A four-dimensional correlated frailty model

by

Nicole Giard

A dissertation submitted to the
Faculty of Mathematics and Natural Sciences
of the
University of Rostock

in partial fulfillment
of the requirements for the degree of

doctor rerum naturalium

Rostock, February 2001

Preface

The work presented here was carried out at the Max Planck Institute for Demographic Research in Rostock, Germany. It deals with questions from different fields of statistics and addresses theoretical aspects as well as practical aspects.
From a theoretical point of view the work can be classified within the field of Survival Analysis. The general frame of the work is made up of random effects models and multivariate survival models. The main result is the generalization of the bivariate correlated frailty model suggested by Yashin et al. (1993). Possible practical applications of the model presented here can be found in genetic epidemiology and twin research.
On the one hand, I investigate theoretical properties of the model like identifiability, consistency and asymptotic normality of parameter estimates. On the other hand practical aspects of the application of the model are discussed. This includes a simulation study regarding the asymptotic properties of maximum likelihood estimates, a simulation study comparing the frailty approach with the more traditional liability approach and, last but not least, the application of the model to a real data set.
The dissertation includes seven chapters and an appendix. The first chapter gives a short introduction to the various fields of mathematical statistics the work touches on. It supplies basic theoretical results that are later explored in proofs and it introduces the most important notations and terminology.
In the second chapter I define the four-dimensional correlated frailty model itself. I derive the survival function and the likelihood function, and I explain the observation scheme of current health status data that is used throughout the dissertation.
Chapters 3 and 4 are devoted to theoretical properties of the new model. In the third chapter I prove identifiability, whereas in the fourth chapter I give sufficient conditions for the consistency and asymptotic normality of maximum likelihood estimates.
In the fifth chapter I perform a simulation study that investigates the asymptotic behavior of maximum likelihood estimates in the four-dimensional cor-

related frailty model when the estimation is based on current health status data. The empirical distribution of the estimates is investigated by simple descriptive means as well as by tests of normality based on skewness and kurtosis.
The sixth chapter is devoted to the application of the four-dimensional correlated frailty model to a data set on prostate cancer in twins. I show that it is possible to use the model in order to evaluate the influence of the genes and the environment on the development of a disease and on life span.
In the last chapter I present another simulation study that aims at the comparison of the frailty concept with liability concept used in genetic epidemiology. Based on simulated data I show that with the help of the correlated frailty model one can eliminate the confounding effect of the age at death on the disease status relationship of twins whereas this is not possible with the liability approach. The advantages and drawbacks of both approaches are discussed.
The appendix contains a summary of notations and explicit representations of functions used in the theoretical proofs. It gives technically extensive lemmata and their proofs which, if included in the text, would make the reading of the dissertation as a whole very difficult. The appendix also includes a short description of the numerical algorithms used, various figures and tables.
I wish to express my appreciation to my supervisors, Prof. Anatoli I. Yashin from the Max Planck Institute for Demographic Research and Prof. Friedrich Liese from the Department of Mathematics at the University of Rostock. I wish to thank them for their advice and their support of this work. I am grateful to Dr. Andreas Wienke and Ivan A. Iachine for many fruitful discussions and constructive criticism which influenced the presentation of my results. I wish to thank Paul Lichtenstein from the Department of Medical Epidemiology at the Karolinska Institute in Stockholm, Sweden, for providing me with the data on Swedish twins. I extend my thanks to the entire staff of the Max Planck Institute for Demographic Research for their overall support of this project. I am grateful to the Max Planck Institute and especially its director, Prof. James W. Vaupel, who gave me the opportunity to perform this research. Finally, I wish to thank my husband, Sven Guttmann, and my children, Josephine and Carl, for their understanding and moral support.

Rostock, February 2001.

Nicole Giard

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Survival analysis

The term survival analysis summarizes statistical models and methods for analyzing lifetime data. The variable of interest can, for instance, be the life span of an individuum, the time until a machine fails or the time of remission of a disease after treatment with a certain drug. In general the time until a certain event occurs is called survival time.
The statistician describes survival time with a random variable T that is defined on a probability space and takes non-negative real values. One is interested in the distribution of T that can be specified with the *survival function*

$$S(x) = P(T > x) \ .$$

Let T have a density function $f(x) = -S'(x)$. Then one can write

$$S(x) = \exp\left\{ - \int\limits_{[0,x)} \mu(u)\, du \right\} = \exp\{-H(x)\}$$

where

$$\mu(x) = \frac{f(x)}{S(x)} = -\frac{d \ln S(x)}{dx} = \lim_{\triangle x \to 0+} \frac{P(x < T \leq x + \triangle\, x | T > x)}{\triangle\, x}$$

is called the *hazard function*. It describes the instantaneous rate of failure at time x. The function $H(x) = \int_{[0,x)} \mu(u)\, du$ is called the *cumulative hazard function*.
In survival analysis the time of interest is not always fully observable. This can be the case if the exact survival time is only known for a portion of the

items in a study. On the one hand for some items it may only be known that the survival times exceed a certain value due to a deterministic or random termination of the study. Such a pattern of observation is called *(right) censoring*. On the other hand some items are actually not included in the study since they failed or died before the beginning of the study, which is called *(left) truncation*.
In practice many other and more complicated forms of incomplete observations are possible; for examples see Lawless (1982) or Miller (1981). A very general definition of the terms censoring and truncation can be found in Turnbull (1976).
The *Kaplan-Meier Estimator* is used as a non-parametric estimator for the survival function from censored data. It was introduced by Kaplan and Meier (1958) and is a generalized maximum likelihood estimator.
Let T and C be two independent non-negative random variables, where T denotes a survival time with survival function S and where C denotes the censoring time. Censored observations of T then consist of realizations of the random variables

$$X = \min(T, C) \quad \text{and} \quad \Delta = \mathrm{I}(T \leq C) \, .$$

Let (x_i, δ_i), $i = 1, \ldots, n$, be the censored data. Observations x_i for which $\delta_i = 0$ are called censored times, and observations for which $\delta_i = 1$ are called uncensored times or failures. Let $x_{(i)}$, $i = 1, \ldots, n$, denote the ordered observed times, i.e. $x_{(1)} \leq \cdots \leq x_{(n)}$. Let $\delta_{(i)}$ denote the corresponding censoring indicator. The convention holds that in case there are ties among the $x_{(i)}$, the failures precede the censored times. The Kaplan-Meier estimator $\hat{S}$ for the survival function S of T can be calculated as

$$\hat{S}(t) = \prod_{i \,:\, x_{(i)} \leq t} \left(\frac{n-i}{n-i+1} \right)^{\delta_{(i)}} \tag{1.1}$$

for $t \leq x_{(n)}$. If $\delta_{(n)} = 1$ then $\hat{S}(t)$ is set equal to zero for $t > x_{(n)}$, and if $\delta_{(n)} = 0$ then $\hat{S}(t)$ is left undefined for $t > x_{(n)}$. Equivalently, $\hat{S}$ can be expressed as

$$\hat{S}(t) = \prod_{j \,:\, t_j \leq t} \left(1 - \frac{d_j}{n_j} \right)$$

where the t_j are the ordered, observed and distinct failure times, that means $t_1 < \cdots < t_k$. d_j denotes the number of failures at time t_j, and n_j is the number of items at risk at time t_j, i.e. the number of items still alive just before t_j.

A detailed description and derivation of the theoretical properties of the Kaplan-Meier estimator can be found in Kotz and Johnson (1988) or Andersen et al. (1993). A general introduction into the field of survival analysis can be found in Kalbfleisch and Prentice (1980) or Cox and Oakes (1984).

1.2 Identifiability

A general problem of statistical analysis is the question of *identifiability*. In statistical inference it is usually assumed that the distribution of interest depends on a set of (possibly infinite dimensional) unknown parameters. "Any statistical procedure developed for estimation of these parameters is meaningful only if the unknown parameters are identifiable." (Prakasa Rao, 1992, p.xi)
One can define *identifiability* in the following way:

Definition 1. Let Y be an observable random variable with values in the measurable space $(\mathfrak{X}, \mathfrak{A})$ and the distribution $Q_\theta \in \mathfrak{Q} = \{Q_\theta, \theta \in \Theta\}$. $\mathfrak{Q}$ is a family of probability distributions defined on $(\mathfrak{X}, \mathfrak{A})$ and indexed by a parameter θ. θ is said to be *identifiable* if for every pair (θ_1, θ_2), with $\theta_1, \theta_2 \in \Theta$ holds: $Q_{\theta_1} = Q_{\theta_2}$ if and only if $\theta_1 = \theta_2$.

Since the parameter set Θ is arbitrary and not necessarily finite dimensional this definition can also be seen from a non-parametric point of view.
If a parameter is identifiable then the estimation of the corresponding distribution Q_θ allows us to make unambiguous conclusions about θ.
If the family $\mathfrak{Q}$ is dominated by a σ-finite measure μ and f_θ denotes the probability density function of Q_θ with respect to μ, then θ is obviously identifiable if for $\theta_1, \theta_2 \in \Theta$

$$f_\theta(x_1) = f_{\theta_2}(x) \qquad \mu\text{-a.s.} \tag{1.2}$$

implies $\theta_1 = \theta_2$.
In the context of survival analysis the statistician is often confronted with the following situation. It is assumed that the distribution of the survival time T of interest belongs to a certain parameterized family $\mathfrak{Q} = \{Q_\theta, \theta \in \Theta\}$. If T could be observed then the identifiability of θ could be proved easily. But in practice T can only be observed incompletely due to censoring or other reasons. That means that one can only observe a variable $X = \varphi(T)$ that is a function φ of T and the distribution of X belongs to a family $\mathfrak{P} = \{P_\theta, \theta \in \Theta\}$ where $P_\theta = Q_\theta \circ \varphi^{-1}$. Even if $Q_{\theta_1} = Q_{\theta_2}$ implies $\theta_1 = \theta_2$ the same conclusion is not necessarily true for P_{θ_1} and P_{θ_2}.

Example 1. Let T_1 and T_2 be two independent random variables that represent two different survival times. T_1 and T_2 can, for instance, be the time until failure of two components of a device. Instead of T_1 and T_2 only $X = \min(T_1, T_2)$ is observable, i.e. the time until failure of the entire device. Let us assume that T_i is exponentially distributed with the parameter λ_i, $i = 1, 2$. Hence T_i has the survival function $S_i(x) = e^{-\lambda_i x}$ and X has the survival function

$$S_X(x) = P(X > x) = P(\min(T_1, T_2) > x) = S_1(x)S_2(x) = e^{-(\lambda_1+\lambda_2)x} \ .$$

Thus the parameters λ_1 and λ_2 cannot be identified by observing X.

Example 2. Let us consider the same situation as in the example above but in addition to X, the minimum of T_1 and T_2, the variable I that takes the value i if $X = T_i$ is also observable. That means that one also knows which component of the device failed first. The distribution of (X, I) is described by the sub-survival functions

$$\check{S}_1(x) = P(X > x, I = 1) = P(T_1 > x, T_1 \leq T_2) = \frac{\lambda_1}{\lambda_1 + \lambda_2} e^{-(\lambda_1+\lambda_2)x}$$

and

$$\check{S}_2(x) = P(X > x, I = 2) = P(T_2 > x, T_1 > T_2) = \frac{\lambda_2}{\lambda_1 + \lambda_2} e^{-(\lambda_1+\lambda_2)x} \ .$$

Let $x_0 > 0$ be an arbitrary value, then one can write

$$\lambda_i = -\frac{1}{x_0} \cdot \ln[\check{S}_1(x_0) + \check{S}_2(x_0)] \cdot \check{S}_i(0) \ , i = 1, 2,$$

thus λ_1 and λ_2 are identifiable if (X, I) is observed.

A very detailed description of various problems of identifiability that occur in different statistical models can be found in Prakasa Rao (1992).

1.3 Frailty models

1.3.1 Univariate model

The notion of *frailty* as a measure of general susceptibility to all causes of death was introduced to describe mortality in heterogeneous populations (Vaupel et al., 1979). It can also be interpreted as the random effect in a random effect model for time variables.

Let T and Z be random variables defined on a probability space $(\Omega, \mathfrak{F}, P)$. T denotes a survival time, such as, for instance, life span. Z represents heterogeneity or risk factors that change individual chances of survival. Let $S(x|z)$ be the *conditional survival function* of T given $Z = z$, i.e.

$$S(x|z) = P(T > x|Z = z) \ .$$

Recall that the following relation holds

$$P(T > x, Z > z) = \int\limits_{(z,\infty)} S(x|s)P_Z(ds) \tag{1.3}$$

where P_Z denotes the probability distribution of Z. If Z has the probability density function $f_Z(z)$ one can calculate the unconditional survival function $S(x)$ of T as

$$S(x) = \int S(x|z)f_Z(z)dz \ .$$

If one can write

$$S(x|z) = \exp\Big\{ - \int\limits_0^x \mu(u, z)\, du \Big\} \tag{1.4}$$

then μ is a given function, called the conditional hazard function of T under $Z = z$.
Usually a proportional hazard model $\mu(x, Z) = Z\mu_0(x)$ with an *underlying or baseline hazard rate* $\mu_0(x)$ and a non-negative random effect Z is used. Z is called *frailty*. This has the interpretation that if an individual has a certain frailty value z, the individual hazard rate is $z\,\mu_0(x)$ and the survival probability can be derived from (1.4). The frailty has a multiplicative effect on the hazard and the unconditional survival function, i.e. the population survival function that usually can be observed, is a mixture over Z.
If $H_0(x) = \int_0^x \mu_0(u)\, du$ denotes the *cumulative baseline hazard function*, the unconditional survival function of T is

$$S(x) = \int \exp\{-zH_0(x)\}\, f_Z(z)dz = Ee^{-ZH_0(x)} = \mathcal{L}(H_0(x)) \tag{1.5}$$

where $\mathcal{L}(v) = Ee^{-vZ}$ is the Laplace transform of Z.
In practical applications the frailty variable is often assumed to be gamma distributed (Vaupel, 1988). The gamma distribution has the advantage of

having a simple density and a simple Laplace transform. A real valued random variable X is called *gamma distributed* with shape parameter k and scale parameter λ if it has the density function:

$$f(x) = \frac{\lambda^k}{\Gamma(k)} x^{k-1} e^{-\lambda x} \, \mathrm{I}(x > 0), \quad \text{notation: } X \sim \Gamma(k, \lambda). \tag{1.6}$$

The expected value of this distribution is $\frac{k}{\lambda}$ and the variance is $\frac{k}{\lambda^2}$. The Laplace transform of the gamma distribution is

$$\mathcal{L}(v) = \left[1 + \frac{v}{\lambda}\right]^{-k}. \tag{1.7}$$

Another advantageous property of the gamma distribution is that if X_1 and X_2 are independent and $X_i \sim \Gamma(k_i, \lambda)$, $i = 1, 2$, then $X_1 + X_2 \sim \Gamma(k_1 + k_2, \lambda)$. In addition $cX_1 \sim \Gamma(k_1, \frac{\lambda}{c})$ for all $c > 0$.
If the frailty is gamma distributed with shape parameter k and scale parameter λ in a univariate frailty model, then

$$S(x) = \left[1 + \frac{H_0(x)}{\lambda}\right]^{-k}$$

holds for the unconditional survival function (see (1.5) and (1.7)). In addition if Z is assumed to have the expected value 1 (this is due to identifiability) and the variance σ^2 then the last formula is reduced to

$$S(x) = [1 + \sigma^2 H_0(x)]^{-1/\sigma^2}. \tag{1.8}$$

The conditional density function $f_Z(z|T > x)$ of Z given $T > x$ is defined as the probability density function corresponding to the conditional survival function $S_Z(z|T > x) = P(Z > z|T > x)$. Applying (1.3) yields

$$P(Z > z|T > x)\, S(x) = P(T > x, Z > z) = \int\limits_{(z,\infty)} S(x|s) f_Z(s)\, ds$$

and thus

$$f(z|T > x) = \frac{S(x|z) f_Z(z)}{S(x)} = \frac{[\lambda + H_0(x)]^k}{\Gamma(k)} z^{k-1} e^{-(\lambda + H_0(x))z}.$$

Hence the frailty among the survivors of age x is gamma distributed again. The conditional distribution of the frailty given $T > x$ is the gamma distribution with same shape parameter k and the scale parameter $\lambda + H_0(x)$.

Unfortunately, however, there are no biological reasons that justify the choice of a gamma distribution for the frailty variable.
The major disadvantage of this univariate approach is that one has to make assumptions concerning the parametric structure of the underlying hazard rate μ_0 in order to make such a univariate frailty model identifiable. Elbers and Ridder (1982) show that in a univariate frailty model, that includes covariates and that assumes a frailty distribution with a finite mean, the underlying hazard rate μ_0 as well as the distribution of Z are identifiable. Elbers and Ridder introduce the covariates into the model by assuming that the baseline hazard rate can be written as

$$\mu(x, \mathbf{w}, Z) = \varphi(\mathbf{w})\, \mu_0(x)\, Z \ .$$

The vector $\mathbf{w}$ denotes the values of the covariates and φ denotes an unknown non-negative function that fulfills some smoothness condition.

1.3.2 Bivariate model

Yashin et al. (1993) extended the univariate frailty model to a bivariate correlated model in a way that is consistent with the univariate definition of frailty.
Let T_1 and T_2 be the survival times of two related individuals. The association between these two survival times can be modeled using the concept of frailty where the univariate interpretation of frailty as individual susceptibility remains the same. Let Z_1 and Z_2 be the frailty variables corresponding to T_1 and T_2 as described in section 1.3.1. The main assumption of the bivariate model is the *conditional independence*.

Definition 2. Let (X, Y, Z) be a vector of three random variables with values in the measurable product space $(\mathfrak{X}_X \times \mathfrak{X}_Y \times \mathfrak{X}_Z, \mathfrak{A}_X \otimes \mathfrak{A}_Y \otimes \mathfrak{A}_Z)$. Let P_Z denote the probability distribution of Z and let $P_X(\cdot|Z = z)$, $P_Y(\cdot|Z = z)$ and $P_{(X,Y)}(\cdot|Z = z)$ denote the conditional distributions of X, Y and (X, Z) given $Z = z$, respectively. The random variables X and Y are said to be *conditionally independent given* Z if

$$P_{(X,Y)}(A \times B|Z = z) = P_X(A|Z = z) \cdot P_Y(B|Z = z)$$

holds for all sets $A \in \mathfrak{A}_X$, $B \in \mathfrak{A}_Y$ and all $z \in \mathfrak{X}_Z$.

The survival times T_1 and T_2 are assumed to be conditionally independent given the frailties Z_1 and Z_2. It is assumed that for the conditional bivariate survival function of (T_1, T_2) given $(Z_1, Z_2) = (z_1, z_2)$ it holds that

$$S(x_1, x_2|z_1, z_2) = S_1(x_1|z_1) \cdot S_2(x_2|z_2) \ , \tag{1.9}$$

where $S_i(x_i|z_i)$ is the conditional survival function of T_i given frailty $Z_i = z_i$. Note that in (1.9) a separation of Z_1 and Z_2 is assumed in addition to the conditional independence. I.e. the conditional distribution of T_1 given Z_1 does not depend on Z_2 and vice versa. Conditional independence as defined above would mean $S(x_1, x_2|z_1, z_2) = S_1(x_1|z_1, z_2) \cdot S_2(x_2|z_1, z_2)$. Under the condition that the frailty variables are fixed, the survival times are independent. If (Z_1, Z_2) has the bivariate density function $f_Z(z_1, z_2)$ one can calculate the unconditional bivariate survival function $S(x_1, x_2)$ of (T_1, T_2) as

$$\begin{aligned} S(x_1, x_2) &= \iint S(x_1, x_2|z_1, z_2)\, f_Z(z_1, z_2)\, dz_1 dz_2 \\ &= \iint S_1(x_1|z_1)\, S_2(x_2|z_2)\, f_Z(z_1, z_2)\, dz_1 dz_2 \end{aligned} \tag{1.10}$$

For the univariate conditional survival functions $S_i(x_i|Z_i)$ it is assumed that (see (1.4))

$$S_i(x|Z_i) = \exp\Big\{ - Z_i \int_0^x \mu_i(u)\, du \Big\} = \exp\{-Z_i H_i(x)\}$$

with underlying baseline hazard rate μ_i and cumulative baseline hazard function H_i. From (1.10) it follows that

$$\begin{aligned} S(x_1, x_2) &= \iint e^{-Z_1 H_1(x_1) - Z_2 H_2(x_2)}\, f_Z(z_1, z_2)\, dz_1 dz_2 \\ &= E e^{-Z_1 H_1(x_1) - Z_2 H_2(x_2)} = \mathcal{L}(H_1(x_1), H_2(x_2)) \end{aligned} \tag{1.11}$$

where $\mathcal{L}(u, v) = E e^{-uZ_1 - vZ_2}$ is the two-dimensional Laplace transform of (Z_1, Z_2).

The identifiability of a bivariate frailty model described in (1.11) is different from that of a univariate model (see (1.5)) when the frailty variables are constructed using independent additive components. Let Y_0, Y_1 and Y_2 be independent non-negative random variables and let $\alpha > 0$ be a real number. Iachine and Yashin (1998) show that if $Z_1 = Y_0 + Y_1$ and $Z_2 = \alpha Y_0 + Y_2$, the model is identifiable under some weak assumptions on $H_i(x)$, $i = 1, 2$, and the distribution of Y_0. They prove that if the unconditional survival function $S(x_1, x_2)$ from (1.11) is known, then the distributions of Y_0, Y_1 and Y_2, the parameter α and the cumulative hazard functions $H_i(x)$, $i = 1, 2$, are identifiable.

Yashin and Iachine (1995b) apply a bivariate correlated frailty model that uses gamma distributed independent components for the construction of the

bivariate frailty distribution on Danish twin survival data. The properties of the model are discussed in Yashin et al. (1995).
Let Y_0, Y_1 and Y_2 be independent random variables which are gamma distributed with the scale parameter 1 and different shape parameters, i.e. $Y_i \sim \Gamma(k_i, 1)$. Y_0 represents the risk factors which the two individuals in a pair share whereas Y_1 and Y_2 represent the non-shared individual risk factors. Let us assume that the frailty variables Z_1 and Z_2 that correspond to the survival times T_1 and T_2 of the pair of individuals can be decomposed:

$$\begin{aligned} Z_1 &= \sigma_1^2(Y_0 + Y_1) \\ Z_2 &= \sigma_2^2(Y_0 + Y_2) \end{aligned} \tag{1.12}$$

Then $Z_i \sim \Gamma(k_0 + k_i, 1/\sigma_i^2)$, $i = 1, 2$ (see page 6). If we use the standard assumption that $EZ_i = 1$, $i = 1, 2$, follows $\mathrm{Var}(Z_i) = \sigma_i^2$. And

$$\varrho = \mathrm{corr}(Z_1, Z_2) = \sigma_1 \sigma_2 k_0 \tag{1.13}$$

holds for the correlation ϱ between the frailty variables Z_1 and Z_2. Since it also holds that $1/\sigma_i^2 = k_0 + k_i$, $i = 1, 2$, there is a one-to-one correspondence between the shape parameters k_i, $i = 0, 1, 2$, on the one hand and the variances σ_i^2, $i = 1, 2$, and the correlation ϱ of the frailty variables on the other hand. The shape parameters must be positive and therefore one gets the following constraint on the correlation ϱ:

$$0 < \varrho < \min\left(\frac{\sigma_1}{\sigma_2}, \frac{\sigma_2}{\sigma_1}\right) \quad . \tag{1.14}$$

To calculate the unconditional bivariate survival function $S(x_1, x_2)$ of (T_1, T_2) one has to know the two-dimensional Laplace transform of (Z_1, Z_2). Using (1.7), (1.11) and (1.12) one gets

$$\begin{aligned} \mathcal{L}(u, v) &= Ee^{-uZ_1 - vZ_2} = Ee^{-(u\sigma_1^2 + v\sigma_2^2)Y_0 - u\sigma_1^2 Y_1 - v\sigma_2^2 Y_2} \\ &= Ee^{-(u\sigma_1^2 + v\sigma_2^2)Y_0}\, Ee^{-u\sigma_1^2 Y_1}\, Ee^{-v\sigma_2^2 Y_2} \\ &= [1 + u\sigma_1^2 + v\sigma_2^2]^{-k_0} [1 + u\sigma_1^2]^{-k_1} [1 + v\sigma_2^2]^{-k_2} \end{aligned}$$

and

$$S(x, y) = \left[1 + \sigma_1^2 H_1(x)\right]^{-k_1} \left[1 + \sigma_2^2 H_2(y)\right]^{-k_2} \left[1 + \sigma_1^2 H_1(x) + \sigma_2^2 H_2(y)\right]^{-k_0} . \tag{1.15}$$

Expressing the shape parameters with σ_1, σ_2 and ϱ, and applying (1.8) yields

$$S(x, y) = [S_1(x)]^{1 - \frac{\sigma_1}{\sigma_2}\varrho} [S_2(y)]^{1 - \frac{\sigma_2}{\sigma_1}\varrho} [S_1^{-\sigma_1^2}(x) + S_2^{-\sigma_2^2}(y) - 1]^{-\frac{\varrho}{\sigma_1 \sigma_2}} \tag{1.16}$$

where S_1 and S_2 are the marginal univariate survival functions of T_1 and T_2, respectively. The second representation of the bivariate survival function makes a semi-parametric approach to the estimation of the parameters of the bivariate frailty distribution possible if one uses a non-parametric estimate for the marginal univariate survival functions (Yashin et al., 1995).
If ϱ takes one of the limits obtained in (1.14), then S in (1.16) is still a bivariate survival function. In case $\varrho = 0$ the survival times are independent. Since identifiability of the model was only shown for $\varrho > 0$ this case will be excluded in all subsequent investigations.
In an application of the bivariate correlated frailty model on same sex twins it should play no role whether a twin is the first or the second in a pair, i.e. the marginal univariate survival functions and the distributions of the frailty variables should be the same. This yields the specification $S_1 = S_2 = S_L$, $H_1 = H_2 = H$ and $\sigma_1 = \sigma_2 = \sigma$. The constraint on the correlation between the frailty variables is $0 < \varrho < 1$, which is a natural constraint for a positive correlation. The bivariate survival function in a twin model is given by

$$\begin{aligned} S(x,y) &= \left[1+\sigma^2 H(x)\right]^{\frac{\varrho-1}{\sigma^2}} \left[1+\sigma^2 H(y)\right]^{\frac{\varrho-1}{\sigma^2}} \left[1+\sigma^2 H(x)+\sigma^2 H(y)\right]^{-\frac{\varrho}{\sigma^2}} \\ &= \left[S_L(x)\, S_L(y)\right]^{1-\varrho} \left[S_L^{-\sigma^2}(x) + S_L^{-\sigma^2}(y) - 1\right]^{-\frac{\varrho}{\sigma^2}}. \end{aligned} \tag{1.17}$$

Since the correlation between the frailty variables is explicitly included in the model it can be estimated separately for monozygotic and dizygotic twins. Therefore it is possible to evaluate the relative importance of genetic and environmental effects on individual frailty (see section 1.5). When data on monozygotic and dizygotic twins are combined different genetic parameters such as the heritability of frailty can be estimated directly. Yashin and Iachine (1995b) obtained an estimate of about 50% for the heritability in frailty representing individual susceptibility to death in a Danish twin population. This estimate was confirmed in the combined analysis of data from 3 Scandinavian twin registers (Iachine et al., 1998).
An additional advantage of the bivariate correlated frailty model is that it is possible to include censored and/or truncated data. It also allows us to include covariates (Yashin et al., 1996; Yashin and Iachine, 1997).
For a rigorous introduction of the terms conditional distribution, conditional survival function and conditional independence see, for instance, Witting (1985) or Chow and Teicher (1988).

1.3.3 Frailty models in the literature

Different forms of multivariate frailty models are discussed and applied in the literature.

The special case where the correlation between frailties is equal to one is called *shared frailty model*. In this model the related individuals share the same relative risk and frailty represents only common risk factors. The interpretation of the variance in frailty and the underlying hazard rates is different to the one in the correlated frailty model. The properties of the shared frailty model are discussed in Hougaard (1995).
A different approach for the introduction of frailty models is based on the theory of counting processes (Andersen, Borgan, Gill, and Keiding, 1993). Nielsen et al. (1992) formulate the shared frailty model in this framework and discuss the EM-Algorithm for the estimation of the variance of the frailty distribution and the underlying cumulative hazard rate. The underlying cumulative hazard rate can be estimated either parametrically or non-parametrically. Iachine (1995) describes the EM-Algorithm for the bivariate correlated frailty model with observed covariates.
For a special case of the model suggested by Nielsen et al., Murphy (1994) proves consistency of the estimators derived by the EM-algorithm in the case where the underlying cumulative hazard rate is estimated non-parametrically. In another paper Murphy (1995) gives the asymptotic distribution of these estimators.
Parner (1996b) generalizes Murphy's proof of consistency for the correlated frailty model and extends the model to allow for covariates. He introduces a correlation between the frailty variables by decomposing them into the sum of a shared and a non-shared component. Parner (1996a) also proves the asymptotic normality of the estimators in the correlated gamma-frailty model that are derived through the EM-algorithm.
Petersen (1998) formulates a general multivariate correlated frailty model using the multivariate counting process framework. His model includes special cases like the shared frailty model suggested by Clayton (1978) and the correlated frailty model suggested by Yashin et al. (1995). It can be applied to litter and adoption designs. The vector of frailty variables is linked by a known design matrix to the intensity process of the event time. The individual frailty is thus a sum of different independent frailty components. Petersen discusses different estimation approaches for his model. For the litter model he describes in detail the EM-Algorithm (Petersen, 1996). In this two-step procedure estimation in the frailty model is regarded as an incomplete data problem. A non-parametric estimator of the cumulative baseline hazard function and estimators for the parameters of the frailty distribution can be derived by this procedure. A simulation study shows properties of the estimation in the litter model. Petersen (1998) states the results of the application of his model to an adoption study.
Pickles et al. (1994) suggest a bivariate frailty model for twins where the

specific environment effects are assumed to be included in the observed covariates. Thus the frailty variables represent only the effects of the common environment and the genes. Pickles and Crouchley (1994) discuss different discrete and continuous frailty distributions and different estimation strategies. In a simulation study Pickles and Crouchley (1995) find that for a multivariate frailty model that represents multiple outcome measurements the estimation of regression type coefficients for the covariates is quite robust to the misspecification of the frailty distribution.
Thomas et al. (1997) propose a correlated frailty model for pedigree data. In this model it is assumed that two frailties belong to each individual, one inherited from each parent, which act multiplicatively on the baseline hazard through a relative risk function that depends on the form of the genetic interaction between frailties. The frailties are assumed to be distributed over the set of integers $\{1, \ldots, K\}$ with some unknown Poisson distributed K. A Markov chain Monte Carlo method is suggested to fit the model. The model is applied to a data set on breast cancer in families.
This work does not deal with the question of using another than the gamma distribution for the frailty variables. One can find discussions of other frailty distributions in the literature. Yashin et al. (1999) compare, for example, different continuous frailty distributions for the bivariate correlated frailty model. Begun et al. (2000) introduce discrete frailty distributions to model the influence of one major gene on life span.

1.4 Asymptotic properties of maximum likelihood estimates

1.4.1 Consistency and asymptotic normality

Let $X_1, X_2, \ldots$ be a sequence of i.i.d. random variables defined on a probability space $(\Omega, \mathfrak{F}, P)$ with values in a measurable space $(\mathfrak{X}, \mathfrak{A})$. Let the distribution of X_1 belong to a family $\mathfrak{Q} = \{Q_{\boldsymbol{\theta}}, \boldsymbol{\theta} \in \Theta\}$ indexed by a parameter $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k)' \in \Theta$, where $\Theta \subset \mathbb{R}^k$ is a Borel set. Any measurable function $\hat{\boldsymbol{\theta}}_n$ from $(\mathfrak{X}^n, \mathfrak{A}^n)$ to $(\Theta, \mathfrak{A}_\Theta)$ is called an *estimator* of $\boldsymbol{\theta}$. Here $\mathfrak{A}_\Theta$ is the sub-σ-algebra of $\mathfrak{B}^k$ generated by Θ. Let $X_1 \sim Q_{\boldsymbol{\theta}_0}$.

Definition 3. The estimator $\hat{\boldsymbol{\theta}}_n$ of $\boldsymbol{\theta}$ is said to be *(weakly) consistent* if $\hat{\boldsymbol{\theta}}_n(X_1, \ldots, X_n)$ converges in probability to $\boldsymbol{\theta}_0$. That means that for any $\varepsilon > 0$

$$\lim_{n \to \infty} P(|\hat{\boldsymbol{\theta}}_n(X_1, \ldots, X_n) - \boldsymbol{\theta}_0| > \varepsilon) = 0 \ .$$

We denote this as: $\hat{\boldsymbol{\theta}}_n \xrightarrow{P} \boldsymbol{\theta}_0$.

Definition 4. A sequence $V_1, V_2, \ldots$ of random variables on a probability space $(\Omega, \mathfrak{F}, P)$ is said to *converge in distribution* to a random variable V if the distribution function F_n of V_n converges to the distribution function F of V at every continuity point of F. We write: $V_n \xrightarrow{\mathcal{L}} V$.

Definition 5. The estimator $\hat{\boldsymbol{\theta}}_n$ of $\boldsymbol{\theta}$ is said to be *asymptotically normal* if there exists a k-dimensional positive semi-definite matrix $\Sigma(\boldsymbol{\theta}_0)$ such that $\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)$ converges in distribution to a random variable $X \sim N(0, \Sigma(\boldsymbol{\theta}_0))$. We denote this as: $\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \xrightarrow{\mathcal{L}} N(0, \Sigma(\boldsymbol{\theta}_0))$.

Let the family $\mathfrak{Q}$ (see page 12) be dominated by a σ-finite measure μ and let $f(x|\boldsymbol{\theta})$ denote the probability density function of $Q_{\boldsymbol{\theta}}$ with respect to μ. The function

$$L(\mathbf{x}|\boldsymbol{\theta}) = L(x_1, \ldots, x_n|\boldsymbol{\theta}) = \prod_{i=1}^{n} f(x_i|\boldsymbol{\theta})$$

is called the *likelihood function* of $\mathbf{X} = (X_1, \ldots, X_n)'$.

Definition 6. Any value $\hat{\boldsymbol{\theta}}_n(\mathbf{x})$ that maximizes the likelihood function

$$L(\mathbf{x}|\hat{\boldsymbol{\theta}}_n(\mathbf{x})) = \sup_{\boldsymbol{\theta} \in \Theta} L(\mathbf{x}|\boldsymbol{\theta}) \tag{1.18}$$

is called a *maximum likelihood estimate* (MLE) of $\boldsymbol{\theta}$.

Remark. If Θ is compact and if $f(x|\boldsymbol{\theta})$ is continuous in x then a measurable solution of (1.18) always exists. (Witting and Nölle, 1970)

The following theorem gives sufficient conditions for the consistency of a maximum likelihood estimate. It can be found in Amemiya (1985).

Theorem 1. *Let $X_1, X_2, \ldots$ be a sequence of i.i.d. random variables defined on a probability space $(\Omega, \mathfrak{F}, P)$ with values in a measurable space $(\mathfrak{X}, \mathfrak{A})$. Let $X_1 \sim Q_{\boldsymbol{\theta}_0} \in \mathfrak{Q} = \{Q_{\boldsymbol{\theta}}, \boldsymbol{\theta} \in \Theta\}$ and let $\mathfrak{Q}$ be dominated by a σ-finite measure μ. $f(x|\boldsymbol{\theta})$ denotes the density function of $Q_{\boldsymbol{\theta}}$ with respect to μ. If*

(A1) the parameter space Θ is a compact subset of $\mathbb{R}^k$,

(B1) $f(x|\boldsymbol{\theta})$ is a continuous function of $\boldsymbol{\theta}$ for all x,

(C1) $\boldsymbol{\theta}$ is identifiable, i.e. $f(x|\boldsymbol{\theta}_1) = f(x|\boldsymbol{\theta}_2)$ a.s. implies $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_2$, and

(D1) $E_{\boldsymbol{\theta}_0} \sup_{\boldsymbol{\theta}\in\Theta} |\ln f(X_1|\boldsymbol{\theta})| = \int \sup_{\boldsymbol{\theta}\in\Theta} |\ln f(x|\boldsymbol{\theta})| f(x|\boldsymbol{\theta}_0)\,\mu(dx) < \infty,$

then the maximum likelihood estimate converges in probability to $\boldsymbol{\theta}_0$.

Remark. It is understood that the sequence of maximum likelihood estimates is chosen as a sequence of measurable functions.

Proof. Define

$$Q_n(\mathbf{x}|\boldsymbol{\theta}) = \ln L(\mathbf{x}|\boldsymbol{\theta}) = \sum_{i=1}^{n} \ln f(x_i|\boldsymbol{\theta})\,.$$

Then $Q_n(\mathbf{x}|\boldsymbol{\theta})$ is a continuous function in $\boldsymbol{\theta}$ and a measurable function of $\mathbf{x}$. Let $Q(\boldsymbol{\theta}) = E_{\boldsymbol{\theta}_0} \ln f(X_1|\boldsymbol{\theta})$. Then under assumptions (B1) and (D1) by a theorem of Amemiya (1985), (Theorem 4.2.1., p.116)

$$\tfrac{1}{n} Q_n(\mathbf{X}|\boldsymbol{\theta}) \xrightarrow{P} Q(\boldsymbol{\theta}) \quad \text{uniformly in } \boldsymbol{\theta}.$$

Jensen's inequality yields

$$E_{\boldsymbol{\theta}_0} \ln \frac{f(X_1|\boldsymbol{\theta})}{f(X_1|\boldsymbol{\theta}_0)} \le \ln E_{\boldsymbol{\theta}_0} \frac{f(X_1|\boldsymbol{\theta})}{f(X_1|\boldsymbol{\theta}_0)} = 0\ .$$

The sign of equality holds if and only if

$$\begin{aligned} &\frac{f(X_1|\boldsymbol{\theta})}{f(X_1|\boldsymbol{\theta}_0)} = E_{\boldsymbol{\theta}_0} \frac{f(X_1|\boldsymbol{\theta})}{f(X_1|\boldsymbol{\theta}_0)} = 1 \quad \text{a.s.} \\ \iff \quad & f(X_1|\boldsymbol{\theta}) = f(X_1|\boldsymbol{\theta}_0) \quad \text{a.s.} \quad \iff \quad \boldsymbol{\theta} = \boldsymbol{\theta}_0\ . \end{aligned}$$

Hence for $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$ holds that:

$$E_{\boldsymbol{\theta}_0} \ln \frac{f(X_1|\boldsymbol{\theta})}{f(X_1|\boldsymbol{\theta}_0)} < 0 \quad \iff \quad E_{\boldsymbol{\theta}} \ln f(X_1|\boldsymbol{\theta}) < E_{\boldsymbol{\theta}_0} \ln f(X_1|\boldsymbol{\theta}_0)\ .$$

Thus $Q(\boldsymbol{\theta})$ attains a unique global maximum at $\boldsymbol{\theta}_0$. The consistency of the maximum likelihood estimate now follows from the application of theorem 4.1.1. of Amemiya (1985), p.106, on the function $Q_n(\mathbf{x}|\boldsymbol{\theta})$. □

Definition 7. Let $\mathfrak{Q} = \{Q_{\boldsymbol{\theta}}, \boldsymbol{\theta} = (\theta_1, \ldots, \theta_k)' \in \Theta\}$ be a family of probability measures on a measurable space $(\mathfrak{X}, \mathfrak{A})$. Let the family $\mathfrak{Q}$ be dominated by a σ-finite measure μ and let $f(x|\boldsymbol{\theta})$ denote the density function of $Q_{\boldsymbol{\theta}}$ with respect to μ. If the partial derivatives $\frac{\partial}{\partial\theta_i} \ln f$ exist for all $\boldsymbol{\theta} \in \Theta$ and $i = 1, \ldots, k$ then the matrix $I(\boldsymbol{\theta}) = \big(I_{ij}(\boldsymbol{\theta})\big)_{i=1,\ldots,k}$ consisting of the elements

$$I_{ij}(\boldsymbol{\theta}) = \int \frac{\partial}{\partial\theta_i} \ln f(x|\boldsymbol{\theta})\, \frac{\partial}{\partial\theta_j} \ln f(x|\boldsymbol{\theta})\, f(x|\boldsymbol{\theta})\, d\mu$$

is the *Fisher's information matrix.*

In the following theorem sufficient conditions for the asymptotic normality of a consistent maximum likelihood estimate are given. The theorem and its proof can be found in Witting and Müller-Funk (1995, Satz 6.35).

Theorem 2. *Let $X_1, X_2, \ldots$ be a sequence of i.i.d. random variables defined on a probability space $(\Omega, \mathfrak{F}, P)$ with values in a measurable space $(\mathfrak{X}, \mathfrak{A})$. Let $X_1 \sim Q_{\boldsymbol{\theta}_0} \in \mathfrak{Q} = \{Q_{\boldsymbol{\theta}}, \boldsymbol{\theta} \in \Theta\}$, $\Theta \subset \mathbb{R}^k$, and let $\mathfrak{Q}$ be dominated by a σ-finite measure μ. $f(x|\boldsymbol{\theta})$ denotes the density function of $Q_{\boldsymbol{\theta}}$ with respect to μ. Let Θ° denote the open kernel of the parameter set Θ. Assume that the following conditions are fulfilled.*

(A2) The second partial derivative $\frac{\partial^2}{\partial\theta_i \partial\theta_j} f(x|\boldsymbol{\theta})$ exists for all $x \in \mathfrak{X}$ and all $i, j = 1, \ldots, k$ and it is continuous on Θ°.

(B2) For all $\boldsymbol{\theta} \in \Theta^\circ$ it holds that $\int \frac{\partial}{\partial\theta_i} f(x|\boldsymbol{\theta})\, d\mu = 0$ and $\int \frac{\partial^2}{\partial\theta_i \partial\theta_j} f(x|\boldsymbol{\theta})\, d\mu = 0$ for $i, j = 1, \ldots, k$.

(C2) For every $\boldsymbol{\theta} \in \Theta^\circ$ there exist a $\delta_{\boldsymbol{\theta}} > 0$ with

$$U(\boldsymbol{\theta}, \delta_{\boldsymbol{\theta}}) = \{\tilde{\boldsymbol{\theta}} : |\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}| \leq \delta_{\boldsymbol{\theta}}\} \subset \Theta^\circ$$

and a measurable function $M(x|\boldsymbol{\theta})$ on $\mathfrak{X}$ with $E_{\boldsymbol{\theta}} M(X_1|\boldsymbol{\theta}) < \infty$ such that $\left|\frac{\partial^2}{\partial\theta_i \partial\theta_j} \ln f(x|\tilde{\boldsymbol{\theta}})\right| \leq M(x|\boldsymbol{\theta})$ for all $\tilde{\boldsymbol{\theta}} \in U(\boldsymbol{\theta}, \delta_{\boldsymbol{\theta}})$ and $i, j = 1, \ldots, k$.

(D2) For the determinant of Fisher's information matrix $\det I(\boldsymbol{\theta}) \neq 0$ holds for all $\boldsymbol{\theta} \in \Theta^\circ$.

Then every consistent maximum likelihood estimate $\hat{\boldsymbol{\theta}}_n$ is asymptotically normal, meaning $\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) \xrightarrow{\mathcal{L}} N(0, I(\boldsymbol{\theta})^{-1})$ for all $\boldsymbol{\theta} \in \Theta^\circ$.

Remark. The conditions (A2) -(C2) imply the existence of Fisher's information matrix since:

$$\int \frac{\partial}{\partial\theta_i} \ln f(x|\boldsymbol{\theta})\, \frac{\partial}{\partial\theta_i} \ln f(x|\boldsymbol{\theta})\, f(x|\boldsymbol{\theta})\, d\mu = \int \frac{\partial^2}{\partial\theta_i \partial\theta_j} f(x|\boldsymbol{\theta})\, d\mu - \int \frac{\partial^2}{\partial\theta_i \partial\theta_j} \ln f(x|\boldsymbol{\theta}) \cdot f(x|\boldsymbol{\theta})\, d\mu\ .$$

The asymptotic normality of maximum likelihood estimates is a very important property that justifies the calculation of asymptotic confidence intervals and the application of tests as the likelihood ratio test.

1.4.2 Tests for normality

There are several tests for the normal distribution. D'Agostino and Stephens (1986) suggest tests that are based on the skewness and the kurtosis of a distribution.

Definition 8. Let X be a random variable with $EX^4 < \infty$. Denote $\mu = EX$ and $\sigma^2 = E(X-\mu)^2$. Then

$$\beta_1 = \frac{E(X-\mu)^3}{\sigma^{3/2}}$$

is called the *skewness* of the distribution of X. The parameter

$$\beta_2 = \frac{E(X-\mu)^4}{\sigma^2}$$

is called the *kurtosis* of the distribution of X.

Remark. For the normal distribution it holds that $\beta_1 = 0$ and $\beta_2 = 3$.

The skewness is a measure of the symmetry of a distribution with respect to the mean value whereas the kurtosis is a measure of the peakedness of a distribution. Both values of $\beta_1 \neq 0$ or values of $\beta_2 \neq 3$ indicate non-normality of a distribution.
Let $X_1, \ldots, X_n$ be a random sample of the random variable X. The statistics that are used to estimate skewness and kurtosis empirically are

$$b_1 = \frac{m_3}{m_2^{3/2}} \quad \text{and} \quad b_2 = \frac{m_4}{m_2^2} \tag{1.19}$$

where $m_k = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^k$ and $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$.
D'Agostino and Stephens (1986, section 9.3.3) give a transformation of the statistic b_1 to an approximately standard normal distributed variable Z_1. They say that the transformation is applicable for sample sizes $n \geq 8$. A two-sided test based on the variable Z_1 would reject the null hypothesis

$$H_0 : X \text{ is normally distributed}$$

with a 0.05 level of significance if $|Z_1| \geq 1.96$. In this situation the alternative

$$H_1 : X \text{ is non-normal with } \beta_1 \neq 0$$

would be accepted.
In the same section D'Agostino and Stephens (1986) also give a transformation of the statistic b_2 to an approximately standard normal distributed variable Z_2. The approximation can be used for $n \geq 20$. Based on Z_2 one can test the null hypothesis

$$H_0 : X \text{ is normally distributed}$$

versus the two sided alternative

$$H_1 : X \text{ is non-normal with } \beta_2 \neq 3.$$

H_0 would be rejected with a significance level of 0.05 if $|Z_2| \geq 1.96$.
The transformations that yield the standard normal variables Z_1 and Z_2 are validated theoretically by the δ-method. Let (X_n) denote a sequence of random variables with $\sqrt{n}(X_n - c) \xrightarrow{\mathcal{L}} N(0, \varsigma)$ for some constants c and ς. Then for every function h that is continuously differentiable in c it holds that $\sqrt{n}(h(X_n) - h(c)) \xrightarrow{\mathcal{L}} N(0, [h'(c)]^2\varsigma)$. The δ-method is proven in Witting and Nölle (1970).
The tests described above are used in section 5.3.2 to check the normality of the distribution of maximum likelihood estimates.

1.5 The use of twins in genetic epidemiology

1.5.1 Basics of quantitative genetics

On the one hand it is surely right that almost every trait a researcher could be interested in varies within a human population. On the other hand nobody doubts that there is resemblance between relatives for some traits. Genetic epidemiology tries to use information about family resemblance in order to gain insight into variation within individuals. It is a science that:

> "deals with etiology, distribution, and control of disease in groups of relatives and with inherited causes of disease in populations." (Encyclopedia of Biostatistics, p.1676)

Twins represent a very special case of relationship. There are two types of twins: monozygotic (MZ) and dizygotic (DZ) twins. MZ twins are genetically identical, whereas DZ twins share 50 % of their genes on average like ordinary siblings. This difference can help to evaluate the relative importance of genes and environment on the variation of a trait.
Traditional methods of analyzing twin data look at the differences between the correlations between MZ and DZ twins. A higher correlation for MZ twins hints at a contribution of genetic factors to differences in a trait.
The basics for the biometrical models that describe the effects of genes lie in Mendelian genetics. It would be beyond the scope of this work to go into detail here, but I would refer the reader to Neale and Cardon (1992), who do this in an easily understandable way.

The main assumptions of biometrical genetics (assortative mating, no genotype-environment correlation, no genotype-environment interaction, no epistasis) yield that the phenotypic variance σ_P^2 can be decomposed as:

$$\sigma_P^2 = \sigma_G^2 + \sigma_U^2 = \sigma_A^2 + \sigma_D^2 + \sigma_C^2 + \sigma_E^2 \ . \tag{1.20}$$

σ_G^2 refers to the part of the phenotypic variation that is due to genetic effects in general and σ_U^2 refers to environmental effects in general. Then one makes the distinction between environmental factors that act between families - the so called shared or common environment (σ_C^2) - and factors that act within families - the non-shared or specific environment (σ_E^2). The variation σ_G^2 is split into σ_A^2, which stands for the part of the variation that is due to additive genetic effects, and σ_D^2, which refers to variation that is due to dominance effects. The additive genetic effects are the sum of the effects of all alleles that influence a trait taken singly. Dominance effects are due to the interaction of alleles at a given locus.
The ratio σ_G^2/σ_P^2 is called the *broad sense heritability* whereas the ratio σ_A^2/σ_P^2 is called the *narrow sense heritability*.
The covariance between the members of a twin pair can be expressed in terms of the portions of the whole phenotypic variance:

$$\text{cov(MZ)} = \sigma_A^2 + \sigma_D^2 + \sigma_C^2 \tag{1.21}$$

$$\text{cov(DZ)} = \tfrac{1}{2}\,\sigma_A^2 + \tfrac{1}{4}\,\sigma_D^2 + \sigma_C^2 \tag{1.22}$$

From the observation of a phenotypic trait in a twin population one can estimate the phenotypic variance and the covariances between the members in the same type of twin pair. In contrast we have 4 unknown parameters: σ_A^2, σ_D^2, σ_C^2 and σ_E^2. In order to make the model identifiable one has to make restrictions on the general model by setting one of the parameters equal to zero. The genetic model where σ_D^2 is set equal to zero is abbreviated as ACE. Similarly the genetic models ADE, AE, DE, CE, DCE and E are explained. Usually in genetic analysis the parameter estimates are derived for different models and then the best fitting model is chosen according to various criteria. One of them is the Akaike Information Criterion (AIC) that is computed

$$AIC = \frac{2}{N}(-\ \text{log-likelihood}\ + p) \ .$$

p is the number of parameters and N the sample size, i.e. the number of twin pairs (Akaike, 1974). The model with the lowest value is said to fit best by AIC. Nested models can be compared using a likelihood ratio test. Neale and Cardon (1992) give a very comprehensive introduction into the methods for analyzing twin data.

The traditional model of genetic epidemiology described above can not usually be applied in survival analysis since one normally has to deal with censored or otherwise incomplete data there. Another limitation of the model is that it can only incorporate twin pairs where the trait of interest is observable for both members of the pair. In studies that investigate a trait in elderly twins, but not the life span of the twins, there are, of course, twin pairs where one twin is already deceased. These pairs are called *broken pairs*. They are usually excluded from a study which can reduce the sample size considerably. This means a waste of information.
In order to study the role of genes in the life span or in the ageing process another approach is needed. For the genetic analysis of life span a bivariate correlated frailty model was successfully applied to Danish twin survival data (Yashin and Iachine, 1995b). This model combines the ideas of survival analysis and demography with those of genetic epidemiology. In this approach it is not the genetic parameters of life span that are estimated but those of an unobserved variable called frailty. Frailty represents the individual susceptibility to disease and death. The estimated heritability in frailty was about 50%. The concept of frailty models is introduced in section 1.3.

1.5.2 The genetic liability model

The methods described in section 1.5.1 can only be applied if a quantitative trait is measured in twins. Many phenotypic traits that are supposed to be influenced by the genes are, however, discrete. This can be, for instance, the presence or absence of a certain disease or different levels of severity of a disease can exist. In order to address questions about the relative importance of both genes and the environment for the development of a phenotypic trait that can only be discriminated into ordered categories, the *liability model* is used.
Under the idea that the individual's liability to develop a disease results from the additive effects of many genetic and environmental factors, it is assumed that an unobservable normally distributed random variable, called *liability*, exists, and one or more *threshold* values that discriminate between the categories. In the case of the dichotomous observation, affected versus non-affected, this means that an individual will develop a disease when his liability value exceeds the threshold value T_h. It is assumed that the liability has the mean value 0 and the variance 1. Figure 1.1 illustrates the density function of the liability variable.
The probability of having a disease can be calculated as $1 - \Phi(T_h)$, where Φ denotes the cumulative density function of the standard normal distribution. When a discrete trait with D different categories is observed in twins then the

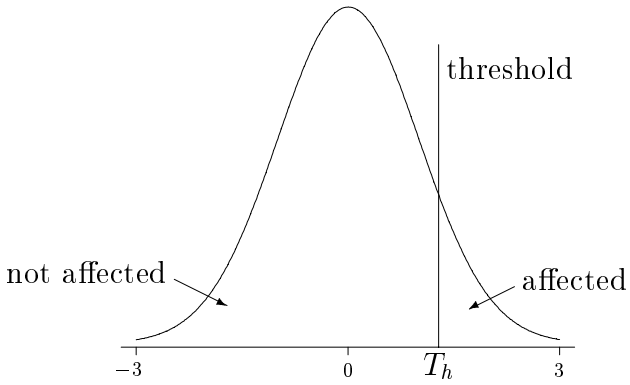


Figure 1.1: Density function of the liability variable.

resulting data set can be expressed as a D-by-D-contingency table. Table 1.1 shows the case of a dichotomous trait.
It is now assumed that the vector of the liability variables of the twins follows a bivariate normal distribution with standard normal margins and the correlation ϱ. Under the assumption that the same threshold value T_h corresponds to each twin it holds that

$$\begin{aligned} P(\text{both twins affected}) &= \int_{T_h}^{\infty}\int_{T_h}^{\infty} \phi(u,v,\varrho)du\,dv \\ P(\text{both twins not affected}) &= \int_{-\infty}^{T_h}\int_{-\infty}^{T_h} \phi(u,v,\varrho)du\,dv \\ P(\text{one twin affected, one not affected}) &= \int_{T_h}^{\infty}\int_{-\infty}^{T_h} \phi(u,v,\varrho)du\,dv \end{aligned} \tag{1.23}$$

where $\phi(u,v,\varrho)$ denotes the density function of the bivariate normal distribution with mean vector $\left(\begin{smallmatrix}0\\0\end{smallmatrix}\right)$ and covariance matrix $\left(\begin{smallmatrix}1 & \varrho\\ \varrho & 1\end{smallmatrix}\right)$.
Given the data in table 1.1 and using (1.23) one can estimate the parameter T_h and ϱ. If this is done for monozygotic and dizygotic twins one gets estimates of the corresponding correlations in liability: ϱ_{MZ} and ϱ_{DZ}. For a dichotomous trait these correlations are called *tetrachoric correlations*.
If there are D categories for the trait of interest then $D-1$ threshold values are introduced and the probabilities of the different cells in the D-by-D contingency table are calculated analogous to (1.23) and one has to integrate over the corresponding intervals.

Table 1.1: General two-by-two contingency table.

twin 1 \ twin 2	affected	not affected	total
affected	n_{11}	n_{10}	$n_{11} + n_{10}$
not affected	n_{01}	n_{00}	$n_{01} + n_{00}$
total	$n_{11} + n_{01}$	$n_{10} + n_{00}$	n

If equations (1.20), (1.21) and (1.22) are divided by the phenotypic variance σ_P^2 then one gets:

$$\begin{aligned} \varrho_{MZ} &= a^2 + d^2 + c^2 \ , \\ \varrho_{DZ} &= \tfrac{1}{2}\, a^2 + \tfrac{1}{4}\, d^2 + c^2 \text{ and} \\ 1 &= a^2 + d^2 + c^2 + e^2 \ , \end{aligned}$$

where $a^2 = \sigma_A^2/\sigma_P^2$ is the proportion of variance associated with additive genetic effects (narrow sense heritability), $d^2 = \sigma_D^2/\sigma_P^2$ is the proportion of variance associated with dominance effects, $c^2 = \sigma_C^2/\sigma_P^2$ is the proportion of variance associated with shared environmental effects and $e^2 = \sigma_E^2/\sigma_P^2$ is the proportion of variance associated with non-shared environmental effects.
Analogous to the case of a quantitative phenotypic trait, different genetic models can be fitted to the correlations in liability and their fit can be compared using the Akaike Information Criterion.
The disadvantages of the liability analysis are the same as for the case of a quantitative trait, i.e. correction for truncation and censoring is not possible and broken pairs are excluded. In Chapter 7 the liability approach is compared with the frailty approach. With the help of simulated data it is shown that the influence of genes on life span can confound the results derived by using the liability analysis. This means that one might find influence of the genes on the liability of the development of a certain disease but originally only the age at death is influenced genetically.

1.5.3 Other measures for the similarities in twin pairs

Let us consider the case of a dichotomous trait, say a certain disease. Let D_i denote the disease status of twin i, $i = 1, 2$, in a pair ($D_i = 1$ if twin i is affected, $D_i = 0$ if twin i is not affected). The observations that one gets for such type of trait in a twin study are presented in table 1.1.
A twin pair is called *concordant* if both members of the pair have the same disease status, otherwise the pair is called *discordant*.

A widely used measure to describe the similarity of the twins in a pair is the *probandwise concordance rate* R. The probandwise concordance rate is a measure that characterizes the conditional probability that a twin is affected given that his or her co-twin is affected. From the data one can estimate the two conditional probabilities $P(D_1 = 1|D_2 = 1)$ and $P(D_2 = 1|D_1 = 1)$. In order to have one single measure for comparison the probandwise concordance rate is used. R is calculated by weighting the conditional probability that twin 1 is affected given that twin 2 is affected and the conditional probability that twin 2 is affected given that twin 1 is affected according to the chance that twin 2 is affected, and the chance that twin 1 is affected, respectively (McGue, 1992):

$$R = \frac{P(D_1 = 1|D_2 = 1) \cdot P(D_2 = 1) + P(D_2 = 1|D_1 = 1) \cdot P(D_1 = 1)}{P(D_1 = 1) + P(D_2 = 1)} .$$

From table 1.1 it follows that the probandwise concordant rate can be estimated as

$$\hat{R} = \frac{2\,n_{11}}{2\,n_{11} + n_{10} + n_{01}} . \tag{1.24}$$

A higher concordance rate for MZ twins compared to DZ twins indicates an influence of the genes on the development of the disease.
Another figure used to describe the dependence within a twin pair is the *relative risk* of the disease for persons whose co-twin has the disease. This approach is based on logistic regression, i.e. the conditional probability to be affected given the health status of the co-twin is expressed as

$$P(D_1 = 1|D_2 = j) = \frac{e^{\alpha+\beta\cdot j}}{1 + e^{\alpha+\beta\cdot j}} .$$

The value e^β is called the relative risk. It is an odds ratio since

$$e^\beta = \frac{\text{odds}(P(D_1 = 1|D_2 = 1))}{\text{odds}(P(D_1 = 1|D_2 = 0))}$$

where $\text{odds}(x) = \frac{x}{1-x}$. Since one can estimate $P(D_1 = 1|D_2 = 1)$ as $\frac{n_{11}}{n_{11}+n_{01}}$ and $P(D_1 = 1|D_2 = 0)$ as $\frac{n_{10}}{n_{10}+n_{00}}$ (see table 1.1), it follows for the estimate of the relative risk that:

$$e^{\hat{\beta}} = \frac{n_{11}\,n_{00}}{n_{10}\,n_{01}} . \tag{1.25}$$

If D_1 and D_2 are independent then $e^\beta = 1$. If $e^\beta > 1$ then this indicates interdependence between the different health status levels of the twins. Again a higher relative risk for monozygotic versus dizygotic twins indicates the influence of genes.

Chapter 2

A correlated frailty model for describing the ageing process

2.1 Introduction of the model

2.1.1 Basic structure

The ageing process is described as a stochastic process consisting of the three states "healthy", "ill", and "deceased". Such a process can be specified by the multivariate distribution of its jump times, i.e. the times where an individual changes from one state to another. Figure 2.1 shows a sample path of the ageing process.

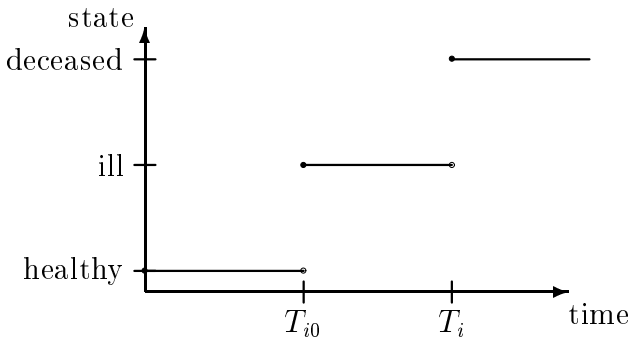


Figure 2.1: 3-state-process and its jumping times

The ageing of an individual can be described with the time he stays in a "healthy" state (T_{i0}) and the life span (T_i). If there is an underlying unobservable frailty variable corresponding to each of these times, then one

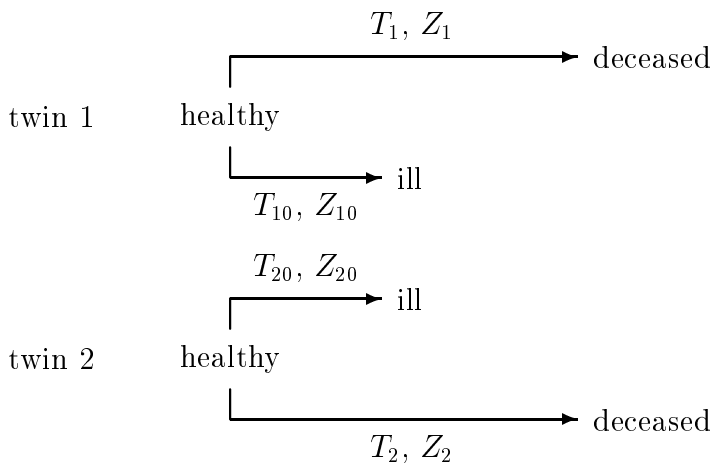


Figure 2.2: Four-dimensional correlated frailty model for a twin pair

represents factors that influence susceptibility to disease and the other represents factors that change individual chances of survival.
The ageing of a twin pair can be described in this way by a four-dimensional correlated frailty model. Figure 2.2 illustrates such a model. T_{i0} is the time spent in a healthy state and T_i is the life span of twin i, $i = 1, 2$. The frailty variables Z_{i0} and Z_i, $i = 1, 2$, correspond to these time periods respectively. We assume that it makes no difference whether a twin is the first or the second born in a pair. Therefore the bivariate distributions of (T_{10}, T_1) and (T_{20}, T_2) as well as those of (Z_{10}, Z_1) and (Z_{20}, Z_2) are assumed to be equal. Let S_0 and S_L denote the univariate survival functions of T_{i0}, $i = 1, 2$, and T_i, $i = 1, 2$, respectively. $S_L(x)$ is the probability of surviving age x and $S_0(x)$ is the probability of still being healthy at age x. A proportional hazard model with the underlying baseline hazard rate μ_0 and the cumulative baseline hazard rate $H_0(x) = \int_0^x \mu_0(u)\, du$ is assumed for the conditional survival function of T_{i0} given Z_{i0}. The conditional survival function of T_i given Z_i is defined similarly, with the underlying baseline hazard rate μ and the cumulative baseline hazard rate $H(x) = \int_0^x \mu(u)\, du$:

$$P(T_{i0} > x|Z_{i0}) = \exp\{-Z_{i0}H_0(x)\} \quad \text{and} \quad P(T_i > x|Z_i) = \exp\{-Z_iH(x)\}\,. \tag{2.1}$$

It is assumed that the times T_{i0}, $i = 1, 2$, and T_i, $i = 1, 2$, are conditionally independent given the frailties and that the conditioning on the frailty

variables can be stratified.

$$P(T_{10} > s_1, T_{20} > s_2, T_1 > t_1, T_2 > t_2 | Z_{10}, Z_{20}, Z_1, Z_2) = \\ P(T_{10} > s_1 | Z_{10}) \cdot P(T_{20} > s_2 | Z_{20}) \cdot P(T_1 > t_1 | Z_1) \cdot P(T_2 > t_2 | Z_2) \quad (2.2)$$

This assumption means that the dependence between the times T_{10}, T_{20}, T_1 and T_2 can be completely explained by the dependence between the corresponding frailty variables.
Equations (2.1) and (2.2) yield for the unconditional survival function S of $(T_{10}, T_{20}, T_1, T_2)$:

$$S(s_1, s_2, t_1, t_2) = P(T_{10} > s_1, T_{20} > s_2, T_1 > t_1, T_2 > t_2) \\ = E \exp\{-Z_{10} H_0(s_1) - Z_{20} H_0(s_2) - Z_1 H(t_1) - Z_2 H(t_2)\} . \quad (2.3)$$

Thus, to calculate the unconditional four-dimensional survival function S of $(T_{10}, T_{20}, T_1, T_2)$ one has to specify the distribution of $(Z_{10}, Z_{20}, Z_1, Z_2)$ or, more precisely, the Laplace transform of this random vector.

2.1.2 Modeling of the frailty distribution

To construct the joint distribution of the four frailty variables one has to model their dependence. There is a wide range of dependence structures that can be described by different four-dimensional distributions. A very clear structure can be reached by linear combinations of independent random variables. Correspondingly one can represent the frailty variables as the sum of independent random variables. Let Y_i, $i = 0, 1, \ldots, 4$, and Y_{ij}, $i = 1, 2$, $j = 0, 1$, be independent random variables. They represent different parts of all risk factors that influence health and survival. Their exact interpretation gives us the following table.

Y_0	risk factors that influence health and survival and that are shared by both twins
Y_i	risk factors that influence health and survival only for twin i, $i = 1, 2$
Y_3	risk factors that influence only health and that are shared by both twins
Y_4	risk factors that influence only survival and that are shared by both twins
Y_{ij}	remaining individual risk factors that influence only health or survival and that are not shared by the twins

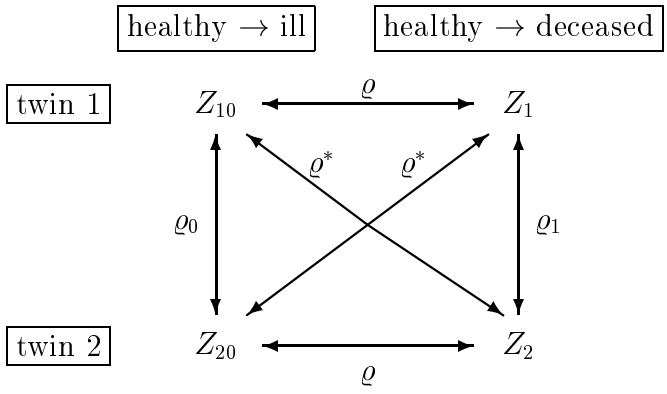


Figure 2.3: Correlation structure of the four-dimensional frailty distribution

Assumption 1. Our model assumption is that the frailty variables admit the representation:

$$\begin{aligned} Z_{10} &= \sigma_0^2(Y_0 + Y_1 + Y_3 + Y_{10}) \qquad Z_1 = \sigma^2(Y_0 + Y_1 + Y_4 + Y_{11}) \\ Z_{20} &= \sigma_0^2(Y_0 + Y_2 + Y_3 + Y_{20}) \qquad Z_2 = \sigma^2(Y_0 + Y_2 + Y_4 + Y_{21}) \end{aligned} \tag{2.4}$$

It is assumed that the Y_i and Y_{ij} are independent gamma-distributed random variables with the scale parameter 1 and different shape parameters. Since the sequence of the twins in a pair is arbitrary the distribution of Y_1 and Y_2, Y_{10} and Y_{20}, and Y_{11} and Y_{21}, respectively, should be the same.

$$\begin{aligned} Y_1 \sim \Gamma(k_1, 1), \quad Y_2 \sim \Gamma(k_1, 1), \quad Y_i \sim \Gamma(k_i, 1), i = 0, 3, 4, \\ Y_{i0} \sim \Gamma(k_{10}, 1), i = 1, 2, \quad Y_{i1} \sim \Gamma(k_{11}, 1), i = 1, 2. \end{aligned} \tag{2.5}$$

In addition, it is assumed that the frailty variables Z_{10}, Z_{20}, Z_1 and Z_2 have the mean value 1.

Notice that the representation (2.4) is similar to a two-way random effects model of analysis of variance. In the context of analysis of variance, Y_0 would be the common mean effect, Y_1 and Y_2 would be the effect of twin, Y_3 and Y_4 would be the effect of health state, and the Y_{ij} would be the error terms. If one applies the properties of the gamma distribution (see page 6) from (2.4) it follows that the frailty variables are also gamma distributed. For the variances it holds that: $\sigma_0^2 = \mathrm{Var}(Z_{10}) = \mathrm{Var}(Z_{20})$ and $\sigma^2 = \mathrm{Var}(Z_1) = \mathrm{Var}(Z_2)$. Denote the correlations between the frailty variables in the following way: $\varrho_0 = \mathrm{corr}(Z_{10}, Z_{20})$, $\varrho_1 = \mathrm{corr}(Z_1, Z_2)$, $\varrho = \mathrm{corr}(Z_{10}, Z_1) = \mathrm{corr}(Z_{20}, Z_2)$, $\varrho^* = \mathrm{corr}(Z_{10}, Z_2) = \mathrm{corr}(Z_{20}, Z_1)$. Figure 2.3 illustrates the correlation structure.

There is a one-to-one correspondence between the shape parameters of the components of the frailty variables and the variances and correlations of the frailty variables.

Lemma 1. *For the shape parameters k_0, k_1, k_3, k_4, k_{10}, k_{11}, the variances σ_0^2, σ^2 and the correlations ϱ_0, ϱ_1, ϱ, ϱ^* it holds that:*

$$\begin{aligned} \sigma_0 &= (k_0 + k_1 + k_3 + k_{10})^{(-1/2)} \\ \sigma &= (k_0 + k_1 + k_4 + k_{11})^{(-1/2)} \\ \varrho_0 &= \sigma_0^2(k_0 + k_3) \\ \varrho_1 &= \sigma^2(k_0 + k_4) \\ \varrho &= \sigma_0\sigma(k_0 + k_1) \\ \varrho^* &= \sigma_0\sigma k_0 \ . \end{aligned} \tag{2.6}$$

Proof. From (2.5) it follows that (see page 6)

$$EY_1 = EY_2 = k_1,\ EY_i = k_i, i = 0, 3, 4, \text{ and } EY_{ij} = k_{1j}, i = 1, 2, j = 0, 1,$$

and

$$\begin{aligned} \mathrm{Var}(Y_1) = \mathrm{Var}(Y_2) = k_1,\ \mathrm{Var}(Y_i) = k_i, i = 0, 3, 4, \text{ and} \\ \mathrm{Var}(Y_{ij}) = k_{1j}, i = 1, 2, j = 0, 1. \end{aligned}$$

The assumption of independence and (2.4) yield therefore

$$\begin{aligned} EZ_{10} = EZ_{20} &= \sigma_0^2(k_0 + k_1 + k_3 + k_{10}) \\ EZ_1 = EZ_2 &= \sigma^2(k_0 + k_1 + k_4 + k_{11}) \\ \mathrm{cov}(Z_{10}, Z_{20}) &= \sigma_0^4(\mathrm{Var}(Y_0) + \mathrm{Var}(Y_3)) = \sigma_0^4(k_0 + k_3) \\ \mathrm{cov}(Z_1, Z_2) &= \sigma^4(\mathrm{Var}(Y_0) + \mathrm{Var}(Y_4)) = \sigma^4(k_0 + k_3) \\ \mathrm{cov}(Z_{10}, Z_1) &= \sigma_0^2\sigma^2(\mathrm{Var}(Y_0) + \mathrm{Var}(Y_1)) = \sigma_0^2\sigma^2(k_0 + k_1) \\ \mathrm{cov}(Z_{20}, Z_2) &= \sigma_0^2\sigma^2(\mathrm{Var}(Y_0) + \mathrm{Var}(Y_2)) = \sigma_0^2\sigma^2(k_0 + k_1) \\ \mathrm{cov}(Z_{10}, Z_2) = \mathrm{cov}(Z_{20}, Z_1) &= \sigma_0^2\sigma^2\,\mathrm{Var}(Y_0) = \sigma_0^2\sigma^2 k_0 \ . \end{aligned}$$

The assertion of the lemma is then obvious. □

Remark. The representation of the shape parameters as a function of the variances and correlations is given in the Appendix in (A.1). The shape parameters must be positive. Therefore from (A.1) one gets the following constraints on the variances and correlations:

$$\begin{aligned} 0 < &\varrho^* < \varrho \\ \varrho^*\sigma_0 < &\varrho_0\sigma < \sigma - (\varrho - \varrho^*)\sigma_0 \\ \varrho^*\sigma < &\varrho_1\sigma_0 < \sigma_0 - (\varrho - \varrho^*)\sigma \end{aligned} \tag{2.7}$$

2.1.3 The survival function of $(\mathbf{T_{10}, T_{20}, T_1, T_2})$

Together with the constraints (2.7) the four-dimensional correlated gamma frailty model is described by the unconditional four-dimensional survival function of $(T_{10}, T_{20}, T_1, T_2)$.

Proposition 1. *Let $(T_{10}, T_{20}, T_1, T_2)$ be a vector of four survival times and $(Z_{10}, Z_{20}, Z_1, Z_2)$ be a vector of four random variables, both defined on a probability space $(\Omega, \mathfrak{F}, P)$. The vector $(Z_{10}, Z_{20}, Z_1, Z_2)$ is distributed according to assumption 1. It is assumed that the univariate conditional survival functions of T_{i0} given Z_{i0}, $i = 1, 2$, and T_i given Z_i, $i = 1, 2$, respectively, are given by (2.1) and that the conditional survival function of $(T_{10}, T_{20}, T_1, T_2)$ given $(Z_{10}, Z_{20}, Z_1, Z_2)$ fulfills (2.2). Under these assumptions the unconditional four-dimensional survival function S of $(T_{10}, T_{20}, T_1, T_2)$ is given by*

$$
\begin{aligned}
&S(s_1, s_2, t_1, t_2) \\
&= \left[\left(1 + \sigma_0^2 H_0(s_1)\right)\left(1 + \sigma_0^2 H_0(s_2)\right)\right]^{-k_{10}} \left[\left(1 + \sigma^2 H(t_1)\right)\left(1 + \sigma^2 H(t_2)\right)\right]^{-k_{11}} \\
&\quad \cdot \left[\left(1 + \sigma_0^2 H_0(s_1) + \sigma^2 H(t_1)\right)\left(1 + \sigma_0^2 H_0(s_2) + \sigma^2 H(t_2)\right)\right]^{-k_1} \\
&\quad \cdot \left[1 + \sigma_0^2 H_0(s_1) + \sigma_0^2 H_0(s_2)\right]^{-k_3} \left[1 + \sigma^2 H(t_1) + \sigma^2 H(t_2)\right]^{-k_4} \\
&\quad \cdot \left[1 + \sigma_0^2 H_0(s_1) + \sigma_0^2 H_0(s_2) + \sigma^2 H(t_1) + \sigma^2 H(t_2)\right]^{-k_0} \qquad (2.8)
\end{aligned}
$$

$$
\begin{aligned}
&= \left[S_0(s_1) S_0(s_2)\right]^{1-\varrho_0-(\varrho-\varrho^*)\frac{\sigma_0}{\sigma}} \cdot \left[S_L(t_1) S_L(t_2)\right]^{1-\varrho_1-(\varrho-\varrho^*)\frac{\sigma}{\sigma_0}} \\
&\quad \cdot \left[\left(S_0^{-\sigma_0^2}(s_1) + S_L^{-\sigma^2}(t_1) - 1\right)\left(S_0^{-\sigma_0^2}(s_2) + S_L^{-\sigma^2}(t_2) - 1\right)\right]^{-\frac{\varrho-\varrho^*}{\sigma_0\sigma}} \\
&\quad \cdot \left[S_0^{-\sigma_0^2}(s_1) + S_0^{-\sigma_0^2}(s_2) - 1\right]^{-\frac{\varrho_0}{\sigma_0^2}+\frac{\varrho^*}{\sigma_0\sigma}} \left[S_L^{-\sigma^2}(t_1) + S_L^{-\sigma^2}(t_2) - 1\right]^{-\frac{\varrho_1}{\sigma^2}+\frac{\varrho^*}{\sigma_0\sigma}} \\
&\quad \cdot \left[S_0^{-\sigma_0^2}(s_1) + S_0^{-\sigma_0^2}(s_2) + S_L^{-\sigma^2}(t_1) + S_L^{-\sigma^2}(t_2) - 3\right]^{-\frac{\varrho^*}{\sigma_0\sigma}} \qquad (2.9)
\end{aligned}
$$

where S_0 denotes the univariate survival function of T_{i0}, $i = 1, 2$, and S_L denotes the univariate survival function of T_i, $i = 1, 2$.

Proof. (2.3) shows that one has to calculate the four-dimensional Laplace transform of $(Z_{10}, Z_{20}, Z_1, Z_2)$ to get the survival function of $(T_{10}, T_{20}, T_1, T_2)$. From (2.4), (2.5) and (1.7) it follows that

$$
\begin{aligned}
&Ee^{-u_1 Z_{10} - u_2 Z_{20} - v_1 Z_1 - v_2 Z_2} \\
&= Ee^{-(\sigma_0^2 u_1 + \sigma_0^2 u_2 + \sigma^2 v_1 + \sigma^2 v_2) Y_0} Ee^{-(\sigma_0^2 u_1 + \sigma^2 v_1) Y_1} Ee^{-(\sigma_0^2 u_2 + \sigma^2 v_2) Y_2} \\
&\quad \cdot Ee^{-(\sigma_0^2 u_1 + \sigma_0^2 u_2) Y_3} Ee^{-(\sigma^2 v_1 + \sigma^2 v_2) Y_4} Ee^{-\sigma_0^2 u_1 Y_{10}} Ee^{-\sigma_0^2 u_2 Y_{20}} Ee^{-\sigma^2 v_1 Y_{11}} Ee^{-\sigma^2 v_2 Y_{21}}
\end{aligned}
$$

$$
\begin{aligned}
=&[1+\sigma_0^2u_1+\sigma_0^2u_2+\sigma^2v_1+\sigma^2v_2]^{-k_0}[1+\sigma_0^2u_1+\sigma^2v_1]^{-k_1}[1+\sigma_0^2u_2+\sigma^2v_2]^{-k_1}\\
&\cdot[1+\sigma_0^2u_1+\sigma_0^2u_2]^{-k_3}[1+\sigma^2v_1+\sigma^2v_2]^{-k_4}[(1+\sigma_0^2u_1)(1+\sigma_0^2u_2)]^{-k_{10}}\\
&\cdot[(1+\sigma^2v_1)(1+\sigma^2v_2)]^{-k_{11}}.
\end{aligned}
$$

Together with (2.3) this yields (2.8). From (2.1) and (1.8) it follows that

$$S_0(x)=[1+\sigma_0^2H_0(x)]^{-1/\sigma_0^2} \quad \text{and} \quad S_L(x)=[1+\sigma^2H(x)]^{-1/\sigma^2}.$$

This implies

$$\sigma_0^2H_0(x)=S_0^{-\sigma_0^2}(x)-1 \quad \text{and} \quad \sigma^2H(x)=S_L^{-\sigma^2}(x)-1\ . \tag{2.10}$$

Expressing the k_i and k_{ij} in terms of the variances and correlations according to (A.1) and substituting (2.10) into (2.8) yields (2.9). □

Remark. The second representation (2.9) could be called semi-parametric since it shows the functional dependence of the survival function S on the marginal univariate survival functions S_0 and S_L, which may be possibly estimated non-parametrically from univariate data (see also section 2.3).

Remark. The marginal bivariate survival functions can easily be derived from (2.8) and (2.9). They are all of the type of a bivariate correlated gamma frailty model (see (1.16)).

$$
\begin{aligned}
S_T(s,t)&=P(T_{10}>s,T_1>t)=P(T_{20}>s,T_2>t)\\
&=[S_0(s)]^{1-\frac{\sigma_0}{\sigma}\varrho}\,[S_L(t)]^{1-\frac{\sigma}{\sigma_0}\varrho}\left[S_0^{-\sigma_0^2}(s)+S_L^{-\sigma^2}(t)-1\right]^{-\frac{\varrho}{\sigma_0\sigma}} \qquad (2.11)\\
S_{0*}(s_1,s_2)&=P(T_{10}>s_1,T_{20}>s_2)\\
&=[S_0(s_1)S_0(s_2)]^{1-\varrho_0}\left[S_0^{-\sigma_0^2}(s_1)+S_0^{-\sigma_0^2}(s_2)-1\right]^{-\frac{\varrho_0}{\sigma_0^2}} \qquad (2.12)\\
S_{L*}(t_1,t_2)&=P(T_1>t_1,T_2>t_2)\\
&=[S_L(t_1)S_L(t_2)]^{1-\varrho_1}\left[S_L^{-\sigma^2}(t_1)+S_L^{-\sigma^2}(t_2)-1\right]^{-\frac{\varrho_1}{\sigma^2}} \qquad (2.13)\\
S_D(s,t)&=P(T_{10}>s,T_2>t)=P(T_{20}>s,T_1>t)\\
&=[S_0(s)]^{1-\frac{\sigma_0}{\sigma}\varrho^*}\,[S_L(t)]^{1-\frac{\sigma}{\sigma_0}\varrho^*}\left[S_0^{-\sigma_0^2}(s)+S_L^{-\sigma^2}(t)-1\right]^{-\frac{\varrho^*}{\sigma_0\sigma}} \qquad (2.14)
\end{aligned}
$$

2.1.4 Definition of the model

The four-dimensional correlated frailty model was introduced thus far under the assumption of positive variances σ_0^2 and σ^2 and strict less-than-signs in

the constraints in (2.7). The constraints can be modified in the following way without changing the properties of the model:

$$\begin{aligned} r \leq & \varrho^* \leq \varrho \\ \varrho^* \sigma_0 \leq & \varrho_0 \sigma \leq \sigma - (\varrho - \varrho^*)\sigma_0 \\ \varrho^* \sigma \leq & \varrho_1 \sigma_0 \leq \sigma_0 - (\varrho - \varrho^*)\sigma \end{aligned} \tag{2.15}$$

Here r is a positive arbitrary constant. Allowing for equality in the inequalities is equivalent to allowing the parameters k_i, $i = 1, 3, 4$, and k_{ij}, $i = 1, 2$, $j = 0, 1$, to take the value zero.
If, for example, the shape parameter k_1 is zero this corresponds to the limiting case that the variable Y_1 takes the value zero with probability one. This can be interpreted that there is no variability in the risk factors that influence health and survival only for one twin.
The functions (2.8) and (2.9) remain four-dimensional survival functions under the constraints (2.15) as well.
In the following the constraints (2.15) are used. Thus the parameter space Θ_r is

$$\Theta_r = \{(\sigma_0, \sigma, \varrho_0, \varrho_1, \varrho, \varrho^*) : 0 \leq \varrho_0, \varrho_1, \varrho, \varrho^* \leq 1, \quad 0 < \sigma_0, \sigma \text{ and } (2.15)\}.$$

Definition 9. A vector of non-negative random variables $(T_{10}, T_{20}, T_1, T_2)$ is said to be distributed according to a *four-dimensional correlated gamma frailty model* if a parameter vector $(\sigma_0, \sigma, \varrho_0, \varrho_1, \varrho, \varrho^*) \in \Theta_r$ and univariate survival functions S_0 and S_L exist such that the four-dimensional survival function of $(T_{10}, T_{20}, T_1, T_2)$ is given by (2.8).

2.2 Current health status data

2.2.1 Description of the observation scheme

It is quite unrealistic to assume that one can get exact information about how long people stay healthy and it is often difficult to define or distinguish the real onset of a disease. In addition, life span information is usually censored, and it may happen that people die before they fall ill. That means that one cannot observe the vector $(T_{10}, T_{20}, T_1, T_2)$ directly but can observe other variables that are a function of this vector. This loss of information can imply the loss of the identifiability of the model. It has to be mentioned here that the model includes a special competing risk situation. The observable data never contain information about T_{i0} if $T_i < T_{i0}$, i.e. if an individual dies before getting ill we don't know anything about the disease. The situation is

different from the usual competing risk model. The latter would mean that the minimum of T_{i0} and T_i is always observable. One might call the situation for the four-dimensional correlated frailty model a half-competing risk situation. It is also important to note that in general a model of dependent competing risks is not identifiable (Prakasa Rao, 1992, section 7.4), whereas it can be shown that a more realistic pattern of censored observations in the four-dimensional correlated frailty model is sufficient for the identifiability of the model (see section 3.3). We call this pattern of observation current health status data.

Let us assume that one has carried out a cross-sectional study where one examined the health state of twin pairs. This means that at a certain moment in time the health of twins is investigated. Assume further that it is possible to obtain the exact life span of twins of the same birth cohorts as the observed ones who have already died from a population registry. Pairs where one or two of the twins are already deceased are included in the study.

Naturally one knows the age or the hypothetical age of every twin individual at the time of the survey. For living individuals this is their real age and for deceased individuals this is the age they would have been on the day of the survey. The difference of the age or the hypothetical age of an individual at the time of the survey and the time of birth will be referred to as A, the *age at the time of observation*, from now on. Since twins are usually born on the same day, A is the same for each twin in a pair. If one includes different birth cohorts of twins, i.e. one has a random time of birth and a deterministic time for the survey, then A is a random variable. It is a censoring variable here.

It is important to distinguish this observation scheme from the usual model of right censoring where the end of the observation period is random. For the type of observation used here the time of observation is deterministic but the time of birth is random. Therefore the age at the time of observation, that is the difference between those two times, is also random. As a consequence the range of birth cohorts influences the proportion of healthy, ill and deceased individuals at the time of the survey.

For twin i, $i = 1, 2$, let T_{i0} denote the time the twin stays healthy and let T_i denote the life span (see figure 2.2). A twin is healthy at the time of the survey if $T_{i0} > A$ and $T_i > A$, he or she is ill if $T_{i0} \leq A$ and $T_i > A$, and he or she is deceased if $T_i \leq A$. It is assumed that for deceased persons there is no information whether they had a certain disease or not.

In the context of this work *current health status data* means that if A is the age of a twin pair at the time of observation then one can, for twin i, $i = 1, 2$,

observe the current health state:

$$C_i = \begin{cases} 0 & \text{if the twin is healthy} \iff T_{i0} > A \text{ and } T_i > A, \\ 1 & \text{if the twin is ill} \iff T_{i0} \leq A \text{ and } T_i > A, \\ 2 & \text{if the twin is deceased} \iff T_i \leq A, \end{cases} \tag{2.16}$$

and the censored life span:

$$X_i = \min(T_i, A) \qquad i = 1, 2. \tag{2.17}$$

The observation for a twin pair is the vector (X_1, X_2, C_1, C_2). X_1 and X_2 are positive real numbers, C_1 and C_2 can take the value 0, 1 or 2.
There are 6 possible cases of observations:

1. both twins are healthy and alive:
   $(X_1, X_2, C_1, C_2) = (A, A, 0, 0)$
2. one twin is ill, the other is healthy, both are alive:
   $(X_1, X_2, C_1, C_2) = (A, A, 1, 0)$ or $(A, A, 0, 1)$
3. both twins are ill and alive:
   $(X_1, X_2, C_1, C_2) = (A, A, 1, 1)$
4. one twin is healthy and alive, the other twin is deceased:
   $(X_1, X_2, C_1, C_2) = (T_1, A, 2, 0)$ or $(A, T_2, 0, 2)$
5. one twin is ill and alive, the other twin is deceased:
   $(X_1, X_2, C_1, C_2) = (T_1, A, 2, 1)$ or $(A, T_2, 1, 2)$
6. both twins are deceased:
   $(X_1, X_2, C_1, C_2) = (T_1, T_2, 2, 2)$

Current health status data are a special case of incomplete observations.

2.2.2 Calculation of the likelihood function

Under the assumption that the distribution of the vector $(T_{10}, T_{20}, T_1, T_2)$ can be described with a correlated frailty model and choosing an appropriate measure on the sample space allows us to calculate the density function f_X of (X_1, X_2, C_1, C_2). The sample space of (X_1, X_2, C_1, C_2) is

$$(\mathcal{X}, \mathfrak{A}) = (\mathbb{R}_+^2 \times \{0, 1, 2\} \times \{0, 1, 2\}, \mathfrak{B}^2 \otimes \mathcal{P}^{\{0,1,2\}} \otimes \mathcal{P}^{\{0,1,2\}}) \,. \tag{2.18}$$

Definition 10. Let λ^1 and λ^2 denote the Lebesgue measure on $\mathbb{R}^1$ and $\mathbb{R}^2$, respectively. Let $A, B \in \mathfrak{B}^1$ be two Borel sets. Let a function $\mu^* : (\mathcal{X}, \mathfrak{A}) \to \mathbb{R}_+$ be defined by

$$\mu^*(A \times B \times \{i\} \times \{j\}) = \begin{cases} \lambda^1(A \cap B) & \text{if } i \leq 1 \text{ and } j \leq 1, \\ \lambda^2(A \times B) & \text{if } i = 2 \text{ or } j = 2, \end{cases}$$

and $\mu^*(A \times B \times C \times D) = \sum_{i \in C, j \in D} \mu^*(A \times B \times \{i\} \times \{j\})$. Then μ is defined as the extension of μ^* to a measure on $(\mathcal{X}, \mathfrak{A})$.

Corollary 1. *For every measurable function $g : (\mathcal{X}, \mathfrak{A}) \to (\mathbb{R}^1, \mathfrak{B}^1)$ it holds that*

$$\int g(x_1, x_2, i, j) d\mu = \begin{cases} \int g(x, x, i, j)\, d\lambda^1 & \text{if } i \leq 1 \text{ and } j \leq 1, \\ \int g(x_1, x_2, i, j)\, d\lambda^2 & \text{if } i = 2 \text{ or } j = 2. \end{cases} \tag{2.19}$$

Proposition 2. *Let $(T_{10}, T_{20}, T_1, T_2)$ be a random vector of 4 survival times, let S denote the survival function of $(T_{10}, T_{20}, T_1, T_2)$ and let the vector have a Lebesgue density. The random variable A and the vector (X_1, X_2, C_1, C_2) are defined as on page 32. Let the age at the time of observation A be independent of $(T_{10}, T_{20}, T_1, T_2)$. If A has got the Lebesgue density f_A and the survival function S_A then the density function of the vector (X_1, X_2, C_1, C_2) with respect to μ as defined in definition 10 is given by*

$$\begin{aligned}
f_X(x_1, x_2, c_1, c_2) &= S(x_1, x_2, x_1, x_2) f_A(x_1)\, \mathrm{I}(x_1 = x_2)\, \mathrm{I}(c_1 = 0)\, \mathrm{I}(c_2 = 0) \\
&\quad + \big(S(0, x_2, x_1, x_2) - S(x_1, x_2, x_1, x_2)\big) f_A(x_1)\, \mathrm{I}(x_1 = x_2)\, \mathrm{I}(c_1 = 1)\, \mathrm{I}(c_2 = 0) \\
&\quad + \big(S(x_1, 0, x_1, x_2) - S(x_1, x_2, x_1, x_2)\big) f_A(x_2)\, \mathrm{I}(x_1 = x_2)\, \mathrm{I}(c_1 = 0)\, \mathrm{I}(c_2 = 1) \\
&\quad + \big(S(0, 0, x_1, x_2) - S(0, x_2, x_1, x_2) - S(x_1, 0, x_1, x_2) + S(x_1, x_2, x_1, x_2)\big) \\
&\qquad \cdot f_A(x_1)\, \mathrm{I}(x_1 = x_2)\, \mathrm{I}(c_1 = 1)\, \mathrm{I}(c_2 = 1) \\
&\quad - \left.\frac{\partial S(s_1, s_2, t_1, t_2)}{\partial t_1}\right|_{\substack{s_1 = 0, s_2 = x_2 \\ t_1 = x_1, t_2 = x_2}} \cdot f_A(x_2)\, \mathrm{I}(x_1 \leq x_2)\, \mathrm{I}(c_1 = 2)\, \mathrm{I}(c_2 = 0) \\
&\quad - \left.\frac{\partial S(s_1, s_2, t_1, t_2)}{\partial t_2}\right|_{\substack{s_1 = x_1, s_2 = 0 \\ t_1 = x_1, t_2 = x_2}} \cdot f_A(x_1)\, \mathrm{I}(x_1 \geq x_2)\, \mathrm{I}(c_1 = 0)\, \mathrm{I}(c_2 = 2) \\
&\quad + \left[\left.\frac{\partial S(s_1, s_2, t_1, t_2)}{\partial t_1}\right|_{\substack{s_1 = 0, s_2 = x_2 \\ t_1 = x_1, t_2 = x_2}} - \left.\frac{\partial S(s_1, s_2, t_1, t_2)}{\partial t_1}\right|_{\substack{s_1 = s_2 = 0 \\ t_1 = x_1, t_2 = x_2}}\right] \\
&\qquad \cdot f_A(x_2)\, \mathrm{I}(x_1 \leq x_2)\, \mathrm{I}(c_1 = 2)\, \mathrm{I}(c_2 = 1)
\end{aligned}$$

$$
\begin{aligned}
&+ \left[\left. \frac{\partial S(s_1, s_2, t_1, t_2)}{\partial t_2} \right|_{\substack{s_1 = x_1, s_2 = 0 \\ t_1 = x_1, t_2 = x_2}} - \left. \frac{\partial S(s_1, s_2, t_1, t_2)}{\partial t_2} \right|_{\substack{s_1 = s_2 = 0 \\ t_1 = x_1, t_2 = x_2}} \right] \\
&\qquad \cdot f_A(x_1)\, \mathrm{I}(x_1 \geq x_2)\, \mathrm{I}(c_1 = 1)\, \mathrm{I}(c_2 = 2) \\
&+ \left. \frac{\partial^2 S(s_1, s_2, t_1, t_2)}{\partial t_1 \partial t_2} \right|_{\substack{s_1 = s_2 = 0 \\ t_1 = x_1, t_2 = x_2}} \cdot S_A(\max(x_1, x_2))\, \mathrm{I}(c_1 = 2)\, \mathrm{I}(c_2 = 2)
\end{aligned}
\tag{2.20}
$$

Proof. f_X is the desired density if for all x, y and all $i, j \in \{0, 1, 2\}$ the following holds:

$$
P(X_1 > x, X_2 > y, C_1 = i, C_2 = j) = \int \mathrm{I}_{(x,\infty)}(x_1)\, \mathrm{I}_{(y,\infty)}(x_2) f_X(x_1, x_2, i, j)\, d\mu.
$$

This can be checked for the different cases using (2.19). The explicit formulas are given in Appendix B.1 on page 114. □

2.3 Estimation strategies

For current health status data (see section 2.2.1) the log-likelihood function in the four-dimensional correlated gamma frailty model can be expressed as a function depending on the parameters of the four-dimensional frailty distribution σ_0, σ, ϱ_0, ϱ_1, ϱ and ϱ^* and the univariate survival functions S_0 and S_L (see (2.9) and (2.20)). The proof of identifiability shows that the univariate survival functions can be identified non-parametrically, i.e. without any parametric specification (see section 3.3). This property opens up different possibilities for the estimation of the parameters in the model and also provides us with the opportunity of evaluating the fit of the model. On the one hand, one can choose a parametric specification for the univariate survival functions S_0 and S_L or equivalently for the underlying hazard rates μ_0 and μ. Then the model becomes a purely parametric one and common methods of parameter estimation, such as maximum likelihood estimation, can be applied.
One has censored information for the life span of every individual, where the age at the time of observation A is the censoring variable. On the other hand, therefore, one can calculate the Kaplan-Meier-Estimator for the univariate survival function S_L from the data. One can use the values of this non-parametric estimator in the log-likelihood function and choose a parametric specification for the function S_0 or μ_0. This approach could be called a partly semi-parametric estimation strategy for the model.

It is important to keep in mind that the standard errors that one would get with traditional methods using this semi-parametric approach do not include the error that is created by using a non-parametric estimator for S_L. Therefore one has to use other methods, such as a bootstrap approach, to get standard errors of the parameter estimates.
Using a fully parametric approach has got the advantage that well-known theoretical results about the asymptotic distribution of the estimators exist. Thus standard errors can be derived directly. The disadvantage is the higher number of parameters that have to be estimated simultaneously and of course one makes restrictions on the shape of the survival functions S_0 and S_L. A semi-parametric approach does not make such restrictions but requires additional computational efforts to calculate the non-parametric estimate. To get standard errors for the remaining parameter estimates is not straightforward and is usually also computationally demanding.
If the parameter estimates from the semi-parametric approach and a fully parametric approach differ substantially, this can be a hint that the chosen parametric specification for S_L is inappropriate.
It is not possible to calculate the Kaplan-Meier-Estimator for the survival function S_0 on the basis of current health status data since one knows neither the censored time of staying healthy nor the censoring indicator. In addition people may die due to other causes of death than the ones under consideration. That means that the risk of dying and the risk of getting ill are two dependent competing risks here.
This work does not deal with the question of finding a non-parametric estimate for the survival function S_0 on the basis of current health status data. If this were possible an approach that uses a non-parametric estimate for S_0 as well as for S_L could be called a fully semi-parametric approach.
Such an approach would be the most general one in the sense that it makes no restrictions on the shape of the univariate survival functions S_0 and S_L. With regard to the properties of the remaining parameter estimates, the estimates of the variances and correlations of the frailty variables, it would be, however, the most complicated one. Theoretical results about the asymptotic behavior of these estimates do not yet exist. Additional computational efforts for the calculation of the non-parametric estimates of the univariate survival functions are also necessary.
For practical applications it is not easy to decide whether a fully parametric or a semi-parametric approach is the most useful one. The decision depends on the aims of a study, the ascertainment of the data (truncation, censoring), and the computational resources that are available.

Chapter 3

Identifiability of a correlated gamma frailty model

In the ideal situation, where one has exact information about $(T_{10}, T_{20}, T_1, T_2)$ the identifiability of the four-dimensional correlated frailty model that is described in section 2.1 is a direct consequence of the identifiability in the bivariate correlated frailty model (see Iachine and Yashin, 1998) since all bivariate marginal distributions correspond to this model (see (2.11)-(2.14)). But this means that T_{i0} and T_i cannot be interpreted as age at onset of disease and life span, respectively, since then the problem of dependent competing risks arises. In this chapter it is shown that current health status data as described in section 2.2.1 are sufficient for the identifiability of the four-dimensional correlated frailty model. As preparation for the main proof the identifiability of the bivariate correlated frailty with censored observations is proven.

3.1 Univariate censored life span information

The following lemmata show that under independent (right) censoring the marginal univariate distributions of the survival time as well as the distribution of the censoring variable are identifiable.

Lemma 2. *Let S be the continuous survival function of a non-negative random variable V. Let V have the Lebesgue density f. Then*

$$-\int_{[0,x)} \frac{f(u)}{S(u)} du = \ln S(x)$$

for all x with $S(x) > 0$.

Proof. Let $\varphi(t) = t^{-1}$. Since $\varphi(S(u)) \leq S(x)^{-1}$ for all $u \leq x$ it follows that $E(\mathrm{I}_{[0,x)}(V)\varphi(S(V))) < \infty$ if $S(x) > 0$. It holds that

$$\int \mathrm{I}_{[0,x)}(u)\frac{f(u)}{S(u)}\,du \;=\; \int\limits_{[0,x)} \varphi(S(u))\,P^V(du) \;=\; \int\limits_{[0,1-S(x))} \varphi(1-t)\,dt \;=\; -\ln S(x)$$

by Hilfssatz 2 in Giard (1997). □

Lemma 3. *Let T denote a survival time, i.e. a non-negative random variable, and let T have the survival function S_L and the Lebesgue density f_L. The non-negative random variable A is a censoring variable. Assume that A has the Lebesgue density f_A and the survival function S_A. It is assumed that T and A are independent and that the variables*

$$X = \min(T, A) \quad \textit{and} \quad \Delta = \mathrm{I}(T \leq A)$$

are observable. Let $\bar{\mu}$ be the measure on $(\mathbb{R}^1 \times \{0,1\}, \mathfrak{B}^1 \otimes \mathcal{P}^{\{0,1\}})$ that is the product of the Lebesgue measure λ^1 with the counting measure on $\{0,1\}$, i.e. $\bar{\mu}(A \times \{i\}) = \lambda^1(A)$ for $i = 0, 1$. The density function f_X^{**} of (X, Δ) with respect to $\bar{\mu}$ is given by:

$$f_X^{**}(x,\delta) = S_L(x)\,f_A(x)\,\mathrm{I}(\delta = 0) + f_L(x)\,S_A(x)\,\mathrm{I}(\delta = 1)\,. \tag{3.1}$$

Proof. See for example Lawless (1982), p.35f. □

Lemma 4. *Let the random variables T, A, X and Δ and the functions S_L, f_L, S_A, f_A and f_X^{**} be defined as in lemma 3. Then the functions $S_L(x)$ and $S_A(x)$ can be identified for all x with $0 \leq x \leq d$ with the help of the distribution of (X, Δ), where $d = \inf\{x : S_A(x) = 0 \text{ or } S_L(x) = 0\}$. That means that if two different pairs of survival functions (S_A^1, S_L^1) and (S_A^2, S_L^2) exist that yield the same distribution for (X, Δ) then it follows $S_L^1(x) = S_L^2(x)$ and $S_A^1(x) = S_A^2(x)$ for all x with $0 \leq x \leq d$.*

Proof. Let us assume that (S_A^1, S_L^1) and (S_A^2, S_L^2) yield the same distribution for (X, Δ). Consequently the corresponding density functions are a.s. equal. From (3.1) it then follows that

$$S_L^1(x)\,f_A^1(x) = S_L^2(x)\,f_A^2(x) \qquad \text{and} \tag{3.2}$$

$$f_L^1(x)\,S_A^1(x) = f_L^2(x)\,S_A^2(x) \qquad \text{a.s.} \tag{3.3}$$

Since the distribution of (X, Δ) is the same for both pairs of survival functions this is also true for the distribution of X, therefore

$$P(X > x) = P(T > x, A > x) = S_L^1(x)\, S_A^1(x) = S_L^2(x)\, S_A^2(x) \tag{3.4}$$

for all x. Let

$$\tilde{d} = \inf\{x : S_L^1(x) = 0 \text{ or } S_L^2(x) = 0 \text{ or } S_A^1(x) = 0 \text{ or } S_A^2(x) = 0\}$$

Since the functions S_L^i, S_A^i, $i = 1, 2$, are continuous it follows that $\tilde{d} > 0$. After multiplying (3.2) with $S_A^1(x)$ and applying (3.4) one gets

$$\frac{f_A^1(x)}{S_A^1(x)} = \frac{f_A^2(x)}{S_A^2(x)}$$

a.s. in $[0, \tilde{d})$. Thus from lemma 2 it follows that

$$S_A^1(x) = S_A^2(x) = S_A(x) \quad \text{for} \quad 0 \le x < \tilde{d}. \tag{3.5}$$

(3.4) yields then that

$$S_L^1(x) = S_L^2(x) = S_L(x) \quad \text{for} \quad 0 \le x < \tilde{d}. \tag{3.6}$$

In addition, $\tilde{d} = d = \inf\{x : S_A(x) = 0 \text{ or } S_L(x) = 0\}$ follows from (3.5) and (3.6). □

3.2 Identifiability in the bivariate correlated frailty model

Let us assume that one investigates the life span of twins. Individuals may already be deceased or still alive at the time of observation which means that the individual data is right censored. It is assumed that the censoring variable A acts for both twins in a pair, i.e. the survival of the twins in a pair is investigated at the same time. If T_1 and T_2 are the life spans of the twins in a pair then the observation for a twin pair is $(X_1, X_2, \Delta_1, \Delta_2)$ where

$$X_i = \min(T_i, A) \quad \text{and} \quad \Delta_i = \mathrm{I}(T_i \le A), \quad i = 1, 2. \tag{3.7}$$

It is assumed that A is independent of the survival times T_i (independent censoring). The sample space of $(X_1, X_2, \Delta_1, \Delta_2)$ is

$$(\mathcal{X}^*, \mathfrak{A}^*) = (\mathbb{R}_+^2 \times \{0,1\} \times \{0,1\}, \mathfrak{B}^2 \otimes \mathcal{P}^{\{0,1\}} \otimes \mathcal{P}^{\{0,1\}}) \quad .$$

Let λ^1 and λ^2 denote the Lebesgue measure on $\mathbb{R}^1$ and $\mathbb{R}^2$, respectively. Let $A, B \in \mathfrak{B}^1$ be two Borel sets. Let $\tilde{\mu}$ be a measure on $(\mathcal{X}^*, \mathfrak{A}^*)$ with

$$\tilde{\mu}(A \times B \times \{i\} \times \{j\}) = \begin{cases} \lambda^1(A \cap B) & \text{if } i = 0 \text{ and } j = 0, \\ \lambda^2(A \times B) & \text{if } i = 1 \text{ or } j = 1. \end{cases} \tag{3.8}$$

One can verify that for every measurable function $g : (\mathcal{X}^*, \mathfrak{A}^*) \to (\mathbb{R}^1, \mathfrak{B}^1)$ it holds

$$\int g(x_1, x_2, i, j) d\tilde{\mu} = \begin{cases} \int g(x, x, i, j)\, d\lambda^1 & \text{if } i = 0 \text{ and } j = 0, \\ \int g(x_1, x_2, i, j)\, d\lambda^2 & \text{if } i = 1 \text{ or } j = 1. \end{cases} \tag{3.9}$$

The following lemma gives the density function of the vector $(X_1, X_2, \Delta_1, \Delta_2)$ with respect to $\tilde{\mu}$.

Lemma 5. *Let (T_1, T_2) be a random vector of two survival times with the bivariate survival function S_{L*}. Let the vector have the Lebesgue density f_{L*}. The common censoring variable A is assumed to be independent of (T_1, T_2). Censored observations of (T_1, T_2) are defined according to (3.7). If A has got the Lebesgue density f_A and the survival function S_A then the density function f_X^* of the vector $(X_1, X_2, \Delta_1, \Delta_2)$ with respect to $\tilde{\mu}$ is given by*

$$\begin{aligned} f_X^*(x_1, x_2, \delta_1, \delta_2) = {} & S_{L*}(x_1, x_2) f_A(x_1)\, \mathrm{I}(x_1 = x_2)\, \mathrm{I}(\delta_1 = 0)\, \mathrm{I}(\delta_2 = 0) \\ & - \frac{\partial S_{L*}(x_1, x_2)}{\partial x_2} f_A(x_1)\, \mathrm{I}(x_2 \le x_1)\, \mathrm{I}(\delta_1 = 0)\, \mathrm{I}(\delta_2 = 1) \\ & - \frac{\partial S_{L*}(x_1, x_2)}{\partial x_1} f_A(x_2)\, \mathrm{I}(x_1 \le x_2)\, \mathrm{I}(\delta_1 = 1)\, \mathrm{I}(\delta_2 = 0) \\ & + f_{L*}(x_1, x_2) S_A(\max(x_1, x_2))\, \mathrm{I}(\delta_1 = 1)\, \mathrm{I}(\delta_2 = 1)\ . \end{aligned} \tag{3.10}$$

Proof. f_X^* is the desired density if for all x, y and all $i, j \in \{0, 1\}$ it holds that

$$P(X_1 > x, X_2 > y, \Delta_1 = i, \Delta_2 = j) = \int \mathrm{I}_{(x,\infty)}(x_1)\, \mathrm{I}_{(y,\infty)}(x_2) f_X^*(x_1, x_2, i, j)\, d\tilde{\mu}.$$

This can be checked for the different cases using (3.9). □

The representation of the density function in (3.10) is independent of the form of the bivariate survival function S_{L*}. If S_{L*} is specified according to a bivariate correlated frailty model then censored observations are sufficient for the identifiability of the model. This is proven in the following proposition.

Proposition 3. *Let (T_1, T_2) be a random vector of two survival times and let the vector be distributed according to a bivariate correlated gamma frailty model. This means that the bivariate survival function S_{L*} of (T_1, T_2) is given by*

$$S_{L*}(x,y) = [S_L(x)\, S_L(y)]^{1-\varrho}\, [S_L^{-\sigma^2}(x) + S_L^{-\sigma^2}(y) - 1]^{-\frac{\varrho}{\sigma^2}}$$

for some constants $\sigma > 0$ and $\varrho > 0$ and a univariate survival function S_L. It is assumed that S_L has a Lebesgue density f_L.
Let A denote the common censoring variable. Assume that A is non-negative, that it has the Lebesgue density f_A and the survival function S_A. A is assumed to be independent of (T_1, T_2).
If the univariate survival function S_L of T_i, $i = 1, 2$, is strictly monotone in an interval I and in this interval it holds that $f_A(x) > 0$, then the model is identifiable with the help of censored observations $X_i = \min(T_i, A)$ and $\Delta_i = \mathrm{I}(T_i \le A)$, $i = 1, 2$. This means that the parameters of the model ϱ and σ^2 and the value of the univariate survival function S_L on $[0, d]$ are uniquely determined, where $d = \inf\{x : S_A(x) = 0 \text{ or } S_L(x) = 0\}$.

Proof. If one looks only at (X_1, Δ_1) one has univariate censored observations as in section 3.1. From lemma 4 it follows that $S_L(x)$, $S_A(x)$, $f_L(x)$ and $f_A(x)$ are identifiable for all $x \in [0, d]$.
Therefore from (3.10) it follows that the functions

$$\begin{aligned} & S_{L*}(x,x) && \text{for } 0 \le x \le d \text{ and } f_A(x) > 0, \\ \frac{\partial S_{L*}(x,y)}{\partial y} &= \frac{\partial S_{L*}(y,x)}{\partial x} && \text{for } 0 \le y \le x \le d \text{ and } f_A(x) > 0, \\ f_{L*}(x,y) &= \frac{\partial^2 S_{L*}(x,y)}{\partial x \partial y} && \text{for } \max(x,y) < d \end{aligned}$$

are identifiable with the help of the distribution of $(X_1, X_2, \Delta_1, \Delta_2)$. The exact representation of these functions can be found in (A.20), (A.23) and (A.25) in Appendix A.1. Since

$$\begin{aligned} \frac{\partial S_{L*}(x,y)}{\partial y} &= S_{L*}(x,y)\, \frac{\partial \ln S_{L*}(x,y)}{\partial y} \quad \text{and} \\ \frac{\partial S_{L*}^2(x,y)}{\partial x \partial y} &= S_{L*}(x,y) \left[\frac{\partial \ln S_{L*}(x,y)}{\partial x}\, \frac{\partial \ln S_{L*}(x,y)}{\partial y} + \frac{\partial^2 \ln S_{L*}(x,y)}{\partial x \partial y} \right], \end{aligned}$$

the functions

$$g(x) = \left. \frac{\partial \ln S_{L*}(x,y)}{\partial y} \right|_{x=y} = -\left[1 - \varrho + \frac{\varrho}{2 - S_L^{\sigma^2}(x)} \right] \frac{f_L(x)}{S_L(x)}$$

and

$$h(x) = \left.\frac{\partial^2 \ln S_{L*}(x,y)}{\partial x \partial y}\right|_{x=y} = \frac{\varrho\sigma^2}{[2 - S_L^{\sigma^2}(x)]^2} \left[\frac{f_L(x)}{S_L(x)}\right]^2$$

are identifiable for all x with $0 \leq x < d$ and $f_A(x) > 0$. Applying the uniqueness of $f_L(x)$ and $S_L(x)$ for these x yields that the functions

$$\tilde{g}(x) = \varrho \frac{1}{2 - S_L^{\sigma^2}(x)} - \varrho$$

and

$$\tilde{h}(x) = \varrho\,\sigma^2 \frac{1}{[2 - S_L^{\sigma^2}(x)]^2}$$

are identifiable for all x with $0 \leq x < d$, $f_A(x) > 0$ and $f_L(x) > 0$. Some transformations yield

$$\tilde{h}(x) = \frac{\sigma^2}{\varrho}\,\tilde{g}(x)^2 + 2\,\sigma^2\,\tilde{g}(x) + \varrho\,\sigma^2 \ .$$

If one uses this equation with 3 different x_1, x_2, x_3 one gets a system of linear equations with uniquely determined coefficients for $\frac{\sigma^2}{\varrho}$, $2\,\sigma^2$ and $\varrho\,\sigma^2$. The value of the determinant of the matrix of coefficients of this system is

$$\big(\tilde{g}(x_1) - \tilde{g}(x_2)\big)\big(\tilde{g}(x_1) - \tilde{g}(x_3)\big)\big(\tilde{g}(x_2) - \tilde{g}(x_3)\big) \ .$$

It holds $\tilde{g}(x_1) = \tilde{g}(x_2)$ if and only if $S_L(x_1) = S_L(x_2)$. Without loss of generality one can assume that the interval $I \subset [0, d]$. Since $S_L(x)$ is strictly monotone on the interval I, in this interval it holds that $f_A(x) > 0$ and $f_L(x) > 0$. Thus 3 different x_1, x_2, x_3 exist for which the functions $\tilde{g}$ and $\tilde{h}$ are identified and the value of the determinant is unequal to zero. Then the system of linear equation has a unique solution. That means that σ and ϱ are identifiable. □

Corollary 2. *If the univariate survival function has the representation*

$$S_L(x) = [1 + \sigma^2 H(x)]^{-1/\sigma^2} \ .$$

with the underlying cumulative hazard rate H, then from the identifiability of S_L and σ follows the identifiability of H.

3.3 Identifiability in the four-dimensional correlated frailty model

In this section it is proven that current health status data as described in section 2.2.1 are sufficient for the identifiability of the four-dimensional correlated gamma frailty model which was defined in section 2.1.4.
The following auxiliary lemma is needed in the main proof. It gives the density function of (X_1, C_1), which is the current health status information for one individual.

Lemma 6. *Let (T_{10}, T_1) be a vector of two survival times. Let S_T denote the bivariate survival function of (T_{10}, T_1) and let the vector have a Lebesgue density. Assume that T_1 has a Lebesgue density f_L and let S_L denote its survival function.*
The random variable A denotes the age at the time of observation. It is assumed to be independent of (T_{10}, T_1). Assume that A has the Lebesgue density f_A and the survival function S_A.
Current health status data is defined as

$$X = \min(T_1, A) \quad \textit{and} \quad C = \begin{cases} 0 & \textit{if } T_{10} > A \textit{ and } T_1 > A, \\ 1 & \textit{if } T_{10} \leq A \textit{ and } T_1 > A, \\ 2 & \textit{if } T_1 \leq A. \end{cases}$$

If λ^1 is the Lebesgue measure on $\mathbb{R}^1$ and $A \in \mathfrak{B}^1$ is a Borel set then let $\hat{\mu}$ be the measure on $(\mathbb{R}_+ \times \{0,1,2\}, \mathfrak{B}^1 \otimes \mathcal{P}^{\{0,1,2\}})$ with $\hat{\mu}(A \times \{i\}) = \lambda^1(A)$. The density function f_T of (X, C) with respect to $\hat{\mu}$ is given by

$$\begin{aligned} f_T(x,c) = {} & S_T(x,x) f_A(x)\, \mathrm{I}(c=0) \\ & + \big[S_L(x) - S_T(x,x)\big] f_A(x)\, \mathrm{I}(c=1) + S_A(x) f_L(x)\, \mathrm{I}(c=2)\,. \end{aligned} \tag{3.11}$$

Proof. f_T from (3.11) is the desired density function if for all x and $i = 0, 1, 2$ it holds that

$$P(X > x, C = i) = \int \mathrm{I}_{(x,\infty)}(u) f_T(u,i)\, d\hat{\mu}.$$

This can be checked for the different cases using

$$\int g(x,i) d\hat{\mu} = \int g(x,i)\, d\lambda^1 \tag{3.12}$$

for every measurable function g. □

The following theorem gives the main identifiability result for the four-dimensional correlated gamma frailty model.

Theorem 3. *Let $(T_{10}, T_1, T_{20}, T_2)$ be a random vector of 4 survival times and let the vector be distributed according to a four-dimensional correlated gamma frailty model (see definition 9). Let S denote the survival function of $(T_{10}, T_1, T_{20}, T_2)$, S_0 is the univariate survival function of T_{i0}, $i = 1, 2$, and S_L is the univariate survival function of T_i, $i = 1, 2$.*
The random variable A denotes the age at the time of observation. It is assumed to be independent of $(T_{10}, T_1, T_{20}, T_2)$ and to be non-negative. Assume that A has the Lebesgue density f_A and the survival function S_A. Let furthermore $d = \inf\{x : S_A(x) = 0 \text{ or } S_L(x) = 0\}$.
If an interval $I \subset [0, d]$ exists such that

(A) the univariate survival function S_L is strictly monotone and twice differentiable on I and $f_A(x) > 0$ and $0 < S_0(x) < 1$ for all $x \in I$, and

(B) if for all c with $0 < c < 1$ there is at least one $x \in I$ such that

$$S_L^{-\sigma^2}(x) \neq \frac{S_0^{-\sigma_0^2}(x) - 1}{S_0^{-c\,\sigma_0^2}(x) - 1} \tag{3.13}$$

then the model is identifiable with the help of current health status data

$$X_i = \min(T_i, A) \quad \text{and} \quad C_i = \begin{cases} 0 & \text{if } T_{i0} > A \text{ and } T_i > A, \\ 1 & \text{if } T_{i0} \leq A \text{ and } T_i > A, \\ 2 & \text{if } T_i \leq A, \end{cases} \quad i = 1, 2.$$

This means that the parameters of the model ϱ_0, ϱ_1, ϱ, ϱ^*, σ_0 and σ and the univariate survival functions S_0 and S_L are uniquely determined on $[0, d]$. The values $S_0(x)$ can only be identified for those $x \in [0, d]$ with $f_A(x) > 0$.

Proof. The proof is constructed stepwise. If the identifiability of one parameter was shown this property is used in order to show the identifiability of the next parameter.

Step 1 - Identifiability of σ, ϱ_1, S_L and f_A:
If one uses only the censored life span X_i and

$$\Delta_i = \begin{cases} 0 & \text{if } C_i = 0 \text{ or } C_i = 1, \\ 1 & \text{if } C_i = 2, \end{cases} \quad i = 1, 2\ ,$$

for every twin, one gets the case of a bivariate correlated gamma frailty model with censored observations. From proposition 3 it follows that the parameters σ and ϱ_1 and the values of the univariate survival function S_L on $[0,d]$ are identifiable with the help of this part of the whole information. Furthermore from lemma 4 it follows that f_A is uniquely determined for all $x \in [0,d]$.

Step 2 - Identifiability of $\frac{\varrho^*}{\sigma_0}$ and $S_0^{-\sigma_0^2}$:

The knowledge of the distribution of (X_1, X_2, C_1, C_2) implies the identifiability of the density function f_X from (2.20). The explicit representation of the functions that are included in f_X is given in Appendix A.1 in equations (A.9)-(A.26). From (A.9), (A.10), (A.13), (A.18), (A.19) and (A.24) it follows that the functions

$$S(x,x,x,x) \quad \text{for } 0 \le x \le d, \tag{3.14}$$

$$S(0,x,x,x) \quad \text{for } 0 \le x \le d, \quad \text{and} \tag{3.15}$$

$$\left.\frac{\partial S(s_1,s_2,t_1,t_2)}{\partial t_1}\right|_{\substack{s_1=0,\\ s_2=t_2=x}} \quad \text{for } 0 \le t_1 \le x \le d, \tag{3.16}$$

are uniquely determined for all x with $f_A(x) > 0$. The uniqueness of (3.14) follows from that of the marginal density of the pairs where both twins are healthy, that of (3.15) from the identifiability of the marginal density of the pairs where one twin is healthy and one twin is ill, and finally the uniqueness of (3.16) from that of the marginal density of the pairs where one twin is healthy and the other is deceased.
Since for every twice differentiable positive function f

$$\frac{df(x)}{dx} = f(x)\frac{d\,\ln f(x)}{dx} \quad \text{and} \quad \frac{d^2 f(x)}{dx^{\,2}} = f(x)\left(\left[\frac{d\ln f(x)}{dx}\right]^2 + \frac{d^2 \ln f(x)}{dx^{\,2}}\right),$$

and applying (3.14), (3.15) and (3.16) one can conclude that the functions

$$g_1(x) = \left.\frac{\partial}{\partial t_1}\ln S(s_1,s_2,t_1,t_2)\right|_{\substack{s_1=0,\\ s_2=t_1=t_2=x}} \quad \text{and}$$

$$g_2(x) = \left.\frac{\partial^2}{\partial t_1^2}\ln S(s_1,s_2,t_1,t_2)\right|_{\substack{s_1=0,\\ s_2=t_1=t_2=x}}$$

are identifiable and defined for all $x \in I$ from assumption (A). Without loss of generality it is assumed that $0 < S_L(x) < 1$ for $x \in I$. $S_L(x)$ and therefore also the derivatives of this function for $x \in I$, σ and ϱ_1 have already been identified in step 1. Therefore one can conclude from the explicit

representation of g_1 and g_2 that the functions

$$\begin{aligned} h_1(x) &= \frac{S_L^{1+\sigma^2}(x)}{S_L'(x)}\, g_1(x) - (1-\varrho_1) S_L^{\sigma^2}(x) - \varrho_1 \big[2\, S_L^{-\sigma^2}(x) - 1\big]^{-1} \\ &= \varrho^* \frac{\sigma}{\sigma_0} \left\{ \big[S_0^{-\sigma_0^2}(x) + 2\, S_L^{-\sigma^2}(x) - 2\big]^{-1} - \big[2\, S_L^{-\sigma^2}(x) - 1\big]^{-1} \right\} \end{aligned}$$

and

$$\begin{aligned} h_2(x) &= \frac{g_2(x) - g_1(x) \left[-(1+\sigma^2) S_L^{-1}(x) S_L'(x) + \dfrac{S_L''(x)}{S_L'(x)} \right]}{\sigma^2\, S_L^{-2(1+\sigma^2)}(x)\, {S_L'}^2(x)} \\ &\quad - (1-\varrho_1) S_L^{2\sigma^2}(x) - \varrho_1 \big[2\, S_L^{-\sigma^2}(x) - 1\big]^{-2} \\ &= \varrho^* \frac{\sigma}{\sigma_0} \left\{ \big[S_0^{-\sigma_0^2}(x) + 2\, S_L^{-\sigma^2}(x) - 2\big]^{-2} - \big[2\, S_L^{-\sigma^2}(x) - 1\big]^{-2} \right\} \end{aligned}$$

are identifiable for all $x \in I$. This leads to the identifiability of $S_0^{-\sigma_0^2}(x)$ for all $x \in I$ and that of $\frac{\varrho^*}{\sigma_0}$ since

$$S_0^{-\sigma_0^2}(x) = \frac{h_1(x)}{h_2(x) - h_1(x)\big[2\, S_L^{-\sigma^2}(x) - 1\big]^{-1}} - 2\, S_L^{-\sigma^2}(x) + 2$$

and

$$\frac{\varrho^*}{\sigma_0} = \frac{1}{\sigma} \frac{h_1^2(x)}{h_2(x) - h_1(x)\big[2\, S_L^{-\sigma^2}(x) - 1\big]^{-1}}\,.$$

Step 3 - Identifiability of $\frac{\varrho_0}{\sigma_0^2}$:

From lemma 6 it follows that using information from only one person yields the identifiability of $S_T(x,x)$ from (2.11) for all x with $f_A(x) > 0$. Taking into account the results just derived and the identifiability of $S(x,x,x,x)$ one sees that the functions

$$\begin{aligned} h_3(x) &= \ln S_T(x,x) - \ln S_L(x) \\ &= -\frac{1}{\sigma_0^2} \ln S_0^{-\sigma_0^2}(x) + \frac{\varrho}{\sigma_0 \sigma} \ln \left[\frac{S_0^{-\sigma_0^2}(x)\, S_L^{-\sigma^2}(x)}{S_0^{-\sigma_0^2}(x) + S_L^{-\sigma^2}(x) - 1} \right] \end{aligned} \tag{3.17}$$

and

$$\begin{aligned} h_4(x) &= \ln S(x,x,x,x) - 2h_3(x) + 2\,\frac{1-\varrho_1}{\sigma^2}\ln S_L^{-\sigma^2}(x) + \frac{\varrho_1}{\sigma^2}\ln\big[2S_L^{-\sigma^2}(x)-1\big] \\ &\quad + \frac{\varrho^*}{\sigma_0\sigma}\ln\left[\frac{S_0^{-2\sigma_0^2}(x)\,S_L^{-2\sigma^2}(x)\big[2S_0^{-\sigma_0^2}(x)+2S_L^{-\sigma^2}(x)-3\big]}{\big[S_0^{-\sigma_0^2}(x)+S_L^{-\sigma^2}(x)-1\big]\big[2S_0^{-\sigma_0^2}(x)-1\big]\big[2S_L^{-\sigma^2}(x)-1\big]}\right] \\ &= \frac{\varrho_0}{\sigma_0^2}\ln\left[\frac{S_0^{-2\sigma_0^2}(x)}{2S_0^{-\sigma_0^2}(x)-1}\right] \end{aligned}$$

are identifiable and finite for $x \in I$. Since the coefficient for $\frac{\varrho_0}{\sigma_0^2}$ is unequal to zero on I this term is uniquely determined.

Step 4 - Identifiability of σ_0^2 and ϱ:
(3.17) yields a system of linear equations with known coefficients for $\frac{1}{\sigma_0^2}$ and $\frac{\varrho}{\sigma_0\sigma}$ if 2 different $x_1, x_2 \in I$ are inserted in the formula. The value $D(x_1,x_2)$ of the determinant of the matrix of coefficients for this system is

$$D(x_1,x_2) = \ln S_0^{-\sigma_0^2}(x_1)\,\ln S_0^{-\sigma_0^2}(x_2)\big(\tilde{H}(x_2)-\tilde{H}(x_1)\big)$$

where

$$\tilde{H}(x) = \frac{\ln\big[S_L^{-\sigma^2}(x)\big] - \ln\big[S_0^{-\sigma_0^2}(x)+S_L^{-\sigma^2}(x)-1\big]}{\ln S_0^{-\sigma_0^2}(x)}\,.$$

Therefore $D(x_1,x_2)$ can only be zero for all $x_1, x_2 \in I$ if the function $\tilde{H}$ is constant on I. Since $-1 < \tilde{H}(x) < 0$ from assumption (B) it follows that $\tilde{H}(x)$ is not constant on I and thus $D(x_1,x_2) \neq 0$ for at least one pair of distinct numbers in I. Thus $\frac{1}{\sigma_0^2}$ and $\frac{\varrho}{\sigma_0\sigma}$ and therefore also σ_0 and ϱ are uniquely determined.

Step 5 - Identifiability of S_0, ϱ^* and ϱ_0:
Taking into account the results of step 2 and 3 the identifiability of ϱ^* and ϱ_0 is now obvious.
The function

$$S_T(x,x) = \big[S_0(x)\big]^{1-\frac{\sigma_0}{\sigma}\varrho}\big[S_L(x)\big]^{1-\frac{\sigma}{\sigma_0}\varrho}\big[S_0^{-\sigma_0^2}(x)+S_L^{-\sigma^2}(x)-1\big]^{-\frac{\varrho}{\sigma_0\sigma}} \qquad (3.18)$$

is known for all $x \in [0,d]$ with $f_A(x) > 0$ and $S_L(x)$, σ_0, σ and ϱ are identifiable. $S_T(x,x)$ is strictly monotone increasing in $S_0(x)$. Therefore, for every $x \in [0,d]$ with $f_A(x) > 0$ the value $S_0(x)$ is uniquely determined by (3.18). □

Remark. If an interval $I = (0, K)$ with some constant $K > 0$ exists such that assumption (A) of theorem 3 is fulfilled for all $x \in I$, then assumption (B) of theorem 3 is automatically fulfilled. Assume that a constant c with $0 < c < 1$ exists such that for all $x \in I$ in (3.13) the equal sign holds. Then taking the limit $x \to 0$ on both sides of (3.13) yields $c = 1$, which is a contradiction.

Remark. If for all intervals $I \subseteq [0, d]$ for which assumption (A) of theorem 3 is fulfilled, assumption (B) of theorem 3 is not fulfilled, then only $S_L(x)$, $S_0^{-\sigma_0^2}(x)$, σ, ϱ_1, c, $\dfrac{\varrho_0}{\sigma_0^2}$, $\dfrac{\varrho^*}{\sigma_0}$ and $\gamma = -\dfrac{1}{\sigma_0^2} + (1-c)\dfrac{\varrho}{\sigma_0 \sigma}$ are identifiable with the help of current health status data.

Chapter 4

Asymptotic properties of maximum likelihood estimates in a correlated gamma frailty model

4.1 Consistency in the four-dimensional model when current health status data is used

In this section a theorem is proven that gives sufficient conditions for the consistency of maximum likelihood estimates in the four-dimensional correlated gamma frailty model in the case where the estimation is based on current health status data. A completely parametric approach is assumed, i.e. that one chooses parametric representations for the univariate survival functions S_0 and S_L and that one estimates these parameters together with the variances and correlations of the frailty variables. Consistency is not shown for the variances and correlations of the frailty variables directly, but rather for the shape parameters of the components of the frailty variables. A one-to-one mapping between these parameters and the variances and correlations exists (see (2.6) and (A.1)).
The first subsection summarizes all technical assumptions used in the proof of the main theorem. The second subsection defines the parameter set for which consistency is proven in detail. Finally, the third subsection contains the main theorem about consistency of maximum likelihood estimates in the four-dimensional correlated gamma frailty model.

4.1.1 Assumptions

First the assumptions on the parameterization of the univariate survival functions are given.

Assumption 2. Let $\Lambda = \{\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_{K_u})' : \boldsymbol{\lambda} \in \Lambda\}$ be a compact subset of $\mathbb{R}^{K_u}$. Let $\mathfrak{H} = \{H(.|\boldsymbol{\lambda}) : \boldsymbol{\lambda} \in \Lambda\}$ be a family of cumulative hazard functions such that for all $\boldsymbol{\lambda} \in \Lambda$:

(a) $H(.|\boldsymbol{\lambda}) : \mathbb{R}_+ \to \mathbb{R}_+$,

(b) $H(0|\boldsymbol{\lambda}) = 0$,

(c) $\lim_{x\to\infty} H(x|\boldsymbol{\lambda}) = \infty$,

(d) $H(x|\boldsymbol{\lambda})$ is a monotone increasing, continuous function in x, and,

(e) $H(x|\boldsymbol{\lambda})$ is a continuous function of $\boldsymbol{\lambda}$ for all $x \in \mathbb{R}_+$.

Assumption 3. Let $\mathfrak{H} = \{H(.|\boldsymbol{\lambda}) : \boldsymbol{\lambda} \in \Lambda\}$ be a family of cumulative hazard functions that fulfills assumption 2. It is assumed that

(a) $H(x|\boldsymbol{\lambda})$ is differentiable in x with $\mu(x|\boldsymbol{\lambda}) = \frac{\partial}{\partial x} H(x|\boldsymbol{\lambda})$ for all $\boldsymbol{\lambda} \in \Lambda$, and that

(b) $\mu(x|\boldsymbol{\lambda})$ is a continuous function of $\boldsymbol{\lambda}$ for all $x \in \mathbb{R}_+$.

Assumption 4. Let $\mathfrak{H}_0 = \{H_0(.|\boldsymbol{\lambda}) : \boldsymbol{\lambda} \in \Lambda\}$ and $\mathfrak{H} = \{H(.|\boldsymbol{\lambda}) : \boldsymbol{\lambda} \in \Lambda\}$ be two families of cumulative hazard functions that fulfill assumption 2. In addition it is assumed that $\mathfrak{H}$ fulfills assumption 3. Let m_s, M_s be two constants with $0 < m_s < M_s < \infty$. Then two families of univariate survival functions can be defined by

$$\mathfrak{S}_0 = \{S_0(x|\sigma_0, \boldsymbol{\lambda}) = [1 + \sigma_0^2 H_0(x|\boldsymbol{\lambda})]^{-1/\sigma_0^2} : H_0 \in \mathfrak{H}_0, m_s \leq \sigma_0 \leq M_s\} \text{ and}$$
$$\mathfrak{S} = \{S_L(x|\sigma, \boldsymbol{\lambda}) = [1 + \sigma^2 H(x|\boldsymbol{\lambda})]^{-1/\sigma^2} : H \in \mathfrak{H}, m_s \leq \sigma \leq M_s\}.$$

It is assumed that a constant $C_0 > 0$ exists such that for all $\boldsymbol{\lambda}_0 \in \Lambda$, all σ_0, σ with $m_s \leq \sigma_0, \sigma \leq M_s$ and all $x \geq 0$ it holds that:

(a) $\dfrac{\mu(x|\boldsymbol{\lambda}_0)}{1 + \sigma^2 H(x|\boldsymbol{\lambda}_0)} \leq C_0$,

(b) $\sup_{\boldsymbol{\lambda}\in\Lambda} |\mu(x|\boldsymbol{\lambda})| \cdot S_L(x|\sigma, \boldsymbol{\lambda}_0) \leq C_0$,

(c) $\sup_{\boldsymbol{\lambda}\in\Lambda} \ln[1 + M_s^2 H_0(x|\boldsymbol{\lambda})] \cdot S_0(x|\sigma_0, \boldsymbol{\lambda}_0) \leq C_0$ and
$\sup_{\boldsymbol{\lambda}\in\Lambda} \ln[1 + M_s^2 H_0(x|\boldsymbol{\lambda})] \cdot S_L(x|\sigma, \boldsymbol{\lambda}_0) \leq C_0$,

(d) $\sup_{\boldsymbol{\lambda}\in\Lambda} \ln[1 + M_s^2 H(x|\boldsymbol{\lambda})] \cdot S_L(x|\sigma, \boldsymbol{\lambda}_0) \leq C_0$, and,

(e) $\sup_{\boldsymbol{\lambda}\in\Lambda} |\ln H_0(x|\boldsymbol{\lambda})| \cdot (1 - S_0(x|\sigma_0, \boldsymbol{\lambda}_0)\, S_0(x|\sigma_0, \boldsymbol{\lambda}_0) \leq C_0$ and
$\sup_{\boldsymbol{\lambda}\in\Lambda} |\ln H_0(x|\boldsymbol{\lambda})| \cdot (1 - S_0(x|\sigma_0, \boldsymbol{\lambda}_0)\, S_L(x|\sigma, \boldsymbol{\lambda}_0) \leq C_0$.

The following assumption ensures the identifiability of the parameters of the univariate survival functions and that of the parameters of the distribution of the frailty variables.

Assumption 5. Let $\mathfrak{H}_0 = \{H_0(.|\boldsymbol{\lambda}) : \boldsymbol{\lambda} \in \Lambda\}$ and $\mathfrak{H} = \{H(.|\boldsymbol{\lambda}) : \boldsymbol{\lambda} \in \Lambda\}$ be two families of cumulative hazard functions that fulfill assumption 2. Let $\mathfrak{H}$ fulfill assumption 3. Assume that

$$H_0(x|\boldsymbol{\lambda}_1) = H_0(x|\boldsymbol{\lambda}_2) \quad \text{and} \quad H(x|\boldsymbol{\lambda}_1) = H(x|\boldsymbol{\lambda}_2) \tag{4.1}$$

for all x in an arbitrary, non-degenerated interval $\tilde{I}$ and for $\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2 \in \Lambda$ implies $\boldsymbol{\lambda}_1 = \boldsymbol{\lambda}_2$.
Let m_s, M_s be two constants with $0 < m_s < M_s < \infty$. Assume further that an interval I exists such that for all $\boldsymbol{\lambda} \in \Lambda$ it holds that

(a) $\mu(x|\boldsymbol{\lambda}) > 0$, $0 < H_0(x|\boldsymbol{\lambda}) < \infty$, $\mu(x|\boldsymbol{\lambda})$ is differentiable in x for all $x \in I$, and

(b) for all c with $0 < c < 1$ there is at least one $x \in I$ such that

$$1 + \sigma^2 H(x|\boldsymbol{\lambda}) \neq \frac{\sigma_0^2 H_0(x|\boldsymbol{\lambda})}{[1 + \sigma_0^2 H_0(x|\boldsymbol{\lambda})]^c - 1}$$

for all σ_0, σ with $m_s \leq \sigma_0, \sigma \leq M_s$.

Now the technical assumptions on the distribution of the random variable A, the age at the time of observation, are given.

Assumption 6. It is assumed that A is a non-negative random variable which has the Lebesgue density f_A and the univariate survival function S_A. For these functions it holds that

(a) $EA = \int x f_A(x)\, dx < \infty$, $EA^2 = \int x^2 f_A(x)\, dx < \infty$,

(b) $\int |\ln f_A(x)|\, f_A(x)\, dx < \infty$,

(c) $\int x\, |\ln f_A(x)|\, f_A(x)\, dx < \infty$ and

(d) $\iint |\ln S_A(\max(x_1, x_2))|\, S_A(\max(x_1, x_2))\, dx_1 dx_2 < \infty$.

Remark. The convention $g(x)\, \mathrm{I}(x \in B) = 0$ for every (possibly infinite) function $g(x)$, every set B and every $x \notin B$ holds for all considerations in this chapter.

4.1.2 Definition of the parameter space

The identifiability of the four-dimensional correlated frailty model was proven when S_0 and S_L are arbitrary univariate survival functions and if the parameter vector $(\sigma_0, \sigma, \varrho_0, \varrho_1, \varrho, \varrho^*)$ belongs to a parameter set Θ_r that is defined as

$$\Theta_r = \{(\sigma_0, \sigma, \varrho_0, \varrho_1, \varrho, \varrho^*) : 0 \leq \varrho_0, \varrho_1, \varrho, \varrho^* \leq 1, \quad 0 < \sigma_0, \sigma \text{ and } (2.15)\}\ .$$

Let m_k and M_k be two arbitrary constants with $0 < m_k < M_k < \infty$ and define

$$\Theta_k = \{(k_0, k_1, k_3, k_4, k_{10}, k_{11}) : m_k \leq k_0 \leq M_k, 0 \leq k_1, k_3, k_4, k_{10}, k_{11} \leq M_k\}. \tag{4.2}$$

Consistency will be proven for every parametric specification of S_0 and S_L that fulfills the assumptions 2, 3, 4 and 5. The parameter set $\Theta = \Theta_k \times \Lambda$ and the parameter vector $\boldsymbol{\theta} = (k_0, k_1, k_3, k_4, k_{10}, k_{11}, \boldsymbol{\lambda}')'$ are used. Here k_0, k_1, k_3, k_4, k_{10} and k_{11} are the shape parameters of the gamma distributed components of the frailty variables (see (2.5)). Since there is a one-to-one mapping between the parameter vectors $(\sigma_0, \sigma, \varrho_0, \varrho_1, \varrho, \varrho^*)$ and $(k_0, k_1, k_3, k_4, k_{10}, k_{11})$ (see (2.6) and (A.1)) it makes no difference which of them is used in the proof of consistency. The MLE for the vector $(\sigma_0, \sigma, \varrho_0, \varrho_1, \varrho, \varrho^*)$ is consistent if and only if the MLE for the vector $(k_0, k_1, k_3, k_4, k_{10}, k_{11})$ is consistent.

Lemma 7. *Set* $r = (1 + 3\frac{M_k}{m_k})^{-1}$ *and let* $(k_0, k_1, k_3, k_4, k_{10}, k_{11}) \in \Theta_k$ *be arbitrary. For the corresponding vector* $(\sigma_0, \sigma, \varrho_0, \varrho_1, \varrho, \varrho^*)$ *that is calculated using* (2.6) *it follows that* $(\sigma_0, \sigma, \varrho_0, \varrho_1, \varrho, \varrho^*) \in \Theta_r$.

Proof. From (2.6) $0 \leq \varrho_0, \varrho_1, \varrho, \varrho^* \leq 1$ and $\sigma_0, \sigma > 0$ obviously follow. Using (2.6) one can calculate

$$\begin{aligned}
\varrho - \varrho^* &= \sigma_0 \sigma k_1 \geq 0 \\
\varrho_0 \sigma - \varrho^* \sigma_0 &= \sigma_0^2 \sigma k_3 \geq 0 \\
\varrho_1 \sigma_0 - \varrho^* \sigma &= \sigma_0 \sigma^2 k_4 \geq 0 \\
(1 - \varrho_0)\sigma - (\varrho - \varrho^*)\sigma_0 &= \sigma_0^2 \sigma k_{10} \geq 0 \\
(1 - \varrho_1)\sigma_0 - (\varrho - \varrho^*)\sigma &= \sigma_0 \sigma^2 k_{11} \geq 0 \ .
\end{aligned}$$

To prove (2.15) it remains to be shown that $\varrho^* \geq r$. It holds that

$$\begin{aligned}
\varrho^* &= \frac{k_0}{\sqrt{k_0 + k_1 + k_3 + k_{10}}\sqrt{k_0 + k_1 + k_4 + k_{11}}} \\
&= \left[1 + \frac{k_1 + k_3 + k_{10}}{k_0}\right]^{-1/2} \left[1 + \frac{k_1 + k_4 + k_{11}}{k_0}\right]^{-1/2} \geq (1 + 3\tfrac{M_k}{m_k})^{-1} = r \ .
\end{aligned}$$

□

Remark. The mapping of Θ_k into the variances and correlations leads to a subset of Θ_r where the variances are bounded since $\frac{1}{2} M_k^{-1/2} \leq \sigma_0, \sigma \leq m_k^{-1/2}$ if $(k_0, k_1, k_3, k_4, k_{10}, k_{11}) \in \Theta_k$.

4.1.3 Proof of consistency

Theorem 4. *Let m_k and M_k be two constants with $0 < m_k < M_k < \infty$. Let $\Theta = \Theta_k \times \Lambda$ where Θ_k is given by (4.2) and $\Lambda = \{\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_{K_u})' : \boldsymbol{\lambda} \in \Lambda\}$ is a compact subset of $\mathbb{R}^{K_u}$. Set $m_s = \frac{1}{2} M_k^{-1/2}$, $M_s = m_k^{-1/2}$,*

$$\sigma_0 = (k_0 + k_1 + k_3 + k_{10})^{-1/2} \quad \textit{and} \quad \sigma = (k_0 + k_1 + k_4 + k_{11})^{-1/2} \tag{4.3}$$

for $(k_0, k_1, k_3, k_4, k_{10}, k_{11}) \in \Theta_k$. $\mathfrak{H}_0 = \{H_0(.|\boldsymbol{\lambda}) : \boldsymbol{\lambda} \in \Lambda\}$ and $\mathfrak{H} = \{H(.|\boldsymbol{\lambda}) : \boldsymbol{\lambda} \in \Lambda\}$ are two families of cumulative hazard functions that fulfill assumption 2. In addition it is assumed that $\mathfrak{H}$ fulfills assumption 3. Let

$$\begin{aligned}
\mathfrak{S}_0 &= \{S_0(x|\sigma_0, \boldsymbol{\lambda}) = [1 + \sigma_0^2 H_0(x|\boldsymbol{\lambda})]^{-1/\sigma_0^2} : H_0 \in \mathfrak{H}_0, m_s \leq \sigma_0 \leq M_s\} \textit{ and} \\
\mathfrak{S} &= \{S_L(x|\sigma, \boldsymbol{\lambda}) = [1 + \sigma^2 H(x|\boldsymbol{\lambda})]^{-1/\sigma^2} : H \in \mathfrak{H}, m_s \leq \sigma \leq M_s\} \textit{ be}
\end{aligned}$$

two families of univariate survival functions that fulfill assumptions 4 and 5. Let $P_{\boldsymbol{\theta}}$ denote the probability distribution on $(\mathbb{R}_+^4, \mathfrak{B}_+^4)$ that corresponds to the four-dimensional survival function:

$$S(s_1, s_2, t_1, t_2|\boldsymbol{\theta}) = [S_0(s_1|\sigma_0, \boldsymbol{\lambda}) S_0(s_2|\sigma_0, \boldsymbol{\lambda})]^{\sigma_0^2 k_{10}} [S_L(t_1|\sigma, \boldsymbol{\lambda}) S_L(t_2|\sigma, \boldsymbol{\lambda})]^{\sigma^2 k_{11}}$$

$$\cdot \left[S_0^{-\sigma_0^2}(s_1|\sigma_0,\boldsymbol{\lambda})+S_0^{-\sigma_0^2}(s_2|\sigma_0,\boldsymbol{\lambda})-1\right]^{-k_3}\left[S_L^{-\sigma^2}(t_1|\sigma,\boldsymbol{\lambda})+S_L^{-\sigma^2}(t_2|\sigma,\boldsymbol{\lambda})-1\right]^{-k_4}$$
$$\cdot \left[\left(S_0^{-\sigma_0^2}(s_1|\sigma_0,\boldsymbol{\lambda})+S_L^{-\sigma^2}(t_1|\sigma,\boldsymbol{\lambda})-1\right)\left(S_0^{-\sigma_0^2}(s_2|\sigma_0,\boldsymbol{\lambda})+S_L^{-\sigma^2}(t_2|\sigma,\boldsymbol{\lambda})-1\right)\right]^{-k_1}$$
$$\cdot \left[S_0^{-\sigma_0^2}(s_1|\sigma_0,\boldsymbol{\lambda})+S_0^{-\sigma_0^2}(s_2|\sigma_0,\boldsymbol{\lambda})+S_L^{-\sigma^2}(t_1|\sigma,\boldsymbol{\lambda})+S_L^{-\sigma^2}(t_2|\sigma,\boldsymbol{\lambda})-3\right]^{-k_0}$$

and let $\mathfrak{P} = \{P_{\boldsymbol{\theta}} : \boldsymbol{\theta} \in \Theta\}$. *If* $(T_{10}, T_{20}, T_1, T_2) \sim P_{\boldsymbol{\theta}}$ *then the vector* (X_1, X_2, C_1, C_2) *is defined by*

$$X_i = \min(T_i, A) \quad \textit{and} \quad C_i = \begin{cases} 0 & \textit{if } T_{i0} > A \textit{ and } T_i > A, \\ 1 & \textit{if } T_{i0} \leq A \textit{ and } T_i > A, \\ 2 & \textit{if } T_i \leq A, \end{cases} \quad i = 1, 2.$$

where A *is a non-negative random variable that has the Lebesgue density* f_A *and the univariate survival function* S_A. *It is assumed that* A *is independent of* $(T_{10}, T_{20}, T_1, T_2)$ *and that it fulfills assumption 6. It is assumed further that the interval* I *from assumption 5 can be chosen in such a way that* $f_A(x) > 0$ *for all* $x \in I$. *If* $(T_{10}, T_{20}, T_1, T_2) \sim P_{\boldsymbol{\theta}}$ *let* $Q_{\boldsymbol{\theta}}$ *denote the distribution of* (X_1, X_2, C_1, C_2) *and let* $\mathfrak{Q} = \{Q_{\boldsymbol{\theta}} : \boldsymbol{\theta} \in \Theta\}$.
A σ*-finite measure* μ *on the sample space of* (X_1, X_2, C_1, C_2) *exists that dominates the family* $\mathfrak{Q}$. *Let* $f_X(.|\boldsymbol{\theta})$ *denote the density of* $Q_{\boldsymbol{\theta}}$ *with respect to* μ *and let* $(x_1^i, x_2^i, c_1^i, c_2^i)$ *denote the* i^{th} *observation. Then every sequence of measurable maximum likelihood estimates* $(\hat{\boldsymbol{\theta}}_n)$ *of* $\boldsymbol{\theta}$ *that is derived by maximizing the likelihood function*

$$L(\mathbf{x}_1, \mathbf{x}_2, \mathbf{c}_1, \mathbf{c}_2|\boldsymbol{\theta}) = \prod_{i=1}^{n} f_X(x_1^i, x_2^i, c_1^i, c_2^i|\boldsymbol{\theta})$$

is consistent. This means that if $\boldsymbol{\theta}_0$ *denotes the true parameter and* $(\hat{\boldsymbol{\theta}}_n)$ *the sequence of maximum likelihood estimates then* $\hat{\boldsymbol{\theta}}_n \xrightarrow{P} \boldsymbol{\theta}_0$.

Proof. The proof is based on theorem 1. The measure μ can be chosen as described in definition 10. The density function f_X of (X_1, X_2, C_1, C_2) with respect to μ is given by (2.20). The explicit representation of the functions that are included in the density can be found in Appendix A.1. Now the assumptions of theorem 1 are verified.

Assumption (A1)
is a direct consequence of the definition of Θ, Θ_k and Λ.

Assumption (B1)

follows from (4.3), from the representation of the functions that are included in the density f_X (see formulas (A.9) - (A.26)) and from assumptions 2(e) and 3(b).

Assumption (C1):

From assumptions 5(a) and 5(b) and the assumptions on f_A one can conclude that the assumptions (A) and (B) of theorem 3 are fulfilled for all $\boldsymbol{\lambda} \in \Lambda$ on an interval I. Using lemma 7 and theorem 3 thus yields that from

$$f_X(x_1, x_2, c_1, c_2|\sigma_0^1, \sigma^1, \varrho_0^1, \varrho_1^1, \varrho^1, \varrho^{*1}, \boldsymbol{\lambda}_1) = f_X(x_1, x_2, c_1, c_2|\sigma_0^2, \sigma^2, \varrho_0^2, \varrho_1^2, \varrho^2, \varrho^{*2}, \boldsymbol{\lambda}_2) \quad \mu\text{-a.s.}$$

follows $\sigma_0^1 = \sigma_0^2$, $\sigma^1 = \sigma^2$, $\varrho_0^1 = \varrho_0^2$, $\varrho_1^1 = \varrho_1^2$, $\varrho^1 = \varrho^2$, $\varrho^{*1} = \varrho^{*2}$ and $S_0(x|\sigma_0^1, \boldsymbol{\lambda}_1) = S_0(x|\sigma_0^2, \boldsymbol{\lambda}_2)$ and $S_L(x|\sigma^1, \boldsymbol{\lambda}_1) = S_L(x|\sigma^2, \boldsymbol{\lambda}_2)$ at least on an interval I. The identifiability of the vector $(k_0, k_1, k_3, k_4, k_{10}, k_{11})$ is therefore obvious. From the representation of S_0 and S_L one can furthermore conclude that $H_0(x|\boldsymbol{\lambda}_1) = H_0(x|\boldsymbol{\lambda}_2)$ and $H(x|\boldsymbol{\lambda}_1) = H(x|\boldsymbol{\lambda}_2)$ at least on an interval I. Thus (4.1) implies $\boldsymbol{\lambda}_1 = \boldsymbol{\lambda}_2$.

Assumption (D1):

From (4.3) and the definition of Θ_k it follows that

$$m_s \leq \sigma_0 \leq M_s \quad \text{and} \quad m_s \leq \sigma \leq M_s \; . \tag{4.4}$$

It holds that

$$\begin{aligned} E_{\boldsymbol{\theta}_0} \sup_{\boldsymbol{\theta}\in\Theta} |\ln f_X(X_1, X_2, C_1, C_2|\boldsymbol{\theta})| &= \int \sup_{\boldsymbol{\theta}\in\Theta} |\ln f_X(x_1, x_2, c_1, c_2|\boldsymbol{\theta})| f_X(x_1, x_2, c_1, c_2|\boldsymbol{\theta}_0) d\mu \\ &= \sum_{i=0}^{2}\sum_{j=0}^{2} \int \sup_{\boldsymbol{\theta}\in\Theta} |\ln f_X(x_1, x_2, i, j|\boldsymbol{\theta})| \, f_X(x_1, x_2, i, j|\boldsymbol{\theta}_0) \, d\mu \\ &= \sum_{i=0}^{2}\sum_{j=0}^{2} I(i,j) \; . \end{aligned}$$

Viewing (A.9)-(A.26) it is easy to verify that singularities of the integrands in the integrals $I(i,j)$ can only occur for $x_1, x_2 \to 0$ or $x_1, x_2 \to \infty$. The proof of the finiteness of these integrals aims therefore at using the assumptions made in section 4.1.1 to estimate the behavior of these singularities and to show that despite the singularities of the integrand the integral itself is finite. Since the structure of the integrand varies for the different cases of i, j, these cases have to be investigated separately. The finiteness of the single integrals $I(i,j)$ is proven in Appendix B.2. □

4.2 Consistency of maximum likelihood estimates in a bivariate correlated gamma frailty model

The bivariate correlated frailty model was introduced in section 1.3.2. This section contains a theorem about the consistency of maximum likelihood estimates in this model. The first subsection recalls the definition of the model, whereas the second subsection gives the theorem itself and its proof.

4.2.1 Definition of the model and the parameter space

Definition 11. Let (T_1, T_2) be a random vector of two survival times with marginal survival function $S_L(x) = P(T_i > x)$, $i = 1, 2$. The vector (T_1, T_2) is said to be distributed according to a *bivariate correlated gamma frailty model* if two parameters ϱ_1 and σ with $0 \leq \varrho_1 \leq 1$ and $\sigma > 0$ exist, such that the bivariate survival function S_{L*} of the vector (T_1, T_2) can be represented as

$$\begin{aligned} S_{L*}(x, y) &= P(T_1 > x, T_2 > y) \\ &= [S_L(x) S_L(y)]^{1-\varrho_1} \left[S_L^{-\sigma^2}(x) + S_L^{-\sigma^2}(y) - 1 \right]^{-\frac{\varrho_1}{\sigma^2}} . \end{aligned}$$

In the following it is assumed that censored observations of the survival times T_1 and T_2 are available. That means that a censoring time A exists and that the random vector $(X_1, X_2, \Delta_1, \Delta_2)$ with

$$X_i = \min(T_i, A) \quad \text{and} \quad \Delta_i = \mathrm{I}(T_i \leq A), \quad i = 1, 2, \tag{4.5}$$

is observable.
If A is independent of (T_1, T_2), if it has the survival function S_A and the Lebesgue density function f_A, then a measure $\tilde{\mu}$ exists that dominates the distribution of $(X_1, X_2, \Delta_1, \Delta_2)$. The density of $(X_1, X_2, \Delta_1, \Delta_2)$ with respect to $\tilde{\mu}$ was derived in section 3.2 and is given in equation (3.10).
Identifiability of the parameters ϱ_1 and σ and of the univariate survival function S_L with the help of the distribution of $(X_1, X_2, \Delta_1, \Delta_2)$ was proven only if $\varrho_1 > 0$ (see proposition 3). One of the assumptions for the consistency of MLEs are the compactness of the parameter space and the identifiability of the parameter vector. This is the reason why it assumed in the following that a constant m_r exists such that $m_r \leq \varrho_1 \leq 1$. In addition the parameter σ is restricted to an interval $[m_s, M_s]$.

As in the four-dimensional model a completely parametric estimation approach is assumed. That means that a parametric specification of the univariate survival function S_L is used. Let the parameter $\boldsymbol{\lambda}$ of S_L belong to a compact subset Λ of $\mathbb{R}^{K_u}$. The complete parameter vector is therefore $\boldsymbol{\vartheta} = (\sigma, \varrho_1, \boldsymbol{\lambda}')' = (\vartheta_1, \ldots, \vartheta_{K_p})'$ where $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_{K_u})'$. The parameter set under consideration is

$$\hat{\Theta} = \{\boldsymbol{\vartheta} = (\sigma, \varrho, \boldsymbol{\lambda}')' : m_s \leq \sigma \leq M_s, m_r \leq \varrho_1 \leq 1, \boldsymbol{\lambda} \in \Lambda\} \,. \tag{4.6}$$

The following lemma shows the relation between the parameter set $\hat{\Theta}$ and the parameter set Θ that was used in the proof of consistency for the four-dimensional model.

Lemma 8. *Let $\hat{\Theta}$ be defined as in (4.6) and set $m_k = \frac{m_r}{M_s^2}$ and $M_k = \frac{1}{m_s^2}$. Define the parameter set $\tilde{\Theta}$ by*

$$\begin{aligned}\tilde{\Theta} = \{\boldsymbol{\theta} =& (k_0, k_1, k_3, k_4, k_{10}, k_{11}, \boldsymbol{\lambda}')' : m_k \leq k_0 \leq M_k, 0 \leq k_{11} \leq M_k, \\ & k_1 = k_3 = k_4 = k_{10} = 0, \boldsymbol{\lambda} \in \Lambda\}.\end{aligned}$$

It obviously holds that $\tilde{\Theta} \subset \Theta$ where $\Theta = \Theta_k \times \Lambda$ and Θ_k is defined in (4.2). While replacing σ with $(k_0 + k_{11})^{-\frac{1}{2}}$ and ϱ_1 with $\frac{k_0}{k_0 + k_{11}}$ every function $h(\boldsymbol{\vartheta})$ can be written as a function $\tilde{h}(\boldsymbol{\theta})$ where $\boldsymbol{\theta} \in \tilde{\Theta}$.

Proof. The following relation holds

$$\sigma = (k_0 + k_{11})^{-\frac{1}{2}} \text{ and } \varrho_1 = \frac{k_0}{k_0 + k_{11}} \text{ iff } k_0 = \frac{\varrho_1}{\sigma^2} \text{ and } k_{11} = \frac{1 - \varrho_1}{\sigma^2} \,.$$

Since

$$\begin{aligned} m_k = \frac{m_r}{M_s^2} \leq \frac{\varrho_1}{\sigma^2} \leq \frac{1}{m_s^2} = M_k \quad \text{and} \\ 0 \leq \frac{1 - \varrho_1}{\sigma^2} \leq \frac{1 - m_r}{m_s^2} = (1 - m_r) M_k \leq M_k \end{aligned}$$

the assertion of the lemma follows. □

Corollary 3. *For the functions $h(\boldsymbol{\vartheta})$ and $\tilde{h}(\boldsymbol{\theta})$ defined in lemma 8 it holds that:*

$$\sup_{\boldsymbol{\vartheta} \in \hat{\Theta}} h(\boldsymbol{\vartheta}) \leq \sup_{\boldsymbol{\theta} \in \tilde{\Theta}} \tilde{h}(\boldsymbol{\theta}) \leq \sup_{\boldsymbol{\theta} \in \Theta} \tilde{h}(\boldsymbol{\theta}) \,.$$

4.2.2 Proof of consistency of MLEs

The consistency of maximum likelihood estimates in the bivariate correlated frailty model is in principle a special cases of the consistency of the maximum likelihood estimates in the four-dimensional model which was shown in section 4.1.3. f_X^* denotes the density function of $(X_1, X_2, \Delta_1, \Delta_2)$ with respect to $\tilde{\mu}$ (see (3.10)). f_X denotes the density function of (X_1, X_2, C_1, C_2) with respect to μ which was used in theorem 4 (see (2.20)). Using lemma 8 one can conclude that the following relation holds:

$$\begin{aligned}
f_X^*(x_1, x_2, 0, 0|\boldsymbol{\vartheta}) &= f_X(x_1, x_2, 0, 0|\boldsymbol{\theta}) + f_X(x_1, x_2, 1, 0|\boldsymbol{\theta}) + f_X(x_1, x_2, 0, 1|\boldsymbol{\theta}) \\
&\quad + f_X(x_1, x_2, 1, 1|\boldsymbol{\theta}) \\
f_X^*(x_1, x_2, 1, 0|\boldsymbol{\vartheta}) &= f_X(x_1, x_2, 2, 0|\boldsymbol{\theta}) + f_X(x_1, x_2, 2, 1|\boldsymbol{\theta}) \\
f_X^*(x_1, x_2, 0, 1|\boldsymbol{\vartheta}) &= f_X(x_1, x_2, 0, 2|\boldsymbol{\theta}) + f_X(x_1, x_2, 1, 2|\boldsymbol{\theta}) \\
f_X^*(x_1, x_2, 1, 1|\boldsymbol{\vartheta}) &= f_X(x_1, x_2, 2, 2|\boldsymbol{\theta})
\end{aligned}$$

where $\boldsymbol{\theta} \in \tilde{\Theta}$.
Since for the bivariate model only assumptions on the parameterization of the univariate survival function S_L have to be formulated, a separate theorem about the consistency of maximum likelihood estimates in the bivariate correlated frailty model is given here. Some results that have been derived in the proof of theorem 4 are used in the proof of the theorem in this section.

Theorem 5. *Let Λ be a compact subset of $\mathbb{R}^{K_u}$. Let m_r, m_s and M_s be three arbitrary constants with $0 < m_r < 1$ and $0 < m_k < M_k < \infty$. Set $\boldsymbol{\vartheta} = (\sigma, \varrho_1, \boldsymbol{\lambda}')'$ and define the parameter set $\hat{\Theta}$ by (4.6).*
Let $\mathfrak{H} = \{H(.|\boldsymbol{\lambda}) : \boldsymbol{\lambda} \in \Lambda\}$ be a family of cumulative hazard functions that fulfills assumptions 2 and 3. Assume that $H(x|\boldsymbol{\lambda}_1) = H(x|\boldsymbol{\lambda}_2)$ for all x in an arbitrary interval I and for $\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2 \in \Lambda$ implies $\boldsymbol{\lambda}_1 = \boldsymbol{\lambda}_2$. A family of univariate survival functions is defined by

$$\mathfrak{S} = \left\{ S_L(x|\sigma, \boldsymbol{\lambda}) = [1 + \sigma^2 H(x|\boldsymbol{\lambda})]^{-1/\sigma^2} : H \in \mathfrak{H}, m_s \leq \sigma \leq M_s \right\}.$$

It is assumed that the family $\mathfrak{S}$ fulfills the assumptions 4(a), 4(b) and 4(d). Let $P_{\boldsymbol{\vartheta}}$ denote the probability distribution on $(\mathbb{R}_+^2, \mathfrak{B}_+^2)$ that corresponds to the bivariate survival function:

$$S_{L*}(x, y|\boldsymbol{\vartheta}) = [S_L(x|\sigma, \boldsymbol{\lambda}) S_L(y|\sigma, \boldsymbol{\lambda})]^{1-\varrho_1} \left[S_L^{-\sigma^2}(x|\sigma, \boldsymbol{\lambda}) + S_L^{-\sigma^2}(y|\sigma, \boldsymbol{\lambda}) - 1 \right]^{-\frac{\varrho_1}{\sigma^2}}$$

and denote $\hat{\mathfrak{P}} = \{P_{\boldsymbol{\vartheta}} : \boldsymbol{\vartheta} \in \hat{\Theta}\}$. If $(T_1, T_2) \sim P_{\boldsymbol{\vartheta}}$ then the random vector $(X_1, X_2, \Delta_1, \Delta_2)$ is defined by

$$X_i = \min(T_i, A) \quad \text{and} \quad \Delta_i = \mathrm{I}(T_i \leq A), \quad i = 1, 2.$$

A is a non-negative random variable that has the Lebesgue density f_A and the univariate survival function S_A. It is assumed that A is independent of (T_1, T_2) and that it fulfills assumption 6. Furthermore, it is assumed that an interval I with $f_A(x) > 0$ and $\mu(x|\boldsymbol{\lambda}) > 0$ for all $x \in I$ and all $\boldsymbol{\lambda} \in \Lambda$ exists. Let $Q_{\boldsymbol{\vartheta}}$ denote the distribution of $(X_1, X_2, \Delta_1, \Delta_2)$ and let $\hat{\Omega} = \{Q_{\boldsymbol{\vartheta}} : \boldsymbol{\vartheta} \in \hat{\Theta}\}$. A σ-finite measure $\tilde{\mu}$ on the sample space of $(X_1, X_2, \Delta_1, \Delta_2)$ exists that dominates the family $\hat{\Omega}$. Let $f_X^*(.|\boldsymbol{\vartheta})$ denote the density of $Q_{\boldsymbol{\vartheta}}$ with respect to $\tilde{\mu}$ and let $(x_1^i, x_2^i, \delta_1^i, \delta_2^i)$ denote the i^{th} observation. Then every sequence of measurable maximum likelihood estimates $(\hat{\boldsymbol{\vartheta}}_n)$ of $\boldsymbol{\vartheta}$ that is derived by maximizing the likelihood function

$$L^*(\mathbf{x}_1, \mathbf{x}_2, \boldsymbol{\delta}_1, \boldsymbol{\delta}_2|\boldsymbol{\vartheta}) = \prod_{i=1}^{n} f_X^*(x_1^i, x_2^i, \delta_1^i, \delta_2^i|\boldsymbol{\vartheta})$$

is consistent. This means that if $\boldsymbol{\vartheta}_0$ denotes the true parameter and $(\hat{\boldsymbol{\vartheta}}_n)$ the sequence of maximum likelihood estimates, then $\hat{\boldsymbol{\vartheta}}_n \xrightarrow{P} \boldsymbol{\vartheta}_0$.

Proof. As for the four-dimensional model the proof is based on theorem 1. The measure $\tilde{\mu}$ and the density function f_X^* are given in (3.8) and (3.10), respectively. The explicit representation of all parts of f_X^* can be found in (A.20), (A.23) and (A.25). Now the assumptions of theorem 1 are verified.

Assumption (A1) follows from the definition of $\hat{\Theta}$.

Assumption (B1)
is a consequence of the form of the density function f_X^* (see (3.10), (A.20), (A.23) and (A.25)) and from assumptions 2(e) and 3(b).

Assumption (C1):
From theorem 3 one can conclude that from

$$f_X^*(x_1, x_2, \delta_1, \delta_2|\sigma_1, \varrho_{11}, \boldsymbol{\lambda}_1) = f_X^*(x_1, x_2, \delta_1, \delta_2|\sigma_2, \varrho_{12}, \boldsymbol{\lambda}_2) \quad \tilde{\mu}\text{-a.s.}$$

it follows that $\sigma_1 = \sigma_2$, $\varrho_{11} = \varrho_{12}$ and $S_L(x|\sigma_1, \boldsymbol{\lambda}_1) = S_L(x|\sigma_2, \boldsymbol{\lambda}_2)$ at least on an interval I. The representation of S_L thus yields that $H(x|\boldsymbol{\lambda}_1) = H(x|\boldsymbol{\lambda}_2)$ at least on an interval I. It was assumed that this implies $\boldsymbol{\lambda}_1 = \boldsymbol{\lambda}_2$.

Assumption (D1):
One can write:

$$\begin{aligned} &E_{\boldsymbol{\vartheta}_0} \sup_{\boldsymbol{\vartheta} \in \hat{\Theta}} |\ln f_X^*(X_1, X_2, \Delta_1, \Delta_2|\boldsymbol{\vartheta})| \\ &\quad = \sum_{i,j=0,1} \int \sup_{\boldsymbol{\vartheta} \in \hat{\Theta}} |\ln f_X^*(x_1, x_2, i, j|\boldsymbol{\vartheta})|\, f_X^*(x_1, x_2, i, j|\boldsymbol{\vartheta}_0)\, d\tilde{\mu} = \sum_{i,j=0,1} \tilde{I}(i,j). \end{aligned}$$

From (3.10), (A.20), (A.23), (A.25) and lemma 8 follows that the parts of the density function f_X^* can be represented as

$$f_X^*(x_1, x_2, 0, 0|\boldsymbol{\vartheta}) = f_{11}(x_1, x_2|\boldsymbol{\theta}) f_A(x_1)\, \mathrm{I}(x_1 = x_2) \tag{4.7}$$
$$f_X^*(x_1, x_2, 1, 0|\boldsymbol{\vartheta}) = f_{21}(x_1, x_2|\boldsymbol{\theta})\, f_A(x_2)\, \mathrm{I}(x_1 \leq x_2) \tag{4.8}$$
$$f_X^*(x_1, x_2, 1, 1|\boldsymbol{\vartheta}) = f_{22}(x_1, x_2|\boldsymbol{\theta})\, S_A(\max(x_1, x_2)) \tag{4.9}$$

where $\boldsymbol{\theta} \in \tilde{\Theta} \subset \Theta$. The finiteness of the single integrals $\tilde{I}(i, j)$ will be proven in Appendix B.2. The proof uses results derived in the proof of theorem 4. □

4.3 Asymptotic normality in a bivariate correlated frailty model

In this section the asymptotic normality of maximum likelihood estimates in the bivariate correlated frailty model is proven. Functions and notations that have been used throughout the whole section are summarized in Appendix A.1. The first subsection summarizes the assumptions that have not yet been necessary to prove consistency. The second subsection includes some lemmata that prepare the proof of asymptotic normality. These lemmata are very technical and therefore their proofs are given in Appendix B.3. Finally the third subsection gives the theorem about asymptotic normality in the bivariate correlated frailty model and its proof.

4.3.1 Assumptions

The following assumptions, together with the assumptions used in the proof of consistency in the bivariate model (see theorem 5), are sufficient for the asymptotic normality of maximum likelihood estimates.

Assumption 7. Let $\mathfrak{H} = \{H(.|\boldsymbol{\lambda}) : \boldsymbol{\lambda} \in \Lambda\}$ be a family of cumulative hazard functions that fulfills assumptions 2 and 3. Assume that the functions $H(x|\boldsymbol{\lambda})$ and $\mu(x|\boldsymbol{\lambda}) = \frac{\partial}{\partial x} H(x|\boldsymbol{\lambda})$ are twice continuously differentiable in $\boldsymbol{\lambda}$. Let m_s, M_s be two constants with $0 < m_s < M_s < \infty$. A family of univariate survival functions is defined by

$$\mathfrak{S} = \{S_L(x|\sigma, \boldsymbol{\lambda}) = [1 + \sigma^2 H(x|\boldsymbol{\lambda})]^{-1/\sigma^2} : H \in \mathfrak{H}, m_s \leq \sigma \leq M_s\}\,. \tag{4.10}$$

Let $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_{K_u})$. It is assumed that for all $\sigma \in [m_s, M_s]$, all $\tilde{\boldsymbol{\lambda}} \in \Lambda$ and all $x \geq 0$ the following is fulfilled:

(a) A constant $\tilde{C}_0 > 0$ exists such that for $\alpha = \frac{1}{2}, 1$ it holds

$$\sup_{\boldsymbol{\lambda} \in \Lambda} \ln[1 + M_s^2 H(x|\boldsymbol{\lambda})] \cdot S_L^{\alpha}(x|\sigma, \tilde{\boldsymbol{\lambda}}) \leq \tilde{C}_0 \ .$$

(b) A constant $C_1 > 0$ exists such that for $\alpha = \frac{1}{4}, \frac{1}{2}, 1$ it holds that

$$\sup_{\boldsymbol{\lambda} \in \Lambda} \left| \frac{\partial \ln H(x|\boldsymbol{\lambda})}{\partial \lambda_j} \right| \cdot S_L^{\alpha}(x|\sigma, \tilde{\boldsymbol{\lambda}}) \leq C_1 \quad \text{for } j \in \{1, \ldots, K_u\}.$$

(c) A constant $C_2 > 0$ exists such that for $\alpha = \frac{1}{2}, 1$ it holds that

$$\sup_{\boldsymbol{\lambda} \in \Lambda} \left| \frac{\partial^2 \ln H(x|\boldsymbol{\lambda})}{\partial \lambda_k \partial \lambda_l} \right| \cdot S_L^{\alpha}(x|\sigma, \tilde{\boldsymbol{\lambda}}) \leq C_2 \quad \text{for } k, l \in \{1, \ldots, K_u\}.$$

(d) A constant $\hat{C}_1 > 0$ exists such that for $\alpha = \frac{1}{4}, \frac{1}{2}, 1$ it holds that

$$\sup_{\boldsymbol{\lambda} \in \Lambda} \left| \frac{\partial \ln \mu(x|\boldsymbol{\lambda})}{\partial \lambda_j} \right| \cdot S_L^{\alpha}(x|\sigma, \tilde{\boldsymbol{\lambda}}) \leq \hat{C}_1 \quad \text{for } j \in \{1, \ldots, K_u\}.$$

(e) A constant $\hat{C}_2 > 0$ exists such that for $\alpha = \frac{1}{2}, 1$ it holds that

$$\sup_{\boldsymbol{\lambda} \in \Lambda} \left| \frac{\partial^2 \ln \mu(x|\boldsymbol{\lambda})}{\partial \lambda_k \partial \lambda_l} \right| \cdot S_L^{\alpha}(x|\sigma, \tilde{\boldsymbol{\lambda}}) \leq \hat{C}_2 \quad \text{for } k, l \in \{1, \ldots, K_u\}.$$

Assumption 8. Let $\mathfrak{H}$ be a family of cumulative hazard functions that fulfills assumption 7. A family of univariate survival functions S_L is defined according to (4.10) and a family of bivariate survival functions is defined by

$$S_{L*}(x, y|\boldsymbol{\vartheta}) = [S_L(x|\boldsymbol{\vartheta}) S_L(y|\boldsymbol{\vartheta})]^{1-\varrho_1} \left[S_L^{-\sigma^2}(x|\boldsymbol{\vartheta}) + S_L^{-\sigma^2}(y|\boldsymbol{\vartheta}) - 1 \right]^{-\frac{\varrho_1}{\sigma^2}}$$

where $\boldsymbol{\vartheta} = (\sigma, \varrho_1, \boldsymbol{\lambda}')' \in \hat{\Theta}$ from (4.6). A density function f_X^* on the measurable space $(\mathbb{R}_+^2 \times \{0,1\} \times \{0,1\}, \mathfrak{B}^2 \otimes \mathcal{P}^{\{0,1\}} \otimes \mathcal{P}^{\{0,1\}})$ with respect to the measure $\tilde{\mu}$ from (3.8) is then given by (3.10). It is assumed that for all $\boldsymbol{\vartheta} \in \hat{\Theta}^{\circ}$ the determinant of Fisher's information matrix $I(\boldsymbol{\vartheta}) = \left(I_{ij}(\boldsymbol{\vartheta})\right)_{i=1,\ldots,K_p}$ is unequal to zero, where

$$I_{ij}(\boldsymbol{\vartheta}) = \int \frac{\partial \ln f_X^*(x_1, x_2, \delta_1, \delta_2|\boldsymbol{\vartheta})}{\partial \vartheta_i} \, \frac{\partial \ln f_X^*(x_1, x_2, \delta_1, \delta_2|\boldsymbol{\vartheta})}{\partial \vartheta_j} \, f_X^*(x_1, x_2, \delta_1, \delta_2|\boldsymbol{\vartheta}) \, d\tilde{\mu}.$$

4.3.2 Corollaries and lemmata

Corollary 4. *From* (4.7), (4.8) *and* (4.9) *it follows that the elements of Fisher's information matrix used in assumption 8 are given by*

$$
\begin{aligned}
I_{ij}(\boldsymbol{\vartheta}) = &\int_0^\infty \frac{\partial \ln f_{11}(x,x|\boldsymbol{\vartheta})}{\partial \vartheta_i} \frac{\partial \ln f_{11}(x,x|\boldsymbol{\vartheta})}{\partial \vartheta_j} f_{11}(x,x,|\boldsymbol{\vartheta}) f_A(x)\, dx \\
&+2\int_0^\infty \int_0^{x_2} \frac{\partial \ln f_{21}(x_1,x_2|\boldsymbol{\vartheta})}{\partial \vartheta_i} \frac{\partial \ln f_{21}(x_1,x_2|\boldsymbol{\vartheta})}{\partial \vartheta_j} f_{21}(x_1,x_2,|\boldsymbol{\vartheta}) f_A(x_2)\, dx_1 dx_2 \\
&+\iint \frac{\partial \ln f_{22}(x_1,x_2|\boldsymbol{\vartheta})}{\partial \vartheta_i} \frac{\partial \ln f_{22}(x_1,x_2|\boldsymbol{\vartheta})}{\partial \vartheta_j} f_{22}(x_1,x_2,|\boldsymbol{\vartheta}) S_A(\max(x_1,x_2))\, dx_1 dx_2
\end{aligned}
$$

where the functions f_{11}, f_{21} *and* f_{22} *can be found in* (A.20), (A.23) *and* (A.25), *respectively.*

The following technical lemmata prepare the proof of the main theorem about asymptotic normality in the bivariate correlated gamma frailty model. The lemmata are based on the representation

$$
\begin{aligned}
f_X^*(x,y,0,0|\boldsymbol{\vartheta}) &= g_{0,1}(x,y|\boldsymbol{\vartheta}) f_A(x)\, \mathrm{I}(x=y) \\
f_X^*(x,y,0,1|\boldsymbol{\vartheta}) &= g_{1,1}(y,x|\boldsymbol{\vartheta}) f_A(x)\, \mathrm{I}(x\geq y) \\
f_X^*(x,y,1,0|\boldsymbol{\vartheta}) &= g_{1,1}(x,y|\boldsymbol{\vartheta}) f_A(y)\, \mathrm{I}(x\leq y) \\
f_X^*(x,y,1,1|\boldsymbol{\vartheta}) &= g_{2,1}(x,y|\boldsymbol{\vartheta}) S_A(\max(x,y))
\end{aligned}
\tag{4.11}
$$

for the parts of the density function f_X^*. The functions $g_{0,\alpha}$, $g_{1,\alpha}$ and $g_{2,\alpha}$ are defined in (A.31), (A.32) and (A.34), respectively. The proof of the lemmata can be found in Appendix B.3.

Lemma 9. *Let* $\mathfrak{H} = \{H(.|\boldsymbol{\lambda}) : \boldsymbol{\lambda} \in \Lambda\}$ *be a family of cumulative hazard functions. Let* $\mathfrak{S}$ *be a family of univariate survival functions defined according to* (4.10) *and let assumptions 2, 3, 7(a), 7(b) and 7(c) be fulfilled. Let the function* $g_{0,\alpha}(x,y|\boldsymbol{\vartheta})$ *be defined by* (A.31). *Then for* $i,j \in \{1,\ldots,K_p\}$ *two constants* $c_0 > 0$ *and* $d_0 > 0$ *exist such that*

$$
\left|\frac{\partial g_{0,\alpha}(x,y|\boldsymbol{\vartheta})}{\partial \vartheta_i}\right| \leq c_0 \quad \textit{for } \alpha = \tfrac{1}{2}, 1 \textit{ and } \quad \left|\frac{\partial^2 g_{0,\alpha}(x,y|\boldsymbol{\vartheta})}{\partial \vartheta_i \partial \vartheta_j}\right| \leq d_0 \quad \textit{for } \alpha = 1,
$$

for all $x,y \geq 0$ *and all* $\boldsymbol{\vartheta} \in \hat{\Theta}$.

Proof. The proof of lemma 9 can be found in Appendix B.3 on page 133. □

Lemma 10. *Let $\mathfrak{H} = \{H(.|\boldsymbol{\lambda}) : \boldsymbol{\lambda} \in \Lambda\}$ be a family of cumulative hazard functions. Let $\mathfrak{S}$ be a family of univariate survival functions defined according to* (4.10) *and let assumptions 2, 3, 4(a) and 7 be fulfilled. Let the functions $g_{1,\alpha}(x, y|\boldsymbol{\vartheta})$ and $g_{2,\alpha}(x, y|\boldsymbol{\vartheta})$ be defined by* (A.32) *and* (A.34)*, respectively. Then positive constants G_1, H_1, G_2 and H_2 exist such that for all $i, j \in \{1, \ldots, K_p\}$ it holds that*

$$\left|\frac{\partial g_{1,\alpha}(x, y|\boldsymbol{\vartheta})}{\partial \vartheta_i}\right| \leq G_1 \text{ and } \left|\frac{\partial^2 g_{1,\alpha}(x, y|\boldsymbol{\vartheta})}{\partial \vartheta_i \partial \vartheta_j}\right| \leq H_1 \text{ for } \alpha = \tfrac{1}{2}, 1 \text{ and}$$
$$\left|\frac{\partial g_{2,1}(x, y|\boldsymbol{\vartheta})}{\partial \vartheta_i}\right| \leq G_2 \text{ and } \left|\frac{\partial^2 g_{2,1}(x, y|\boldsymbol{\vartheta})}{\partial \vartheta_i \partial \vartheta_j}\right| \leq H_2$$

for all $x, y \geq 0$ and all $\boldsymbol{\vartheta} \in \hat{\Theta}$.

Proof. The proof of lemma 10 can be found in Appendix B.3 on page 141. □

Lemma 11. *Let $\mathfrak{H} = \{H(.|\boldsymbol{\lambda}) : \boldsymbol{\lambda} \in \Lambda\}$ be a family of cumulative hazard functions that fulfills assumptions 2 and 3. Let the family of univariate survival functions $\mathfrak{S}$ be defined by* (4.10) *and let assumptions 7(a), 7(b) and 7(c) be fulfilled. Let the function $g_{0,\alpha}(x, y|\boldsymbol{\vartheta})$ be defined by* (A.31)*. Then two functions $M_0(x, y)$ and $\tilde{M}_0(x, y)$ and two constants $\tilde{c}_0 > 0$ and $\tilde{d}_0 > 0$ exist such that*

$$M_0(x, y) \cdot g_{0,\alpha}(x, y|\tilde{\boldsymbol{\vartheta}}) \leq \tilde{c}_0 \qquad \text{for } \alpha = \tfrac{1}{2}, 1 \text{ and}$$
$$\tilde{M}_0(x, y) \cdot g_{0,\alpha}(x, y|\tilde{\boldsymbol{\vartheta}}) \leq \tilde{d}_0 \qquad \text{for } \alpha = 1,$$

for all $x, y \geq 0$ and all $\tilde{\boldsymbol{\vartheta}} \in \hat{\Theta}$ and it holds that

$$\left|\frac{\partial \ln g_{0,1}(x, y|\boldsymbol{\vartheta})}{\partial \vartheta_i}\right| \leq M_0(x, y) \quad \text{and} \quad \left|\frac{\partial^2 \ln g_{0,1}(x, y|\boldsymbol{\vartheta})}{\partial \vartheta_i \partial \vartheta_j}\right| \leq \tilde{M}_0(x, y)$$

for $i, j \in \{1, \ldots, K_p\}$, for all $x, y \geq 0$ and all $\boldsymbol{\vartheta} \in \hat{\Theta}$.

Proof. The proof of lemma 11 can be found in Appendix B.3 on page 142. □

Lemma 12. *Let $\mathfrak{H} = \{H(.|\boldsymbol{\lambda}) : \boldsymbol{\lambda} \in \Lambda\}$ be a family of cumulative hazard functions that fulfills assumptions 2 and 3. Let the family of univariate survival functions $\mathfrak{S}$ be defined by* (4.10) *and let assumption 7 be fulfilled. The functions $g_{0,1}(x, y|\boldsymbol{\vartheta})$ and $g_{1,\alpha}(x, y|\boldsymbol{\vartheta})$ are defined by* (A.31) *and* (A.32)*, respectively. Two functions $M_1(x, y)$ and $\tilde{M}_1(x, y)$ and two constants $\tilde{c}_1 > 0$ and $\tilde{d}_1 > 0$ exist such that*

$$M_1(x, y) \cdot g_{0,\tilde{\alpha}}(x, y|\tilde{\boldsymbol{\vartheta}}) \leq \tilde{c}_1 \quad \text{for } \tilde{\alpha} = \tfrac{1}{2}, 1 \text{ and} \quad \tilde{M}_1(x, y) \cdot g_{0,1}(x, y|\tilde{\boldsymbol{\vartheta}}) \leq \tilde{d}_1$$

for all $x, y \geq 0$ *and all* $\tilde{\boldsymbol{\vartheta}} \in \hat{\Theta}$ *and it holds*

$$\left|\frac{\partial \ln g_{1,\alpha}(x,y|\boldsymbol{\vartheta})}{\partial \vartheta_i}\right| \leq \frac{M_1(x,y)}{1-\varrho_1} \quad \text{for } \alpha = \tfrac{1}{2}, 1 \text{ and}$$
$$\left|\frac{\partial^2 \ln g_{1,1}(x,y|\boldsymbol{\vartheta})}{\partial \vartheta_i \partial \vartheta_j}\right| \leq \frac{\tilde{M}_1(x,y)}{(1-\varrho_1)^2} \quad \text{for } \alpha = \tfrac{1}{2}, 1,$$

for $i, j \in \{1, \ldots, K_p\}$, *for all* $x, y \geq 0$ *and all* $\boldsymbol{\vartheta} = (\sigma, \varrho_1, \boldsymbol{\lambda}')' \in \hat{\Theta}$.

Proof. The proof of lemma 12 can be found in Appendix B.3 on page 143. □

Lemma 13. *Let* $\mathfrak{H} = \{H(.|\boldsymbol{\lambda}) : \boldsymbol{\lambda} \in \Lambda\}$ *be a family of cumulative hazard functions that fulfills assumptions 2 and 3. Let the family of univariate survival functions* $\mathfrak{S}$ *be defined by* (4.10) *and let assumption 7 be fulfilled. The functions* $g_{0,1}(x,y|\boldsymbol{\vartheta})$ *and* $g_{2,1}(x,y|\boldsymbol{\vartheta})$ *are defined by* (A.31) *and* (A.34), *respectively. Then a function* $\tilde{M}_2(x,y)$ *and a constant* $\tilde{d}_2 > 0$ *exist such that*

$$\tilde{M}_2(x,y) \cdot g_{0,1}(x,y|\tilde{\boldsymbol{\vartheta}}) \leq \tilde{d}_2$$

for all $x, y \geq 0$ *and all* $\tilde{\boldsymbol{\vartheta}} \in \hat{\Theta}$ *and it holds*

$$\left|\frac{\partial^2 \ln g_{2,1}(x,y|\boldsymbol{\vartheta})}{\partial \vartheta_i \partial \vartheta_j}\right| \leq \frac{\tilde{M}_2(x,y)}{(1-\varrho_1)^2}$$

for $i, j \in \{1, \ldots, K_p\}$, *for all* $x, y \geq 0$ *and all* $\boldsymbol{\vartheta} = (\sigma, \varrho_1, \boldsymbol{\lambda}')' \in \hat{\Theta}$.

Proof. The proof of lemma 13 can be found in Appendix B.3 on page 146. □

4.3.3 Proof of asymptotic normality

Before the main theorem is formulated some properties of the density function f_X^* of censored observations in the bivariate correlated gamma frailty model are shown.

Proposition 4. *Let* $\mathfrak{H} = \{H(.|\boldsymbol{\lambda}) : \boldsymbol{\lambda} \in \Lambda\}$ *be a family of cumulative hazard functions and let* $\mathfrak{S}$ *be a family of univariate survival functions defined according to* (4.10). *Let assumptions 2, 3, 4(a) and 7 be fulfilled. The density function* f_X^* *on* $(\mathbb{R}_+^2 \times \{0,1\} \times \{0,1\}, \mathfrak{B}^2 \otimes \mathcal{P}^{\{0,1\}} \otimes \mathcal{P}^{\{0,1\}})$ *is given by* (4.11) *and the dominating measure* $\tilde{\mu}$ *is defined in* (3.8). *If for the function* $f_A(x)$ *it holds that* $\int x^2 f_A(x)\, dx < \infty$ *then two measurable functions* G *and* H *with*

$\int G(x,y,\delta_1,\delta_2)\,d\tilde{\mu} < \infty$ and $\int H(x,y,\delta_1,\delta_2)\,d\tilde{\mu} < \infty$ exist such that for all $i,j \in \{1,\ldots,K_p\}$, all $x,y \geq 0$, for $\delta_1,\delta_2 \in \{0,1\}$ and all $\boldsymbol{\vartheta} \in \hat{\Theta}$ it holds that

$$\left|\frac{\partial f_X^*(x,y,\delta_1,\delta_2|\boldsymbol{\vartheta})}{\partial \vartheta_i}\right| \leq G(x,y,\delta_1,\delta_2) \quad \textit{and}$$
$$\left|\frac{\partial^2 f_X^*(x,y,\delta_1,\delta_2|\boldsymbol{\vartheta})}{\partial \vartheta_i \partial \vartheta_j}\right| \leq H(x,y,\delta_1,\delta_2)\ .$$

Proof. One can set

$$\begin{aligned} G(x,y,\delta_1,\delta_2) = c_0 \cdot f_A(x)\,\mathrm{I}(x=y)\,\mathrm{I}(\delta_1=0)\,\mathrm{I}(\delta_2=0) \\ + G_1 \cdot f_A(x)\,\mathrm{I}(y\leq x)\,\mathrm{I}(\delta_1=0)\,\mathrm{I}(\delta_2=1) \\ + G_1 \cdot f_A(y)\,\mathrm{I}(x\leq y)\,\mathrm{I}(\delta_1=1)\,\mathrm{I}(\delta_2=0) \\ + G_2 \cdot S_A(\max(x,y))\,\mathrm{I}(\delta_1=1)\,\mathrm{I}(\delta_2=1) \end{aligned}$$

where the constants c_0, G_1 and G_2 are chosen according to the lemmata 9 and 10. Then from (4.11) and corollary 7 it follows that

$$\left|\frac{\partial f_X^*(x,y,\delta_1,\delta_2|\boldsymbol{\vartheta})}{\partial \vartheta_i}\right| \leq G(x,y,\delta_1,\delta_2)$$

and it holds that

$$\begin{aligned} \int G(x,y,\delta_1,\delta_2)\,d\tilde{\mu} &= c_0 \int f_A(x)\,dx + G_1 \iint f_A(x)\,\mathrm{I}(y\leq x)\,dy\,dx \\ &\quad + G_1 \iint f_A(y)\,\mathrm{I}(x\leq y)\,dx\,dy + G_2 \iint S_A(\max(x,y))\,dx\,dy \\ &= c_0 + 2\,G_1 \int x f_A(x)\,dx + G_2 \int x^2 f_A(x)\,dx < \infty\ . \end{aligned}$$

Analogously one can set

$$\begin{aligned} H(x,y,\delta_1,\delta_2) = d_0 \cdot f_A(x)\,\mathrm{I}(x=y)\,\mathrm{I}(\delta_1=0)\,\mathrm{I}(\delta_2=0) \\ + H_1 \cdot f_A(x)\,\mathrm{I}(y\leq x)\,\mathrm{I}(\delta_1=0)\,\mathrm{I}(\delta_2=1) \\ + H_1 \cdot f_A(y)\,\mathrm{I}(x\leq y)\,\mathrm{I}(\delta_1=1)\,\mathrm{I}(\delta_2=0) \\ + H_2 \cdot S_A(\max(x,y))\,\mathrm{I}(\delta_1=1)\,\mathrm{I}(\delta_2=1) \end{aligned}$$

where the constants d_0, H_1 and H_2 are chosen according to the lemmata 9 and 10. □

Corollary 5. *For the density function f_X^* of the vector $(X_1, X_2, \Delta_1, \Delta_2)$, for all $i, j \in \{1, \dots, K_p\}$ and all $\boldsymbol{\vartheta} \in \hat{\Theta}$ under the mentioned assumptions it holds that*

$$\int \frac{\partial}{\partial \vartheta_i} f_X^*(x, y, \delta_1, \delta_2 | \boldsymbol{\vartheta}) \, d\tilde{\mu} = 0 \text{ and } \int \frac{\partial^2}{\partial \vartheta_i \partial \vartheta_j} f_X^*(x, y, \delta_1, \delta_2 | \boldsymbol{\vartheta}) = 0 \ .$$

Proof. The corollary is a direct consequence of lemma 17 and proposition 4. □

Proposition 5. *Let $\hat{\Theta}^\circ$ denote the open kernel of the parameter set $\hat{\Theta}$, i.e.*

$$\hat{\Theta}^\circ = \{(\sigma, \varrho_1, \boldsymbol{\lambda}) : m_s < \sigma < M_s, m_r < \varrho_1 < 1, \boldsymbol{\lambda} \in \Lambda^\circ\}.$$

Let $\mathfrak{H} = \{H(.|\boldsymbol{\lambda}) : \boldsymbol{\lambda} \in \Lambda\}$ be a family of cumulative hazard functions and let $\mathfrak{S}$ be a family of univariate survival functions defined according to (4.10). Let assumptions 2, 3, 4(a) and 7 be fulfilled. Let (T_1, T_2) be a random vector which is distributed according to a bivariate correlated gamma frailty model (see definition 11). The random vector $(X_1, X_2, \Delta_1, \Delta_2)$ is defined by (4.5). The density function f_X^* of $(X_1, X_2, \Delta_1, \Delta_2)$ is given by (4.11) where the dominating measure $\tilde{\mu}$ is defined in (3.8). If for the density function $f_A(x)$ of the censoring variable it holds that $\int x^2 f_A(x) \, dx < \infty$, then for every $\boldsymbol{\vartheta} \in \hat{\Theta}^\circ$ a $\delta_{\boldsymbol{\vartheta}} > 0$ with $U(\boldsymbol{\vartheta}, \delta_{\boldsymbol{\vartheta}}) \subset \hat{\Theta}^\circ$ and a measurable function $M(.|\boldsymbol{\vartheta})$ with $E_{\boldsymbol{\vartheta}} M(X_1, X_2, \Delta_1, \Delta_2 | \boldsymbol{\vartheta}) < \infty$ exist such that

$$\left| \frac{\partial^2 \ln f_X^*(x, y, \delta_1, \delta_2 | \tilde{\boldsymbol{\vartheta}})}{\partial \vartheta_k \partial \vartheta_l} \right| \leq M(x, y, \delta_1, \delta_2 | \boldsymbol{\vartheta}) \tag{4.12}$$

for all $x, y \geq 0$, $\delta_1, \delta_2 \in \{0, 1\}$, all $\tilde{\boldsymbol{\vartheta}} \in U(\boldsymbol{\vartheta}, \delta_{\boldsymbol{\vartheta}})$ and all $k, l \in \{1, \dots, K_p\}$.

Proof. One can set

$$\begin{aligned} M(x, y, \delta_1, \delta_2 | \boldsymbol{\vartheta}) = \tilde{M}_0(x, y) \, \mathrm{I}(\delta_1 = 0) \, \mathrm{I}(\delta_2 = 0) + \frac{\tilde{M}_1(x, y)}{(1 - \varrho_1)^2} \, \mathrm{I}(\delta_1 = 1) \, \mathrm{I}(\delta_2 = 0) \\ + \frac{\tilde{M}_1(y, x)}{(1 - \varrho_1)^2} \, \mathrm{I}(\delta_1 = 0) \, \mathrm{I}(\delta_2 = 1) + \frac{\tilde{M}_2(x, y)}{(1 - \varrho_1)^2} \, \mathrm{I}(\delta_1 = 1) \, \mathrm{I}(\delta_2 = 1) \end{aligned}$$

where the functions $\tilde{M}_0$, $\tilde{M}_1$ and $\tilde{M}_2$ are chosen according to the lemmata 11, 12 and 13, respectively.

$0 \leq g_{1,1}(x,y|\boldsymbol{\vartheta}) \leq K_1 g_{0,1}(x,y|\boldsymbol{\vartheta})$ and $0 \leq g_{2,1}(x,y|\boldsymbol{\vartheta}) \leq (K_2+K_1^2) g_{0,1}(x,y|\boldsymbol{\vartheta})$ follow from lemma 20, (A.32) and (A.34). Hence it holds that

$$\begin{aligned}
E_{\boldsymbol{\vartheta}} M(X_1,X_2,\Delta_1,\Delta_2|\boldsymbol{\vartheta}) &= \int M(x,y,\delta_1,\delta_2|\boldsymbol{\vartheta}) f_X^*(x,y,\delta_1,\delta_2|\boldsymbol{\vartheta})\, d\tilde{\mu} \\
&= \sum_{i,j=0,1} \iint M(x,y,i,j|\boldsymbol{\vartheta}) f_X^*(x,y,i,j|\boldsymbol{\vartheta})\, d\tilde{\mu} \\
&\leq \int \tilde{M}_0(x,x) g_{0,1}(x,x|\boldsymbol{\vartheta}) f_A(x)\, dx \\
&\quad + \frac{2\,K_1}{(1-\varrho_1)^2} \iint \tilde{M}_1(x,y) g_{0,1}(x,y|\boldsymbol{\vartheta})\, \mathrm{I}(x \leq y) f_A(y)\, dx\, dy \\
&\quad + \frac{K_2+K_1^2}{(1-\varrho_1)^2} \iint \tilde{M}_2(x,y) g_{0,1}(x,y|\boldsymbol{\vartheta}) S_A(\max(x,y))\, dx\, dy \\
&\leq \tilde{d}_0 \int f_A(x)\, dx + \frac{2\,\tilde{d}_1 K_1}{(1-\varrho_1)^2} \iint \mathrm{I}(x \leq y) f_A(y)\, dx\, dy \\
&\quad + \frac{(K_2+K_1^2)\tilde{d}_2}{(1-\varrho_1)^2} \iint S_A(\max(x,y))\, dx\, dy \\
&= \tilde{d}_0 + \frac{2\,\tilde{d}_1 K_1}{(1-\varrho_1)^2} EA + \frac{(K_2+K_1^2)\tilde{d}_2}{(1-\varrho_1)^2} EA^2 < \infty.
\end{aligned}$$

It remains to proof that the inequality (4.12) is fulfilled for all $\tilde{\boldsymbol{\vartheta}} \in U(\boldsymbol{\vartheta}, \delta_{\boldsymbol{\vartheta}})$ for some $\delta_{\boldsymbol{\vartheta}}$. Let $\delta_{\boldsymbol{\vartheta}} = 1-\varrho_1$, then for all $\tilde{\varrho}_1$ with $|\tilde{\varrho}_1 - \varrho_1| < \delta_{\boldsymbol{\vartheta}}$ it holds that $(1-\tilde{\varrho}_1)^{-2} \leq (1-\varrho_1)^{-2}$. Using the lemmata 11, 12 and 13, respectively, and (A.39) one can therefore conclude that

$$\begin{aligned}
\Big|\frac{\partial^2 \ln f_X^*(x,x,0,0|\tilde{\boldsymbol{\vartheta}})}{\partial\vartheta_k \partial\vartheta_l}\Big| &\leq \Big|\frac{\partial^2 \ln g_{0,1}(x,x|\tilde{\boldsymbol{\vartheta}})}{\partial\vartheta_k \partial\vartheta_l}\Big| \leq \tilde{M}_0(x,x) = M(x,x,0,0|\boldsymbol{\vartheta})\ , \\
\Big|\frac{\partial^2 \ln f_X^*(x,y,1,0|\tilde{\boldsymbol{\vartheta}})}{\partial\vartheta_k \partial\vartheta_l}\Big| &\leq \Big|\frac{\partial^2 \ln g_{1,1}(x,y|\tilde{\boldsymbol{\vartheta}})}{\partial\vartheta_k \partial\vartheta_l}\Big| \\
&\leq \frac{\tilde{M}_1(x,y)}{(1-\tilde{\varrho}_1)^2} \leq \frac{\tilde{M}_1(x,y)}{(1-\varrho_1)^2} = M(x,y,1,0|\boldsymbol{\vartheta}) \text{ for } x \leq y, \text{ and} \\
\Big|\frac{\partial^2 \ln f_X^*(x,y,1,1|\tilde{\boldsymbol{\vartheta}})}{\partial\vartheta_k \partial\vartheta_l}\Big| &\leq \Big|\frac{\partial^2 \ln g_{2,1}(x,y|\tilde{\boldsymbol{\vartheta}})}{\partial\vartheta_k \partial\vartheta_l}\Big| \\
&\leq \frac{\tilde{M}_2(x,y)}{(1-\tilde{\varrho}_1)^2} \leq \frac{\tilde{M}_2(x,y)}{(1-\varrho_1)^2} = M(x,y,1,1|\boldsymbol{\vartheta})
\end{aligned}$$

for all $\tilde{\boldsymbol{\vartheta}} \in U(\boldsymbol{\vartheta}, \delta_{\boldsymbol{\vartheta}})$. If $U(\boldsymbol{\vartheta}, \delta_{\boldsymbol{\vartheta}}) \not\subset \hat{\Theta}^\circ$ then obviously a $\tilde{\delta}_{\boldsymbol{\vartheta}} < \delta_{\boldsymbol{\vartheta}}$ exists such that $U(\boldsymbol{\vartheta}, \tilde{\delta}_{\boldsymbol{\vartheta}}) \subset \hat{\Theta}^\circ$ and all inequalities are still valid if $\tilde{\boldsymbol{\vartheta}} \in U(\boldsymbol{\vartheta}, \tilde{\delta}_{\boldsymbol{\vartheta}})$. □

Now the proof of asymptotic normality of maximum likelihood estimates in the bivariate correlated frailty model is given. It is based on theorem 2.

Theorem 6. *Let $\mathfrak{H} = \{H(x|\boldsymbol{\lambda}) : \boldsymbol{\lambda} \in \Lambda\}$ be a family of cumulative hazard functions that fulfills assumptions 2 and 3. Assume that $H(x|\boldsymbol{\lambda}_1) = H(x|\boldsymbol{\lambda}_2)$ for all x in an arbitrary non-degenerated interval I and for $\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2 \in \Lambda$ implies $\boldsymbol{\lambda}_1 = \boldsymbol{\lambda}_2$. Let Λ be a compact subset of $\mathbb{R}^{K_u}$. The parameter set $\hat{\Theta}$ is defined by*

$$\hat{\Theta} = \{(\sigma, \varrho_1, \boldsymbol{\lambda}')' : m_s \leq \sigma \leq M_s, m_r \leq \varrho_1 \leq 1, \boldsymbol{\lambda} \in \Lambda\}$$

where m_s, M_s and m_r are constants with $0 < m_s < M_s < \infty$ and $0 < m_r < 1$. Let (T_1, T_2) be a vector of two survival times with marginal survival function:

$$S_L(x|\boldsymbol{\vartheta}) = P(T_i > x) = \left[1 + \sigma^2 H(x|\boldsymbol{\lambda})\right]^{-1/\sigma^2} \quad , i = 1, 2,$$

and bivariate survival function:

$$\begin{aligned} S_{L*}(x, y|\boldsymbol{\vartheta}) &= P(T_1 > x, T_2 > y) \\ &= \left[S_L(x|\boldsymbol{\vartheta}) S_L(y|\boldsymbol{\vartheta})\right]^{1-\varrho_1} \left[S_L^{-\sigma^2}(x|\boldsymbol{\vartheta}) + S_L^{-\sigma^2}(y|\boldsymbol{\vartheta}) - 1\right]^{-\frac{\varrho_1}{\sigma^2}}. \end{aligned}$$

Denote $\mathfrak{S} = \{S_L(x|\boldsymbol{\vartheta}) : \boldsymbol{\vartheta} \in \hat{\Theta}\}$ and let assumptions 4(a), 4(b), 7 and 8 be fulfilled.
Let A be a non-negative censoring time with survival function S_A and density function f_A. A is assumed to be independent of (T_1, T_2) and it fulfills assumption 6. Furthermore, it is assumed that an interval I with $f_A(x) > 0$ and $\mu(x|\boldsymbol{\lambda}) > 0$ for all $x \in I$ and all $\boldsymbol{\lambda} \in \Lambda$ exists.
The random vector $(X_1, X_2, \Delta_1, \Delta_2)$ is defined by

$$X_i = \min(T_i, A) \quad \text{and} \quad \Delta_i = \mathrm{I}(T_i \leq A), \quad i = 1, 2.$$

A σ-finite measure $\tilde{\mu}$ on the sample space of $(X_1, X_2, \Delta_1, \Delta_2)$ exists such that the density function of the vector $(X_1, X_2, \Delta_1, \Delta_2)$ with respect to $\tilde{\mu}$ can be written as:

$$\begin{aligned} f_X^*(x_1, x_2, \delta_1, \delta_2|\boldsymbol{\vartheta}) = {} & S_{L*}(x_1, x_2|\boldsymbol{\vartheta}) f_A(x_1) \,\mathrm{I}(x_1 = x_2) \,\mathrm{I}(\delta_1 = 0) \,\mathrm{I}(\delta_2 = 0) \\ & - \frac{\partial S_{L*}(x_1, x_2|\boldsymbol{\vartheta})}{\partial x_2} f_A(x_1) \,\mathrm{I}(x_2 \leq x_1) \,\mathrm{I}(\delta_1 = 0) \,\mathrm{I}(\delta_2 = 1) \\ & - \frac{\partial S_{L*}(x_1, x_2|\boldsymbol{\vartheta})}{\partial x_1} f_A(x_2) \,\mathrm{I}(x_1 \leq x_2) \,\mathrm{I}(\delta_1 = 1) \,\mathrm{I}(\delta_2 = 0) \\ & + \frac{\partial^2 S_{L*}(x_1, x_2|\boldsymbol{\vartheta})}{\partial x_1 \partial x_2} S_A(\max(x_1, x_2)) \,\mathrm{I}(\delta_1 = 1) \,\mathrm{I}(\delta_2 = 1). \end{aligned}$$

Let $(x_1^i, x_2^i, \delta_1^i, \delta_2^i)$ *denote the* i^{th} *observation. The maximum likelihood estimate* $\hat{\boldsymbol{\vartheta}}_n$ *of the parameter vector* $\boldsymbol{\vartheta}_0 = (\sigma, \varrho_1, \boldsymbol{\lambda}')'$ *that is derived by maximizing the likelihood function*

$$L^*(\mathbf{x}_1, \mathbf{x}_2, \boldsymbol{\delta}_1, \boldsymbol{\delta}_2 | \boldsymbol{\vartheta}) = \prod_{i=1}^{n} f_X^*(x_1^i, x_2^i, \delta_1^i, \delta_2^i | \boldsymbol{\vartheta})$$

is asymptotically normal for all $\boldsymbol{\vartheta}_0 \in \hat{\Theta}^\circ$.
That means $\sqrt{n}(\hat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}_0) \xrightarrow{\mathcal{L}} N(0, I(\boldsymbol{\vartheta_0})^{-1})$ *where Fisher's information matrix* $I(\boldsymbol{\vartheta_0})$ *is given in corollary 4.*

Proof. The measure $\tilde{\mu}$ is defined in (3.8). The formula of the density function f_X^* is derived in lemma 5. The consistency of the maximum likelihood estimate is a direct consequence of theorem 5. Now the conditions of theorem 2 are verified.

Assumption (A2) follows from the representation of f_X^*, S_{L*}, S_L and assumption 7.

Assumption (B2) is a consequence of corollary 5.

Assumption (C2) was proven in proposition 5.

Assumption (D2) follows from assumption 8. □

4.4 Concluding remarks

The assumptions made in sections 4.1.1 and 4.3.1 are technical. They might be relaxed in the future by replacing the constant bounds with some measurable and integrable functions which have to fulfill additional conditions. It is easy to verify that all assumptions are fulfilled for

$$\mathfrak{H} = \{H(x) = \tfrac{a}{b}\left(e^{bx} - 1\right), m_a \leq a \leq M_a, m_b \leq b \leq M_b\}$$

as the family of cumulative baseline hazard functions. This corresponds to the Gompertz-parameterization, which has been used in applications (Yashin et al., 1994).
The proof of asymptotic normality of maximum likelihood estimates in the four-dimensional correlated frailty model is not given in this work. It is one of the many future tasks to be solved. In principle, the way of proving asymptotical normality in the four-dimensional model is not different from that of the bivariate model but it requires many more technical consideration. As one can verify in Appendix A.1 all functions which are included in the density function of current health status data in the four-dimensional model

are smooth and all derivatives with respect to the parameters exist and can be calculated. To give the reader the chance of understanding the main principle and to avoid getting lost in technical details the main reason that a proof of asymptotic normality is given only for the bivariate model here. Furthermore, such a long and technical proof would destroy the conception of this work, which intends to present both the theory as well as the application of the model.
Chapter 5 gives the results of a simulation study that empirically investigates the asymptotic behavior of maximum likelihood estimates in the four-dimensional correlated frailty model when estimation is based on current health status data. These results confirm the conjecture that asymptotic normality holds in the four-dimensional model as well.
Parner (1996a) also gives a proof of asymptotic normality in the correlated gamma frailty model, but he uses the counting process approach and shows asymptotic normality for a non-parametric maximum likelihood estimate that is based on the EM-Algorithm. So his model and his technical methods are different from those presented here.

Chapter 5

Simulation of the distribution of MLEs in the four-dimensional correlated gamma frailty model

In this chapter the results of a simulation study are presented. The empirical distribution of maximum likelihood estimates in the four-dimensional correlated gamma frailty model was investigated with the help of repeated simulations of data sets. This was done for the case where the current health status is observed (see section 2.2.1). The simulations were carried out for different sample sizes, different distributions of age at the time of observation and using the fully parametric and the partly semi-parametric estimation strategies, respectively (see section 2.3).

5.1 Mechanism of simulation

The simulation of uncensored data sets, i.e. realizations of the random vector $(T_{10}, T_1, T_{20}, T_2)$, was done as the first step. The values of the parameters of the frailty distribution were chosen as $\sigma_0 = 1$, $\sigma = 1$, $\varrho_0 = 0.7$, $\varrho_1 = 0.6$, $\varrho = 0.4$ and $\varrho^* = 0.3$ for all simulations. According to the definition of the model (see definition 9) the values of $(T_{10}, T_1, T_{20}, T_2)$ were calculated using

$$S_0(x) = [1 + \sigma_0^2 \frac{a_0}{b_0}(e^{b_0 x} - 1)]^{-1/\sigma_0^2} \quad \text{with } a_0 = 0.002 \text{ and } b_0 = 0.09,$$

and

$$S_L(x) = [1 + \sigma^2 \frac{a}{b}(e^{bx} - 1)]^{-1/\sigma^2} \quad \text{with } a = 0.0005 \text{ and } b = 0.06$$

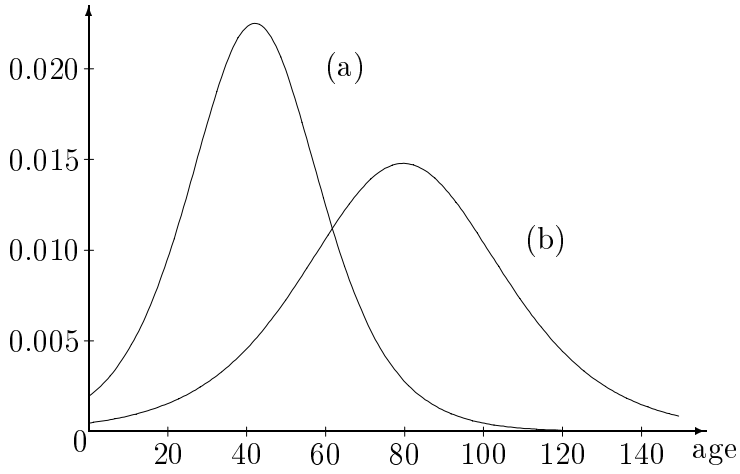


Figure 5.1: Density function of the age at onset of disease (a) and of the life span (b).

as the univariate survival functions. The expected value for the life span is then 80.47 and the standard deviation is 28.99. The expected value for the age at onset of disease resulting from the described choice of the parameters is 43.26 (standard deviation 18.62). This would correspond to a disease with a very early onset. But, instead of thinking of age at onset of disease, one might also consider age at the beginning of menopause or of any other condition that remains until the end of life. To avoid confusion the terms age at onset of disease or time of staying healthy are used throughout this chapter. Figure 5.1 shows the density functions of the age at onset of disease and of the life span.
The next step was the transformation of the uncensored data sets to data sets of restricted observations. After simulating A, the age at the time of observation, current health status data (X_1, X_2, C_1, C_2) was calculated as:

$$C_i = \begin{cases} 0 & \text{if } T_{i0} > A \text{ and } T_i > A \text{ (twin is healthy)}, \\ 1 & T_{i0} \leq A \text{ and } T_i > A \text{ (twin is ill)}, \\ 2 & T_i \leq A \text{ (twin is deceased)}, \end{cases}$$

and

$$X_i = \min(T_i, A) \qquad i = 1, 2.$$

The age at the time of observation was assumed to be uniformly distributed over different intervals. That means it is assumed that in a certain time interval the number of twin births is uniformly distributed. In other words

Table 5.1: Mean proportion of individuals in the health states "healthy", "ill" and "deceased" for different choices of the distribution of the age at the time of observation (in %)

A uniformly distributed over	healthy	ill	deceased
$[40, 100]$	12.9	49.0	38.1
$[30, 110]$	17.6	42.9	39.6
$[0, 120]$	34.4	31.4	34.2
$[0, 80]$	51.4	34.8	13.8
$[0, 250]$	16.5	15.6	67.9

The standard deviation for all numbers is smaller than 0.5, results are derived by the help of 300 simulated data sets per case.

different birth cohorts of twins are observed at a fixed point in time and the number of twin pairs in each birth cohort is more or less the same. The intervals $[40, 100]$, $[30, 110]$, $[0, 120]$, $[0, 80]$ and $[0, 250]$ were used.
Table 5.1 gives the mean proportions of healthy, ill and deceased individuals for the different choices of the distribution of A. If the age at the time of observation is distributed over $[40, 100]$ or $[30, 110]$ then most individuals are ill or deceased. For $[0, 120]$ about one third of all individuals is in each of the health states. Observing younger people (interval $[0, 80]$) yields a large proportion of healthy individuals. If A is distributed over $[0, 250]$ then naturally most people are deceased. Remember that for a deceased person A is the age that he or she would have been at the time of observation.
The sample size for the single data sets was 1000, 5000 or 10,000 pairs. Maximum likelihood estimates of the parameters were calculated for each data set using the likelihood function given in (2.20) and the methods described in Appendix C.1.1. In the fully parametric estimation approach estimates of the full parameter vector

$$(\sigma_0, \sigma, \varrho_0, \varrho_1, \varrho, \varrho^*, a_0, b_0, a_1, b_1)$$

were derived. In the partly semi-parametric approach the Kaplan-Meier Estimator was calculated for S_L using X_i, $i = 1, 2$, as the observed censored time and $\mathrm{I}(C_i = 2)$ as the censoring indicator. The formula for the Kaplan-Meier Estimator is given in (1.1). The values of the Kaplan-Meier Estimator were substituted in the formula for the likelihood function and the maximum likelihood estimates of the remaining parameters $(\sigma_0, \sigma, \varrho_0, \varrho_1, \varrho, \varrho^*, a_0, b_0)$ were calculated.

Table 5.2: Mean value and standard deviation for the maximum likelihood estimate of σ_0, value used in simulation: $\sigma_0 = 1$.

sample size	method	interval of A [40,100]	[30,110]	[0,120]	[0,80]	[0,250]
1000	para	1.041 (.323)	1.032 (.254)	1.010 (.190)	1.019 (.172)	1.019 (.269)
	semi	1.037 (.383)	1.028 (.238)	1.007 (.192)	0.988 (.188)	1.033 (.252)
5000	para	1.021 (.195)	0.991 (.109)	0.992 (.091)	0.999 (.086)	1.002 (.123)
	semi	1.023 (.174)	0.999 (.125)	0.995 (.091)	0.996 (.095)	1.007 (.126)
10,000	para	1.012 (.123)	0.999 (.084)	0.992 (.060)	1.001 (.064)	0.997 (.090)
	semi	1.009 (.114)	1.010 (.084)	0.996 (.062)	0.990 (.065)	0.990 (.087)

para = fully parametric , semi = partly semi-parametric

The procedure of simulation and estimation was repeated 250 times. As a consequence it is possible to make conclusions about the empirical distribution of the maximum likelihood estimates. It is also possible to evaluate the influence of the sample size, the distribution of the age at the time of observation and of the estimation strategy on the properties of the estimates.

5.2 Descriptive statistics of the estimates

Tables 5.2 - 5.11 show the mean value and the standard deviation of the estimates for the different samples sizes, distributions of A, and estimation approaches.

$\boldsymbol{\sigma_0}$:
σ_0^2 denotes the variance of the frailty that represents risk factors which influence the age at onset of disease.
The estimation approach has no visible influence on the standard deviation or the bias of the maximum likelihood estimate of the parameter σ_0 (see table 5.2). This is not surprising since the survival function S_0 is estimated parametrically for both approaches. In general the bias of the estimates is small.
The standard deviation is the smallest one in the cases that A is distributed over $[0, 80]$ or $[0, 120]$. These are the cases with the highest proportion of healthy individuals (see table 5.1). The main reason that the standard deviation is smallest for the intervals mentioned is that these intervals are the most "informative" for the distribution of the age at onset of disease (see

Table 5.3: Mean value and standard deviation for the maximum likelihood estimate of σ, value used in simulation: $\sigma = 1$.

sample size	method	interval of A				
		[40,100]	[30,110]	[0,120]	[0,80]	[0,250]
1000	para	1.006 (.146)	0.994 (.119)	0.994 (.103)	1.079 (.332)	0.997 (.060)
	semi	1.070 (.380)	1.059 (.346)	1.063 (.316)	1.272 (.921)	1.012 (.214)
5000	para	1.004 (.069)	1.001 (.053)	0.995 (.052)	1.014 (.190)	0.999 (.025)
	semi	0.994 (.198)	1.011 (.172)	0.998 (.152)	1.072 (.287)	1.002 (.090)
10,000	para	1.004 (.048)	1.001 (.039)	0.998 (.037)	1.007 (.141)	1.000 (.017)
	semi	0.991 (.153)	1.014 (.129)	1.008 (.122)	1.062 (.268)	1.001 (.060)

para = fully parametric , semi = partly semi-parametric

figure 5.1(a)). Consequently the standard deviation must be the highest for the interval $[40, 100]$ which is confirmed by the results.
If one wants to have an estimate for σ_0 with an acceptably small standard deviation for small sample sizes then one has to chose an appropriate range of birth cohorts.
As expected the standard deviation as well as the bias decrease when the sample size increases.

$\boldsymbol{\sigma}$:
The parameter σ denotes the standard deviation of the frailty that summarizes risks which change individual chances of surviving.
The estimation approach has a large influence on the magnitude of the standard deviation of the MLE for this parameter. The standard deviation of the estimate resulting from a fully parametric estimation approach is about one third of the standard deviation of the estimate resulting from a partly semi-parametric estimation approach (see table 5.3). This is due to the fact that in the semi-parametric approach the information about this parameter comes only from the four-dimensional frailty distribution whereas in the fully parametric approach additional information about σ is available since the parameter σ also occurs in the representation of S_L.
The smallest standard deviation yields the case where A is distributed over the interval $[0, 250]$. For this case, on the one hand, most people are deceased and, on the other hand, this interval is the most "informative" for S_L (see figure 5.1(b)). On the contrary the standard deviation is the highest for the interval $[0, 80]$ where the proportion of deceased individuals is the smallest

Table 5.4: Mean value and standard deviation for the maximum likelihood estimate of ϱ_0, value used in simulation: $\varrho_0 = 0.7$.

sample size	method	interval of A				
		[40,100]	[30,110]	[0,120]	[0,80]	[0,250]
1000	para	0.727 (.185)	0.711 (.170)	0.686 (.172)	0.714 (.159)	0.673 (.231)
	semi	0.730 (.193)	0.709 (.165)	0.700 (.175)	0.720 (.175)	0.687 (.215)
5000	para	0.705 (.113)	0.704 (.087)	0.699 (.090)	0.696 (.077)	0.699 (.116)
	semi	0.704 (.105)	0.707 (.090)	0.705 (.088)	0.704 (.083)	0.705 (.119)
10,000	para	0.701 (.078)	0.701 (.067)	0.700 (.064)	0.699 (.054)	0.705 (.091)
	semi	0.700 (.072)	0.692 (.066)	0.695 (.064)	0.701 (.052)	0.700 (.087)

para = fully parametric , semi = partly semi-parametric

(see table 5.1). In addition the interval $[0, 80]$ does not even include the mean value corresponding to S_L.
A sample size of 1000 pairs is not sufficient to get a standard deviation that is smaller than one third of the mean value for the MLE of σ with the semi-parametric approach (exception A is distributed over $[0, 250]$). The situation can be improved if an adequate parametric specification is used in a fully parametric approach.
An increasing sample size generally yields a smaller standard deviation and a smaller bias.

ϱ_0 :
ϱ_0 denotes the correlation of the frailty variables influencing the transition from healthy to ill. This parameter is the most important one for the genetic analysis of the risk factors that influence this transition.
Looking at table 5.4 the most surprising fact is that the distribution of A has no pronounced effect on either the standard deviation or the bias of the MLE of ϱ_0. The estimation approach has also almost no influence. This means that the estimate of ϱ_0 is quite robust against the choice of the observed birth cohorts and the estimation approach. This would be an advantage for applications. It has to be noted here that for all investigated distributions of A (with the exception of the interval $[0, 250]$) the proportion of ill individuals is quite high (see table 5.1). This situation is obviously not realistic even for common diseases but it might be for other characteristics, such as menopause. The highest standard deviation yields the case where A is distributed over $[0, 250]$. This seems reasonable since the proportion of ill individuals is the

Table 5.5: Mean value and standard deviation for the maximum likelihood estimate of ϱ_1, value used in simulation: $\varrho_1 = 0.6$.

sample size	method	interval of A				
		[40,100]	[30,110]	[0,120]	[0,80]	[0,250]
1000	para	0.615 (.137)	0.608 (.125)	0.612 (.110)	0.635 (.259)	0.609 (.073)
	semi	0.657 (.247)	0.640 (.222)	0.643 (.199)	0.617 (.289)	0.622 (.130)
5000	para	0.605 (.065)	0.602 (.054)	0.602 (.055)	0.629 (.191)	0.602 (.033)
	semi	0.653 (.172)	0.622 (.136)	0.618 (.115)	0.629 (.238)	0.604 (.059)
10,000	para	0.603 (.045)	0.599 (.038)	0.601 (.039)	0.621 (.151)	0.599 (.022)
	semi	0.633 (.135)	0.608 (.102)	0.606 (.089)	0.623 (.226)	0.602 (.040)

para = fully parametric , semi = partly semi-parametric

smallest for this case.
In general the standard deviation of the estimate is already relatively small for a sample size of 1000 pairs.
As one would expect, the standard deviation and the bias decrease when the sample size is increased.

ϱ_1 :
The parameter ϱ_1 denotes the correlation of the frailty variables that describe the risk factors which influence the survival of an individual. This parameter is the one that is used for the genetic analysis of these risk factors.
It can be seen in table 5.5 that the situation of the MLE of ϱ_1 is completely different from that of the MLE of ϱ_0. The estimation approach and the distribution of A have an influence on the standard deviation and the bias of the estimate.
The standard deviation of the estimate resulting from the fully parametric approach is about one half of the standard deviation of the estimate resulting from the partly semi-parametric approach. A possible explanation of this effect is that the quality of the estimate of the parameter σ influences the performance of the estimate of ϱ_1 since these two parameters describe the bivariate distribution of the life spans of the twins in a pair.
The bias for the estimate in the fully parametric approach is also smaller. The smallest standard deviation yields the case where A is distributed over the interval $[0, 250]$. That is the case with the highest proportion of deceased individuals. The case with the smallest proportion of deceased individuals is the case with the highest standard deviation.

Table 5.6: Mean value and standard deviation for the maximum likelihood estimate of ϱ, value used in simulation: $\varrho = 0.4$.

sample size	method	interval of A				
		[40,100]	[30,110]	[0,120]	[0,80]	[0,250]
1000	para	0.420 (.226)	0.406 (.235)	0.407 (.235)	0.396 (.251)	0.388 (.248)
	semi	0.404 (.235)	0.379 (.247)	0.426 (.256)	0.418 (.270)	0.368 (.282)
5000	para	0.441 (.163)	0.428 (.154)	0.428 (.171)	0.434 (.162)	0.396 (.203)
	semi	0.427 (.165)	0.433 (.161)	0.435 (.188)	0.444 (.173)	0.418 (.203)
10,000	para	0.417 (.128)	0.435 (.132)	0.423 (.145)	0.433 (.138)	0.399 (.136)
	semi	0.423 (.152)	0.439 (.146)	0.437 (.153)	0.445 (.163)	0.431 (.165)

para = fully parametric , semi = partly semi-parametric

It is necessary to have a reasonable number of deceased individuals in order to get an acceptable estimate of the parameter ϱ_1 for a sample size of 1000 pairs. The standard deviation and the bias decrease when the sample size increases.

$\boldsymbol{\varrho}$:
The parameter ϱ denotes the correlation of the frailty representing susceptibility to disease of one twin and of the frailty representing susceptibility to death of the same twin. This parameter is not identifiable from univariate data without making parametric assumptions on the underlying hazard rates. The situation is comparable to that of dependent competing risks where it is not possible to identify the distribution of two survival times by observing their minimum (Prakasa Rao, 1992, section 7.4). While observing current health status you do not observe the minimum of the time of staying healthy and the life span but another function of these two dependent survival times (see also page 30).
The parameter ϱ is a parameter that represents the dependence of two variables belonging to one individual but that is estimated from data on two individuals. This characteristic is reflected by the properties of the MLE (see table 5.6).
The MLE of ϱ has a quite large standard deviation even for a sample size of 10,000 pairs. The bias is also quite high for this large sample size. That means that the consistency of the estimate which was proven in Chapter 4 holds only for higher sample sizes than the ones considered here.
The choice of the distribution of the age at the time of observation has

Table 5.7: Mean value and standard deviation for the maximum likelihood estimate of ϱ^*, value used in simulation: $\varrho^* = 0.3$.

sample size	method	interval of A				
		[40,100]	[30,110]	[0,120]	[0,80]	[0,250]
1000	para	0.301 (.181)	0.304 (.193)	0.302 (.197)	0.296 (.207)	0.296 (.226)
	semi	0.301 (.196)	0.291 (.199)	0.316 (.225)	0.312 (.232)	0.271 (.245)
5000	para	0.302 (.091)	0.303 (.096)	0.294 (.114)	0.294 (.112)	0.280 (.160)
	semi	0.306 (.093)	0.302 (.104)	0.310 (.120)	0.307 (.105)	0.297 (.160)
10,000	para	0.305 (.066)	0.299 (.072)	0.293 (.087)	0.297 (.071)	0.286 (.118)
	semi	0.304 (.076)	0.308 (.072)	0.299 (.084)	0.309 (.085)	0.305 (.115)

para = fully parametric , semi = partly semi-parametric

no pronounced effect on the standard deviation. The same holds for the estimation method. The standard deviation and the bias resulting from the fully parametric approach is a little bit smaller than that resulting from a partly semi-parametric approach. But it is only a small difference.
The parameter ϱ is the one which is most difficult to estimate. Even with high sample sizes the properties of the estimate are bad. In applications one should be very careful with the interpretation of the estimate for this parameter.

ϱ^* :
ϱ^* denotes the correlation of the frailty corresponding to the transition from healthy to ill of one twin and of the frailty corresponding to the transition from healthy to deceased of the other twin. Contrary to the parameter ϱ, the parameter ϱ^* describes the association of two variables belonging to two different individuals.
The estimation approach has almost no effect on the standard deviation and the bias of the MLE of ϱ^* (see table 5.7).
The distribution of age at the time of observation has no large influence on bias or standard deviation, either - especially for a sample size of 1000 pairs. The smallest standard deviation yields the case where A is distributed over the interval $[40, 100]$. This is the case with highest proportion of pairs where one twin is ill and the other is deceased. The case with the lowest proportion of these pairs, the interval $[0, 250]$, leads to the largest standard deviation.
The standard deviation is falling with a rising sample size. For a sample size of 1000 pairs it is still too high to make meaningful interpretations of the

Table 5.8: Mean value and standard deviation for the maximum likelihood estimate of $a_0 \times 100$, value used in simulation: $a_0 = 0.002$.

sample size	method	interval of A [40,100]	[30,110]	[0,120]	[0,80]	[0,250]
1000	para	0.234 (.171)	0.206 (.109)	0.202 (.063)	0.196 (.052)	0.195 (.081)
	semi	0.243 (.182)	0.200 (.107)	0.201 (.065)	0.206 (.055)	0.194 (.088)
5000	para	0.206 (.105)	0.203 (.050)	0.201 (.029)	0.199 (.024)	0.199 (.041)
	semi	0.202 (.101)	0.202 (.058)	0.201 (.030)	0.200 (.025)	0.200 (.040)
10,000	para	0.200 (.072)	0.200 (.041)	0.202 (.021)	0.200 (.018)	0.201 (.029)
	semi	0.198 (.069)	0.193 (.041)	0.200 (.021)	0.201 (.017)	0.201 (.029)

para = fully parametric , semi = partly semi-parametric

estimate of the parameter. The bias, however, is already small for a sample size of 1000 pairs.

a_0 and b_0 :
The parameters a_0, b_0 and σ_0 describe the shape of the survival function S_0. The estimation approach has no pronounced effect on the properties of the MLEs of a_0 and b_0 (see tables 5.8 and 5.9). This seems natural since the survival function S_0 is estimated parametrically for both approaches.
As already discussed for the parameter σ_0, the distribution of A influences the properties of the estimates in such a way that the smallest standard deviation and bias are derived for the most "informative" interval for S_0 - the interval $[0, 80]$. On the contrary the less "informative" interval $[40, 100]$ gives the largest standard deviation and bias.
As expected both standard deviation and bias decrease when the sample size is increased. To get acceptable estimates of the parameters a_0 and b_0 one has to choose either an adequate interval or a relatively high sample size.

a_1 and b_1 :
The parameters a_1, b_1 and σ are used for the parametric specification of the survival function S_L. Of course it is only possible to get MLEs of a_1 and b_1 with the fully parametric estimation approach (see tables 5.10 and 5.11).
The smallest standard deviation is derived in the case where A is distributed over the interval $[0, 250]$. The proportion of deceased individuals is the highest for this interval and it is the most "informative" for S_L (see figure 5.1(b)).The largest standard deviation consequently yields the interval $[0, 80]$.

Table 5.9: Mean value and standard deviation for the maximum likelihood estimate of $b_0 \times 10$, value used in simulation: $b_0 = 0.09$.

sample size	method	interval of A [40,100]	[30,110]	[0,120]	[0,80]	[0,250]
1000	para	1.019 (.486)	0.967 (.314)	0.923 (.158)	0.928 (.135)	0.956 (.249)
	semi	1.056 (.681)	0.966 (.288)	0.925 (.170)	0.906 (.142)	0.967 (.235)
5000	para	0.956 (.272)	0.904 (.106)	0.901 (.074)	0.906 (.061)	0.910 (.099)
	semi	0.949 (.220)	0.913 (.124)	0.903 (.071)	0.906 (.067)	0.911 (.102)
10,000	para	0.925 (.143)	0.908 (.084)	0.898 (.048)	0.904 (.046)	0.902 (.071)
	semi	0.925 (.137)	0.922 (.086)	0.903 (.048)	0.900 (.044)	0.900 (.067)

para = fully parametric , semi = partly semi-parametric

Table 5.10: Mean value and standard deviation for the maximum likelihood estimate of $a_1 \times 1000$, value used in simulation: $a_1 = 0.0005$.

sample size	interval of A [40,100]	[30,110]	[0,120]	[0,80]	[0,250]
1000	0.506 (.075)	0.502 (.074)	0.503 (.070)	0.486 (.092)	0.509 (.072)
5000	0.499 (.036)	0.499 (.033)	0.502 (.036)	0.497 (.049)	0.501 (.030)
10,000	0.498 (.025)	0.500 (.023)	0.501 (.027)	0.500 (.035)	0.500 (.021)

Table 5.11: Mean value and standard deviation for the maximum likelihood estimate of $b_1 \times 10$, value used in simulation: $b_1 = 0.06$.

sample size	interval of A [40,100]	[30,110]	[0,120]	[0,80]	[0,250]
1000	0.602 (.039)	0.601 (.036)	0.601 (.032)	0.620 (.063)	0.599 (.029)
5000	0.602 (.019)	0.601 (.016)	0.600 (.017)	0.605 (.032)	0.600 (.012)
10,000	0.601 (.013)	0.600 (.012)	0.600 (.013)	0.602 (.024)	0.600 (.008)

The standard deviation decreases with an increasing sample size. In general for a sample size of 1000 pairs the standard deviation is already acceptable small. The bias is small for all situations.

5.3 Empirical distribution of the estimates

5.3.1 Graphical analysis

Figures C.1 - C.9 in Appendix C.2 show histograms for the empirical distribution of the MLEs in the four-dimensional correlated frailty model for different situations. The histograms confirm the empirical results discussed in section 5.2.
The behavior of the estimates of the parameter σ and ϱ_1 depends strongly on the estimation approach. For the other parameter the estimation approach has no significant influence. For all parameters a higher sample size leads to a decrease in the variance and a better approximation to the normal curve.
The interval on which the age at the time of observation is distributed has an impact on the distribution of most parameters. The only exception here are the parameters ϱ_0 and ϱ.
A sample size of 1000 pairs is only sufficient for an acceptably well-distributed MLE if A is distributed suitably (parameter σ_0, a_0, b_0, a_1, b_1) and an adequate parametric specification is chosen for S_L (parameter σ, ϱ_1). For all other parameters higher sample sizes are necessary to get useful estimates.
The estimate of the parameter ϱ is the one with the worst properties. The approximation to the normal curve is quite good for a sample size of 10,000 pairs but the variance is still very large for this high sample size.
For most parameters the approximation to the normal curve is good for high sample sizes. This is an indication of the asymptotic normality of the maximum likelihood estimates.

5.3.2 Test of normality based on skewness and kurtosis

To make statements about the normality of the distribution of the maximum likelihood estimates in the four-dimensional frailty model, tests based on skewness and kurtosis were used. These tests are explained in section 1.4.2.
In table C.1 (see page 164) the empirical estimate of the skewness of the maximum likelihood estimates is given for the different distributions of A and for the sample sizes 1000 and 10,000 pairs. The skewness was calculated using formula (1.19).
The value of the skewness was used to test the null hypothesis

$$H_0 : \text{The MLE is normally distributed.}$$

versus the two-sided alternative

$$H_1 : \text{The MLE is non-normally distributed with } \beta_1 \neq 0.$$

In table C.1 the asterisk indicates the cases where the null hypothesis is rejected with a 0.05 level of significance.
Except for the case where A is distributed over the interval $[40, 100]$ the null hypothesis of normality is rejected less for the higher sample size of 10,000 pairs than for a sample size of 1000 pairs. That means that a higher sample size leads to a less skewed distribution of the estimate.
Except for the parameter σ the estimation approach has no visible effect on the number of rejections. For the parameter σ the null hypothesis is more often rejected for the semi-parametric estimation approach. The reason for this is the same as for the smaller bias and standard deviation in the fully parametric case. While estimating σ in this approach one gets additional information about the parameter from the estimation of S_L.
The parameters with the most skewed distribution are σ_0, ϱ_1 and b_0. For all other parameters a sample size of at least 10,000 pairs is sufficient to accept normality. But for the parameters σ_0, ϱ_1 and b_0 for this sample size the asymptotic normality still isn't reached. The exception to the last statement is the case where the age at time of observation is distributed on the shortest interval ($[40, 110]$). For this case a sample size of 10,000 pairs is not generally sufficient to accept normality.
Table C.2 (see page 165) shows the values of the kurtosis estimated using formula 1.19. The results are given again for the different distributions of A, the sample sizes 1000 and 10,000 pairs and for the two different estimation approaches.
Based on the kurtosis the null hypothesis

$$H_0 : \text{The MLE is normally distributed.}$$

was tested versus the two-sided alternative

$$H_1 : \text{The MLE is non-normally distributed with } \beta_2 \neq 3.$$

Rejection of the null hypothesis with a 0.05 level of significance is indicated by an asterisk.
The number of rejections of normality decreases when the sample size is increased. In general for a sample size of 1000 pairs normality is rejected for most parameters, whereas it is accepted for most parameters for a sample size of 10,000 pairs.

There is no pronounced effect of the estimation approach for all parameters with the exception of parameter ϱ_1. The null hypothesis of normality is more often rejected if the estimate of ϱ_1 is derived in a partly semi-parametric estimation approach. A possible explanation for this might be that the quality of the estimate of σ influences the properties of the estimate of ϱ_1.
The parameter for which normality is rejected most often is the parameter ϱ. Non-normality is also accepted quite often for the parameter b_0.
With regard to kurtosis non-normality also occurs most often for the case that the age at the time of observation is distributed over $[40, 100]$.

5.3.3 Asymptotical variance of the estimates

Let $f(x|\theta)$ be a density function depending on a one-dimensional parameter θ. From theorem 2 it follows that under certain smoothness conditions on f a constant $\alpha(\theta)$ exists such that for the maximum likelihood estimate $\hat{\theta}_n$ of θ it holds that $\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{\mathcal{L}} N(0, \alpha^2(\theta))$. From a heuristic point of view, the variance of $\sqrt{n}(\hat{\theta}_n - \theta)$ should be more or less constant for high values of the sample size n.
Table C.3 (see page 167) shows the empirical variances of $\sqrt{n}(\hat{\theta}_n - \theta)$ for the maximum likelihood estimates of the different parameters of the four-dimensional correlated frailty model. The behavior of the these values for a rising sample size is investigated. If, on the one hand, the empirical variance is relatively constant then this can be a hint that the asymptotic normality of the maximum likelihood estimate is valid. On the other hand, if it varies widely then this can be hint that the asymptotic distribution of the estimate does not follow the standard behavior. If the asymptotic normality of the maximum likelihood estimate was theoretically proven then a varying variance of $\sqrt{n}(\hat{\theta}_n - \theta)$ would mean that the asymptotic behavior is valid only for very high sample sizes.

$\boldsymbol{\sigma_0}$:
The values of the empirical variance of $\sqrt{n}(\hat{\sigma}_{0n} - \sigma_0)$ are given in the second row of table C.3. A more or less constant value for all sample sizes can only be found for the case where A is distributed over the interval $[0, 120]$. The values of the empirical variance are similar for $n = 5000$ and $n = 10{,}000$ if A is distributed over $[0, 80]$ and $[0, 250]$. This confirms the results of the tests for normality carried out in section 5.3.2. For $n = 1000$ asymptotic normality is not yet reached but for $n = 10{,}000$ the distribution of the MLE of σ_0 seems to approach the normal distribution better.

$\boldsymbol{\sigma}$:
The values of the empirical variance of $\sqrt{n}(\hat{\sigma}_n - \sigma)$ can be found in the

third row of table C.3. For the fully parametric estimation approach the values are relatively similar for all sample sizes except for the case that A is distributed over $[0, 80]$. For the semi-parametric approach the situation is worse. Looking at the values it is not surprising that the normality of $\hat{\sigma}_n$ was rejected more often for the semi-parametric estimation approach in section 5.3.2. In general the properties of the MLE of σ which were described on page 74 are confirmed. The especially bad behavior of the estimate when A is distributed over $[0, 80]$ is not surprising if one looks at figure C.2. The pattern of the empirical distribution is far from being normal even for the sample size $n = 5000$.

ϱ_0 :
The values of the empirical variance of $\sqrt{n}(\hat{\varrho}_{0n} - \varrho_0)$ are given in the fourth row of table C.3. For most cases the values for $n = 5000$ and $n = 10{,}000$ are quite similar, whereas the values for $n = 1000$ are different. Combining this with the results derived in section 5.3.2 one can conclude that asymptotic normality seems to hold for the MLE of ϱ_0 and that the asymptotical behavior seems to be valid at least for $n = 10{,}000$.

ϱ_1 :
The fifth row of table C.3 shows the empirical variance of $\sqrt{n}(\hat{\varrho}_{1n} - \varrho_1)$. The influence of the estimation approach on the properties of the estimate that have been discussed in section 5.2 are also obvious here. For the fully parametric approach the values are more or less constant for all sample sizes except for the case where A is distributed over the interval $[0, 80]$. For the semi-parametric approach something like constant values can only be found for the case where A is distributed over the interval $[0, 250]$. This confirms the results derived in section 5.3.2 exactly. The poor properties of the estimator when A is distributed over $[0, 80]$ are also reflected in figure C.4.

ϱ :
The values of the empirical variance of $\sqrt{n}(\hat{\varrho}_n - \varrho)$ can be found in the sixth row of table C.3. One cannot find any pattern of similar values here. That confirms the fact that the estimate of the parameter ϱ is the one with the worst properties and that even a sample size of 10,000 pairs is not sufficient to get an acceptable estimate (see also figure C.5).

ϱ^* :
The results for the parameter ϱ^* can be found in the seventh row of table C.3. There is no clear pattern for the parametric estimation approach, whereas for the semi-parametric estimation approach the values for $n = 5000$ and $n = 10{,}000$ are quite similar for the cases where A is distributed over $[30, 110]$,

$[0, 120]$ and $[0, 250]$. This confirms the results of section 5.3.2. A high sample size is necessary to get an approximately normally distributed estimate of ϱ^*.

a_0 and b_0 :
The values of the empirical variance of $\sqrt{n}(\hat{a}_{0n} - a_0)$ and $\sqrt{n}(\hat{b}_{0n} - b_0)$ are given in rows 8 and 9 of table C.3. For the cases where A is distributed over $[0, 80]$ and $[0, 120]$ the values are similar for all sample sizes. For most other cases this is true at least for $n = 5000$ and $n = 10{,}000$. This confirms the results of section 5.3.2. An approximately normal distribution can be excepted at least for high sample sizes.

a_1 and b_1 :
The values of the empirical variance of $\sqrt{n}(\hat{a}_{1n} - a_1)$ and $\sqrt{n}(\hat{b}_{1n} - b_1)$ can be found in rows 10 and 11 of table C.3. At least for the sample sizes $n = 5000$ and $n = 10{,}000$ these values are more or less constant. Again the results of section 5.3.2 are confirmed. An approximately normal distribution seems to be valid for high sample sizes.

5.4 Summary

In general the empirical results for the maximum likelihood estimates confirm expected properties of the estimates about consistency and asymptotic normality.
A higher sample size improves the properties of the estimates. In general, the proportion of healthy, ill and deceased individuals has a high influence on the behavior of the estimates. The estimation yields acceptable results for the parameter of the univariate survival function S_0 only if the age at the time of observation is distributed over a reasonable interval.
A suitable parametric choice for the survival function S_L can improve the behavior of the MLE of the parameters σ and ϱ_1. For all other parameters the semi-parametric estimation approach performs as well as the fully parametric approach.
The parameter ϱ seems to be the one with the worst properties. One should be careful when interpreting the estimates of this parameter.
The parameter ϱ_0 is the most important for the genetic analysis of the risk factors that influence susceptibility to disease. The estimator of this parameter seems to be quite robust, at least for the situations investigated here.
In general for a fixed sample size the distribution of a maximum likelihood estimate is unknown. For sufficiently smooth density functions asymptotic normality of MLEs is usually true. For the bivariate correlated gamma frailty model asymptotic normality of the maximum likelihood estimates was proven

in theorem 6. Asymptotic normality of the maximum likelihood estimates holds probably also in the four-dimensional model. On the one hand this could be supported partly by the simulation study described in this chapter. On the other hand one has to keep in mind that the four-dimensional model is much more complicated. The number of parameters is higher in the four-dimensional model. Consequently an approximate normal distribution of the maximum likelihood estimates might not yet be valid for the sample sizes that have been investigated here but for higher sample sizes.
Asymptotic normality of maximum likelihood estimates is an important theoretical property that justifies the construction of confidence intervals and tests. If, for example, one wants to construct a test for the null hypothesis $\varrho^* = 0$ then one has to calculate the (asymptotic) α-quantile of a test statistic. The results of the simulation study suggest that the distribution of the MLE of ϱ^* is not approximately normal for smaller sample sizes. The conclusion must then be that it is better to use other methods such as a bootstrap approach to calculate an α-quantile for a test. Similar conclusions can be made concerning the validity of asymptotic confidence intervals for the estimates for small or medium sample sizes.

Chapter 6

Application of the four-dimensional correlated frailty model to prostate cancer data

6.1 Data

The applicability of the four-dimensional correlated frailty model was tested with data on prostate cancer in male Swedish twins. The data set was created by merging the Swedish Twin Registry with the Swedish Cancer Registry and the Swedish Mortality Registry.
The twin pairs come from the "old cohort" in the Swedish Twin Registry, these are the birth cohorts from 1886 to 1925. The twins were tracked down by contacting all parishes in Sweden. The "old cohort" includes all like-sexed twin pairs where both twins responded to a questionnaire sent out in 1961. Therefore the data is left-truncated since both twins of a pair had to be alive in 1961 in order to be included. The determination of the zygosity of a pair was based on the responses to the question "During childhood, were you and your twin partner 'alike as two peas in a pod' or not more alike than other siblings in general?". This method was validated using blood samples and showed correct classification for 95% of the pairs (Cederlöf et al., 1961).
The cancer follow-up extended from 1958 to 1997. The determination of a diagnosis of cancer and of the vital status was based on the unique national registration number that is assigned to each Swedish citizen.
The day of birth and the zygosity is known for every twin pair. For individuals who died during the observation period (1/1/1962 - 20/12/2000), the

Table 6.1: Proportion of pairs with no, one, or two affected twins

zygosity	no twin with diagnosis	one twin with diagnosis	both twins with diagnosis	total
MZ	1412	200	33	1645
	85.8 %	12.2 %	2.0 %	100 %
DZ	2554	401	18	2973
	85.9 %	13.5 %	0.6 %	100 %

day of death is included in the data, as is the day of diagnosis for individuals with prostate cancer during the follow-up (1/1/1958 - 31/12/1997).
There is information on 9,236 individuals, i.e. 4,618 pairs of twins (1,645 monozygotic, 2,973 dizygotic). A cancer diagnosis was registered for 703 individuals (7.6%). The empirical mean of the observed ages at cancer diagnosis is 74.1 years (standard deviation 7.4 years).
Table 6.1 shows the proportion of pairs where a diagnosis of cancer was registered for no, one, or both twins for the different zygosities. The proportion of pairs with two affected twins is higher for monozygotic twins. This could be an indication of a genetic influence.
The data provide more information than the type of data described in section 2.2.1, since there is information about the time of diagnosis, which could be interpreted as time of onset of disease. The censored life span is used for every individual in the analysis and the age at onset of disease is used for people with a diagnosis of cancer. For individuals without a cancer diagnosis who were deceased at the end of the observation period it is assumed that their age at onset of disease is greater than their age at death. This means that here the risk of death and the risk of getting cancer are two competing risks. Finally, individuals who are still alive at the end of the observation period and who do not have a diagnosis of cancer contribute with the information that they can only have had a diagnosis before the beginning of the follow-up or will have one after the end of the follow-up.
Four different types of observation are possible for every individual:

1. A person died during the observation period and no cancer diagnosis is registered. Thus only the age at death is known for this person.

2. A person died during the observation period and a cancer diagnosis is registered. Then the age at diagnosis and the age at death are known.

3. A person is still alive at the end of the observation period and there is no diagnosis of cancer. Then one only knows the censored life span.

4. A person is still alive at the end of the observation period and there is a diagnosis of cancer. Then one knows the censored life span and the age at diagnosis.

Tables 6.2 and 6.3 show the cross-tabulation of these types of observation in a twin pair for monozygotic and dizygotic twins. In the majority of pairs both members are deceased and do not have a diagnosis of cancer.

Table 6.2: Proportions of different types of observations for MZ twins in %

twin 1 \ twin 2	1	2	3	4	total
1	56.5	4.5	8.9	0.5	70.5
2	3.6	1.6	0.9	0.2	6.2
3	9.8	1.1	10.6	0.5	22.1
4	0.5	0.1	0.5	0.1	1.3
total	70.4	7.3	20.9	1.4	100.0

1 = deceased without diagnosis of cancer, 2 = deceased with diagnosis of cancer,
3 = alive without diagnosis of cancer, 4 = alive with diagnosis of cancer

Table 6.3: Proportions of different types of observations for DZ twins in %

twin 1 \ twin 2	1	2	3	4	total
1	54.2	4.9	9.1	0.6	68.9
2	4.7	0.5	0.8	0.1	6.1
3	11.8	0.9	10.7	0.4	23.9
4	0.7	0.0	0.4	0.0	1.1
total	71.4	6.4	21.1	1.2	100.0

1 = deceased without diagnosis of cancer, 2 = deceased with diagnosis of cancer,
3 = alive without diagnosis of cancer, 4 = alive with diagnosis of cancer

6.2 Methods and results

The data are left truncated pairwise since both twins in a pair had to survive the year 1961. That means that the conditional probability to be included in the study for an individual also depends on the survival of the co-twin. The life spans of the twins in pair are not independent. Their common distribution is described by the bivariate survival function S_{L*} (see (2.13)).

For this reason the usual Kaplan-Meier-Estimator is not appropriate for the estimation of the survival function S_L. Thus the partly semi-parametric approach that is described in section 2.3 could not be used and a parametric specification had to be chosen for S_0 as well as for S_L.

$$S(x) = e^{-dx}[1 + s^2 H(x)]^{-1/s^2} \quad \text{with} \quad H(x) = \frac{a}{b}(e^{bx} - 1) + cx \tag{6.1}$$

was used as the most general parameterization for the univariate survival functions (indices are ignored here). The survival function S in (6.1) is derived by using the conditional hazard rate $\mu(x, Z) = Z(a\, e^{bx} + c) + d$. This model is a combination of the gamma-Makeham-parameterization

$$\mu(x, Z) = Z\, \mu_{01}(x) + d = Z\, a\, e^{bx} + d$$

and the Gompertz-Makeham-parameterization

$$\mu(x, Z) = Z\, \mu_{02}(x) = Z\, (a\, e^{bx} + c)\ .$$

Both parameterizations have been used in lifetime data analysis (Yashin et al., 1994). The fit of different submodels of (6.1) was compared using the Akaike Information Criterion. Let (S1) denote the submodel of (6.1) with $c = 0$, $d = 0$ and $s^2 = \sigma^2$ (variance of the corresponding frailty variable):

$$S(x) = [1 + \sigma^2\, \frac{a}{b}(e^{bx} - 1)]^{-1/\sigma^2} \tag{S1}$$

and let (S2) denote the submodel with $c = 0$ and $d = 0$:

$$S(x) = [1 + s^2\, \frac{a}{b}(e^{bx} - 1)]^{-1/s^2}. \tag{S2}$$

First a separate analysis for monozygotic and dizygotic twins was done, fitting combinations of different submodels of (6.1) for the univariate survival functions S_0 and S_L to the data. We determined maximum likelihood estimates of the parameters of the four-dimensional frailty distribution and of the univariate survival functions.
For monozygotic twins the parametric specifications (S1) for S_0 and (S2) for S_L fitted best according to AIC. For dizygotic twins a model with the parameterization (S2) for S_0 and S_L provided the best fit. The parameter estimates and their standard errors are given in the rows MZ and DZ of table 6.4.
Then a joint analysis of all data was carried out, in which the same parameters were estimated under the assumption of equal marginal distributions for monozygotic and dizygotic twin individuals. In the joint analysis the

Table 6.4: Results of separate and joint analysis. Estimators of the variances and correlations of the frailty variables and of the parameters of the univariate survival functions S_0 and S_L together with their standard errors (in brackets).

	σ_0	σ	ϱ_0	ϱ_1	ϱ	ϱ^*
MZ	3.738	1.670	0.369	0.477	0.379	0.119
	(0.558)	(0.114)	(0.077)	(0.039)	(0.054)	(0.041)
DZ	2.773	2.400	0.046	0.145	0.771	0.040
	(0.497)	(0.189)	(0.063)	(0.026)	(0.145)	(0.055)
MZ & DZ	2.935	1.918	0.280^{MZ}	0.427^{MZ}	0.560	0.130^{MZ}
	(0.508)	(0.116)	(0.089)	(0.035)	(0.106)	(0.057)
			0.055^{DZ}	0.182^{DZ}		0.036^{DZ}
			(0.064)	(0.028)		(0.043)

	s_0	$a_0 \cdot 10^9$	b_0	s_1	$a_1 \cdot 10^5$	b_1
MZ	3.738	0.237	0.255	0.259	2.488	0.102
	(0.558)	(0.382)	(0.025)	(0.107)	(0.704)	(0.004)
DZ	1.714	1.260	0.233	0.261	2.563	0.102
	(0.283)	(1.352)	(0.017)	(0.075)	(0.493)	(0.003)
MZ & DZ	2.264	2.035	0.221	0.270	2.545	0.102
	(0.402)	(1.733)	(0.013)	(0.060)	(0.409)	(0.002)

parametric specification (S2) was used for all univariate survival functions. Assuming equal values for σ_0, σ, ϱ, s_0, a_0, b_0, s_1, a_1 and b_1 for monozygotic and dizygotic twins, respectively, yields the estimation results given in the rows MZ & DZ of table 6.4.

The genetic analysis of the data is the last step. Five different genetic models were explored: ACE, AE, ADE, DE, CE. A refers to additive genetic effects, D to genetic effects due to dominance, C to shared environmental factors, and E to non-shared environmental factors. The basic assumptions and notations of quantitative genetics are introduced in section 1.5.

Let a^2 denote the proportion of variance associated with additive genetic effects. This is called narrow sense heritability. d^2, c^2 and e^2 are similarly defined as proportions of the phenotypic variance associated with the corresponding genetic or non-genetic effects. In quantitative genetics one generally assumes that for a phenotypic trait the correlations between monozygotic and

dizygotic twins can be expressed as

$$\begin{aligned} \varrho_{MZ} &= a^2 + d^2 + c^2 \ , \\ \varrho_{DZ} &= \tfrac{1}{2}\, a^2 + \tfrac{1}{4}\, d^2 + c^2 \end{aligned}$$

and that the normalizing equation

$$1 = a^2 + d^2 + c^2 + e^2$$

holds. Since only 3 of the parameters a^2, d^2, c^2 and e^2 are determined by these equations, one can only look at the 5 genetic models mentioned above. Let a_0^2, d_0^2, c_0^2 and e_0^2 denote the genetic parameters of the frailty that corresponds to the transition from healthy to ill. The same parameters of the frailty that corresponds to the transition from healthy to deceased are denoted a_L^2, d_L^2, c_L^2 and e_L^2. The four-dimensional frailty model allows for the joint genetic analysis of the frailty of becoming ill and the frailty of dying. One can thus combine the 5 genetic models and carry out the estimation procedure for a total of 25 different models. In each of these models a subset of the parameters $\{a_0^2, d_0^2, c_0^2, e_0^2, a_L^2, d_L^2, c_L^2, e_L^2\}$ is estimated. Since these models are not nested, the best-fitting one was chosen using the Akaike Information Criterion.
For the possible 25 combinations of the different genetic models, maximum likelihood estimates were calculated for σ_0, σ, ϱ, ϱ^*, the parameters of S_0 and S_L and for the genetic parameters. It turned out that a DE-model for the transition from healthy to ill ($d_0^2 = 0.27\,(0.09)$) and an AE-model for the transition from healthy to deceased ($a_L^2 = 0.41\,(0.03)$) was the best-fitting combination.
The maximum likelihood estimation and the calculation of standard errors of the MLEs are based on the methods described in Appendix C.1.

6.3 Results of the liability analysis applied to the data set

To compare the results derived by estimating the genetic parameters of frailty with those from the traditional analysis of the data set a liability analysis was carried out (see section 1.5.2).
The information that had to be extracted from the data set on prostate cancer in Swedish twins in order to perform the liability analysis is given in table 6.1.
The analysis was done using the program Mx. Mx is a matrix algebra interpreter and numerical optimizer for structural equation modeling and other

Table 6.5: Estimates of the tetrachoric correlations and the threshold values (95% confidence intervals in brackets).

ϱ_{MZ}	ϱ_{DZ}	T_{MZ}	T_{DZ}	χ^2 (df)	p	AIC
0.416	0.033	1.399	1.450	2.31 (2)	0.32	-1.70
(.290,.531)	(-.092,.155)	(1.332, 1.467)	(1.403,1.498)			
0.411	0.033	1.433	1.433	3.75 (3)	0.29	-2.25
(.289,.525)	(-.092,.157)	(1.395, 1.473)	(1.395,1.473)			

ϱ_{MZ} = correlation in liability for MZ twins, ϱ_{DZ} = correlation in liability for DZ twins, T_{MZ} = threshold value for MZ twins, T_{DZ} = threshold value for DZ twins

types of statistical modeling of data (Neale, 1995). It includes a special built-in fit function for the calculation of the maximum likelihood estimates for the tetrachoric correlation and the threshold values given a two-by-two contingency table. Mx also gives confidence intervals of the parameter estimates and a χ^2-statistic that allows us to evaluate the fit of the model as well as to compare nested models. The χ^2-statistic measures the discrepancy between the cell frequencies predicted by the model (bivariate normal distributed liability with the same threshold value for each individual in a pair) and those actually observed in the data. If this discrepancy is to large, i.e. the corresponding p-value is smaller than .05, then one would conclude that the model fails to predict the data.

In the first step of the analysis the tetrachoric correlations and threshold values were estimated for the prostate cancer data set. The results are given in table 6.5. In the last row of table 6.5 the parameters are estimated under the assumption of equal threshold values for MZ and DZ twins. The null hypothesis $H_0 : T_{MZ} = T_{DZ}$ can be tested by calculating the difference between the two χ^2 statistics given in the table. The resulting test statistic is asymptotically χ^2-distributed with one degree of freedom. Its value here is 1.44 which corresponds to a p-value of 0.23, thus the null hypothesis can be accepted. This means that MZ and DZ twins regarded as individuals have the same pattern of disease distribution.

The value of the correlation in liability is larger for the MZ twins which can hint at genes playing a role in the development of prostate cancer.

In the second step of the liability analysis different genetic models are fitted to the data and the genetic parameters are estimated using Mx again. The results are given in table 6.6. The best fitting model according to AIC is the DE model. As for the estimation of the threshold values and the tetrachoric correlations the χ^2-statistic given in table 6.6 tests the goodness of fit of the

Table 6.6: Estimates of the genetic parameters of liability (95% confidence intervals in brackets).

model	a^2	d^2	c^2	e^2	χ^2 (df)	p	AIC
ACE	0.341 (.193,.450)	-	0.000 (-)	0.659 (.550,.774)	9.80 (3)	0.02	3.80
ADE	0.000 (-)	0.396 (.061,.509)	-	0.604 (.492,.726)	4.89 (3)	0.18	-1.10
AE	0.341 (.226,.450)	-	-	0.659 (.550,.774)	9.80 (4)	0.04	1.80
DE	-	0.396 (.274,.509)	-	0.604 (.492,.726)	4.89 (4)	0.30	-3.11
CE	-	-	0.209 (.121,.295)	0.791 (.705,.879)	21.52 (4)	<0.01	13.52

a^2 = proportion of additive genetic variance, d^2 = proportion of dominance variance, c^2 = proportion of shared environment variance, e^2 = proportion of non-shared environment variance

model based on the difference between observed and expected cell frequencies. A p-value smaller than .05 would imply the rejection of the model. In an ACE model the c^2 and in an ADE the a^2 is estimated at zero. Both parameters are not significant in the corresponding model. But a DE model fits better than the AE model. The model without genetic effects (CE) fits badly.
The liability analysis yields that there are genetic effects on the liability to prostate cancer. These effects are due to dominance effects.

6.4 Discussion

The analysis of the Swedish twin data on prostate cancer suggests that genes have an influence on frailty in the development of the disease, which is due to dominance effects. The genetic influence on the frailty corresponding to the life span is moderate and results from additive genetic effects.
The latter confirm results derived by Yashin and Iachine (1995b) for Danish twins who estimated a heritability of about 50% for the frailty corresponding to life span.
The findings from the prostate data must be judged critically, however. First, using the age at diagnosis to calculate the age at onset of disease is a questionable procedure. Second, the number of affected people, i.e. the number

of people with a diagnosis of cancer is small. This means that the estimators of the parameters for the transition from healthy to ill have large errors, so one should be careful interpreting them.
Ahlbom et al. (1997) applied traditional methods of quantitative genetics on the same prostate cancer data set. They estimate correlations of liability and the relative risk of cancer for twins with an affected co-twin compared with twins with a non-affected co-twin. They find that "prostate cancer displays a clear familial effect that is almost accounted for by heritable effects" and an indication of non-additive heritable effects. The limitations of the data set are discussed thoroughly in the paper.
Another study of prostate cancer in Swedish twins by Grönberg et al. (1994) used other statistical methods (calculation of concordance rates and correlation of liability) but it also finds that "genetic factors might be important in prostate cancer development" on the basis of differences in proband concordance rates and correlations in liability between monozygotic and dizygotic twins. They calculate correlations of liability of 0.40 and -0.05 for monozygotic and dizygotic twins, respectively. The corresponding estimates of the correlation between the frailty variables influencing the transition from healthy to ill are 0.28 and 0.06, respectively (see table 6.4, row MZ & DZ, estimates for ϱ_0). These values are of the same magnitude.
The results of the liability analysis coincide with the results of the frailty analysis for this special prostate cancer data set. Both analyses show that there is a genetic influence on the development of prostate cancer and this influence might be explained with nonadditive genetic effects. As shown in Chapter 7 the conclusions made from liability analysis can differ from those made from frailty analysis. The genetic influence on the age at death can be a confounder for the disease-status relationship in a twin pair. When applying the 3-state-frailty model one can distinguish whether there are genetic effects on the development of disease or on the life span or on both. With the liability approach this distinction is not possible.
Applying the traditional method of calculating tetrachoric correlations in liability on the data described in section 6.1 means that one only uses the information given in table 6.1. In doing this one ignores, on the one hand, the fact the some of the living twins might still receive a cancer diagnosis after the end of the study (right censoring) and that there might have been cases of prostate cancer before the beginning of the observation period (left truncation). On the other hand, one does not use the additional information about the age at onset of the disease. All these facts are accounted for in the four-dimensional correlated frailty model.

Chapter 7

A comparison of liability analysis with frailty analysis

The liability approach (see section 1.5.2) and the frailty approach (see section 2.1) are conceptually similar in the sense that in both models it is assumed that an unobserved random variable exists that influences the probability of being affected by a disease. Both variables, liability and frailty, represent the susceptibility of getting ill.
The liability is assumed to be standard normal distributed and the probability of being ill depends on the fact whether the value of liability is above or below a certain fixed threshold value. In this work the frailty variable is assumed to be gamma distributed with mean 1 and it influences the probability of getting ill through the conditional hazard function. In both analyses the genetic parameters of the underlying non-observable variables are estimated by applying the structural equations from quantitative genetics on the correlations of liability and frailty, respectively.
In this chapter the two approaches are compared using simulated data. The advantages of the frailty approach are illustrated.

7.1 Simulation of data sets

The data sets for this simulation study were structured using the data set on prostate cancer in Swedish twins (see section 6.1) as an example. This means that there is the same mechanism of truncation and censoring, that different birth cohorts of twins are investigated in a certain observation period, and that the time of the onset of disease is known.

7.1.1 Uncensored data

In the first step uncensored data sets were simulated according to the four-dimensional correlated gamma frailty model (see definition 9). That means that for a given choice of the parameters of the four-dimensional frailty distribution $(\sigma_0, \sigma, \varrho_0, \varrho_1, \varrho, \varrho^*)$ and of the univariate survival functions S_0 and S_L realizations of the random vector $(T_{10}, T_{20}, T_1, T_2)$ were calculated. This was done for separate data sets for MZ and DZ twins and for different genetic models for the frailty variables.
For all data sets the following parameters had the same value in the simulation: $\sigma_0 = 1.5$, $\sigma = 1$, $\varrho = 0.3$, $\varrho^*_{MZ} = \varrho^*_{DZ} = 0.05$. The univariate survival functions have been chosen to be

$$S_0(x) = [1 + \sigma_0^2 \frac{a_0}{b_0}(e^{b_0 x} - 1)]^{-1/\sigma_0^2} \quad \text{with } a_0 = 10^{-7} \text{ and } b_0 = 0.14, \text{ and}$$

$$S_L(x) = [1 + \sigma^2 \frac{a}{b}(e^{bx} - 1)]^{-1/\sigma^2} \quad \text{with } a = 4 \cdot 10^{-5} \text{ and } b = 0.1.$$

This choice corresponds with an expected value of 107.4 for the time until the onset of disease and an expected value of 78.3 for the life span. The reason for this choice of the parameters of the univariate survival functions is the assumption that the risk of falling ill and the risk of dying are two dependent competing risks here. That means that everybody would eventually get the disease if he or she would live long enough but most people die before developing the disease under study. Consequently the observed mean age at the onset of disease is smaller then the expected value of the age at the onset of disease.
The values of the parameters ϱ_0 and ϱ_1 used in the simulation are given in table 7.1. For the first pair of data sets (one data set for MZ, one for DZ twins) it was assumed that the frailty corresponding to the transition from healthy to ill and the frailty corresponding to the transition from healthy to deceased can both be described by a model that consists only of additive genetic effects. This is equivalent to setting the correlation between the MZ twins twice as high as the correlation between the DZ twins (see (1.21) and (1.22)). The first pair of data sets has the code gg.
For the second pair of data sets (code eg) it was assumed that the frailty corresponding to the transition from healthy to ill can be described by a model consisting only of shared environmental effects. That means that the correlation between the MZ twins equals the correlation between the DZ twins. The genetic model for the frailty corresponding to the transition from healthy to deceased remained the same.
The reversed choice of genetic models was done for the third pair of data sets (code ge). I.e. the genetic model for the frailty of becoming ill is a purely

Table 7.1: Values of the correlations ϱ_0 and ϱ_1 used in the simulation.

code	genetic model healthy → ill	genetic model healthy → deceased	ϱ_0	ϱ_1
gg	AE	AE	0.6^{MZ} 0.3^{DZ}	0.8^{MZ} 0.4^{DZ}
eg	CE	AE	0.45^{MZ} 0.45^{DZ}	0.8^{MZ} 0.4^{DZ}
ge	AE	CE	0.6^{MZ} 0.3^{DZ}	0.6^{MZ} 0.6^{DZ}
ee	CE	CE	0.45^{MZ} 0.45^{DZ}	0.6^{MZ} 0.6^{DZ}

ϱ_0 = correlation of frailties for transition from healthy to ill,
ϱ_1 = correlation of frailties for transition from healthy to deceased

additive genetic one and the genetic model for the frailty of dying is a purely shared environmental one.
In the last pair of data sets (code ee) the correlations are the same for MZ and DZ twins for both transitions, i.e. the frailty of becoming ill as well as the frailty of dying are described by a model consisting only of shared environmental effects.
The simulated sample size was always 2000 MZ and 4000 DZ pairs.

7.1.2 Restricted observations

In the second step of simulating data a transformation procedure was applied to get observations as in the real prostate cancer data set.
First, for every twin pair the age at left truncation was simulated as the realization of a random variable that is uniformly distributed on the interval $[15, 80]$. Then a fixed observation period of 60 units was assumed. If A_T denotes the age at left truncation then the age at the end of the observation period A_C can be calculated as $A_C = A_T + 60$. All pairs where $T_1 \leq A_T$ or $T_2 \leq A_T$, meaning one of the twins died before the beginning of the observation period, were removed from the data set. This is exactly the mechanism of pairwise left truncation that holds for the prostate cancer data set.
Second, for every twin the type of observation $\tilde{C}_i$, $i = 1, 2$, was calculated as

(see page 88):

$$\tilde{C}_i = \begin{cases} 1 & \text{if the twin is deceased and no diagnosis exists} \\ & \iff (T_{i0} \leq A_T \text{ or } T_{i0} > T_i) \text{ and } T_i \leq A_C, \\ 2 & \text{if the twin is deceased and a diagnosis exists} \\ & \iff A_T < T_{i0} \leq T_i \leq A_C, \\ 3 & \text{if the twin is alive and no diagnosis exists} \\ & \iff (T_{i0} \leq A_T \text{ or } T_{i0} > A_C) \text{ and } T_i > A_C, \\ 4 & \text{if the twin is alive and a diagnosis exists} \\ & \iff A_T < T_{i0} \leq A_C \text{ and } T_i > A_C. \end{cases}$$

The information about the age at death of every individual is included in the censored life span:

$$X_i = \min(T_i, A_C) \quad , i = 1, 2.$$

The age at the end of the observation period A_C is the censoring variable here. The censored age at the onset of disease is not known but

$$\tilde{X}_i = \begin{cases} \min(T_{i0}, X_i) & \text{if } T_{i0} > A_T, \\ X_i & \text{if } T_{i0} \leq A_T. \end{cases}$$

The exact age at the onset of disease is only known if it is smaller than the censored life span (competing risk situation) and larger than the age at left truncation. Due to the truncation the sample size is reduced. The new sample sizes are given in table 7.2.

Table 7.2: Sample size in the transformed data sets in pairs.

code	MZ	DZ
gg	1633	3189
eg	1632	3221
ge	1596	3272
ee	1597	3261

For every twin pair an observation in the transformed data set consists of a realization of the vector $(\tilde{X}_1, \tilde{X}_2, X_1, X_2, \tilde{C}_1, \tilde{C}_2, A_T, A_C)$. This new data set is used in the frailty analysis, that means that maximum likelihood estimates of the parameter are derived using this data.

Table 7.3: Estimates of the correlations in liability and the threshold values for the simulated data sets (95% confidence intervals in brackets).

code	ϱ_{MZ}	ϱ_{DZ}	T_h	χ^2 (df)	p	AIC
gg	0.507 (.395,.607)	0.161 (.050,.268)	1.411 (1.373, 1.450)	2.26 (3)	0.52	-3.74
eg	0.476 (.360,.581)	0.313 (.211,.410)	1.432 (1.393, 1.472)	3.37 (3)	0.34	-2.63
ge	0.343 (.212,.464)	0.297 (.198,.391)	1.405 (1.367, 1.444)	1.68 (3)	0.64	-4.32
ee	0.250 (.103,.389)	0.334 (.239,.425)	1.419 (1.380, 1.458)	0.43 (3)	0.93	-5.57

ϱ_{MZ} = correlation in liability for MZ twins, ϱ_{DZ} = correlation in liability for DZ twins, T_h threshold value for MZ and DZ twins

For the liability analysis the two-by-two contingency tables containing the information "a diagnosis exists" ($\tilde{C}_i = 2$ or $\tilde{C}_i = 4$) versus "no diagnosis exists" ($\tilde{C}_i = 1$ or $\tilde{C}_i = 3$) for the twins in a pair are calculated.

7.2 Results of the liability analysis

The analysis of the simulated data sets (the two-by-two contingency tables) with respect to the liability approach was done in the same way as for the data set on prostate cancer in twins (see section 6.3). The interpretation of the χ^2-statistics and p-values given in tables 7.3, 7.4 and 7.5 is the same as in this analysis. They test the goodness of fit of the model based on the difference between observed and expected cell frequencies. A p-value smaller than .05 would imply the rejection of the model.
First the correlations in liability and the threshold values were estimated. The results are given in table 7.3. The table shows the estimates under the assumption of equal threshold values for MZ and DZ twins. A significant difference between the threshold values of MZ and DZ twins could not be found for any of the four pairs of data sets .
One would conclude that there are genetic influences on the liability in the disease if the correlation between MZ twins is substantially higher than the correlation between DZ twins. This is true for the cases where the model for the frailty corresponding to the transition from healthy to deceased is additive genetic (gg and eg). Whether the model for the frailty corresponding

Table 7.4: Estimates of the genetic parameters of liability - case gg (95% confidence intervals in brackets).

model	a^2	d^2	c^2	e^2	χ^2 (df)	p	AIC
ACE	0.469 (.302,.564)	-	0.000 (-)	0.531 (.436,.634)	4.54 (3)	0.21	-1.46
ADE	0.136 (.000,.530)	0.371 (.000,.606)	-	0.493 (.393,.605)	2.26 (3)	0.52	-3.74
AE	0.469 (.366,.564)	-	-	0.531 (.436,.634)	4.54 (4)	0.34	-3.46
DE	-	0.514 (.407,.611)	-	0.486 (.389,.593)	2.61 (4)	0.62	-5.39
CE	-	-	0.311 (.232,.387)	0.689 (.613,.768)	21.19 (4)	<0.001	13.19

a^2 = proportion of additive genetic variance, d^2 = proportion of dominance variance, c^2 = proportion of shared environment variance, e^2 = proportion of non-shared environment variance

to the transition from healthy to ill includes genetic factors or not has no pronounced effect on the correlations in liability. The same pattern can be found for the cases where the model for the frailty in dying contains no genetic effects (ge and ee). Then the correlations in liability are of the same magnitude for MZ and DZ twins regardless of whether there are genetic effects for the frailty in falling ill or not. More detailed information can be derived by the fitting of different genetic models.

The results of fitting different genetic models for the liability are given in the tables 7.4 and 7.5 for the two cases where the frailty of dying is modeled to be influenced by additive genetic factors (gg and eg) .

In the case that both frailty variables include only additive genetic effects (case gg) the best fitting model for the liability according to AIC is a DE model ($d^2 = 0.514\,(0.407, 0.611)$), see table 7.4. This model has also the highest p-value. In an ACE model the estimate of c^2 is zero, as well as the estimate of a^2 in an ADE model. These results would yield the conclusion that there are genetic influences on the liability of developing a disease.

In the case where the model for frailty of dying includes genetic effects again but the model for the frailty of falling ill does not include genetic effects, the best fitting model for the liability is an AE model ($a^2 = 0.509\,(0.411, 0.599)$), see table 7.5. Looking at the results from this analysis one would conclude

Table 7.5: Estimates of the genetic parameters of liability - case eg (95% confidence intervals in brackets).

model	a^2	d^2	c^2	e^2	χ^2 (df)	p	AIC
ACE	0.325 (.023,.578)	-	0.150 (.000,.376)	0.524 (.420,.640)	3.37 (3)	0.34	-2.63
ADE	0.509 (.268,.599)	0.000 (.000,.254)	-	0.492 (.402,.590)	5.01 (3)	0.17	-0.99
AE	0.509 (.411,.599)	-	-	0.492 (.402,.590)	5.01 (4)	0.29	-2.99
DE	-	0.522 (.417,.617)	-	0.478 (.383,.583)	15.95 (4)	0.003	7.95
CE	-	-	0.379 (.303,.452)	0.621 (.548,.698)	7.83 (4)	0.10	-0.17

a^2 = proportion of additive genetic variance, d^2 = proportion of dominance variance, c^2 = proportion of shared environment variance, e^2 = proportion of non-shared environment variance

that there are genetic influences on the development of the disease but the data were modeled in such a way that no such influences occur. The reason for this misinterpretation is that the genetic influence on the life span confounds the disease-status relationship of the twins. One cannot recognize this confounding by using the liability approach.
For the two cases where the model for the frailty corresponding to the transition from healthy to deceased contains no genetic effects, the best fitting model for the liability is the CE model: $c^2 = 0.313\,(0.235, 0.389)$ for the case ge and $c^2 = 0.308\,(0.229, 0.385)$ for the case ee. In both situations the AIC is the smallest and the p-value is the highest for the CE model. Thus for both cases one would conclude that there are no genetic effects on the liability of developing the disease. In the situation where the frailty of falling ill is modeled to be influenced by the genes this again shows the confounding of the relation between the ages at death of the twins.

7.3 Results of the frailty analysis

The analysis of the transformed data sets using the frailty approach was carried out like for the real data on prostate cancer (see section 6.2).
In the first step maximum likelihood estimates for the correlations and vari-

ances of the frailty variables ($\sigma_0, \sigma, \varrho_0, \varrho_1, \varrho, \varrho^*$) and for the parameter of the univariate survival functions (a_0, b_0, a, b) were calculated. This was first done separately for the MZ and DZ twins and then in a joint analysis these parameters were estimated under the assumption of equal marginal distributions for MZ and DZ twin individuals. The latter assumption is equivalent to introducing the constraint that the parameter estimates of σ_0, σ, ϱ, a_0, b_0, a and b are equal for MZ and DZ twins. The null hypothesis

$$H_0 : \sigma_0^{MZ} = \sigma_0^{DZ}, \sigma^{MZ} = \sigma^{DZ}, \varrho^{MZ} = \varrho^{DZ}, a_0^{MZ} = a_0^{DZ},$$
$$b_0^{MZ} = b_0^{DZ}, a^{MZ} = a^{DZ}, b^{MZ} = b^{DZ}$$

was tested using a likelihood ratio test. It could be accepted for all four pairs of data sets.

Table 7.6 shows the results of the joint analysis. Compared to the values used in the data simulation the values of the estimates are acceptable. The standard errors of the estimates of the parameters belonging to the transition from healthy to ill have larger standard errors than the estimates belonging to the transition from healthy to deceased. This is due to the small proportion of affected people (about 8 % for all data sets).

For all four pairs of data sets the estimates of the correlations of the frailty variables representing susceptibility to illness or death confirm the mechanism with which the data sets where simulated. If a frailty variable was modeled to be genetically influenced then the corresponding correlation in frailty is higher for the MZ twins than for the DZ twins, if there were only shared environmental effects modeled, then the corresponding correlation in frailty is of the same magnitude for both MZ and DZ twins.

In the second step a more detailed analysis that fits combinations of different genetic models should show whether it is possible to investigate the nature of genetic and environmental influences with the help of the frailty approach. Let analogous to the analysis of the prostate cancer data (see page 92) a_0^2, d_0^2, c_0^2 and e_0^2 denote the genetic parameters of the frailty that corresponds to the transition from healthy to ill and let a_L^2, d_L^2, c_L^2 and e_L^2 denote the same parameters of the frailty that corresponds to the transition from healthy to deceased. The 25 possible combinations of the genetic models ACE, ADE, AE, DE and CE were fitted to the four pairs of data sets. That means that in addition to the maximum likelihood estimates of the parameter σ_0, σ, ϱ, ϱ^*, a_0, b_0, a_1, b_1 the maximum likelihood estimates of the subset of $\{a_0^2, d_0^2, c_0^2, e_0^2, a_L^2, d_L^2, c_L^2, e_L^2\}$ that corresponds to a certain combination of genetic models are calculated. Table 7.7 gives the latter estimates for the best fitting models according to AIC for the four pairs of data sets.

Table 7.6: Results of joint analysis. MLEs of the variances and correlations of the frailty variables and of the parameters of the univariate survival functions S_0 and S_L together with their standard errors (in brackets).

code	σ_0	σ	ϱ_0	ϱ_1	ϱ	ϱ^*	$a_0 \cdot 10^7$	b_0	$a_1 \cdot 10^5$	b_1
values used in simulation	1.5	1	see table 7.1		0.3	0.05	1	0.14	4	0.1
gg	1.468 (0.141)	1.005 (0.026)	0.659^{MZ} (0.161) 0.269^{DZ} (0.159)	0.752^{MZ} (0.038) 0.445^{DZ} (0.037)	0.311 (0.077)	0.077^{MZ} (0.123) 0.059^{DZ} (0.110)	0.947 (0.416)	0.140 (0.006)	3.869 (0.534)	0.101 (0.002)
eg	1.422 (0.131)	1.006 (0.026)	0.492^{MZ} (0.135) 0.435^{DZ} (0.174)	0.792^{MZ} (0.036) 0.416^{DZ} (0.038)	0.346 (0.072)	0.055^{MZ} (0.100) 0.037^{DZ} (0.123)	1.181 (0.501)	0.138 (0.006)	3.581 (0.492)	0.102 (0.002)
ge	1.299 (0.139)	1.081 (0.027)	0.542^{MZ} (0.162) 0.254^{DZ} (0.080)	0.614^{MZ} (0.040) 0.582^{DZ} (0.031)	0.434 (0.088)	0.122^{MZ} (0.142) $3{\cdot}10^{-5\,\mathrm{DZ}}$ $(6{\cdot}10^{-4})$	1.193 (0.532)	0.138 (0.006)	2.924 (0.429)	0.105 (0.002)
ee	1.501 (0.174)	0.999 (0.026)	0.393^{MZ} (0.202) 0.528^{DZ} (0.127)	0.623^{MZ} (0.045) 0.592^{DZ} (0.033)	0.358 (0.075)	0.150^{MZ} (0.134) 0.140^{DZ} (0.089)	0.963 (0.459)	0.141 (0.006)	3.581 (0.495)	0.102 (0.002)

In the case where the frailty variables that represent susceptibility to illness or death are both modeled to be influenced genetically (case gg), a combination of an AE model with an ACE model fits as well as a combination of a DE model with an ACE model. Therefore the frailty analysis yields that there are genetic influences on the frailty of falling ill whose nature is not clear, and that there are additive genetic effects and small environmental effects on the frailty of dying.
For the other three cases, the best fitting combination of genetic models according to AIC is precisely the combination of genetic models used in the simulation of the data.

Table 7.7: Results of genetic analysis. MLEs of the genetic parameter for the best fitting model according to AIC (standard errors in brackets).

code	genetic models*	a_0^2	d_0^2	c_0^2	e_0^2	a_L^2	c_L^2	e_L^2
gg	AE × ACE	0.632 (0.143)	-	-	0.368 (0.143)	0.614 (0.102)	0.138 (0.081)	0.248 (0.038)
	DE × ACE	-	0.682 (0.153)	-	0.318 (0.153)	0.619 (0.102)	0.133 (0.081)	0.248 (0.038)
eg	CE × AE	-	-	0.477 (0.107)	0.523 (0.107)	0.796 (0.032)	-	0.204 (0.032)
ge	AE × CE	0.523 (0.120)	-	-	0.477 (0.120)	-	0.593 (0.026)	0.407 (0.026)
ee	CE × CE	-	-	0.490 (0.109)	0.510 (0.109)	-	0.602 (0.028)	0.398 (0.028)

*Best fitting combination of genetic models for frailty according to AIC.

7.4 Conclusions

The comparison of the liability approach with the frailty approach shows that both methods have their advantages and drawbacks.
To apply the liability analysis for a certain disease one only needs the information affected versus non-affected for twins of different zygosities. The calculations are numerically relatively easy.

The method has some drawbacks. With the liability approach it is not possible to account for truncation and censoring in the data set. If, in addition to the health status, information about age at onset of disease or age at death is available then this information is not used in the analysis. Finally, due to the confounding effect of the age at death, misinterpretations are possible. On the one hand, genetic influences on the development of a disease might be concluded that do not exist and the conclusion is due to the genetic influence on the age at death. On the other hand, genetic influences on the liability to disease might not be detected.
The frailty approach has the advantage that it is possible to distinguish whether there are genetic effects for the frailty of falling ill and/or the frailty of dying. It is easy to account for truncation and censoring. Whether it is possible to make conclusions about the nature of genetic effects (additive or non-additive) depends on the quality of the data set. In general one should be very careful about making such conclusions.
Another advantage of the frailty model is the ability to handle multiple study waves, at the same time increasing the number of observations without changing the model.
The main drawback of the frailty approach is that it demands additional information at least about the age at death but preferably also about the age at onset of disease, and that the required sample sizes are at least moderate. In addition a rather large proportion of ill people is necessary which restricts the approach to common diseases. The frailty analysis is computationally demanding, additional efforts for getting maximum likelihood estimates are necessary.

Appendix A

Notations

A.1 Functions

The following functions are used throughout the entire work, most notably in the proofs of the theorems about identifiability, consistency and asymptotic normality. To make the understanding of these proofs easier, the functions are collected here and their explicit representation is given.

Shape parameter of the components of the frailty variables

$$\begin{aligned}
k_0 &= k_0(\sigma_0, \sigma, \varrho^*) = \frac{\varrho^*}{\sigma_0\sigma} \\
k_1 &= k_1(\sigma_0, \sigma, \varrho, \varrho^*) = \frac{\varrho - \varrho^*}{\sigma_0\sigma} \\
k_3 &= k_3(\sigma_0, \sigma, \varrho_0, \varrho^*) = \frac{\varrho_0}{\sigma_0^2} - \frac{\varrho^*}{\sigma_0\sigma} \\
k_4 &= k_4(\sigma_0, \sigma, \varrho_1, \varrho^*) = \frac{\varrho_1}{\sigma^2} - \frac{\varrho^*}{\sigma_0\sigma} \\
k_{10} &= k_{10}(\sigma_0, \sigma, \varrho_0, \varrho, \varrho^*) = \frac{1 - \varrho_0}{\sigma_0^2} - \frac{\varrho - \varrho^*}{\sigma_0\sigma} \\
k_{11} &= k_{11}(\sigma_0, \sigma, \varrho_1, \varrho, \varrho^*) = \frac{1 - \varrho_1}{\sigma^2} - \frac{\varrho - \varrho^*}{\sigma_0\sigma}
\end{aligned} \tag{A.1}$$

Parts of the four-dimensional survival function

$$\begin{aligned}
L_0(s_1, s_2, t_1, t_2) &= L_0(s_1, s_2, t_1, t_2|\boldsymbol{\theta}) \\
&= 1 + \sigma_0^2 H_0(s_1|\boldsymbol{\lambda}) + \sigma_0^2 H_0(s_2|\boldsymbol{\lambda}) + \sigma^2 H(t_1|\boldsymbol{\lambda}) + \sigma^2 H(t_2|\boldsymbol{\lambda})
\end{aligned} \tag{A.2}$$

$$L_1(s,t) = L_1(s,t|\boldsymbol{\theta}) = 1 + \sigma_0^2 H_0(s|\boldsymbol{\lambda}) + \sigma^2 H(t|\boldsymbol{\lambda}) \tag{A.3}$$

$$L_3(s_1,s_2) = L_3(s_1,s_2|\boldsymbol{\theta}) = 1 + \sigma_0^2 H_0(s_1|\boldsymbol{\lambda}) + \sigma_0^2 H_0(s_2|\boldsymbol{\lambda}) \tag{A.4}$$

$$L_4(t_1,t_2) = L_4(t_1,t_2|\boldsymbol{\theta}) = 1 + \sigma^2 H(t_1|\boldsymbol{\lambda}) + \sigma^2 H(t_2|\boldsymbol{\lambda}) \tag{A.5}$$

$$L_{10}(s) = L_{10}(s|\boldsymbol{\theta}) = 1 + \sigma_0^2 H_0(s|\boldsymbol{\lambda}) \tag{A.6}$$

$$L_{11}(t) = L_{11}(t|\boldsymbol{\theta}) = 1 + \sigma^2 H(t|\boldsymbol{\lambda}) \tag{A.7}$$

The four-dimensional survival function

In addition to the representations of the four-dimensional survival function given in (2.8) and (2.9) this function can be represented as

$$\begin{aligned} S(s_1,s_2,t_1,t_2|\boldsymbol{\theta}) = {} & L_{10}^{-k_{10}}(s_1|\boldsymbol{\theta}) L_{10}^{-k_{10}}(s_2|\boldsymbol{\theta}) L_{11}^{-k_{11}}(t_1|\boldsymbol{\theta}) L_{11}^{-k_{11}}(t_2|\boldsymbol{\theta}) L_1^{-k_1}(s_1,t_1|\boldsymbol{\theta}) \\ & \cdot L_1^{-k_1}(s_2,t_2|\boldsymbol{\theta}) L_3^{-k_3}(s_1,s_2|\boldsymbol{\theta}) L_4^{-k_4}(t_1,t_2|\boldsymbol{\theta}) L_0^{-k_0}(s_1,s_2,t_1,t_2|\boldsymbol{\theta}) \end{aligned} \tag{A.8}$$

The likelihood function of current health status data

The likelihood function of current health status data is given in (2.20). The parts of this function can be represented as

$$f_X(x_1,x_2,0,0|\boldsymbol{\theta}) = f_{00}(x_1,x_2|\boldsymbol{\theta}) f_A(x_1)\, \mathrm{I}(x_1 = x_2) \tag{A.9}$$

$$f_X(x_1,x_2,1,0|\boldsymbol{\theta}) = \big(f_{10}(x_1,x_2|\boldsymbol{\theta}) - f_{00}(x_1,x_2|\boldsymbol{\theta})\big) f_A(x_1)\, \mathrm{I}(x_1 = x_2) \tag{A.10}$$

$$f_X(x_1,x_2,0,1|\boldsymbol{\theta}) = f_X(x_2,x_1,1,0|\boldsymbol{\theta}) \tag{A.11}$$

$$\begin{aligned} f_X(x_1,x_2,1,1|\boldsymbol{\theta}) = {} & \big(f_{11}(x_1,x_2|\boldsymbol{\theta}) - 2\, f_{10}(x_1,x_2|\boldsymbol{\theta}) + f_{00}(x_1,x_2|\boldsymbol{\theta})\big) \\ & \cdot f_A(x_1)\, \mathrm{I}(x_1 = x_2) \end{aligned} \tag{A.12}$$

$$f_X(x_1,x_2,2,0|\boldsymbol{\theta}) = f_{20}(x_1,x_2|\boldsymbol{\theta}) f_A(x_2)\, \mathrm{I}(x_1 \le x_2) \tag{A.13}$$

$$f_X(x_1,x_2,0,2|\boldsymbol{\theta}) = f_X(x_2,x_1,2,0|\boldsymbol{\theta}) \tag{A.14}$$

$$f_X(x_1,x_2,2,1|\boldsymbol{\theta}) = \big(f_{21}(x_1,x_2|\boldsymbol{\theta}) - f_{20}(x_1,x_2|\boldsymbol{\theta})\big) f_A(x_2)\, \mathrm{I}(x_1 \le x_2) \tag{A.15}$$

$$f_X(x_1,x_2,1,2|\boldsymbol{\theta}) = f_X(x_2,x_1,2,1|\boldsymbol{\theta}) \tag{A.16}$$

$$f_X(x_1,x_2,2,2|\boldsymbol{\theta}) = f_{22}(x_1,x_2|\boldsymbol{\theta}) S_A(\max(x_1,x_2)) \tag{A.17}$$

where (omitting the dependence on $\boldsymbol{\theta}$)

$$\begin{aligned} f_{00}(x_1,x_2|\boldsymbol{\theta}) &= S(x_1,x_2,x_1,x_2) \\ &= L_{10}^{-k_{10}}(x_1) L_{10}^{-k_{10}}(x_2) L_{11}^{-k_{11}}(x_1) L_{11}^{-k_{11}}(x_2) L_1^{-k_1}(x_1,x_1) L_1^{-k_1}(x_2,x_2) \\ &\quad \cdot L_3^{-k_3}(x_1,x_2) L_4^{-k_4}(x_1,x_2) L_0^{-k_0}(x_1,x_2,x_1,x_2) \end{aligned} \tag{A.18}$$

$$
\begin{aligned}
f_{10}(x_1, x_2|\boldsymbol{\theta}) &= S(0, x_2, x_1, x_2) \\
&= L_{10}^{-k_{10}-k_3}(x_2) L_{11}^{-k_{11}-k_1}(x_1) L_{11}^{-k_{11}}(x_2) L_1^{-k_1}(x_2, x_2) \\
&\quad \cdot L_4^{-k_4}(x_1, x_2) L_0^{-k_0}(0, x_2, x_1, x_2) \qquad \text{(A.19)}
\end{aligned}
$$

$$
\begin{aligned}
f_{11}(x_1, x_2|\boldsymbol{\theta}) &= S(0, 0, x_1, x_2) = S_{L*}(x_1, x_2) \\
&= L_{11}^{-\frac{1-\varrho_1}{\sigma^2}}(x_1) L_{11}^{-\frac{1-\varrho_1}{\sigma^2}}(x_2) L_4^{-\frac{\varrho_1}{\sigma^2}}(x_1, x_2) \qquad \text{(A.20)} \\
&= L_{11}^{-k_{11}-k_1}(x_1) L_{11}^{-k_{11}-k_1}(x_2) L_4^{-k_4-k_0}(x_1, x_2) \qquad \text{(A.21)}
\end{aligned}
$$

$$
\begin{aligned}
f_{20}(x_1, x_2|\boldsymbol{\theta}) &= -\frac{\partial S(s_1, s_2, t_1, t_2)}{\partial t_1}\bigg|_{\substack{s_1=0, t_1=x_1 \\ s_2=t_2=x_2}} \\
&= f_{10}(x_1, x_2|\boldsymbol{\theta})\, \sigma^2 \mu(x_1) \left\{ \frac{k_{11}+k_1}{L_{11}(x_1)} + \frac{k_4}{L_4(x_1, x_2)} + \frac{k_0}{L_0(0, x_2, x_1, x_2)} \right\}
\end{aligned} \qquad \text{(A.22)}
$$

$$
\begin{aligned}
f_{21}(x_1, x_2|\boldsymbol{\theta}) &= -\frac{\partial S(s_1, s_2, t_1, t_2)}{\partial t_1}\bigg|_{\substack{s_1=s_2=0 \\ t_1=x_1, t_2=x_2}} = -\frac{\partial S_{L*}(x_1, x_2)}{\partial x_1} \\
&= f_{11}(x_1, x_2|\boldsymbol{\theta})\, \mu(x_1) \left\{ \frac{1-\varrho_1}{L_{11}(x_1)} + \frac{\varrho_1}{L_4(x_1, x_2)} \right\} \qquad \text{(A.23)} \\
&= f_{11}(x_1, x_2|\boldsymbol{\theta})\, \sigma^2 \mu(x_1) \left\{ \frac{k_{11}+k_1}{L_{11}(x_1)} + \frac{k_4+k_0}{L_4(x_1, x_2)} \right\} \qquad \text{(A.24)}
\end{aligned}
$$

$$
\begin{aligned}
f_{22}(x_1, x_2|\boldsymbol{\theta}) &= -\frac{\partial^2 S(s_1, s_2, t_1, t_2)}{\partial t_1 \partial t_2}\bigg|_{\substack{s_1=s_2=0 \\ t_1=x_1, t_2=x_2}} = -\frac{\partial^2 S_{L*}(x_1, x_2)}{\partial x_1 \partial x_2} \\
&= f_{11}(x_1, x_2|\boldsymbol{\theta})\, \mu(x_1)\mu(x_2) \left\{ \frac{\sigma^2 \varrho_1}{L_4^2(x_1, x_2)} \right. \\
&\qquad \left. + \left[\frac{1-\varrho_1}{L_{11}(x_1)} + \frac{\varrho_1}{L_4(x_1, x_2)} \right] \left[\frac{1-\varrho_1}{L_{11}(x_2)} + \frac{\varrho_1}{L_4(x_1, x_2)} \right] \right\}
\end{aligned} \qquad \text{(A.25)}
$$

$$
\begin{aligned}
&= f_{11}(x_1, x_2|\boldsymbol{\theta})\, \sigma^4 \mu(x_1)\mu(x_2) \left\{ \frac{k_4+k_0}{L_4^2(x_1, x_2)} \right. \\
&\qquad \left. + \left[\frac{k_{11}+k_1}{L_{11}(x_1)} + \frac{k_4+k_0}{L_4(x_1, x_2)} \right] \left[\frac{k_{11}+k_1}{L_{11}(x_2)} + \frac{k_4+k_0}{L_4(x_1, x_2)} \right] \right\}
\end{aligned} \qquad \text{(A.26)}
$$

Functions used in the proof of asymptotic normality

$$l_0(\boldsymbol{\vartheta}) = l_0(\sigma, \varrho_1) = \frac{\varrho_1}{\sigma^2} \tag{A.27}$$

$$l_1(\boldsymbol{\vartheta}) = l_1(\sigma, \varrho_1) = \frac{1 - \varrho_1}{\sigma^2} \tag{A.28}$$

$$L_4(x, y|\boldsymbol{\vartheta}) = L_4(x, y|\sigma, \boldsymbol{\lambda}) = 1 + \sigma^2 H(x|\boldsymbol{\lambda}) + \sigma^2 H(y|\boldsymbol{\lambda}) \tag{A.29}$$

$$L_{11}(x|\boldsymbol{\vartheta}) = L_{11}(x|\sigma, \boldsymbol{\lambda}) = 1 + \sigma^2 H(x|\boldsymbol{\lambda}) \tag{A.30}$$

$$g_{0,\alpha}(x, y|\boldsymbol{\vartheta}) = S^{\alpha}_{L*}(x, y|\boldsymbol{\vartheta}) = [L_{11}(x|\boldsymbol{\vartheta}) L_{11}(y|\boldsymbol{\vartheta})]^{-\alpha\, l_1(\boldsymbol{\vartheta})}\, [L_4(x, y|\boldsymbol{\vartheta})]^{-\alpha\, l_0(\boldsymbol{\vartheta})} \tag{A.31}$$

$$g_{1,\alpha}(x, y|\boldsymbol{\vartheta}) = h_1(x, y|\boldsymbol{\vartheta}) g_{0,\alpha}(x, y|\boldsymbol{\vartheta}) \tag{A.32}$$

$$h_1(x, y|\boldsymbol{\vartheta}) = \sigma^2 \mu(x|\boldsymbol{\lambda}) \left\{ \frac{l_1(\boldsymbol{\vartheta})}{L_{11}(x|\boldsymbol{\vartheta})} + \frac{l_0(\boldsymbol{\vartheta})}{L_4(x, y|\boldsymbol{\vartheta})} \right\} \tag{A.33}$$

$$g_{2,\alpha}(x, y|\boldsymbol{\vartheta}) = h_2(x, y|\boldsymbol{\vartheta}) g_{0,\alpha}(x, y|\boldsymbol{\vartheta}) + g_{1,\alpha/2}(x, y|\boldsymbol{\vartheta}) g_{1,\alpha/2}(y, x|\boldsymbol{\vartheta}) \tag{A.34}$$

$$h_2(x, y|\boldsymbol{\vartheta}) = \sigma^4 \mu(x|\boldsymbol{\lambda}) \mu(y|\boldsymbol{\lambda}) \frac{l_0(\boldsymbol{\vartheta})}{L_4^2(x, y|\boldsymbol{\vartheta})} \tag{A.35}$$

$$h_{11}(x|\boldsymbol{\vartheta}) = \sigma^2 \mu(x|\boldsymbol{\lambda}) L_{11}^{-1}(x|\boldsymbol{\vartheta}) \tag{A.36}$$

$$h_{10}(x, y|\boldsymbol{\vartheta}) = \sigma^2 \mu(x|\boldsymbol{\lambda}) L_4^{-1}(x, y|\boldsymbol{\vartheta}) \tag{A.37}$$

$$h_{20}(x, y|\boldsymbol{\vartheta}) = \sigma^4 \mu(x|\boldsymbol{\lambda}) \mu(y|\boldsymbol{\lambda}) L_4^{-2}(x, y|\boldsymbol{\vartheta}) \tag{A.38}$$

Parts of the density function in the bivariate model

$$\begin{aligned}
f_X^*(x, y, 0, 0|\boldsymbol{\vartheta}) &= g_{0,1}(x, y|\boldsymbol{\vartheta}) f_A(x)\, \mathrm{I}(x = y) \\
f_X^*(x, y, 0, 1|\boldsymbol{\vartheta}) &= g_{1,1}(y, x|\boldsymbol{\vartheta}) f_A(x)\, \mathrm{I}(y \leq x) \\
f_X^*(x, y, 1, 0|\boldsymbol{\vartheta}) &= g_{1,1}(x, y|\boldsymbol{\vartheta}) f_A(y)\, \mathrm{I}(x \leq y) \\
f_X^*(x, y, 1, 1|\boldsymbol{\vartheta}) &= g_{2,1}(x, y|\boldsymbol{\vartheta}) S_A(\max(x, y))
\end{aligned} \tag{A.39}$$

A.2 List of symbols and abbreviations

ADE, ACE, AE, DE, CE, DCE, E	genetic models (see page 18)
AIC	Akaike Information Criterion (see page 18)
a.s.	almost surely
DZ	dizygotic
gg, ge, eg, ee	codes for different combinations of genetic models in the simulation study in Chapter 7
iff	if and only if
i.i.d.	independent, identical distributed
MLE	maximum likelihood estimate
MZ	monozygotic
a^2, d^2, c^2, e^2	proportion of the parts of the phenotypic variance (see page 21)
A, C	age at the time of observation, censoring time
A_T, A_C	age at left truncation/right censoring
b_1	empirical estimate of the skewness (see page 16)
b_2	empirical estimate of the kurtosis (see page 16)
C_i, $\tilde{C}_i$	current health status (see pages 32 and 99)
EX, $E_{\theta_0}X$	expectation of a random variable X
f_A, f_X, f_Z, f_θ, $f(.\|\theta)$	probability density function
F, F_n	distribution function
H, H_0, H_i	cumulative hazard function
H_0, H_1	null hypothesis, alternative hypothesis
I, $[a,b]$	interval of real numbers
$I(\theta) = \big(I_{ij}(\theta)\big)_{i=1,\dots,k}$	Fisher's information matrix
$L(\boldsymbol{x}\|\theta)$	likelihood function
m_k	$= \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})$, k-th standardized moment
n, N	number of observations, sample size
$N(0, \Sigma(\theta_0))$	multivariate normal distribution with mean vector zero and covariance matrix $\Sigma(\theta_0)$
P	probability measure
P_θ, P_Z, Q_θ	probability distribution
$P_X(.\|Z = z)$	conditional distribution of X given $Z = z$
R, $\hat{R}$	probandwise concordance rate and its empirical estimate (see page 22)
S_A, S_L, S_X, S_0, S_i	univariate survival function

S_T, S_{L*}, S_D, S_{0*}	bivariate survival function (see page 29)
$\hat{S}(t)$	Kaplan-Meier Estimator
$S(x\|Z)$, $S_i(x\|Z_i)$	conditional survival function
T, T_i, T_{i0}	survival time
T_h, T_{MZ}, T_{DZ}	threshold value
$U(\theta, \delta)$	$= \{\tilde{\theta} : \|\tilde{\theta} - \theta\| \leq \delta\}$, closed neighborhood of θ
X, X_i, Y, Y_i, Y_{ij}	random variable
Z, Z_i, Z_{ij}	frailty variable
β_1	skewness (see page 16)
β_2	kurtosis (see page 16)
$\Gamma(k, \lambda)$	Gamma-distribution with shape parameter k and scale parameter λ (see page 6)
Δ, Δ_i	censoring indicator
θ, θ_i, ϑ, ϑ_i, λ_i	one-dimensional parameter
$\boldsymbol{\theta}$, $\boldsymbol{\vartheta}$, $\boldsymbol{\lambda}$	multi-dimensional parameter vector
$\hat{\theta}_n$, $\hat{\theta}_n(X_1, \ldots, X_n)$	MLE of θ
$\hat{\boldsymbol{\theta}}_n$, $\hat{\boldsymbol{\theta}}_n(X_1, \ldots, X_n)$	MLE of $\boldsymbol{\theta}$
Θ, Θ_r, Θ_k	parameter set
Θ°	open kernel of the set Θ
λ^k	Lebesgue measure on $(\mathbb{R}^k, \mathfrak{B}^k)$
μ, μ_0, μ_i	baseline hazard function
μ, $\hat{\mu}$, $\tilde{\mu}$, $\bar{\mu}$, μ^*	σ-finite measure
ϱ_{MZ}, ϱ_{DZ}	correlation of a phenotypic trait in MZ/DZ twins
ϱ_0, ϱ_1, ϱ, ϱ^*	correlation of frailty variables
σ^2, σ_0^2, σ_i^2	variance of frailty variables
σ_P^2	phenotypic variance
σ_A^2, σ_D^2, σ_C^2, σ_E^2	parts of the phenotypic variance (see page 18)
$\Sigma(\theta_0)$	covariance matrix
Φ	distribution function of the standard normal distribution
$(\Omega, \mathfrak{F}, P)$	probability space
$\mathfrak{A}$, $\mathfrak{A}^*$, $\mathfrak{F}$	σ-algebra
$\mathfrak{A}_X \otimes \mathfrak{A}_Y$	product σ-algebra
$\mathfrak{A}_\Theta$	σ-algebra generated by the Borel set Θ
$\mathfrak{B}^k$	σ-algebra of the k-dimensional Borel sets $\subset \mathbb{R}^k$
$\mathfrak{B}_+^k$	σ-algebra of the k-dimensional Borel sets $\subset \mathbb{R}_+^k$
$\mathfrak{H}$, $\mathfrak{H}_0$	family of cumulative hazard functions
$\mathfrak{P}$, $\mathfrak{Q}$	family of probability distributions

$(\mathfrak{X}, \mathfrak{A})$, $(\mathfrak{X}^*, \mathfrak{A}^*)$	measurable space
$(\mathfrak{X}_X \times \mathfrak{X}_Y, \mathfrak{A}_X \otimes \mathfrak{A}_Y)$	product space
$(\mathfrak{X}^n, \mathfrak{A}^n)$	n-fold product space
$\mathcal{L}(v) = Ee^{-vZ}$	Laplace transform of the univariate random variable Z
$\mathcal{P}^{\{0,1,2\}}$	set of all subsets of $\{0, 1, 2\}$
$\mathbb{R}^k$	k-dimensional real vectors
$\mathbb{R}^k_+$	non-negative k-dimensional real vectors
$\mathrm{corr}(X, Y)$	correlation of the random variable X and Y
$\mathrm{cov}(X, Y)$	covariance of the random variable X and Y
$\mathrm{I}(x \in A)$, $\mathrm{I}_A(x)$	indicator function with respect to the set A
$\mathrm{odds}(x)$	$= \frac{x}{1-x}$, odds ratio of x
$\mathrm{Var}(X)$	variance of a random variable X
$X \sim Q_\theta$	random variable X has probability distribution Q_θ
$V_n \xrightarrow{P} V$	convergence in probability
$V_n \xrightarrow{\mathcal{L}} V$	convergence in distribution
$\frac{\partial}{\partial \theta_i}$, $\frac{\partial^2}{\partial \theta_i \partial \theta_j}$	partial derivative
$f(x)\big\|_{x=a}$	value of the function f at $x = a$
$f'(x)$	derivative of f with respect to x
$(\theta_1, \ldots, \theta_k)'$	transpose of a vector
$\square$	end of proof, q.e.d.

Appendix B

Lemmata and Proofs

B.1 Calculation of the likelihood function for current health status data

Proof of proposition 2

1.

$$\begin{aligned} P(X_1 > x, & X_2 > y, C_1 = 0, C_2 = 0) \\ &= P(A > x, A > y, T_{10} > A, T_{20} > A, T_1 > A, T_2 > A) \\ &= \int \mathrm{I}_{(x,\infty)\cap(y,\infty)}(a)\, S(a,a,a,a)\, f_A(a)\, da \\ &= \int \mathrm{I}_{(x,\infty)}(x_1)\, \mathrm{I}_{(y,\infty)}(x_2) f_X(x_1, x_2, 0, 0)\, d\mu \end{aligned}$$

2.

$$\begin{aligned} P(X_1 > x, & X_2 > y, C_1 = 1, C_2 = 0) \\ &= P(A > x, A > y, T_{10} \leq A, T_{20} > A, T_1 > A, T_2 > A) \\ &= \int \mathrm{I}_{(x,\infty)\cap(y,\infty)}(a) \left[S(0,a,a,a) - S(a,a,a,a) \right] f_A(a)\, da \\ &= \int \mathrm{I}_{(x,\infty)}(x_1)\, \mathrm{I}_{(y,\infty)}(x_2) f_X(x_1, x_2, 1, 0)\, d\mu \end{aligned}$$

3.

$$\begin{aligned}
&P(X_1 > x, X_2 > y, C_1 = 1, C_2 = 1) \\
&\quad = P(A > x, A > y, T_{10} \le A, T_{20} \le A, T_1 > A, T_2 > A) \\
&\quad = \int \mathrm{I}_{(x,\infty)\cap(y,\infty)}(a) \left[S(0,0,a,a) - S(a,0,a,a) \right. \\
&\qquad\qquad \left. - S(0,a,a,a) + S(a,a,a,a) \right] f_A(a)\, da \\
&\quad = \int \mathrm{I}_{(x,\infty)}(x_1)\, \mathrm{I}_{(y,\infty)}(x_2) f_X(x_1, x_2, 1, 1)\, d\mu
\end{aligned}$$

4.

$$\begin{aligned}
&P(X_1 > x, X_2 > y, C_1 = 2, C_2 = 0) \\
&\quad = P(A > y, T_{20} > A, A \ge T_1 > x, T_2 > A) \\
&= \iint \mathrm{I}_{(x,\infty)}(t)\, \mathrm{I}_{(y,\infty)}(a)\, \mathrm{I}(t \le a) \left[-\frac{\partial S(s_1, s_2, t_1, t_2)}{\partial t_1}\bigg|_{\substack{s_1=0, s_2=a \\ t_1=t, t_2=a}} \right] f_A(a)\, da\, dt \\
&\quad = \int \mathrm{I}_{(x,\infty)}(x_1)\, \mathrm{I}_{(y,\infty)}(x_2) f_X(x_1, x_2, 2, 0)\, d\mu
\end{aligned}$$

5.

$$\begin{aligned}
&P(X_1 > x, X_2 > y, C_1 = 2, C_2 = 1) \\
&\quad = P(A > y, T_{20} \le A, A \ge T_1 > x, T_2 > A) \\
&\quad = \iint \mathrm{I}_{(x,\infty)}(t)\, \mathrm{I}_{(y,\infty)}(a)\, \mathrm{I}(t \le a) \\
&\quad \cdot \left[\frac{\partial S(s_1, s_2, t_1, t_2)}{\partial t_1}\bigg|_{\substack{s_1=0, s_2=a \\ t_1=t, t_2=a}} - \frac{\partial S(s_1, s_2, t_1, t_2)}{\partial t_1}\bigg|_{\substack{s_1=s_2=0 \\ t_1=t, t_2=a}} \right] f_A(a)\, da\, dt \\
&\quad = \int \mathrm{I}_{(x,\infty)}(x_1)\, \mathrm{I}_{(y,\infty)}(x_2) f_X(x_1, x_2, 2, 1)\, d\mu
\end{aligned}$$

6.

$$\begin{aligned}
&P(X_1 > x, X_2 > y, C_1 = 2, C_2 = 2) \\
&\quad = P(A \ge T_1 > x, A \ge T_2 > y) \\
&= \iint \mathrm{I}_{(x,\infty)}(t_1)\, \mathrm{I}_{(y,\infty)}(t_2) S_A(\max(t_1, t_2)) \frac{\partial^2 S(s_1, s_2, t_1, t_2)}{\partial t_1 \partial t_2}\bigg|_{s_1=s_2=0} dt_1\, dt_2 \\
&\quad = \int \mathrm{I}_{(x,\infty)}(x_1)\, \mathrm{I}_{(y,\infty)}(x_2) f_X(x_1, x_2, 2, 2)\, d\mu
\end{aligned}$$

□

B.2 Consistency

In the following lemmata some inequalities are given that are used in the proof of consistency of MLEs in the four-dimensional correlated frailty model.

Lemma 14. *Let $\alpha_1, \alpha_2, \ldots, \alpha_N$ be some arbitrary non-negative real numbers, then it holds that:*

$$\ln\Big(1 + \sum_{i=1}^{N} \alpha_i\Big) \leq \sum_{i=1}^{N} \ln(1 + \alpha_i) \; .$$

Lemma 15. *Let α and k be some arbitrary real numbers with $0 \leq \alpha < 1$ and $k > 0$, then it holds that:*

$$|\ln(1 - \alpha^k)| \leq (k+1)\,|\ln(1-\alpha)| + |\ln k|.$$

Proof. The two cases $k \geq 1$ and $k < 1$ are investigated separately.

$\mathbf{k \geq 1}$: The function $f(\alpha) = 1 - \alpha^k - (1-\alpha)^k$ is monotone increasing for $0 \leq \alpha \leq \frac{1}{2}$ and monotone decreasing for $1 \geq \alpha \geq \frac{1}{2}$. Since $f(0) = f(1) = 0$ it follows that $f(\alpha) \geq 0$ for $0 \leq \alpha \leq 1$. Hence $(1-\alpha)^k \leq 1 - \alpha^k$ which is equivalent to

$$|\ln(1 - \alpha^k)| \leq k\,|\ln(1-\alpha)| \leq (k+1)\,|\ln(1-\alpha)| + |\ln k| \; .$$

$\mathbf{k < 1}$: The function $f(\alpha) = 1 - \alpha^k - k(1-\alpha)$ is monotone decreasing for $0 \leq \alpha \leq 1$ and $f(1) = 0$. Therefore $f(\alpha) \geq 0$ for $0 \leq \alpha \leq 1$ which means $1 - \alpha^k \geq k\,(1-\alpha)$. It follows that

$$|\ln(1 - \alpha^k)| \leq |\ln(1-\alpha)| - \ln k \leq (k+1)\,|\ln(1-\alpha)| + |\ln k| \; .$$

□

Lemma 16. *Let α and k be some arbitrary real numbers with $0 \leq \alpha \leq 1$ and $k > 0$, then it holds that:*

$$1 - \alpha^k \leq (k+1)(1-\alpha) \; .$$

Proof. The cases $k < 1$ and $k \geq 1$ are investigated separately.

$\mathbf{k \geq 1}$: The function $f(\alpha) = k(1-\alpha) - 1 + \alpha^k$ is monotone decreasing for $0 \leq \alpha \leq 1$ and $f(1) = 0$. Hence

$$1 - \alpha^k \leq k(1-\alpha) \leq (k+1)(1-\alpha) \; .$$

$\mathbf{k < 1}$: It is easy to verify that

$$1 - \alpha^k \leq 1 - \alpha \leq (k+1)(1-\alpha) \; .$$

□

Proof of theorem 4

The finiteness of the single integrals $I(i,j)$ is now proven. To make the proof easier to understand the dependence of the functions on the parameter vectors $\boldsymbol{\theta}$ or $\boldsymbol{\lambda}$ is omitted where it is obvious and would not be informative for the conclusions. Let $\boldsymbol{\theta}_0 = (\tilde{k}_0, \tilde{k}_1, \tilde{k}_3, \tilde{k}_4, \tilde{k}_{10}, \tilde{k}_{11}, \boldsymbol{\lambda}_0')'$ be an arbitrary element of Θ (the true parameter) and set $\tilde{\sigma}_0 = (\tilde{k}_0 + \tilde{k}_1 + \tilde{k}_3 + \tilde{k}_{10})^{-1/2}$ and $\tilde{\sigma} = (\tilde{k}_0 + \tilde{k}_1 + \tilde{k}_4 + \tilde{k}_{11})^{-1/2}$.

$\mathbf{I(0,0)}$: Using (A.9) one can estimate for $x_1 = x_2 = x$

$$|\ln f_X(x,x,0,0|\boldsymbol{\theta})| \leq |\ln f_{00}(x,x|\boldsymbol{\theta})| + |\ln f_A(x)|$$

and (A.18) leads to

$$\begin{aligned}|\ln f_{00}(x_1,x_2|\boldsymbol{\theta})| \leq\ & k_{10}\big(\ln L_{10}(x_1) + \ln L_{10}(x_2)\big) + k_{11}\big(\ln L_{11}(x_1) + \ln L_{11}(x_2)\big) \\ & + k_1\big(\ln L_1(x_1,x_1) + \ln L_1(x_2,x_2)\big) + k_3 \ln L_3(x_1,x_2) \\ & + k_4 \ln L_4(x_1,x_2) + k_0 \ln L_0(x_1,x_2,x_1,x_2)\ .\end{aligned}$$

Applying (A.2)-(A.7), lemma 14, the definition of Θ_k and (4.4) yields

$$\begin{aligned}|\ln f_{00}(x_1,x_2|\boldsymbol{\theta})| &\leq 4\,M_k\big(\ln L_{10}(x_1) + \ln L_{10}(x_2) + \ln L_{11}(x_1) + \ln L_{11}(x_2)\big) \\ &\leq 4\,M_k\big(\ln[1 + M_s^2 H_0(x_1|\boldsymbol{\lambda})] + \ln[1 + M_s^2 H_0(x_2|\boldsymbol{\lambda})] \\ &\qquad + \ln[1 + M_s^2 H(x_1|\boldsymbol{\lambda})] + \ln[1 + M_s^2 H(x_2|\boldsymbol{\lambda})]\big)\ .\end{aligned} \tag{B.1}$$

From (A.18) it follows that

$$f_{00}(x_1,x_2|\boldsymbol{\theta}) \leq \min\big(S_0(x_1|\sigma_0,\boldsymbol{\lambda}), S_0(x_2|\sigma_0,\boldsymbol{\lambda}), S_L(x_1|\sigma,\boldsymbol{\lambda}), S_L(x_2|\sigma,\boldsymbol{\lambda})\big)$$

and therefore using assumptions 4(c) and 4(d) one can conclude

$$\sup_{\boldsymbol{\theta}\in\Theta} |\ln f_{00}(x_1,x_2|\boldsymbol{\theta})| \cdot f_{00}(x_1,x_2|\boldsymbol{\theta}_0) \leq 16\,M_k C_0 = C_{00}$$

and thus

$$\sup_{\boldsymbol{\theta}\in\Theta} |\ln f_X(x,x,0,0|\boldsymbol{\theta})| \cdot f_X(x,x,0,0|\boldsymbol{\theta}_0) \leq \big(C_{00} + |\ln f_A(x)|\big) f_A(x)\ .$$

Finally, using assumption 6(b), one gets

$$I(0,0) \leq C_{00} + \int |\ln f_A(x)| f_A(x)\,dx < \infty\ .$$

$\mathbf{I}(\mathbf{1},\mathbf{0}) = \mathbf{I}(\mathbf{0},\mathbf{1})$: From (A.10), (A.18) and (A.19) it follows that

$$|\ln f_X(x,x,1,0|\boldsymbol{\theta})| \leq |\ln f_{10}(x,x|\boldsymbol{\theta})| + |\ln[1-T(x,x|\boldsymbol{\theta})]| + |\ln f_A(x)| \tag{B.2}$$

where

$$\begin{aligned} T(x_1,x_2|\boldsymbol{\theta}) &= \frac{f_{00}(x_1,x_2|\boldsymbol{\theta})}{f_{10}(x_1,x_2|\boldsymbol{\theta})} \\ &= L_{10}^{-k_{10}}(x_1)\left[\frac{L_{10}(x_2)}{L_3(x_1,x_2)}\right]^{k_3}\left[\frac{L_{11}(x_1)}{L_1(x_1,x_1)}\right]^{k_1}\left[\frac{L_0(0,x_2,x_1,x_2)}{L_0(x_1,x_2,x_1,x_2)}\right]^{k_0}. \end{aligned} \tag{B.3}$$

Hence $T(x_1,x_2|\boldsymbol{\theta}) \leq \left[\frac{L_0(0,x_2,x_1,x_2)}{L_0(x_1,x_2,x_1,x_2)}\right]^{m_k} \leq 1$ and with lemma 15 one gets

$$\begin{aligned} |\ln[1-T(x_1,x_2|\boldsymbol{\theta})]| &\leq -\ln\left(1-\left[\frac{L_0(0,x_2,x_1,x_2)}{L_0(x_1,x_2,x_1,x_2)}\right]^{m_k}\right) \\ &\leq -(m_k+1)\ln\left(1-\frac{L_0(0,x_2,x_1,x_2)}{L_0(x_1,x_2,x_1,x_2)}\right) + |\ln m_k| \\ &= (m_k+1)\left|\ln\frac{\sigma_0^2 H_0(x_1|\boldsymbol{\lambda})}{L_0(x_1,x_2,x_1,x_2)}\right| + |\ln m_k| \\ &\leq (m_k+1)\big(2\,|\ln\sigma_0| + |\ln H_0(x_1|\boldsymbol{\lambda})| + \ln L_0(x_1,x_2,x_1,x_2) + |\ln m_k|. \end{aligned}$$

Using the notation $\bar{C}_{10} = 2\,(m_k+1)(|\ln m_s| + |\ln M_s|) + |\ln m_k|$, (4.4), (A.2) and lemma 14 now yields

$$\begin{aligned} |\ln[1-T(x_1,x_2|\boldsymbol{\theta})]| \leq \bar{C}_{10} + (m_k+1)\big(&|\ln H_0(x_1|\boldsymbol{\lambda})| + \ln[1+M_s^2 H_0(x_1|\boldsymbol{\lambda})] \\ &+ \ln[1+M_s^2 H_0(x_2|\boldsymbol{\lambda})] + \ln[1+M_s^2 H(x_1|\boldsymbol{\lambda})] + \ln[1+M_s^2 H(x_2|\boldsymbol{\lambda})]\big). \end{aligned} \tag{B.4}$$

Analogously to (B.1) one can estimate

$$\begin{aligned} |\ln f_{10}(x_1,x_2|\boldsymbol{\theta})| \leq 4\,M_k\big(&\ln[1+M_s^2 H_0(x_2|\boldsymbol{\lambda})] + \ln[1+M_s^2 H(x_1|\boldsymbol{\lambda})] \\ &+ \ln[1+M_s^2 H(x_2|\boldsymbol{\lambda})]\big)\,. \end{aligned} \tag{B.5}$$

Since $\frac{L_3(x_1,x_2)}{L_{10}(x_1)L_{10}(x_2)} \leq 1$, $\frac{L_1(x_1,x_1)}{L_{10}(x_1)L_{11}(x_1)} \leq 1$ and $\frac{L_0(x_1,x_2,x_1,x_2)}{L_{10}(x_1)L_0(0,x_2,x_1,x_2)} \leq 1$, from (B.3) and (4.3) it follows that

$$T(x_1,x_2|\boldsymbol{\theta}) \geq [L_{10}(x_1)]^{-k_0-k_1-k_3-k_{10}} = S_0(x_1|\sigma_0,\boldsymbol{\lambda})\,. \tag{B.6}$$

In addition it holds that

$$f_{10}(x_1,x_2|\boldsymbol{\theta}) \leq \min\big(S_0(x_2|\sigma_0,\boldsymbol{\lambda}), S_L(x_1|\sigma,\boldsymbol{\lambda}), S_L(x_2|\sigma,\boldsymbol{\lambda})\big)\,. \tag{B.7}$$

From (B.4), (B.6) and (B.7) one can conclude

$$\begin{aligned}\sup_{\boldsymbol{\theta}\in\Theta}|\ln[1-T(x,x|\boldsymbol{\theta})]|\cdot\big(f_{10}(x,x|\boldsymbol{\theta}_0)-f_{00}(x,x|\boldsymbol{\theta}_0)\big) &\le \bar{C}_{10}\\ &+(m_k+1)\sup_{\boldsymbol{\lambda}\in\Lambda}|\ln H_0(x|\boldsymbol{\lambda})|\cdot(1-S_0(x|\tilde{\sigma}_0,\boldsymbol{\lambda}_0)\,S_0(x|\tilde{\sigma}_0,\boldsymbol{\lambda}_0)\\ &+(m_k+1)\sup_{\boldsymbol{\lambda}\in\Lambda}\ln[1+M_s^2H_0(x|\boldsymbol{\lambda})]\cdot S_0(x|\tilde{\sigma}_0,\boldsymbol{\lambda}_0)\\ &+2\,(m_k+1)\sup_{\boldsymbol{\lambda}\in\Lambda}\ln[1+M_s^2H(x|\boldsymbol{\lambda})]\cdot S_L(x|\tilde{\sigma},\boldsymbol{\lambda}_0)\end{aligned}$$

and thus assumptions 4(c), 4(d) and 4(e) yield that

$$\begin{aligned}\sup_{\boldsymbol{\theta}\in\Theta}|\ln[1-T(x,x|\boldsymbol{\theta})]|\cdot\big(f_{10}(x,x|\boldsymbol{\theta}_0)-f_{00}(x,x|\boldsymbol{\theta}_0)\big)&\\ \le \bar{C}_{10}+4\,(m_k+1)C_0=C_{10}\,.&\end{aligned}\tag{B.8}$$

From (B.5), (B.7) and assumptions 4(c) and 4(d) it follows that

$$\sup_{\boldsymbol{\theta}\in\Theta}|\ln f_{10}(x_1,x_2|\boldsymbol{\theta})|\cdot f_{10}(x_1,x_2|\boldsymbol{\theta}_0)\le 12\,M_kC_0\,.\tag{B.9}$$

Finally from (A.10), (B.2), (B.8) and (B.9) one gets

$$\sup_{\boldsymbol{\theta}\in\Theta}|\ln f_X(x,x,1,0|\boldsymbol{\theta})|\cdot f_X(x,x,1,0|\boldsymbol{\theta}_0)\le\big(12\,M_kC_0+C_{10}+|\ln f_A(x)|\big)f_A(x)$$

and therefore

$$I(1,0)\le 12\,M_kC_0+C_{10}+\int|\ln f_A(x)|f_A(x)\,dx<\infty$$

if assumption 6(b) is taken into account.

$\mathbf{I(1,1)}$: Using (A.12), (A.19), (A.21) and (B.3) one can write

$$f_X(x_1,x_2,1,1|\boldsymbol{\theta})=f_{11}(x_1,x_2|\boldsymbol{\theta})\,R(x_1,x_2|\boldsymbol{\theta})\,f_A(x_1)\,\mathrm{I}(x_1=x_2)\tag{B.10}$$

where

$$R(x_1,x_2|\boldsymbol{\theta})=1-2\,\tilde{T}(x_1,x_2|\boldsymbol{\theta})+T(x_1,x_2|\boldsymbol{\theta})\,\tilde{T}(x_1,x_2|\boldsymbol{\theta})$$

and

$$\begin{aligned}\tilde{T}(x_1,x_2|\boldsymbol{\theta})&=\frac{f_{10}(x_1,x_2|\boldsymbol{\theta})}{f_{11}(x_1,x_2|\boldsymbol{\theta})}\\ &=L_{10}^{-k_{10}-k_3}(x_2)\left[\frac{L_{11}(x_2)}{L_1(x_2,x_2)}\right]^{k_1}\left[\frac{L_4(x_1,x_2)}{L_0(0,x_2,x_1,x_2)}\right]^{k_0}.\end{aligned}\tag{B.11}$$

Since $\frac{L_3(x,x)}{L_{10}^2(x)} \leq 1$ and $\frac{L_4(x,x)L_0(x,x,x,x)}{L_0^2(0,x,x,x)} \leq 1$ it holds that

$$\frac{\tilde{T}(x,x|\boldsymbol{\theta})}{T(x,x|\boldsymbol{\theta})} = \left[\frac{L_3(x,x)}{L_{10}^2(x)}\right]^{k_3}\left[\frac{L_4(x,x)L_0(x,x,x,x)}{L_0^2(0,x,x,x)}\right]^{k_0} \leq 1\,.$$

It follows that $\tilde{T}(x,x|\boldsymbol{\theta}) \leq T(x,x|\boldsymbol{\theta}) \leq 1$ and

$$[1-\tilde{T}(x,x|\boldsymbol{\theta})]^2 \leq R(x,x|\boldsymbol{\theta}) \leq 1$$

and therefore

$$|\ln R(x,x|\boldsymbol{\theta})| = -\ln R(x,x|\boldsymbol{\theta}) \leq 2\,|\ln[1-\tilde{T}(x,x|\boldsymbol{\theta})]|\,.$$

From (B.11) one can conclude

$$\tilde{T}(x_1,x_2|\boldsymbol{\theta}) \leq \left[\frac{L_4(x_1,x_2)}{L_0(0,x_2,x_1,x_2)}\right]^{m_k} \leq 1$$

and using lemma 15 one can then estimate

$$\begin{aligned}
|\ln[1-\tilde{T}(x_1,x_2|\boldsymbol{\theta})]| &\leq -\ln\left(1-\left[\frac{L_4(x_1,x_2)}{L_0(0,x_2,x_1,x_2)}\right]^{m_k}\right)\\
&\leq (m_k+1)\left|\ln\left[1-\frac{L_4(x_1,x_2)}{L_0(0,x_2,x_1,x_2)}\right]\right| + |\ln m_k|\\
&= (m_k+1)\left|\ln\frac{\sigma_0^2 H_0(x_2|\boldsymbol{\lambda})}{L_0(0,x_2,x_1,x_2)}\right| + |\ln m_k|\\
&\leq (m_k+1)\big(2\,|\ln\sigma_0| + |\ln H_0(x_2|\boldsymbol{\lambda})| + \ln L_0(0,x_2,x_1,x_2) + |\ln m_k|.
\end{aligned}$$

Using again the notation $\bar{C}_{10} = 2\,(m_k+1)(|\ln m_s| + |\ln M_s|) + |\ln m_k|$, (4.4),(A.2) and lemma 14 now yields

$$\begin{aligned}
|\ln[1-\tilde{T}(x_1,x_2|\boldsymbol{\theta})]| \leq \bar{C}_{10} + (m_k+1)\big(&|\ln H_0(x_2|\boldsymbol{\lambda})| + \ln[1+M_s^2 H_0(x_2|\boldsymbol{\lambda})]\\
&+\ln[1+M_s^2 H(x_1|\boldsymbol{\lambda})] + \ln[1+M_s^2 H(x_2|\boldsymbol{\lambda})]\big)
\end{aligned} \tag{B.12}$$

Applying (A.21) and lemma 14 it is easy to verify that

$$|\ln f_{11}(x_1,x_2|\boldsymbol{\theta})| \leq 4\,M_k\big(\ln[1+M_s^2 H(x_1|\boldsymbol{\lambda})] + \ln[1+M_s^2 H(x_2|\boldsymbol{\lambda})]\big) \tag{B.13}$$

and

$$f_{11}(x_1,x_2|\boldsymbol{\theta}) \leq \min\big(S_L(x_1|\sigma,\boldsymbol{\lambda}), S_L(x_2|\sigma,\boldsymbol{\lambda})\big)\,. \tag{B.14}$$

Since $\frac{L_1(x_1,x_2)}{L_{10}(x_2)L_{11}(x_2)} \leq 1$ and $\frac{L_0(0,x_2,x_1,x_2)}{L_{10}(x_2)L_4(x_1,x_2)} \leq 1$ using (B.11) and (4.3) one can estimate

$$\tilde{T}(x_1, x_2) \geq [L_{10}(x_2)]^{-k_0-k_1-k_3-k_{10}} = S_0(x_2|\sigma_0, \boldsymbol{\lambda}) \tag{B.15}$$

and therefore

$$R(x, x|\boldsymbol{\theta}) \leq 1 - \tilde{T}(x, x|\boldsymbol{\theta}) \leq 1 - S_0(x|\sigma_0, \boldsymbol{\lambda}) \; . \tag{B.16}$$

From (B.12), (B.14) and (B.16) it follows that

$$\begin{aligned}
\sup_{\boldsymbol{\theta}\in\Theta} |\ln[1 - \tilde{T}(x, x|\boldsymbol{\theta})]| \cdot f_{11}(x, x|\boldsymbol{\theta}_0)\, R(x, x|\boldsymbol{\theta}_0) &\leq \bar{C}_{10} \\
&+ (m_k + 1) \sup_{\boldsymbol{\lambda}\in\Lambda} |\ln H_0(x|\boldsymbol{\lambda})| \cdot (1 - S_0(x|\tilde{\sigma}_0, \boldsymbol{\lambda}_0))\, S_L(x|\tilde{\sigma}, \boldsymbol{\lambda}_0) \\
&+ (m_k + 1) \sup_{\boldsymbol{\lambda}\in\Lambda} \ln[1 + M_s^2 H_0(x|\boldsymbol{\lambda})] \cdot S_L(x|\tilde{\sigma}, \boldsymbol{\lambda}_0) \\
&+ 2\,(m_k + 1) \sup_{\boldsymbol{\lambda}\in\Lambda} \ln[1 + M_s^2 H(x|\boldsymbol{\lambda})] \cdot S_L(x|\tilde{\sigma}, \boldsymbol{\lambda}_0)
\end{aligned}$$

and thus assumptions 4(c), 4(d) and 4(e) yield that

$$\begin{aligned}
2 \sup_{\boldsymbol{\theta}\in\Theta} |\ln[1 - \tilde{T}(x, x|\boldsymbol{\theta})]| \cdot f_{11}(x, x|\boldsymbol{\theta}_0)\, R(x, x|\boldsymbol{\theta}_0) & \\
\leq 2\,\bar{C}_{10} + 8\,(m_k + 1)C_0 = C_{11} \; . &
\end{aligned} \tag{B.17}$$

From (B.13), (B.14) and assumption 4(d) one can conclude that

$$\begin{aligned}
\sup_{\boldsymbol{\theta}\in\Theta} |\ln f_{11}(x_1, x_2|\boldsymbol{\theta})| \cdot f_{11}(x_1, x_2|\boldsymbol{\theta}_0) & \\
\leq 4M_k \big(\sup_{\boldsymbol{\lambda}\in\Lambda} \ln[1 + M_s^2 H(x_1|\boldsymbol{\lambda})] \cdot S_L(x_1|\tilde{\sigma}, \boldsymbol{\lambda}_0) & \\
+ \sup_{\boldsymbol{\lambda}\in\Lambda} \ln[1 + M_s^2 H(x_2|\boldsymbol{\lambda})] \cdot S_L(x_2|\tilde{\sigma}, \boldsymbol{\lambda}_0) \big) \leq 8\, M_k C_0 \; . &
\end{aligned} \tag{B.18}$$

Finally from (B.10), (B.17) and (B.18) one gets

$$\sup_{\boldsymbol{\theta}\in\Theta} |\ln f_X(x, x, 1, 1|\boldsymbol{\theta})| \cdot f_X(x, x, 1, 1|\boldsymbol{\theta}_0) \leq \big(8\, M_k C_0 + C_{11} + |\ln f_A(x)|\big) f_A(x)$$

and, using assumption 6(b), that yields

$$I(1, 1) \leq 8\, M_k C_0 + C_{11} + \int |\ln f_A(x)| f_A(x)\, dx < \infty \; .$$

$\mathbf{I(2, 0) = I(0, 2)}$: From (A.13) and (A.22) it follows that

$$f_X(x_1, x_2, 2, 0|\boldsymbol{\theta}) = f_{10}(x_1, x_2|\boldsymbol{\theta})\, \mu(x_1|\boldsymbol{\lambda})\, R_1(x_1, x_2|\boldsymbol{\theta})\, f_A(x_2)\, \mathrm{I}(x_1 \leq x_2) \tag{B.19}$$

where

$$R_1(x_1, x_2|\boldsymbol{\theta}) = \sigma^2 \left\{ \frac{k_{11} + k_1}{L_{11}(x_1)} + \frac{k_4}{L_4(x_1, x_2)} + \frac{k_0}{L_0(0, x_2, x_1, x_2)} \right\} . \tag{B.20}$$

Using (4.3), (A.2), (A.5) and (A.7) one can conclude

$$\begin{aligned} R_1(x_1, x_2|\boldsymbol{\theta}) &= \frac{1 - (k_4 + k_0)\sigma^2}{L_{11}(x_1)} + \frac{k_4\sigma^2}{L_4(x_1, x_2)} + \frac{k_0\sigma^2}{L_0(0, x_2, x_1, x_2)} \\ &\leq \frac{1}{L_{11}(x_1)} \leq 1 . \end{aligned} \tag{B.21}$$

This yields

$$\begin{aligned} |\ln R_1(x_1, x_2|\boldsymbol{\theta})| = -\ln R_1(x_1, x_2|\boldsymbol{\theta}) &\leq -\ln \frac{m_k\sigma^2}{L_0(0, x_2, x_1, x_2)} \\ &\leq |\ln m_k| + 2\,|\ln \sigma| + \ln L_0(0, x_2, x_1, x_2) . \end{aligned}$$

With the notation $C_{20} = |\ln m_k| + 2\,|\ln m_s| + 2\,|\ln M_s|$ and applying (A.2) and lemma 14 one can thus estimate

$$\begin{aligned} |\ln R_1(x_1, x_2|\boldsymbol{\theta})| \leq C_{20} &+ \ln[1 + M_s^2 H_0(x_2|\boldsymbol{\lambda})] + \ln[1 + M_s^2 H(x_1|\boldsymbol{\lambda})] \\ &+ \ln[1 + M_s^2 H(x_2|\boldsymbol{\lambda})] . \end{aligned}$$

From (B.19) and (B.5) it therefore follows for $x_1 \leq x_2$:

$$\begin{aligned} |\ln f_X(x_1, x_2, 2, 0|\boldsymbol{\theta})| \leq C_{20} &+ |\ln \mu(x_1|\boldsymbol{\lambda})| + (1 + 4\,M_k)\ln[1 + M_s^2 H_0(x_2|\boldsymbol{\lambda})] \\ &+ (1 + 4\,M_k)\big(\ln[1 + M_s^2 H(x_1|\boldsymbol{\lambda})] + \ln[1 + M_s^2 H(x_2|\boldsymbol{\lambda})]\big) + |\ln f_A(x_2)| . \end{aligned}$$

Using (B.19), (B.21), (B.7) and assumption 4(a) leads to

$$\begin{aligned} &f_X(x_1, x_2, 2, 0|\boldsymbol{\theta}) \\ &\qquad \leq C_0 \min\big(S_0(x_2|\sigma_0, \boldsymbol{\lambda}), S_L(x_1|\sigma, \boldsymbol{\lambda}), S_L(x_2|\sigma, \boldsymbol{\lambda})\big) f_A(x_2)\, \mathrm{I}(x_1 \leq x_2) . \end{aligned}$$

Hence it holds for $x_1 \leq x_2$ that

$$\begin{aligned} \sup_{\boldsymbol{\theta}\in\Theta} |\ln f_X(x_1, x_2, 2, 0|\boldsymbol{\theta})| \cdot f_X(x_1, x_2, 2, 0|\boldsymbol{\theta}_0) \leq C_0 \Big[& C_{20} + |\ln f_A(x_2)| \\ &+ \sup_{\boldsymbol{\lambda}\in\Lambda} |\ln \mu(x_1|\boldsymbol{\lambda})| \cdot S_L(x_1|\tilde{\sigma}, \boldsymbol{\lambda}_0) \\ &+ (1 + 4\,M_k) \sup_{\boldsymbol{\lambda}\in\Lambda} \ln[1 + M_s^2 H_0(x_2|\boldsymbol{\lambda})] \cdot S_0(x_2|\tilde{\sigma}_0, \boldsymbol{\lambda}_0) \\ &+ (1 + 4\,M_k) \sup_{\boldsymbol{\lambda}\in\Lambda} \ln[1 + M_s^2 H(x_1|\boldsymbol{\lambda})] \cdot S_L(x_1|\tilde{\sigma}, \boldsymbol{\lambda}_0) \\ &+ (1 + 4\,M_k) \sup_{\boldsymbol{\lambda}\in\Lambda} \ln[1 + M_s^2 H(x_2|\boldsymbol{\lambda})] \cdot S_L(x_2|\tilde{\sigma}, \boldsymbol{\lambda}_0) \Big] f_A(x_2) \end{aligned}$$

and assumptions 4(b), 4(c) and 4(d) yield that for $x_1 \leq x_2$

$$\sup_{\boldsymbol{\theta}\in\Theta} |\ln f_X(x_1, x_2, 2, 0|\boldsymbol{\theta})| \cdot f_X(x_1, x_2, 2, 0|\boldsymbol{\theta}_0) \leq C_0 \left(C_{20} + (4 + 12\, M_k)C_0 + |\ln f_A(x_2)|\right) f_A(x_2) \; .$$

Finally, using assumptions 6(a) and 6(c), we get

$$I(2,0) \leq [C_{20} + (4 + 12\, M_k)C_0]C_0\, EA + C_0 \int x\, |\ln f_A(x)| f_A(x)\, dx < \infty \; .$$

$\mathbf{I(2,1) = I(1,2)}$: Using (A.15), (A.22) and (A.24) one can write

$$f_X(x_1, x_2, 2, 1|\boldsymbol{\theta}) = f_{11}(x_1, x_2|\boldsymbol{\theta})\, \mu(x_1|\boldsymbol{\lambda})\, \tilde{R}(x_1, x_2|\boldsymbol{\theta})\, f_A(x_2)\, \mathrm{I}(x_1 \leq x_2) \tag{B.22}$$

with

$$\tilde{R}(x_1, x_2|\boldsymbol{\theta}) = R_2(x_1, x_2|\boldsymbol{\theta}) - \tilde{T}(x_1, x_2|\boldsymbol{\theta})\, R_1(x_1, x_2|\boldsymbol{\theta}) \tag{B.23}$$

where $\tilde{T}(x_1, x_2|\boldsymbol{\theta})$ and $R_1(x_1, x_2|\boldsymbol{\theta})$ are given by (B.11) and (B.20), respectively, and it holds that

$$R_2(x_1, x_2|\boldsymbol{\theta}) = \sigma^2 \left\{ \frac{k_{11} + k_1}{L_{11}(x_1)} + \frac{k_4 + k_0}{L_4(x_1, x_2)} \right\} \; . \tag{B.24}$$

From (4.3), (A.5) and (A.7) it follows that

$$R_2(x_1, x_2|\boldsymbol{\theta}) = \frac{1 - (k_4 + k_0)\sigma^2}{L_{11}(x_1)} + \frac{(k_4 + k_0)\sigma^2}{L_4(x_1, x_2)} \leq \frac{1}{L_{11}(x_1)} \leq 1 \; . \tag{B.25}$$

Thus (B.23) yields

$$\tilde{R}(x_1, x_2|\boldsymbol{\theta}) \leq R_2(x_1, x_2|\boldsymbol{\theta}) \leq 1 \; . \tag{B.26}$$

In addition, from (B.20) and (B.24) one can conclude that

$$\begin{aligned} \hat{R}(x_1, x_2|\boldsymbol{\theta}) &= R_2(x_1, x_2|\boldsymbol{\theta}) - R_1(x_1, x_2|\boldsymbol{\theta}) \\ &= k_0 \left[\frac{1}{L_4(x_1, x_2)} - \frac{1}{L_0(0, x_2, x_1, x_2)} \right] \geq 0 \end{aligned} \tag{B.27}$$

and therefore

$$\begin{aligned}\tilde{R}(x_1, x_2|\boldsymbol{\theta}) &= [1 - \tilde{T}(x_1, x_2|\boldsymbol{\theta})]\, R_2(x_1, x_2|\boldsymbol{\theta}) + \tilde{T}(x_1, x_2|\boldsymbol{\theta})\, \hat{R}(x_1, x_2|\boldsymbol{\theta}) \\ &\geq [1 - \tilde{T}(x_1, x_2|\boldsymbol{\theta})]\, R_2(x_1, x_2|\boldsymbol{\theta})\ .\end{aligned} \tag{B.28}$$

Together with (B.26) we get

$$|\ln \tilde{R}(x_1, x_2|\boldsymbol{\theta})| \leq |\ln[1 - \tilde{T}(x_1, x_2|\boldsymbol{\theta})]| + |\ln R_2(x_1, x_2|\boldsymbol{\theta})|\ .$$

From (B.24) it follows that

$$\begin{aligned}|\ln R_2(x_1, x_2|\boldsymbol{\theta})| = -\ln R_2(x_1, x_2|\boldsymbol{\theta}) &\leq -\ln \frac{m_k \sigma^2}{L_4(x_1, x_2)} \\ &\leq |\ln m_k| + 2\,|\ln \sigma| + \ln L_4(x_1, x_2)\ .\end{aligned}$$

Using the notation $C_{20} = |\ln m_k| + 2\,|\ln m_s| + 2\,|\ln M_s|$ again and applying lemma 14 and (A.5) one can thus estimate

$$|\ln R_2(x_1, x_2|\boldsymbol{\theta})| \leq C_{20} + \ln[1 + M_s^2 H(x_1|\boldsymbol{\lambda})] + \ln[1 + M_s^2 H(x_2|\boldsymbol{\lambda})]\ . \tag{B.29}$$

Hence using (B.12) one can conclude

$$\begin{aligned}|\ln \tilde{R}(x_1, x_2|\boldsymbol{\theta})| \leq \bar{C}_{10} + C_{20} &+ (m_k + 1)\big(|\ln H_0(x_2|\boldsymbol{\lambda})| + \ln[1 + M_s^2 H_0(x_2|\boldsymbol{\lambda})]\big) \\ &+ (m_k + 2)\big(\ln[1 + M_s^2 H(x_1|\boldsymbol{\lambda})] + \ln[1 + M_s^2 H(x_2|\boldsymbol{\lambda})]\big)\ .\end{aligned} \tag{B.30}$$

Since $L_{10}(x_2) L_4(x_1, x_2) \geq L_0(0, x_2, x_1, x_2)$ and applying lemma 16, from (B.27) we get

$$\begin{aligned}\hat{R}(x_1, x_2|\boldsymbol{\theta}) &= \frac{k_0}{L_4(x_1, x_2)} \left[1 - \frac{L_4(x_1, x_2)}{L_0(0, x_2, x_1, x_2)}\right] \\ &\leq \frac{M_k}{L_{11}(x_1)} \left[1 - \frac{1}{L_{10}(x_2)}\right] = \frac{M_k}{L_{11}(x_1)}[1 - S_0^{\sigma_0^2}(x_2|\sigma_0, \boldsymbol{\lambda})] \\ &\leq \frac{M_k(M_s^2 + 1)}{L_{11}(x_1)}[1 - S_0(x_2|\sigma_0, \boldsymbol{\lambda})]\ .\end{aligned} \tag{B.31}$$

From (B.15), (B.25), (B.28) and (B.31) it follows that

$$\tilde{R}(x_1, x_2|\boldsymbol{\theta}) \leq (1 - S_0(x_2|\sigma_0, \boldsymbol{\lambda}))\, \frac{1 + M_k(M_s^2 + 1)}{L_{11}(x_1)}$$

and this together with (B.14), (B.22) and assumption 4(a) leads to

$$f_X(x_1,x_2,2,1|\boldsymbol{\theta}) \le \hat{C}_{21}\,(1-S_0(x_2|\sigma_0,\boldsymbol{\lambda})) \cdot \min\big(S_L(x_1|\sigma,\boldsymbol{\lambda}),S_L(x_2|\sigma,\boldsymbol{\lambda})\big) f_A(x_2)\,\mathrm{I}(x_1\le x_2)\,. \quad \text{(B.32)}$$

where $\hat{C}_{21} = C_0\,[1+M_k(M_s^2+1)]$. The inequalities (B.30) and (B.32) yield

$$\begin{aligned}\sup_{\boldsymbol{\theta}\in\Theta}|\ln\tilde{R}(x_1,x_2|\boldsymbol{\theta})|\cdot f_X(x_1,x_2,2,1|\boldsymbol{\theta}_0) &\le \hat{C}_{21} f_A(x_2)\,\mathrm{I}(x_1\le x_2)\\ \cdot\Big[&(m_k+1)\sup_{\boldsymbol{\lambda}\in\Lambda}|\ln H_0(x_2|\boldsymbol{\lambda})|\cdot(1-S_0(x_2|\tilde{\sigma}_0,\boldsymbol{\lambda}_0))S_L(x_2|\tilde{\sigma},\boldsymbol{\lambda}_0)\\ &+(m_k+1)\sup_{\boldsymbol{\lambda}\in\Lambda}\ln[1+M_s^2H_0(x_2|\boldsymbol{\lambda})]\cdot S_L(x_2|\tilde{\sigma},\boldsymbol{\lambda}_0)\\ &+(m_k+2)\sup_{\boldsymbol{\lambda}\in\Lambda}\ln[1+M_s^2H(x_1|\boldsymbol{\lambda})]\cdot S_L(x_1|\tilde{\sigma},\boldsymbol{\lambda}_0)\\ &+(m_k+2)\sup_{\boldsymbol{\lambda}\in\Lambda}\ln[1+M_s^2H(x_2|\boldsymbol{\lambda})]\cdot S_L(x_2|\tilde{\sigma},\boldsymbol{\lambda}_0)+\bar{C}_{10}+C_{20}\Big]\end{aligned}$$

and using assumptions 4(c), 4(d) and 4(e) one can thus conclude that

$$\begin{aligned}&\sup_{\boldsymbol{\theta}\in\Theta}|\ln\tilde{R}(x_1,x_2|\boldsymbol{\theta})|\cdot f_X(x_1,x_2,2,1|\boldsymbol{\theta}_0)\\ &\quad\le \hat{C}_{21}[\bar{C}_{10}+C_{20}+2(2\,m_k+3)C_0]f_A(x_2)\,\mathrm{I}(x_1\le x_2)\\ &\quad= C_{21}\,f_A(x_2)\,\mathrm{I}(x_1\le x_2)\,.\end{aligned}$$

From (B.18) and (B.32) it follows that

$$\sup_{\boldsymbol{\theta}\in\Theta}|\ln f_{11}(x_1,x_2|\boldsymbol{\theta})|\cdot f_X(x_1,x_2,2,1|\boldsymbol{\theta}_0) \le 8\,M_kC_0\,\hat{C}_{21}\,f_A(x_2)\,\mathrm{I}(x_1\le x_2)\,.$$

We can now, using (B.22), (B.32) and assumption 4(b), estimate

$$\begin{aligned}&\sup_{\boldsymbol{\theta}\in\Theta}|\ln f_X(x_1,x_2,2,1|\boldsymbol{\theta})|\cdot f_X(x_1,x_2,2,1|\boldsymbol{\theta}_0)\\ &\qquad\le (8\,M_kC_0\hat{C}_{21}+C_0+C_{21}+\hat{C}_{21}|\ln f_A(x_2)|)\,f_A(x_2)\end{aligned}$$

for $x_1\le x_2$, which finally yields

$$I(2,1)\le (8\,M_kC_0\hat{C}_{21}+C_0+C_{21})EA+\hat{C}_{21}\int x\,|\ln f_A(x)|f_A(x)\,dx<\infty$$

if we take assumptions 6(a) and 6(c) into account.

$\mathbf{I(2,2)}$: From (A.17) and (A.26) it follows that

$$f_X(x_1,x_2,2,2|\boldsymbol{\theta}) = f_{11}(x_1,x_2|\boldsymbol{\theta})\,\mu(x_1|\boldsymbol{\lambda})\,\mu(x_2|\boldsymbol{\lambda})R_3(x_1,x_2|\boldsymbol{\theta})S_A(\max(x_1,x_2)) \quad \text{(B.33)}$$

with

$$R_3(x_1, x_2|\boldsymbol{\theta}) = \frac{\sigma^4(k_0 + k_4)}{L_4^2(x_1, x_2)} + R_2(x_1, x_2|\boldsymbol{\theta}) \cdot R_2(x_2, x_1|\boldsymbol{\theta}) \; .$$

The function $R_2(x_1, x_2|\boldsymbol{\theta})$ is given in (B.24). Using (4.3), (A.5) and (A.7) one can verify that

$$\frac{(m_k + m_k^2)m_s^4}{L_4^2(x_1, x_2)} \leq R_3(x_1, x_2|\boldsymbol{\theta}) \leq \frac{M_s^2 + 4}{L_{11}(x_1)L_{11}(x_2)} \; . \tag{B.34}$$

Hence it holds that

$$\begin{aligned} |\ln R_3(x_1, x_2|\boldsymbol{\theta})| &\leq \left|\ln \frac{(m_k + m_k^2)m_s^4}{L_4^2(x_1, x_2)}\right| + \left|\ln \frac{M_s^2 + 4}{L_{11}(x_1)L_{11}(x_2)}\right| \\ &\leq |\ln[(m_k + m_k^2)m_s^4]| + |\ln(M_s^2 + 4)| + 2\,|\ln L_4(x_1, x_2)| \\ &\quad + |\ln L_{11}(x_1)| + |\ln L_{11}(x_2)| \; . \end{aligned}$$

Using the notation $\bar{C}_{22} = |\ln[(m_k + m_k^2)m_s^4]| + |\ln(M_s^2 + 4)|$ and applying lemma 14 yields

$$|\ln R_3(x_1, x_2|\boldsymbol{\theta})| \leq \bar{C}_{22} + 3\,\ln[1 + M_s^2 H(x_1|\boldsymbol{\lambda})] + 3\,\ln[1 + M_s^2 H(x_2|\boldsymbol{\lambda})] \; . \tag{B.35}$$

From (B.34) and assumption 4(a) it follows that

$$\mu(x_1|\boldsymbol{\lambda})\,\mu(x_2|\boldsymbol{\lambda}) R_3(x_1, x_2|\boldsymbol{\theta}) \leq C_0^2(M_s^2 + 4)$$

and applying this to (B.33) leads to the estimation

$$f_X(x_1, x_2, 2, 2|\boldsymbol{\theta}) \leq C_0^2(M_s^2 + 4) f_{11}(x_1, x_2|\boldsymbol{\theta})\, S_A(\max(x_1, x_2)) \; . \tag{B.36}$$

Using (B.14) and (B.35) one can thus conclude

$$\begin{aligned} \sup_{\boldsymbol{\theta}\in\Theta} |\ln R_3(x_1, x_2|\boldsymbol{\theta})| \cdot f_X(x_1, x_2, 2, 2|\boldsymbol{\theta}_0) &\leq C_0^2(M_s^2 + 4)\, S_A(\max(x_1, x_2)) \\ &\cdot \Big[\bar{C}_{22} + 3 \sup_{\boldsymbol{\lambda}\in\Lambda} \ln[1 + M_s^2 H(x_1|\boldsymbol{\lambda})] \cdot S_L(x_1|\tilde{\sigma}, \boldsymbol{\lambda}_0) \\ &\qquad + 3 \sup_{\boldsymbol{\lambda}\in\Lambda} \ln[1 + M_s^2 H(x_2|\boldsymbol{\lambda})] \cdot S_L(x_2|\tilde{\sigma}, \boldsymbol{\lambda}_0)\Big] \; . \end{aligned}$$

Applying assumption 4(d) yields

$$\sup_{\boldsymbol{\theta}\in\Theta} |\ln R_3(x_1, x_2|\boldsymbol{\theta})|\, f_X(x_1, x_2, 2, 2|\boldsymbol{\theta}_0) \leq C_{22}\, S_A(\max(x_1, x_2)) \tag{B.37}$$

where $C_{22} = C_0^2(M_s^2+4)(\bar{C}_{22}+6C_0)$. Using assumption 4(b) and (B.14) one can estimate

$$\sup_{\boldsymbol{\lambda}\in\Lambda} |\ln \mu(x_1|\boldsymbol{\lambda})| \cdot f_{11}(x_1,x_2|\boldsymbol{\theta}_0) \le C_0 \tag{B.38}$$

From (B.18), (B.33), (B.36), (B.37) and (B.38) it follows that

$$\begin{aligned}\sup_{\boldsymbol{\theta}\in\Theta} &|\ln f_X(x_1,x_2,2,2|\boldsymbol{\theta})| \cdot f_X(x_1,x_2,2,2|\boldsymbol{\theta}_0) \\ &\le C_0^2(M_s^2+4)\big(\hat{C}_{22} + |\ln S_A(\max(x_1,x_2))|\big) S_A(\max(x_1,x_2)). \end{aligned}\tag{B.39}$$

with $\hat{C}_{22} = 8M_kC_0 + 2\,C_0 + C_{22}$. Since $\iint S_A(\max(x_1,x_2))\,dx_1dx_2 = EA^2$ it therefore holds that

$$\begin{aligned} I(2,2) &\le C_0^2(M_s^2+4)\hat{C}_{22}EA^2 \\ &\quad + C_0^2(M_s^2+4)\iint |\ln S_A(\max(x_1,x_2))|\, S_A(\max(x_1,x_2))\,dx_1dx_2 < \infty \end{aligned}$$

where assumptions 6(a) and 6(d) are taken into account. □

Proof of theorem 5

The finiteness of the single integrals $\tilde{I}(i,j)$ will now be proven.

$\tilde{\mathbf{I}}(\mathbf{0},\mathbf{0})$: From corollary 3, (B.14), (B.18) and (4.7) it follows that

$$\sup_{\boldsymbol{\vartheta}\in\hat{\Theta}} |\ln f_X^*(x,x,0,0|\boldsymbol{\vartheta})| \cdot f_X^*(x,x,0,0|\boldsymbol{\vartheta}_0) \le \big(8\,M_kC_0 + |\ln f_A(x)|\big) f_A(x)$$

and therefore assumption 6(b) yields

$$\tilde{I}(0,0) \le 8\,M_kC_0 + \int |\ln f_A(x)| f_A(x)\,dx < \infty\ .$$

$\tilde{\mathbf{I}}(\mathbf{1},\mathbf{0}) = \tilde{\mathbf{I}}(\mathbf{0},\mathbf{1})$: Using (4.8) and (A.23) one can write

$$f_X^*(x_1,x_2,1,0|\boldsymbol{\vartheta}) = f_{11}(x_1,x_2|\boldsymbol{\theta})\,\mu(x_1|\boldsymbol{\lambda})\,R_2(x_1,x_2|\boldsymbol{\theta})\,f_A(x_2)\,\mathrm{I}(x_1\le x_2)$$

where $R_2(x_1,x_22|\boldsymbol{\theta})$ is given in (B.24). From (B.14), (B.25) and assumption 4(a) it follows that

$$f_X^*(x_1,x_2,1,0|\boldsymbol{\vartheta}) \le C_0 \min\big(S_L(x_1|\sigma,\boldsymbol{\lambda}), S_L(x_2|\sigma,\boldsymbol{\lambda})\big) f_A(x_2)\,\mathrm{I}(x_1\le x_2)\ .$$

In addition from (B.13), (B.29) and corollary 3 one can conclude that for $x_1 \leq x_2$:

$$\sup_{\boldsymbol{\vartheta}\in\hat{\Theta}} |\ln f_X^*(x_1, x_2, 0, 0|\boldsymbol{\vartheta})| \leq C_{20} + \sup_{\boldsymbol{\lambda}\in\Lambda} |\ln \mu(x_1|\boldsymbol{\lambda})| + |\ln f_A(x_2)| + (4\,M_k + 1)\big(\sup_{\boldsymbol{\lambda}\in\Lambda} \ln[1 + M_s^2 H(x_1|\boldsymbol{\lambda})] + \sup_{\boldsymbol{\lambda}\in\Lambda} \ln[1 + M_s^2 H(x_2|\boldsymbol{\lambda})]\big) .$$

Therefore assumptions 4(b) and 4(d) yield for $x_1 \leq x_2$

$$\sup_{\boldsymbol{\vartheta}\in\hat{\Theta}} |\ln f_X^*(x_1, x_2, 1, 0|\boldsymbol{\vartheta})| \cdot f_X^*(x_1, x_2, 1, 0|\boldsymbol{\vartheta}_0) \leq C_0\big(C_{20} + (8\,M_k + 3)C_0 + |\ln f_A(x_2)|\big) f_A(x_2) .$$

Finally we get

$$\tilde{I}(1, 0) \leq C_0[C_{20} + (8\,M_k + 3)C_0]EA + C_0 \int x\,|\ln f_A(x)| f_A(x)\,dx < \infty$$

if we take assumptions 6(a) and 6(c) into account.

$\tilde{\mathbf{I}}(\mathbf{1}, \mathbf{1})$: From (4.9) and (A.17) it follows that

$$f_X^*(x_1, x_2, 1, 1|\boldsymbol{\vartheta}) = f_X(x_1, x_2, 2, 2|\boldsymbol{\theta}) .$$

Thus using corollary 3 and (B.39) one can conclude

$$\sup_{\boldsymbol{\vartheta}\in\hat{\Theta}} |\ln f_X^*(x_1, x_2, 1, 1|\boldsymbol{\vartheta})| \cdot f_X^*(x_1, x_2, 1, 1|\boldsymbol{\vartheta}_0) \leq C_0^2(M_s^2 + 4)\big(\hat{C}_{22} + |\ln S_A(\max(x_1, x_2))|\big) S_A(\max(x_1, x_2)) .$$

Using assumptions 6(a) and 6(d) this yields

$$\tilde{I}(1, 1) \leq C_0^2(M_s^2 + 4)\hat{C}_{22}EA^2 + C_0^2(M_s^2 + 4) \iint |\ln S_A(\max(x_1, x_2))| S_A(\max(x_1, x_2))\,dx_1 dx_2 < \infty .$$

□

B.3 Asymptotic normality in a bivariate correlated frailty model

First some basic corollaries and lemmata that are applied in the proof of asymptotic normality of maximum likelihood estimates in the bivariate correlated gamma frailty model are presented.

Corollary 6. *From* (A.5) *and* (A.7) *it follows that*

$$1 \leq L_{11}(x|\boldsymbol{\vartheta}) \leq L_4(x,y|\boldsymbol{\vartheta}) \;, \tag{B.40}$$

$$0 \leq \ln L_4(x,y|\boldsymbol{\vartheta}) \leq \ln L_{11}(x|\boldsymbol{\vartheta}) + \ln L_{11}(y|\boldsymbol{\vartheta}) \text{ and} \tag{B.41}$$

$$\Big|\frac{\partial \ln L_4(x,y|\boldsymbol{\vartheta})}{\partial \vartheta_i}\Big| \leq \Big|\frac{\partial \ln L_{11}(x|\boldsymbol{\vartheta})}{\partial \vartheta_i}\Big| + \Big|\frac{\partial \ln L_{11}(y|\boldsymbol{\vartheta})}{\partial \vartheta_i}\Big| \tag{B.42}$$

for $i \in \{1,\ldots,K_p\}$.

Corollary 7. *From* (A.39) *it follows that*

$$\begin{aligned}
\Big|\frac{\partial}{\partial \vartheta_i} f_X^*(x,y,0,0|\boldsymbol{\vartheta})\Big| &\leq \Big|\frac{\partial}{\partial \vartheta_i} g_{0,1}(x,y|\boldsymbol{\vartheta})\Big| \cdot f_A(x)\,\mathrm{I}(x=y)\;,\\
\Big|\frac{\partial^2}{\partial \vartheta_i \partial \vartheta_j} f_X^*(x,y,0,0|\boldsymbol{\vartheta})\Big| &\leq \Big|\frac{\partial^2}{\partial \vartheta_i \partial \vartheta_j} g_{0,1}(x,y|\boldsymbol{\vartheta})\Big| \cdot f_A(x)\,\mathrm{I}(x=y)\;,\\
\Big|\frac{\partial}{\partial \vartheta_i} f_X^*(x,y,1,0|\boldsymbol{\vartheta})\Big| &\leq \Big|\frac{\partial}{\partial \vartheta_i} g_{1,1}(x,y|\boldsymbol{\vartheta})\Big| \cdot f_A(y)\,\mathrm{I}(x\leq y)\;,\\
\Big|\frac{\partial^2}{\partial \vartheta_i \partial \vartheta_j} f_X^*(x,y,1,0|\boldsymbol{\vartheta})\Big| &\leq \Big|\frac{\partial^2}{\partial \vartheta_i \partial \vartheta_j} g_{1,1}(x,y|\boldsymbol{\vartheta})\Big| \cdot f_A(y)\,\mathrm{I}(x\leq y)\;,\\
\Big|\frac{\partial}{\partial \vartheta_i} f_X^*(x,y,1,1|\boldsymbol{\vartheta})\Big| &\leq \Big|\frac{\partial}{\partial \vartheta_i} g_{2,1}(x,y|\boldsymbol{\vartheta})\Big| \cdot S_A(\max(x,y)) \text{ and}\\
\Big|\frac{\partial^2}{\partial \vartheta_i \partial \vartheta_j} f_X^*(x,y,1,1|\boldsymbol{\vartheta})\Big| &\leq \Big|\frac{\partial^2}{\partial \vartheta_i \partial \vartheta_j} g_{2,1}(x,y|\boldsymbol{\vartheta})\Big| \cdot S_A(\max(x,y))
\end{aligned}$$

for $i,j \in \{1,\ldots,K_u\}$.

Lemma 17. *Let* $f(x|\boldsymbol{\theta})$ *be a family of density functions dominated by a measure* μ. *Let* $x \in \mathfrak{X}$ *from a measurable space* $(\mathfrak{X},\mathfrak{A})$ *and let* $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^k$. *If for all* $i,j \in \{1,\ldots,k\}$ *and all* $\boldsymbol{\theta} \in \Theta$ *the partial derivatives* $\frac{\partial f(x|\boldsymbol{\theta})}{\partial \theta_i}$ *and* $\frac{\partial^2 f(x|\boldsymbol{\theta})}{\partial \theta_i \partial \theta_j}$ *exist and are continuous, and if there exist two measurable functions* $G(x)$ *and* $H(x)$ *such that for all* $\boldsymbol{\theta} \in \Theta$, *all* $x \in \mathfrak{X}$ *and all* $i,j \in \{1,\ldots,k\}$

$$\Big|\frac{\partial f(x|\boldsymbol{\theta})}{\partial \theta_i}\Big| \leq G(x) \quad \text{and} \quad \Big|\frac{\partial^2 f(x|\boldsymbol{\theta})}{\partial \theta_i \partial \theta_j}\Big| \leq H(x)\;,$$

where $\int G(x)\,d\mu < \infty$ *and* $\int H(x)\,d\mu < \infty$, *then for all* $i,j \in \{1,\ldots,k\}$ *it holds that*

$$\begin{aligned}
\int \frac{\partial f(x|\boldsymbol{\theta})}{\partial \theta_i}\,d\mu &= \frac{\partial}{\partial \theta_i}\int f(x|\boldsymbol{\theta})\,d\mu = 0 \quad \text{and}\\
\int \frac{\partial^2 f(x|\boldsymbol{\theta})}{\partial \theta_i \partial \theta_j}\,d\mu &= \frac{\partial^2}{\partial \theta_i \partial \theta_j}\int f(x|\boldsymbol{\theta})\,d\mu = 0\;.
\end{aligned}$$

Proof. Without loss of generality let $\boldsymbol{\theta} = \theta$ be one-dimensional. Let $g_n(x|\theta) = \frac{f(x|\theta+h_n)-f(x|\theta)}{h_n}$ for a null sequence (h_n). Then $g_n(x) \xrightarrow{n\to\infty} \frac{\partial f(x|\theta)}{\partial\theta}$ and

$$g_n(x|\theta) = \frac{\partial f(x|\theta_n^*)}{\partial\theta}$$

for some intermediate value θ_n^*. Therefore $|g_n(x|\theta)| \leq G(x)$ for all x and one can apply Lebesgue's Dominated Convergence Theorem (see Witting, 1985). It follows that

$$\lim_{n\to\infty} \int g_n(x|\theta)\, d\mu = \int \lim_{n\to\infty} g_n(x|\theta)\, d\mu = \int \frac{\partial f(x|\theta)}{\partial\theta}\, d\mu$$

and

$$\int g_n(x|\theta)\, d\mu = \frac{1}{h_n}\left[\int f(x|\theta+h_n)\, d\mu - \int f(x|\theta)\, d\mu\right] = 0$$

yields the first assertion.
Now let $\boldsymbol{\theta} = (\theta_1, \theta_2)$ be two-dimensional and

$$\tilde{g}_n(x|\theta_1,\theta_2) = \frac{1}{h_n}\left[\frac{\partial f(x|\theta_1,\theta_2+h_n)}{\partial\theta_1} - \frac{\partial f(x|\theta_1,\theta_2)}{\partial\theta_1}\right] .$$

The second assertion can be proved completely analogous to the first one. □

Lemma 18. *Let $l_0(\boldsymbol{\vartheta})$ and $l_1(\boldsymbol{\vartheta})$ be defined by* (A.27)*. A constant K exists such that for all $\boldsymbol{\vartheta} \in \hat{\Theta}$, for $k,l \in \{1,\ldots,K_p\}$ and for $i = 0,1$ it holds that*

$$|l_i(\boldsymbol{\vartheta})| \leq K\ , \quad \left|\frac{\partial l_i(\boldsymbol{\vartheta})}{\partial\vartheta_k}\right| \leq K \quad \textit{and} \quad \left|\frac{\partial^2 l_i(\boldsymbol{\vartheta})}{\partial\vartheta_k\partial\vartheta_l}\right| \leq K\ . \tag{B.43}$$

Proof. l_0, l_1, $\frac{\partial l_0}{\partial\vartheta_k}$, $\frac{\partial l_1}{\partial\vartheta_k}$, $\frac{\partial^2 l_0}{\partial\vartheta_k\partial\vartheta_l}$ and $\frac{\partial^2 l_1}{\partial\vartheta_k\partial\vartheta_l}$ are continuous functions of σ and ϱ_1 and the parameter set $[m_s, M_s] \times [m_r, 1]$ is compact. □

Lemma 19. *Let $\mathfrak{H} = \{H(.|\boldsymbol{\lambda}) : \boldsymbol{\lambda} \in \Lambda\}$ be a family of cumulative hazard functions that fulfills assumptions 2 and 3. Let a family of univariate survival functions $\mathfrak{S}$ be defined according to* (4.10) *and let assumptions 7(a), 7(b) and 7(c) be fulfilled. Then a positive constant K_0 exists such that for all $\tilde{\boldsymbol{\vartheta}} \in \hat{\Theta}$, all $x \geq 0$ and for $i,j \in \{1,...,K_p\}$ it holds that*

$$\sup_{\boldsymbol{\vartheta}\in\hat{\Theta}} \ln L_{11}(x|\boldsymbol{\vartheta}) \cdot S_L^{\alpha}(x|\tilde{\boldsymbol{\vartheta}}) \leq K_0 \qquad \textit{for } \alpha = \tfrac{1}{2}, 1, \tag{B.44}$$

$$\sup_{\boldsymbol{\vartheta}\in\hat{\Theta}} \left|\frac{\partial \ln L_{11}(x|\boldsymbol{\vartheta})}{\partial\vartheta_i}\right| \cdot S_L^{\alpha}(x|\tilde{\boldsymbol{\vartheta}}) \leq K_0 \qquad \textit{for } \alpha = \tfrac{1}{4}, \tfrac{1}{2}, 1, \tag{B.45}$$

$$\sup_{\boldsymbol{\vartheta}\in\hat{\Theta}} \left|\frac{\partial^2 \ln L_{11}(x|\boldsymbol{\vartheta})}{\partial\vartheta_i\partial\vartheta_j}\right| \cdot S_L^{\alpha}(x|\tilde{\boldsymbol{\vartheta}}) \leq K_0 \qquad \textit{for } \alpha = \tfrac{1}{2}, 1. \tag{B.46}$$

Proof. The inequality (B.44) is a direct consequence of assumption 7(a). The derivatives have to be considered separately.

Inequality (B.45):

1. $\vartheta_i = \varrho_1$: $\dfrac{\partial \ln L_{11}(x|\boldsymbol{\vartheta})}{\partial \varrho_1} = 0$ thus $\sup\limits_{\boldsymbol{\vartheta}\in\hat{\Theta}} \left| \dfrac{\partial \ln L_{11}(x|\boldsymbol{\vartheta})}{\partial \varrho_1} \right| \cdot S_L^{\alpha}(x|\tilde{\boldsymbol{\vartheta}}) = 0.$

2. $\vartheta_i = \sigma$: $\dfrac{\partial \ln L_{11}(x|\boldsymbol{\vartheta})}{\partial \sigma} = \dfrac{2\sigma H(x|\boldsymbol{\lambda})}{1+\sigma^2 H(x|\boldsymbol{\lambda})} \Rightarrow \left| \dfrac{\ln L_{11}(x|\boldsymbol{\vartheta})}{\partial \sigma} \right| \leq \dfrac{2}{\sigma} \leq \dfrac{2}{m_s}$
   and
   $$\sup_{\boldsymbol{\vartheta}\in\hat{\Theta}} \left| \frac{\partial \ln L_{11}(x|\boldsymbol{\vartheta})}{\partial \sigma} \right| \cdot S_L^{\alpha}(x|\tilde{\boldsymbol{\vartheta}}) \leq \frac{2}{m_s} \ .$$

3. $\vartheta_i = \lambda_j$: $\dfrac{\partial \ln L_{11}(x|\boldsymbol{\vartheta})}{\partial \lambda_j} = \dfrac{\partial \ln H(x|\boldsymbol{\lambda})}{\partial \lambda_j} \cdot \dfrac{\sigma^2 H(x|\boldsymbol{\lambda})}{1+\sigma^2 H(x|\boldsymbol{\lambda})} \leq \dfrac{\partial \ln H(x|\boldsymbol{\lambda})}{\partial \lambda_j}$
   From assumption 7(b) it follows that
   $$\sup_{\boldsymbol{\vartheta}\in\hat{\Theta}} \left| \frac{\ln L_{11}(x|\boldsymbol{\vartheta})}{\partial \lambda_j} \right| \cdot S_L^{\alpha}(x|\tilde{\boldsymbol{\vartheta}}) \leq \sup_{\boldsymbol{\lambda}\in\Lambda} \left| \frac{\partial \ln H(x|\boldsymbol{\lambda})}{\partial \lambda_j} \right| \cdot S_L^{\alpha}(x|\tilde{\boldsymbol{\vartheta}}) \leq C_1 \ .$$

Inequality (B.46):

1. $\vartheta_i = \varrho_1$ or $\vartheta_j = \varrho_1$: Then $\dfrac{\partial^2 \ln L_{11}(x|\boldsymbol{\vartheta})}{\partial \vartheta_i \partial \vartheta_j} = 0$ and therefore
   $$\sup_{\boldsymbol{\vartheta}\in\hat{\Theta}} \left| \frac{\partial \ln L_{11}(x|\boldsymbol{\vartheta})}{\partial \vartheta_i \partial \vartheta_j} \right| \cdot S_L^{\alpha}(x|\tilde{\boldsymbol{\vartheta}}) = 0 \ .$$

2. $\vartheta_i = \sigma$ and $\vartheta_j = \sigma$: $\dfrac{\partial^2 \ln L_{11}(x|\boldsymbol{\vartheta})}{\partial \sigma^2} = \dfrac{2H(x|\boldsymbol{\lambda})}{1+\sigma^2 H(x|\boldsymbol{\lambda})} - \dfrac{4\sigma^2 H^2(x|\boldsymbol{\lambda})}{[1+\sigma^2 H(x|\boldsymbol{\lambda})]^2}$
   and hence
   $$\left| \frac{\partial^2 \ln L_{11}(x|\boldsymbol{\vartheta})}{\partial \sigma^2} \right| \leq \frac{2}{\sigma^2} + \frac{4}{\sigma^2} \leq \frac{6}{m_s^2}$$
   and
   $$\sup_{\boldsymbol{\vartheta}\in\hat{\Theta}} \left| \frac{\partial^2 \ln L_{11}(x|\boldsymbol{\vartheta})}{\partial \sigma^2} \right| \cdot S_L^{\alpha}(x|\tilde{\boldsymbol{\vartheta}}) \leq \frac{6}{m_s^2} \ .$$

3. $\vartheta_i = \sigma$ and $\vartheta_j = \lambda_k$: Since

$$\frac{\partial^2 \ln L_{11}(x|\boldsymbol{\vartheta})}{\partial\sigma\partial\lambda_k} = \left[\frac{\sigma^2 H(x|\boldsymbol{\lambda})}{1+\sigma^2 H(x|\boldsymbol{\lambda})} - \frac{\sigma^4 H^2(x|\boldsymbol{\lambda})}{[1+\sigma^2 H(x|\boldsymbol{\lambda})]^2}\right] \cdot \frac{2}{\sigma} \cdot \frac{\partial \ln H(x|\boldsymbol{\lambda})}{\partial\lambda_k}$$

it follows that $\left|\frac{\partial^2 \ln L_{11}(x|\boldsymbol{\vartheta})}{\partial\sigma\partial\lambda_k}\right| \leq \frac{4}{m_s}\left|\frac{\partial \ln H(x|\boldsymbol{\lambda})}{\partial\lambda_k}\right|$ and using assumption 7(b) yields

$$\sup_{\boldsymbol{\vartheta}\in\hat{\Theta}}\left|\frac{\partial^2 \ln L_{11}(x|\boldsymbol{\vartheta})}{\partial\sigma\partial\lambda_k}\right| \cdot S_L^{\alpha}(x|\tilde{\boldsymbol{\vartheta}}) \leq \frac{4}{m_s} C_1 \ .$$

4. $\vartheta_i = \lambda_k$ and $\vartheta_j = \lambda_l$: It holds that

$$\frac{\partial^2 \ln L_{11}(x|\boldsymbol{\vartheta})}{\partial\lambda_k\partial\lambda_l} = \frac{\partial^2 \ln H(x|\boldsymbol{\lambda})}{\partial\lambda_k\partial\lambda_l}\frac{\sigma^2 H(x|\boldsymbol{\lambda})}{1+\sigma^2 H(x|\boldsymbol{\lambda})} + \frac{\partial \ln H(x|\boldsymbol{\lambda})}{\partial\lambda_k}\frac{\partial \ln H(x|\boldsymbol{\lambda})}{\partial\lambda_l}\left[\frac{\sigma^2 H(x|\boldsymbol{\lambda})}{1+\sigma^2 H(x|\boldsymbol{\lambda})} - \frac{\sigma^4 H^2(x|\boldsymbol{\lambda})}{[1+\sigma^2 H(x|\boldsymbol{\lambda})]^2}\right]$$

and therefore

$$\left|\frac{\partial^2 \ln L_{11}(x|\boldsymbol{\vartheta})}{\partial\lambda_k\partial\lambda_l}\right| \leq \left|\frac{\partial^2 \ln H(x|\boldsymbol{\lambda})}{\partial\lambda_k\partial\lambda_l}\right| + 2\left|\frac{\partial \ln H(x|\boldsymbol{\lambda})}{\partial\lambda_k}\right|\left|\frac{\partial \ln H(x|\boldsymbol{\lambda})}{\partial\lambda_l}\right| .$$

Using the assumptions 7(b) and 7(c) it follows that

$$\sup_{\boldsymbol{\vartheta}\in\hat{\Theta}}\left|\frac{\partial^2 \ln L_{11}(x|\boldsymbol{\vartheta})}{\partial\lambda_k\partial\lambda_l}\right| \cdot S_L^{\alpha}(x|\tilde{\boldsymbol{\vartheta}}) \leq C_2 + C_1^2 \ .$$

□

Lemma 20. *Let $\mathfrak{H} = \{H(.|\boldsymbol{\lambda}) : \boldsymbol{\lambda} \in \Lambda\}$ be a family of cumulative hazard functions that fulfills assumption 4(a). The parameter set $\hat{\Theta}$ and the functions $g_{0,\alpha}(x,y|\boldsymbol{\vartheta})$, $h_1(x,y|\boldsymbol{\vartheta})$ and $h_2(x,y|\boldsymbol{\vartheta})$ are defined according to* (4.6), (A.31), (A.33) *and* (A.35), *respectively. Obviously $g_{0,\alpha}(x,y|\boldsymbol{\vartheta}) \leq 1$ holds for all $x,y \geq 0$, all $\boldsymbol{\vartheta} \in \hat{\Theta}$ and all α. In addition two positive constants K_1 and K_2 exist such that $h_i(x,y|\boldsymbol{\vartheta}) \leq K_i$, $i = 1,2$, for all $x,y \geq 0$ and all $\boldsymbol{\vartheta} \in \hat{\Theta}$.*

Proof. Using assumption 4(a), (B.40) and (B.43) one can conclude that $h_1(x,y|\boldsymbol{\vartheta}) \leq 2\,M_s^2 K C_0 = K_1$ and $h_2(x,y|\boldsymbol{\vartheta}) \leq M_s^2 K C_0^2 = K_2$. □

Lemma 9 is given on page 61. Some inequalities derived in the proof of this lemma are used in the proof of the subsequent lemmata. This is reason that the proof of lemma 9 is given first.

Proof of lemma 9

Since

$$\frac{\partial g_{0,\alpha}}{\partial \vartheta_i} = \frac{\partial \ln g_{0,\alpha}}{\partial \vartheta_i}\, g_{0,\alpha} = \alpha\, \frac{\partial \ln g_{0,1}}{\partial \vartheta_i}\, g_{0,\alpha}$$

and

$$\ln g_{0,1}(x,y|\boldsymbol{\vartheta}) = -l_1(\boldsymbol{\vartheta}) \ln L_{11}(x|\boldsymbol{\vartheta}) - l_1(\boldsymbol{\vartheta}) \ln L_{11}(y|\boldsymbol{\vartheta}) - l_0(\boldsymbol{\vartheta}) \ln L_4(x,y|\boldsymbol{\vartheta}) \tag{B.47}$$

using (B.41), (B.42) and (B.43) it follows that

$$\begin{aligned}
\Big|\frac{\partial g_{0,\alpha}(x,y|\boldsymbol{\vartheta})}{\partial \vartheta_i}\Big| &\le K\Big(\ln L_{11}(x|\boldsymbol{\vartheta}) + \ln L_{11}(y|\boldsymbol{\vartheta}) + \ln L_4(x,y|\boldsymbol{\vartheta}) \\
&\quad + \Big|\frac{\partial \ln L_{11}(x|\boldsymbol{\vartheta})}{\partial \vartheta_i}\Big| + \Big|\frac{\partial \ln L_{11}(y|\boldsymbol{\vartheta})}{\partial \vartheta_i}\Big| + \Big|\frac{\partial \ln L_4(x,y|\boldsymbol{\vartheta})}{\partial \vartheta_i}\Big|\Big) g_{0,\alpha}(x,y|\boldsymbol{\vartheta}) \\
&\le 2K\Big(\ln L_{11}(x|\boldsymbol{\vartheta}) + \ln L_{11}(y|\boldsymbol{\vartheta}) \\
&\qquad + \Big|\frac{\partial \ln L_{11}(x|\boldsymbol{\vartheta})}{\partial \vartheta_i}\Big| + \Big|\frac{\partial \ln L_{11}(y|\boldsymbol{\vartheta})}{\partial \vartheta_i}\Big|\Big) g_{0,\alpha}(x,y|\boldsymbol{\vartheta})
\end{aligned}$$

Applying $g_{0,\alpha}(x,y|\boldsymbol{\vartheta}) \le S_L^{\alpha}(x|\boldsymbol{\vartheta})$ and $g_{0,\alpha}(x,y|\boldsymbol{\vartheta}) \le S_L^{\alpha}(y|\boldsymbol{\vartheta})$, respectively, yields

$$\begin{aligned}
\Big|\frac{\partial g_{0,\alpha}(x,y|\boldsymbol{\vartheta})}{\partial \vartheta_i}\Big| &\le 2K\big(\ln L_{11}(x|\boldsymbol{\vartheta})\, S_L^{\alpha}(x|\boldsymbol{\vartheta}) + \ln L_{11}(y|\boldsymbol{\vartheta})\, S_L^{\alpha}(y|\boldsymbol{\vartheta})\big) \\
&\quad + 2K\Big(\Big|\frac{\partial \ln L_{11}(x|\boldsymbol{\vartheta})}{\partial \vartheta_i}\Big|\, S_L^{\alpha}(x|\boldsymbol{\vartheta}) + \Big|\frac{\partial \ln L_{11}(y|\boldsymbol{\vartheta})}{\partial \vartheta_i}\Big|\, S_L^{\alpha}(y|\boldsymbol{\vartheta})\Big)
\end{aligned}$$

and therefore from (B.44) and (B.45) it follows that

$$\Big|\frac{\partial g_{0,\alpha}(x,y|\boldsymbol{\vartheta})}{\partial \vartheta_i}\Big| \le 8KK_0 = c_0 \quad \text{for } \alpha = \tfrac{1}{2}, 1. \tag{B.48}$$

It holds that

$$\begin{aligned}
\frac{\partial^2 g_{0,1}}{\partial \vartheta_i \partial \vartheta_j} &= \Big[\frac{\partial^2 \ln g_{0,1}}{\partial \vartheta_i \partial \vartheta_j} + \frac{\partial \ln g_{0,1}}{\partial \vartheta_i} \cdot \frac{\partial \ln g_{0,1}}{\partial \vartheta_j}\Big] \cdot g_{0,1} \\
&= \frac{\partial^2 \ln g_{0,1}}{\partial \vartheta_i \partial \vartheta_j} \cdot g_{0,1} + 4\, \frac{\partial g_{0,1/2}}{\vartheta_i} \cdot \frac{\partial g_{0,1/2}}{\vartheta_j} \quad .
\end{aligned}$$

From (B.47) using (B.43) and (B.48) it follows that

$$\begin{aligned}\left|\frac{\partial^2 g_{0,1}(x,y|\boldsymbol{\vartheta})}{\partial\vartheta_i\partial\vartheta_j}\right| \leq K\, g_{0,1}(x,y|\boldsymbol{\vartheta}) \Bigg(&\ln L_{11}(x|\boldsymbol{\vartheta}) + \ln L_{11}(y|\boldsymbol{\vartheta}) + \ln L_4(x,y|\boldsymbol{\vartheta}) \\ &+ \left|\frac{\partial \ln L_{11}(x|\boldsymbol{\vartheta})}{\partial\vartheta_i}\right| + \left|\frac{\partial \ln L_{11}(y|\boldsymbol{\vartheta})}{\partial\vartheta_i}\right| + \left|\frac{\partial \ln L_4(x,y|\boldsymbol{\vartheta})}{\partial\vartheta_i}\right| \\ &+ \left|\frac{\partial \ln L_{11}(x|\boldsymbol{\vartheta})}{\partial\vartheta_j}\right| + \left|\frac{\partial \ln L_{11}(y|\boldsymbol{\vartheta})}{\partial\vartheta_j}\right| + \left|\frac{\partial \ln L_4(x,y|\boldsymbol{\vartheta})}{\partial\vartheta_j}\right| \\ &+ \left|\frac{\partial^2 \ln L_{11}(x|\boldsymbol{\vartheta})}{\partial\vartheta_i\partial\vartheta_j}\right| + \left|\frac{\partial^2 \ln L_{11}(y|\boldsymbol{\vartheta})}{\partial\vartheta_i\partial\vartheta_j}\right| + \left|\frac{\partial^2 \ln L_4(x,y|\boldsymbol{\vartheta})}{\partial\vartheta_i\partial\vartheta_j}\right|\Bigg) \\ &+ 4\,c_0^2\ .\end{aligned}$$

One can write

$$\begin{aligned}\frac{\partial^2 \ln L_4(x,y)}{\partial\vartheta_i\partial\vartheta_j} = &-\frac{1}{L_4^2(x,y)}\left(\frac{\partial L_{11}(x)}{\partial\vartheta_i} + \frac{\partial L_{11}(y)}{\partial\vartheta_i}\right)\left(\frac{\partial L_{11}(x)}{\partial\vartheta_j} + \frac{\partial L_{11}(y)}{\partial\vartheta_j}\right) \\ &+ \frac{L_{11}(x)}{L_4(x,y)}\left(\frac{\partial^2 \ln L_{11}(x)}{\partial\vartheta_i\partial\vartheta_j} + \frac{\partial \ln L_{11}(x)}{\partial\vartheta_i}\cdot\frac{\partial \ln L_{11}(x)}{\partial\vartheta_j}\right) \\ &+ \frac{L_{11}(y)}{L_4(x,y)}\left(\frac{\partial^2 \ln L_{11}(y)}{\partial\vartheta_i\partial\vartheta_j} + \frac{\partial \ln L_{11}(y)}{\partial\vartheta_i}\cdot\frac{\partial \ln L_{11}(y)}{\partial\vartheta_j}\right)\end{aligned}$$

and therefore it holds that

$$\begin{aligned}\left|\frac{\partial^2 \ln L_4(x,y)}{\partial\vartheta_i\partial\vartheta_j}\right| \leq &\left[\left|\frac{\partial \ln L_{11}(x)}{\partial\vartheta_i}\right| + \left|\frac{\partial \ln L_{11}(y)}{\partial\vartheta_i}\right|\right]\left[\left|\frac{\partial \ln L_{11}(x)}{\partial\vartheta_j}\right| + \left|\frac{\partial \ln L_{11}(y)}{\partial\vartheta_j}\right|\right] \\ &+ \left|\frac{\partial^2 \ln L_{11}(x)}{\partial\vartheta_i\partial\vartheta_j}\right| + \left|\frac{\partial^2 \ln L_{11}(y)}{\partial\vartheta_i\partial\vartheta_j}\right| \\ &+ \left|\frac{\partial \ln L_{11}(x)}{\partial\vartheta_i}\right|\cdot\left|\frac{\partial \ln L_{11}(x)}{\partial\vartheta_j}\right| + \left|\frac{\partial \ln L_{11}(y)}{\partial\vartheta_i}\right|\cdot\left|\frac{\partial \ln L_{11}(y)}{\partial\vartheta_j}\right| \\ \leq\ &2\left[\left|\frac{\partial \ln L_{11}(x)}{\partial\vartheta_i}\right| + \left|\frac{\partial \ln L_{11}(y)}{\partial\vartheta_i}\right|\right]\left[\left|\frac{\partial \ln L_{11}(x)}{\partial\vartheta_j}\right| + \left|\frac{\partial \ln L_{11}(y)}{\partial\vartheta_j}\right|\right] \\ &+ \left|\frac{\partial^2 \ln L_{11}(x)}{\partial\vartheta_i\partial\vartheta_j}\right| + \left|\frac{\partial^2 \ln L_{11}(y)}{\partial\vartheta_i\partial\vartheta_j}\right| \qquad \text{(B.49)}\end{aligned}$$

and one can estimate using (B.41) and (B.42)

$$\begin{aligned}\Big|\frac{\partial^2 g_{0,1}(x,y|\boldsymbol{\vartheta})}{\partial\vartheta_i\partial\vartheta_j}\Big| \le 2\,K\,g_{0,1}(x,y|\boldsymbol{\vartheta})\Bigg(&\ln L_{11}(x|\boldsymbol{\vartheta})+\ln L_{11}(y|\boldsymbol{\vartheta})\\ &+\Big|\frac{\partial \ln L_{11}(x|\boldsymbol{\vartheta})}{\partial\vartheta_i}\Big|+\Big|\frac{\partial \ln L_{11}(y|\boldsymbol{\vartheta})}{\partial\vartheta_i}\Big|\\ &+\Big|\frac{\partial \ln L_{11}(x|\boldsymbol{\vartheta})}{\partial\vartheta_j}\Big|+\Big|\frac{\partial \ln L_{11}(y|\boldsymbol{\vartheta})}{\partial\vartheta_j}\Big|\\ &+\Big|\frac{\partial^2 \ln L_{11}(x|\boldsymbol{\vartheta})}{\partial\vartheta_i\partial\vartheta_j}\Big|+\Big|\frac{\partial^2 \ln L_{11}(y|\boldsymbol{\vartheta})}{\partial\vartheta_i\partial\vartheta_j}\Big|\\ +\Bigg[\Big|\frac{\partial \ln L_{11}(x|\boldsymbol{\vartheta})}{\partial\vartheta_i}\Big|+\Big|\frac{\partial \ln L_{11}(y|\boldsymbol{\vartheta})}{\partial\vartheta_i}\Big|\Bigg]&\Bigg[\Big|\frac{\partial \ln L_{11}(x|\boldsymbol{\vartheta})}{\partial\vartheta_j}\Big|+\Big|\frac{\partial \ln L_{11}(y|\boldsymbol{\vartheta})}{\partial\vartheta_j}\Big|\Bigg]\Bigg)\\ &+4\,c_0^2\,.\end{aligned}$$

Applying $g_{0,1}(x,y|\boldsymbol{\vartheta}) \le S_L(x|\boldsymbol{\vartheta})$ and $g_{0,1}(x,y|\boldsymbol{\vartheta}) \le S_L(y|\boldsymbol{\vartheta})$, respectively, yields

$$\begin{aligned}\Big|\frac{\partial^2 g_{0,1}(x,y|\boldsymbol{\vartheta})}{\partial\vartheta_i\partial\vartheta_j}\Big| &\le 2\,K\big(\ln L_{11}(x|\boldsymbol{\vartheta})S_L(x|\boldsymbol{\vartheta})+\ln L_{11}(y|\boldsymbol{\vartheta})S_L(y|\boldsymbol{\vartheta})\big)\\ &+2\,K\Big(\Big|\frac{\partial \ln L_{11}(x|\boldsymbol{\vartheta})}{\partial\vartheta_i}\Big|S_L(x|\boldsymbol{\vartheta})+\Big|\frac{\partial \ln L_{11}(y|\boldsymbol{\vartheta})}{\partial\vartheta_i}\Big|S_L(y|\boldsymbol{\vartheta})\Big)\\ &+2\,K\Big(\Big|\frac{\partial \ln L_{11}(x|\boldsymbol{\vartheta})}{\partial\vartheta_j}\Big|S_L(x|\boldsymbol{\vartheta})+\Big|\frac{\partial \ln L_{11}(y|\boldsymbol{\vartheta})}{\partial\vartheta_j}\Big|S_L(y|\boldsymbol{\vartheta})\Big)\\ &+2\,K\Big(\Big|\frac{\partial^2 \ln L_{11}(x|\boldsymbol{\vartheta})}{\partial\vartheta_i\partial\vartheta_j}\Big|S_L(x|\boldsymbol{\vartheta})+\Big|\frac{\partial^2 \ln L_{11}(y|\boldsymbol{\vartheta})}{\partial\vartheta_i\partial\vartheta_j}\Big|S_L(y|\boldsymbol{\vartheta})\Big)\\ &+2\,K\Big[\Big|\frac{\partial \ln L_{11}(x|\boldsymbol{\vartheta})}{\partial\vartheta_i}\Big|S_L^{1/2}(x|\boldsymbol{\vartheta})+\Big|\frac{\partial \ln L_{11}(y|\boldsymbol{\vartheta})}{\partial\vartheta_i}\Big|S_L^{1/2}(y|\boldsymbol{\vartheta})\Big]\\ &\quad\cdot\Big[\Big|\frac{\partial \ln L_{11}(x|\boldsymbol{\vartheta})}{\partial\vartheta_j}\Big|S_L^{1/2}(x|\boldsymbol{\vartheta})+\Big|\frac{\partial \ln L_{11}(y|\boldsymbol{\vartheta})}{\partial\vartheta_j}\Big|S_L^{1/2}(y|\boldsymbol{\vartheta})\Big]\\ &\quad+4\,c_0^2\,.\end{aligned}$$

From (B.44), (B.45) and (B.46) it follows that

$$\Big|\frac{\partial^2 g_{0,1}(x,y|\boldsymbol{\vartheta})}{\partial\vartheta_i\partial\vartheta_j}\Big| \le 16KK_0+8KK_0^2+4\,c_0^2 = d_0.$$

□

Lemma 21. *Let $\mathfrak{H} = \{H(.|\boldsymbol{\lambda}) : \boldsymbol{\lambda} \in \Lambda\}$ be a family of cumulative hazard functions. Let $\mathfrak{S}$ be a family of univariate survival functions defined according to* (4.10) *and let assumptions 2, 3, 4(a) and 7 be fulfilled. Let the*

function $g_{0,\alpha}(x,y|\boldsymbol{\vartheta})$, $h_1(x,y|\boldsymbol{\vartheta})$ *and* $h_2(x,y|\boldsymbol{\vartheta})$ *be defined by* (A.31), (A.33) *and* (A.35), *respectively. Then for* $i,j \in \{1,\dots,K_p\}$ *positive constants* c_1, d_1, c_2 *and* d_2 *exist such that for* $\alpha = \frac{1}{2}, 1$

$$\left|\frac{\partial h_1(x,y|\boldsymbol{\vartheta})}{\partial \vartheta_i}\right| \cdot g_{0,\alpha}(x,y|\boldsymbol{\vartheta}) \le c_1 \ , \quad \left|\frac{\partial^2 h_1(x,y|\boldsymbol{\vartheta})}{\partial \vartheta_i \partial \vartheta_j}\right| \cdot g_{0,\alpha}(x,y|\boldsymbol{\vartheta}) \le d_1 \ ,$$
$$\left|\frac{\partial h_2(x,y|\boldsymbol{\vartheta})}{\partial \vartheta_i}\right| \cdot g_{0,\alpha}(x,y|\boldsymbol{\vartheta}) \le c_2 \text{ and } \left|\frac{\partial^2 h_2(x,y|\boldsymbol{\vartheta})}{\partial \vartheta_i \partial \vartheta_j}\right| \cdot g_{0,1}(x,y|\boldsymbol{\vartheta}) \le d_2$$

for all $x,y \ge 0$ *and all* $\boldsymbol{\vartheta} \in \hat{\Theta}$.

Proof. One can write

$$h_1(x,y|\boldsymbol{\vartheta}) = l_1(\boldsymbol{\vartheta}) \cdot h_{11}(x|\boldsymbol{\vartheta}) + l_0(\boldsymbol{\vartheta}) \cdot h_{10}(x,y|\boldsymbol{\vartheta})$$

and therefore it holds that

$$\frac{\partial h_1}{\partial \vartheta_i} = \frac{\partial l_1}{\partial \vartheta_i} \cdot h_{11} + \frac{\partial l_0}{\partial \vartheta_i} \cdot h_{10} + l_1 \cdot \frac{\partial \ln h_{11}}{\partial \vartheta_i} \cdot h_{11} + l_0 \cdot \frac{\partial \ln h_{10}}{\partial \vartheta_i} \cdot h_{10} \ . \quad \text{(B.50)}$$

From assumption 4(a) and (B.40) it follows that $h_{1k}(x,y|\boldsymbol{\vartheta}) \le M_s^2 C_0 = \tilde{C}$ for $k = 0, 1$, all $x,y \ge 0$ and all $\boldsymbol{\vartheta} \in \hat{\Theta}$. Using (B.43) one can estimate

$$\left|\frac{\partial h_1(x,y|\boldsymbol{\vartheta})}{\partial \vartheta_i}\right| \le 2\,K\tilde{C} + K\tilde{C}\left|\frac{\partial \ln h_{11}(x|\boldsymbol{\vartheta})}{\partial \vartheta_i}\right| + K\tilde{C}\left|\frac{\partial \ln h_{10}(x,y|\boldsymbol{\vartheta})}{\partial \vartheta_i}\right| . \quad \text{(B.51)}$$

It holds that (see (A.36))

$$\ln h_{11}(x|\boldsymbol{\vartheta}) = 2\ln\sigma + \ln\mu(x|\boldsymbol{\lambda}) - \ln L_{11}(x|\boldsymbol{\lambda}) \quad \text{(B.52)}$$

and consequently

$$\left|\frac{\partial \ln h_{11}(x|\boldsymbol{\vartheta})}{\partial \vartheta_i}\right| \le 2\left|\frac{\partial \ln\sigma}{\partial \vartheta_i}\right| + \left|\frac{\partial \ln\mu(x|\boldsymbol{\lambda})}{\partial \vartheta_i}\right| + \left|\frac{\partial \ln L_{11}(x|\boldsymbol{\vartheta})}{\partial \vartheta_i}\right| .$$

If $\vartheta_i \ne \sigma$ then $\left|\frac{\partial \ln\sigma}{\partial \vartheta_i}\right| = 0$ and if $\vartheta_i = \sigma$ then $\left|\frac{\partial \ln\sigma}{\partial \vartheta_i}\right| = \frac{1}{\sigma} \le \frac{1}{m_s}$. Using $g_{0,\alpha}(x,y|\boldsymbol{\vartheta}) \le S_L^\alpha(x|\boldsymbol{\vartheta})$ one can conclude that

$$\left|\frac{\partial \ln h_{11}(x|\boldsymbol{\vartheta})}{\partial \vartheta_i}\right| \cdot g_{0,\alpha}(x,y|\boldsymbol{\vartheta}) \le \frac{2}{m_s} + \left|\frac{\partial \ln\mu(x|\boldsymbol{\lambda})}{\partial \vartheta_i}\right| \cdot S_L^\alpha(x|\boldsymbol{\vartheta}) + \left|\frac{\partial \ln L_{11}(x|\boldsymbol{\vartheta})}{\partial \vartheta_i}\right| \cdot S_L^\alpha(x|\boldsymbol{\vartheta})$$

and therefore assumption 7(d) and (B.45) yield

$$\left|\frac{\partial \ln h_{11}(x|\boldsymbol{\vartheta})}{\partial \vartheta_i}\right| \cdot g_{0,\alpha}(x,y|\boldsymbol{\vartheta}) \leq \frac{2}{m_s} + \hat{C}_1 + K_0 = c_{11} \text{ for } \alpha = \tfrac{1}{4}, \tfrac{1}{2}, 1. \quad \text{(B.53)}$$

Since (see (A.37))

$$\ln h_{10}(x|\boldsymbol{\vartheta}) = 2\ln\sigma + \ln\mu(x|\boldsymbol{\lambda}) - \ln L_4(x,y|\boldsymbol{\lambda}) \quad \text{(B.54)}$$

with (B.42) it follows that

$$\left|\frac{\partial \ln h_{10}(x,y|\boldsymbol{\vartheta})}{\partial \vartheta_i}\right| \leq 2\left|\frac{\partial \ln\sigma}{\partial \vartheta_i}\right| + \left|\frac{\partial \ln\mu(x|\boldsymbol{\lambda})}{\partial \vartheta_i}\right| + \left|\frac{\partial \ln L_{11}(x|\boldsymbol{\vartheta})}{\partial \vartheta_i}\right| + \left|\frac{\partial \ln L_{11}(y|\boldsymbol{\vartheta})}{\partial \vartheta_i}\right|.$$

Using assumption 7(d) and (B.45) one gets

$$\left|\frac{\partial \ln h_{10}(x,y|\boldsymbol{\vartheta})}{\partial \vartheta_i}\right| \cdot g_{0,\alpha}(x,y|\boldsymbol{\vartheta}) \leq \frac{2}{m_s} + \hat{C}_1 + 2\,K_0 = c_{10} \text{ for } \alpha = \tfrac{1}{4}, \tfrac{1}{2}, 1. \quad \text{(B.55)}$$

Inequalities (B.51), (B.53) and (B.55) yield

$$\left|\frac{\partial h_1(x,y|\boldsymbol{\vartheta})}{\partial \vartheta_i}\right| \cdot g_{0,\alpha}(x,y|\boldsymbol{\vartheta}) \leq K\tilde{C}[2 + c_{11} + c_{10}] = c_1 \text{ for } \alpha = \tfrac{1}{4}, \tfrac{1}{2}, 1.$$

From (B.50) it follows that

$$\begin{aligned}\frac{\partial^2 h_1}{\partial\vartheta_i\partial\vartheta_j} = {} & \frac{\partial^2 l_1}{\partial\vartheta_i\partial\vartheta_j}\cdot h_{11} + \frac{\partial^2 l_0}{\partial\vartheta_i\partial\vartheta_j}\cdot h_{10} + \frac{\partial l_1}{\partial\vartheta_i}\cdot\frac{\partial \ln h_{11}}{\partial\vartheta_j}\cdot h_{11} + \frac{\partial l_0}{\partial\vartheta_i}\cdot\frac{\partial \ln h_{10}}{\partial\vartheta_j}\cdot h_{10} \\ & + \frac{\partial l_1}{\partial\vartheta_j}\cdot\frac{\partial \ln h_{11}}{\partial\vartheta_i}\cdot h_{11} + \frac{\partial l_0}{\partial\vartheta_j}\cdot\frac{\partial \ln h_{10}}{\partial\vartheta_i}\cdot h_{10} \\ & + l_1\cdot\frac{\partial^2 \ln h_{11}}{\partial\vartheta_i\partial\vartheta_j}\cdot h_{11} + l_0\cdot\frac{\partial^2 \ln h_{10}}{\partial\vartheta_i\partial\vartheta_j}\cdot h_{10} \\ & + l_1\cdot\frac{\partial \ln h_{11}}{\partial\vartheta_i}\cdot\frac{\partial \ln h_{11}}{\partial\vartheta_j}\cdot h_{11} + l_0\cdot\frac{\partial \ln h_{10}}{\partial\vartheta_i}\cdot\frac{\partial \ln h_{10}}{\partial\vartheta_j}\cdot h_{10}\,.\end{aligned}$$

Using (B.43), (B.53) and (B.55) one can conclude

$$\begin{aligned}\left|\frac{\partial^2 h_1(x,y|\boldsymbol{\vartheta})}{\partial\vartheta_i\partial\vartheta_j}\right| \cdot g_{0,\alpha}(x,y|\boldsymbol{\vartheta}) \leq {} & K\tilde{C}[2 + 2c_{11} + 2c_{10} + c_{10}^2 + c_{11}^2] \\ & + K\tilde{C}\left(\left|\frac{\partial^2 \ln h_{11}(x|\boldsymbol{\vartheta})}{\partial\vartheta_i\partial\vartheta_j}\right| + \left|\frac{\partial^2 \ln h_{10}(x,y|\boldsymbol{\vartheta})}{\partial\vartheta_i\partial\vartheta_j}\right|\right) g_{0,\alpha}(x,y|\boldsymbol{\vartheta}) \text{ for } \alpha = \tfrac{1}{2}, 1.\end{aligned} \quad \text{(B.56)}$$

From (B.52) it follows that

$$\left|\frac{\partial^2 \ln h_{11}(x|\boldsymbol{\vartheta})}{\partial\vartheta_i\partial\vartheta_j}\right| \leq 2\left|\frac{\partial^2 \ln \sigma}{\partial\vartheta_i\partial\vartheta_j}\right| + \left|\frac{\partial^2 \ln \mu(x|\boldsymbol{\lambda})}{\partial\vartheta_i\partial\vartheta_j}\right| + \left|\frac{\partial^2 \ln L_{11}(x|\boldsymbol{\vartheta})}{\partial\vartheta_i\partial\vartheta_j}\right|.$$

If $\vartheta_i \neq \sigma$ or $\vartheta_j \neq \sigma$ then $\left|\frac{\partial^2 \ln \sigma}{\partial\vartheta_i\partial\vartheta_j}\right| = 0$ and if $\vartheta_i = \vartheta_j = \sigma$ then $\left|\frac{\partial^2 \ln \sigma}{\partial\sigma^2}\right| = \frac{1}{\sigma^2} \leq \frac{1}{m_s^2}$. Since $g_{0,\alpha}(x, y|\boldsymbol{\vartheta}) \leq S_L^\alpha(x|\boldsymbol{\vartheta})$ one gets

$$\begin{aligned}\left|\frac{\partial^2 \ln h_{11}(x|\boldsymbol{\vartheta})}{\partial\vartheta_i\partial\vartheta_j}\right| \cdot g_{0,\alpha}(x, y|\boldsymbol{\vartheta}) \leq \frac{2}{m_s^2} &+ \left|\frac{\partial^2 \ln \mu(x|\boldsymbol{\lambda})}{\partial\vartheta_i\partial\vartheta_j}\right| \cdot S_L^\alpha(x|\boldsymbol{\vartheta}) \\ &+ \left|\frac{\partial^2 \ln L_{11}(x|\boldsymbol{\vartheta})}{\partial\vartheta_i\partial\vartheta_j}\right| \cdot S_L^\alpha(x|\boldsymbol{\vartheta})\end{aligned}$$

and thus assumption 7(e) and (B.46) yield

$$\left|\frac{\partial^2 \ln h_{11}(x|\boldsymbol{\vartheta})}{\partial\vartheta_i\partial\vartheta_j}\right| \cdot g_{0,\alpha}(x, y|\boldsymbol{\vartheta}) \leq \frac{2}{m_s^2} + \hat{C}_2 + K_0 = d_{11} \text{ for } \alpha = \tfrac{1}{2}, 1. \quad \text{(B.57)}$$

From equation (B.54) it follows that

$$\left|\frac{\partial^2 \ln h_{10}(x, y|\boldsymbol{\vartheta})}{\partial\vartheta_i\partial\vartheta_j}\right| \leq 2\left|\frac{\partial^2 \ln \sigma}{\partial\vartheta_i\partial\vartheta_j}\right| + \left|\frac{\partial^2 \ln \mu(x|\boldsymbol{\lambda})}{\partial\vartheta_i\partial\vartheta_j}\right| + \left|\frac{\partial^2 \ln L_4(x, y|\boldsymbol{\vartheta})}{\partial\vartheta_i\partial\vartheta_j}\right|$$

and together with (B.49), $\left|\frac{\partial^2 \ln \sigma}{\partial\sigma^2}\right| = \frac{1}{\sigma^2} \leq \frac{1}{m_s^2}$, $g_{0,\alpha}(x, y|\boldsymbol{\vartheta}) \leq S_L^\alpha(x|\boldsymbol{\vartheta})$ and $g_{0,\alpha}(x, y|\boldsymbol{\vartheta}) \leq S_L^\alpha(y|\boldsymbol{\vartheta})$ this yields

$$\begin{aligned}\left|\frac{\partial^2 \ln h_{10}(x, y|\boldsymbol{\vartheta})}{\partial\vartheta_i\partial\vartheta_j}\right| \cdot g_{0,\alpha}(x, y|\boldsymbol{\vartheta}) \leq \frac{2}{m_s^2} &+ \left|\frac{\partial^2 \ln \mu(x|\boldsymbol{\lambda})}{\partial\vartheta_i\partial\vartheta_j}\right| S_L^\alpha(x|\boldsymbol{\vartheta}) \\ &+ 2\left(\left|\frac{\partial^2 \ln L_{11}(x|\boldsymbol{\vartheta})}{\partial\vartheta_i\partial\vartheta_j}\right| S_L^\alpha(x|\boldsymbol{\vartheta}) + \left|\frac{\partial^2 \ln L_{11}(y|\boldsymbol{\vartheta})}{\partial\vartheta_i\partial\vartheta_j}\right| S_L^\alpha(y|\boldsymbol{\vartheta})\right) \\ &+ 2\left[\left|\frac{\partial \ln L_{11}(x|\boldsymbol{\vartheta})}{\partial\vartheta_i}\right| S_L^{\alpha/2}(x|\boldsymbol{\vartheta}) + \left|\frac{\partial \ln L_{11}(y|\boldsymbol{\vartheta})}{\partial\vartheta_i}\right| S_L^{\alpha/2}(y|\boldsymbol{\vartheta})\right] \\ &\cdot \left[\left|\frac{\partial \ln L_{11}(x|\boldsymbol{\vartheta})}{\partial\vartheta_j}\right| S_L^{\alpha/2}(x|\boldsymbol{\vartheta}) + \left|\frac{\partial \ln L_{11}(y|\boldsymbol{\vartheta})}{\partial\vartheta_j}\right| S_L^{\alpha/2}(y|\boldsymbol{\vartheta})\right]\end{aligned}$$

resulting in

$$\left|\frac{\partial^2 \ln h_{10}(x, y|\boldsymbol{\vartheta})}{\partial\vartheta_i\partial\vartheta_j}\right| \cdot g_{0,\alpha}(x, y|\boldsymbol{\vartheta}) \leq \frac{2}{m_s^2} + \hat{C}_2 + 4K_0 + 8K_0^2 = d_{10} \quad \text{(B.58)}$$

for $\alpha = \frac{1}{2}, 1$, if assumption 7(e), (B.45) and (B.46) are taken into account. From (B.56), (B.57) and (B.58) it follows that

$$\begin{aligned}\left|\frac{\partial^2 h_1(x,y|\boldsymbol{\vartheta})}{\partial\vartheta_i\partial\vartheta_j}\right| \cdot g_{0,\alpha}(x,y|\boldsymbol{\vartheta}) \leq & K\tilde{C}[2+2c_{11}+2c_{10}+c_{10}^2+c_{11}^2+d_{10}+d_{11}] \\ & = d_1 \text{ for } \alpha = \tfrac{1}{2}, 1.\end{aligned}$$

From (A.35) and (A.38) it follows that

$$h_2(x,y|\boldsymbol{\vartheta}) = l_0(\boldsymbol{\vartheta}) \cdot h_{20}(x,y|\boldsymbol{\vartheta})$$

and therefore it holds that

$$\frac{\partial h_2}{\partial\vartheta_i} = \frac{\partial l_0}{\vartheta_i} \cdot h_{20} + l_0 \cdot \frac{\partial \ln h_{20}}{\partial\vartheta_i} \cdot h_{20} \ . \tag{B.59}$$

Since (B.40) and assumption 4(a) yield $h_{20}(x,y|\boldsymbol{\vartheta}) \leq M_s^4\, C_0^2$, using (B.43) one can conclude

$$\left|\frac{\partial h_2(x,y|\boldsymbol{\vartheta})}{\partial\vartheta_i}\right| \leq K\, M_s^4\, C_0^2 \left[1 + \left|\frac{\partial \ln h_{20}(x,y|\boldsymbol{\vartheta})}{\partial\vartheta_i}\right|\right] \ . \tag{B.60}$$

It holds that (see (A.38))

$$\ln h_{20}(x,y|\boldsymbol{\vartheta}) = 4\ln\sigma + \ln\mu(x|\boldsymbol{\vartheta}) + \ln\mu(y|\boldsymbol{\vartheta}) - 2\ln L_4(x,y|\boldsymbol{\vartheta}) \tag{B.61}$$

and hence with (B.42) it follows that

$$\begin{aligned}\left|\frac{\partial \ln h_{20}(x,y|\boldsymbol{\vartheta})}{\partial\vartheta_i}\right| \leq & 4 \cdot \left|\frac{\partial \ln\sigma}{\partial\vartheta_i}\right| + \left|\frac{\partial \ln\mu(x|\boldsymbol{\vartheta})}{\partial\vartheta_i}\right| + \left|\frac{\partial \ln\mu(y|\boldsymbol{\vartheta})}{\partial\vartheta_i}\right| \\ & + 2 \cdot \left|\frac{\partial \ln L_{11}(x|\boldsymbol{\vartheta})}{\partial\vartheta_i}\right| + 2 \cdot \left|\frac{\partial \ln L_{11}(y|\boldsymbol{\vartheta})}{\partial\vartheta_i}\right| \ .\end{aligned}$$

If $\vartheta_i \neq \sigma$ then $\left|\frac{\partial \ln\sigma}{\partial\vartheta_i}\right| = 0$ and if $\vartheta_i = \sigma$ then $\left|\frac{\partial \ln\sigma}{\partial\vartheta_i}\right| = \frac{1}{\sigma} \leq \frac{1}{m_s}$. Using $g_{0,\alpha}(x,y|\boldsymbol{\vartheta}) \leq S_L^\alpha(x|\boldsymbol{\vartheta})$ and $g_{0,\alpha}(x,y|\boldsymbol{\vartheta}) \leq S_L^\alpha(y|\boldsymbol{\vartheta})$, respectively, one gets

$$\begin{aligned}\left|\frac{\partial \ln h_{20}(x,y|\boldsymbol{\vartheta})}{\partial\vartheta_i}\right| \cdot g_{0,\alpha}(x,y|\boldsymbol{\vartheta}) \leq & \left|\frac{\partial \ln\mu(x|\boldsymbol{\lambda})}{\partial\vartheta_i}\right| \cdot S_L^\alpha(x|\boldsymbol{\vartheta}) + \left|\frac{\partial \ln\mu(y|\boldsymbol{\lambda})}{\partial\vartheta_i}\right| \cdot S_L^\alpha(y|\boldsymbol{\vartheta}) \\ & + 2\left|\frac{\partial \ln L_{11}(x|\boldsymbol{\vartheta})}{\partial\vartheta_i}\right| \cdot S_L^\alpha(x|\boldsymbol{\vartheta}) + 2\left|\frac{\partial \ln L_{11}(y|\boldsymbol{\vartheta})}{\partial\vartheta_i}\right| \cdot S_L^\alpha(y|\boldsymbol{\vartheta}) + \frac{4}{m_s}\end{aligned}$$

and therefore assumption 7(d) and (B.45) yield

$$\left|\frac{\partial \ln h_{20}(x,y|\boldsymbol{\vartheta})}{\partial\vartheta_i}\right| \cdot g_{0,\alpha}(x,y|\boldsymbol{\vartheta}) \leq \frac{4}{m_s} + 2\,\hat{C}_1 + 2\,K_0 = c_{20} \text{ for } \alpha = \tfrac{1}{2}, 1. \tag{B.62}$$

Together with (B.60) this leads to

$$\left|\frac{\partial h_2(x,y|\boldsymbol{\vartheta})}{\partial\vartheta_i}\right| \cdot g_{0,\alpha}(x,y|\boldsymbol{\vartheta}) \leq K\, M_s^4\, C_0^4[1+c_{20}] = c_2 \text{ for } \alpha = \tfrac{1}{2}, 1.$$

From (B.59) it follows that

$$\begin{aligned}\frac{\partial^2 h_2}{\partial\vartheta_i\partial\vartheta_j} = \frac{\partial^2 l_0}{\partial\vartheta_i\partial\vartheta_j}\cdot h_{20} + \frac{\partial l_0}{\partial\vartheta_i}\cdot\frac{\partial \ln h_{20}}{\partial\vartheta_j}\cdot h_{20} + \frac{\partial l_0}{\partial\vartheta_j}\cdot\frac{\partial \ln h_{20}}{\partial\vartheta_i}\cdot h_{20} \\ + l_0\cdot\frac{\partial^2 \ln h_{20}}{\partial\vartheta_i\partial\vartheta_j}\cdot h_{20} + l_0\cdot\frac{\partial \ln h_{20}}{\partial\vartheta_i}\cdot\frac{\partial \ln h_{20}}{\partial\vartheta_j}\cdot h_{20}\ .\end{aligned}$$

and therefore using $|h_{20}| \leq M_s^4\, C_0^2$, (B.43) and (B.62) one can conclude

$$\begin{aligned}&\left|\frac{\partial^2 h_2(x,y|\boldsymbol{\vartheta})}{\partial\vartheta_i\partial\vartheta_j}\right| \cdot g_{0,1}(x,y|\boldsymbol{\vartheta}) \\ &\qquad \leq K\, M_s^4\, C_0^2\left[1 + 2\,c_{20} + c_{20}^2 + \left|\frac{\partial^2 \ln h_{20}(x,y|\boldsymbol{\vartheta})}{\partial\vartheta_i\partial\vartheta_j}\right| \cdot g_{0,1}(x,y|\boldsymbol{\vartheta})\right]\ .\end{aligned}$$

Equation (B.61) yields

$$\begin{aligned}\left|\frac{\partial^2 \ln h_{20}(x,y|\boldsymbol{\vartheta})}{\partial\vartheta_i\partial\vartheta_j}\right| \leq 4\cdot\left|\frac{\partial^2 \ln\sigma}{\partial\vartheta_i\partial\vartheta_j}\right| + \left|\frac{\partial^2 \ln\mu(x|\boldsymbol{\vartheta})}{\partial\vartheta_i\partial\vartheta_j}\right| + \left|\frac{\partial^2 \ln\mu(y|\boldsymbol{\vartheta})}{\partial\vartheta_i\partial\vartheta_j}\right| \\ + \left|\frac{\partial^2 \ln L_4(x,y|\boldsymbol{\vartheta})}{\partial\vartheta_i\partial\vartheta_j}\right|\ .\end{aligned}$$

Since $\left|\frac{\partial^2 \ln\sigma}{\partial\sigma^2}\right| = \frac{1}{\sigma^2} \leq \frac{1}{m_s^2}$, $g_{0,1}(x,y|\boldsymbol{\vartheta}) \leq S_L(x|\boldsymbol{\vartheta})$ and $g_{0,1}(x,y|\boldsymbol{\vartheta}) \leq S_L(y|\boldsymbol{\vartheta})$, respectively, using (B.49) it follows that

$$\begin{aligned}&\left|\frac{\partial^2 \ln h_{20}(x,y|\boldsymbol{\vartheta})}{\partial\vartheta_i\partial\vartheta_j}\right| \cdot g_{0,1}(x,y|\boldsymbol{\vartheta}) \leq \frac{4}{m_s^2} \\ &\quad + \left|\frac{\partial^2 \ln\mu(x|\boldsymbol{\lambda})}{\partial\vartheta_i\partial\vartheta_j}\right| S_L(x|\boldsymbol{\vartheta}) + \left|\frac{\partial^2 \ln\mu(y|\boldsymbol{\lambda})}{\partial\vartheta_i\partial\vartheta_j}\right| S_L(y|\boldsymbol{\vartheta}) \\ &\quad + 2\left(\left|\frac{\partial^2 \ln L_{11}(x|\boldsymbol{\vartheta})}{\partial\vartheta_i\partial\vartheta_j}\right| S_L(x|\boldsymbol{\vartheta}) + \left|\frac{\partial^2 \ln L_{11}(y|\boldsymbol{\vartheta})}{\partial\vartheta_i\partial\vartheta_j}\right| S_L(y|\boldsymbol{\vartheta})\right) \\ &\quad + 2\left[\left|\frac{\partial \ln L_{11}(x|\boldsymbol{\vartheta})}{\partial\vartheta_i}\right| S_L^{1/2}(x|\boldsymbol{\vartheta}) + \left|\frac{\partial \ln L_{11}(y|\boldsymbol{\vartheta})}{\partial\vartheta_i}\right| S_L^{1/2}(y|\boldsymbol{\vartheta})\right] \\ &\qquad \cdot\left[\left|\frac{\partial \ln L_{11}(x|\boldsymbol{\vartheta})}{\partial\vartheta_j}\right| S_L^{1/2}(x|\boldsymbol{\vartheta}) + \left|\frac{\partial \ln L_{11}(y|\boldsymbol{\vartheta})}{\partial\vartheta_j}\right| S_L^{1/2}(y|\boldsymbol{\vartheta})\right]\end{aligned}$$

and therefore assumption 7(e), (B.45) and (B.46) yield

$$\left|\frac{\partial^2 \ln h_{20}(x,y|\boldsymbol{\vartheta})}{\partial\vartheta_i\partial\vartheta_j}\right| \cdot g_{0,1}(x,y|\boldsymbol{\vartheta}) \leq \frac{4}{m_s^2} + 2\,\hat{C}_2 + 4\,K_0 + 8\,K_0^2 = d_{20}$$

and thus follows

$$\left|\frac{\partial^2 h_2(x,y|\boldsymbol{\vartheta})}{\partial\vartheta_i\partial\vartheta_j}\right| \cdot g_{0,1}(x,y|\boldsymbol{\vartheta}) \leq K\,M_s^4\,C_0^2\left[1 + 2\,c_{20} + c_{20}^2 + d_{20}\right] = d_2\ .$$

□

Now the proofs of the technical lemmata that prepare the proof of the main theorem about asymptotic normality in the bivariate correlated gamma frailty model are given.

Proof of lemma 10

From (A.31) it follows that

$$\frac{\partial g_{0,\alpha}(x,y|\boldsymbol{\vartheta})}{\partial\vartheta_i} = 2\,g_{0,\alpha/2}(x,y|\boldsymbol{\vartheta}) \cdot \frac{\partial g_{0,\alpha/2}(x,y|\boldsymbol{\vartheta})}{\partial\vartheta_i}\ . \tag{B.63}$$

Hence from (A.32) one can conclude that

$$\begin{aligned}
\left|\frac{\partial g_{1,\alpha}}{\partial\vartheta_i}\right| &\leq \left|\frac{\partial h_1}{\partial\vartheta_i}\right| \cdot g_{0,\alpha} + h_1 \cdot \left|\frac{\partial g_{0,\alpha}}{\partial\vartheta_i}\right| \quad \text{and} \\
\left|\frac{\partial^2 g_{1,\alpha}}{\partial\vartheta_i\partial\vartheta_j}\right| &\leq \left|\frac{\partial^2 h_1}{\partial\vartheta_i\partial\vartheta_j}\right| \cdot g_{0,\alpha} + 2\left|\frac{\partial h_1}{\partial\vartheta_i}\right| \cdot g_{0,\frac{\alpha}{2}} \cdot \left|\frac{\partial g_{0,\frac{\alpha}{2}}}{\partial\vartheta_j}\right| \\
&\quad + 2\left|\frac{\partial h_1}{\partial\vartheta_j}\right| \cdot g_{0,\frac{\alpha}{2}} \cdot \left|\frac{\partial g_{0,\frac{\alpha}{2}}}{\partial\vartheta_i}\right| + h_1 \cdot \left|\frac{\partial^2 g_{0,\alpha}}{\partial\vartheta_i\partial\vartheta_j}\right|
\end{aligned}$$

and therefore lemmata 20, 9 and 21 yield

$$\begin{aligned}
\left|\frac{\partial g_{1,\alpha}(x,y|\boldsymbol{\vartheta})}{\partial\vartheta_i}\right| &\leq c_1 + K_1\,c_0 = G_1 \quad \text{and} \\
\left|\frac{\partial^2 g_{1,\alpha}(x,y|\boldsymbol{\vartheta})}{\partial\vartheta_i\partial\vartheta_j}\right| &\leq d_1 + 4\,c_1\,c_0 + K_0\,d_0 = H_1\ .
\end{aligned}$$

Denote $\tilde{g}_{1,\alpha}(x,y) = g_{1,\alpha}(y,x)$. From (A.34) and (B.63) it follows that

$$\left|\frac{\partial g_{2,\alpha}}{\partial\vartheta_i}\right| \leq \left|\frac{\partial h_2}{\partial\vartheta_i}\right| \cdot g_{0,\alpha} + h_2 \cdot \left|\frac{\partial g_{0,\alpha}}{\partial\vartheta_i}\right| + \left|\frac{\partial g_{1,\alpha/2}}{\partial\vartheta_i}\right| \cdot \tilde{g}_{1,\alpha/2} + g_{1,\alpha/2} \cdot \left|\frac{\partial \tilde{g}_{1,\alpha/2}}{\partial\vartheta_i}\right|$$

and

$$\begin{aligned}\left|\frac{\partial^2 g_{2,\alpha}}{\partial\vartheta_i\partial\vartheta_j}\right| \le \left|\frac{\partial^2 h_2}{\partial\vartheta_i\partial\vartheta_j}\right|\cdot g_{0,\alpha} + 2\left|\frac{\partial h_2}{\partial\vartheta_i}\right| g_{0,\alpha/2}\cdot\left|\frac{\partial g_{0,\alpha/2}}{\partial\vartheta_j}\right| + 2\left|\frac{\partial h_2}{\partial\vartheta_j}\right| g_{0,\alpha/2}\cdot\left|\frac{\partial g_{0,\alpha/2}}{\partial\vartheta_i}\right| \\ + h_2\cdot\left|\frac{\partial^2 g_{0,\alpha}}{\partial\vartheta_i\partial\vartheta_j}\right| + \left|\frac{\partial^2 g_{1,\alpha/2}}{\partial\vartheta_i\partial\vartheta_j}\right|\cdot\tilde{g}_{1,\alpha/2} + \left|\frac{\partial g_{1,\alpha/2}}{\partial\vartheta_i}\right|\cdot\left|\frac{\partial\tilde{g}_{1,\alpha/2}}{\partial\vartheta_j}\right| \\ + \left|\frac{\partial g_{1,\alpha/2}}{\partial\vartheta_j}\right|\cdot\left|\frac{\partial\tilde{g}_{1,\alpha/2}}{\partial\vartheta_i}\right| + g_{1,\alpha/2}\cdot\left|\frac{\partial^2\tilde{g}_{1,\alpha/2}}{\partial\vartheta_i\partial\vartheta_j}\right| \quad .\end{aligned}$$

Obviously it holds that $g_{1,\alpha} \le h_1$ and $\tilde{g}_{1,\alpha} \le h_1$ (see equation (A.32)). Hence, applying lemmata 20, 9, 21 and 10 yields

$$\begin{aligned}\left|\frac{\partial g_{2,1}(x,y|\boldsymbol{\vartheta})}{\partial\vartheta_i}\right| &\le c_2 + K_2\,c_0 + 2\,K_1\,G_1 = G_2 \quad \text{and} \\ \left|\frac{\partial^2 g_{2,1}(x,y|\boldsymbol{\vartheta})}{\partial\vartheta_i\partial\vartheta_j}\right| &\le d_2 + 4\,c_2\,c_0 + K_2\,d_0 + 2\,H_1\,K_1 + 2\,G_1^2 = H_2 \ .\end{aligned}$$

□

Proof of lemma 11

From equation (A.31) it follows that

$$\ln g_{0,1}(x,y|\boldsymbol{\vartheta}) = -l_1(\boldsymbol{\vartheta})\ln L_{11}(x|\boldsymbol{\vartheta}) - l_1(\boldsymbol{\vartheta})\ln L_{11}(y|\boldsymbol{\vartheta}) - l_0(\boldsymbol{\vartheta})\ln L_4(x,y|\boldsymbol{\vartheta})$$

and therefore using (B.43) one can conclude

$$\begin{aligned}\left|\frac{\partial\ln g_{0,1}(x,y|\boldsymbol{\vartheta})}{\partial\vartheta_i}\right| \le K\Bigg(&\ln L_{11}(x|\boldsymbol{\vartheta}) + \ln L_{11}(y|\boldsymbol{\vartheta}) + \ln L_4(x,y|\boldsymbol{\vartheta}) \\ &+ \left|\frac{\partial\ln L_{11}(x|\boldsymbol{\vartheta})}{\partial\vartheta_i}\right| + \left|\frac{\partial\ln L_{11}(y|\boldsymbol{\vartheta})}{\partial\vartheta_i}\right| + \left|\frac{\partial\ln L_4(x,y|\boldsymbol{\vartheta})}{\partial\vartheta_i}\right|\Bigg)\end{aligned}$$

and

$$\begin{aligned}\left|\frac{\partial^2\ln g_{0,1}(x,y|\boldsymbol{\vartheta})}{\partial\vartheta_i\partial\vartheta_j}\right| \le K\Bigg(&\ln L_{11}(x|\boldsymbol{\vartheta}) + \ln L_{11}(y|\boldsymbol{\vartheta}) + \ln L_4(x,y|\boldsymbol{\vartheta}) \\ &+ \left|\frac{\partial\ln L_{11}(x|\boldsymbol{\vartheta})}{\partial\vartheta_i}\right| + \left|\frac{\partial\ln L_{11}(y|\boldsymbol{\vartheta})}{\partial\vartheta_i}\right| + \left|\frac{\partial\ln L_4(x,y|\boldsymbol{\vartheta})}{\partial\vartheta_i}\right| \\ &+ \left|\frac{\partial\ln L_{11}(x|\boldsymbol{\vartheta})}{\partial\vartheta_j}\right| + \left|\frac{\partial\ln L_{11}(y|\boldsymbol{\vartheta})}{\partial\vartheta_j}\right| + \left|\frac{\partial\ln L_4(x,y|\boldsymbol{\vartheta})}{\partial\vartheta_j}\right| \\ &+ \left|\frac{\partial^2\ln L_{11}(x|\boldsymbol{\vartheta})}{\partial\vartheta_i\partial\vartheta_j}\right| + \left|\frac{\partial^2\ln L_{11}(y|\boldsymbol{\vartheta})}{\partial\vartheta_i\partial\vartheta_j}\right| + \left|\frac{\partial^2\ln L_4(x,y|\boldsymbol{\vartheta})}{\partial\vartheta_i\partial\vartheta_j}\right|\Bigg) \ .\end{aligned}$$

Hence (B.42) and (B.49) yield

$$\Big|\frac{\partial \ln g_{0,1}(x,y|\boldsymbol{\vartheta})}{\partial \vartheta_i}\Big| \leq M_0(x,y)$$

with

$$M_0(x,y) = 2\,K\,\Big[\sup_{\boldsymbol{\vartheta}\in\hat{\Theta}} \ln L_{11}(x|\boldsymbol{\vartheta}) + \sup_{\boldsymbol{\vartheta}\in\hat{\Theta}} \ln L_{11}(y|\boldsymbol{\vartheta}) + \sup_{\boldsymbol{\vartheta}\in\hat{\Theta}}\Big|\frac{\partial \ln L_{11}(x|\boldsymbol{\vartheta})}{\partial \vartheta_i}\Big| + \sup_{\boldsymbol{\vartheta}\in\hat{\Theta}}\Big|\frac{\partial \ln L_{11}(y|\boldsymbol{\vartheta})}{\partial \vartheta_i}\Big|\Big]$$

and

$$\Big|\frac{\partial^2 \ln g_{0,1}(x,y|\boldsymbol{\vartheta})}{\partial \vartheta_i \partial \vartheta_j}\Big| \leq \tilde{M}_0(x,y)$$

with

$$\begin{aligned}\tilde{M}_0(x,y) = 2\,K\,\Big(&\sup_{\boldsymbol{\vartheta}\in\hat{\Theta}} \ln L_{11}(x|\boldsymbol{\vartheta}) + \sup_{\boldsymbol{\vartheta}\in\hat{\Theta}} \ln L_{11}(y|\boldsymbol{\vartheta}) + \sup_{\boldsymbol{\vartheta}\in\hat{\Theta}}\Big|\frac{\partial \ln L_{11}(x|\boldsymbol{\vartheta})}{\partial \vartheta_i}\Big| \\ &+ \sup_{\boldsymbol{\vartheta}\in\hat{\Theta}}\Big|\frac{\partial \ln L_{11}(y|\boldsymbol{\vartheta})}{\partial \vartheta_i}\Big| + \sup_{\boldsymbol{\vartheta}\in\hat{\Theta}}\Big|\frac{\partial \ln L_{11}(x|\boldsymbol{\vartheta})}{\partial \vartheta_j}\Big| + \sup_{\boldsymbol{\vartheta}\in\hat{\Theta}}\Big|\frac{\partial \ln L_{11}(y|\boldsymbol{\vartheta})}{\partial \vartheta_j}\Big| \\ &+ \sup_{\boldsymbol{\vartheta}\in\hat{\Theta}}\Big|\frac{\partial^2 \ln L_{11}(x|\boldsymbol{\vartheta})}{\partial \vartheta_i \partial \vartheta_j}\Big| + \sup_{\boldsymbol{\vartheta}\in\hat{\Theta}}\Big|\frac{\partial^2 \ln L_{11}(y|\boldsymbol{\vartheta})}{\partial \vartheta_i \partial \vartheta_j}\Big| \\ &+ \Big[\sup_{\boldsymbol{\vartheta}\in\hat{\Theta}}\Big|\frac{\partial \ln L_{11}(x)}{\partial \vartheta_i}\Big| + \sup_{\boldsymbol{\vartheta}\in\hat{\Theta}}\Big|\frac{\partial \ln L_{11}(y)}{\partial \vartheta_i}\Big|\Big] \\ &\quad \cdot \Big[\sup_{\boldsymbol{\vartheta}\in\hat{\Theta}}\Big|\frac{\partial \ln L_{11}(x)}{\partial \vartheta_j}\Big| + \sup_{\boldsymbol{\vartheta}\in\hat{\Theta}}\Big|\frac{\partial \ln L_{11}(y)}{\partial \vartheta_j}\Big|\Big]\Big)\,.\end{aligned}$$

Since $g_{0,\alpha}(x,y|\boldsymbol{\vartheta}) \leq S_0^{\alpha}(x|\boldsymbol{\vartheta})$ and $g_{0,\alpha}(x,y|\boldsymbol{\vartheta}) \leq S_0^{\alpha}(y|\boldsymbol{\vartheta})$ applying (B.44), (B.45) and (B.46) leads to

$$\begin{aligned} M_0(x,y)\cdot g_{0,\alpha}(x,y|\boldsymbol{\vartheta}) &\leq 8\,KK_0 = \tilde{c}_0 \qquad \text{and} \\ \tilde{M}_0(x,y)\cdot g_{0,\alpha}(x,y|\boldsymbol{\vartheta}) &\leq 8\,KK_0(2+K_0) = \tilde{d}_0\ . \end{aligned}$$

□

Proof of lemma 12

From (A.32) and (A.33) it follows that

$$\ln g_{1,\alpha}(x,y|\boldsymbol{\vartheta}) = \ln \mu(x|\boldsymbol{\lambda}) + \ln R_1(x,y|\boldsymbol{\vartheta}) + \alpha\ \ln g_{0,1}(x,y|\boldsymbol{\vartheta}) \qquad \text{(B.64)}$$

with

$$R_1(x,y|\boldsymbol{\vartheta}) = \frac{1-\varrho_1}{L_{11}(x|\boldsymbol{\vartheta})} + \frac{\varrho_1}{L_4(x,y|\boldsymbol{\vartheta})} = \frac{\tilde{k}_1(\boldsymbol{\vartheta})}{L_{11}(x|\boldsymbol{\vartheta})} + \frac{\tilde{k}_0(\boldsymbol{\vartheta})}{L_4(x,y|\boldsymbol{\vartheta})} \ . \tag{B.65}$$

For $l = 0, 1$ it is obvious that

$$0 \leq \tilde{k}_l(\boldsymbol{\vartheta}) \leq 1, \quad \Big|\frac{\partial \tilde{k}_l(\boldsymbol{\vartheta})}{\partial \vartheta_i}\Big| \leq 1 \quad \text{and} \quad \frac{\partial^2 \tilde{k}_l(\boldsymbol{\vartheta})}{\partial \vartheta_i \partial \vartheta_j} = 0 \quad . \tag{B.66}$$

It holds that

$$\frac{1}{R_1(x,y|\boldsymbol{\vartheta})}\left[\frac{1}{L_{11}(x|\boldsymbol{\vartheta})} + \frac{1}{L_4(x,y|\boldsymbol{\vartheta})}\right] \leq \frac{2}{1-\varrho_1} \ . \tag{B.67}$$

Lemma 11 and (B.64) yield

$$\Big|\frac{\partial \ln g_{1,\alpha}(x,y|\boldsymbol{\vartheta})}{\partial \vartheta_i}\Big| \leq \Big|\frac{\partial \ln \mu(x|\boldsymbol{\lambda})}{\partial \vartheta_i}\Big| + \Big|\frac{\partial \ln R_1(x,y|\boldsymbol{\vartheta})}{\partial \vartheta_i}\Big| + M_0(x,y) \quad \text{and} \tag{B.68}$$

$$\Big|\frac{\partial^2 \ln g_{1,\alpha}(x,y|\boldsymbol{\vartheta})}{\partial \vartheta_i \partial \vartheta_j}\Big| \leq \Big|\frac{\partial^2 \ln \mu(x|\boldsymbol{\lambda})}{\partial \vartheta_i \partial \vartheta_j}\Big| + \Big|\frac{\partial^2 \ln R_1(x,y|\boldsymbol{\vartheta})}{\partial \vartheta_i \partial \vartheta_j}\Big| + \tilde{M}_0(x,y) \ . \tag{B.69}$$

From (B.65) it follows that

$$\frac{\partial \ln R_1}{\partial \vartheta_i} = \frac{1}{R_1}\left(\frac{\partial \tilde{k}_1}{\partial \vartheta_i}\frac{1}{L_{11}} + \frac{\partial \tilde{k}_0}{\partial \vartheta_i}\frac{1}{L_4} - \frac{\tilde{k}_1}{L_{11}}\frac{\partial \ln L_{11}}{\partial \vartheta_i} - \frac{\tilde{k}_0}{L_4}\frac{\partial \ln L_4}{\partial \vartheta_i}\right) \ .$$

Using (B.66), (B.42) and (B.67) one can conclude

$$\begin{aligned}
&\Big|\frac{\partial \ln R_1(x,y|\boldsymbol{\vartheta})}{\partial \vartheta_i}\Big| \\
&\quad \leq \frac{1}{R_1(x,y|\boldsymbol{\vartheta})}\left(\frac{1}{L_{11}(x|\boldsymbol{\vartheta})} + \frac{1}{L_4(x,y|\boldsymbol{\vartheta})} + \frac{\left|\frac{\partial \ln L_{11}(x|\boldsymbol{\vartheta})}{\partial \vartheta_i}\right|}{L_{11}(x|\boldsymbol{\vartheta})} + \frac{\left|\frac{\partial \ln L_4(x,y|\boldsymbol{\vartheta})}{\partial \vartheta_i}\right|}{L_4(x,y|\boldsymbol{\vartheta})}\right) \\
&\quad \leq \frac{\frac{1}{L_{11}(x|\boldsymbol{\vartheta})} + \frac{1}{L_4(x,y|\boldsymbol{\vartheta})}}{R_1(x,y|\boldsymbol{\vartheta})}\left(1 + \Big|\frac{\partial \ln L_{11}(x|\boldsymbol{\vartheta})}{\partial \vartheta_i}\Big| + \Big|\frac{\partial \ln L_{11}(y|\boldsymbol{\vartheta})}{\partial \vartheta_i}\Big|\right) \\
&\quad \leq \frac{2}{1-\varrho_1}\left(1 + \Big|\frac{\partial \ln L_{11}(x|\boldsymbol{\vartheta})}{\partial \vartheta_i}\Big| + \Big|\frac{\partial \ln L_{11}(y|\boldsymbol{\vartheta})}{\partial \vartheta_i}\Big|\right) \ .
\end{aligned} \tag{B.70}$$

(B.68) and (B.70) yield that

$$\Big|\frac{\partial \ln g_{1,\alpha}(x,y|\boldsymbol{\vartheta})}{\partial \vartheta_i}\Big| \leq \frac{M_1(x,y)}{1-\varrho_1}$$

where

$$M_1(x,y) = \sup_{\boldsymbol{\lambda}\in\Lambda}\Big|\frac{\partial \ln \mu(x|\boldsymbol{\lambda})}{\partial\vartheta_i}\Big| + M_0(x,y) + 2\left(1 + \sup_{\boldsymbol{\vartheta}\in\hat{\Theta}}\Big|\frac{\partial \ln L_{11}(x|\boldsymbol{\vartheta})}{\partial\vartheta_i}\Big| + \sup_{\boldsymbol{\vartheta}\in\hat{\Theta}}\Big|\frac{\partial \ln L_{11}(y|\boldsymbol{\vartheta})}{\partial\vartheta_i}\Big|\right) .$$

Since $g_{0,\tilde{\alpha}}(x,y|\boldsymbol{\vartheta}) \le S_0^{\tilde{\alpha}}(x|\boldsymbol{\vartheta})$ and $g_{0,\tilde{\alpha}}(x,y|\boldsymbol{\vartheta}) \le S_0^{\tilde{\alpha}}(y|\boldsymbol{\vartheta})$, applying assumption 7(d), (B.45) and lemma 11 leads to

$$M_1(x,y)\cdot g_{0,\tilde{\alpha}}(x,y|\boldsymbol{\vartheta}) \le \hat{C}_1 + \tilde{c}_0 + 2\,(1+2\,K_0) = \tilde{c}_1 .$$

It holds that

$$\frac{\partial^2 \ln R_1}{\partial\vartheta_i\partial\vartheta_j} = \frac{1}{R_1}\frac{\partial^2 R_1}{\partial\vartheta_i\partial\vartheta_j} + \frac{\partial \ln R_1}{\partial\vartheta_i}\frac{\partial \ln R_1}{\partial\vartheta_j} . \tag{B.71}$$

(B.65) leads to

$$\begin{aligned}\frac{\partial^2 R_1}{\partial\vartheta_i\partial\vartheta_j} = &-\frac{\partial\tilde{k}_1}{\partial\vartheta_i}\frac{1}{L_{11}}\frac{\partial \ln L_{11}}{\partial\vartheta_j} - \frac{\partial\tilde{k}_0}{\partial\vartheta_i}\frac{1}{L_4}\frac{\partial \ln L_4}{\partial\vartheta_j} - \frac{\partial\tilde{k}_1}{\partial\vartheta_j}\frac{1}{L_{11}}\frac{\partial \ln L_{11}}{\partial\vartheta_i}\\ &- \frac{\partial\tilde{k}_0}{\partial\vartheta_j}\frac{1}{L_4}\frac{\partial \ln L_4}{\partial\vartheta_i} + \frac{\tilde{k}_1}{L_{11}}\frac{\partial \ln L_{11}}{\partial\vartheta_i}\frac{\partial \ln L_{11}}{\partial\vartheta_j} + \frac{\tilde{k}_0}{L_4}\frac{\partial \ln L_4}{\partial\vartheta_i}\frac{\partial \ln L_4}{\partial\vartheta_j}\\ &- \frac{\tilde{k}_1}{L_{11}}\frac{\partial^2 \ln L_{11}}{\partial\vartheta_i\partial\vartheta_j} - \frac{\tilde{k}_0}{L_4}\frac{\partial^2 \ln L_4}{\partial\vartheta_i\partial\vartheta_j}.\end{aligned}$$

From (B.42), (B.49) and (B.66) thus follows

$$\begin{aligned}\Big|\frac{\partial^2 R_1(x,y|\boldsymbol{\vartheta})}{\partial\vartheta_i\partial\vartheta_j}\Big| \le &\left[\frac{1}{L_{11}(x|\boldsymbol{\vartheta})} + \frac{1}{L_4(x,y|\boldsymbol{\vartheta})}\right]\\ &\cdot\left\{\Big|\frac{\partial \ln L_{11}(x|\boldsymbol{\vartheta})}{\partial\vartheta_i}\Big| + \Big|\frac{\partial \ln L_{11}(y|\boldsymbol{\vartheta})}{\partial\vartheta_i}\Big| + \Big|\frac{\partial \ln L_{11}(x|\boldsymbol{\vartheta})}{\partial\vartheta_j}\Big| + \Big|\frac{\partial \ln L_{11}(y|\boldsymbol{\vartheta})}{\partial\vartheta_j}\Big|\right.\\ &+ \left[\Big|\frac{\partial \ln L_{11}(x|\boldsymbol{\vartheta})}{\partial\vartheta_i}\Big| + \Big|\frac{\partial \ln L_{11}(y|\boldsymbol{\vartheta})}{\partial\vartheta_i}\Big|\right]\left[\Big|\frac{\partial \ln L_{11}(x|\boldsymbol{\vartheta})}{\partial\vartheta_j}\Big| + \Big|\frac{\partial \ln L_{11}(y|\boldsymbol{\vartheta})}{\partial\vartheta_j}\Big|\right]\\ &\left.+ \Big|\frac{\partial^2 \ln L_{11}(x|\boldsymbol{\vartheta})}{\partial\vartheta_i\partial\vartheta_j}\Big| + \Big|\frac{\partial^2 \ln L_{11}(y|\boldsymbol{\vartheta})}{\partial\vartheta_i\partial\vartheta_j}\Big|\right\}\end{aligned}$$

and therefore (B.67), (B.71) and (B.70) yield

$$\begin{aligned}\Big|\frac{\partial^2 \ln R_1(x|\boldsymbol{\vartheta})}{\partial\vartheta_i\partial\vartheta_j}\Big| \le &\frac{2}{1-\varrho_1}\left[\Big|\frac{\partial^2 \ln L_{11}(x|\boldsymbol{\vartheta})}{\partial\vartheta_i\partial\vartheta_j}\Big| + \Big|\frac{\partial^2 \ln L_{11}(y|\boldsymbol{\vartheta})}{\partial\vartheta_i\partial\vartheta_j}\Big|\right]\\ &+ \frac{8}{(1-\varrho_1)^2}\left(1 + \Big|\frac{\partial \ln L_{11}(x|\boldsymbol{\vartheta})}{\partial\vartheta_i}\Big| + \Big|\frac{\partial \ln L_{11}(y|\boldsymbol{\vartheta})}{\partial\vartheta_i}\Big|\right)\\ &\cdot\left(1 + \Big|\frac{\partial \ln L_{11}(x|\boldsymbol{\vartheta})}{\partial\vartheta_j}\Big| + \Big|\frac{\partial \ln L_{11}(y|\boldsymbol{\vartheta})}{\partial\vartheta_j}\Big|\right) .\end{aligned}$$

Denote $\hat{R}_i(x) = \sup_{\boldsymbol{\vartheta}\in\hat{\Theta}} \Big| \frac{\partial \ln L_{11}(x|\boldsymbol{\vartheta})}{\partial \vartheta_i} \Big|$ and $\hat{R}_{ij}(x) = \sup_{\boldsymbol{\vartheta}\in\hat{\Theta}} \Big| \frac{\partial^2 \ln L_{11}(x|\boldsymbol{\vartheta})}{\partial \vartheta_i \partial \vartheta_j} \Big|$ then, using (B.69), one can estimate

$$\Big| \frac{\partial^2 \ln g_{1,\alpha}(x,y|\boldsymbol{\vartheta})}{\partial \vartheta_i \partial \vartheta_j} \Big| \leq \frac{\tilde{M}_1(x,y)}{(1-\varrho_1)^2}$$

with

$$\tilde{M}_1(x,y) = \sup_{\boldsymbol{\vartheta}\in\hat{\Theta}} \Big| \frac{\partial^2 \ln \mu(x|\boldsymbol{\lambda})}{\partial \vartheta_i \partial \vartheta_j} \Big| + 2\,\hat{R}_{ij}(x) + 2\,\hat{R}_{ij}(y) + 8\,[1+\hat{R}_i(x)+\hat{R}_i(y)][1+\hat{R}_j(x)+\hat{R}_j(y)] + \tilde{M}_0(x,y)\ .$$

Using $g_{0,\tilde{\alpha}}(x,y|\boldsymbol{\vartheta}) \leq S_0^{\tilde{\alpha}}(x|\boldsymbol{\vartheta})$, $g_{0,\tilde{\alpha}}(x,y|\boldsymbol{\vartheta}) \leq S_0^{\tilde{\alpha}}(y|\boldsymbol{\vartheta})$, assumption 7(e), (B.45), (B.46) and lemma 11 it follows

$$\tilde{M}_1(x,y)\cdot g_{0,\tilde{\alpha}}(x,y|\boldsymbol{\vartheta}) \leq \hat{C}_2 + 4\,K_0 + 8\,[1+2\,K_0]^2 + \tilde{d}_0 = \tilde{d}_1\ .$$

□

Proof of lemma 13

Denote $\tilde{g}_{1,\alpha}(x,y|\boldsymbol{\vartheta}) = g_{1,\alpha}(y,x|\boldsymbol{\vartheta})$. From (A.34) it follows that

$$\begin{aligned}
\frac{\partial^2 \ln g_{2,1}}{\partial\vartheta_i\partial\vartheta_j} = {}& \\
& \frac{h_2\,g_{0,1}}{g_{2,1}}\Big[\frac{\partial^2\ln h_2}{\partial\vartheta_i\partial\vartheta_j} + \frac{\partial^2\ln g_{0,1}}{\partial\vartheta_i\partial\vartheta_j} + \Big(\frac{\partial\ln h_2}{\partial\vartheta_i}+\frac{\partial\ln g_{0,1}}{\partial\vartheta_i}\Big)\Big(\frac{\partial\ln h_2}{\partial\vartheta_j}+\frac{\partial\ln g_{0,1}}{\partial\vartheta_j}\Big)\Big] \\
+ & \frac{g_{1,\frac12}\,\tilde{g}_{1,\frac12}}{g_{2,1}}\Big[\frac{\partial^2\ln g_{1,\frac12}}{\partial\vartheta_i\partial\vartheta_j} + \frac{\partial^2\ln \tilde{g}_{1,\frac12}}{\partial\vartheta_i\partial\vartheta_j} + \Big(\frac{\partial\ln g_{1,\frac12}}{\partial\vartheta_i}+\frac{\partial\ln \tilde{g}_{1,\frac12}}{\partial\vartheta_i}\Big)\Big(\frac{\partial\ln g_{1,\frac12}}{\partial\vartheta_j}+\frac{\partial\ln \tilde{g}_{1,\frac12}}{\partial\vartheta_j}\Big)\Big] \\
& - \frac{1}{g_{2,1}^2}\Big[\Big(\frac{\partial\ln h_2}{\partial\vartheta_i}+\frac{\partial\ln g_{0,1}}{\partial\vartheta_i}\Big)h_2\,g_{0,1} + \Big(\frac{\partial\ln g_{1,\frac12}}{\partial\vartheta_i}+\frac{\partial\ln \tilde{g}_{1,\frac12}}{\partial\vartheta_i}\Big)g_{1,\frac12}\,\tilde{g}_{1,\frac12}\Big] \\
& \quad\cdot\Big[\Big(\frac{\partial\ln h_2}{\partial\vartheta_j}+\frac{\partial\ln g_{0,1}}{\partial\vartheta_j}\Big)h_2\,g_{0,1} + \Big(\frac{\partial\ln g_{1,\frac12}}{\partial\vartheta_j}+\frac{\partial\ln \tilde{g}_{1,\frac12}}{\partial\vartheta_j}\Big)g_{1,\frac12}\,\tilde{g}_{1,\frac12}\Big]\ .
\end{aligned}$$

Using the notation $N_1(x,y) = M_1(y,x)$ and $\tilde{N}_1(x,y) = \tilde{M}_1(y,x)$, lemma 11 and lemma 12 one can thus estimate (the dependence on x, y and ϑ is omitted

here)

$$\begin{aligned}\Big|\frac{\partial^2 \ln g_{2,1}}{\partial\vartheta_i\partial\vartheta_j}\Big| \leq{}& \Big|\frac{\partial^2 \ln h_2}{\partial\vartheta_i\partial\vartheta_j}\Big| + \tilde{M}_0 + \left[\Big|\frac{\partial \ln h_2}{\partial\vartheta_i}\Big| + M_0\right]\left[\Big|\frac{\partial \ln h_2}{\partial\vartheta_j}\Big| + M_0\right] \\ &+ \frac{\tilde{M}_1 + \tilde{N}_1 + (M_1+N_1)^2}{(1-\varrho_1)^2} \\ &+ \left[\Big|\frac{\partial \ln h_2}{\partial\vartheta_i}\Big| + M_0 + \frac{M_1+N_1}{1-\varrho_1}\right]\left[\Big|\frac{\partial \ln h_2}{\partial\vartheta_j}\Big| + M_0 + \frac{M_1+N_1}{1-\varrho_1}\right] .\end{aligned} \tag{B.72}$$

Equation (A.35) yields

$$\ln h_2(x,y|\boldsymbol{\vartheta}) = 2\ln\sigma + \ln\varrho_1 + \ln\mu(x|\boldsymbol{\vartheta}) + \ln\mu(y|\boldsymbol{\vartheta}) - 2\ln L_4(x,y|\boldsymbol{\vartheta})$$

and therefore with (B.42) it follows that

$$\begin{aligned}\Big|\frac{\partial \ln h_2(x,y|\boldsymbol{\vartheta})}{\partial\vartheta_i}\Big| \leq{}& \frac{2}{m_s} + \frac{1}{\varrho_1} + \Big|\frac{\partial \ln\mu(x|\boldsymbol{\vartheta})}{\partial\vartheta_i}\Big| + \Big|\frac{\partial \ln\mu(y|\boldsymbol{\vartheta})}{\partial\vartheta_i}\Big| \\ &+ 2\Big|\frac{\partial \ln L_{11}(x|\boldsymbol{\vartheta})}{\partial\vartheta_i}\Big| + 2\Big|\frac{\partial \ln L_{11}(y|\boldsymbol{\vartheta})}{\partial\vartheta_i}\Big|\end{aligned} \tag{B.73}$$

and with (B.49) it follows that

$$\begin{aligned}\Big|\frac{\partial^2 \ln h_2(x,y|\boldsymbol{\vartheta})}{\partial\vartheta_i\partial\vartheta_j}\Big| \leq{}& \frac{2}{m_s^2} + \frac{1}{\varrho_1^2} + \Big|\frac{\partial^2 \ln\mu(x|\boldsymbol{\vartheta})}{\partial\vartheta_i\partial\vartheta_j}\Big| + \Big|\frac{\partial^2 \ln\mu(y|\boldsymbol{\vartheta})}{\partial\vartheta_i\partial\vartheta_j}\Big| \\ &+ \Big|\frac{\partial^2 \ln L_{11}(x|\boldsymbol{\vartheta})}{\partial\vartheta_i\partial\vartheta_j}\Big| + \Big|\frac{\partial^2 \ln L_{11}(y|\boldsymbol{\vartheta})}{\partial\vartheta_i\partial\vartheta_j}\Big| \\ &+2\left[\Big|\frac{\partial \ln L_{11}(x|\boldsymbol{\vartheta})}{\partial\vartheta_i}\Big| + \Big|\frac{\partial \ln L_{11}(y|\boldsymbol{\vartheta})}{\partial\vartheta_i}\Big|\right]\left[\Big|\frac{\partial \ln L_{11}(x|\boldsymbol{\vartheta})}{\partial\vartheta_j}\Big| + \Big|\frac{\partial \ln L_{11}(y|\boldsymbol{\vartheta})}{\partial\vartheta_j}\Big|\right] .\end{aligned} \tag{B.74}$$

If one denotes $\hat{R}_i(x) = \sup_{\boldsymbol{\vartheta}\in\hat{\Theta}}\Big|\frac{\partial \ln L_{11}(x|\boldsymbol{\vartheta})}{\partial\vartheta_i}\Big|$ and $\hat{R}_{ij}(x) = \sup_{\boldsymbol{\vartheta}\in\hat{\Theta}}\Big|\frac{\partial^2 \ln L_{11}(x|\boldsymbol{\vartheta})}{\partial\vartheta_i\partial\vartheta_j}\Big|$, $\tilde{R}_i(x) = \sup_{\boldsymbol{\vartheta}\in\hat{\Theta}}\Big|\frac{\partial \ln\mu(x|\boldsymbol{\vartheta})}{\partial\vartheta_i}\Big|$ and $\tilde{R}_{ij}(x) = \sup_{\boldsymbol{\vartheta}\in\hat{\Theta}}\Big|\frac{\partial^2 \ln\mu(x|\boldsymbol{\vartheta})}{\partial\vartheta_i\partial\vartheta_j}\Big|$ then one can conclude from (B.72), (B.73) and (B.74) that

$$\Big|\frac{\partial^2 \ln g_{2,1}(x,y|\boldsymbol{\vartheta})}{\partial\vartheta_i\partial\vartheta_j}\Big| \leq \frac{\tilde{M}_2(x,y)}{(1-\varrho_1)^2}$$

where

$$\begin{aligned}\tilde{M}_2(x,y) &= \tfrac{1}{m_r^2} + \tfrac{2}{m_s^2} + \tilde{R}_{ij}(x) + \tilde{R}_{ij}(y) + \hat{R}_{ij}(x) + \hat{R}_{ij}(y) \\ &\quad + 2(\hat{R}_i(x) + \hat{R}_i(y))(\hat{R}_j(x) + \hat{R}_j(y)) + \tilde{M}_0(x,y) \\ &\quad + \Big[\tfrac{1}{m_r} + \tfrac{2}{m_2} + \tilde{R}_i(x) + \tilde{R}_i(y) + 2\,\hat{R}_i(x) + 2\,\hat{R}_i(y) + M_0(x,y)\Big] \\ &\qquad \Big[\tfrac{1}{m_r} + \tfrac{2}{m_2} + \tilde{R}_j(x) + \tilde{R}_j(y) + 2\,\hat{R}_j(x) + 2\,\hat{R}_j(y) + M_0(x,y)\Big] \\ &\quad + \tilde{M}_1(x,y) + \tilde{M}_1(y,x) + (M_1(x,y) + M_1(y,x))^2 \\ &+ \Big[\tfrac{1}{m_r} + \tfrac{2}{m_2} + \tilde{R}_i(x) + \tilde{R}_i(y) + 2\hat{R}_i(x) + 2\hat{R}_i(y) + M_0(x,y) + M_1(x,y) + M_1(y,x)\Big] \\ &\cdot \Big[\tfrac{1}{m_r} + \tfrac{2}{m_2} + \tilde{R}_j(x) + \tilde{R}_j(y) + 2\hat{R}_j(x) + 2\hat{R}_j(y) + M_0(x,y) + M_1(x,y) + M_1(y,x)\Big].\end{aligned}$$

Using $g_{0,1}(x,y|\boldsymbol{\vartheta}) \leq S_0(x|\boldsymbol{\vartheta})$, $g_{0,1}(x,y|\boldsymbol{\vartheta}) \leq S_0(y|\boldsymbol{\vartheta})$, the assumptions 7(d) and 7(e), (B.45), (B.46) and the lemmata 11 and 12 one can finally conclude that

$$\begin{aligned}\tilde{M}_2(x,y)\cdot g_{0,1}(x,y|\boldsymbol{\vartheta}) &\leq 1 + \tfrac{1}{m_s} + 2\hat{C}_2 + 2K_0 + 8K_0^2 + \tilde{d}_0 + [1 + \tfrac{2}{m_s} + 2\hat{C}_1 + 4K_0 + \tilde{C}_0]^2 \\ &\quad + 2\tilde{d}_1 + 4\tilde{c}_1^2 + [1 + \tfrac{2}{m_s} + 2\hat{C}_1 + 4K_0 + \tilde{C}_0 + 2\tilde{c}_1]^2 = \tilde{d}_2\ .\end{aligned}$$

□

Appendix C

Results of simulations

C.1 Numerical algorithms used in maximum likelihood estimation

The maximum likelihood estimation described in this work was performed with the help of the software GAUSS (http://www.aptech.com). GAUSS is a data analysis environment based on the GAUSS Matrix Programming Language. It provides several application packages, among them modules for maximum likelihood estimation (MAXLIK) and constrained maximum likelihood estimation (CML). This section gives a short and very general introduction to the numerical algorithms that have been used. A detailed description can be found in GAUSS Applications (1996) or Judd (1998).

C.1.1 Maximization of the log-likelihood function

The maximum likelihood package (MAXLIK) includes several procedures that iteratively find the maximum of the log-likelihood function

$$\ln L(\mathbf{x}|\boldsymbol{\theta}) = \ln L(x_1, \ldots, x_n|\boldsymbol{\theta}) = \sum_{i=1}^{n} \ln f(x_i|\boldsymbol{\theta})$$

with respect to $\boldsymbol{\theta}$. Here $f(.|\boldsymbol{\theta})$ is the density function and $x_1, \ldots, x_n$ are the individual observations.
If $\boldsymbol{\theta}_k$ is the current parameter vector, then the succeeding vector is

$$\boldsymbol{\theta}_{k+1} = \boldsymbol{\theta}_k + \rho_k \boldsymbol{\delta}_k$$

where $\boldsymbol{\delta}_k$ is the direction vector and ρ_k is the scalar step length. The starting value $\boldsymbol{\theta}_0$ is provided by the user.

There are several methods of finding a direction vector and a step length in each iteration. Define

$$\Psi(\boldsymbol{\theta}) = \frac{\partial \ln L}{\partial \boldsymbol{\theta}} \quad \text{and} \quad \Sigma(\boldsymbol{\theta}) = \frac{\partial^2 \ln L}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}$$

to be the gradient and the Hessian matrix of the log-likelihood function, respectively. $\Psi(\boldsymbol{\theta})$ is the vector of the first partial derivatives and $\Sigma(\boldsymbol{\theta})$ is the matrix of the second partial derivatives of $\ln L$ with respect to the components of $\boldsymbol{\theta}$.
The most important methods of calculating a direction vector provided by GAUSS are the Newton-Algorithm and the BFGS-Algorithm. The Newton-Algorithm is based on an approximation of $\ln L$ by a quadratic function. The new direction vector is calculated as

$$\boldsymbol{\delta}_k = \Sigma(\boldsymbol{\theta}_k)^{-1} \Psi(\boldsymbol{\theta}_k) \ .$$

It may be very expensive to compute the Hessian matrix or its inverse at every iteration, and in addition the Hessian matrix might be ill-conditioned. For these reasons other algorithms exist that do not calculate the Hessian matrix directly at each iteration, but rather approximate it by another positive definite matrix. One of them is the algorithm of Broyden, Fletcher, Goldfarb, and Shanno (BFGS). In this algorithm the new direction is calculated as

$$\boldsymbol{\delta}_k = \mathbf{F}_k \Psi(\boldsymbol{\theta}_k)$$

where $\mathbf{F}_k$ is a positive definite matrix which is easier to compute. The matrix $\mathbf{F}_k$ is updated after calculating the vector $\boldsymbol{\theta}_{k+1}$ as

$$\mathbf{F}_{k+1} = \mathbf{F}_k - \frac{\mathbf{F}_k (\boldsymbol{\theta}_{k+1} - \boldsymbol{\theta}_k)(\boldsymbol{\theta}_{k+1} - \boldsymbol{\theta}_k)' \mathbf{F}_k}{(\boldsymbol{\theta}_{k+1} - \boldsymbol{\theta}_k)' \mathbf{F}_k (\boldsymbol{\theta}_{k+1} - \boldsymbol{\theta}_k)} + \frac{\boldsymbol{\psi}_k \boldsymbol{\psi}_k'}{\boldsymbol{\psi}_k' (\boldsymbol{\theta}_{k+1} - \boldsymbol{\theta}_k)}$$

where $\boldsymbol{\psi}_k = \Psi(\boldsymbol{\theta}_{k+1}) - \Psi(\boldsymbol{\theta}_k)$. A common choice for the starting matrix $\mathbf{F}_0$ is the identity matrix I.
The iterative methods described of course assume the existence of the second partial derivatives of the log-likelihood function and the positive definiteness of the Hessian matrix. In GAUSS the derivatives can be supplied analytically by the user. Otherwise they are calculated numerically.
GAUSS provides different methods of computing the step length ρ_k. All these methods attempt to find a value for ρ that increases the function

$$m(\rho) = \ln L(\boldsymbol{x} | \boldsymbol{\theta}_k + \rho \boldsymbol{\delta}_k) \ .$$

A necessary condition for $\boldsymbol{\theta}^*$ to be a local maximum of $\ln L$ is $\Psi(\boldsymbol{\theta}^*) = 0$. Based on this the iterative procedure of calculating the MLE is stopped by GAUSS if the normed gradient $||\Psi(\boldsymbol{\theta}_k)||/(1 + |\ln L(\boldsymbol{\theta}_k)|)$ is small.
The numerical algorithms that are used in the maximum likelihood package (MAXLIK) of GAUSS are described in the corresponding manual (GAUSS Applications, 1996).
Many statistical models contain constraints on the parameters. One way to avoid the problems which arise with those models is to carry out the maximum likelihood estimation for a transformation of the parameter vector. This was done for the four-dimensional correlated frailty model by estimating the shape parameters of the components of the frailty variables instead of the correlations and variances of the frailty variables. However, optimization algorithms also exist, that allow for an incorporation of constraints on the parameters.
The GAUSS package Constrained Maximum Likelihood (CML) provides procedures for the calculation of MLEs of statistical models with general constraints on the parameters. These constraints can be linear or nonlinear, equality or inequality constraints. A detailed description of the algorithms can be found in GAUSS Applications (1995). The theory of constrained nonlinear optimization is summarized in Judd (1998).

C.1.2 Calculation of standard errors

As can be seen in theorem 2, under certain conditions on the density function $f(.|\boldsymbol{\theta})$ and on the parameter space, for the asymptotic distribution of the MLEs it holds that

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) \xrightarrow{\mathcal{L}} N(0, I(\boldsymbol{\theta})^{-1}) \; . \tag{C.1}$$

Here $I(\boldsymbol{\theta})$ denotes Fisher's information matrix. (C.1) is equivalent to

$$\sqrt{n}\, I(\boldsymbol{\theta})^{1/2}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) \xrightarrow{\mathcal{L}} N(0, \mathrm{I}) \; .$$

If the Law of Large Numbers (see Sen and Singer, 1993, Theorem 2.3.6) can be applied then

$$\begin{aligned} \tfrac{1}{n}\Sigma(\boldsymbol{\theta}) = \frac{1}{n}\frac{\partial^2 \ln L(\mathbf{X}|\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} &= \frac{1}{n}\sum_{i=1}^{n}\frac{\partial^2 \ln f(X_i|\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} \\ &\xrightarrow{P} E\frac{\partial^2 \ln f(X_1|\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} = I(\boldsymbol{\theta}) \; . \end{aligned}$$

Consequently $\frac{1}{\sqrt{n}}\,\Sigma(\boldsymbol{\theta})^{1/2} \xrightarrow{P} I(\boldsymbol{\theta})^{1/2}$. Then based on Hilfssatz 6.39 in Witting and Müller-Funk (1995) one can conclude that

$$\Sigma(\boldsymbol{\theta})^{1/2}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) \xrightarrow{\mathcal{L}} N(0, \mathrm{I}) \; .$$

From the existence of continuous second partial derivatives of $\ln f(.|\boldsymbol{\theta})$ with respect to all components of $\boldsymbol{\theta}$ the continuity of $\Sigma(\boldsymbol{\theta})$ in $\boldsymbol{\theta}$ follows. The consistency of the MLE and Satz 5.107 in Witting and Müller-Funk (1995) then yield $\Sigma(\hat{\boldsymbol{\theta}}_n)^{1/2} \xrightarrow{P} \Sigma(\boldsymbol{\theta})^{1/2}$. From a second application of Hilfssatz 6.39 (Witting and Müller-Funk, 1995) it therefore follows that

$$\Sigma(\hat{\boldsymbol{\theta}}_n)^{1/2}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) \xrightarrow{\mathcal{L}} N(0, \mathrm{I}) \; .$$

This implies $\mathrm{Var}(\Sigma(\hat{\boldsymbol{\theta}}_n)^{1/2}\hat{\boldsymbol{\theta}}_n) \to \mathrm{I}$ for $n \to \infty$. Hence the approximation

$$\mathrm{Var}(\hat{\boldsymbol{\theta}}_n) \approx \Sigma(\hat{\boldsymbol{\theta}}_n)^{-1}$$

is valid for large values of n. It is used to calculate (approximate) standard errors of the maximum likelihood estimates. That means that the covariance matrix of the maximum likelihood estimates is approximated with the inverse of the Hessian matrix.
This approximation is also used for constrained maximum likelihood estimation. Usually it is assumed that the standard statistical inference described above is valid as long as the parameter estimates are "sufficiently distant" from the constraint boundaries, and if the sample size is high.
Schoenberg (1996) discusses statistical inference in models with constrained parameters. He also explains the methods for analyzing the distribution of the estimated parameters which are implemented in the GAUSS module CML.

C.2 Histograms of the empirical distribution of MLEs in the four-dimensional model

On the following pages the figures which contain the histograms of the empirical distribution of the maximum likelihood estimates in the four-dimensional correlated gamma frailty model are given (see Chapter 5). In addition each histogram includes the normal density function that corresponds to the estimated values of mean and standard deviation. The interval axis shows the range of the values of the estimate whereas the scale axis shows the number of estimates with a value in a certain interval. Each figure gives the results for a different parameter. The structure of the figures is basically the same.

Each line of charts represents the histograms for a different distribution of the age at the time of observation (A). The first and the third column show the results derived in the fully parametric estimation approach, whereas the second and the fourth column do the same for the semi-parametric estimation approach. The sample size used to get the results in the first two columns was 1000 pairs and for the last two columns a sample size of either 5000 or 10,000 pairs was used.
The caption under each chart gives the interval over which A is distributed, the sample size, and the estimation approach. All charts for one parameter have the same scale so that they can be compared directly.
Since there are only MLEs for the parameter a_1 and b_1 in the fully parametric estimation approach, the results for these two parameters are presented together in figure C.9. Instead of the estimation approach the caption under each chart includes the parameter.
The figures are discussed in section 5.3.1 on page 81.

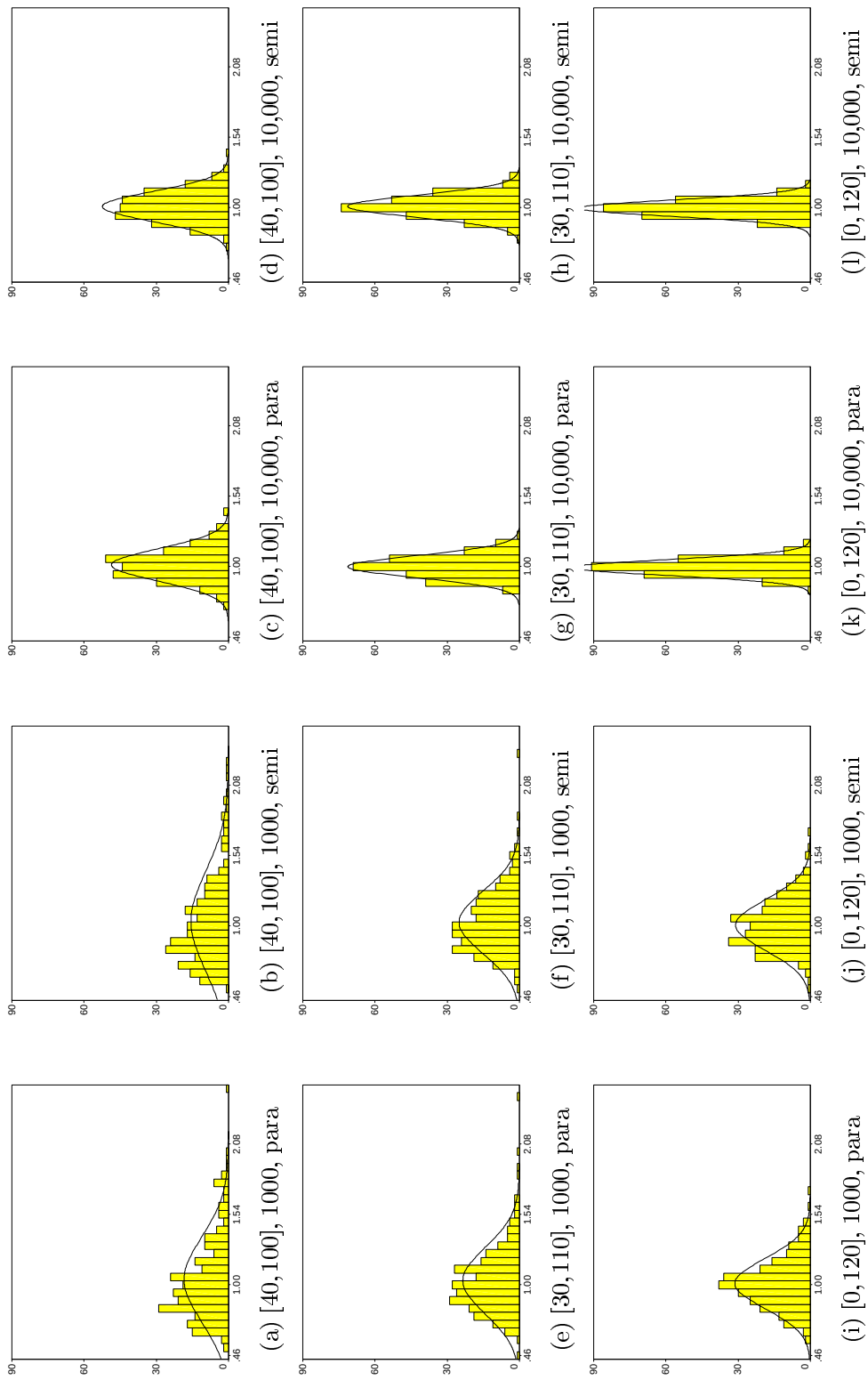
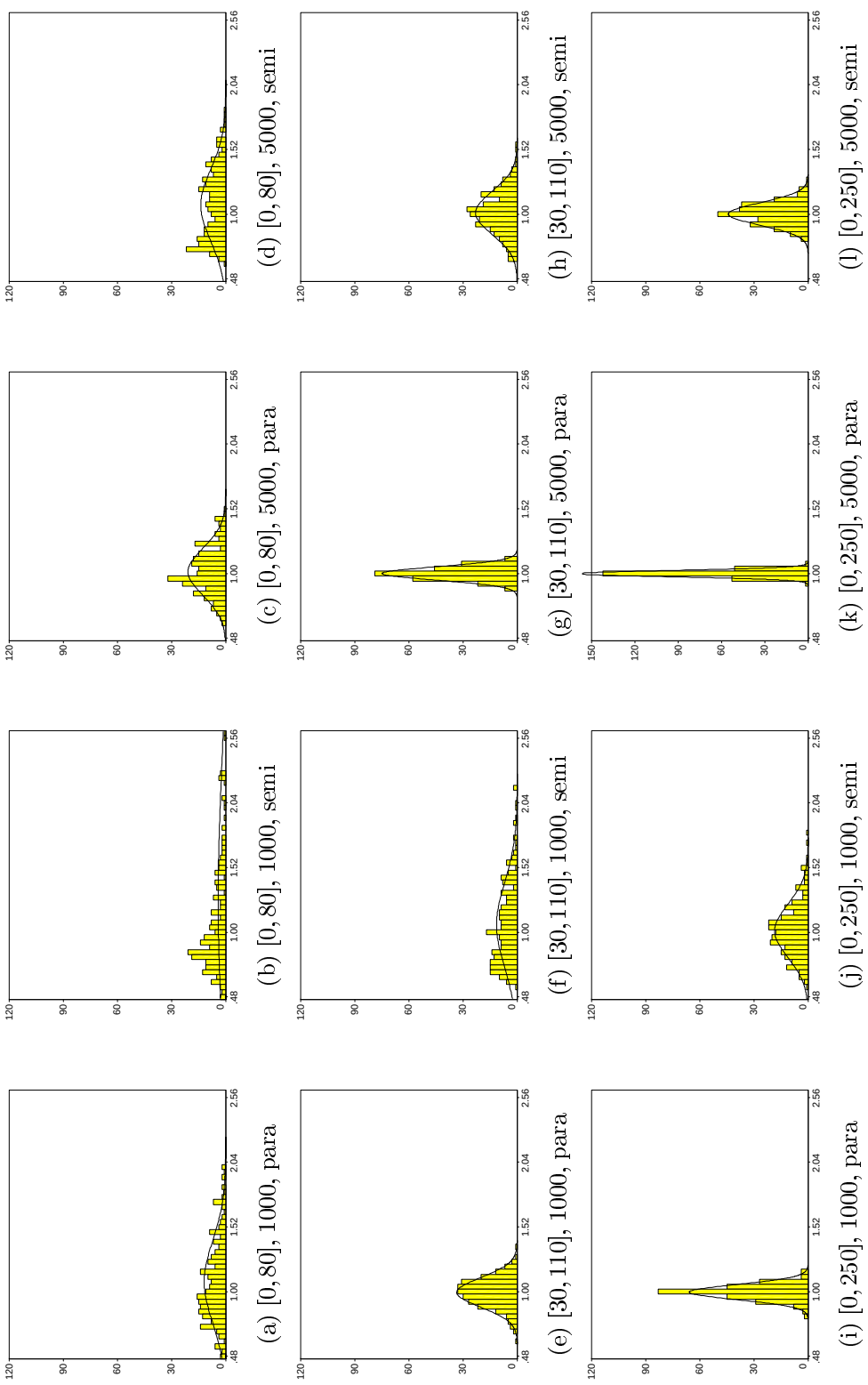


Figure C.1: Empirical distribution of the MLEs of σ_0



(a) $[0, 80]$, 1000, para (b) $[0, 80]$, 1000, semi (c) $[0, 80]$, 5000, para (d) $[0, 80]$, 5000, semi

(e) $[30, 110]$, 1000, para (f) $[30, 110]$, 1000, semi (g) $[30, 110]$, 5000, para (h) $[30, 110]$, 5000, semi

(i) $[0, 250]$, 1000, para (j) $[0, 250]$, 1000, semi (k) $[0, 250]$, 5000, para (l) $[0, 250]$, 5000, semi

Figure C.2: Empirical distribution of the MLEs of σ

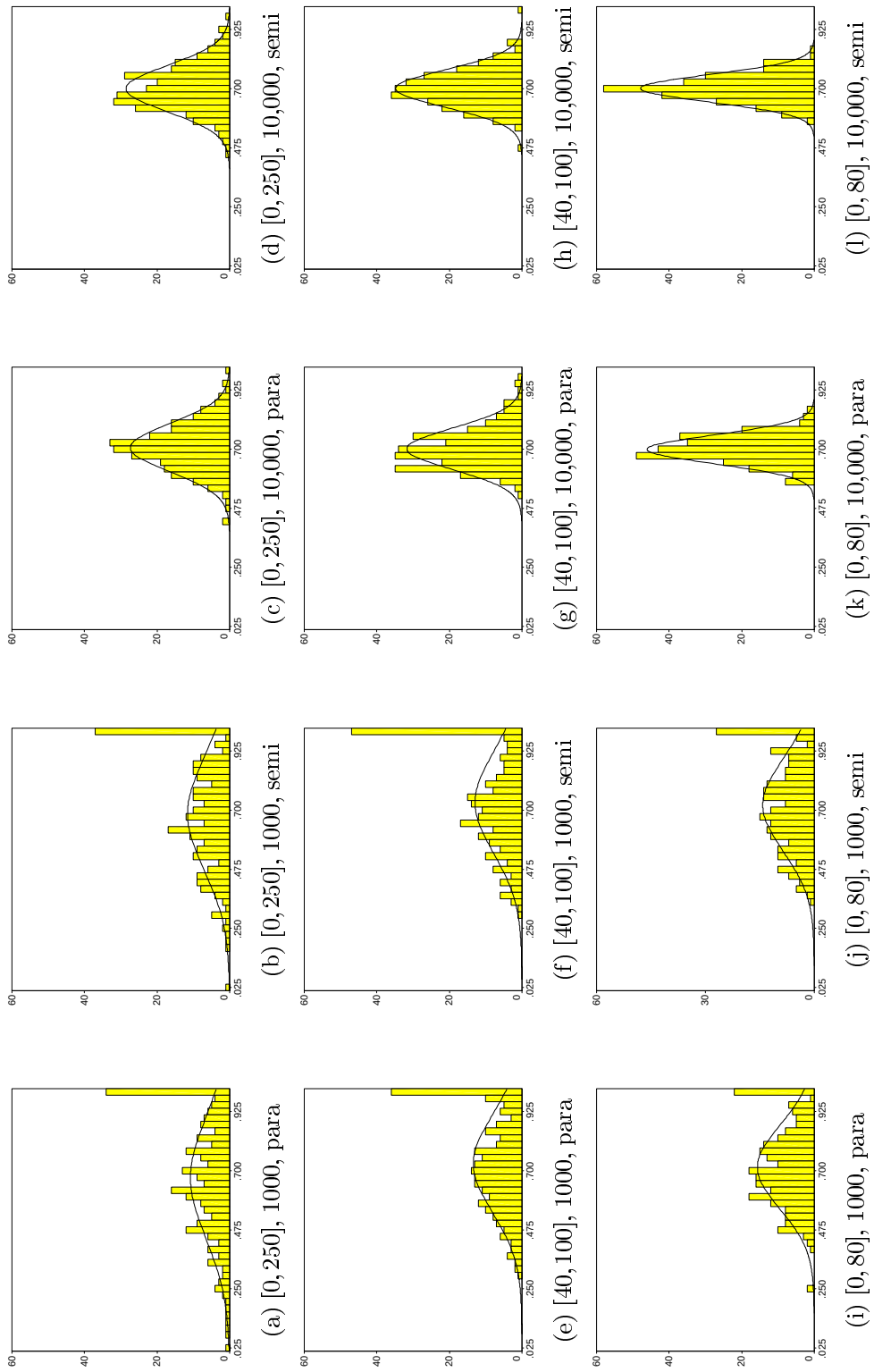
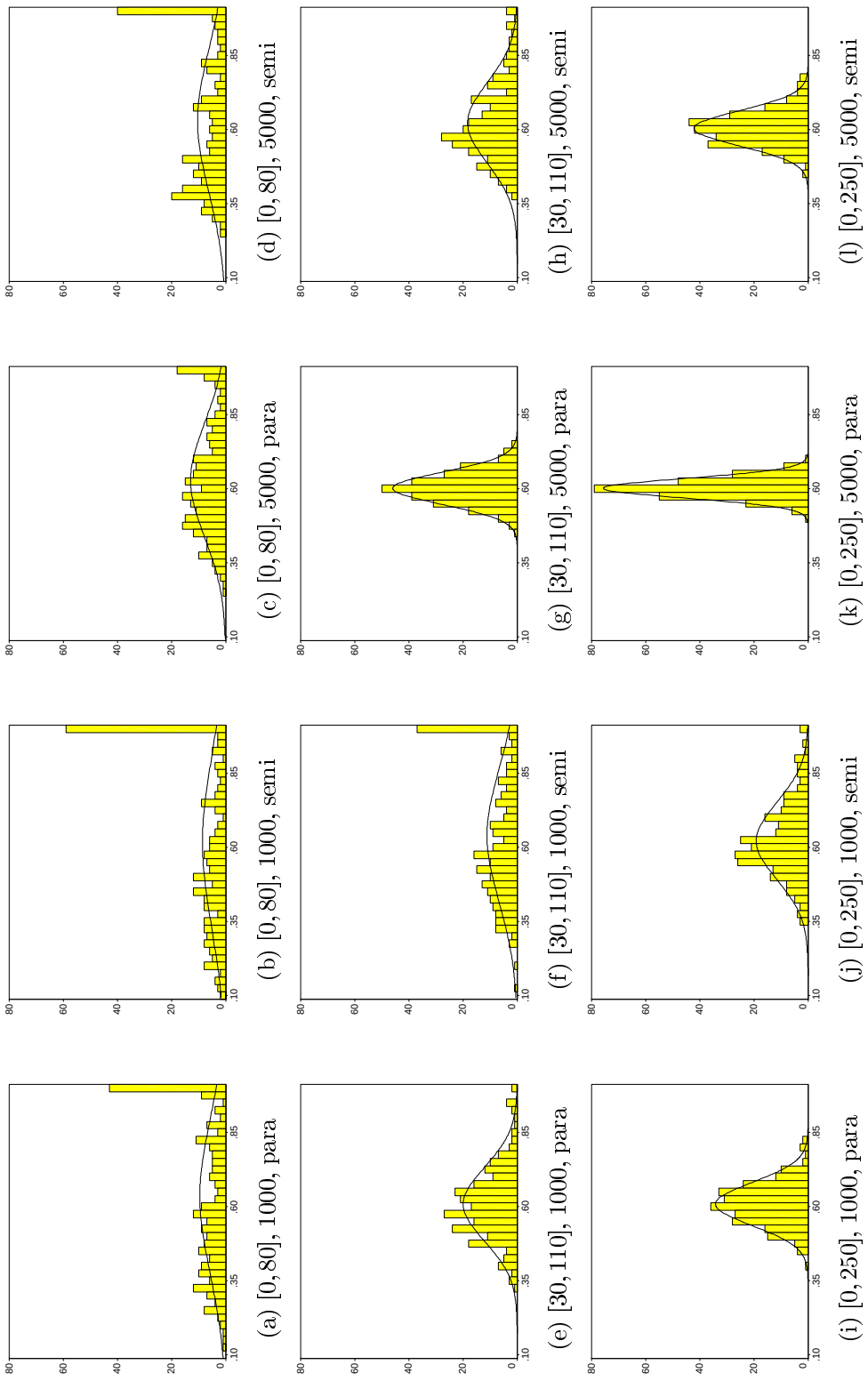


Figure C.3: Empirical distribution of the MLEs of ϱ_0

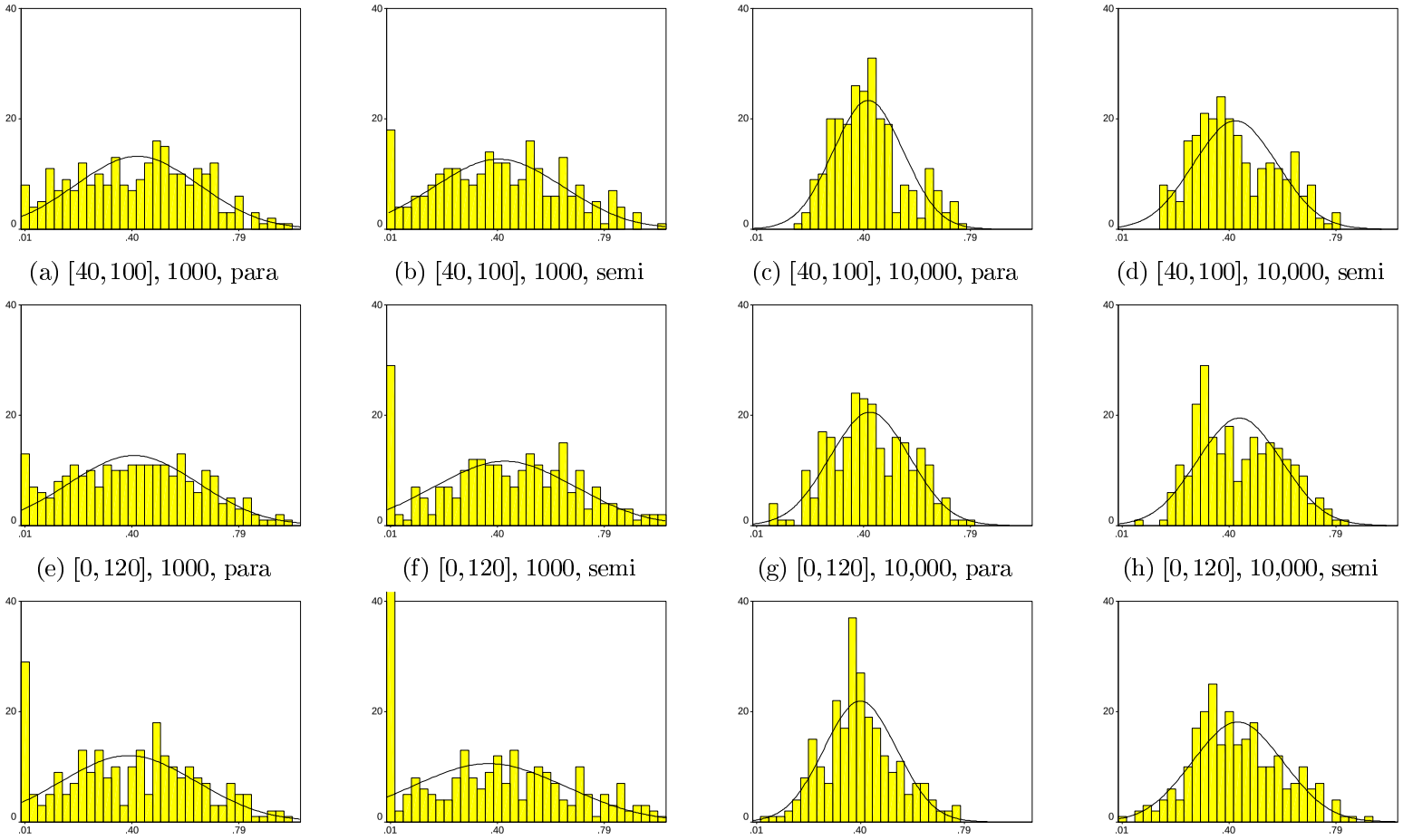


(a) $[0, 80]$, 1000, para (b) $[0, 80]$, 1000, semi (c) $[0, 80]$, 5000, para (d) $[0, 80]$, 5000, semi

(e) $[30, 110]$, 1000, para (f) $[30, 110]$, 1000, semi (g) $[30, 110]$, 5000, para (h) $[30, 110]$, 5000, semi

(i) $[0, 250]$, 1000, para (j) $[0, 250]$, 1000, semi (k) $[0, 250]$, 5000, para (l) $[0, 250]$, 5000, semi

Figure C.4: Empirical distribution of the MLEs of ϱ_1

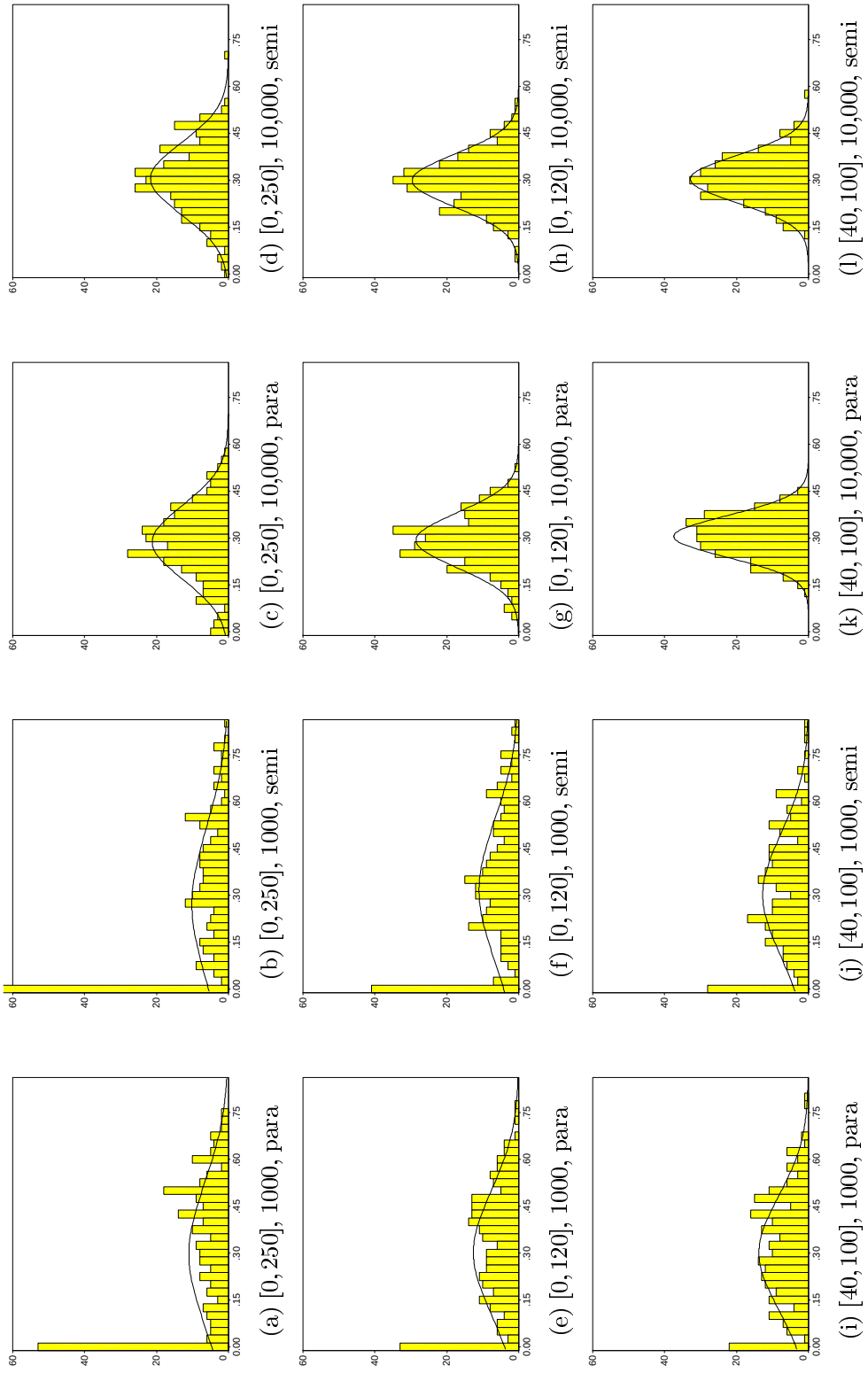


(a) [40,100], 1000, para (b) [40,100], 1000, semi (c) [40,100], 10,000, para (d) [40,100], 10,000, semi

(e) [0,120], 1000, para (f) [0,120], 1000, semi (g) [0,120], 10,000, para (h) [0,120], 10,000, semi

(i) [0,250], 1000, para (j) [0,250], 1000, semi (k) [0,250], 10,000, para (l) [0,250], 10,000, semi

Figure C.5: Empirical distribution of the MLEs of ϱ

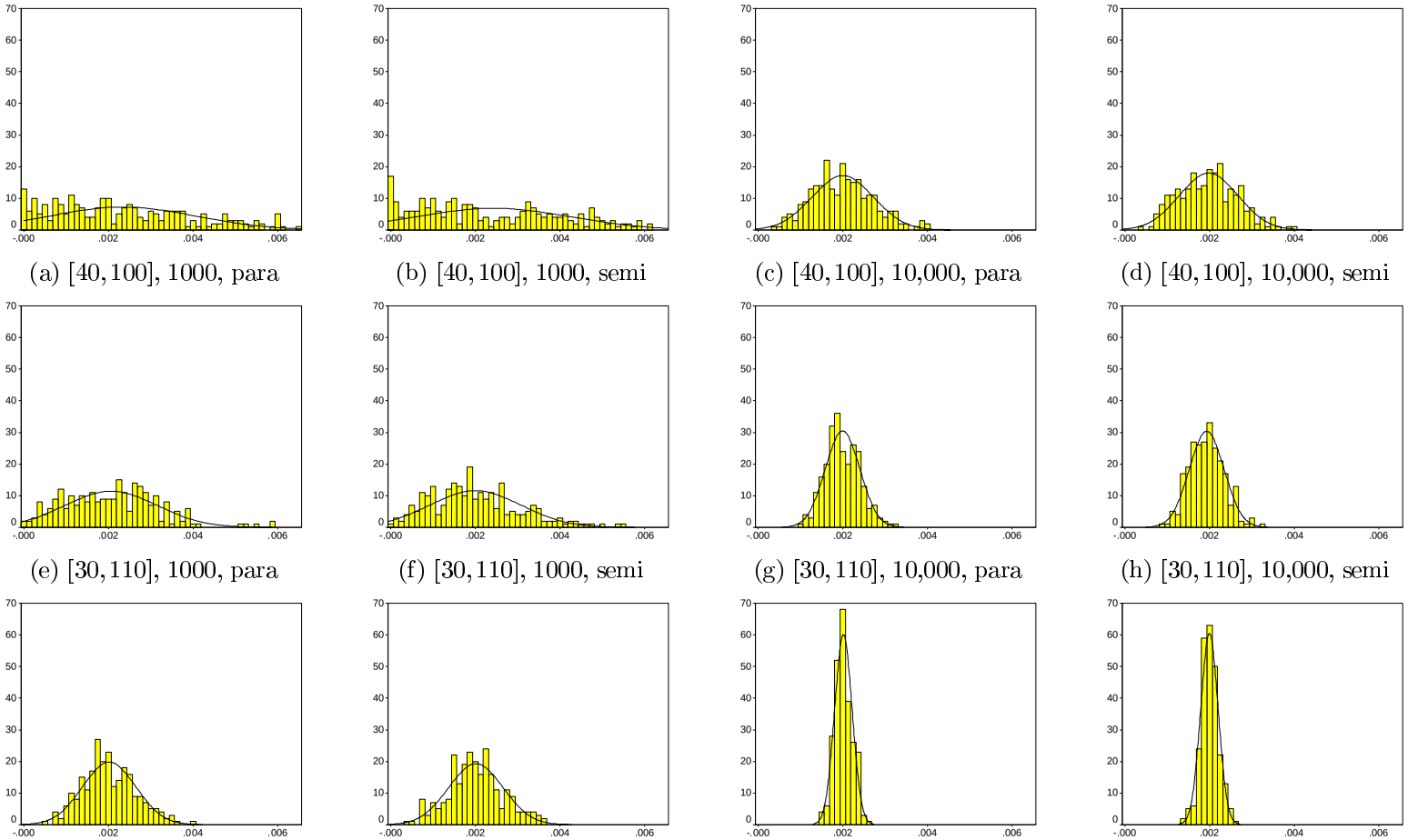


(a) $[0, 250]$, 1000, para (b) $[0, 250]$, 1000, semi (c) $[0, 250]$, 10,000, para (d) $[0, 250]$, 10,000, semi

(e) $[0, 120]$, 1000, para (f) $[0, 120]$, 1000, semi (g) $[0, 120]$, 10,000, para (h) $[0, 120]$, 10,000, semi

(i) $[40, 100]$, 1000, para (j) $[40, 100]$, 1000, semi (k) $[40, 100]$, 10,000, para (l) $[40, 100]$, 10,000, semi

Figure C.6: Empirical distribution of the MLEs of ϱ^*

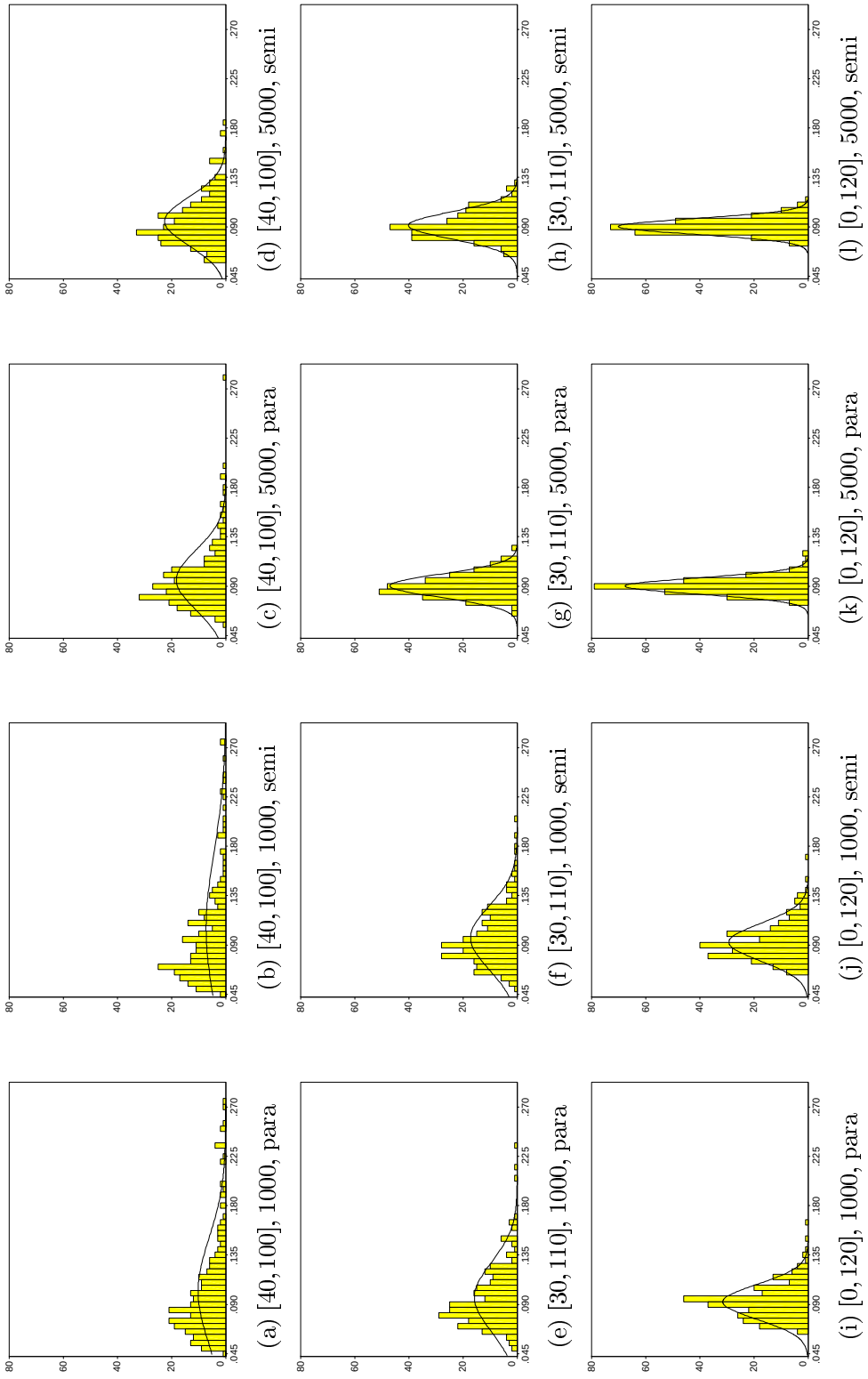


(a) [40,100], 1000, para (b) [40,100], 1000, semi (c) [40,100], 10,000, para (d) [40,100], 10,000, semi

(e) [30,110], 1000, para (f) [30,110], 1000, semi (g) [30,110], 10,000, para (h) [30,110], 10,000, semi

(i) [0,120], 1000, para (j) [0,120], 1000, semi (k) [0,120], 10,000, para (l) [0,120], 10,000, semi

Figure C.7: Empirical distribution of the MLEs of a_0

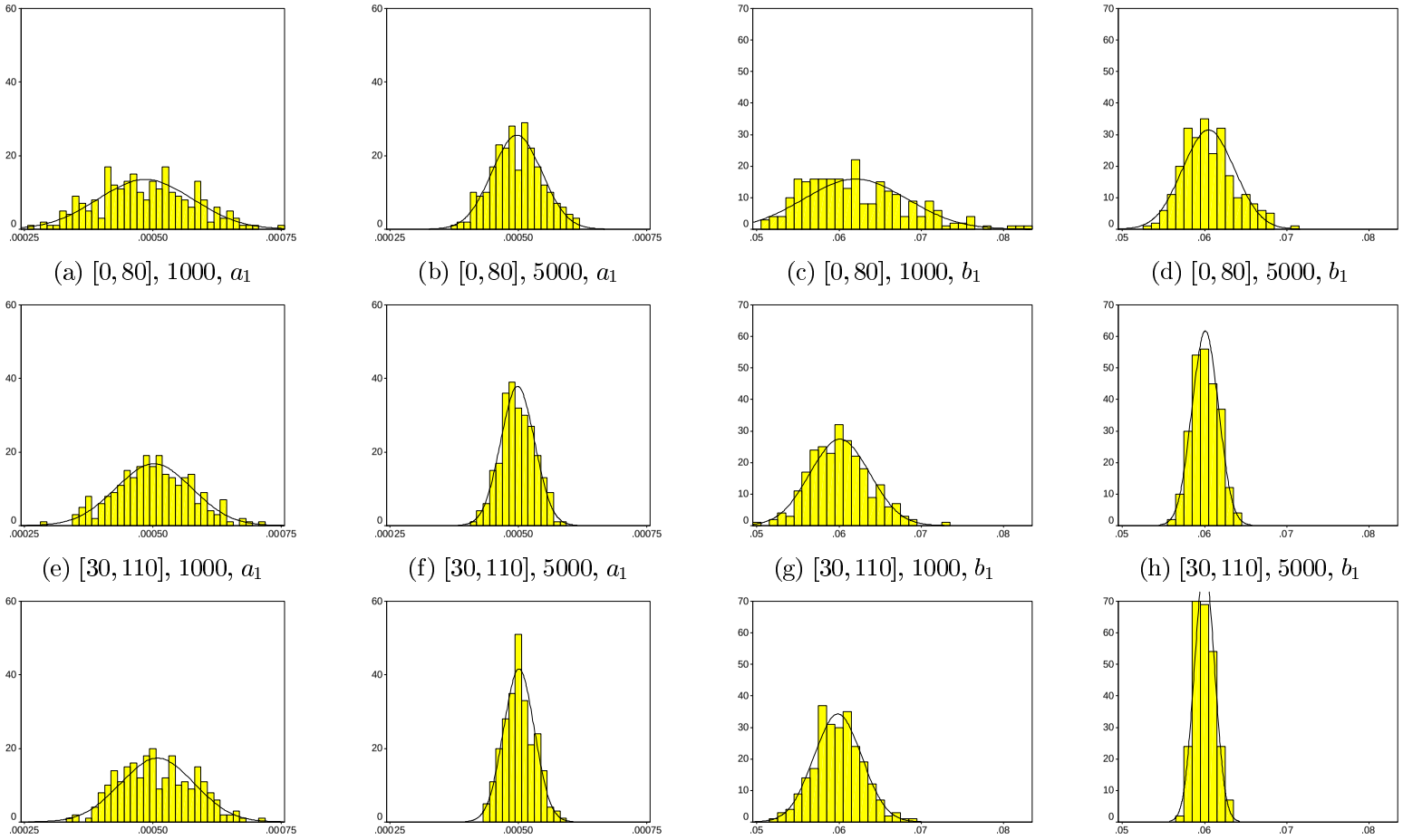


(a) $[40, 100]$, 1000, para (b) $[40, 100]$, 1000, semi (c) $[40, 100]$, 5000, para (d) $[40, 100]$, 5000, semi

(e) $[30, 110]$, 1000, para (f) $[30, 110]$, 1000, semi (g) $[30, 110]$, 5000, para (h) $[30, 110]$, 5000, semi

(i) $[0, 120]$, 1000, para (j) $[0, 120]$, 1000, semi (k) $[0, 120]$, 5000, para (l) $[0, 120]$, 5000, semi

Figure C.8: Empirical distribution of the MLEs of b_0



(a) $[0, 80]$, 1000, a_1 (b) $[0, 80]$, 5000, a_1 (c) $[0, 80]$, 1000, b_1 (d) $[0, 80]$, 5000, b_1

(e) $[30, 110]$, 1000, a_1 (f) $[30, 110]$, 5000, a_1 (g) $[30, 110]$, 1000, b_1 (h) $[30, 110]$, 5000, b_1

(i) $[0, 250]$, 1000, a_1 (j) $[0, 250]$, 5000, a_1 (k) $[0, 250]$, 1000, b_1 (l) $[0, 250]$, 5000, b_1

Figure C.9: Empirical distribution of the MLEs of a_1 and b_1

C.3 Skewness and kurtosis of the MLEs in the four-dimensional model

Table C.1 shows the empirical estimate of the skewness of the distribution of the maximum likelihood estimates in the four-dimensional correlated gamma frailty model. The empirical distribution of the MLEs was derived in the simulation study described in Chapter 5. The values are given for the different distributions of the age at the time of observation A and for the sample sizes 1000 and 10,000 pairs. In each cell of the table the upper number gives the skewness of the estimate derived in the fully parametric estimation approach and the lower number gives the value of the skewness for the estimate derived in the partly semi-parametric approach. That is the reason that there is only one line of values for the parameters of S_L, i.e. for a_1 and b_1.
In table C.2 one can find the empirical estimate of the kurtosis of the distribution of the maximum likelihood estimates in the four-dimensional correlated gamma frailty model. The results are given again for the different distributions of A, the sample sizes 1000 and 10,000 pairs and for the two different estimation approaches. The structure of the table is the same as the structure of table C.1.
Based on the values given in tables C.1 and C.2 tests for normality of the empirical distribution of the MLEs have been carried out. The asterisk (*) indicates rejection of normality with 0.05 significance level in a two sided test. The tests are described in section 1.4.2.
The results presented in tables C.1 and C.2 are discussed in section 5.3.2 on page 81.

Table C.1: Skewness of the maximum likelihood estimates. The first line for each parameter gives results from fully parametric estimation approach, the second gives results from partly semi-parametric estimation approach.

parameter	$A \sim [40, 100]$		$A \sim [30, 110]$		$A \sim [0, 120]$		$A \sim [0, 80]$		$A \sim [0, 250]$	
	1000	10,000	1000	10,000	1000	10,000	1000	10,000	1000	10,000
σ_0	1.195*	0.326*	1.422*	0.073	0.471*	0.112	0.519*	0.268	1.061*	0.137
	1.779*	0.271	1.085*	0.118	0.571*	0.219	0.303*	0.031	0.574*	0.220
σ	0.183	0.279	-0.253	-0.144	0.182	0.061	0.630*	0.107	-0.239	0.109
	0.977*	0.053	1.222*	-0.168	0.828*	0.164	3.468*	0.393*	0.622	-0.175
ϱ_0	-0.060	0.719*	0.017	0.253	-0.218	0.141	0.080	-0.099	-0.336*	0.090
	-0.093	0.382*	-0.016	0.162	-0.067	0.174	-0.010	0.069	-0.245	0.267
ϱ_1	0.635*	0.382*	0.529*	0.492*	0.220	0.061	0.070	0.838*	0.203	-0.162
	0.122	0.808*	0.287	1.399*	0.328*	0.829*	0.055	0.472*	0.531*	0.204
ϱ	0.019	0.606*	0.105	0.344*	0.089	0.047	0.268	0.187	0.102	0.297
	0.102	0.358*	0.312*	0.305*	-0.018	0.252	0.114	0.205	0.253	0.273
ϱ^*	0.112	-0.143	0.156	-0.071	0.018	-0.129	0.581*	-0.132	0.062	-0.217
	0.261	0.225	0.442*	0.038	0.240	0.143	0.451*	0.198	0.523*	-0.031
a_0	0.638*	0.331*	0.514*	0.270	0.323*	0.184	0.433*	-0.099	0.388*	0.083
	0.524*	0.225	0.632*	0.260	0.147	0.062	0.289	0.159	0.444*	0.274
b_0	2.119*	0.701*	2.714*	0.299	0.936*	0.180	0.463*	0.424*	2.178*	0.357*
	3.290*	0.717*	2.501*	0.397	1.052*	0.282	0.627*	0.065	0.924*	0.255*
a_1	0.013	-0.094	0.050	0.065	0.361*	-0.149	0.089	-0.170	0.171	0.175
b_1	0.490*	0.310*	0.277	0.020	0.149	0.225	0.677*	0.464*	0.124	0.080

* indicates rejection of normality with 0.05 significance level, two sided test.

Table C.2: Kurtosis of the maximum likelihood estimates. The first line for each parameter gives results from fully parametric estimation approach, the second gives results from partly semi-parametric estimation approach.

parameter	$A \sim [40, 100]$		$A \sim [30, 110]$		$A \sim [0, 120]$		$A \sim [0, 80]$		$A \sim [0, 250]$	
	1000	10,000	1000	10,000	1000	10,000	1000	10,000	1000	10,000
σ_0	4.746*	3.367	7.416*	2.571	3.648	3.318	4.076*	3.445	5.808*	3.214
	7.457*	2.949	6.139*	2.759	3.519	2.600	2.839	3.114	3.155	3.684*
σ	2.928	3.163	3.374	2.920	2.923	2.999	2.927	2.800	3.808*	2.881
	3.642	2.778	5.519*	3.129	3.584	2.970	19.401*	2.741	3.669	2.695
ϱ_0	2.097*	3.569	2.105*	3.138	2.820	2.716	2.560	3.023	2.461*	3.481
	2.045*	3.765*	2.540	2.649	2.360*	2.787	2.050*	2.670	2.406*	3.197
ϱ_1	3.534	2.855	3.572	3.118	2.939	2.861	1.692*	3.220	3.210	2.519
	1.667*	3.268	1.977*	5.850*	2.265*	4.165*	1.644*	1.791*	3.268	2.737
ϱ	2.188*	2.913	2.349*	2.302*	2.210*	2.484*	2.124*	2.181*	2.156*	3.058
	2.278*	2.266*	2.332*	1.991*	2.228*	2.148*	2.010*	1.906*	2.026*	2.770
ϱ^*	2.330*	2.399*	2.427*	3.276	2.026	2.934	2.827	2.887	1.734*	2.939
	2.428*	3.096	2.749	2.974	2.178*	3.118	2.422*	2.942	2.275*	2.983
a_0	2.697	2.818	3.531	2.897	2.917	2.784	3.209	3.020	2.921	2.827
	2.365*	2.610	3.169	2.989	2.843	3.455	2.781	2.841	2.720	3.345
b_0	9.146*	3.799*	17.050*	2.925	4.915*	2.937	3.336	3.417	13.437*	3.240
	17.309*	3.723*	16.890*	3.193	4.654*	3.286	3.175	3.064	3.735*	3.521
a_1	2.845	3.030	2.796	3.046	3.023	2.893	2.501*	3.035	2.399*	3.243
b_1	3.367	3.284	3.317	2.981	2.930	3.025	3.213	3.009	3.113	2.946

* indicates rejection of normality with 0.05 significance level, two sided test.

C.4 Asymptotic behavior of the variances of the MLEs in the four-dimensional model

Table C.3 shows the empirical variances of $\sqrt{n}(\hat{\theta}_n - \theta)$ for the maximum likelihood estimates of the different parameters of the four-dimensional correlated gamma frailty model. This is done for the different sample sizes $n = 1000$, 5000, 10,000, different distributions of the age at the time of observation, and the two different estimation approaches. In each cell of the table the upper number gives value derived in the fully parametric estimation approach and the lower number gives the value derived in the partly semi-parametric approach. There are only results of the fully parametric estimation approach for the parameters a_1 and b_1.
The figures given in table C.3 are discussed in section 5.3.3 on page 83.

Table C.3: Asymptotic behavior of the empirical variances of the MLEs. The first line for each parameter gives results from fully parametric estimation approach, the second gives results from partly semi-parametric estimation approach.

variance of	$A \sim [40,100]$			$A \sim [30,110]$			$A \sim [0,120]$			$A \sim [0,80]$			$A \sim [0,250]$		
	1000	5000	10,000	1000	5000	10,000	1000	5000	10,000	1000	5000	10,000	1000	5000	10,000
$\sqrt{n}(\hat{\sigma}_{0n} - \sigma_0)$	104.7	190.0	150.9	64.4	59.9	70.2	36.2	41.7	36.4	29.7	37.2	40.3	72.5	75.3	80.6
	146.4	151.1	130.1	56.7	77.5	70.0	37.0	41.0	38.6	35.2	45.2	41.7	63.3	79.2	74.9
$\sqrt{n}(\hat{\sigma}_n - \sigma)$	21.4	24.1	23.0	14.1	14.1	15.5	10.6	13.6	13.3	110.0	180.0	199.0	3.7	3.1	3.0
	144.0	196.6	233.4	119.8	147.9	167.3	99.9	114.9	148.9	848.5	411.8	719.9	45.8	40.4	36.1
$\sqrt{n}(\hat{\varrho}_{0n} - \varrho_0)$	34.4	64.2	61.5	28.7	37.4	44.3	29.4	40.7	40.4	25.2	29.9	29.1	53.3	67.3	82.6
	37.1	55.2	51.2	27.1	40.3	42.9	30.7	38.3	40.4	30.7	34.4	27.0	46.1	71.0	75.8
$\sqrt{n}(\hat{\varrho}_{1n} - \varrho_1)$	18.8	21.4	19.9	15.5	14.6	14.4	12.1	14.9	15.5	67.0	183.2	226.7	5.3	5.4	4.9
	61.2	147.7	182.1	49.4	91.8	104.5	39.5	65.6	79.5	83.3	283.3	512.1	16.8	17.4	16.0
$\sqrt{n}(\hat{\varrho}_n - \varrho)$	51.1	132.8	163.9	55.3	118.7	174.2	55.3	146.6	211.1	63.2	131.2	189.5	61.6	205.2	185.5
	55.3	135.8	231.0	60.8	128.9	212.7	65.4	176.7	234.8	72.7	150.3	265.5	79.7	206.3	271.7
$\sqrt{n}(\hat{\varrho}_n^* - \varrho^*)$	32.7	41.3	44.1	37.1	46.1	52.0	39.0	64.9	75.8	42.8	62.6	49.9	51.0	127.2	138.0
	38.3	42.9	57.4	39.4	54.5	52.0	50.8	71.8	70.7	53.9	55.2	72.8	60.0	127.6	131.6
$\sqrt{n}(\hat{a}_{0n} - a_0)$	29.3	54.7	52.4	11.8	12.6	16.7	3.9	4.2	4.3	2.7	2.9	3.2	6.6	8.5	8.6
$\times$ 10,000	32.9	51.0	47.9	11.5	16.6	16.8	4.2	4.6	4.2	3.0	3.2	3.0	7.7	8.1	8.5
$\sqrt{n}(\hat{b}_{0n} - b_0)$	2.36	3.70	2.05	0.98	0.56	0.70	0.25	0.27	0.23	0.18	0.19	0.21	0.62	0.49	0.51
	4.64	2.43	1.89	0.83	0.77	0.74	0.29	0.25	0.23	0.20	0.23	0.20	0.55	0.52	0.45
$\sqrt{n}(\hat{a}_{1n} - a_1)$ $\times 10^6$	5.6	6.3	6.2	5.5	5.4	5.5	4.9	6.6	7.4	8.5	11.8	12.3	5.1	4.5	4.4
$\sqrt{n}(\hat{b}_{1n} - b_1)$ $\times 1000$	15.3	17.1	17.0	13.2	13.0	13.8	10.3	14.3	16.2	39.1	50.0	55.7	8.4	7.5	7.0

Bibliography

A. Ahlbom, P. Lichtenstein, H. Malmström, M. Feychting, K. Hemminki, and N. L. Pedersen. Cancer in twins: Genetic and nongenetic familial risk factors. *Journal of the National Cancer Institute*, 89:287–293, 1997.

M. Aigner. *Diskrete Mathematik*. Vieweg, Braunschweig/Wiesbaden, 1996.

H. Akaike. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6):716–723, 1974.

T. Amemiya. *Advanced Econometrics*. Harvard University Press, Cambridge, Massachusetts, 1985.

P. K. Andersen, Ø. Borgan, R. D. Gill, and N. Keiding. *Statistical Models Based on Counting Processes*. Springer, New York, 1993.

P. Armitage and T. Colton, editors. *Encyclopedia of Biostatistics*. Wiley, Chichester, 1998.

H. Bauer. *Wahrscheinlichkeitstheorie und Grundzüge der Maßtheorie*. Walter de Gruyter, Berlin, 1974.

A. Z. Begun, I. A. Iachine, and A. I. Yashin. Genetic nature of individual frailty: comparison of two approaches. *Twin Research*, 3:51–57, 2000.

J. Bellach, P. Franken, E. Warmuth, and W. Warmuth. *Maß, Integral und bedingter Erwartungswert*. Akademie-Verlag, Berlin, 1978.

R. Cederlöf, L. Friberg, L. Jonsson, and L. Kaij. Studies on similarity diagnosis in twins with the aid of mailed questionnaires. *Acta Geneticae Medicae et Gemellologiae*, 11:338–362, 1961.

R. Cederlöf and U. Lorich. The Swedish Twin Registry. In W. E. Nance, G. Allen, and P. Parisi, editors, *Twin Research: Biology and Epidemiology*, pages 189–195. Alan R. Liss, New York, 1978.

Y. S. Chow and H. Teicher. *Probability Theory*. Springer, New York, 1988.

K. Christensen, N. V. Holm, M. McGue, et al. A Danish population-based twin study on general health in the elderly. *Journal of Aging and Health*, 11(1):49–64, 1999.

K. Christensen, M. McGue, A. I. Yashin, et al. Genetic and environmental influences on physical abilities among Danish twins aged 75 years and older. Draft, 1997.

D. G. Clayton. A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika*, 65(1):141 – 151, 1978.

D. R. Cox and D. Oakes. *Analysis of Survival Data*. Chapman & Hall, London, 1984.

R. B. D'Agostino and M. A. Stephens. *Goodness-of-fit techniques*. Marcel Dekker, New York, 1986.

C. Elbers and G. Ridder. True and spurious duration dependence: The identifiability of the proportional hazard model. *Review of Economic Studies*, XLIX:403–409, 1982.

GAUSS Applications. *Constraint Maximum Likelihood - Application Module*. Aptech Systems, Inc., Maple Valley, WA, 1995.

GAUSS Applications. *Maximum Likelihood Estimation - Application Module*. Aptech Systems, Inc., Maple Valley, WA, 1996.

N. Giard. Konsistenz von Schätzungen im Gumbel-Morgenstern-Modell der abhängigen Zensur. Diplomarbeit, University of Rostock, 1997.

N. Giard, P. Lichtenstein, and A. I. Yashin. A multistate model for the genetic analysis of the ageing process. Submitted to Statistics in Medicine, 1999.

B. Gompertz. On the nature of the function expressive of the law of human mortality. In D. Smith and N. Keyfitz, editors, *Mathematical Demography - Selected Papers*, pages 279–282. Springer, Berlin, 1825.

H. Grönberg, L. Damber, and J.-E. Damber. Studies of genetic factors in prostate cancer in a twin population. *The Journal of Urology*, 152:1484–1489, 1994.

A. M. Herskind, M. McGue, N. V. Holm, et al. The heritability of human longevity: a population-based study of 2872 Danish twin pairs born 1870-1900. *Human Genetics*, 97:319–323, 1996a.

A. M. Herskind, M. McGue, I. A. Iachine, et al. Untangling genetic influences on smoking, body mass index and longevity: a multivariate study of 2464 Danish twins followed for 28 years. *Human Genetics*, 98:467–475, 1996b.

P. Hougaard. Frailty models for survival data. *Lifetime Data Analysis*, 1: 255–273, 1995.

I. A. Iachine. Parameter estimation in the bivariate correlated frailty model with observed covariates via the EM-Algorithm. Working Paper Series: Population Studies of Aging, No. 16, Center for Health and Social Policy, Odense University, 1995.

I. A. Iachine, N. V. Holm, J. R. Harris, et al. How heritable is individual susceptibility to death? The results of an analysis of survival data on Danish, Swedish and Finish twins. *Twin Research*, 1:196–205, 1998.

I. A. Iachine and A. I. Yashin. Identifiability of bivariate frailty models based on additive independent components. Research Report 8, Department of Statistics and Demography, Odense University, 1998.

K. L. Judd. *Numerical Methods in Economics.* MIT Press, Cambridge, Massachsetts, 1998.

J. D. Kalbfleisch and R. L. Prentice. *The Statistical Analysis of Failure Time Data.* John Wiley & Sons, New York, 1980.

E. L. Kaplan and P. Meier. Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, 53:457–481, 1958.

S. Kotz and N. L. Johnson, editors. *Encyclopedia of Statistical Sciences.* Wiley, New York, 1988.

J. F. Lawless. *Statistical Models and Methods for Lifetime Data.* John Wiley & Sons, New York, 1982.

P. Lichtenstein, N. V. Holm, P. K. Verkasalo, et al. Environmental and heritable factors in the causation of cancer. *The New England Journal of Medicine*, 343(2):78–85, 2000.

M. McGue. When assessing twin concordance, use the probandwise not the pairwise rate. *Schizophrenia Bulletin*, 18:171–176, 1992.

M. McGue. Odense workshop on the analysis of twin data. Private notes, August 1997.

M. McGue, J. W. Vaupel, N. Holm, et al. Longevity is moderately heritable in a sample of Danish twins born 1870-1880. *Journal of Gerontology*, 48 (6):B237–B244, 1993.

R. G. Miller. *Survival Analysis*. John Wiley & Sons, New York, 1981.

S. A. Murphy. Consistency in a proportional hazards model incorporating a random effect. *The Annals of Statistics*, 22(2):712–731, 1994.

S. A. Murphy. Asymptotic theory for the frailty model. *The Annals of Statistics*, 23(1):182–198, 1995.

M. C. Neale. *Mx: Statistical Modeling*. Box 710 MCV, Richmond, VA 23298: Department of Psychiatry, 3rd edition, 1995.

M. C. Neale and L.R. Cardon. *Methodology for Genetic Studies of Twins and Families*. Kluwer, Dordrecht, 1992.

G. G. Nielsen, R. D. Gill, P. K. Andersen, and T. I. A. Sørensen. A counting process approach to maximum likelihood estimation in frailty models. *Scandinavian Journal of Statistics*, 19:25–43, 1992.

D. Oakes. Bivariate survival models induced by frailties. *Journal of the American Statistical Association*, 84(406):487–493, 1989.

E. Parner. Asymptotic normality in the correlated gamma-frailty model. Research Report 346, Department of Theoretical Statistics, University of Aarhus, 1996a.

E. Parner. Consistency in the correlated gamma-frailty model. Research Report 345, Department of Theoretical Statistics, University of Aarhus, 1996b.

J. H. Petersen. *Analyzing Genetic and Environmental Influences on Mortality*. Ph.D. thesis, University of Copenhagen, 1996.

J. H. Petersen. An additive frailty model for correlated life times. *Biometrics*, 54(2):646–661, 1998.

A. Pickles and R. Crouchley. Generalizations and applications of frailty models for survival and event data. *Statistical Methods in Medical Research*, 3: 263–278, 1994.

A. Pickles and R. Crouchley. A comparison of frailty models for multivariate survival data. *Statistics in Medicine*, 14:1447–1461, 1995.

A. Pickles, R. Crouchley, E. Simonoff, et al. Survival models for developmental genetic data: age of onset of puberty and antisocial behavior in twins. *Genetic Epidemiology*, 11:155–170, 1994.

B. L. S. Prakasa Rao. *Identifiability in Stochastic Models: Characterization of Probability Distributions*. Academic Press, New York, 1992.

B. Rosner. *Fundamentals of Biostatistics*. Duxbury Press, Belmont, 4 edition, 1995.

R. Schoenberg. CML. *http://faculty.washington.edu/rons/*, 1996.

P. K. Sen and J. M. Singer. *Large sample methods in statistics*. Chapman & Hall, New York, 1993.

C. Smith. Concordance in twins: methods and interpretations. *American Journal of Human Genetics*, 26:454–466, 1974.

D. C. Thomas, S. Richardson, J. Gauderman, and J. Pitkaniemi. A Bayesian approach to multipoint mapping in nuclear families. *Genetic Epidemiology*, 14(6):903–908, 1997.

B. W. Turnbull. The empirical distribution function with arbitrarily grouped, censored and truncated data. *J. R. Stat. Soc., Ser. B*, 38(3):290–295, 1976.

J. W. Vaupel. Inherited frailty and longevity. *Demography*, 25(2):277–287, 1988.

J. W. Vaupel, K. G. Manton, and E. Stallard. The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography*, 16(3): 439–454, 1979.

H. Witting. *Mathematische Statistik*, volume 1. Teubner, Stuttgart, 1985.

H. Witting and U. Müller-Funk. *Mathematische Statistik*, volume 2. Teubner, Stuttgart, 1995.

H. Witting and G. Nölle. *Angewandte Mathematische Statistik: Optimale finite und asymptotische Verfahren*. Teubner, Leipzig, 1970.

A. I. Yashin, A. Z. Begun, and I. A. Iachine. Genetic factors in susceptibility to death: a comparitive analysis of bivariate survival models. *Journal of Epidemiology and Biostatistics*, 4(1):53–60, 1999.

A. I. Yashin and I. Iachine. How long can humans live? Lower bound for biological limit of human longevity calculated from Danish twin data using correlated frailty model. *Mechanisms of Ageing and Development*, 80:147–169, 1995a.

A. I. Yashin and I. A. Iachine. Genetic analysis of durations: Correlated frailty model applied to survival of Danish twins. *Genetic Epidemiology*, 12:529–538, 1995b.

A. I. Yashin and I. A. Iachine. Survival of related individuals: an extension of some fundamental results of heterogeneity analysis. *Mathematical Population Studies*, 5(4):321–339, 1995c.

A. I. Yashin and I. A. Iachine. How frailty models can be used for evaluating longevity limits: taking advantage of an interdisciplinary approach. *Demography*, 34(1):31–48, 1997.

A. I. Yashin, K. G. Manton, and I. A. Iachine. Genetic and environmental factors in duration studies: multivariate frailty models and estimation strategies. *Journal of Epidemiology and Biostatistics*, 1(2):115–120, 1996.

A. I. Yashin, J. W. Vaupel, and I. A. Iachine. Correlated individual frailty: An advantageous approach to survival analysis of bivariate data. Working Paper Series: Population Studies of Aging, No. 7, Center for Health and Social Policy, Odense University, 1993.

A. I. Yashin, J. W. Vaupel, and I. A. Iachine. A duality in aging: the equivalence of mortality models based on radically different concepts. *Mechanisms of Ageing and Development*, 74:1–14, 1994.

A. I. Yashin, J. W. Vaupel, and I. A. Iachine. Correlated individual frailty: an advantageous approach to survival analysis of bivariate data. *Mathematical Population Studies*, 5(2):145–159, 1995.

Summary

A model that can help in assessing the overall contributions of the genes and the environment to the ageing of an individual is presented in this dissertation. Here it is formulated for twin pairs, but it can easily be generalized for other groups of related individuals. The model can be placed in the framework of multivariate survival models and random effects models. It combines the ideas of survival analysis and demography with those of genetic epidemiology.

The ageing process is simplified as a 3-state process consisting of the states "healthy", "ill" or "disabled", and "deceased". In the approach presented here it is not the genetic parameters of the age at onset of disease or disability, and life span that are estimated but those of unobserved latent variables called frailty. The frailty variables represent the individual susceptibility to disease or death.

The main advantage of the model is the type of data one can use. As one observation unit it is possible to include groups of related individuals with some or all members deceased at the time of observation. Such groups would be excluded in an analysis that only examines disease status. The current health status of the living individuals and the exact life span of individuals who are already deceased are the only information necessary for the application of the model.

Theoretical as well as practical aspects of the proposed model are discussed in the dissertation. These are, on the one hand, theoretical properties such as the identifiability, the consistency, or the asymptotic normality of parameter estimates. On the other hand practical questions that arise in the application of the model are also considered. On the base of simulated data, properties of maximum likelihood estimates and the advantages of the model in comparison to a more traditional approach are discussed. Finally, the model is applied to a real data set.

The dissertation leads to new insights of the modeling and analysis of dependent and not fully observable survival times.

Zusammenfassung

In dieser Arbeit wird ein Modell vorgestellt, das dabei helfen kann, den allgemeinen Einfluß der Gene und der Umwelt auf das Altern einer Person zu untersuchen. Das Modell wird speziell für Zwillingspaare formuliert, es kann aber auch für andere Gruppen miteinander verwandter Personen verallgemeinert werden. Die Arbeit hat ihren Platz in der multivariaten Lebensdaueranalyse und im Bereich der Modelle mit zufälligen Effekten. Methoden der Lebensdaueranalyse werden mit Methoden der Demografie und der genetischen Epidemiologie kombiniert.
Der Alterungsprozeß wird vereinfacht als ein Prozeß modelliert, der aus den drei Zuständen "gesund", "krank" und "verstorben" besteht. Mit dem vorgestellten Modellierungsansatz werden nicht die genetischen Parameter des Alters zum Erkrankungszeitpunkt bzw. der Lebensdauer direkt bestimmt, sondern die genetischen Parameter von gewissen, nicht beobachtbaren zufälligen Größen, die *Frailty* genannt werden. Die *Frailty*-Variablen repräsentieren die individuelle Anfälligkeit für Krankheit bzw. Tod.
Der größte Vorteil des vorgestellten Modelles ist die Art von Daten, für die es anwendbar ist. Es ist möglich Gruppen von Personen miteinzubeziehen, in denen ein oder mehrere Gruppenmitglieder zum Beobachtungszeitpunkt bereits verstorben sind. In einer Analyse, die sich ausschließlich mit dem Gesundheitszustand befaßt, würden solche Gruppen von Individuen ausgeschlossen werden. Die einzige Information, die für die Anwendbarkeit des Modelles erforderlich ist, ist der aktuelle Gesundheitszustand der lebenden Personen sowie die Lebensdauer der verstorbenen Personen.
In der Arbeit werden sowohl theoretische als auch praktische Aspekte des präsentierten Modelles diskutiert. Das sind zum einen theoretische Eigenschaften wie die Identifizierbarkeit, die Konsistenz oder die asymptotische Normalität von Parameterschätzern. Zum anderen sind es praktische Fragen, die die Anwendbarkeit des Modelles betreffen. Auf der Basis von simulierten Daten werden Eigenschaften von Maximum-Likelihood-Schätzern und die Vorteile des Modelles gegenüber einer herkömmlichen Analysemethode betrachtet. Das Modell wird schließlich auch auf reale Daten angewendet.
Diese Arbeit führt zu neuen Erkenntnissen über die Modellierung und Analyse von abhängigen, nicht vollständig beobachtbaren Überlebenszeiten.

Curriculum vitae

Name:	Nicole Giard
Office Address:	Max Planck Institute for Demographic Research
	Doberaner Str. 114
	18057 Rostock
Date of birth:	11 April 1973
Place of birth:	Potsdam, Germany

Professional experience:

since 5/97	Ph.D. student at the Max Planck Institute for Demographic Research, Rostock Member of the Laboratory for the Development of Advanced Statistical Methods

University education:

4/97	diploma
10/91 to 4/97	University of Rostock, Faculty of Mathematics and Natural Sciences major: Mathematics, minor: Biology specialization: Mathematical Statistics

School education:

6/91	*Abitur*
9/87 to 6/91	Extended secondary school (*EOS*) specialized in Mathematics and Natural Sciences, Kleinmachnow
9/81 to 8/87	Secondary school (*POS*) with specialized instruction in Russian, Potsdam
9/79 to 8/81	Primary school, Potsdam and Caputh

Distinctions:

5/92 to 4/97	holder of a Friedrich Ebert Foundation Scholarship

Lebenslauf

Name:	Nicole Giard
Adresse:	Max-Planck-Institut für demografische Forschung Doberaner Str. 114 18057 Rostock
Geburtsdatum:	11. April 1973
Geburtsort:	Potsdam

Berufliche Tätigkeit:

seit 5/97	Doktorandin am Max-Planck-Institut für demografische Forschung Arbeitsgruppe zur Weiterentwicklung mathematisch-statistischer Methoden

Studium:

4/97	Diplom
5/92 bis 4/97	Stipendiatin der Friedrich-Ebert-Stiftung
10/91 bis 4/97	Universität Rostock, Mathematisch-Naturwissenschaftliche Fakultät Hauptfach: Mathematik, Nebenfach: Biologie Spezialisierungsrichtung: Mathematische Statistik

Schulbildung:

6/91	Abitur
9/87 bis 6/91	Erweiterte Spezialoberschule mathematisch-naturwissenschaftlich-technischer Richtung Kleinmachnow
9/81 bis 8/87	Polytechnische Oberschule mit erweitertem Russischunterricht Potsdam
9/79 bis 8/81	Polytechnische Oberschule in Potsdam, seit 10/80 in Caputh

Erklärung

Hiermit erkläre ich, daß ich die vorliegende Arbeit selbständig angefertigt und keine außer den angegebenen Hilfsmitteln dazu verwendet habe.

Rostock, den 2. Februar 2001