lead to a bad estimate of the covariance matrix of the cluster-specific frailty parameters. Table 9 shows the results of first-order frailtyHL (HL01) estimations in the high-censoring setting. The results of coxph and frailtyHL (second-order approximation) are the same as above and are there for comparison. Reducing the order of approximation significantly improves the estimates in terms of bias. In case of β, bias, standard error and mean squared error are identical to coxph. The bias of θ reduces significantly and comes close to that of the coxph estimation but is still a little bit higher. The SSE, however, is increased. Hence, at least some proportion of the heavy bias is due to the inclusion of F. Note, that the elements of F are either from $\boldsymbol{H_v}^{-1}$ or negative first- and second-order derivatives of $\boldsymbol{H_v}$. This is an indication that there is an issue with $\boldsymbol{H_v}$ as suspected which might also be responsible for worse performance in the first-order approximation of frailtyHL.

The additional information in frailtyHL, be it the first- or the second-order approximation, does not lead to better estimations in the case of small clusters. Higher-order approximations even lead to terrible results in the lognormal case. With respect to the real-world example, the results of the coxph estimation should be regarded as much more reliable than those of frailtyHL.

Table 9: Goodness Measures in High-Censoring Setting Including First-order Approximation

Parameter	**SB**	**SSE**	**SRMSE**
$\hat{\beta}^{HL}$	0.042	0.355^*	0.357^*
$\hat{\beta}^{HL01}$	0.028^*	0.360	0.361
$\hat{\beta}^{coxph}$	0.028^*	0.360	0.361
$\hat{\theta}^{HL}$	-0.4305	0.475^*	0.641
$\hat{\theta}^{HL01}$	-0.096	0.608	0.615
$\hat{\theta}^{coxph}$	-0.079^*	0.597	0.603^*

* marks better performing approach for corresponding criterion

9 Conclusion

The aim of this thesis was to asses the difference between the frailtyHL approach and the coxph estimation. The frailtyHL approach estimated the frailty variances to be bigger than the coxph estimation for the gamma frailty for the twin dataset. The opposite has been the case for the log-normal model. However, both approaches coincided in very general inferences: the frailty variances of monozygotic twins are bigger than for dizygotic twins and the log-normal model is a better fit than the gamma model.

From the twin dataset, it can be learned that the choice of distribution should represent the specific problem at hand. The log-normal distribution resulted in the mass of people being closer together (by the comparison of the first and third quartile of the frailty distribution) despite a bigger variance than in the gamma model. However, the right tail of the log-normal distribution was heavier. This is an interesting feature for modelling survival times: On the hand, it allowed the log-normal frailty model to better account for people who died early. On the other hand, there was less mass for frailties very close to zero than in the gamma distribution which also seemed to be a better fit as there are fewer people who are getting extremely old. Consequently, the tail behaviour can be an interesting feature when it comes to the choice of the frailty distribution. Heavy tails might be a good fit for certain problems: A heavy right tail if there are relatively many events occurring very early or lots of mass close to zero if there are a lot of events happening extremely late. The hardly surprising conclusion is that the frailty distribution should be a good representation of reality. This is a contra argument against the use of black-box solutions - as happened here - in essence. The frailtyHL approach could be an interesting standardised approach to model lots of frailty distributions. Identifying the scenarios where it performs badly and finding countermeasures is, hence, highly desirable.

With respect to more detailed estimations, one should be cautious. It is tempting to believe that a higher-order approximation and more computational details lead to better

estimators, at least in terms of bias. However, this does not seem to be the case. For data with small cluster size the higher-order of approximation led to strongly biased results for log-normal frailty. That also means that the estimates resulting from frailtyHL with second-order approximation to the likelihood for the twin dataset should be regarded as unreliable. It would be interesting to see if this is reversed if the cluster size increases for a log-normal frailty model. The Hessian of the log-frailties is the main candidate to be the reason for bad performance but to a big proportion through its influence on the second-order term. Estimation based on the first-order approximation delivered much better results in terms of bias. However, results from the coxph function were still better despite the restriction that the Hessian of the log frailties is a diagonal matrix.

References

Aalen, Odd O.; Borgan, Ørnulf, and Gjessing, Håkon K. *Survival and Event History Analysis: A Process Point of View*. Springer, New York, 2008.

Anderson, Jon; Louis, Thomas; Holm, Niels, and Harvald, Bent. Time-Dependent Association Measures for Bivariate Survival Distributions. *Journal of the American Statistical Association*, 87(419):641–650, 1992.

Antoniou, Antonis and Easton, Douglas. Risk Prediction Models for Familial Breast Cancer. *Future Oncology*, 2(2):257–274, 2006.

Bender, Carl and Orszag, Steven. *Advanced Mathematical Methods for Scientists and Engineers*. McGraw-Hill Book Company, New York, 1978.

Breslow, Norman. Covariance Analysis of Censored Survival Data. *Biometrics*, 30(1): 89–99, 1974.

Burnham, Kenneth and Anderson, David. *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. Springer, New York, 2nd edition, 2002.

Collet, David. *Modelling Survival Data in Medical Research*. CRC Press, 3rd edition, 2015.

Duchateau, Luc and Janssen, Paul. *The Frailty Model*. Springer, New York, 2008.

Fahrmeier, Ludwig; Kneib, Thomas; Land, Stefen, and Marx, Brian. *Regression: Models, Methods and Applications*. Springer, Heidelberg, 2013.

Ha, Il Do; Jeong, Jong-Hyeon, and Lee, Youngjo. *Statistical Modelling of Survival Data with Random Effects: H-Likelihood Approach*. Springer, Singapore, 2017.

Ha, Il Do; Noh, Maengseok; Kim, Jiwoong, and Lee, Youngjo. frailtyHL: Frailty Models via Hierarchical Likelihood: R package version 2.2. 2018. URL `https://CRAN.R-project.org/package=frailtyHL`.

Hauge, M.; Harvald, B.; Fischer, M.; Gotlieb-Jensen, K.; Juel-Nielsen, N.; Raebild, I.; Shapiro, R., and Videbech, T. The Danish Twin Register. *Acta geneticae medicae et gemellologiae*, 17(2):315–332, 1968.

Hougaard, Philip. *Analysis of Multivariate Survival Data*. Springer, New York, 2000.

Hougaard, Philip; Harvald, Bent, and Holm, Niels. Measuring the Similarities Between the Lifetimes of Adult Dansih Twins Born Between 1881-2930. *Journal of the American Statistical Association*, 87(417):17–24, 1992.

James, Gareth; Witten, Daniela, and Hastie, Trevor, Tibshirani, Robert, . *An Introduction to Statistical Learning: with Applications in R*. Springer, New York, 6th edition, 2015.

Kendall, Maurice. A New Measure of Rank Correlation. *Biometrika*, 30(1-2):81–93, 1938.

Lee, Youngjo and Nelder, John. Hierarchical Generalized Linear Models. *Journal of the Royal Statistical Society B*, 58(4):619–678, 1996.

Lee, Youngjo and Nelder, John. Likelihood for Random Effect Models. *Statistical and Operational Research Transactions*, 29(2):141–182, 2005.

Lee, Youngjo; Nelder, John, and Pawitan, Yudi. *Generalized Linear Models with Random Effects: Unified Analysis via H-Likelihood*. CRC Press, Boca Raton, 2nd edition, 2017.

Link, Carol. Confidence Intervals for the Survival Function Using Cox's Proportional-Hazard Model with covariates. *International Biometric Society*, 40(3):601–609, 1984.

R Core Team, . *R: A Language and Environment for Statistical Computing*. Vienna, 2013.

Ripatti, Samuli and Palmgren, Juni. Estimation of Multivariate Frailty Models Using Penalized Partial Likelihood. *Biometrics*, 56(4):1016–1022, 2000.

Therneau, Terry and Grambsch, Patricia. A Package for Survival Analysis in S: version 2.38. 2015. URL `https://CRAN.R-project.org/package=survival`.

Therneau, Terry; Grambsch, Patricia, and Pankratz, Shane. Penalized Survival Models and Frailty. *Journal of Computational and Graphical Statistics*, 12(1):156–175, 2003.

Wang, Weijing and Wells, Martin. Estimation of Kendall's Tau Under Censoring. *Statistica Sinica*, 20(4):1199–1215, 2000.

Acknowledgments

My thanks are going to be short. But they are honest!

Thanks to my sister Simone. Hard times would be harder without you. Good times too!

Thanks to Riccarda for plenty of support and tons of kindness!

Thanks to Dr.Unkel for support regarding administrative issues, help with this thesis and quick answers.

Thanks to Rene! Your pressure paid off.

Selbstständigkeitserklärung

Ich versichere, dass ich die Arbeit selbständig und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Alle Stellen, die wörtlich oder sinngemäß aus Veröffentlichungen oder anderen Quellen entnommen sind, sind als solche kenntlich gemacht. Die schriftliche und elektronische Form der Arbeit stimmen überein.

Ort, Datum: _____

Unterschrift: _____

Maximilian Bardo