University of Nebraska Medical Center
DigitalCommons@UNMC

Theses & Dissertations

Graduate Studies

Fall 12-18-2020

Statistical Modeling of Survival Data Using Frailty Models

Adams Kusi Appiah
University of Nebraska Medical Center

Tell us how you used this information in this short survey.
Follow this and additional works at: https://digitalcommons.unmc.edu/etd

Recommended Citation
Kusi Appiah, Adams, "Statistical Modeling of Survival Data Using Frailty Models" (2020). *Theses & Dissertations*. 502.
https://digitalcommons.unmc.edu/etd/502

This Dissertation is brought to you for free and open access by the Graduate Studies at DigitalCommons@UNMC. It has been accepted for inclusion in Theses & Dissertations by an authorized administrator of DigitalCommons@UNMC. For more information, please contact digitalcommons@unmc.edu.

STATISTICAL MODELING OF SURVIVAL DATA USING FRAILTY MODELS

by

Adams Kusi Appiah

A DISSERTATION

Presented to the Faculty of

the University of Nebraska Graduate College

in Partial Fulfillment of the Requirements

for the Degree of Doctor of Philosophy

Biostatistics

Graduate Program

Under the Supervision of Professor Hongying (Daisy) Dai

University of Nebraska Medical Center

Omaha, Nebraska

November, 2020

Supervisory Committee:

Gleb Haynatzki, Ph.D. Xiaoyue Zoe Cheng, Ph.D. Yeongjin Gwon, Ph.D.

ACKNOWLEDGEMENTS

I am grateful to my supervisor Professor Hongying Dai for her guidance throughout writing the dissertation, whose expertise was invaluable during the entire process. I would like to thank Professor Gleb Haynatzki for his encouragement and insights on many issues. I would also my committee members, Professor Yeongjin Gwon and Professor Xiaoyue Zoe Cheng, for their valuable suggestions and inputs for this dissertation. The department of biostatistics faculty and staff have all added positively to the comfortable study environment during my Ph. D. program are very much appreciated. Finally, I would like to thank my wife, Ramatu Adam, for her unconditional love and support. I also thank my dear friends and follow graduate students for all the help, care and fun I have received, especially Morshed Alam, Nirosha Rathnayake, Karl Bisselou Moukagna, Nicholas Hein, Moses New-Aaron, and Ekow Lewis.

STATISTICAL MODELING OF SURVIVAL DATA USING FRAILTY MODELS

Adams Kusi Appiah, Ph.D.

University of Nebraska, 2020

Supervisor: Hongying (Daisy) Dai, Ph.D.

In many clinical trials, time-to-event endpoints are often adopted to demonstrate a clinically convincing effect of treatments appropriately. These variables might be clustered or correlated because of certain common features, such as genetic traits or shared environmental factors or repeated events. Observations from the same cluster are assumed to be correlated because they usually share specific unobserved characteristics. Ignoring the correlations between the survival times may lead to incorrect estimates of parameters of interest and invalid statistical inferences. The scientific interest may lie in the estimation of treatment effect while accounting for the correlated event times. This dissertation proposes a shared frailty model to fit correlated or clustered survival data and investigates the effect on the corresponding estimated regression coefficients. In this work, we propose new methods using hierarchical likelihood (h-likelihood) to fit a wide range of frailty models, in which the latent frailties are treated as “parameters” and estimated jointly with other parameters of interest. The adjusted profile likelihood is adopted to estimate the frailty parameter. In this dissertation, we (1) propose effective bias correction methods for the h-likelihood estimators under the shared gamma frailty models; (2) extend the h-likelihood to log-logistic frailty model, a non-exponential family distribution, and describe the total derivative approach to estimate the model parameters; (3) propose a flexible log-skew normal distribution as the frailty distribution to model the dependency in multivariate survival data. The performance of the proposed models is examined via Monte Carlo simulations. We illustrate our methods using kidney infection and cow mastitis data.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

AFT	Accelerated failure time
cAIC	Conditional Akaike information criterion
EM	Expectation and maximization
HGLM	Poisson hierarchical generalized linear models
HL	Hierarchical likelihood
h-likelihood	h-likelihood
ILS	Iterative least squares
IWLS	Iterative weighted least squares
MCMC	Markov Chain Monte Carlo
MHLE	Maximum h-likelihood estimator
PH	Proportional hazards
PPL	Penalized partial likelihood
PVF	power variance function

LIST OF NOTATIONS

Symbol	Description
det	Determinant
exp	Exponential
lim	Limit
log	Logarithm of base e
var	Variance
$f(.)$	probability density function
$h(.)$	Hierarchical log-likelihood
T	Transpose
$\boldsymbol{\beta}$	Vector of fixed effects
$(\boldsymbol{u}, \boldsymbol{v})$	Vector of frailty variates (random effects)
$\boldsymbol{X}$	Model matrix of fixed effects
$\boldsymbol{Z}$	Model matrix of random effects
$\boldsymbol{\delta}$	Censoring indicator
θ	A generic parameter indicating any frailty parameter to be estimated
$\lambda(.)$	Hazard function
$\lambda_0(.)$	Baseline hazard function
$\Lambda_0(.)$	Cumulative baseline hazard function
$\boldsymbol{\eta}$	Linear predictor in a hazard function
$I(.)$	Indicator function
$\boldsymbol{S}(.)$	Score statistic
∂	Partial derivative
$\boldsymbol{H}(\boldsymbol{\beta}, \boldsymbol{v})$	Hessian matrix of $(\boldsymbol{\beta}, \boldsymbol{v})$ based on the data $(\boldsymbol{v})$

LIST OF TABLES

LIST OF FIGURES

Chapter 1

General introduction

1.1. Introduction

Time-to-event data or survival data are common in applied research such as medicine, engineering, economics, etc. For example, the time from diagnosis or start of treatment to death of cancer patients, time to infection after exposure to disease, lifetimes of equipment, duration of unemployment, etc. A distinguishing feature of time-to-event data is the possible presence of censoring. Censoring occurs when information about a subject's survival time is incomplete. In this dissertation, we focus on right-censored data. The right-censoring happens when a subject is lost to follow-up before an event occurs or the event does not occur within the study period. The class of statistical techniques developed to deal with time-to-event data is known as survival analysis.

1.2 Basic survival functions

Suppose that we have a random sample of n subjects, $i = 1,2, \dots, n$. Let Y_i be a non-negative random variable representing the survival time from a homogeneous population and C_i be the censoring time. Throughout this dissertation, we assume non-informative censoring (i.e., the censoring time distribution is unrelated to the parameter of interest from the failure time distribution) and independent censoring mechanisms. Let $T_i = \min(Y_i, C_i)$ and δ_i be the censoring indicator equal to 1 if $T_i = Y_\mathrm{i}$ and 0 if $T_i = C_i$. The survival function and the hazard function are two probability distributions of T_i that are particularly essential in survival applications. The survival function, denoted by $S(t)$, is defined as the probability that the T_i exceeds the specified time t. That is

$$S(t) = \Pr(T_i \geq t), \quad 0 < t < \infty, \tag{1.1}$$

where $S(t)$ is a monotone non-increasing left continuous function with $S(0) = 1$ and

$\lim_{t \to \infty} S(t) = 0$. The hazard function, $\lambda(t)$, gives the instantaneous rate of failure at time t on condition that individual surviving up to t, and is given by,

$$\lambda(t) = \lim_{\Delta t \to 0} \frac{\Pr(t \leq T_i < t + \Delta t | T_i \geq t)}{\Delta t}. \tag{1.2}$$

The relationship between $S(t)$ and $\lambda(t)$ can be written as,

$$S(t) = \exp\left(-\int_0^t \lambda(u)\, du\right). \tag{1.3}$$

1.3 Proportional hazard (PH) models

The proportional hazards (PH) model is the most widely used survival regression model to investigate the presence of a vector of explanatory variables that may affect time-to-event through the hazard function. It assumes that the covariates have a multiplicative effect on the hazard and that, this effect is constant over time.

The proportional hazard (PH) model can be written as,

$$\lambda_i(t) = \lambda_0(t) \exp(\boldsymbol{x}_i^T \boldsymbol{\beta}), \tag{1.4}$$

where $\lambda_i(t)$ is the hazard for subject i at time t, hazard function; $\exp(\boldsymbol{x}_i^T \boldsymbol{\beta})$ is the relative risk of subject i, where $\boldsymbol{x}_i = (x_{i1}, x_{i2}, x_{i3}, \dots, x_{ip})^T$ is a $p \times 1$ vector of explanatory variables, $\boldsymbol{\beta}$ is the associated vector of fixed unknown regression parameter, and T is a transpose. For two individuals with covariate vectors $\boldsymbol{x}$ and $\boldsymbol{x}^*$, the ratio of their hazard rates is

$$\frac{\lambda_0(t) \exp(\boldsymbol{x}_i^T \boldsymbol{\beta})}{\lambda_0(t) \exp(\boldsymbol{x}_i^{*T} \boldsymbol{\beta})} = \exp[(\boldsymbol{x}_i^T - \boldsymbol{x}_i^{*T})\boldsymbol{\beta}], \tag{1.5}$$

which is a constant. That is, the conditional hazard functions have a fixed ratio over time.

The Cox PH modelCox (1992) is the most popularly used statistical method for analyzing time-to-event data. The Cox PH model assumes a semiparametric form for the hazard $\lambda_i(t)$ in equation (1.4). That is, $\lambda_0(t)$ is an unspecified common baseline hazard function. The $\boldsymbol{\beta}$ parameters may be estimated by maximizing the partial likelihood.(Cox 1975) Let $T_{(1)}, T_{(2)}, \dots, T_{(K)}$ be the distinct failure times, then partial likelihood function is defined as

$$L(\boldsymbol{\beta}) = \prod_{k=1}^{K} \left(\frac{\exp\left(\boldsymbol{x}_k^T \boldsymbol{\beta}\right)}{\sum_{l \in \mathcal{R}_{(k)}} \exp(\boldsymbol{x}_l^T \boldsymbol{\beta})} \right)^{\delta_k}, \tag{1.6}$$

where $\mathcal{R}_{(k)}$ is the risk set of subjects at the time $T_{(k)}$. That is the set of individuals who have not failed or been censored by that time.

Certain non-negative probability distributions can be used to describe the functional form of the baseline hazard, λ_0, in the equation (1.4) leading to parametric survival models that are frequenlty used to analyze time-event-data. The exponential and Weibull models, Gompertz, log-logistic, for example, are widely used. The Weibull regression model, which has been successfully employed in many fields, including reliability and medical studies, can be considered as an attractive alternative to the Cox PH model in analyzing survival data. It is the most extensively used parametric model in time-to-event data analysis in both physical and social sciences. The reason being that, the Weibull regression model can be expressed as both accelerated failure time (AFT) regression model and PH regression model, so both hazard ratios and time ratios can be estimated. The classical maximum likelihood approach can be used to obtain the estimates of parameters in parametric survival models. The maximum likelihood function,

$$L(.) = \prod_{i=1}^{n} \left\{\lambda_0(t) \exp\left(\boldsymbol{x}_i^T \boldsymbol{\beta}\right)\right\}^{\delta_i} \exp\left(-\Lambda_0(t) \exp\left(\boldsymbol{x}_i^T \boldsymbol{\beta}\right)\right), \tag{1.7}$$

where $\Lambda_0(\mathrm{t}) = \int_0^t \lambda_0(u)du$ is the cumulative baseline hazard function. For example, for the Weibull parametric survival models, $\Lambda_0(\mathrm{t}) = \rho t^{\gamma}$. The estimation of the parameters are obtained by solving the score equations

$$S_j(\boldsymbol{\xi}) = \frac{\partial \log L(\boldsymbol{\xi})}{\partial \xi_j} = 0, j = 1,2,3,$$

where $\boldsymbol{\xi} = (\rho, \gamma, \boldsymbol{\beta}^T)^T$.

However, analytical solutions to the score equations are intractable and require numerical methods such as the Newton-Raphson algorithm.

1.4 Multivariate survival data

The traditional applications and development of the classical PH survival analysis techniques assume that survival times of different subjects are independent. Multivariate survival data frequently occur in medical research, especially in clinical trials and cohort studies. Examples include clustered survival data, recurrent events, competing risk events, etc. These survival times may be correlated due to some natural, artificial clustering or shared environmental factors of subjects, or repeated events that may influence the same cluster or subject's failure times. Thus, the classical Cox PH or the parametric (e.g., Weibull PH) models may not be appropriate since the assumption of independence may not be valid.

Several statistical methods have been proposed for the analysis of multivariate survival data. One of the approach is to use the marginal models where the covariates effects are specified unconditionally. Those methods were discussed in.(Wei, Lin et al. 1989, Prentice and Cai 1992, Pipper and Martinussen 2003) However, these marginal models have been used in different settings. One can also adopt a general framework such as the counting process(Aalen 1978, Aalen and Hoem 1978, Prentice and Cai 1992) for analyzing multivariate survival data. The fixed effects models, where the clusters are introduced as fixed effects, have also been used in the literature.(Petersen 1998, Yin and Cai 2004, Yin 2007)

To model multivariate survival data, a natural approach is to incorporate random effects called frailties into the proportional hazard model to account for the dependency on survival outcomes.(Vaupel, Manton et al. 1979, Aalen 1988, Aalen 1994) The frailty model introduced by(Clayton 1978) has become increasingly popular in analyzing multivariate survival data. In this dissertation, we mainly focus on the shared frailty model, where all subjects within the same group share common frailty, and the frailties of different groups are independent.

1.5 Shared frailty models

The shared frailty model for right-censored failure time can be described as follows. Let Y_{ij} be the failure-time variable corresponding to individual or repeated event j $(j = 1,2, \dots, n_i)$ from cluster or subject i $(i = 1,2, \dots, G)$. Thus the total sample size is $N = \sum_{i=1}^{G} n_i$. Let C_{ij} be the non-informative right-censoring time and independent of Y_{ij}, $T_{ij} = \min(y_{ij}, C_{ij})$, δ_{ij} is the censoring indicator with δ_{ij} equal to 1 if $T_{ij} = y_{ij}$ and 0 if $T_{ij} = C_{ij}$, and $\boldsymbol{x}_{ij} = (x_{ij}, \dots, x_{ijp})^T$ is a $p \times 1$ covariate vector associated with the fixed-effect parameters $\boldsymbol{\beta}$. The conditional hazard function of T_{ij} given the unobserved frailty random variable u_i and $\boldsymbol{x}_{ij}$, can be written as

$$\lambda_{ij}(t_{ij}|u_i, \boldsymbol{x}_{ij}) = \lambda_0(t) \exp(\boldsymbol{x}_{ij}^T \boldsymbol{\beta})\, u_i. \tag{1.8}$$

The $u_i, i = 1, \dots, G$ are independent, identically distributed random variables with some common density function $f(u_i; \theta)$, where θ is the parameter of the frailty distribution. In addition to the independent and non-informative censoring mechanism assumption, these two assumptions are extended to the frailty models. That is, given $U_i = u_i$, the pairs $\{(T_{ij}, C_{ij})\ j = 1, \dots, n_i\}$ are conditionally independent and both T_{ij} and C_{ij} are also conditionally independent for $j = 1, \dots, n_i$, and given $U_i = u_i$, $\{(T_{ij}, C_{ij})\ j = 1, \dots, n_i\}$ are conditionally non-informative of T_{ij}.

A Cox PH model with shared frailty is where baseline hazard function $\lambda_0(t)$ in equation (1.8) is unspecified, whereas a parametric distribution can be assumed to the baseline hazard leading to parametric frailty models. The natural parametric distribution is the Weibull because it allows for both the PH and AFT models.

A variety of probability distributions has been proposed for the frailty u_i. The frequently studied frailty distributions belong to the Hougaard's(1986) power variance function (PVF) family. The gamma, inverse Gaussian, positive stable, and compound Poisson distribution are all members of this family. These frailty densities lead to tractable integration due to the closed form of Laplace approximation of these distributions.(Hanagal 2009) However, except for gamma frailty distribution, these tractable integrations are of a much more complex form. A further important frailty distribution is the lognormal distribution. However, this distribution is not a member of the power variance function family, and thus, it does not have a simple expression for the Laplace transform. Numerical integration is often used to approximate the integral.

Techniques to fit frailty models for multivariate survival data have been proposed in the literature. The expectation-maximization (EM) algorithm(Dempster, Laird et al. 1977) can be used to obtain parameter estimates in the semiparametric frailty model. In this approach, the EM method considers full data likelihood, which is a function of the observed event times and the unobserved random variables. The expectation step is often approximated using Laplace approximation. The following authors discussed the EM algorithm-based estimation approach.(Klein 1992, Nielsen, Gill et al. 1992, Xue and Brookmeyer 1996, Sastry 1997, Cortiñas and Burzykowski 2005, Yu 2006, Chen, Tong et al. 2009) The Bayesian version of the EM algorithm has also been proposed, where the expectation step is approximated with Markov Chain Monte Carlo (MCMC) methods.(Vaida and Xu 2000, Ripatti, Larsen et al. 2002)

The penalized partial likelihood (PPL) approach is another commonly used approach to estimate parameters in frailty models. In this approach, the full data likelihood consists of two parts. The first part consists of the likelihood of the data given the frailties. The second part corresponds to the distribution of the frailties. In most cases, the distribution of the frailties is considered as the penalty term. However, one can consider penalizing the baseline hazard function. A detailed discussion of PPL approaches can be found in.(McGilchrist and Aisbett 1991, McGilchrist 1993, Ripatti and Palmgren 2000, Therneau, Grambsch et al. 2003, Duchateau and Janssen 2004, Wang 2006) Rondeau, Commenges et al. (2003) discussed inference methods based on the penalized full likelihood.

Other estimation methods in frailty models include the Bayesian approach(Sinha 1993, Sahu, Dey et al. 1997, Yin and Ibrahim 2005, Komárek and Lesaffre 2008), and the generalized estimating equations approach (Cai and Prentice 1995, Zhang 2006).

The hierarchical-likelihood (a.k.a. h-likelihood), proposed byLee and Nelder (1996), has been recently applied to estimate parameters in both semiparametric and parametric frailty models.(Ha, Lee et al. 2001, Ha and Lee 2003) For the lognormal and gamma frailty models,(Ha, Lee et al. 2001, Ha and Lee 2005, Ha, Sylvester et al. 2011, Jeon, Hsu et al. 2012, Ha, Jeong et al. 2018) and Weibull frailty model(Wang, Xu et al. 2011) developed estimation procedures using Cox's(Cox 1975, Cox 1992) partial likelihood. Ha and Lee (2003) proposed the hierarchical-likelihood approach to analyze both parametric and semiparametric frailty models. Hanagal (2010) used the h-likelihood to estimate the Weibull and lognormal frailty models with Weibull parametric baseline hazard. Ha and Lee (2005) applied the h-likelihood estimation method to multilevel frailty models. In this dissertation, we develop h-likelihood estimators for estimating parameters in shared frailty models.

1.6. Hierarchical likelihood estimators

The h-likelihood for frailty models initially developed byHa, Lee et al. (2001) to describe frailty models is defined as follow:

First, define the h-likelihood function. Let $i = 1,2, \dots, G$ be clusters or individuals where each cluster or each individual has $j = 1,2, \dots, n_i$ observations or repeated events. FollowingHa, Lee et al. (2001) the contribution, h_i say, of the j^{th} observation or repeated events in the i^{th} cluster or by i^{th} individual is given by the logarithm of the joint density of $(T_{ij}, \delta_{ij}, u_i)$

$$h_{ij} = h_{ij}(\boldsymbol{\beta}, \lambda_0, \theta; t_{ij}, \delta_{ij}, u_i) = \log[\mathcal{L}_{1ij}(.\,|u_i)\mathcal{L}_{2i}(u_i; \theta)], \tag{1.9}$$

where $\mathcal{L}_{1ij}$ is the conditional density of (T_{ij}, δ_{ij}) given $U_i = u_i$ with parameters $(\boldsymbol{\beta}, \lambda_0)$ and $\mathcal{L}_{2i}$ is the density of U_i with parameter θ. By the conditional independence of T_{ij}, C_{ij} given $U_i = u_i$, and the non-informative censoring assumption, we have,

$$\mathcal{L}_{1ij}(t_{ij}, \delta_{ij}, \lambda_0, \boldsymbol{\beta}|x_{ij}, u_i) = \{\lambda_0(t) \exp(\boldsymbol{x}_{ij}^T \boldsymbol{\beta})\, u_i\}^{\delta_{ij}} \exp\{-\Lambda_0(t_{ij}) \exp(\boldsymbol{x}_{ij}^T \boldsymbol{\beta})\, u_i\}. \tag{1.10}$$

The $\mathcal{L}_{1ij}$ in the above equation becomes the ordinary censored-data likelihood given $U_i = u_i$, while $\Lambda_0(.)$ is the conditional cumulative baseline hazard function of T_{ij} given $U_i = u_i$. Thus, the h-likelihood for the frailty is

$$h = \sum_{ij} \ell_{1ij} + \sum_{i} \ell_{2i}, \tag{1.11}$$

where $\ell_{1ij} = \ell_{1ij}(.\,|u_i) = \log(\lambda_0(t)) + \log u_i + \boldsymbol{x}_{ij}^T \boldsymbol{\beta} - \Lambda_0(t_{ij}) \exp(\boldsymbol{x}_{ij}^T \boldsymbol{\beta})\, u_i$ is the logarithm of the conditional density of function for t_{ij} and δ_{ij} given $U_i = u_i$ and ℓ_{2i} is the logarithm of the density of function $f(u_i; \theta)$.

A unique aspect of the h-likelihood approach is that it avoids multi-dimensional integration of the latent frailty variates when obtaining the parameter estimates, the frailty variates $\boldsymbol{u} = (u_1, \dots, u_G)^T$ are treated as parameters and jointly estimated along with the parameters of interests $\{\boldsymbol{\beta}, \theta, \Lambda_0(.)\}$. The h-likelihood approach eliminates the unspecified baseline hazard via profile likelihood,

$$h^* = \sum_{ij} \ell^*_{1ij} + \sum_{i} \ell_{2i}, \tag{1.12}$$

and the Newton Raphson method is adopted to jointly estimate $\boldsymbol{\tau} = (\boldsymbol{\beta}, \boldsymbol{u})$. The frailty parameters are obtained using the adjusted profile likelihood, which is given by,

$$h_A^* = h^*|_{\boldsymbol{\tau}=\hat{\boldsymbol{\tau}}} + \frac{1}{2}\log\{\det(2\pi \boldsymbol{H}^{-1})\}\Big|_{\boldsymbol{\tau}=\hat{\boldsymbol{\tau}}}. \tag{1.13}$$

where $\boldsymbol{H} = -\frac{\partial^2 h^*}{\partial \boldsymbol{\tau}^2}$ is the asymptotic covariance matrix of $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{u}} - \boldsymbol{u}$. The Newton-Raphson method can be used to solve for θ in $\frac{\partial h_A^*}{\partial \theta}$.

1.7 Dissertation organization

This dissertation investigates h-likelihood approaches for estimating the regression parameters of Cox and Weibull proportional hazard (PH) regression models applied to multivariate survival data. It is organized into five chapters. Chapter 1 presents a general introduction, including literature reviews of past work on the multivariate survival data, an introduction to PH regression models. The next three parts are papers in the form to be submitted to journals. The final section summarises the results of the previous chapters and discusses additional issues.

Chapter 2

Bias reduction methods in the hierarchical likelihood approach for shared gamma frailty model in clustered failure time data

Abstract

Shared frailty models with both parametric and non-parametric baseline hazard functions are widely used for the analyses of survival data. A hierarchical likelihood (h-likelihood) approach has been developed for estimating the regression parameters and frailty variates, in which the latent frailties are treated as “parameters” and estimated jointly with other parameters of interest. The h-likelihood estimators generally perform well in various frailty models. However, they are known to be biased for non-normal random effects. Existing modifications to the h-likelihood employ the total derivative and second-order Laplace approximation, which is computationally intensive with complicated mathematical derivations. In this work, we propose two effective bias correction methods for the h-likelihood estimators under the shared gamma frailty models. The first method modifies the adjusted profile likelihood by adding a logarithmic transformation of the variance of the frailty parameter to avoid zero estimates in the frailty parameter. The second approach modifies the score function of the adjusted profile likelihood. Thus, in the two modifications, we avoid the use of the total derivative and second-order Laplace approximation. Simulation studies show that the proposed approaches reduce the bias in the h-likelihood estimators, especially for the estimate of the frailty parameter (from 30% to 3% when the frailty variance and sample size are small). Applications of both methods are illustrated using recurrent kidney infection data. Furthermore, the proposed bias correction methods can be extended to a broad class of frailty distributions and complex models such as joint modeling and competing risks.

Keywords: bias correction; hierarchical likelihood; frailty model; clustered survival data; profile likelihood.

2.1. Introduction

Clustered or grouped survival data frequently occur in many medical research studies, especially in clinical trials and cohort studies. For example, a typical clinical trial study may consist of multiple participating centers or hospitals. Observations within clusters (e.g., centers, hospitals, etc.) may be correlated due to some natural, artificial clustering or shared environmental factors of subjects that may influence the failure times of the same cluster. The natural approach is to incorporate random effects or shared frailties to account for within-cluster homogeneity in outcomes(Wienke 2010, Enki, Noufaily et al. 2014). Frailties, which are random effects in survival models, have been widely used for the analysis of clustered failure time data.

The semiparametric frailty model, which assumes a non-parametric baseline hazard function, plays an important role in modeling clustered survival data. The gamma frailty model has received enormous considerations due to its mathematical convenience.(Clayton 1978, Keiding, Andersen et al. 1997) Other choices of frailty models have been developed.(Duchateau and Janssen 2007, Hougaard 2012) The parametric frailty models, where the baseline hazard function is assumed to follow a parametric distribution, have also been proposed.(Keiding, Andersen et al. 1997) When the assumption of the parametric distribution is valid, inferences lead to smaller standard errors for the hazard ratios and the survival time quantities compared to the non-parametric baseline hazard model due to the existence of sufficient statistics in the parametric case. There is no sufficient statistics in the non-parametric baseline hazard model. However, when the correct distribution of the baseline hazards is uncertain, the use of non-parametric models is desirable.

Challenges remain in fitting frailty models for multivariate survival data. The expectation-maximization (EM) algorithm(Dempster, Laird et al. 1977) has been the fundamental tool for obtaining parameter estimates in the semiparametric frailty model. In this framework, the EM method considers full data likelihood, which is a function of the observed event times and the unobserved random variables (treated as missing). The E-step involves the computation of the full likelihood with respect to the observed data. This expectation is often approximated using numerical integration since, in most cases, an analytic solution does not exist. The EM algorithm-based estimation procedure has been developed for the gamma frailty model.(Klein 1992) However, such an approach is shown to have a finite sample underestimation of the model parameters.(Barker and Henderson 2005) Inferences for the log-normal frailty has also been developed, where the conditional expectation of frailty given the observed data are usually computed using numerical integration.(Sastry 1997, Cortinas Abrahantes and Burzykowski 2005) The EM algorithm can also be conducted using the Bayesian framework, where the expectation step is approximated by MCMC methods.(Vaida and Xu 2000, Ripatti, Larsen et al. 2002) However, EM and other alternative approaches require complex numerical integration due to intractable integration, which can be computationally intensive, especially when the number of clusters is large. Therneau, Grambsch et al. (2003) proposed a penalized estimation method for frailty models by utilizing the partial likelihood function. This procedure leads to simple estimating equations but results in an underestimation of variances of the fixed effects parameters.(Ripatti and Palmgren 2000)

The hierarchical-likelihood (a.k.a. h-likelihood), proposed by Lee and Nelder (1996), has been extended to estimate parameters in frailty models.(Ha, Lee et al. 2001, Ha, Sylvester et al. 2011, Wang, Xu et al. 2011) A unique aspect of the h-likelihood approach is that it avoids multi-dimensional integration with respect to the latent frailty variates when obtaining the parameter estimates. Thus the frailty variates are jointly estimated along with the parameters of interest. This property is particularly appealing when the complex dependence structure among clustered failure times requires estimation of multiple frailties. The h-likelihood estimators generally perform well in various frailty models. However, they are substantially biased for non-normal random effects such as gamma frailty models.(Ha, Lee et al. 2001)

In this paper, we propose two effective bias reduction methods for the h-likelihood estimators in survival analysis. The remaining of the paper is organized as follows. Section 2.2 introduces the gamma frailty model as a proof of concept and the h-likelihood estimators for both non-parametric and parametric baseline hazards survival models. In Section 2.3, we propose two bias correction approaches for the gamma frailty model. The proposed bias correction methods are also applicable to a broad class of frailty distributions and complex models such as joint modeling and competing risks. Section 2.4 presents the results of a simulation study in which the performance of the proposed methods is evaluated in cases when the baseline hazard is left unspecified or parametrically specified. Both proposed bias correction methods are also applied to the recurrent kidney infection data, and the results are discussed in Section 2.5. The discussion, presented in Section 2.6, concludes the paper.

2.2. Gamma shared frailty model

Let T_{ij} be the failure-time variable corresponding to individual j $(j = 1, 2, \dots, n_i)$ from cluster i $(i = 1, 2, \dots, G)$, and C_{ij} be a non-informative right-censoring time that is independent of T_{ij}. Let $y_{ij} = \min(T_{ij}, C_{ij})$ be the observed failure times, and $\delta_{ij} = I(T_{ij} \leq C_{ij})$, where $I(.)$ is the indicator function. In particular, y_{ij} are observations on T_{ij}

when censoring is not present. Below, all vectors are column-vectors, whereas their transposed (denoted by the superscript ^{T}) are row-vectors. The shared frailty model for gamma frailty can be written as,

$$\lambda_{ij}(t_{ij}|u_i, \boldsymbol{x}_{ij}) = \lambda_0(t_{ij})u_i \exp(\boldsymbol{x}_{ij}^T\boldsymbol{\beta}), \tag{2.1}$$

where $\lambda_0(t_{ij})$ is the baseline hazard function, $\boldsymbol{x}_{ij} = (x_{ij1}, \dots, x_{ijp})^T$ is a vector of fixed covariates, and $\boldsymbol{\beta}$ is a $p \times 1$ vector of unknown regression parameters. Assume that the unobserved frailties $U_i's$, are independently and identically distributed gamma random variables, where mean is set to be 1 to avoid identifiability issues, and unknown variance is θ. The probability density function of U_i is given as follows,

$$f(u_i;\theta) = \frac{1}{\Gamma(1/\theta)\theta^{\frac{1}{\theta}}} u_i^{\left(\frac{1}{\theta}-1\right)} \exp\left(-\frac{u_i}{\theta}\right). \tag{2.2}$$

2.3 H-likelihood estimation

The h-likelihood for shared frailty(Ha, Lee et al. 2001) is given by

$$h = \sum_{ij} \ell_{1ij} + \sum_{i} \ell_{2i}, \tag{2.3}$$

where ℓ_{1ij} is the logarithm of the conditional likelihood in T_{ij} and δ_{ij} with parameters $(\boldsymbol{\beta}, \lambda_0)$ given $U_i = u_i$, and ℓ_{2i} is the log density function of $U_i = u_i$ with parameter θ, shown below.

Let $v_i = \log(u_i)$, and $\boldsymbol{v} = (v_1, v_2, \dots, v_G)^T$. Also, let $\boldsymbol{z}_{ij} = (z_{ij1}, z_{ij2}, \dots, z_{ijG})^T$ be a $G \times 1$ cluster indicator vector. By defining $\eta_{ij} = \boldsymbol{x}_{ij}^T\boldsymbol{\beta} + \boldsymbol{z}_{ij}^T\boldsymbol{v}$, we have

$$\ell_{1ij} = \delta_{ij}\{\log\lambda_0(y_{ij}) + \eta_{ij}\} - \Lambda_0(y_{ij})\exp(\eta_{ij}),$$

$$\ell_{2i} = \{v_i - \exp(v_i)\}\theta^{-1} - \log\Gamma\left(\frac{1}{\theta}\right) - \theta^{-1}\log\theta,$$

where $\Lambda_0(.)$ is the cumulative baseline hazard function.

A unique aspect of the h-likelihood approach is that, instead of integrating out the latent frailty variates $\boldsymbol{v}$ are treated as parameters and jointly estimated along with $\{\boldsymbol{\beta}, \theta, \Lambda_0(.)\}$.

2.3.1. Non-parametric baseline hazard models

In this section, we will describe the h-likelihood estimation process in the semiparametric gamma frailty model. The semiparametric model is desirable when the underlying functional form of the baseline hazard is unknown. Suppose that the baseline hazard $\lambda_0(t_{ij})$ in equation (2.1) is unspecified. Given $(\boldsymbol{\beta}, \boldsymbol{v})$, and θ, one can solve the score equations

$$\frac{\partial h}{\partial \lambda_{0k}} = 0, \qquad k = 1, \dots, K,$$

to obtain the non-parametric maximum hierarchical likelihood estimator of λ_{0k},

$$\hat{\lambda}_0(y_{(k)}) = \frac{d_{(k)}}{\sum_{ij \in \mathcal{R}(y_{(k)})} \exp(\eta_{ij})}, \tag{2.4}$$

where $y_{(k)}$ is the k^{th} smallest distinct event time. Thus, $\widehat{\Lambda}_0(y_{ij}) = \sum_{k: y_{(k)} \le t} \hat{\lambda}_0(y_{(k)})$ where $d_{(k)}$ is the number of events at $y_{(k)}$ and $\mathcal{R}(y_{(k)}) = \{(i,j): y_{ij} \ge y_{(k)}\}$ is the risk set at $y_{(k)}$. This estimator is an extension of the estimator of the baseline cumulative hazard function for the Cox model to the frailty model.(Breslow 1972, Breslow 1974) After substituting in the estimated baseline hazard, the kernel of the profile h-likelihood $h^* = h|_{\Lambda_0(t) = \widehat{\Lambda}_0(t)}$ satisfies:

$$h^* \propto \sum_{ij} \delta_{ij} \eta_{ij} - \sum_{k: y(k) \le t} d_{(k)} \log \left[\sum_{ij \in \mathcal{R}(y_{(k)})} \exp(\eta_{ij}) \right] + \sum_{i} \ell_{2i}. \tag{2.5}$$

Given the frailty parameter θ, the maximum h-likelihood estimators (MHLE) of $\boldsymbol{\tau} = (\boldsymbol{v}, \boldsymbol{\beta})$ can be obtained by using the Newton-Raphson algorithm, as follows

$$\boldsymbol{\tau}^{(m+1)} \equiv \begin{pmatrix} \widehat{\boldsymbol{\beta}}^{(m+1)} \\ \widehat{\boldsymbol{v}}^{(m+1)} \end{pmatrix} = \begin{pmatrix} \widehat{\boldsymbol{\beta}}^{(m)} \\ \widehat{\boldsymbol{v}}^{(m)} \end{pmatrix} + \left(\boldsymbol{H}^{-1}(\boldsymbol{\tau})\boldsymbol{S}(\boldsymbol{\tau})\right)\Big|_{\boldsymbol{\tau}=(\boldsymbol{\beta},\boldsymbol{v})=\left(\widehat{\boldsymbol{\beta}}^{(m)},\widehat{\boldsymbol{v}}^{(m)}\right)}, \tag{2.6}$$

where $\widehat{\boldsymbol{\beta}}^{(m)}$ and $\widehat{\boldsymbol{v}}^{(m)}$ represent the estimates of $\boldsymbol{\beta}$ and $\boldsymbol{v}$ at the m^{th} iteration, $\boldsymbol{H}(\boldsymbol{\tau}) = -\frac{\partial^2 h^*}{\partial \boldsymbol{\tau}^2}$ is the $(p+G) \times (p+G)$ observed information matrix, and $\boldsymbol{S}(\boldsymbol{\tau}) = \left(\frac{\partial h^*}{\partial \boldsymbol{\beta}}, \frac{\partial h^*}{\partial \boldsymbol{v}}\right)^T$ is the score function. Details of the mathematical derivation of $\boldsymbol{S}(\boldsymbol{\tau})$ and $\boldsymbol{H}(\boldsymbol{\tau})$ are given in Appendix A.

After direction calculation, the first-order Laplace approximation to the adjusted profile marginal likelihood, $h_A^* = \log\{\int \exp(h^*) d\boldsymbol{v}\}$, can be written as

$$h_A^* = h^*|_{\boldsymbol{\tau}=\hat{\boldsymbol{\tau}}} + \frac{1}{2}\log\{\det(2\pi \boldsymbol{H}^{-1})\}\Big|_{\boldsymbol{\tau}=\hat{\boldsymbol{\tau}}}. \tag{2.7}$$

The adjusted profile h-likelihood seeks to approximate the restricted likelihood of θ by accounting for the estimation of $\boldsymbol{\beta}$ and $\boldsymbol{v}$. Then, the estimator of frailty parameter $\hat{\theta}$ can be obtained by solving the equation,

$$\frac{\partial h_A^*}{\partial \theta} = 0, \tag{2.8}$$

and checking that $\hat{\theta}$ is indeed the unique maximum of h_A^*.

2.3.2. Parametric baseline hazard models

In the previous section, we outlined the h-likelihood fitting procedure for cases when the functional form of the baseline hazard is unknown. In this section, we will layout the h-likelihood estimation process when the baseline hazard function follows a parametric distribution. Hence, the distribution of survival time can be estimated. The probability distributions of certain non-negative random variables can be used to describe the functional form of the baseline hazard $\lambda_0(t_{ij})$ in the equation (2.1), leading to parametric survival models that are commonly used to analyze time-to-event data. For example, the exponential, Weibull, Gompertz, log-logistic are broad families of probability distributions that can be specified.

We consider estimating the parameters in the parametric frailty models via Poisson hierarchical generalized linear models (HGLM). We can thus, utilize the iterative weighted least square (IWLS) technique for the estimation of the fixed effect and random effect parameters.(Ha and Lee 2003) The Poison representation of the parametric gamma frailties is as follows. We add zero to the ℓ_{1ij} in (2.3). That is, we add $\log\left[\frac{\Lambda_0(y_{ij})}{\Lambda_0(y_{ij})}\right]$ to ℓ_{1ij}, and after some algebraic derivations, we obtain,

$$\ell_{1ij} = \delta_{ij}\{\log\Lambda_0(y_{ij}) + \eta_{ij}\} - \Lambda_0(y_{ij})\exp(\eta_{ij}) + \delta_{ij}\log\left[\frac{\lambda_0(y_{ij})}{\Lambda_0(y_{ij})}\right],$$

$$= \delta_{ij}\log(\mu_{ij}) - \mu_{ij} + \delta_{ij}\log\left[\frac{\lambda_0(y_{ij})}{\Lambda_0(y_{ij})}\right], \tag{2.9}$$

where $\mu_{ij} = \Lambda_0(y_{ij})\exp(\eta_{ij})$ and $\eta_{ij} = \boldsymbol{x}_{ij}^T\boldsymbol{\beta} + \boldsymbol{z}_{ij}^T\boldsymbol{v}$.

The expression $\delta_{ij}\log(\mu_{ij}) - \mu_{ij}$ is similar to the kernel of the loglikelihood function of conditional Poisson for δ_{ij} given $\boldsymbol{V} = \boldsymbol{v}$ with mean μ_{ij}. Note that the last term $\log\left[\frac{\lambda_0(y_{ij})}{\Lambda_0(y_{ij})}\right]$ is independent of both $\boldsymbol{\beta}$ and $\boldsymbol{v}$.

As an illustration, and assuming an exponential parametric baseline hazard function, the term $\frac{\lambda_0(y_{ij})}{\Lambda_0(y_{ij})}$ becomes $\frac{1}{y_{ij}}$ where $\lambda_0(y_{ij}) = \rho$ and $\Lambda_0(y_{ij}) = \rho y_{ij}$ and no extra parameters are involved. It follows that

$$\log(\mu_{ij}) = \log\rho + \log y_{ij} + \boldsymbol{x}_{ij}^T\boldsymbol{\beta} + \boldsymbol{z}_{ij}^T\boldsymbol{v}.$$

We consider $\log \rho$ as an intercept β_0 on letting $\beta_0 = \log \rho$. Thus, we have

$$\log(\mu_{ij}) = \log y_{ij} + \boldsymbol{x}_{ij}^{*T}\boldsymbol{\beta}^* + \boldsymbol{z}_{ij}^T\boldsymbol{v},$$

where $\boldsymbol{x}_{ij}^{*T} = \left(x_{ij0}, \boldsymbol{x}_{ij}^T\right)^T$ with $x_{ij0} = (1, \dots, 1)^T$, $\boldsymbol{\beta}^* = (\beta_0, \boldsymbol{\beta}^T)^T$. Thus, the exponential baseline hazard frailty model can be directly fitted using PHGLM with the offset $\log y_{ij}$.

In a second example, we consider a Weibull baseline hazard frailty model where $\lambda_0(y_{ij}) = \gamma\rho y_{ij}^{\gamma-1}$ and $\Lambda_0(y_{ij}) = \rho y_{ij}^{\gamma}$. Therefore the term $\frac{\lambda_0(y_{ij})}{\Lambda_0(y_{ij})}$ becomes $\frac{\gamma}{y_{ij}}$, which depends on the unknown parameter γ. We have

$$\log(\mu_{ij}) = \gamma \log y_{ij} + \boldsymbol{x}_{ij}^{*T}\boldsymbol{\beta}^* + \boldsymbol{z}_{ij}^T\boldsymbol{v}.$$

We can combine the Weibull shape parameter γ with the frailty variates $\boldsymbol{v}$ by letting $z_{ij0} = \log y_{ij}$, $v_0 = \gamma$, $\boldsymbol{z}_{ij}^* = \left(z_{ij0}, \boldsymbol{z}_{ij}^T\right)^T$, $\boldsymbol{v}^* = (v_0, \boldsymbol{v}^T)^T$ and $\log(\mu_{ij}) = \boldsymbol{x}_{ij}^{*T}\boldsymbol{\beta}^* + \boldsymbol{z}_{ij}^{*T}\boldsymbol{v}^*$.

Thus, frailty models with exponential and Weibull hazard functions can be fitted via PHGLM using available standard statistical software. This approach can be extended to fitting parametric models with other baseline hazard functions such as Gompertz, extreme value distribution, etc.

Calculating the estimators will be simplified if matrices are used instead of the summations. Let $\boldsymbol{\eta}^* = \boldsymbol{X}^*\boldsymbol{\beta}^* + \boldsymbol{Z}^*\boldsymbol{v}^*$ where $\boldsymbol{X}^*$ is the $N \times (p+1)$ a matrix whose ij^{th} row vector is $\boldsymbol{x}_{ij}^{*T}$, $\boldsymbol{Z}^*$ is the $N \times (G+1)$ group indicator matrix whose ij^{th} row vector is $\boldsymbol{z}_{ij}^{*T}$. Let $\boldsymbol{\delta}$ be $N \times 1$ vector of δ_{ij}, and $\boldsymbol{\mu}^*$ be an $N \times 1$ vector equal to $\exp(\boldsymbol{\eta}^*)$. Given the frailty parameter θ, the MHLE for $\boldsymbol{\beta}^*$ and $\boldsymbol{v}^*$ in the parametric frailty models are obtained by solving the system of equations

$$\frac{\partial h}{\partial \boldsymbol{\beta}} = \boldsymbol{X}^{*T}(\boldsymbol{\delta} - \boldsymbol{\mu}^*), \tag{2.10}$$

$$\frac{\partial h}{\partial \boldsymbol{v}^*} = \boldsymbol{Z}^{*T}(\boldsymbol{\delta} - \boldsymbol{\mu}^*) + \boldsymbol{R}^*, \tag{2.11}$$

where $\boldsymbol{R}^* = \frac{\partial}{\partial \boldsymbol{v}^*}(d_i \log v_0, \ell_{2i})^T = (v_0^{-1} d_i, \boldsymbol{R})$, with $d_i = \sum_{ij} \delta_{ij}$, and $\boldsymbol{R} = \frac{\ell_{2i}}{\partial \boldsymbol{v}}$.

Given the frailty parameter θ, the IWLS method can be used to solve the above system for $\widehat{\boldsymbol{\beta}}^*$ and $\widehat{\boldsymbol{v}}^*$.

The IWLS equation for $(\boldsymbol{\beta}^*, \boldsymbol{v}^*)$ in the parametric gamma frailty model is given by

$$\begin{pmatrix} \boldsymbol{X^*}^T \boldsymbol{W} \boldsymbol{X}^* & \boldsymbol{X^*}^T \boldsymbol{W} \boldsymbol{Z}^* \\ \boldsymbol{Z}^{*T} \boldsymbol{W} \boldsymbol{X}^* & \boldsymbol{Z}^{*T} \boldsymbol{W} \boldsymbol{Z}^* + \boldsymbol{Q}^* \end{pmatrix} \begin{pmatrix} \widehat{\boldsymbol{\beta}^*} \\ \widehat{\boldsymbol{v}^*} \end{pmatrix} = \begin{pmatrix} \boldsymbol{X^*}^T \boldsymbol{W} \boldsymbol{w}^* \\ \boldsymbol{Z^*}^T \boldsymbol{W} \boldsymbol{w}^* + \boldsymbol{U}^* \end{pmatrix}, \tag{2.12}$$

where $\boldsymbol{w}^* = \boldsymbol{\eta} + \boldsymbol{W}^{-1}(\boldsymbol{\delta} - \boldsymbol{\mu})$, $\boldsymbol{Q}^*$ is the $(G+1) \times (G+1)$ diagonal matrix whose i^{th} element is $\left(-\frac{\partial \boldsymbol{R}^*}{\partial \boldsymbol{v}^*}\right)$, $\boldsymbol{U}^* = \boldsymbol{Q}^* \boldsymbol{v}^* + \boldsymbol{R}^*$ and $\boldsymbol{W} = \mathrm{diag}\{\exp(\boldsymbol{\eta}^*)\}$.

The two score equations (2.10) and (2.11) can also be expressed in a more compact form as

$$\frac{\partial h^*}{\partial \boldsymbol{\tau}^*} = \boldsymbol{E}^T(\boldsymbol{\delta} - \boldsymbol{\mu}^*) + \boldsymbol{B}, \tag{2.13}$$

where $\boldsymbol{E} = (\boldsymbol{X}^*, \boldsymbol{Z}^*)$ and $\boldsymbol{B} = (\boldsymbol{0}^T, \boldsymbol{R}^*)^T$. Thus, $\boldsymbol{E}\boldsymbol{\tau}^* = \boldsymbol{\eta} = \boldsymbol{X}^* \boldsymbol{\beta}^* + \boldsymbol{Z}^* \boldsymbol{v}^*$.

Next, from (2.10) and (2.11), we obtain the negative second partial derivatives with respect to $\boldsymbol{\beta}^*$ and $\boldsymbol{v}^*$,

$$-\frac{\partial^2 h^*}{\partial {\boldsymbol{\beta}^*}^2} = \boldsymbol{X^*}^T \boldsymbol{W} \boldsymbol{X}^*,$$

$$-\frac{\partial^2 h^*}{\partial \boldsymbol{\beta}^* \partial \boldsymbol{v}^*} = \boldsymbol{X^*}^T \boldsymbol{W} \boldsymbol{Z}^*,$$

$$-\frac{\partial^2 h^*}{\partial \boldsymbol{v}^* \partial \boldsymbol{\beta}^*} = \boldsymbol{Z}^{*T} \boldsymbol{W} \boldsymbol{X}^*,$$

$$-\frac{\partial^2 h^*}{\partial {\boldsymbol{v}^*}^2} = \boldsymbol{Z}^{*T} \boldsymbol{W} \boldsymbol{Z}^* + \boldsymbol{Q}^*,$$

where

$$\boldsymbol{Q}^* = \left\{-\frac{\partial \boldsymbol{R}^*}{\partial \boldsymbol{v}^*}\right\}.$$

Therefore,

$$\boldsymbol{H}^* = \begin{pmatrix} \boldsymbol{X}^{*T}\boldsymbol{W}\boldsymbol{X}^* & \boldsymbol{X}^{*T}\boldsymbol{W}\boldsymbol{Z}^* \\ \boldsymbol{Z}^{*T}\boldsymbol{W}\boldsymbol{X}^* & \boldsymbol{Z}^{*T}\boldsymbol{W}\boldsymbol{Z}^* + \boldsymbol{Q}^* \end{pmatrix}, \tag{2.14}$$

which can be written in a simple form as

$$\boldsymbol{H}^* = \boldsymbol{E}^{\boldsymbol{T}}\boldsymbol{W}^*\boldsymbol{E} + \boldsymbol{F}, \tag{2.15}$$

where $\boldsymbol{F} = \boldsymbol{BD}(\boldsymbol{0}, \boldsymbol{Q}^*)$ is a block diagonal matrix.

From $\widehat{\boldsymbol{\tau}^*} = \boldsymbol{\tau}^* + \boldsymbol{H}^{*-1}\left(\frac{\partial h^*}{\partial \boldsymbol{\tau}^*}\right)$, (13), and (15), we obtain,

$$\left(\boldsymbol{E}^{\boldsymbol{T}}\boldsymbol{W}\boldsymbol{E} + \boldsymbol{F}\right)\widehat{\boldsymbol{\tau}^*} = \left(\boldsymbol{E}^{\boldsymbol{T}}\boldsymbol{W}\boldsymbol{E} + \boldsymbol{F}\right)\boldsymbol{\tau}^* + \boldsymbol{E}^{\boldsymbol{T}}(\boldsymbol{\delta} - \boldsymbol{\mu}) + \boldsymbol{B},$$

$$= (\boldsymbol{E}^T\boldsymbol{W}\boldsymbol{E})\boldsymbol{\tau}^* + \boldsymbol{E}^{\boldsymbol{T}}(\boldsymbol{\delta} - \boldsymbol{\mu}) + \boldsymbol{b},$$

$$= \boldsymbol{E}^T\boldsymbol{W}\boldsymbol{w}^* + \boldsymbol{b},$$

where $\boldsymbol{b} = \boldsymbol{F}\boldsymbol{\tau}^* + \boldsymbol{B}$ and $\boldsymbol{w}^* = \boldsymbol{\eta} + \boldsymbol{W}^{-1}(\boldsymbol{\delta} - \boldsymbol{\mu}^*)$.

The asymptotic covariance matrix for $(\widehat{\boldsymbol{\tau}^*} - \boldsymbol{\tau}^*)$ is given by $\boldsymbol{H}^{*-1} = \left(-\frac{\partial^2 h}{\partial \boldsymbol{\tau}^{*2}}\right)^{-1}$. So, the upper left-hand corner of $\boldsymbol{H}^{*-1}$ gives the asymptotic variance-covariance matrix of $\widehat{\boldsymbol{\beta}^*}$ as

$$\mathrm{var}\left(\widehat{\boldsymbol{\beta}^*}\right) = \left(\boldsymbol{X}^{*T}\boldsymbol{\Sigma}^{-1}\boldsymbol{X}^*\right)^{-1},$$

where $\boldsymbol{\Sigma} = \boldsymbol{W}^{-1} + \boldsymbol{Z}^*\boldsymbol{U}^{*-1}\boldsymbol{Z}^{*T}$.

The adjusted profile hierarchical likelihood,

$$h_A = h|_{\boldsymbol{\tau}^* = \widehat{\boldsymbol{\tau}^*}} + \frac{1}{2}\log\{\det(2\pi\boldsymbol{H}^{-1})\}\Big|_{\boldsymbol{\tau}^* = \widehat{\boldsymbol{\tau}^*}}. \tag{2.16}$$

is used for estimation of the frailty parameter θ by solving

$$\frac{\partial h_A}{\partial \theta} = 0.$$

2.4. Bias-corrected maximum hierarchical likelihood estimator (MHLE)

The h-likelihood estimators generally perform well in various frailty models. However, they are known to be biased for non-normal random effects. Existing modifications to the h-likelihood first employ the total derivative approach, which considers $\hat{v}$ as a function of θ. Thus, the term $\partial\hat{v}/\partial\theta$ is added in the calculation since there is a direct dependence between $\hat{\boldsymbol{v}}$ when computing $\frac{\partial h_A^*}{\partial \theta}$.(Ha and Lee 2003) To further remove the bias, the second-order Laplace approximation is adopted.(Ha and Lee 2003) However, the total derivative and second-order Laplace approximation are computationally intensive by involving complex terms with complicated mathematical derivations.(Shun 1997, Ha and Lee 2003) Thus, the existing bias correction method cannot directly be applied to h-likelihood estimation for a broad class of distributions, such as correlated gamma frailty models.

In this chapter, we aim to provide adequate modifications of the adjusted profile h-likelihood estimators to improve the estimation for the parameter in the frailty distribution, as well as the estimates of regression coefficients. Remarkably, our proposed methods to remove the bias in the h-likelihood estimator are immediately applicable to a general class of frailty models, including correlated frailty models and joint modeling.

2.4.1 Bias Corrected profile h-likelihood function (BC-HL)

The estimate of the gamma frailty parameter θ is biased in a finite sample.(Ha and Lee 2003, Wang, Xu et al. 2011) Thus, the accuracy of the estimators of other parameters may be impacted. Under the linear mixed model framework, Morris(Morris 2006) proposed an adjustment on the classical likelihood function. We expand this adjustment procedure to the estimation of the h-likelihood dispersion parameter. The idea is to add an extra term to the adjusted profile h-likelihood function h_A^* for estimating θ to avoid zero estimates.

For our frailty model, we define the corrected profile as follows,

$$h_c = h_A^* + \log(\det(\boldsymbol{\Sigma}_{\boldsymbol{\theta}})), \tag{2.17}$$

where $\boldsymbol{\Sigma}_{\boldsymbol{\theta}}$ is a variance-covariance matrix for correlated random effects model. In our gamma frailty case, $\boldsymbol{\Sigma}_{\boldsymbol{\theta}}$ is θ. In multivariate frailty model as shown in Section 2.3.3, $\boldsymbol{\Sigma}_{\boldsymbol{\theta}}$ could be variance-covariance matrix to accommodate complex models.

The corrected profile h-likelihood has the following properties:

1. $\exp(h_c) = \exp(h_A^*) \det(\boldsymbol{\Sigma}_{\boldsymbol{\theta}}) \geq 0.$
2. $\exp(h_c) = 0$ only if $\det(\boldsymbol{\Sigma}_{\boldsymbol{\theta}}) = 0$. This ensures that the zero estimates of the dispersion parameter could be avoided.(Li and Lahiri 2010)

2.4.2 Modified score function to correct bias for profile h-likelihood (SC-HL)

The second proposed modification seeks to prevent downward bias in the adjusted profile likelihood estimate of the gamma frailty parameter, especially when the sample size is small. We modify the score function of the adjusted profile h-likelihood for estimating $\boldsymbol{\Sigma}_{\boldsymbol{\theta}}$ by adding an extra term to remove the bias estimate in the frailty parameter. Based on the bias reduction in fixed-effect models proposed by Firth(Firth 1993), we modified the adjusted profile likelihood score function equation. The goal is to prevent bias before computing $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\theta}}$, by adding a modification $M(\boldsymbol{\Sigma}_{\boldsymbol{\theta}})$, of order $O(1)$, to the adjusted profile likelihood score function. This leads to modified score adjusted likelihood equation,

$$U_m(\mathbf{\Sigma_\theta}) = U(\mathbf{\Sigma_\theta}) + M(\mathbf{\Sigma_\theta}) = \mathbf{0}, \tag{2.18}$$

where $U(\mathbf{\Sigma_\theta}) = \frac{\partial h_A^*}{\partial \mathbf{\Sigma_\theta}}$,

The solution of (2.18) gives the estimate $\widehat{\mathbf{\Sigma}}_{\mathbf{\theta m}}$, where $\widehat{\mathbf{\Sigma}}_{\mathbf{\theta m}}$ is the bias-corrected estimate of $\mathbf{\Sigma_\theta}$. The modification is chosen in such a way that

$$E_{\mathbf{\Sigma_\theta}}\left(\widehat{\mathbf{\Sigma}}_{\mathbf{\theta m}} - \mathbf{\Sigma_\theta}\right) = O(n^{-2}). \tag{2.19}$$

The latter can be achieved using Taylor's expansion for $U_m\left(\widehat{\mathbf{\Sigma}}_{\mathbf{\theta m}}\right)$ around $\mathbf{0}$ and finding an expression for $\left(\widehat{\mathbf{\Sigma}}_{\mathbf{\theta m}} - \mathbf{\Sigma_\theta}\right)$. By imposing condition (2.19), we find

$$M(\mathbf{\Sigma_\theta}) = -I(\mathbf{\Sigma_\theta})b(\mathbf{\Sigma_\theta}), \tag{2.20}$$

where $I(\mathbf{\Sigma_\theta}) = E_{\mathbf{\Sigma_\theta}}\left(-{\partial^2 l_{2i}}\middle/{\partial \mathbf{\Sigma_\theta^2}}\right)$ and the expected value is used to remove the first-order bias of $\widehat{\mathbf{\Sigma}}_{\mathbf{\theta m}}$. Thus, $M(\mathbf{\Sigma_\theta})$ does not depend on the observed sample. The bias $b(\mathbf{\Sigma_\theta})$ is given by

$$b(\mathbf{\Sigma_\theta}) = -\frac{1}{2}I(\mathbf{\Sigma_\theta})^{-2}\left\{v_{f,g,h} + v_{f,gh}\right\} = O(n^{-1}),$$

and

$$v_{f,g,h} = E_{\mathbf{\Sigma_\theta}}\left\{U_f U_g U_h\right\}, \quad v_{f,gh} = E_{\mathbf{\Sigma_\theta}}\left\{U_f U_{gh}\right\},$$

where U_f and U_{gh} denote the first and second derivative of l_{2i}. Here, l_{2i} is the log of the frailty density function.

Firth noticed the connection between $M(\mathbf{\Sigma_\theta})$ and Bayesian model-based priors. In full exponential models, the estimator $\widehat{\mathbf{\Sigma}}_{\mathbf{\theta m}}$ coincides with the mode of the posterior distribution obtained using Jeffrey's non-informative prior $|I(\mathbf{\Sigma_\theta})|^{\frac{1}{2}}$. To extend the Firth's method, we proposed to use Jeffrey's non-informative prior modification function $M(\mathbf{\Sigma_\theta})$ given by

$$M(\mathbf{\Sigma_\theta}) = |I(\mathbf{\Sigma_\theta})|^{\frac{1}{2}}. \tag{2.21}$$

The mathematical derivation of $M(\mathbf{\Sigma_\theta})$ can be found in Appendix B.

2.4.3 Extension to other frailty distributions and joint modeling

The proposed bias correction methods (BC-HL and SC-HL) in Section 2.3.1-2.3.2 can be directly extended to a broad class of frailty distributions and complex models such as joint modeling and competing risks. In Section 2.3.1-2.3.2, we illustrate our method using the gamma frailty mode as a proof of concept. However, the proposed methods can be applied to other frailty distribution such as the inverse Gaussian (IG) distribution,(Hougaard 1984) positive stable distribution,(Hougaard 1986) and Weibull distribution(Wang, Xu et al. 2011) that have been introduced as frailty distributions.

Extension example 1 (other (normal) and non-normal frailty distribution): In the equation (2.2), we can propose the IG frailty distribution for the U_i's with the probability density function with mean equal to 1 and unknown variance θ,

$$f(u_i;\theta) = \left[\frac{1}{2\pi\theta}\right]^{\frac{1}{2}} u_i^{-\frac{3}{2}} \exp\left[-\frac{(u_i-1)^2}{2u_i\theta}\right]. \tag{2.22}$$

The h-likelihood estimators of the IG frailty model or positive stable frailty model, or Weibull frailty model may lead to bias estimate of the model parameter since the distribution is non-normal. Thus, our approaches (BC-HL and SC-HL) can be extended to remove the bias in the h-likelihood estimators when the frailty distribution is not normal.

Extension example 2 (multivariate frailty distribution): The shared frailty model in equation (2.1) can be extended to describe complicated dependencies between survival times by introducing additional random effects. Here, a frailty model with more than one random component is of interest to model multilevel structures or hierarchical clustering of the data. The conditional hazard function of the multilevel frailty model may be written as

$$\lambda_{ij}\left(t_{ij}|\boldsymbol{X},\boldsymbol{\beta},\boldsymbol{Z}_1,\ldots,\boldsymbol{Z}_S,\boldsymbol{v}_1,\ldots,\boldsymbol{v}_S,\right)=\lambda_0\left(t_{ij}\right)\exp(\boldsymbol{X\beta}+\boldsymbol{Z}_1\boldsymbol{v}_1+\boldsymbol{Z}_2\boldsymbol{v}_2+\cdots+\boldsymbol{Z}_S\boldsymbol{v}_S)\,, \qquad (2.23)$$

where $\boldsymbol{Z}_r(r=1,2,..,S)$ are $N\times G_r$ with respect to the model matrices of $G_r\times 1$ frailty distribution. The aforementioned nested frailty or multicomponent model may result in a biased estimate of the variance component parameter.(Rondeau, Filleul et al. 2006) Therefore, interest lies in eliminating such bias, especially in finite samples. Our approaches in Section 2.3.1 and 2.3.2 to modify the h-likelihood estimator can be applied to reduce the bias in multicomponent frailty models.

Extension example 3 (joint modeling): One of the extensions to the frailty model in equation (2.1) is the consideration of different types of baseline hazard functions λ_{0r} $(r=1,\ldots,S)$ for multivariate survival times $\left(T_{ij1},\ldots,T_{ijS}\right)$. The shared frailty assumes that the censoring and event times are conditionally independent given the frailties. This assumption is violated when individuals who experience competing risk from different types of events might have different censoring mechanisms. However, traditional competing risk model is unable to tease out different censoring mechanisms. To address this issue, a generalization of the shared frailty model that incorporates competing risks as well as independent censoring is proposed. Thus, the cause-specific frailty model is given by

$$\lambda_{ijr}\left(t_{ijr}|v_i,\boldsymbol{x}_{ij}\right)=\lambda_{0r}\left(t_{ijr}\right)\exp\left(\boldsymbol{x}_{ij}^T\boldsymbol{\beta}_r+v_{ir}\right), \qquad (2.24)$$

where λ_{0r} is the baseline hazard function for event type r in cluster i and $\boldsymbol{\beta}_r$ is the $p\times 1$ regression parameter for the r^{th} time-to-event variable. If there is only one time-to-event variable, $r = 1$ then the model above reduces to the shared frailty model in equation (2.1). The h-likelihood approach treats random terms as unknown parameters. Thus, the sample size increases as the frailty parameters increase. This might lead to non-negligible bias to the coefficient parameter estimation. Therefore, it is worth investigating an appropriate bias reduction method, and our method to remove bias in the h-likelihood estimator can be applied.

2.5. Simulation studies

We conducted simulation studies to evaluate the finite sample performance of the proposed bias reduction methods of the h-likelihood estimators (BC-HL and SC-HL) and compared them with the h-likelihood estimators (HL)(Ha, Lee et al. 2001) for estimating the frailty parameter in the semiparametric gamma frailty model. BC-HL refers to the method described in Section 2.3.1, and SC-HL refers to the method described in Section 2.3.2.

The frailties $u_i, i = 1, .., G,$ were generated from a gamma distribution with mean 1 and variance parameter $\theta = 0.5$, 1.0, and 2.0. Given $U_i = u_i$, the independent survival times T_{ij}, $j = 1, \dots, n_i$ were generated from an exponential distribution with parameter λ_{ij}. Without loss of generality, we consider only one predictor. The regression parameter β is related to the hazard rate λ_{ij} via the frailty model

$$\lambda_{ij} = \lambda_0(t) u_i \exp(\beta x_{ij}), \tag{2.25}$$

where $\lambda_0(t) = 1.0$, $\beta = 1.0$. We set $x_{ij} = 0$ for the first $G/2$ individuals to form the control group, and $x_{ij} = 1$ for the remaining $G/2$ to form the treatment group. Thus, the survival time is

$$T_{ij} = \left(\frac{-\log U_{ij}}{u_i \exp(\beta x_{ij})} \right), U_{ij} \sim Uniform\ (0,1). \tag{2.26}$$

Given the $u_i's$, the corresponding censoring time C_{ij} was generated from a uniform distribution $U(0, l)$ with parameter l empirically determined to achieve approximately the desired censoring rate of 30%. We used a sample size $N = G * n_i$ with $N = 100$ with $(G;\ n_i) = (25;\ 4)$ and $N = 200$ with $(G;\ n_i) = (50;\ 4)$. From the sample of 250 replications of simulated data, we computed the mean, the standard deviation, the mean of the estimated standard error, and the mean squared error for $\hat{\beta}$. The sample standard deviation is calculated by $\left\{\frac{1}{249}\sum_{i=1}^{250}\left(\hat{\beta}_i - \bar{\beta}\right)^2\right\}^{1/2}$ where $\bar{\beta} = 250^{-1}\sum_{i=1}^{250}\hat{\beta}_i$, and the sample mean squared error is given by $\left\{\frac{1}{250}\sum_{i=1}^{250}\left(\hat{\beta}_i - \beta\right)^2\right\}$. The sample standard error was obtained from the observed information $\boldsymbol{H}$ in (6). Also, we calculated the empirical coverage probability for a nominal 95% confidence interval for β. Similarly, the mean, standard deviation, and mean squared error for $\hat{\theta}$ were calculated for the frailty parameter θ. For the comparison of the estimation methods for θ, we also compute relative percent bias, denoted by $\%bias = \left\{\frac{mean(\hat{\theta})-\theta}{\theta}\right\} \times 100$. The empirical assessment was conducted using the R Software Version 4.0.2.(R Core Team 2020)

The results from fitting the semiparametric frailty models, where the baseline hazard function is assumed unknown, are summarized in Table 2.1 and Table 2.2. To study the bias of omitting random effects, the semiparametric Cox proportional hazard model is estimated too. From the same simulation settings, we also fit the exponential and Weibull parametric frailty models, where the baseline hazard was specified to follow the exponential distribution and Weibull distribution, respectively, and the results are summarized in Table 2.3 and Table 2.4.

Table 2.1 summarizes the simulation results for the regression coefficient $\hat{\beta}$ under the semiparametric frailty models. Here, the baseline hazard function is non-parametric, and the frailty follows gamma distribution. We found that when the cluster level is small, and the frailty variance is 0.5 or 1.0, the h-likelihood slightly underestimated the fixed effect parameter. Our proposed methods (BC-HL and SC-HL) reduced the bias in the h-likelihood estimator from 7% to 4% when the cluster level is 25 and the frailty variance is 0.5. In all simulated scenarios, the proposed methods and the h-likelihood estimators provide a satisfactory estimate for the regression fixed effect parameter. As expected, the estimates of the regression coefficient are hugely underestimated when the random effect is ignored in the semiparametric Cox proportional hazard model. These results confirm the importance of incorporating frailty term into survival model to describe heterogeneity between clusters or individuals.

Tables 2.2 lists the results from the simulation study for the frailty parameter θ under the semiparametric frailty models. Our results show that the h-likelihood estimate of the frailty parameter is underestimated. The bias is more pronounced when the variance of frailty and cluster level is small but reduces as the sample size increases (as expected). In contrast, the proposed modifications (BC-HL and SC-HL) reduced the bias in the variance of frailty substantially, and the bias is minimal when the cluster level increases. It important to note that our proposed methods decreased the bias in the h-likelihood from 30% to 3% when the frailty variance is 0.5 and cluster level 25. For the proposed method in which one term with dispersion parameter is added to the adjusted profile likelihood (BC-HL), the bias and mean squared error were consistently the lowest across all situations we considered. Also, the bias and MSE tend to zero in all the methods (HL, BC-HL, SC-HL) when cluster level increases. Thus, as a general rule, the results indicate that the h-likelihood and the proposed modification estimates of the gamma frailty models demonstrate good asymptotic properties.

Next, we investigated the performance of our procedure when the baseline hazard functions are misspecified. We assumed that the correct model follows the Weibull-gamma frailty models; that is, the correct baseline hazard and frailty distributions are Weibull and gamma, respectively. Given the gamma frailties $u_i, i = 1,..,G$, as shown above, the independent survival times were generated from Weibull distribution with scale parameter $u_i \exp(\beta_0 + \beta_1 x_{ij})$ and shape parameter $\gamma = 1.5$, indicating an increasing hazard. The parameters β_0, β_1, and γ are related to the hazard rate λ_{ij}^* via the frailty model

$$\lambda_{ij}^* = \lambda_0(t) u_i \exp(\beta_0 + \beta_1 x_{ij}). \tag{2.27}$$

where $\lambda_0(t) = \gamma y_{ij}^{\gamma-1}$ and $\beta_0 = \beta_1 = 1.0$. The simulation setting is the same as that of the semiparametric gamma frailty model above. We fitted both parametric and semiparametric frailty models to each of the 250 simulated datasets.

Table 2.3 summarizes the simulation results from fitting the exponential-gamma and Weibull-gamma parametric frailty models, where the correct baseline hazard follows exponential, and the frailty distribution is gamma. As expected, fitting both exponential-gamma and Weibull-gamma frailty models gives good overall results of the estimate of the regression fixed effect parameter. However, the h-likelihood leads to the underestimation of the frailty parameter. Our methods to modify the h-likelihood reduce this bias. For example, in the exponential-gamma models, our proposed methods (BC-HL and SC-HL) reduced the bias in the h-likelihood (HL) for the frailty parameter estimate from 27% to 6% and 11%, respectively, when the true frailty parameter was 0.5, and the cluster level was 25. Furthermore, the frailty parameter bias decreased from 12% in the HL to 2% in the BC-HL approach and 4% in the SC-HL approach when the

Weibull-gamma models were fitted considering the true frailty variance was 2.0, and the cluster level was 25.

When the correct baseline hazard follows Weibull, and the frailty distribution is gamma, the simulation results are summarized in Table 2.4. Our results show that when the Weibull-gamma models are right, we find that the exponential-gamma fittings are very bad. Our methods (BC-HL and SC-HL) to modify the h-likelihood reduced the bias in the h-likelihood in all the scenarios considered. The simulation results of Table 2.4 indicate that the misspecified baseline hazard creates substantial bias. In a nutshell, the correct model specifications about baseline hazards are crucial for valid inferences. If these are wrongly specified, the parameter estimates suffer from significant biases. Thus, when the correct baseline hazards are uncertain, the use of non-parametric models is desirable. When parametric frailty models are used, model-checking for baseline hazard is necessary. If correctly specified, the gain of information would be somewhat higher for frailty parameter estimation and therefore provides correct standard error estimation for regression parameter estimation.

We also have found that the resulting pattern from fitting the semiparametric models under the correct Weibull-gamma are similar to those evident in the correct exponential-gamma as shown in Table 2.1 and Table 2.2. That is, our proposed methods (BC-HL and SC-HL) shrunk the bias of the h-likelihood (HL).

Table 2.1 Simulation results on the estimation of the regression parameter in the semiparametric model where the baseline hazard is non-parametric and the frailty distribution is gamma.

θ	(G, n_i)	Method	Regression Parameter ($\beta = 1.0$)			
			Mean	SD (se)	MSE	95% CP
0.5	25; 4	Cox	0.765	0.314 (0.249)	0.153	0.784
		HL	0.928	0.371 (0.353)	0.142	0.932
		BC-HL	0.958	0.382 (0.384)	0.147	0.956
		SC-HL	0.952	0.380 (0.377)	0.146	0.956
	50; 4	Cox	0.784	0.243 (0.175)	0.105	0.668
		HL	1.017	0.293 (0.272)	0.086	0.948
		BC-HL	1.032	0.296 (0.283)	0.089	0.952
		SC-HL	1.029	0.296 (0.281)	0.088	0.952
1.0	25; 4	Cox	0.606	0.403 (0.246)	0.317	0.556
		HL	0.938	0.498 (0.453)	0.250	0.916
		BC-HL	0.957	0.507 (0.478)	0.258	0.924
		SC-HL	0.953	0.505 (0.473)	0.256	0.924
	50; 4	Cox	0.624	0.296 (0.172)	0.228	0.436
		HL	1.032	0.408 (0.352)	0.167	0.908
		BC-HL	1.042	0.412 (0.361)	0.171	0.912
		SC-HL	1.040	0.412 (0.359)	0.171	0.912
2.0	25; 4	Cox	0.473	0.423 (0.244)	0.456	0.428
		HL	0.991	0.597 (0.577)	0.355	0.948
		BC-HL	1.006	0.607 (0.599)	0.367	0.952
		SC-HL	1.004	0.605 (0.595)	0.365	0.948
	50; 4	Cox	0.430	0.315 (0.171)	0.424	0.224
		HL	0.985	0.509 (0.468)	0.258	0.936
		BC-H	0.992	0.512 (0.477)	0.261	0.940
		SC-HL	0.991	0.511 (0.476)	0.261	0.940

The simulation is conducted with 250 replications for each of the sample sizes N, G is th e number of clusters, n_i is the number of members in each cluster. Cox is the Cox propo rtional hazard model ignoring frailty, HL is the original h-likelihood method, BC-HL is the proposed bias-corrected HL, SC-HL is the proposed score function modification on the H L. Mean, SD, and MSE indicates the mean, standard deviation, and mean squared error for $\hat{\beta}$, respectively. Also, se and 95% CP indicate the mean of estimated standard errors for $\hat{\beta}$ and empirical coverage probability for a nominal 95% of the confidence interval for β.

Table 2.2 Simulation results on the estimation of the frailty parameter θ in the semiparametric model where the baseline hazard is non-parametric and the frailty distribution is gamma.

θ	(G, n_i)	Method	Mean	SD	MSE	%bias
0.5	25; 4	HL	0.348	0.185	0.057	-30.4
		BC-HL	0.483	0.187	0.035	-3.4
		SC-HL	0.451	0.188	0.037	-9.8
	50; 4	HL	0.440	0.158	0.029	-12.0
		BC-HL	0.505	0.159	0.025	1.0
		SC-HL	0.489	0.159	0.025	-2.2
1.0	25; 4	HL	0.831	0.284	0.109	-16.9
		BC-HL	0.982	0.298	0.089	-1.8
		SC-HL	0.951	0.299	0.091	-4.9
	50; 4	HL	0.943	0.233	0.057	-5.7
		BC-HL	1.018	0.239	0.057	1.8
		SC-HL	1.002	0.239	0.057	0.2
2.0	25; 4	HL	1.656	0.380	0.262	-17.2
		BC-HL	1.852	0.403	0.184	-7.4
		SC-HL	1.821	0.403	0.194	-9.0
	50; 4	HL	1.877	0.305	0.108	-6.2
		BC-HL	1.976	0.314	0.099	-2.4
		SC-HL	1.962	0.314	0.099	-3.8

The simulation is conducted with 250 replications for each of the sample sizes N, G is th e number of clusters, n_i is the number of members in each cluster. Cox is the Cox propo rtional hazard model ignoring frailty, HL is the original h-likelihood method, BC-HL is the proposed bias-corrected HL, SC-HL is the proposed score function modification on the H L. Mean and SD indicates the mean and standard deviation for $\hat{\theta}$. MSE is the mean squa red error.

Table 2.3 Simulation results on the estimation of the regression parameter in the exponential-gamma frailty models where the correct baseline hazard is exponential, and the frailty distribution is gamma

θ	(G, n_i)	Model	Method	Baseline hazard Parameters				Regression Parameter ($\beta = 1.0$)				Frailty Parameter	
				$\rho = 1.0$		$\gamma = 1.0$							
0.5				Mean	MSE	Mean	MSE	Mean	SD (se)	MSE	95% CP	Mean	MSE
	25; 4	E-G	HL	1.115	0.079	-	-	0.942	0.365(0.348)	0.136	0.924	0.364	0.048
			BC-HL	1.148	0.092	-	-	0.949	0.369(0.372)	0.138	0.948	0.471	0.030
			SC-HL	1.141	0.089	-	-	0.947	0.368(0.366)	0.137	0.944	0.445	0.033
		W-G	HL	1.135	0.108	1.017	0.013	0.954	0.388(0.358)	0.152	0.924	0.380	0.054
			BC-HL	1.184	0.135	1.040	0.015	0.977	0.397(0.386)	0.158	0.944	0.506	0.041
			SC-HL	1.174	0.129	1.035	0.015	0.972	0.395 (0.380)	0.156	0.936	0.476	0.041
	50; 4	E-G	HL	0.916	0.034	-	-	1.029	0.283 (0.268)	0.081	0.936	0.455	0.021
			BC-HL	0.929	0.032	-	-	1.032	0.283(0.276)	0.081	0.948	0.506	0.019
			SC-HL	0.926	0.033	-	-	1.031	0.283(0.275)	0.081	0.940	0.494	0.019
		W-G	HL	0.917	0.035	1.000	0.006	1.030	0.293 (0.274)	0.087	0.932	0.457	0.025
			BC-HL	0.934	0.034	1.011	0.006	1.041	0.296 (0.284)	0.089	0.948	0.517	0.024
			SC-HL	0.931	0.034	1.001	0.079	1.039	0.296 (0.282)	0.089	0.948	0.503	0.024
1.0													
	25; 4	E-G	HL	1.015	0.122	-	-	0.953	0.497(0.451)	0.249	0.928	0.869	0.076
			BC-HL	1.038	0.128	-	-	0.955	0.497 (0.469)	0.248	0.928	0.988	0.063
			SC-HL	1.034	0.126	-	-	0.955	0.497(0.465)	0.248	0.928	0.963	0.065
		W-G	HL	1.026	0.135	1.011	0.012	0.960	0.511(0.461)	0.261	0.920	0.886	0.091
			BC-HL	1.060	0.151	1.028	0.013	0.976	0.519 (0.483)	0.269	0.924	1.029	0.086
			SC-HL	1.054	0.148	1.025	0.013	0.973	0.517 (0.479)	0.267	0.920	1.000	0.085
	50; 4	E-G	HL	0.866	0.073	-	-	1.056	0.414 (0.353)	0.174	0.900	0.994	0.035
			BC-HL	0.875	0.072	-	-	1.057	0.415 (0.356)	0.175	0.900	1.052	0.040
			SC-HL	0.874	0.073	-	-	1.057	0.415 (0.358)	0.175	0.900	1.041	0.038
		W-G	HL	0.867	0.075	0.990	0.005	1.044	0.410 (0.357)	0.170	0.916	0.984	0.051
			BC-HL	0.880	0.075	0.999	0.005	1.053	0.414 (0.365)	0.174	0.920	1.054	0.057
			SC-HL	0.877	0.075	0.997	0.005	1.052	0.413 (0.364)	0.173	0.920	1.040	0.055
2.0													
	25; 4	E-G	HL	0.942	0.151	-	-	1.012	0.593 (0.580)	0.350	0.948	1.774	0.141
			BC-HL	0.960	0.156	-	-	1.012	0.594(0.595)	0.351	0.948	1.927	0.104
			SC-HL	0.958	0.155	-	-	1.012	0.594(0.592)	0.351	0.948	1.904	0.108
		W-G	HL	0.942	0.173	0.997	0.012	1.011	0.605 (0.587)	0.365	0.952	1.767	0.186
			BC-HL	0.971	0.186	1.011	0.013	1.024	0.615(0.606)	0.377	0.952	1.954	0.150
			SC-HL	0.967	0.184	1.009	0.013	1.022	0.613 (0.603)	0.375	0.952	1.926	0.153
	50; 4	E-G	HL	0.871	0.140	-	-	1.027	0.530 (0.478)	0.281	0.920	2.027	0.064
			BC-HL	0.879	0.141	-	-	1.026	0.530 (0.485)	0.281	0.920	2.105	0.077
			SC-HL	0.878	0.141	-	-	1.026	0.530 (0.485)	0.281	0.920	2.094	0.075
		W-G	HL	0.863	0.152	0.976	0.006	0.999	0.513 (0.476)	0.262	0.940	1.967	0.089
			BC-HL	0.876	0.157	0.983	0.006	1.005	0.516(0.485)	0.265	0.940	2.062	0.097
			SC-HL	0.874	0.156	0.982	0.006	1.004	0.516(0.483)	0.265	0.940	2.048	0.095

E-G, BC-EG, and SC-E denote exponential-gamma frailty model, where the baseline ha zard is assumed exponential and the frailty is gamma, the proposed bias-corrected EG, score function modification of EG, respectively. Similarly, W-G, BC-WG, and SC-WG rep resent Weibull-gamma where the baseline hazard is assumed Weibull and the frailty is g amma, the proposed bias-corrected WG, score function modification of WG, respectively . HL is the original h-likelihood method, BC-HL is the proposed bias-corrected HL, SC-H L is the proposed score function modification on the HL.

Table 2.4 Simulation results on the estimation of the model parameters in the Weibull-p gamma frailty models where the correct baseline hazard Weibull with shape parameter $\gamma = 1.5$ and the frailty distribution is gamma

θ	(G, n_i)	Model	Method	Baseline hazard Parameters				Regression Parameter ($\beta = 1.0$)				Frailty Parameter	
				$\rho = 1.0$		$\gamma = 1.5$							
0.5				Mean	MSE	Mean	MSE	Mean	SD (se)	MSE	95%CP	Mean	MSE
	25; 4	E-G	HL	0.791	0.065	-	-	0.758	0.279 (0.292)	0.136	0.868	0.185	0.115
			BC-HL	0.818	0.055	-	-	0.762	0.277 (0.326)	0.133	0.932	0.291	0.058
			SC-HL	0.812	0.057	-	-	0.761	0.278 (0.320)	0.134	0.928	0.265	0.069
		W-G	HL	0.899	0.071	1.513	0.028	1.055	0.386 (0.392)	0.151	0.952	0.507	0.048
			BC-HL	0.933	0.073	1.547	0.032	1.077	0.395 (0.418)	0.161	0.960	0.635	0.068
			SC-HL	0.925	0.073	1.540	0.030	1.072	0.393 (0.412)	0.159	0.956	0.605	0.062
	50; 4	E-G	HL	0.837	0.037	-	-	0.760	0.202 (0.207)	0.099	0.760	0.161	0.122
			BC-HL	0.853	0.032	-	-	0.765	0.201 (0.221)	0.096	0.796	0.216	0.087
			SC-HL	0.849	0.033	-	-	0.764	0.202 (0.208)	0.096	0.784	0.203	0.095
		W-G	HL	0.971	0.033	1.511	0.013	1.056	0.280 (0.278)	0.081	0.960	0.466	0.024
			BC-HL	0.991	0.034	1.529	0.014	1.069	0.283 (0.288)	0.085	0.960	0.527	0.024
			SC-HL	0.986	0.034	1.524	0.014	1.066	0.282 (0.286)	0.084	0.956	0.512	0.023
1.0													
	25; 4	E-G	HL	0.813	0.072	-	-	0.727	0.377(0.362)	0.215	0.836	0.438	0.345
			BC-HL	0.834	0.065	-	-	0.728	0.374 (0.383)	0.213	0.872	0.543	0.238
			SC-HL	0.829	0.066	-	-	0.728	0.375 (0.383)	0.214	0.860	0.518	0.262
		W-G	HL	0.986	0.136	1.500	0.025	1.006	0.508 (0.466)	0.258	0.932	0.890	0.098
			BC-HL	1.021	0.151	1.528	0.026	1.022	0.517 (0.489)	0.267	0.940	1.034	0.095
			SC-HL	1.015	0.148	1.522	0.026	1.019	0.515 (0.484)	0.265	0.940	1.005	0.094
	50; 4	E-G	HL	0.738	0.082	-	-	0.772	0.283 (0.278)	0.132	0.864	0.522	0.245
			BC-HL	0.747	0.078	-	-	0.773	0.282 (0.285)	0.131	0.872	0.571	0.201
			SC-HL	0.745	0.079	-	-	0.772	0.282 (0.283)	0.131	0.872	0.560	0.211
		W-G	HL	0.839	0.066	1.480	0.011	1.074	0.392 (0.361)	0.159	0.928	0.996	0.047
			BC-HL	0.852	0.064	1.493	0.012	1.083	0.395 (0.370)	0.163	0.940	1.067	0.053
			SC-HL	0.850	0.066	1.491	0.012	1.081	0.395 (0.368)	0.162	0.936	1.053	0.052
2.0													
	25; 4	E-G	HL	0.765	0.107	-	-	0.703	0.441(0.479)	0.281	0.920	1.075	0.912
			BC-HL	0.778	0.102	-	-	0.704	0.436 (0.495)	0.277	0.936	1.198	0.704
			SC-HL	0.776	0.103	-	-	0.704	0.437 (0.492)	0.278	0.932	1.175	0.743
		W-G	HL	0.982	0.214	1.513	0.027	1.004	0.603 (0.604)	0.362	0.944	1.854	0.208
			BC-HL	1.012	0.233	1.536	0.029	1.018	0.609(0.624)	0.370	0.956	2.045	0.212
			SC-HL	1.008	0.231	1.533	0.028	1.017	0.608 (0.621)	0.369	0.956	2.017	0.210
	50; 4	E-G	HL	0.705	0.117	-	-	0.749	0.374 (0.375)	0.203	0.888	1.203	0.659
			BC-HL	0.712	0.114	-	-	0.748	0.373 (0.382)	0.202	0.888	1.263	0.568
			SC-HL	0.710	0.114	-	-	0.748	0.374 (0.381)	0.202	0.888	1.252	0.584
		W-G	HL	0.848	0.112	1.472	0.012	1.053	0.507 (0.478)	0.259	0.924	1.980	0.072
			BC-HL	0.860	0.112	1.483	0.012	1.060	0.511 (0.487)	0.263	0.928	2.075	0.081
			SC-HL	0.858	0.112	1.482	0.012	1.059	0.510 (0.485)	0.263	0.928	2.062	0.079

E-G, BC-EG, and SC-E denote the exponential-gamma frailty model, where the baseline hazard is assumed exponential, and the frailty is gamma, the proposed bias-corrected E-G, score function modification of E-G, respectively. Similarly, W-G, BC-WG, and SC-WG represent Weibull-gamma where the baseline hazard is assumed Weibull and the frailty is gamma, the proposed bias-corrected W-G, score function modification of W-G, respectively. HL is the original h-likelihood method, BC-HL is the proposed bias-corrected HL, SC-HL is the proposed score function modification on the HL.

2.6. Data example

The kidney infectious data(Hanagal 2020) is related to the recurrence time to event, which is infection at the point of insertion of the catheter for 38 kidney patients using portable dialysis equipment. For each patient, the first and second recurrence time (in days) of infection from the time of insertion of the catheter until it has to be removed owing to infection was recorded. The catheter may have to be removed for reasons other than kidney infection, and this was regarded as censoring. The data consist of demographics variables age and sex, and three dichotomous disease type variables: Glomerulo Neptiritis (GN), Acute Neptiritis (AN), and Polycyatic Kidney Disease (PKD). The standard Cox model, the extended Cox Anderson-Gill (A-G) method,(Andersen and Gill 1982) the h-likelihood approach, and the proposed bias reduction methods were fitted to this data. The A-G method generalizes the Cox model, for analyzing data when all dependence between subsequent events is induced by time-dependent covariates. We report the results from fitting the semiparametric frailty models, where the baseline hazard is non-parametric, and the frailty is gamma distribution. The conditional Akaike's Information Criterion (cAIC) are reported for models considered. Details of the computation of cAIC can be found in Appendix C.

From Table 2.5, we find that sex is the only significant effect, which indicates that females have a lower infection rate than males. The absolute value of the coefficient estimator of the effect of sex in the proposed bias-corrected models is larger than in the h-likelihood method. Thus, the modifications of the h-likelihood improve the fixed effect estimates. The estimators have a clear difference between the semiparametric Cox proportional hazard and A-G model and the frailty models. The estimators of the Cox and A-G methods are closer to zero more than the frailty models except for the coefficient of PKD. We found that the fitting of A-G produces overall the smallest standard errors as expected. However, the model fitting is bad from the cAIC. The A-G model fails to describe the kidney data better because, in the kidney data, a subject can be censored in the first follow-up but can have an event in the second follow-up. The A-G method assumes that censoring terminates further occurrence of the event. Thus, the A-G method may not be appropriate for analyzing the kidney data.

From the cAIC values, the two bias reduction methods (BC-HL and SC-HL) provide a better fit to the kidney data. The estimated variances of the frailty distribution are 0.459 and 0.411 from the proposed methods and 0.212 from the h-likelihood method. Based on the simulation results, the traditional h-likelihood method will underestimate the frailty variance, while the proposed methods will have a more precise estimate for this parameter.

The h-likelihood approach allows for inference on the random effects rather than on just estimating the frailty parameters. Predictions and their intervals are important in investigating heterogeneity across centers. The estimation of the standard errors in the h-likelihood approach for the prediction of random effects, which is required to construct $100(1 - \alpha)\%$ prediction intervals are obtained from the lower right-hand corner of $\boldsymbol{H}^{-1}$. However, the adjusted profile likelihood h_A^* in equation (2.8) can give a zero estimate of the frailty parameter when the sample is small(Ha, Jeong et al. 2017), leading to null confidence intervals for $\boldsymbol{v}$. We further show that our proposed methods (BC-HL and SC-HL) are adequate to avoid zero estimates in the frailty parameter. We predict the realizations of the random effects for the kidney catheter data and construct the 95% Wald confidence intervals (CI) of individual frailties of each patient. Figure 1 (a)-(c) displays the estimated frailties of 38 patients and their 95% CI for HL, BC-HL, and SC-HL models. The models included sex since is it the only statistically significant effect covariate. From the plots, we note that the patient’s realized frailty effects on the recurrent times are heterogeneous. Patient 21 was identified to have a significantly lower hazard, and the corresponding 95% CI does not include zero. Thus, we find that a graphical display such as Figure 1 is useful to investigate a particular patient's heterogeneity.

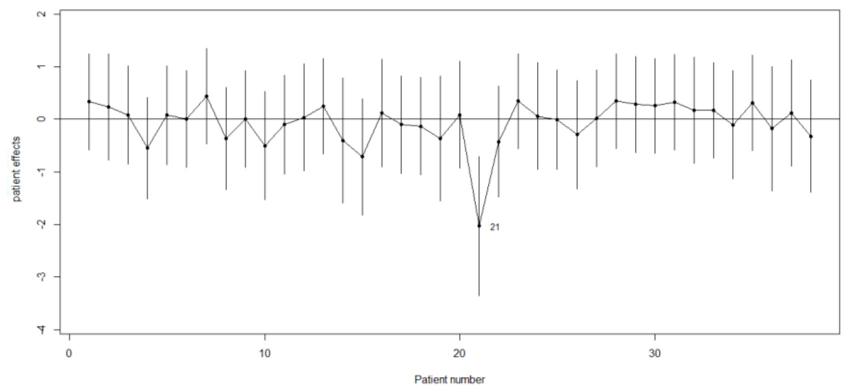
Table 2.5 Results of model fitting for kidney infection data.

Model	Cox	Cox-A-G	HL	BC-HL	SC-HL
Age	0.003 (0.011)	0.005 (0.003)	0.004 (0.013)	0.004 (0.015)	0.004 (0.014)
Sex	-1.472 (0.358)	-0.987 (0.097)	-1.608 (0.417)	-1.756 (0.473)	-1.730 (0.463)
GN	0.089 (0.407)	-0.436 (0.086)	0.157 (0.472)	0.231 (0.543)	0.217 (0.530)
AN	0.352 (0.400)	-0.665 (0.096)	0.375 (0.470)	0.426 (0.546)	0.415 (0.532)
PKD	-1.428 (0.631)	-1.210 (0.119)	-1.154 (0.770)	-0.825 (0.873)	-0.878 (0.858)
$\hat{\theta}$	-	-	0.212	0.459	0.411
cAIC	368.789	12290.78	364.769	362.230	362.583

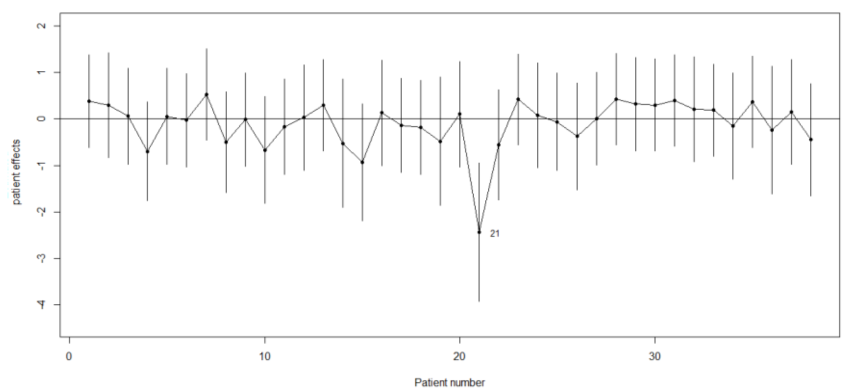
Notes: Sex = 1 for female and 0 for male; GLN = 1 represents glomerulus's nephritis, els e GN = 0; AN = 1 and PKD = 1 represent acute nephritis and polycystic kidney disease (PKD), respectively, else CAN = 0 and PKD = 0, respectively; in parentheses are the stan dard error, which is the square root of diagonal element of $\boldsymbol{H}^{-1}$; θ is the estimator of the f railty parameter. Cox is the Cox proportional hazard model, Cox-A-G is the Anderson Gil l approach, HL is the original h-likelihood method, BC-HL is the proposed bias-corrected HL, SC-HL is the proposed score function modification on the HL

Figure 2. 1. Frailty estimates and the 95% confidence intervals of individual frailties of patients under the gamma frailty model

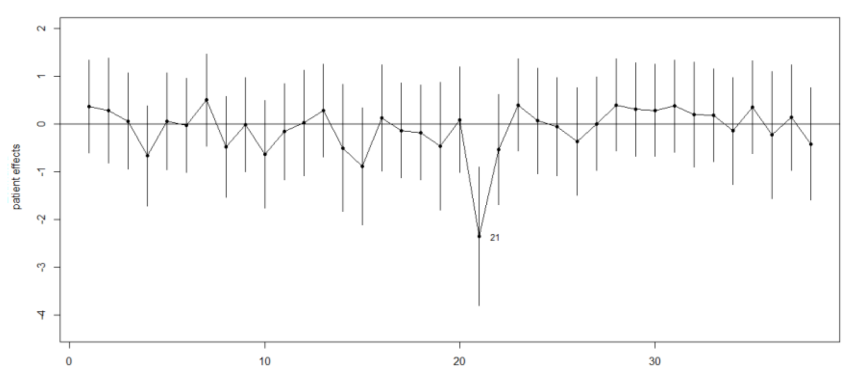
(a) HL method



(b) BC-HL method



(c) SC-HL method



2.7. Discussion

In this chapter, we have considered two simple but effective bias correction methods for the hierarchical likelihood (h-likelihood) approach for the frailty model under both the semiparametric proportional hazard and also the parametric hazard framework. The h-likelihood method has multiple advantages in computation when the frailty distribution is multivariate because it avoids integration over the frailty distribution. Another appealing feature of the h-likelihood is the possibility to conduct statistical inference on the latent frailties, which is often not feasible under the classical maximum likelihood approach. For example, the EM algorithm, focused on only parameter estimation, where the latent frailties are integrated out to obtain the parameter estimates. The h-likelihood generally performs well but can be biased when the frailty distribution is not normal. Simulation study results demonstrate that the proposed bias reduction methods overcome this issue. The efficiency of the proposed bias reduction methods for $\hat{\theta}$ is improved.

For the hierarchical generalized linear models, the previous studies have(Lee and Nelder 2001, Noh and Lee 2007) proposed the use of second-order Laplace approximation to estimate the dispersion parameters. Recently, some studies (Ha and Lee 2003, Wang, Xu et al. 2011, Christian, Ha et al. 2016) discussed the total derivative approach to reduce bias in h-likelihood estimators for frailty models. Their strategy uses the second-order Laplace approximation and $\frac{\partial \hat{v}}{\partial \theta}$ to estimate the dispersion parameters. However, this method could be hard to compute as it involves too many complicated terms and cumbersome mathematical derivations, especially, under competing risks and joint frailty models . Ha, Vaida et al. (2016) considered the adjustment of the h-likelihood estimators in section 2.3.1 for interval estimation for individual frailties of the clusters, not the parameters of the frailty distribution, and the frailty distribution was log-normal. When we extended this adjustment to the gamma frailty to reduce the bias in the h-likelihood estimator of variance in frailty distribution, the bias is reduced substantially. The real data analysis further confirmed the superiority of the proposed method over the h-likelihood by.(Ha, Lee et al. 2001) The bias reduction methods discussed in this paper are straightforward to implement, and integration with the estimation of the correlated frailty model using h-likelihood is an exciting prospect for future research.

Chapter 3

Modeling heterogeneity for clustered survival data by log-logistic distribution

Abstract

We propose a proportional hazard regression model with heterogeneity (frailty or random effect), which is generated by log-logistic distribution. The iterative least square (ILS) approach is adopted to estimate the regression parameters and to predict the realizations of random effects in the frailty model. The adjusted profile hierarchical likelihood is used to estimate the parameter in frailty distribution. We demonstrate via simulation studies that the regression parameter estimates in the log-logistic model are accurate, which is similar to the gamma and lognormal frailty models. We also apply these models to real data as an illustration.

KEYWORDS: iterative least square; frailty model; hierarchical likelihood; clustered survival data

3.1. Introduction

Frailty models have been widely used for the analysis of clustered survival data. The frailty, a common random effect acting multiplicatively on each individual's hazard rate to model the dependence between the survival times. The reason is that the frailty describes the influence of common unknown factors. The introduction of a shared frailty term in each cluster is one way of modeling the dependence of clustered survival times. Such clusters could be, for example, hospitals, families, communities, or treatment centers.

Models with a gamma frailty distribution where the marginal likelihood has a closed form have been frequently used; see, for example.(Androulakis, Koukouvinos et al. 2012, Giussani and Bonetti 2019, Martins, Aerts et al. 2019, Scudilio, Calsavara et al. 2019) The lognormal frailty distributions are also a common choice of frailty modeling due to its natural extension to the multivariate cases.(Xue and Brookmeyer 1996, Wang 2019) The power variance family frailty distribution is a broad class of distributions incorporating the gamma, positive stable, and inverse Gaussian as special cases, and thereby offers a flexible framework for modeling is also proposed in the literature.(Hougaard 1986) The frailty distribution choice is crucial to obtain correct estimates of the dependence structure.(Duchateau and Janssen 2007) However, in many situations, prior information about choosing among the distributions may not be available, and the frailty distribution is neither gamma nor lognormal.

This chapter focuses on multivariate frailty models for clustered data, which are extensions of the Cox proportional hazard model. These models' concept provides a convenient way of introducing unobserved heterogeneity and associations into the Cox model. The shared frailty model, which is a random effect model in survival analysis, is specified by

$$\lambda_{ij}(t_{ij}|u_i, \boldsymbol{x}_{ij}) = \lambda_0(t) u_i \exp(\boldsymbol{x}_{ij}^T \boldsymbol{\beta}), \tag{3.1}$$

where $\lambda_0(t)$ denotes the unspecified baseline hazard function, assumed to be common for all subjects in the study population, $\boldsymbol{x}_{ij} = (x_{ij1}, \ldots, x_{ijp})^T$ is a vector of fixed covariates of subject $j(j = 1, .., n_i)$ in cluster $i(1, \ldots, G)$, T is a transpose, and $\boldsymbol{\beta}$ is a $p \times 1$ vector of unknown regression parameters. The frailty term u_i is assumed to be equal for all individuals in cluster i.

The estimation of parameters in frailty models is often complicated because the marginal likelihood involves an intractable integral. When multiple frailties are involved, the dimensionality of a required integral will be high, and thus numerical integration would not be an ideal estimation method.(Wu and Bentler 2012) Thus we propose h-likelihood for estimating parameters in statistical shared frailty models where the intractable integral is avoided.

The remaining of the chapter is organized as follows. Section 3.2 introduces the logistic frailty model and the h-likelihood estimation process. Section 3.3 presents a simulation study results in which the proposed frailty models' performance is evaluated, and compared to the frequently used gamma frailty model and the lognormal frailty model. Sections 3.4 and 3.5 presents a data example and conclusion, respectively.

3.2. Log-logistic shared frailty model

In frailty models, it is common to specify the distribution of the frailty terms ω_i in equation (3.1). Historically, gamma frailty models dominated the literature because of their mathematical convenience based on the marginal likelihood's explicit form. The gamma frailty, although favorite, may have some drawbacks. That is, it weakens the effect of covariates.(Hougaard 1986) In this case, the convenience, however, may not necessarily assure that the fit is good. To overcome such drawbacks, models with other distributions of the frailties are needed.

We introduce a new class of log-logistic frailty distribution to model multivariate survival data. This is the first attempt such frailty distribution is being used to characterize dependency in correlated survival data to the best of our knowledge. We assume that the failure-time variable T_{ij} corresponding to the j^{th} subject from the ith cluster, C_{ij} is non-informative right-censoring time, independent of T_{ij}, $y_{ij} = \min(T_{ij}, C_{ij})$, and $\delta_{ij} = I(T_{ij} \leq C_{ij})$, where $I(.)$ is the indicator function. The shared frailty model for log-logistic frailty can be written as,

$$\lambda_{ij}(t_{ij}|v_i, \boldsymbol{x}_{ij}) = \lambda_0(t)\exp(\boldsymbol{x}_{ij}^T\boldsymbol{\beta} + v_i), \tag{3.2}$$

where $v_i = \exp(u_i)$ in equation (3.1).

Assume that the unobserved frailties v_i's are independent and identically distributed logistic random variables with mean 0 and unknown frailty θ with probability density function given as follows,

$$f(v_i;\theta) = \frac{\exp\left(-\frac{v_i}{\theta}\right)}{\theta\left(1+\exp\left(-\frac{v_i}{\theta}\right)\right)^2}, \tag{3.3}$$

where $E(V_i) = 0$ and $var(V_i) = \frac{\pi^2\theta^2}{3}$.

The h-likelihood for shared frailty(Ha, Lee et al. 2001, Ha and Lee 2003) is given by

$$h \quad = \sum_{ij} \ell_{1ij} + \sum_{i} \ell_{2i}, \tag{3.4}$$

where ℓ_{1ij} is the logarithm of the conditional likelihood in T_{ij} and δ_{ij} given $V_i = v_i$ with parameters $(\boldsymbol{\beta}, \lambda_0)$ and ℓ_{2i} is the log density function of $V_i = v_i$ with parameter θ. Defining $\eta_{ij} = \boldsymbol{x}_{ij}^T\boldsymbol{\beta} + v_i$, we have,

$$\ell_{1ij} = \delta_{ij}\{\log\lambda_0(y_{ij}) + \eta_{ij}\} - \Lambda_0(y_{ij})\exp(\eta_{ij}),$$

$$\ell_{2i} = -\frac{v_i}{\theta} - \log(\theta) - 2\log\left(1+\exp\left(-\frac{v_i}{\theta}\right)\right),$$

where $\Lambda_0(.) = \int_0^t f(.)dv_i$ is the conditional cumulative baseline hazard function of T_{ij} given $V_i = v_i$.

Suppose that the functional form of $\lambda_0(t)$ is unknown. Furthermore, suppose that the events occur at K distinct ordered event times $T_{(1)}, T_{(2)}, \ldots T_{(K)}$. Given $\boldsymbol{\beta}$, and $\boldsymbol{v} = (v_1, \ldots, v_G)^T$, and θ, the score equations

$$\frac{\partial h}{\partial \lambda_{0k}} = 0, \qquad k = 1, \ldots, K,$$

gives the nonparametric maximum hierarchical likelihood estimator of λ_{0k},

$$\hat{\lambda}_0\big(y_{(k)}\big) = \frac{d_{(k)}}{\sum_{ij\in\mathcal{R}(y_{(k)})} \exp\big(\eta_{ij}\big)}. \tag{3.4}$$

Thus, $\widehat{\Lambda}_0\big(y_{ij}\big) = \sum_{k:y_{(k)}\le t} \hat{\lambda}_0\big(y_{(k)}\big)$, where $d_{(k)}$ is the number of deaths at $y_{(k)}$ and $\mathcal{R}\big(y_{(k)}\big) = \big\{(i,j): y_{ij} \ge T_{(k)}\big\}$ is the risk set at $T_{(k)}$. This estimator is an extension of the estimator(Breslow 1972, Breslow 1974) of the baseline cumulative hazard function for the Cox model to the frailty model. After eliminating the baseline hazard, the kernel of the profile hierarchical likelihood $h^* = h|_{\Lambda_0(t)=\widehat{\Lambda}_0(t)}$ is as follows:

$$h^* \propto \sum_{ij} \delta_{ij}\, \eta_{ij} - \sum_{k:y(k)\le t} d_{(k)} \log\left[\sum_{ij\in\mathcal{R}(y_{(k)})} \exp\big(\eta_{ij}\big)\right] + \sum_i \ell_{2i}. \tag{3.5}$$

3.2.1 Maximum hierarchical likelihood estimator (MHLE) of $\boldsymbol{\tau} = (\boldsymbol{\beta}, \boldsymbol{v})$

Given the frailty parameter θ, the MHLE of $\boldsymbol{\tau} = (\boldsymbol{\beta}, \boldsymbol{v})$ can be obtained by solving the following scores equations,

$$\frac{\partial h^*}{\partial \beta_r} = \frac{\partial}{\partial \beta_r}\left\{\sum_{ij} \delta_{ij}\, \eta_{ij} - \sum_{k:y(k)\le t} d_{(k)} \log\left[\sum_{ij\in\mathcal{R}(y_{(k)})} \exp\big(\eta_{ij}\big)\right] + \sum_i \ell_{2i}\right\}, \qquad r = 1, \ldots, p,$$

$$\frac{\partial h^*}{\partial \beta_r} = \sum_{ij} \delta_{ij}\, \boldsymbol{x}_{ijr} - \sum_k \frac{d_{(k)}}{\sum_{ij\in\mathcal{R}(t_{(k)})} \exp\big(\eta_{ij}\big)} \sum_{ij\in\mathcal{R}(t_{(k)})} \boldsymbol{x}_{ijr} \exp\big(\eta_{ij}\big),$$

$$\frac{\partial h^*}{\partial \beta_r} = \sum_{ij} \delta_{ij}\, \boldsymbol{x}_{ijr} - \widehat{\Lambda}_0(y_{ij}) \sum_{ij \in \mathcal{R}(t_{(k)})} \boldsymbol{x}_{ijr} \exp(\eta_{ij}). \tag{3.6}$$

Similarly,

$$\frac{\partial h^*}{\partial v_i} = \sum_{ij} \delta_{ij} - \widehat{\Lambda}_0(y_{ij}) \sum_{ij \in \mathcal{R}(t_{(k)})} \exp(\eta_{ij}) + \sum_{i} \frac{\partial \ell_{2i}(\theta; v_i)}{\partial v_i}, \tag{3.7}$$

where $\widehat{\Lambda}_0(y_{ij}) = \sum_k \frac{d_{(k)}}{\sum_{ij \in \mathcal{R}(t_{(k)})} \exp(\eta_{ij})}$.

Calculating the estimators will be much easier if matrices are used instead of summations. The following matrices and notations are used for the remainder of this chapter. Let $\boldsymbol{X}$ be a $N \times p$ matrix of p covariates, $\boldsymbol{Z}$ be a $N \times G$ cluster indicator matrix whose ijth row vector is $\boldsymbol{z}_{ij}^T$, where $\boldsymbol{z}_{ij} = (z_{ij1}, z_{ij2}, \dots, z_{ijG})^T$, $\boldsymbol{\delta}$ be a $N \times 1$ vector of δ_{ij}, and $\boldsymbol{\mu}$ be a $N \times 1$ vector with $\widehat{\Lambda}_0(y_{ij}) \exp(\boldsymbol{X}^T\boldsymbol{\beta} + \boldsymbol{Z}^T\boldsymbol{v})$. The vector $\boldsymbol{\mu}$ can be written as a simple form by using a weighted risk indicator matrix $\boldsymbol{M}$, which contains the risk set $R_{(k)}$.(Ha and Lee 2003) Let $\boldsymbol{M} = (R_1, R_2, \dots, R_D)$ be a $N \times D$ at risk indicator matrix where the ijth element is one if $I(y_{ij} \geq y_{(k)})$ and zero otherwise. Define $\boldsymbol{B} = \mathrm{diag}\{\Lambda_0(y_{ij})\}$ as a $D \times D$ diagonal matrix. Let $\boldsymbol{W_1}$ be $N \times N$ diagonal matrix with elements $\exp(\boldsymbol{X}^T\boldsymbol{\beta} + \boldsymbol{Z}^T\boldsymbol{v})$, and let $\boldsymbol{C}$ be a diagonal $D \times D$ matrix where the kth element is $\frac{\left(\hat{\lambda}_0(t_{(k)})\right)^2}{d_{(k)}}$.

The score functions, from equations (3.6) and (3.7) can be written as follows:

$$\frac{\partial h^*}{\partial \boldsymbol{\beta}} = \boldsymbol{X}^T(\boldsymbol{\delta} - \boldsymbol{\mu}), \tag{3.8}$$

$$\frac{\partial h^*}{\partial \boldsymbol{v}} = \boldsymbol{Z}^T(\boldsymbol{\delta} - \boldsymbol{\mu}) + \frac{\partial \ell_{2i}}{\partial \boldsymbol{v}}. \tag{3.9}$$

These are the estimating equations for a Poisson hierarchical generalized linear model, with $\boldsymbol{\delta}$ as the response variable but with the offset $\log\left(\widehat{\Lambda}_0(y_{ij})\right)$. Given the frailty

θ, the joint maximization of for $(\boldsymbol{\beta}, \boldsymbol{v})$ (i.e., $\partial^2 h^* / \partial(\boldsymbol{\beta}, \boldsymbol{v})^2 = 0$) leads to the iterative least squares (ILS) score equations.

The ILS equation for $(\boldsymbol{\beta}, \boldsymbol{v})$ in the log-logistic frailty model is given by

$$\begin{pmatrix} \boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X} & \boldsymbol{X}^T\boldsymbol{W}\boldsymbol{Z} \\ \boldsymbol{Z}^T\boldsymbol{W}\boldsymbol{X} & \boldsymbol{Z}^T\boldsymbol{W}\boldsymbol{Z} + \boldsymbol{Q} \end{pmatrix} \begin{pmatrix} \widehat{\boldsymbol{\beta}} \\ \widehat{\boldsymbol{v}} \end{pmatrix} = \begin{pmatrix} \boldsymbol{X}^T\boldsymbol{w} \\ \boldsymbol{Z}^T\boldsymbol{w} + \boldsymbol{R} \end{pmatrix}, \tag{3.10}$$

where the adjusted dependent variable, $\boldsymbol{w} = \boldsymbol{W}(\boldsymbol{X}^T\boldsymbol{\beta} + \boldsymbol{Z}^T\boldsymbol{v}) + (\boldsymbol{\delta} - \boldsymbol{\mu})$, $\boldsymbol{Q}$ is the $G \times G$ diagonal matrix whose ith element is $-\frac{\partial^2 \ell_{2i}}{\partial v^2}$ and $\boldsymbol{R} = \boldsymbol{Q}\boldsymbol{v} + \frac{\partial \ell_{2i}}{\partial \boldsymbol{v}}$.

The asymptotic covariance matrix for $\hat{\boldsymbol{\tau}} - \boldsymbol{\tau}$ is obtained from $\boldsymbol{H}^{-1}$ where $\boldsymbol{H} = -\partial^2 h^* / \partial(\boldsymbol{\beta}, \boldsymbol{v})^2 = \begin{pmatrix} \boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X} & \boldsymbol{X}^T\boldsymbol{W}\boldsymbol{Z} \\ \boldsymbol{Z}^T\boldsymbol{W}\boldsymbol{X} & \boldsymbol{Z}^T\boldsymbol{W}\boldsymbol{Z} + \boldsymbol{Q} \end{pmatrix}$. So, the upper left-hand corner of $\boldsymbol{H}^{-1}$ gives the asymptotic variance matrix of $\widehat{\boldsymbol{\beta}}$,

$$\mathrm{var}(\widehat{\boldsymbol{\beta}}) = (\boldsymbol{X}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1},$$

where $\boldsymbol{\Sigma} = \boldsymbol{W}^{-1} + \boldsymbol{Z}\boldsymbol{Q}^{-1}\boldsymbol{Z}^T$.

We layout the mathematical derivations of the equation (3.10). The two equations (3.8) and (3.9), can be simply expressed as

$$\frac{\partial h^*}{\partial \boldsymbol{\tau}} = \boldsymbol{E}^T(\boldsymbol{\delta} - \boldsymbol{\mu}) + \boldsymbol{B}, \tag{3.11}$$

where $\boldsymbol{E} = (\boldsymbol{X}, \boldsymbol{Z})$, $\boldsymbol{\tau} = (\boldsymbol{\beta}, \boldsymbol{v})$, and $\boldsymbol{B} = \left(\mathbf{0}^T, \frac{\partial \ell}{\partial v}\right)^T$. Thus, $\boldsymbol{E}\boldsymbol{\tau} = \boldsymbol{\eta} = \boldsymbol{X}^T\boldsymbol{\beta} + \boldsymbol{Z}^T\boldsymbol{v}$.

Next, we calculate the entries of the observed information matrix $\boldsymbol{H}$ of $\boldsymbol{\beta}$ and $\boldsymbol{v}$. Following(Ha and Lee 2003), we define $\boldsymbol{W} = \boldsymbol{W_1}\boldsymbol{B} - (\boldsymbol{W_1}\boldsymbol{M})\boldsymbol{C}(\boldsymbol{W_1}\boldsymbol{M})$.

Next, from (3.8) and (3.9), we have the negative second partial derivatives with respect to $\boldsymbol{\beta}$ and $\boldsymbol{v}$,

$$-\frac{\partial^2 h^*}{\partial \boldsymbol{\beta}^2} = \boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X},$$

$$-\frac{\partial^2 h^*}{\partial \boldsymbol{\beta}\partial \boldsymbol{v}} = \boldsymbol{X}^T\boldsymbol{W}\boldsymbol{Z},$$

$$-\frac{\partial^2 h^*}{\partial \boldsymbol{v}\partial \boldsymbol{\beta}} = \boldsymbol{Z}^T\boldsymbol{W}\boldsymbol{X},$$

$$-\frac{\partial^2 h^*}{\partial \boldsymbol{v}^2} = \boldsymbol{Z}^T\boldsymbol{W}\boldsymbol{Z} + \boldsymbol{Q},$$

leading to

$$\boldsymbol{H} = \begin{pmatrix} -\frac{\partial^2 h^*}{\partial \boldsymbol{\beta}^2} & -\frac{\partial^2 h^*}{\partial \boldsymbol{\beta}\partial \boldsymbol{v}} \\ -\frac{\partial^2 h^*}{\partial \boldsymbol{v}\partial \boldsymbol{\beta}} & -\frac{\partial^2 h^*}{\partial \boldsymbol{v}^2} \end{pmatrix} = \begin{pmatrix} \boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X} & \boldsymbol{X}^T\boldsymbol{W}\boldsymbol{Z} \\ \boldsymbol{Z}^T\boldsymbol{W}\boldsymbol{X} & \boldsymbol{Z}^T\boldsymbol{W}\boldsymbol{Z} + \boldsymbol{Q} \end{pmatrix}. \tag{3.12}$$

The equation (3.12) can be written in a simple form

$$\boldsymbol{H} = \boldsymbol{E}^T\boldsymbol{W}\boldsymbol{E} + \boldsymbol{F}, \tag{3.13}$$

where $\boldsymbol{F} = \boldsymbol{BD}(\boldsymbol{0}, \boldsymbol{Q})$, block diagonal matrix.

From $\hat{\boldsymbol{\tau}} = \boldsymbol{\tau} + \boldsymbol{H}^{-1}\left(\frac{\partial h^*}{\partial \boldsymbol{\tau}}\right)$, (3.11), and (3.13), we obtain

$$(\boldsymbol{E}^T\boldsymbol{W}\boldsymbol{E} + \boldsymbol{F})\hat{\boldsymbol{\tau}} = (\boldsymbol{E}^T\boldsymbol{W}\boldsymbol{E} + \boldsymbol{F})\tau + \boldsymbol{E}^T(\boldsymbol{\delta} - \boldsymbol{\mu}) + \boldsymbol{B},$$

$$= (\boldsymbol{E}^T\boldsymbol{W}\boldsymbol{E})\boldsymbol{\tau} + \boldsymbol{E}^T(\boldsymbol{\delta} - \boldsymbol{\mu}) + \boldsymbol{b},$$

$$\hat{\boldsymbol{\tau}} = \boldsymbol{E}^{\boldsymbol{T}}\boldsymbol{W}\boldsymbol{w} + \boldsymbol{b}, \tag{3.14}$$

where $\boldsymbol{b} = \boldsymbol{F}\boldsymbol{\tau} + \boldsymbol{B}$ and $\boldsymbol{w} = \boldsymbol{\eta} + \boldsymbol{W}^{-\boldsymbol{1}}(\boldsymbol{\delta} - \boldsymbol{\mu})$. This completes the proof of the equation (3.10).

Let $\boldsymbol{P} = \begin{pmatrix} \boldsymbol{X} & \boldsymbol{Z} \\ \boldsymbol{0} & \boldsymbol{I}_G \end{pmatrix}$ and $\boldsymbol{V} = \begin{pmatrix} \boldsymbol{W} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{Q} \end{pmatrix}$, then the ILS equation (3.14) can be written in a new simple matrix form as

$$(\boldsymbol{P}^T\boldsymbol{V}\boldsymbol{P})\hat{\tau} = \boldsymbol{P}^T\boldsymbol{y_0},$$

where $\boldsymbol{y_0} = (\boldsymbol{w}^T, \boldsymbol{R}^T)^T$. Note that $\boldsymbol{H} = -{\partial^2 h^*}/{\partial(\boldsymbol{\beta}, \boldsymbol{v})^2} = \boldsymbol{P}^T\boldsymbol{V}\boldsymbol{P}$.

Mathematical derivations of $\boldsymbol{R} = \boldsymbol{Q}\boldsymbol{v} + \frac{\partial \ell_{2i}}{\partial \boldsymbol{v}}$, are as follows: The logarithm of the density function $f(v_i)$ in equation (3) is

$$\begin{aligned}\ell_2 &= \log f(\boldsymbol{v}; \theta),\\ &= -\frac{\boldsymbol{v}}{\theta} - \log(\theta) - 2\log\left(1 + \exp\left(-\frac{\boldsymbol{v}}{\theta}\right)\right),\end{aligned}$$

Therefore,

$$\begin{aligned}\frac{\partial \ell_2}{\partial \boldsymbol{v}} &= \frac{\partial}{\partial \boldsymbol{v}}\left(-\frac{\boldsymbol{v}}{\theta} - \log(\theta) - 2\log\left(1 + \exp\left(-\frac{\boldsymbol{v}}{\theta}\right)\right)\right),\\ &= -\frac{1}{\theta} - 2\left[\frac{-\frac{1}{\theta}\exp\left(-\frac{\boldsymbol{v}}{\theta}\right)}{1 + \exp\left(-\frac{\boldsymbol{v}}{\theta}\right)}\right],\end{aligned}$$

$$\frac{\partial \ell_2}{\partial \boldsymbol{v}} = -\frac{1}{\theta} + \frac{1}{\theta}2\boldsymbol{\pi_v} = \frac{1}{\theta}[2\boldsymbol{\pi_v} - 1],$$

where $\boldsymbol{\pi_v} = \frac{\exp\left(-\frac{v}{\theta}\right)}{1+\exp\left(-\frac{v}{\theta}\right)}$.

The negative second partial derivative of ℓ_2 is a $G \times G$ diagonal matrix with $\imath$th element,

$$\begin{aligned}\boldsymbol{Q} &= -\frac{\partial^2 \ell_2}{\partial \boldsymbol{v}^2} = -\frac{\partial}{\partial \boldsymbol{v}}\left\{\frac{1}{\theta}[2\boldsymbol{\pi_v} - 1]\right\},\\ &= -\frac{1}{\theta}2\boldsymbol{\pi}'_{\boldsymbol{v}}\,.\end{aligned}$$

But

$$\begin{aligned}\boldsymbol{\pi}'_{\boldsymbol{v}} &= \frac{\partial}{\partial \boldsymbol{v}}\left\{\frac{\exp\left(-\frac{\boldsymbol{v}}{\theta}\right)}{1 + \exp\left(-\frac{\boldsymbol{v}}{\theta}\right)}\right\},\\ &= \frac{-\frac{1}{\theta}\exp\left(-\frac{\boldsymbol{v}}{\theta}\right)\left(1 + \exp\left(-\frac{\boldsymbol{v}}{\theta}\right)\right) - \left(\exp\left(-\frac{\boldsymbol{v}}{\theta}\right)\right)\left(-\frac{1}{\theta}\exp\left(-\frac{\boldsymbol{v}}{\theta}\right)\right)}{\left[1 + \exp\left(-\frac{\boldsymbol{v}}{\theta}\right)\right]^2},\end{aligned}$$

$$= -\frac{1}{\theta}\left[\left[\frac{\exp\left(-\frac{\boldsymbol{v}}{\theta}\right)}{1+\exp\left(-\frac{\boldsymbol{v}}{\theta}\right)}\right] - \left[\frac{\exp\left(-\frac{\boldsymbol{v}}{\theta}\right)}{1+\exp\left(-\frac{\boldsymbol{v}}{\theta}\right)}\right]^2\right],$$

$$\boldsymbol{\pi}'_v = -\frac{1}{\theta}(\boldsymbol{\pi}_v - \boldsymbol{\pi}_v^2).$$

Therefore,

$$\boldsymbol{Q} = -\frac{1}{\theta}2\boldsymbol{\pi}'_v\,,$$

$$\boldsymbol{Q} = \frac{2}{\theta^2}(\boldsymbol{\pi}_v - \boldsymbol{\pi}_v^2).$$

Thus,

$$\boldsymbol{R} = \frac{\boldsymbol{v}}{\theta^2}2[\boldsymbol{\pi}_v - \boldsymbol{\pi}_v^2] + \frac{1}{\theta}[2\boldsymbol{\pi}_v - \mathbf{1}].$$

3.2.2 Maximum hierarchical likelihood estimator (MHLE) of $\boldsymbol{\theta}$

The estimate of the frailty θ is found by maximizingLee and Nelder (1996) adjusted profile h-likelihood(Ha, Lee et al. 2001, Ha and Lee 2003),

$$h_A^* = h^*|_{\boldsymbol{\tau}=\hat{\boldsymbol{\tau}}} + \frac{1}{2}\log\{\det(2\pi\boldsymbol{H}^{-1})\}\Big|_{\boldsymbol{\tau}=\hat{\boldsymbol{\tau}}}. \tag{3.15}$$

The adjusted profile h-likelihood is used to approximate the restricted likelihood of θ that considers the estimation of $\boldsymbol{\beta}$ and $\boldsymbol{v}$. The frailty parameter estimator θ can then be obtained by solving the equation,

$$\frac{\partial h_A^*}{\partial\theta} = 0. \tag{3.16}$$

The gradient vector $\frac{\partial h_A^*}{\partial\theta}$ is

$$\frac{\partial h_A^*}{\partial\theta} = \frac{\partial h^*|_{\boldsymbol{\tau}=\hat{\boldsymbol{\tau}}}}{\partial\theta} - \frac{1}{2}trace\left(\widehat{\boldsymbol{H}}^{-1}\frac{\partial\widehat{\boldsymbol{H}}}{\partial\theta}\right), \tag{3.17}$$

where $\widehat{\boldsymbol{H}} = H|_{\boldsymbol{\tau}=\hat{\boldsymbol{\tau}}}$.

The observed information matrix $-\frac{\partial^2 h_A^*}{\partial\theta^2}$ for the frailty parameter, θ is,

$$\frac{\partial^2 h_A^*}{\partial \theta^2} = \frac{\partial h^*|_{\boldsymbol{\tau}=\hat{\boldsymbol{\tau}}}}{\partial \theta^2} - \frac{\partial}{\partial \theta}\left\{\frac{1}{2} trace\left(\widehat{\boldsymbol{H}}^{-1}\frac{\partial \widehat{\boldsymbol{H}}}{\partial \theta}\right)\right\},$$

$$-\frac{\partial^2 h_A^*}{\partial \theta^2} = -\frac{\partial h^*|_{\boldsymbol{\tau}=\hat{\boldsymbol{\tau}}}}{\partial \theta^2} + \frac{1}{2} trace\left(-\widehat{\boldsymbol{H}}^{-1}\frac{\partial \widehat{\boldsymbol{H}}}{\partial \theta}\widehat{\boldsymbol{H}}^{-1}\frac{\partial \widehat{\boldsymbol{H}}}{\partial \theta} + \widehat{\boldsymbol{H}}^{-1}\frac{\partial^2 \widehat{\boldsymbol{H}}}{\partial \theta^2}\right). \qquad (3.18)$$

Taking partial derivatives of $\frac{\partial h_A^*}{\partial \theta}$ should include $\frac{\partial \widehat{\boldsymbol{\beta}}}{\partial \theta}$ and $\frac{\partial \widehat{\boldsymbol{v}}}{\partial \theta}$. We consider the total derivation of h_A^* since $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{v}}$ are functions of θ. Ignoring $\left(\frac{\partial \widehat{\boldsymbol{\beta}}}{\partial \theta}\right)$ and $\left(\frac{\partial \widehat{\boldsymbol{v}}}{\partial \theta}\right)$ do not work in some cases, such as data with binary covariates and small cluster sizes.(Ha and Lee 2003, Ha, Jeong et al. 2017) However, $\frac{\partial \widehat{\boldsymbol{\beta}}}{\partial \theta}$ can be ignored because there is an indirect dependency between θ and $\widehat{\boldsymbol{\beta}}$ whereas, $\frac{\partial \widehat{\boldsymbol{v}}}{\partial \theta}$ is included because there is a direct dependency between $\widehat{\boldsymbol{v}}$ and θ.

Thus, the total derivative of $\frac{\partial h_A^*}{\partial \theta}$ is

$$\frac{\partial h_A^*}{\partial \theta} = \frac{\partial h_A^*}{\partial \theta} + \left(\frac{\partial h_A^*}{\partial \boldsymbol{\beta}}\bigg|_{\boldsymbol{\beta}=\widehat{\boldsymbol{\beta}}}\right)\frac{\partial \widehat{\boldsymbol{\beta}}}{\partial \theta} + \left(\frac{\partial h_A^*}{\partial \boldsymbol{v}}\bigg|_{\boldsymbol{v}=\widehat{\boldsymbol{v}}}\right)\left(\frac{\partial \widehat{\boldsymbol{v}}}{\partial \theta}\right).$$

The total derivative calculates the derivative of h_A^* with respect to θ where the other arguments h_A^*, $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{v}}$ are allowed to depend on θ; they do not have to remain constant. The total derivation of the first term $\frac{\partial h^*|_{\tau=\hat{\tau}}}{\partial \theta}$ in (3.17) is,

$$\frac{\partial h^*|_{\boldsymbol{\tau}=\hat{\boldsymbol{\tau}}}}{\partial \theta} = \frac{\partial h^*|_{\boldsymbol{\tau}=\hat{\boldsymbol{\tau}}}}{\partial \theta} + \left(\frac{\partial h^*}{\partial \boldsymbol{v}}\bigg|_{\boldsymbol{\tau}=\hat{\boldsymbol{\tau}}}\right)\left(\frac{\partial \widehat{\boldsymbol{v}}}{\partial \theta}\right).$$

$$\frac{\partial h^*|_{\boldsymbol{\tau}=\hat{\boldsymbol{\tau}}}}{\partial \theta} = \frac{\partial h^*|_{\boldsymbol{\tau}=\hat{\boldsymbol{\tau}}}}{\partial \theta}. \qquad (3.19)$$

since $\frac{\partial h^*}{\partial \boldsymbol{v}}\Big|_{\boldsymbol{\tau}=\hat{\boldsymbol{\tau}}} = \mathbf{0}$.

Therefore,

$$\frac{\partial h^*|_{\boldsymbol{\tau}=\hat{\boldsymbol{\tau}}}}{\partial \theta} = \sum_i \frac{\partial \ell_2}{\partial \theta}\bigg|_{\boldsymbol{\tau}=\hat{\boldsymbol{\tau}}} \ ,$$

$$= \sum_i \left[\frac{\partial}{\partial\theta}\left\{-\frac{\hat{\boldsymbol{v}}}{\theta} - \log(\theta) - 2\log\left(1+\exp\left(-\frac{\hat{\boldsymbol{v}}}{\theta}\right)\right)\right\}\right],$$

$$= \sum_i \left[\frac{\hat{\boldsymbol{v}}}{\theta^2} - \frac{1}{\theta} - 2\left(\frac{\frac{\hat{\boldsymbol{v}}}{\theta^2}\exp\left(-\frac{\hat{\boldsymbol{v}}}{\theta}\right)}{1+\exp\left(-\frac{\hat{\boldsymbol{v}}}{\theta}\right)}\right)\right],$$

$$\frac{\partial h^*|_{\boldsymbol{\tau}=\hat{\boldsymbol{\tau}}}}{\partial\theta} = \sum_i \left[\frac{\hat{\boldsymbol{v}}}{\theta^2} - \frac{1}{\theta} - \frac{2\hat{\boldsymbol{v}}}{\theta^2}\hat{\boldsymbol{\pi}}_{\hat{v}}\right].$$

The total derivation of the second term $\frac{\partial\hat{\boldsymbol{H}}}{\partial\theta}$ in (3.17) is more complicated,

$$\frac{\partial\hat{\boldsymbol{H}}}{\partial\theta} = \frac{\partial\hat{\boldsymbol{H}}}{\partial\theta} + \left(\frac{\partial\hat{\boldsymbol{H}}}{\partial\boldsymbol{v}}\right)\left(\frac{\partial\hat{\boldsymbol{v}}}{\partial\theta}\right). \tag{3.20}$$

First, we show how to compute $\frac{\partial\hat{v}}{\partial\theta}$ following Lee et al. (2006). From h^* given θ, let $\hat{\boldsymbol{v}}(\theta)$ be the solution to $g(\theta) = \frac{\partial h^*}{\partial\boldsymbol{v}}\Big|_{\boldsymbol{\tau}=\hat{\boldsymbol{\tau}}} = \boldsymbol{0}$. Then,

$$\frac{\partial g(\theta)}{\partial\theta} = \frac{\partial^2 h^*}{\partial\boldsymbol{v}\partial\theta}\bigg|_{\boldsymbol{\tau}=\hat{\boldsymbol{\tau}}} + \left(\frac{\partial^2 h^*}{\partial\boldsymbol{v}^2}\bigg|_{\boldsymbol{\tau}=\hat{\boldsymbol{\tau}}}\right)\left(\frac{\partial\hat{\boldsymbol{v}}}{\partial\theta}\right) = \boldsymbol{0}.$$

Solving for $\left(\frac{\partial\hat{v}}{\partial\theta}\right)$ gives,

$$\left(\frac{\partial\hat{\boldsymbol{v}}}{\partial\theta}\right) = \left(-\frac{\partial^2 h^*}{\partial\boldsymbol{v}^2}\bigg|_{\boldsymbol{\tau}=\hat{\boldsymbol{\tau}}}\right)^{-1}\left(\frac{\partial^2 h^*}{\partial\boldsymbol{v}\partial\theta}\bigg|_{\boldsymbol{\tau}=\hat{\boldsymbol{\tau}}}\right),$$

$$= \left(\boldsymbol{Z}^T\hat{\boldsymbol{W}}\boldsymbol{Z} + \hat{\boldsymbol{Q}}\right)^{-\boldsymbol{1}}\left(\frac{\partial^2 h^*}{\partial\boldsymbol{v}\partial\theta}\bigg|_{\boldsymbol{\tau}=\hat{\boldsymbol{\tau}}}\right),$$

where $\hat{\boldsymbol{W}} = \boldsymbol{W}|_{\boldsymbol{\tau}=\hat{\boldsymbol{\tau}}}$ and $\hat{\boldsymbol{Q}} = \boldsymbol{Q}|_{\boldsymbol{\tau}=\hat{\boldsymbol{\tau}}}$.

But,

$$\frac{\partial h^*}{\partial\theta} = \frac{\partial l_2}{\partial\theta} = \left(\frac{\boldsymbol{v}}{\theta^2} - \frac{1}{\theta} - \frac{2\boldsymbol{v}}{\theta^2}\boldsymbol{\pi}_{\boldsymbol{v}}\right)\bigg|_{\boldsymbol{\tau}=\hat{\boldsymbol{\tau}}}.$$

Therefore,

$$\widehat{\boldsymbol{K}}' = \frac{\partial^2 h^*}{\partial \boldsymbol{v} \partial \theta}\bigg|_{\boldsymbol{\tau}=\hat{\boldsymbol{\tau}}} = \frac{\partial}{\partial \boldsymbol{v}}\left\{\left(\frac{\boldsymbol{v}}{\theta^2} - \frac{1}{\theta} - \frac{2\boldsymbol{v}}{\theta^2}\boldsymbol{\pi}_{\boldsymbol{v}}\right)\bigg|_{\boldsymbol{\tau}=\hat{\boldsymbol{\tau}}}\right\},$$

$$= \frac{1}{\theta^2} - \frac{2}{\theta^2}\widehat{\boldsymbol{\pi}}_{\hat{\boldsymbol{v}}} - \frac{2\widehat{\boldsymbol{v}}}{\theta^2}\widehat{\boldsymbol{\pi}}'_{\hat{\boldsymbol{v}}},$$

$$\widehat{\boldsymbol{K}}' = \frac{1}{\theta^2} - \frac{2}{\theta^2}\widehat{\boldsymbol{\pi}}_{\hat{\boldsymbol{v}}} + \frac{2}{\theta}\widehat{\boldsymbol{\pi}}'_{\theta}.$$

where

$$\widehat{\boldsymbol{\pi}}'_{\theta} = \frac{\partial}{\partial \theta}\boldsymbol{\pi}_{\boldsymbol{v}}|_{\boldsymbol{\tau}=\hat{\boldsymbol{\tau}}} = \frac{\partial}{\partial \theta}\left\{\frac{\exp\left(-\frac{\widehat{\boldsymbol{v}}}{\theta}\right)}{1+\exp\left(-\frac{\widehat{\boldsymbol{v}}}{\theta}\right)}\right\},$$

$$= \frac{\left(\frac{\widehat{\boldsymbol{v}}}{\theta^2}\exp\left(-\frac{\widehat{\boldsymbol{v}}}{\theta}\right)\right)\left(1+\exp\left(-\frac{\widehat{\boldsymbol{v}}}{\theta}\right)\right)}{\left(1+\exp\left(-\frac{\widehat{\boldsymbol{v}}}{\theta}\right)\right)^2} - \frac{\frac{\widehat{\boldsymbol{v}}}{\theta^2}\exp\left(-\frac{\widehat{\boldsymbol{v}}}{\theta}\right)\left(\exp\left(-\frac{\widehat{\boldsymbol{v}}}{\theta}\right)\right)}{\left(1+\exp\left(-\frac{\widehat{\boldsymbol{v}}}{\theta}\right)\right)^2},$$

$$\widehat{\boldsymbol{\pi}}'_{\theta} = \frac{\widehat{\boldsymbol{v}}}{\theta^2}\left[\widehat{\boldsymbol{\pi}}_{\hat{\boldsymbol{v}}} - \widehat{\boldsymbol{\pi}}^{\mathbf{2}}_{\hat{\boldsymbol{v}}}\right].$$

Hence

$$\left(\frac{\partial \hat{v}}{\partial \theta}\right) = \left(\boldsymbol{Z}^T\widehat{\boldsymbol{W}}\boldsymbol{Z} + \widehat{\boldsymbol{Q}}\right)^{-1}\widehat{\boldsymbol{K}}'. \tag{3.21}$$

Now since $\boldsymbol{X}$ and $\boldsymbol{Z}$ are constant matrices that have no dependence on θ it follows that the total derivative $\frac{\partial \widehat{\boldsymbol{H}}}{\partial \theta}$ is,

$$\frac{\partial \widehat{\boldsymbol{H}}}{\partial \theta} = \begin{pmatrix} \boldsymbol{X}^T\widehat{\boldsymbol{W}}\boldsymbol{X} & \boldsymbol{X}^T\widehat{\boldsymbol{W}}\boldsymbol{Z} \\ \boldsymbol{Z}^T\widehat{\boldsymbol{W}}\boldsymbol{X} & \boldsymbol{Z}^T\widehat{\boldsymbol{W}}\boldsymbol{Z} + \widehat{\boldsymbol{Q}}' \end{pmatrix}, \tag{3.22}$$

where $\widehat{\boldsymbol{W}}' = \frac{\partial \widehat{\boldsymbol{W}}}{\partial \theta}$ and $\widehat{\boldsymbol{Q}}' = \frac{\partial \widehat{\boldsymbol{Q}}}{\partial \theta}$. Since $\widehat{\boldsymbol{W}}$ does not depend on θ, it follows that the total derivative is,

$$\widehat{\boldsymbol{W}}' = \frac{\partial \boldsymbol{W}}{\partial \theta}\bigg|_{\boldsymbol{\tau}=\hat{\boldsymbol{\tau}}} + \left(\frac{\partial \boldsymbol{W}}{\partial \boldsymbol{v}}\bigg|_{\boldsymbol{\tau}=\hat{\boldsymbol{\tau}}}\right)\left(\frac{\partial \widehat{\boldsymbol{v}}}{\partial \theta}\right) = \left(\frac{\partial \boldsymbol{W}}{\partial \boldsymbol{v}}\bigg|_{\boldsymbol{\tau}=\hat{\boldsymbol{\tau}}}\right)\left(\frac{\partial \widehat{\boldsymbol{v}}}{\partial \theta}\right). \tag{3.23}$$

Finally, we calculate the total derivative for $\boldsymbol{Q}$ since it depends on $\boldsymbol{v}$. Thus,

$$\widehat{\boldsymbol{Q}}' = \frac{\partial \widehat{\boldsymbol{Q}}}{\partial \theta} + \frac{\partial \widehat{\boldsymbol{Q}}}{\partial \boldsymbol{v}}\left(\frac{\partial \widehat{\boldsymbol{v}}}{\partial \theta}\right),$$

$$\widehat{\boldsymbol{Q}}' = \widehat{\boldsymbol{Q}}'_{\theta} + \widehat{\boldsymbol{Q}}'_{\widehat{\boldsymbol{v}}}\left(\frac{\partial \widehat{\boldsymbol{v}}}{\partial \theta}\right).$$

The derivation of $\frac{\partial \widehat{\boldsymbol{Q}}}{\partial \theta}$ is,

$$\widehat{\boldsymbol{Q}}'_{\theta} = \frac{\partial}{\partial \theta}\left\{\frac{2}{\theta^2}\left(\widehat{\boldsymbol{\pi}}_{\widehat{\boldsymbol{v}}} - \widehat{\boldsymbol{\pi}}^{\boldsymbol{2}}_{\widehat{\boldsymbol{v}}}\right)\right\},$$

$$\widehat{\boldsymbol{Q}}'_{\theta} = -\frac{4}{\theta^3}\left(\widehat{\boldsymbol{\pi}}_{\widehat{\boldsymbol{v}}} - \widehat{\boldsymbol{\pi}}^{\boldsymbol{2}}_{\widehat{\boldsymbol{v}}}\right) + \frac{2}{\theta^2}\left(\widehat{\boldsymbol{\pi}}'_{\theta} - 2\widehat{\boldsymbol{\pi}}_{\widehat{\boldsymbol{v}}} * \widehat{\boldsymbol{\pi}}'_{\theta}\right).$$

Also, the derivation of $\frac{\partial \widehat{\boldsymbol{Q}}}{\partial \boldsymbol{v}}$ is,

$$\frac{\partial \widehat{\boldsymbol{Q}}}{\partial \boldsymbol{v}} = \frac{\partial}{\partial \boldsymbol{v}}\left\{\left(\frac{2}{\theta^2}\left(\boldsymbol{\pi}_{v} - \boldsymbol{\pi}^{2}_{v}\right)\Big|_{\boldsymbol{\tau}=\widehat{\boldsymbol{\tau}}}\right)\right\},$$

$$\widehat{\boldsymbol{Q}}'_{\widehat{\boldsymbol{v}}} = \frac{2}{\theta^2}\left(\widehat{\boldsymbol{\pi}}'_{\widehat{\boldsymbol{v}}} - 2\widehat{\boldsymbol{\pi}}_{\widehat{\boldsymbol{v}}} * \widehat{\boldsymbol{\pi}}'_{\widehat{\boldsymbol{v}}}\right).$$

The next step is to calculate the terms in the observed information (3.18). First, we compute the total derivative of the first term in (3.18).

$$-\frac{\partial h^*|_{\boldsymbol{\tau}=\widehat{\boldsymbol{\tau}}}}{\partial \theta^2} = \sum_i -\frac{\partial h^*|_{\boldsymbol{\tau}=\widehat{\boldsymbol{\tau}}}}{\partial \theta^2} - \sum_i\left[\left(\frac{\partial^2 h^*|_{\boldsymbol{\tau}=\widehat{\boldsymbol{\tau}}}}{\partial \boldsymbol{v}\partial \theta}\right)\left(\frac{\partial \widehat{\boldsymbol{v}}}{\partial \theta}\right)\right],$$

$$= -\sum_i \frac{\partial}{\partial \theta}\left[\frac{\widehat{\boldsymbol{v}}}{\theta^2} - \frac{1}{\theta} - \frac{2\widehat{\boldsymbol{v}}}{\theta^2}\widehat{\boldsymbol{\pi}}_{\widehat{\boldsymbol{v}}}\right] - \sum_i \frac{\partial}{\partial \boldsymbol{v}}\left[\left(\frac{\widehat{\boldsymbol{v}}}{\theta^2} - \frac{1}{\theta} - \frac{2\widehat{\boldsymbol{v}}}{\theta^2}\widehat{\boldsymbol{\pi}}_{\widehat{\boldsymbol{v}}}\right)\left(\frac{\partial \widehat{\boldsymbol{v}}}{\partial \theta}\right)\right],$$

$$-\frac{\partial h^*|_{\boldsymbol{\tau}=\widehat{\boldsymbol{\tau}}}}{\partial \theta^2} = \sum_i\left[\frac{2\widehat{\boldsymbol{v}}}{\theta^3} - \frac{1}{\theta^2} - \frac{4\widehat{\boldsymbol{v}}}{\theta^3}\widehat{\boldsymbol{\pi}}_{\widehat{\boldsymbol{v}}} + \frac{2\widehat{\boldsymbol{v}}}{\theta^2}\widehat{\boldsymbol{\pi}}'_{\boldsymbol{\theta}}\right] - \sum_i\left[\left(\frac{1}{\theta^2} - \frac{2}{\theta^2}\widehat{\boldsymbol{\pi}}_{\widehat{\boldsymbol{v}}} - \frac{2\widehat{\boldsymbol{v}}}{\theta^2}\widehat{\boldsymbol{\pi}}'_{\widehat{\boldsymbol{v}}}\right)\left(\frac{\partial \widehat{\boldsymbol{v}}}{\partial \theta}\right)\right].$$

The last term needed to calculate in (3.18) is,

$$\frac{\partial^2 \widehat{\boldsymbol{H}}}{\partial \theta^2} = \begin{pmatrix} \boldsymbol{X}^T\widehat{\boldsymbol{W}}''\boldsymbol{X} & \boldsymbol{X}^T\widehat{\boldsymbol{W}}''\boldsymbol{Z} \\ \boldsymbol{Z}^T\widehat{\boldsymbol{W}}''\boldsymbol{X} & \boldsymbol{Z}^T\widehat{\boldsymbol{W}}''\boldsymbol{Z} + \widehat{\boldsymbol{Q}}'' \end{pmatrix}, \tag{3.24}$$

where $\widehat{\boldsymbol{Q}}'' = \frac{\partial^2 \boldsymbol{Q}}{\partial \theta^2}$, $\widehat{\boldsymbol{W}}'' = \frac{\partial^2 \widehat{\boldsymbol{W}}}{\partial \theta^2}$.

The total derivative of $\widehat{\boldsymbol{W}}'' = \frac{\partial^2 \widehat{\boldsymbol{W}}}{\partial \theta^2}$ is

$$\frac{\partial^2 \widehat{\boldsymbol{W}}}{\partial \theta^2} = \left[\left(\frac{\partial \widehat{\boldsymbol{W}}}{\partial \boldsymbol{v}^2}\bigg|_{\boldsymbol{\tau}=\hat{\boldsymbol{\tau}}}\right)\frac{\partial \widehat{\boldsymbol{v}}}{\partial \theta}\right]\frac{\partial \widehat{\boldsymbol{v}}}{\partial \theta} + \left(\frac{\partial \widehat{\boldsymbol{W}}}{\partial \boldsymbol{v}}\bigg|_{\boldsymbol{\tau}=\hat{\boldsymbol{\tau}}}\right)\frac{\partial^2 \widehat{\boldsymbol{v}}}{\partial \theta^2}, \tag{3.25}$$

The $\frac{\partial^2 \widehat{\boldsymbol{W}}}{\partial \boldsymbol{v}^2}$ is found by twice differentiating $\boldsymbol{W} = \boldsymbol{W_1 B} - (\boldsymbol{W_1 M})\boldsymbol{C}(\boldsymbol{W_1 M})$ with respect to $\boldsymbol{v}$.

The second derivation of $\frac{\partial^2 \widehat{\boldsymbol{v}}}{\partial \theta^2}$ is

$$\frac{\partial^2 \widehat{\boldsymbol{v}}}{\partial \theta^2} = -\left(\boldsymbol{Z}^T \widehat{\boldsymbol{W}} \boldsymbol{Z} + \widehat{\boldsymbol{Q}}\right)^{-1}\left(\boldsymbol{Z}^T \widehat{\boldsymbol{W}}' \boldsymbol{Z} + \widehat{\boldsymbol{Q}}'\right)\left(\boldsymbol{Z}^T \widehat{\boldsymbol{W}} \boldsymbol{Z} + \widehat{\boldsymbol{Q}}\right)^{-1}\left(\frac{\partial^2 h^*}{\partial \boldsymbol{v} \partial \theta}\bigg|_{\boldsymbol{\tau}=\hat{\boldsymbol{\tau}}}\right)$$

$$+ \left(\boldsymbol{Z}^T \widehat{\boldsymbol{W}} \boldsymbol{Z} + \widehat{\boldsymbol{Q}}\right)^{-\boldsymbol{1}}\left(\frac{\partial^3 h^*}{\partial \boldsymbol{v} \partial \theta^2}\bigg|_{\boldsymbol{\tau}=\hat{\boldsymbol{\tau}}}\right).$$

Let

$$\widehat{\boldsymbol{K}}'' = \frac{\partial^3 h^*}{\partial \boldsymbol{v} \partial \theta^2}\bigg|_{\boldsymbol{\tau}=\hat{\boldsymbol{\tau}}} = \frac{\partial}{\partial \theta}\widehat{\boldsymbol{K}}' = \frac{\partial}{\partial \theta}\left\{\frac{1}{\theta^2} - \frac{2}{\theta^2}\widehat{\boldsymbol{\pi}}_{\widehat{\boldsymbol{v}}} + \frac{2}{\theta}\widehat{\boldsymbol{\pi}}'_{\theta}\right\},$$

$$\widehat{\boldsymbol{K}}'' = -\frac{2}{\theta^3} + \frac{4}{\theta^3}\widehat{\boldsymbol{\pi}}_{\widehat{\boldsymbol{v}}} - \frac{2}{\theta^2}\widehat{\boldsymbol{\pi}}'_{\theta} - \frac{2}{\theta^2}\widehat{\boldsymbol{\pi}}'_{\theta} + \frac{2}{\theta}\widehat{\boldsymbol{\pi}}''_{\theta},$$

$$\widehat{\boldsymbol{K}}'' = -\frac{2}{\theta^3} + \frac{4}{\theta^3}\widehat{\boldsymbol{\pi}}_{\widehat{\boldsymbol{v}}} - \frac{4}{\theta^2}\widehat{\boldsymbol{\pi}}'_{\theta} + \frac{2}{\theta}\widehat{\boldsymbol{\pi}}''_{\theta}.$$

Therefore

$$\frac{\partial^2 \widehat{\boldsymbol{v}}}{\partial \theta^2} = -\left(\boldsymbol{Z}^T \widehat{\boldsymbol{W}} \boldsymbol{Z} + \boldsymbol{Q}\right)^{-1}\left(\boldsymbol{Z}^T \widehat{\boldsymbol{W}}' \boldsymbol{Z} + \boldsymbol{Q}'\right)\left(\boldsymbol{Z}^T \widehat{\boldsymbol{W}} \boldsymbol{Z} + \boldsymbol{Q}\right)^{-1} K' + (\boldsymbol{Z}^T \boldsymbol{W} \boldsymbol{Z} + \boldsymbol{Q})^{-\boldsymbol{1}} \widehat{\boldsymbol{K}}''.$$

The total derivation of $\widehat{\boldsymbol{Q}}'' = \frac{\partial^2}{\partial \theta^2}\boldsymbol{Q}|_{\boldsymbol{\tau}=\hat{\boldsymbol{\tau}}}$ in (3.24) is

$$\widehat{\boldsymbol{Q}}'' = \frac{\partial}{\partial \theta}\left\{\frac{\partial \widehat{\boldsymbol{Q}}}{\partial \theta} + \frac{\partial \widehat{\boldsymbol{Q}}}{\partial \boldsymbol{v}}\left(\frac{\partial \widehat{\boldsymbol{v}}}{\partial \theta}\right)\right\} + \left(\frac{\partial}{\partial \boldsymbol{v}}\left\{\frac{\partial \widehat{\boldsymbol{Q}}}{\partial \theta} + \frac{\partial \widehat{\boldsymbol{Q}}}{\partial \boldsymbol{v}}\left(\frac{\partial \widehat{\boldsymbol{v}}}{\partial \theta}\right)\right\}\right)\left(\frac{\partial \widehat{\boldsymbol{v}}}{\partial \theta}\right),$$

$$= \frac{\partial^2 \widehat{\boldsymbol{Q}}}{\partial \theta^2} + \frac{\partial^2 \widehat{\boldsymbol{Q}}}{\partial \theta \partial \boldsymbol{v}}\left(\frac{\partial \widehat{\boldsymbol{v}}}{\partial \theta}\right) + \frac{\partial \widehat{\boldsymbol{Q}}}{\partial \boldsymbol{v}}\left(\frac{\partial^2 \widehat{\boldsymbol{v}}}{\partial \theta^2}\right) + \frac{\partial^2 \widehat{\boldsymbol{Q}}}{\partial \boldsymbol{v} \partial \theta}\left(\frac{\partial \widehat{\boldsymbol{v}}}{\partial \theta}\right) + \frac{\partial^2 \widehat{\boldsymbol{Q}}}{\partial \boldsymbol{v}^2}\left(\frac{\partial \widehat{\boldsymbol{v}}}{\partial \theta}\right)^2,$$

$$= \widehat{\boldsymbol{Q}}''_{\theta} + \widehat{\boldsymbol{Q}}''_{\hat{v}\theta}\left(\frac{\partial \widehat{\boldsymbol{v}}}{\partial \theta}\right) + \widehat{\boldsymbol{Q}}'_{\hat{v}}\left(\frac{\partial^2 \widehat{\boldsymbol{v}}}{\partial \theta^2}\right) + \widehat{\boldsymbol{Q}}''_{\hat{v}\theta}\left(\frac{\partial \widehat{\boldsymbol{v}}}{\partial \theta}\right) + \widehat{\boldsymbol{Q}}''_{\hat{v}}\left(\frac{\partial \widehat{\boldsymbol{v}}}{\partial \theta}\right)^2,$$

$$\widehat{\boldsymbol{Q}}'' = \widehat{\boldsymbol{Q}}''_{\theta} + 2 * \widehat{\boldsymbol{Q}}''_{\hat{v}\theta}\left(\frac{\partial \widehat{\boldsymbol{v}}}{\partial \theta}\right) + \widehat{\boldsymbol{Q}}'_{\hat{v}}\left(\frac{\partial^2 \widehat{\boldsymbol{v}}}{\partial \theta^2}\right) + \widehat{\boldsymbol{Q}}''_{\hat{v}}\left(\frac{\partial \widehat{\boldsymbol{v}}}{\partial \theta}\right)^2.$$

The calculation of $\widehat{\boldsymbol{Q}}''_{\theta}$ is

$$\widehat{\boldsymbol{Q}}''_{\theta} = \frac{\partial}{\partial \theta}\left\{-\frac{4}{\theta^3}\left(\widehat{\boldsymbol{\pi}}_{\hat{v}} - \widehat{\boldsymbol{\pi}}^{\mathbf{2}}_{\hat{v}}\right) + \frac{2}{\theta^2}\left(\widehat{\boldsymbol{\pi}}'_{\theta} - 2\widehat{\boldsymbol{\pi}}_{\hat{v}} * \widehat{\boldsymbol{\pi}}'_{\theta}\right)\right\},$$

$$= \frac{12}{\theta^4}\left(\widehat{\boldsymbol{\pi}}_{\hat{v}} - \widehat{\boldsymbol{\pi}}^{\mathbf{2}}_{\hat{v}}\right) - \frac{4}{\theta^3}\left(\widehat{\boldsymbol{\pi}}'_{\theta} - 2\widehat{\boldsymbol{\pi}}_{\hat{v}} * \widehat{\boldsymbol{\pi}}'_{\theta}\right) - \frac{4}{\theta^3}\left(\widehat{\boldsymbol{\pi}}'_{\theta} - 2\widehat{\boldsymbol{\pi}}_{\hat{v}} * \widehat{\boldsymbol{\pi}}'_{\theta}\right)$$
$$+ \frac{2}{\theta^2}\left(\widehat{\boldsymbol{\pi}}''_{\theta} - 2[\widehat{\boldsymbol{\pi}}'_{\theta} \times \widehat{\boldsymbol{\pi}}'_{\theta} + \hat{\pi}_{\hat{v}} \times \widehat{\boldsymbol{\pi}}''_{\theta})]\right),$$

$$= \frac{12}{\theta^4}\left(\widehat{\boldsymbol{\pi}}_{\hat{v}} - \widehat{\boldsymbol{\pi}}^{\mathbf{2}}_{\hat{v}}\right) - \frac{8}{\theta^3}\left(\widehat{\boldsymbol{\pi}}'_{\theta} - 2\widehat{\boldsymbol{\pi}}_{\hat{v}} * \widehat{\boldsymbol{\pi}}'_{\theta}\right) + \frac{2}{\theta^2}\left(\widehat{\boldsymbol{\pi}}''_{\theta} - 2[(\widehat{\boldsymbol{\pi}}'_{\theta})^2 + \hat{\pi}_{\hat{v}} \times \widehat{\boldsymbol{\pi}}''_{\theta})]\right),$$

where

$$\widehat{\boldsymbol{\pi}}''_{\theta} = \frac{\partial}{\partial \theta}\left\{\frac{\boldsymbol{v}}{\theta^2}\left(\widehat{\boldsymbol{\pi}}_{\hat{v}} - \widehat{\boldsymbol{\pi}}^{\mathbf{2}}_{\hat{v}}\right)\right\},$$

$$\widehat{\boldsymbol{\pi}}''_{\theta} = -\frac{2\boldsymbol{v}}{\theta^3}\left(\widehat{\boldsymbol{\pi}}_{\hat{v}} - \widehat{\boldsymbol{\pi}}^{\mathbf{2}}_{\hat{v}}\right) + \frac{\boldsymbol{v}}{\theta^2}\left(\widehat{\boldsymbol{\pi}}'_{\theta} - 2\widehat{\boldsymbol{\pi}}_{\hat{v}} * \widehat{\boldsymbol{\pi}}'_{\theta}\right).$$

The calculation of $\widehat{\boldsymbol{Q}}''_{\hat{v}\theta}$ is

$$\widehat{\boldsymbol{Q}}''_{\hat{v}\theta} = \frac{\partial}{\partial \theta}\left\{\frac{2}{\theta^2}\left(\widehat{\boldsymbol{\pi}}'_{\hat{v}} - 2\widehat{\boldsymbol{\pi}}_{\hat{v}} * \widehat{\boldsymbol{\pi}}'_{\hat{v}}\right)\right\},$$

$$\widehat{\boldsymbol{Q}}''_{v\theta} = -\frac{4}{\theta^3}\left(\widehat{\boldsymbol{\pi}}'_{\hat{v}} - \mathbf{2}\widehat{\boldsymbol{\pi}}_{\hat{v}} * \widehat{\boldsymbol{\pi}}'_{\hat{v}}\right) + \frac{2}{\theta^2}\left(\widehat{\boldsymbol{\pi}}''_{\hat{v}\boldsymbol{\theta}} - \mathbf{2}[\widehat{\boldsymbol{\pi}}_{\boldsymbol{\theta}} * \widehat{\boldsymbol{\pi}}'_{\hat{v}} + \widehat{\boldsymbol{\pi}}_{\hat{v}} * \widehat{\boldsymbol{\pi}}''_{\hat{v}\boldsymbol{\theta}}]\right),$$

where

$$\widehat{\boldsymbol{\pi}}''_{\hat{v}\boldsymbol{\theta}} = \frac{\partial}{\partial \theta}\left\{\left(-\frac{1}{\theta}\left(\boldsymbol{\pi}_{\boldsymbol{v}} - \boldsymbol{\pi}^2_{\boldsymbol{v}}\right)\Big|_{\boldsymbol{\tau}=\hat{\boldsymbol{\tau}}}\right)\right\},$$

$$= \frac{1}{\theta^2}\left(\widehat{\boldsymbol{\pi}}_{\hat{v}} - \widehat{\boldsymbol{\pi}}^{\mathbf{2}}_{\hat{v}}\right) - \frac{1}{\theta}\left(\widehat{\boldsymbol{\pi}}'_{\boldsymbol{\theta}} - \mathbf{2}\widehat{\boldsymbol{\pi}}_{\hat{v}} \times \widehat{\boldsymbol{\pi}}'_{\boldsymbol{\theta}}\right).$$

The calculation of $\widehat{\boldsymbol{Q}}''_{\hat{v}}$ **is**

$$\widehat{\boldsymbol{Q}}''_{\hat{v}} = \frac{\partial}{\partial v}\left\{\frac{2}{\theta^2}(\widehat{\boldsymbol{\pi}}'_{\hat{v}} - 2\widehat{\boldsymbol{\pi}}_{\hat{v}} * \widehat{\boldsymbol{\pi}}'_{\hat{v}})\right\}$$

$$= \frac{2}{\theta^2}(\widehat{\boldsymbol{\pi}}''_{\hat{v}} - 2[\widehat{\boldsymbol{\pi}}'_{\hat{v}} * \widehat{\boldsymbol{\pi}}'_{\hat{v}} + \widehat{\boldsymbol{\pi}}_{\hat{v}} * \widehat{\boldsymbol{\pi}}''_{\hat{v}}])$$

$$= \frac{2}{\theta^2}\left(\widehat{\boldsymbol{\pi}}''_{\hat{v}} - 2\left[(\widehat{\boldsymbol{\pi}}'_{\hat{v}})^{\mathbf{2}} + \widehat{\boldsymbol{\pi}}_{\hat{v}} * \widehat{\boldsymbol{\pi}}''_{\hat{v}}\right]\right)$$

where

$$\widehat{\boldsymbol{\pi}}''_{\hat{v}} = \frac{\partial}{\partial v}\left\{\left(-\frac{1}{\theta}(\boldsymbol{\pi}_{\boldsymbol{v}} - \boldsymbol{\pi}^2_{\boldsymbol{v}})\Big|_{\boldsymbol{\tau}=\hat{\boldsymbol{\tau}}}\right)\right\},$$

$$\widehat{\boldsymbol{\pi}}''_{\hat{v}} = -\frac{1}{\theta}(\widehat{\boldsymbol{\pi}}'_{\hat{v}} - 2\widehat{\boldsymbol{\pi}}_{\hat{v}} * \widehat{\boldsymbol{\pi}}'_{\hat{v}}).$$

Note that the $var(\hat{\theta})$ is obtained from the inverse of $-\frac{\partial^2 h^*_A}{\partial \theta^2}$.

3.3. Simulation studies

We conducted simulation studies to evaluate the proposed frailty model's finite sample performance and compare it to two widely used frailty models (gamma frailty model and lognormal frailty models). Thus, three simulation studies are presented. We assume the frailty model is as follows:

$$\lambda_{ij} = \lambda_0(t)u_i \exp(\beta x_{ij}) = \lambda_0(t)\exp(\beta x_{ij} + v_i), \tag{3.26}$$

where $i = 1,..,G$; $j = 1,2,\dots,n_i$, G is the cluster size, and n_i is the number of individuals in each cluster. The binary independent variable x_{ij} is generated from a Bernoulli distribution with success probability 0.5. Given the v_i's, the corresponding censoring times C_{ij} are generated from a uniform distribution $U(0, l)$ with parameter l empirically determined to achieve approximately the right censoring rate of 20%. We use a sample size $N = G * n_i$ where $N = 60$ with $(G;\ n_i) = (30;\ 2)$, $N = 120$ with $(G;\ n_i) = (30;\ 4)$, $N = 160$ with $(G;\ n_i) = (80;\ 2)$ and $N = 120$ with $(G;\ n_i) = (30;\ 4)$. From 200 replications of simulated data, we compute the mean, the mean squared error for $\hat{\beta}$. The mean squared error is given by $\left\{\frac{1}{200}\sum_{i=1}^{200}(\hat{\beta}_i - \beta)^2\right\}$. Also, we calculate the empirical coverage probability for a nominal 95% confidence interval for β. For the frailty parameter θ; the mean and mean squared error for $\hat{\theta}$ are also given. For the computation, we used R software Version 4.0.2.

3.3.1 Data from the log-logistic frailty model

We assume that the random effects $v_i, i = 1,..,G,$ are generated from a logistic with mean 0 and frailty parameter with the probability density function

$$f(v_i;\theta) = \frac{\exp\left(-\frac{v_i}{\theta}\right)}{\theta\left(1+\exp\left(-\frac{v_i}{\theta}\right)\right)^2} \tag{3.27}$$

Let the standard deviation of random effect be 1.0, then the true value of parameter θ is $\frac{\sqrt{3}}{\pi} = 0.5513289$.

3.3.2 Data from the lognormal frailty model

Assume that v_i comes from the lognormal distribution $LN\ (0,\theta)$, then v_i comes from the normal distribution $N(0,\theta)$. We set $\theta = 1.0$, and other characters of the data generation process are the same as before.

$$f(v_i;\theta) = \frac{1}{\sqrt{2\pi\theta}}\exp\left(-\frac{v_i^2}{\theta}\right). \tag{2.28}$$

3.3.3 Data from the gamma frailty model

Assume that the random effect comes from the gamma probability density

$$f(u_i;\theta) = \frac{1}{\Gamma(1/\theta)\theta^{\frac{1}{\theta}}}u_i^{\frac{1}{\theta}-1}\exp\left(-\frac{u_i}{\theta}\right). \tag{3.29}$$

with $E(u_i) = 1$ and $var(u_i) = \theta$, then $v_i = log(u_i)$ with $var\{log(u_i)\} = \psi^{(1)}(1/\theta)$, where ψ is digamma the function, and $\psi^{(1)}$ is the trigamma function. We want $\sqrt{\psi^{(1)}(1/\theta)} = 1.0$, thus, we set $\theta = 0.70113689$, which means that the standard deviation of the random effects of different data generation processes are all 1.0. Other characteristics of the data generation process are the same as the log-logistic frailty model. The FriailtyHL R-package(Ha et al., 2012) was used for estimating parameters in the gamma frailty model and lognormal frailty models.

We estimated the log-logistic frailty model, the lognormal model, and the gamma frailty model simultaneously for the three generated data. The results are summarized in Tables 3.1–3.3 To study the impact of omitting random effects, the Cox model is estimated too. For the estimate of β, it can be seen that the log-logistic frailty model gives a similar accurate estimator to the lognormal frailty model and the gamma frailty model. The simulation results also indicate that the log-logistic frailty model is robust against misspecification of random effect, which is similar to the lognormal frailty model and the gamma frailty model. From the simulation results, we can conclude that the hierarchical likelihood method is suitable for the log-logistic frailty model outside the exponential family range. The log-logistic frailty model can provide a good choice for correlated survival data besides the lognormal frailty model and the gamma frailty model.

Tables 3.1–3.3 also listed the estimated mean for the frailty-parameter a and the corresponding mean of the frailty variance. All three simulations indicate that the frailty-variance estimates of the three frailty models are reasonable, except when the frailty model was gamma, fitting incorrect lognormal frailty model slightly underestimated the frailty variance.

Table 3.1 Simulation results under Log-logistic data generation process with frailty parameter $\theta = \frac{\sqrt{3}}{\pi} = 0.5513289$.

(G, n_i)	Model	Regression parameter β=1.0			Frailty parameter		Random Effect Std (1.0)	
		Mean	MSE	95% CP	Mean	MSE	Mean	MSE
30;2	Cox	0.703	0.199	0.780	-	-	-	-
	LLM	1.001	0.171	0.955	0.650	0.044	1.180	0.146
	LNM	1.040	0.218	0.940	1.064	0.286	1.064	0.286
	GAM	1.006	0.187	0.945	0.801	0.300	1.073	0.293
30;4	Cox	0.691	0.135	0.685	-	-	-	-
	LLM	0.994	0.059	0.960	0.542	0.012	0.982	0.041
	LNM	1.006	0.062	0.955	0.962	0.053	0.962	0.053
	GAM	0.990	0.059	0.960	0.678	0.067	0.972	0.064
80;2	Cox	0.693	0.128	0.595	-	-	-	-
	LLM	0.978	0.063	0.925	0.599	0.008	1.087	0.027
	LNM	0.993	0.073	0.910	0.966	0.077	0.966	0.077
	GAM	0.984	0.069	0.925	0.676	0.066	0.970	0.064
80;4	Cox	0.698	0.108	0.370	-	-	-	-
	LLM	0.986	0.025	0.950	0.544	0.005	0.987	0.017
	LNM	1.008	0.026	0.935	0.976	0.022	0.976	0.022
	GAM	0.992	0.025	0.935	0.686	0.026	0.984	0.025

Std is standard deviation; G represents the cluster number; ni represents the number of observations per cluster; Cox represents the results of the Cox proportional hazard model; LLM for the log-logistic frailty model, LNM for the lognormal frailty model; GAM for the gamma frailty model.

Table 3.2 Simulation results under Lognormal data generation process with frailty parameter $\theta = 1.0$

(G, n_i)	Model	Regression parameter β=1.0			Frailty parameter		Random Effect Std (1.0)	
		Mean	MSE	95% CP	Mean	MSE	Mean	MSE
30;2	Cox	0.724	0.177	0.850	-	-	-	-
	LLM	1.033	0.166	0.940	0.641	0.033	1.163	0.108
	LNM	1.081	0.211	0.910	1.058	0.242	1.058	0.242
	GAM	1.040	0.178	0.940	0.793	0.230	1.062	0.238
30;4	Cox	0.694	0.143	0.68	-	-	-	-
	LLM	0.976	0.071	0.935	0.570	0.014	1.034	0.046
	LNM	1.001	0.075	0.945	1.011	0.052	1.011	0.052
	GAM	0.992	0.074	0.940	0.758	0.082	1.049	0.076
80;2	Cox	0.669	0.136	0.540	-	-	-	-
	LLM	0.972	0.045	0.975	0.604	0.009	1.095	0.029
	LNM	0.998	0.056	0.960	0.989	0.076	0.989	0.076
	GAM	0.989	0.051	0.960	0.717	0.081	1.007	0.078
80;4	Cox	0.668	0.125	0.265	-	-	-	-
	LLM	0.964	0.019	0.970	0.558	0.005	1.012	0.017
	LNM	0.988	0.018	0.975	1.000	0.019	1.000	0.019
	GAM	0.975	0.019	0.975	0.731	0.029	1.026	0.027

Std is standard deviation; G represents the cluster number; ni represents the number of observations per cluster; Cox represents the results of the Cox proportional hazard model; LLM for the log-logistic frailty model, LNM for the lognormal frailty model; GAM for the gamma frailty model.

Table 3.3 Simulation results under gamma data generation process with variance $\theta = 0.70113689$

(G, n_i)	Model	Regression parameter β=1.0			Frailty parameter		Random Effect Std (1.0)	
		Mean	MSE	95% CP	Mean	MSE	Mean	MSE
30;2	Cox	0.687	0.179	0.815	-	-	-	-
	LLM	0.984	0.155	0.940	0.580	0.019	1.052	0.061
	LNM	0.992	0.188	0.925	0.885	0.199	0.884	0.199
	GAM	1.007	0.180	0.930	0.686	0.199	0.963	0.205
30;4	Cox	0.694	0.143	0.680	-	-	-	-
	LLM	0.976	0.071	0.935	0.570	0.014	1.034	0.046
	LNM	1.001	0.075	0.945	1.011	0.052	1.011	0.052
	GAM	0.992	0.074	0.940	0.758	0.082	1.049	0.076
80;2	Cox	0.679	0.139	0.560	-	-	-	-
	LLM	0.984	0.058	0.955	0.584	0.007	1.060	0.024
	LNM	0.993	0.067	0.920	0.928	0.071	0.928	0.071
	GAM	1.012	0.066	0.940	0.713	0.078	1.006	0.073
80;4	Cox	0.675	0.122	0.315	-	-	-	-
	LLM	0.979	0.022	0.970	0.521	0.004	0.946	0.016
	LNM	0.998	0.023	0.950	0.924	0.023	0.924	0.023
	GAM	1.004	0.022	0.965	0.672	0.024	0.969	0.023

Std is standard deviation; G represents the cluster number; ni represents the number of observations per cluster; Cox represents the results of the Cox proportional hazard model; LLM for the log-logistic frailty model, LNM for the lognormal frailty model; GAM for the gamma frailty model.

3.4. Data example (cow mastitis data)

Mastitis, the udder infection, is economically the most important disease in the western world's dairy sector. Many organisms can cause mastitis, most of the bacteria, such as Escherichia coli, etc. Since each udder quarter is separated from the three other quarters, one quarter might be infected with the other quarters free of infection. In a study by(Adkinson, Ingawa et al. 1993), 100 cows are followed up for infections. This observational study aims to estimate the incidence of the different organisms causing mastitis in the dairy cattle population in Flanders. Also, the correlation between the four udder quarters of a cow's infection is an important parameter for taking preventive measures against mastitis. Much attention should be given to the uninfected udder quarters of a cow with an infected quarter with a high correlation. A milk sample is taken monthly from each quarter and is screened for the presence of different bacteria. We model the time to infection with any bacteria, with the cow being the cluster and the quarter the cluster's experimental unit. Observations are right-censored if no infection occurs before the end of the lactation period, which is roughly 300-350 days but different for every cow, or if the cow is lost to follow-up during the study, for example, due to culling. In the analysis, one covariate is considered. Cow level covariates take the same value for every udder quarter of the cow (e.g., number of calvings or parity). Several studies have shown that prevalence, as well as the incidence of intramammary infections, increase with parity. Several hypotheses have been suggested to explain these findings, e.g., teat end condition deteriorates with increasing parity. Because the teat end is a physical barrier that prevents organisms from invading the udder, impaired teat ends make the udder more vulnerable for intramammary infections. For simplicity, parity is dichotomized into primiparous cows (heifer=1) and multiparous cows (heifer=0).

The log-logistic frailty model, lognormal frailty model, and the gamma frailty model are estimated simultaneously for the cow mastitis data set. Table 3.6., we find that value of the coefficient estimator of heifer in the gamma frailty model is larger than in the log-logistic frailty model and the lognormal frailty model. The estimators have a clear difference between the Cox model and the frailty models, and the estimators of the Cox model are towards zero more than the frailty models. We can also find that the heifer effect is significant only in the gamma frailty model, which indicates that primiparous cows are more susceptible to infection than multiparous cows.

Table 3.4 Analysis of Cow Mastitis data

Model	Heifer		Frailty $\hat{\theta}$	Variance $var(\hat{\theta})$
	$\hat{\beta}$	SE		
Cox	0.145	0.120	-	-
LLM	0.417	0.365	0.882	2.559
LNM	0.448	0.363	2.388	2.388
Gam	0.712	0.332	1.515	3.113

LLM for the log-logistic frailty model, LNM for the lognormal frailty model; GAM for the gamma frailty model.

3.5. Conclusion

This chapter investigated log-logistic frailty distribution, which is out of the exponential family range in the Cox proportional hazard model. To estimate the log-logistic frailty model, the hierarchical likelihood method is used to estimate the regression parameters and predict the realizations of random effects. The adjusted profile hierarchical likelihood is adopted to estimate the frailty parameters. It has been shown that the hierarchical likelihood method can give the accurate estimates of the parameter for the log-logistic frailty model, and the log-logistic frailty model is robust against misspecification of random effect through simulation studies. The estimating process is simplified mainly by using the hierarchical likelihood method, which avoids multidimensional integration over the frailties. The simulation studies indicate that the log-logistic frailty model is suited for multivariate survival data analysis besides the gamma frailty model and the lognormal frailty model. The research in this chapter is a good attempt to apply the hierarchical likelihood to nonexponential distribution.

Chapter 4

Hierarchical likelihood estimation of the log skew normal shared frailty model

Abstract

In this chapter, we present a frailty model using the log-skew normal distribution as the frailty distribution. It is an extension of the popular lognormal frailty model. It includes the lognormal as a special case. This frailty distribution's flexibility makes it possible to detect a complex frailty distribution structure that may otherwise be missed. Due to the intractable integrals in the likelihood function, we propose the hierarchical likelihood estimation method of estimation for the model's parameters, which avoids integrating the likelihood functions. We investigate the properties of the proposed frailty model via a simulation study.

KEYWORDS: skew-normal distribution, multivariate survival data, frailty model, hierarchical likelihood, adjusted profile likelihood

4.1. Introduction

The frailty model has been widely used to analyze multivariate survival data to account for potential correlation among failure times. The frailty variable describes the heterogeneity in the data caused by unknown covariates or randomness in the data. A frailty model for survival data is defined as follows. Let $(t_{ij}, \delta_{ij}, \boldsymbol{x}_{ij})$, $i = 1, \dots, G$, $j = 1, \dots, n_i$, be the failure time, censoring indicator, and a vector covariate of the jth individual in the ith cluster, where $\delta_{ij} = 1$ if t_{ij} is not censored and 0 otherwise. Let u_i denote the unobserved frailty shared by the individuals in the ith cluster usually assumed to be independent and identically distributed random variable with density function $f(u)$. Given u_i, the frailty model specifies that t_{ij} are independent with a proportional hazards function

$$\lambda_{ij}(t_{ij}|u_i) = \lambda_0(t_{ij})u_i \exp(\boldsymbol{x}_{ij}^T\boldsymbol{\beta}), \tag{4.1}$$

where $\lambda_0(t)$ is a baseline hazard function .

4.2. Log skew-normal frailty model

The skew-normal (SN) distribution is an extension of the normal distribution to allow for non-zero skewness by using a shape parameter.(Azzalini 1985) The random variable $\mathcal{Z}$ is said to have a scalar $SN(a)$ distribution if its density is given by

$$f(z; a) = 2\phi(z)\Phi(az), \quad -\infty < z < \infty;\ -\infty < a < \infty, \tag{4.2}$$

where a is the shape parameter which determines the skewness, and ϕ and Φ denote, respectively, the probability density function (PDF) and the cumulative density function (CDF) of a standard Gaussian random variable. When $a > 0$, we have a distribution with positive skewness, and $a < 0$ corresponds to negative skewness; if $a = 0$, we are back to the usual standard normal density. The mean and variance of $\mathcal{Z}$ are

$$E(\mathcal{Z}) = \mu_z = \sqrt{\frac{2}{\pi}}\frac{a}{\sqrt{1 + a^2}},$$

$$Var(\mathcal{Z}) = \sigma_z^2 = 1 - \mu_z^2 = 1 - \frac{2a^2}{\pi(1+a^2)}.$$

A transformation is required to extend the equation (4.2) with the introduction of a location and a scale parameter. Let $v_i = \sigma \frac{z_i - E(\mathcal{Z}_i)}{\sqrt{Var(\mathcal{Z}_i)}} \Rightarrow z_i = \frac{v_i}{\sigma}\sqrt{Var(\mathcal{Z}_i)} + E(\mathcal{Z}_i)$, then the density function of v_i is,

$$f(v_i) = f(z_i; a) \times \frac{\partial z_i}{\partial v_i} = 2\phi(z_i)\Phi(az_i) \times \frac{1}{\sigma}\sqrt{Var(\mathcal{Z}_i)},$$

$$= \frac{2}{\sqrt{2\pi}} \exp\left\{-\frac{z_i^2}{2}\right\} \times \Phi(az_i) \times \frac{1}{\sigma}\sqrt{Var(\mathcal{Z}_i)},$$

where $z_i = \frac{v_i}{\sigma}\sqrt{Var(\mathcal{Z}_i)} + E(\mathcal{Z}_i)$.

We fix $E(V) = \mu = 0$, to avoid identifiability issues. We then define the conditional hazard function of the log-skew normal frailty model as

$$\lambda_{ij}(t_{ij}|v_i) = \lambda_0(t_{ij}) \exp(\boldsymbol{x}_{ij}^T \boldsymbol{\beta} + v_i), \qquad (4.3)$$

where $v_i = \log\, u_i$ in equation (1), λ_0 is an arbitrary baseline hazard function and $v_i, (i = 1, \dots, G)$ are independent and identically skew-normally distributed with mean 0, variance σ^2 and shape parameter a. The complete data likelihood is given v_i's is

$$\mathcal{L}(.) = \prod_{i=1}^{G} \prod_{j=1}^{n_i} \left(\lambda_0 \exp(\boldsymbol{x}_{ij}^T \boldsymbol{\beta} + v_i)\right)^{\delta_{ij}} \exp\left(-\Lambda_0 \exp(\boldsymbol{x}_{ij}^T \boldsymbol{\beta} + v_i)\right) f(v_i), \qquad (4.4)$$

where $\Lambda_0(.|v_i) = \int_0^t f(s)ds$ conditional cumulative baseline hazard function of T_{ij} given $V_i = v_i$.

4.3. Estimation procedure

The h-likelihood(Lee and Nelder 1996, Lee and Nelder 2001) has been applied by several authors to estimate the parameters of Cox random effects models by treating the frailty variates as parameters and estimating jointly with the parameters of interests; see specifically.(Ha, Lee et al. 2001, Ha and Lee 2003, Ha and Lee 2005, Christian, Ha et al. 2016, Ha, Jeong et al. 2017) The h-likelihood for shared frailty is the logarithm of the complete data likelihood in the equation (4.4) and it is given by

$$h = \sum_{ij} \ell_{1ij} + \sum_{i} \ell_{2i}, \tag{4.5}$$

where ℓ_{1ij} is the logarithm of the conditional likelihood in T_{ij} and δ_{ij} given $V_i = v_i$ with parameters $(\boldsymbol{\beta}, \lambda_0)$ and ℓ_{2i} is the log density function of $V_i = v_i$ with parameters $\boldsymbol{\theta} = (a, \sigma)^T$. The logarithm of the conditional density function of (T_{ij}, δ_{ij}) given $V_i = v_i$ is

$$\ell_{1ij} = \delta_{ij}\{\log \lambda_0(y_{ij}) + \eta_{ij}\} - \Lambda_0(y_{ij}) \exp(\eta_{ij}),$$

where $\eta_{ij} = \boldsymbol{x}_{ij}^T \boldsymbol{\beta} + \boldsymbol{v}$, with $\boldsymbol{v} = (v_1, \dots, v_G)^T$ and the $\ell_{2i} = \log f(v_i)$ is the logarithm of the probability density function of the frailty distribution given by

$$\log f(v_i) = \log 2 - \frac{1}{2}\log(2\pi) - \frac{z_i^2}{2} + \log(\Phi(az_i)) + \log\left\{\frac{1}{\sigma}\sqrt{Var(Z_i)}\right\}. \tag{4.6}$$

We will now describe the h-likelihood steps in more detail. Notice that the functional form of $\lambda_0(t_{ij})$ in equation (4.3) is unknown. FollowingBreslow (1972) and Ha, Lee et al. (2001), we define the baseline cumulative hazard function to be a step function with jumps at λ_{0k} the observed event times $t_{(k)}$, defined by

$$\Lambda_0(t) = \sum_{k: t_{(k)} \le t} \lambda_{0k}, \tag{4.7}$$

where $t_{(k)}$ is the kth $(k = 1, \dots, D)$ is the smallest distinct event time among the t_{ij}^*'s and $\lambda_{0k} = \lambda_0(t_{(k)})$. By substituting (4.7) into (4.5), the second term $\sum_{ij} \ell_{1ij}$ in (4.5) becomes

$$\sum_{ij} \ell_{1ij} = \sum_{k} d_{(k)} \log \lambda_{0k} + \sum_{ij} \delta_{ij}\, \eta_{ij} - \sum_{k} \lambda_{0k} \left\{ \sum_{ij \in \mathcal{R}(t_{(k)})} \exp(\eta_{ij}) \right\},$$

where $d_{(k)}$ is the number of events at $t_{(k)}$ and

$$\mathcal{R}(t_{(k)}) = \{ij: t_{ij}^* \geq t_{(k)}\}$$

is the risk set at $t_{(k)}$. As the number of in λ_{0k}'s in $\sum_{ij} \ell_{1ij}$ above increases with the number of events, the function $\lambda_0(t)$ is potentially of high dimension. FollowingHa, Lee et al. (2001), we use a profile h-likelihood after eliminating nuisances λ_{0k}, given by

$$h^* = h|_{\lambda_0=\hat{\lambda}_0} = \sum_{ij} \ell_{1ij}^* + \sum_i \ell_{2i}, \tag{4.8}$$

where

$$\sum_{ij} \ell_{1ij}^* = \sum_{ij} \ell_{1ij}|_{\lambda_0=\hat{\lambda}_0} = \sum_k d_{(k)} \log \hat{\lambda}_{0k} + \sum_{ij} \delta_{ij}\, \eta_{ij} - \sum_k \lambda_{0k}.$$

Here

$$\hat{\lambda}_{0k} = \hat{\lambda}_{0k}(\boldsymbol{\beta}, \boldsymbol{v}) = \frac{d_{(k)}}{\sum_{ij\in\mathcal{R}(t_{(k)})} \exp(\eta_{ij})},$$

are the solutions of the estimating equations, $\partial h / \partial \lambda_{0r} = 0$, for $r = 1, \dots, D$. We thus see that h^* does not depend on λ_0.

Let $\boldsymbol{X}$ and $\boldsymbol{Z}$ be model matrices for $\boldsymbol{\beta}$ and $\boldsymbol{v}$, respectively. The score equations for fixed and random effects $(\boldsymbol{\beta}, \boldsymbol{v})$ given $\boldsymbol{\theta} = (\sigma,\ a)^T$ are given by

$$\frac{\partial h^*}{\partial \boldsymbol{\beta}} = \boldsymbol{X}^T(\boldsymbol{\delta} - \boldsymbol{\mu}),$$

$$\frac{\partial h^*}{\partial \boldsymbol{v}} = \boldsymbol{Z}^T(\boldsymbol{\delta} - \boldsymbol{\mu}) + \frac{\partial l_{2i}}{\partial \boldsymbol{v}}.$$

Here $\boldsymbol{\mu} = \exp(\log \widehat{\Lambda}_0(t^*) + \boldsymbol{\eta})$ with $\boldsymbol{\eta} = \boldsymbol{X\beta} + \boldsymbol{Zv}$, and

$$\widehat{\Lambda}_0(t) = \sum_{k:t_{(k)}\leq t} \hat{\lambda}_{0k}$$

is the Breslow-type estimator of cumulative baseline hazard.

The parameters $\boldsymbol{\tau} = (\boldsymbol{\beta}^T, \boldsymbol{v}^T)^T$ can be estimated via the iterative least square methodLee and Nelder (1996) and Ha and Lee (2003) given by

$$\begin{pmatrix} \boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X} & \boldsymbol{X}^T\boldsymbol{W}\boldsymbol{Z} \\ \boldsymbol{Z}^T\boldsymbol{W}\boldsymbol{X} & \boldsymbol{Z}^T\boldsymbol{W}\boldsymbol{Z}+\boldsymbol{Q} \end{pmatrix}\begin{pmatrix} \widehat{\boldsymbol{\beta}} \\ \widehat{\boldsymbol{v}} \end{pmatrix} = \begin{pmatrix} \boldsymbol{X}^T\boldsymbol{w} \\ \boldsymbol{Z}^T\boldsymbol{w}+\boldsymbol{R} \end{pmatrix}, \tag{4.9}$$

where the adjusted dependent variable, $\boldsymbol{w} = \boldsymbol{W}(\boldsymbol{X}^T\boldsymbol{\beta} + \boldsymbol{Z}^T\boldsymbol{v}) + (\boldsymbol{\delta} - \boldsymbol{\mu})$, and the detailed matrix form of $\boldsymbol{W}$ is given in Appendix B ofHa and Lee (2003), $\boldsymbol{Q}$ is the $G \times G$ diagonal matrix whose ith element is $-\frac{\partial^2 \ell_{2i}}{\partial v^2}$ and $\boldsymbol{R} = \boldsymbol{Q}\boldsymbol{v} + \frac{\partial \ell_{2i}}{\partial v}$.

The asymptotic covariance matrix for $\hat{\boldsymbol{\tau}} - \boldsymbol{\tau}$ is obtained from $\boldsymbol{H}^{-1}$ where $\boldsymbol{H} = -\partial^2 h^* / \partial(\boldsymbol{\beta}, \boldsymbol{v})^2 = \begin{pmatrix} \boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X} & \boldsymbol{X}^T\boldsymbol{W}\boldsymbol{Z} \\ \boldsymbol{Z}^T\boldsymbol{W}\boldsymbol{X} & \boldsymbol{Z}^T\boldsymbol{W}\boldsymbol{Z}+\boldsymbol{Q} \end{pmatrix}$. So, the upper left-hand corner of $\boldsymbol{H}^{-1}$ gives the asymptotic variance matrix of $\widehat{\boldsymbol{\beta}}$,

$$\mathrm{var}(\widehat{\boldsymbol{\beta}}) = (\boldsymbol{X}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1},$$

where $\boldsymbol{\Sigma} = \boldsymbol{W}^{-1} + \boldsymbol{Z}\boldsymbol{Q}^{-1}\boldsymbol{Z}^T$.

4.3.1. Fitting procedure

Let $\boldsymbol{P} = \begin{pmatrix} \boldsymbol{X} & \boldsymbol{Z} \\ \boldsymbol{0} & \boldsymbol{I}_G \end{pmatrix}$ and $\boldsymbol{V} = \begin{pmatrix} \boldsymbol{W} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{Q} \end{pmatrix}$, then the fitting procedure consists of the following two steps:

(i) Estimation of fixed and random effects $= (\boldsymbol{\beta}^T, \boldsymbol{v}^T)^T$. Following Lee et al. (2006), the ILS equations above reduce to a simple explicit form,

$$(\boldsymbol{P}^T\boldsymbol{V}\boldsymbol{P})\hat{\tau} = \boldsymbol{P}^T\boldsymbol{y}_{\boldsymbol{0}},$$

where $\boldsymbol{y}_{\boldsymbol{0}} = (\boldsymbol{w}^T, \boldsymbol{R}^T)^T$. Note that $\boldsymbol{H} = -\partial^2 h^* / \partial(\boldsymbol{\beta}, \boldsymbol{v})^2 = \boldsymbol{P}^T\boldsymbol{V}\boldsymbol{P}$.

(ii) estimation of the parameters $\boldsymbol{\theta}$ can be estimated by maximizingLee and Nelder (1996) adjusted profile h-likelihood,

$$h_A^* = h^*|_{\tau=\hat{\tau}} + \frac{1}{2}\log\{\det(2\pi\boldsymbol{H}^{-1})\}\Big|_{\tau=\hat{\tau}}. \tag{4.10}$$

The adjusted profile h-likelihood is used to approximate the restricted likelihood of θ that considers the estimation of $\boldsymbol{\beta}$ and $\boldsymbol{v}$. The estimating equations of θ is given by

$$\frac{\partial h_A^*}{\partial \boldsymbol{\theta}} = \mathbf{0}. \tag{4.11}$$

The Newton-Raphson method is can be used to find $\hat{\theta}$ the maximum hierarchical likelihood estimator of $\boldsymbol{\theta}$. This requires finding the first and second derivatives of h_A^* with respect $\boldsymbol{\theta}$. Details of the direct calculation of first and second derivatives for the log skew-normal frailty model are available in Appendix 3.

4.4. Numerical study

Based on 250 replications of simulated data, a numerical study is presented to evaluate the proposed h-likelihood estimation method's performance to fit log-skew normal frailty models. The data were generated using the following conditional hazard model

$$\lambda_{ij}(t) = \lambda_0(t)\exp\left(\beta x_{ij} + v_i\right). \tag{4.12}$$

The model corresponds to the setting of $G = 30$ clusters containing $n_i = 2$ or 5 $(i = 1, \dots, G)$ subjects. The cluster effect v_i was generated from the skew-normal distribution with a mean $\mu = 0$, standard deviation $\sigma = 1$, and shape parameter $a = (-4, 0, 4)$. The true regression parameter was taken to be $\beta = 1.0$, and we set x_{ij} to 0 for the first $G/2$ individuals, to form the control group, and x_{ij} to 1 for the remaining $G/2$, to form the treatment group. Given $V_i = v_i$ the independent survival times T_{ij}, $j = 1, \dots, n_i$ are generated from the model in (4.12) with a Weibull baseline hazard, that is $\lambda_0(t) = \rho\gamma t_{ij}^{\gamma-1}$, with $\rho = 1/80$ and $\gamma = 5$. Data were censored using a right-censoring random variable generated from a uniform distribution on $[0, l]$, with l chosen to obtain a percentage of censoring in each simulated dataset around 20%. From the 250 simulations, we computed the mean, standard deviation, and the mean of the SEs for the fixed effects $\hat{\beta}$. For the frailty paramter $\hat{\sigma}$ and for the shape parameter $\hat{a}$ we computed the mean and standard deviation. Further, we computed the coverages of the β (CP %), i.e., the percentage of simulated data for which the 95% confidence interval contains the real parameter. The results of the simulation studies are summarized in Table 4.1.

From Table 4.1, we find that, in general, the fixed effects β are estimated well by the proposed method. Note, however, that fixed effects are robust against misspecification of the frailty distribution. The variance of $\hat{\beta}$ seems to be slightly overestimated since the standard error is slightly greater than the standard deviation. The frailty standard deviation $\hat{\sigma}$ is well estimated. However, the proposed method estimated the shape parameter very poorly.

Table 4.1 Estimated parameters and their estimated and empirical (se) in 250 simulations based on a shared frailty model with G clusters and $n_i = (2,5)$ repetitions per cluster with standard deviation $\sigma = 1.0$ and skew parameter $a = (-4,0,4)$.

		Regression parameter $(\beta = 1.0)$			Frailty Parameter $(\sigma = 1.0)$		Skew Parameter	
					Mean	SD	Mean	SD
a	n_i	Mean	SD (se)	95% CP				
4.0	2	1.020	0.515 (0.516)	0.940	0.980	0.447	0.020	0.039
	5	1.025	0.389 (0.426)	0.956	1.007	0.223	0.008	0.005
0.0	2	1.003	0.518 (0.509)	0.944	0.979	0.345	0.013	0.021
	5	0.998	0.416 (0.425)	0.936	1.008	0.206	0.007	0.002
-4.0	2	1.022	0.547 (0.525)	0.936	1.027	0.369	0.011	0.014
	5	0.992	0.413 (0.419)	0.964	0.989	0.219	0.007	0.0023

Mean, and SD indicates the mean,and standard deviation

4.5. Conclusion

In this chapter, a log skew-normal frailty model is introduced, an extension to the classical lognormal frailty model(McGilchrist and Aisbett 1991, Ha, Lee et al. 2001, Therneau, Grambsch et al. 2003), to include an additional shape parameter. The shape parameter gives more flexibility to the distribution of the unobserved random effects. The flexibility is an important feature of the proposed method because the random effect distribution choice is crucial to obtain a more realistic estimate of the dependence structure.

To fit the proposed model, we developed an h-likelihood algorithm to estimate the model parameters and to predict the realization of the random effects. The proposed approach produced good estimates of the fixed effects and the frailty parameters. However, the adjusted profile likelihood estimated the shape parameter very poorly from the simulation study. This shows that adjusted profile likelihood estimates of $\hat{\sigma}$ are robust with respect to the poor estimates of the shape parameter. The nice behavior of $\hat{\sigma}$ suggests we can combine the hierarchical likelihood method and the traditional maximum likelihood approach in the estimation process, where the hierarchical likelihood method is adopted to estimate parameters $(\boldsymbol{\beta}, \boldsymbol{v}, \sigma)$, and the maximum likelihood is used to estimate the shape parameter α. However, the maximum likelihood function of the skew-normal distribution is monotone in α and thus, the maximum likelihood estimate $\hat{\alpha}$ of the shape parameter α takes on a $\pm\infty$ with a non zero probability(Liseo 1990). To solve this problem, a penalized likelihood approach may be used. We are investigating if combining the h-likelihood with penalized likelihood is necessary to estimate the skewness parameters efficiently.

Chapter 5

General conclusions

In many clinical trials, time-to-event endpoints, which are often adopted to demonstrate a clinically convincing effect of treatments appropriately, maybe clustered or correlated because of certain common features such as genetic traits or shared environmental factors, or repeated events. Observations from the same cluster are assumed to be correlated because they usually share specific unobserved characteristics. When the correlation between survival times is present, the frailty model can explain the relationship between covariates and a time-to-event outcome. The frailty model concept provides a convenient way of introducing unobserved heterogeneity and associations into the classical proportional hazard survival model, a random effect model in survival analysis.

Recently, the h-likelihood estimation procedure, a computationally efficient approach, has been developed to fit these types of models. A drawback of using the h-likelihood is that it can be challenging to implement because of the numerous derivatives that need to be calculated. Once the derivatives are calculated, though, the analysis is computationally efficient. However, the frequently used expectation maximization (EM) algorithm will always be computationally intensive since it involves integration over multidimensional frailty variates to obtain marginal likelihood function.

The current work considered frailty models for modeling dependence in multivariate survival data that arise because individuals in the same group (family, litter, study center) are related to each other, or the individual experience multiple recurrences of the same event. The frailty model assumes that all individuals are susceptible to the event of interest and will eventually experience this event if the follow-up is sufficiently long. However, it may be possible that a fraction of individuals in the population may not be susceptible to the event under study. Thus, certain proportions of subjects in the population who are not expected to experience the events of interest can be considered cured. It may be interesting to consider cure rate frailty models when cure fractions are present.(Kuk and Chen 1992, Stoltenberg, Nordeng et al. 2020) This is a unique model in that it allows for modeling the heterogeneity in risk among those individuals experiencing the event of interest incorporating a surviving fraction.

Another future work area is extending the gamma frailty model introduced in Chapter two to the correlated gamma frailty model. This will allow for the estimation of the variance parameter of the frailties as in shared frailty models and the estimation of an extra parameter for modeling the correlation between frailties in each or cluster. Thus the correlated frailty model becomes a natural extension of the shared frailty approach where subjects in a cluster are assumed correlated but not necessarily shared. It will be interesting to consider bivariate gamma frailty(Wienke, Holm et al. 2003, Fiocco, Putter et al. 2009, Hens, Wienke et al. 2009) to model the kidney catheter data presented in Chapter 2 and describe the h-likelihood approach for estimating the model parameters.

Another interesting scientific question in recurrent events research is whether a fatal event such as death could be correlated with repeated events such as multiple hospitalizations for heart attack and tumor relapse. Here, the usual assumption of noninformative censoring of the recurrent event process by death, required by the shared frailty model, may not be appropriate. This dependence should be accounted for in the joint modeling of recurrent events and deaths. Thus, it would be worthwhile to see the likelihood method applies to the joint frailty models.(Huang and Liu 2007, Belot, Rondeau et al. 2014, Emura, Nakatochi et al. 2017)

References

Aalen, O. O. (1978). "Nonparametric inference for a family of counting processes." The Annals of Statistics: 701-726.

Aalen, O. O. (1988). "Heterogeneity in survival analysis." Statistics in medicine **7**(11): 1121-1137.

Aalen, O. O. (1994). "Effects of frailty in survival analysis." Statistical Methods in Medical Research **3**(3): 227-243.

Aalen, O. O. and J. M. Hoem (1978). "Random time changes for multivariate counting processes." Scandinavian Actuarial Journal **1978**(2): 81-101.

Adkinson, R., et al. (1993). "Distribution of clinical mastitis among quarters of the bovine udder1." Journal of dairy science **76**(11): 3453-3459.

Andersen, P. K. and R. D. Gill (1982). "Cox's regression model for counting processes: a large sample study." The annals of statistics: 1100-1120.

Androulakis, E., et al. (2012). "Estimation and variable selection via frailty models with penalized likelihood." Statistics in Medicine **31**(20): 2223-2239.

Azzalini, A. (1985). "A class of distributions which includes the normal ones." Scandinavian journal of statistics: 171-178.

Barker, P. and R. Henderson (2005). "Small sample bias in the gamma frailty model for univariate survival." Lifetime data analysis **11**(2): 265-284.

Belot, A., et al. (2014). "A joint frailty model to estimate the recurrence process and the disease ‐ specific mortality process without needing the cause of death." Statistics in medicine **33**(18): 3147-3166.

Breslow, N. (1974). "Covariance analysis of censored survival data." Biometrics: 89-99.

Breslow, N. E. (1972). "Discussion of Professor Cox's paper." J Royal Stat Soc B **34**: 216-217.

Cai, J. and R. L. Prentice (1995). "Estimating equations for hazard ratio parameters based on correlated failure time data." Biometrika **82**(1): 151-164.

Chen, M. H., et al. (2009). "A frailty model approach for regression analysis of multivariate current status data." Statistics in Medicine **28**(27): 3424-3436.

Christian, N. J., et al. (2016). "Hierarchical likelihood inference on clustered competing risks data." Statistics in medicine **35**(2): 251-267.

Clayton, D. G. (1978). "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.

Cortinas Abrahantes, J. and T. Burzykowski (2005). "A version of the EM algorithm for proportional hazard model with random effects." Biometrical Journal: Journal of Mathematical Methods in Biosciences **47**(6): 847-862.

Cortiñas, A. J. and T. Burzykowski (2005). "A version of the EM algorithm for proportional hazard model with random effects." Biometrical Journal: Journal of Mathematical Methods in Biosciences **47**(6): 847-862.

Cox, D. R. (1975). "Partial likelihood." Biometrika **62**(2): 269-276.

Cox, D. R. (1992). Regression models and life-tables. Breakthroughs in statistics, Springer: 527-541.

Dempster, A. P., et al. (1977). "Maximum likelihood from incomplete data via the EM algorithm." Journal of the Royal Statistical Society: Series B (Methodological) **39**(1): 1-22.

Dempster, A. P., et al. (1977). "Maximum likelihood from incomplete data via the EM algorithm." Journal of the Royal Statistical Society. Series B (Methodological): 1-38.

Duchateau, L. and P. Janssen (2004). "Penalized partial likelihood for frailties and smoothing splines in time to first insemination models for dairy cows." Biometrics **60**(3): 608-614.

Duchateau, L. and P. Janssen (2007). The Frailty Model, Springer Science & Business Media.

Emura, T., et al. (2017). "A joint frailty-copula model between tumour progression and death for meta-analysis." Statistical methods in medical research **26**(6): 2649-2666.

Enki, D. G., et al. (2014). "A time-varying shared frailty model with application to infectious diseases." The Annals of Applied Statistics: 430-447.

Fiocco, M., et al. (2009). "Meta ‐ analysis of pairs of survival curves under heterogeneity: a Poisson correlated gamma ‐ frailty approach." Statistics in medicine **28**(30): 3782-3797.

Firth, D. (1993). "Bias reduction of maximum likelihood estimates." Biometrika: 27-38.

Giussani, A. and M. Bonetti (2019). "A note on the length-biased Weibull-Gamma frailty survival model." Statistics & probability letters **153**: 32-36.

Ha, D. I., et al. (2018). Statistical modelling of survival data with random effects: h-likelihood approach, Springer.

Ha, D. I. and Y. Lee (2005). "Multilevel mixed linear models for survival data." Lifetime Data Analysis **11**(1): 131-142.

Ha, I. D., et al. (2017). Statistical Modelling of Survival Data with Random Effects, Springer.

Ha, I. D. and Y. Lee (2003). "Estimating frailty models via Poisson hierarchical generalized linear models." Journal of computational and graphical statistics **12**(3): 663-681.

Ha, I. D. and Y. Lee (2005). "Comparison of hierarchical likelihood versus orthodox best linear unbiased predictor approaches for frailty models." Biometrika **92**(3): 717-723.

Ha, I. D. and Y. Lee (2005). "Multilevel mixed linear models for survival data." Lifetime Data Analysis **11**(1): 131-142.

Ha, I. D., et al. (2001). "Hierarchical likelihood approach for frailty models." Biometrika **88**(1): 233-233.

Ha, I. D., et al. (2011). "Frailty modelling for survival data from multi‐centre clinical trials." Statistics in medicine **30**(17): 2144-2159.

Ha, I. D., et al. (2016). "Interval estimation of random effects in proportional hazards models with frailties." Statistical methods in medical research **25**(2): 936-953.

Hanagal, D. D. (2009). "Weibull extension of bivariate exponential regression model with different frailty distributions." Statistical Papers **50**(1): 29-49.

Hanagal, D. D. (2010). "Modeling heterogeneity for bivariate survival data by the Weibull distribution." Statistical Papers **51**(4): 947-958.

Hanagal, D. D. (2020). "Correlated positive stable frailty models." Communications in Statistics-Theory and Methods: 1-17.

Hens, N., et al. (2009). "The correlated and shared gamma frailty model for bivariate current status data: an illustration for cross‐sectional serological data." Statistics in medicine **28**(22): 2785-2800.

Hougaard, P. (1984). "Life table methods for heterogeneous populations: distributions describing the heterogeneity." Biometrika **71**(1): 75-83.

Hougaard, P. (1986). "A class of multivanate failure time distributions." Biometrika **73**(3): 671-678.

Hougaard, P. (1986). "Survival models for heterogeneous populations derived from stable distributions." Biometrika **73**(2): 387-396.

Hougaard, P. (2012). Analysis of Multivariate Survival Data, Springer Science & Business Media.

Huang, X. and L. Liu (2007). "A joint frailty model for survival and gap times between recurrent events." Biometrics **63**(2): 389-397.

Jeon, J., et al. (2012). "Bias correction in the hierarchical likelihood approach to the analysis of multivariate survival data." Biostatistics **13**(3): 384-397.

Keiding, N., et al. (1997). "The role of frailty models and accelerated failure time models in describing heterogeneity due to omitted covariates." Statistics in medicine **16**(2): 215-224.

Klein, J. P. (1992). "Semiparametric estimation of random effects using the Cox model based on the EM algorithm." Biometrics: 795-806.

Komárek, A. and E. Lesaffre (2008). "Bayesian accelerated failure time model with multivariate doubly interval-censored data and flexible distributional assumptions." Journal of the American Statistical Association **103**(482): 523-533.

Kuk, A. Y. and C.-H. Chen (1992). "A mixture model combining logistic regression with proportional hazards regression." Biometrika **79**(3): 531-541.

Lee, Y. and J. A. Nelder (1996). "Hierarchical generalized linear models." Journal of the Royal Statistical Society. Series B (Methodological): 619-678.

Lee, Y. and J. A. Nelder (2001). "Hierarchical generalised linear models: a synthesis of generalised linear models, random‐effect models and structured dispersions." Biometrika **88**(4): 987-1006.

Li, H. and P. Lahiri (2010). "An adjusted maximum likelihood method for solving small area estimation problems." Journal of multivariate analysis **101**(4): 882-892.

Liseo, B. (1990). "The skew-normal class of densities: inferential aspects from a Bayesian viewpoint." Statistica **50**: 59-70.

Martins, A., et al. (2019). "Correlated gamma frailty models for bivariate survival time data." Statistical methods in medical research **28**(10-11): 3437-3450.

McGilchrist, C. and C. Aisbett (1991). "Regression with frailty in survival analysis." Biometrics: 461-466.

McGilchrist, C. A. (1993). "REML estimation for survival models with frailty." Biometrics: 221-225.

Morris, C. N. (2006). "Mixed model prediction and small area estimation." Test **15**(1): 1.

Nielsen, G. G., et al. (1992). "A counting process approach to maximum likelihood estimation in frailty models." Scandinavian journal of Statistics: 25-43.

Noh, M. and Y. Lee (2007). "REML estimation for binary data in GLMMs." Journal of multivariate analysis **98**(5): 896-915.

Petersen, J. H. (1998). "An additive frailty model for correlated life times." Biometrics: 646-661.

Pipper, C. B. and T. Martinussen (2003). "A likelihood based estimating equation for the Clayton–Oakes model with marginal proportional hazards." Scandinavian journal of statistics **30**(3): 509-521.

Prentice, R. and J. Cai (1992). Marginal and conditional models for the analysis of multivariate failure time data. Survival Analysis: State of the Art, Springer**:** 393-406.

Prentice, R. L. and J. Cai (1992). "Covariance and survivor function estimation using censored multivariate failure time data." Biometrika **79**(3): 495-512.

R Core Team (2020). R: A language and environment for statistical computing. Vienna, Austria, R Foundation for Statistical Computing. URL https://www.R-project.org/.

Ripatti, S., et al. (2002). "Maximum likelihood inference for multivariate frailty models using an automated Monte Carlo EM algorithm." Lifetime data analysis **8**(4): 349-360.

Ripatti, S. and J. Palmgren (2000). "Estimation of multivariate frailty models using penalized partial likelihood." Biometrics **56**(4): 1016-1022.

Rondeau, V., et al. (2003). "Maximum penalized likelihood estimation in a gamma-frailty model." Lifetime Data Analysis **9**(2): 139-153.

Rondeau, V., et al. (2006). "Nested frailty models using maximum penalized likelihood estimation." Statistics in medicine **25**(23): 4036-4052.

Sahu, S. K., et al. (1997). "A Weibull Regression Model with Gamma Frailties for Multivariate Survival Data." Lifetime Data Analysis **3**(2): 123-137.

Sastry, N. (1997). "A nested frailty model for survival data, with an application to the study of child survival in northeast Brazil." Journal of the American Statistical Association **92**(438): 426-435.

Scudilio, J., et al. (2019). "Defective models induced by gamma frailty term for survival data with cured fraction." Journal of Applied Statistics **46**(3): 484-507.

Shun, Z. (1997). "Another look at the salamander mating data: A modified Laplace approximation approach." Journal of the American Statistical Association **92**(437): 341-349.

Sinha, D. (1993). "Semiparametric Bayesian analysis of multiple event time data." Journal of the American Statistical Association **88**(423): 979-983.

Stoltenberg, E. A., et al. (2020). "The cure model in perinatal epidemiology." Statistical methods in medical research: 0962280220904092.

Therneau, T. M., et al. (2003). "Penalized survival models and frailty." Journal of computational and graphical statistics **12**(1): 156-175.

Vaida, F. and R. Xu (2000). "Proportional hazards model with random effects." Statistics in medicine **19**(24): 3309-3324.

Vaupel, J. W., et al. (1979). "The impact of heterogeneity in individual frailty on the dynamics of mortality." Demography **16**(3): 439-454.

Wang, J. (2019). "Weighted estimation for multivariate shared frailty models for complex surveys." Lifetime Data Analysis **25**(3): 469-479.

Wang, N., et al. (2011). "Hierarchical likelihood approach for the Weibull frailty model." Journal of Statistical Computation and Simulation **81**(3): 343-356.

Wang, Y. (2006). "Estimation of accelerated failure time models with random effects."

Wei, L.-J., et al. (1989). "Regression analysis of multivariate incomplete failure time data by modeling marginal distributions." Journal of the American Statistical Association **84**(408): 1065-1073.

Wienke, A. (2010). Frailty Models in Survival Analysis, CRC press.

Wienke, A., et al. (2003). "The heritability of cause ‐ specific mortality: a correlated gamma ‐ frailty model applied to mortality due to respiratory diseases in Danish twins born 1870–1930." Statistics in medicine **22**(24): 3873-3887.

Wu, J. and P. M. Bentler (2012). "Application of H-likelihood to factor analysis models with binary response data." Journal of multivariate analysis **106**: 72-79.

Xue, X. and R. Brookmeyer (1996). "Bivariate frailty model for the analysis of multivariate survival time." Lifetime Data Analysis **2**(3): 277-289.

Yin, G. (2007). "Model checking for additive hazards model with multivariate survival data." Journal of Multivariate Analysis **98**(5): 1018-1032.

Yin, G. and J. Cai (2004). "Additive hazards model with multivariate failure time data." Biometrika **91**(4): 801-818.

Yin, G. and J. G. Ibrahim (2005). "A class of Bayesian shared gamma frailty models with multivariate failure time data." Biometrics **61**(1): 208-216.

Yu, B. (2006). "Estimation of shared gamma frailty models by a modified EM algorithm." Computational statistics & data analysis **50**(2): 463-474.

Zhang, X. (2006). "Generalized estimating equations for clustered survival data."

Appendix

Appendix 1

The appendix 1 lays out the mathematical derivation for the gamam frailty model in chapter 2.

A1.1 Maximum hierarchical likelihood estimator (MHLE) of $(\boldsymbol{\beta}, \boldsymbol{v})$

From equation (2.5), we obtain the first partial derivatives

$$\frac{\partial h^*}{\partial \beta_r} = \sum_{ij} \delta_{ij}\, x_{ijr} - \sum_{k} \frac{d_{(k)}}{\sum_{ij\in\mathcal{R}(t_k)} \exp(\eta_{ij})} \sum_{ij\in\mathcal{R}(t_k)} x_{ijr} \exp(\eta_{ij}), \qquad r = 1, \ldots, p,$$

$$= \sum_{ij} \delta_{ij}\, x_{ijr} - \widehat{\Lambda}_0(y_{ij}) \sum_{ij\in\mathcal{R}(t_k)} x_{ijr} \exp(\eta_{ij}),$$

where $\widehat{\Lambda}_0(y_{ij}) = \sum_k \frac{d_{(k)}}{\sum_{ij\in\mathcal{R}(t_k)} \exp(\eta_{ij})}$.

Thus, using matrix notations, we have

$$\frac{\partial h^*}{\partial \boldsymbol{\beta}} = \boldsymbol{X}^T(\boldsymbol{\delta} - \boldsymbol{\mu}).$$

Similarly, we express $\frac{\partial h^*}{\partial v_i}$, $i = 1, \ldots, G$ as

$$\frac{\partial h^*}{\partial \boldsymbol{v}} = \boldsymbol{Z}^T(\boldsymbol{\delta} - \boldsymbol{\mu}) + \boldsymbol{R},$$

where $\boldsymbol{\beta}$ is $p \times 1$ vector of fixed effects, $\boldsymbol{X}$ is $N \times p$ matrix of p covariates, $\boldsymbol{v}$ is $G \times 1$ vector of frailty variate, $\boldsymbol{Z}$ is $N \times G$ cluster indicator matrix, $\boldsymbol{\delta}$ is $N \times 1$ vector of δ_{ij}, $\boldsymbol{\mu}$ is $N \times 1$ vector with $\widehat{\Lambda}_0(y_{ij}) \exp(\boldsymbol{\eta})$ where $\widehat{\Lambda}_0(y_{ij}) = \sum_k \hat{\lambda}_0(t_k) I(y_{(k)} \le y_{ij})$, and $\boldsymbol{R} = \frac{\partial \ell_{2i}}{\partial \boldsymbol{v}}$, and $\boldsymbol{\eta} = \boldsymbol{X}^T\boldsymbol{\beta} + \boldsymbol{Z}^T\boldsymbol{v}$.

The vector $\boldsymbol{\mu}$ can be written as a simple form by using a weighted risk indicator matrix $\boldsymbol{M}$, which contains the risk set $R_{(k)}$. Let $\boldsymbol{M} = (R_1, R_2, \ldots, R_D)$ be a $N \times D$ at-risk indicator matrix where the ij^{th} element in column k is one if $I(y_{ij} \ge y_{(k)})$ and zero otherwise.

Define $\boldsymbol{B} = \mathrm{diag}\{\Lambda_0(y_{ij})\}$ as a $D \times D$ diagonal matrix. Let $\boldsymbol{W_1}$ be $N \times N$ diagonal matrix with elements $\exp(\boldsymbol{\eta})$, and $\boldsymbol{C}$ be a diagonal $D \times D$ matrix where the k^{th} element is $\frac{\left(\hat{\lambda}_0(t_{(k)})\right)^2}{d_{(k)}}$.

Let $\boldsymbol{\Omega} = \boldsymbol{W_1 B} - (\boldsymbol{W_1 M})\boldsymbol{C}(\boldsymbol{W_1 M})$,(Ha and Lee 2003)then, the observed information matrix $\boldsymbol{H}$ in (6) has the following entries:

$$\boldsymbol{H} = \begin{pmatrix} \boldsymbol{X}^T\boldsymbol{\Omega X} & \boldsymbol{X}^T\boldsymbol{\Omega Z} \\ \boldsymbol{Z}^T\boldsymbol{\Omega X} & \boldsymbol{Z}^T\boldsymbol{\Omega Z} + \boldsymbol{Q} \end{pmatrix}.$$

That is,

$$-\frac{\partial^2 h^*}{\partial \boldsymbol{\beta}^2} = \boldsymbol{X}^T\boldsymbol{\Omega X},$$

$$-\frac{\partial^2 h^*}{\partial \boldsymbol{\beta}\partial \boldsymbol{v}} = \boldsymbol{X}^T\boldsymbol{\Omega Z},$$

$$-\frac{\partial^2 h^*}{\partial \boldsymbol{v}\partial \boldsymbol{\beta}} = \boldsymbol{Z}^T\boldsymbol{\Omega X},$$

$$-\frac{\partial^2 h^*}{\partial \boldsymbol{v}^2} = \boldsymbol{Z}^T\boldsymbol{\Omega Z} + \boldsymbol{Q}.$$

where

$$\boldsymbol{Q} = -\frac{\partial^2 \ell_{2i}}{\partial \boldsymbol{v}^2}.$$

A1.2 Maximum hierarchical likelihood estimator (MHLE) of $\boldsymbol{\theta}$

The first order and second-order derivatives of the hierarchical likelihood(Ha, Lee et al. 2001) are

$$\frac{\partial h_A^*}{\partial \theta} = \frac{\partial h^*|_{\boldsymbol{\tau}=\hat{\boldsymbol{\tau}}}}{\partial \theta} - \frac{1}{2}\mathrm{trace}\left(\widehat{\boldsymbol{H}}^{-1}\frac{\partial \widehat{\boldsymbol{H}}}{\partial \theta}\right),$$

where $\widehat{\boldsymbol{H}} = \boldsymbol{H}\big|_{\boldsymbol{\tau}=\hat{\boldsymbol{\tau}}}$. The observed information matrix $-\frac{\partial^2 h_A^*}{\partial \theta^2}$ for the frailty parameter θ is,

$$\frac{\partial^2 h_A^*}{\partial \theta^2} = \frac{\partial h^*|_{\boldsymbol{\tau}=\hat{\boldsymbol{\tau}}}}{\partial \theta^2} - \frac{\partial}{\partial \theta}\left\{\frac{1}{2}\mathrm{trace}\left(\widehat{\boldsymbol{H}}^{-1}\frac{\partial \widehat{\boldsymbol{H}}}{\partial \theta}\right)\right\},$$

$$-\frac{\partial^2 h_A^*}{\partial\theta^2} = -\frac{\partial h^*|_{\boldsymbol{\tau}=\hat{\boldsymbol{\tau}}}}{\partial\theta^2} + \frac{1}{2}\operatorname{trace}\left(-\widehat{\boldsymbol{H}}^{-1}\frac{\partial\widehat{\boldsymbol{H}}}{\partial\theta}\widehat{\boldsymbol{H}}^{-1}\frac{\partial\widehat{\boldsymbol{H}}}{\partial\theta} + \widehat{\boldsymbol{H}}^{-1}\frac{\partial^2\widehat{\boldsymbol{H}}}{\partial\theta^2}\right).$$

The partial derivation of $\frac{\partial\widehat{\boldsymbol{H}}}{\partial\theta}$,

$$\frac{\partial\widehat{\boldsymbol{H}}}{\partial\theta} = \begin{pmatrix} \boldsymbol{X}^T\widehat{\boldsymbol{\Omega}}'\boldsymbol{X} & \boldsymbol{X}^T\widehat{\boldsymbol{\Omega}}'\boldsymbol{Z} \\ \boldsymbol{Z}^T\widehat{\boldsymbol{\Omega}}'\boldsymbol{X} & \boldsymbol{Z}^T\widehat{\boldsymbol{\Omega}}'\boldsymbol{Z} + \boldsymbol{Q}' \end{pmatrix}.$$

where $\widehat{\boldsymbol{\Omega}}' = \frac{\partial\widehat{\boldsymbol{\Omega}}}{\partial\theta}$. Since $\widehat{\boldsymbol{\Omega}}$ does not depend on θ, it follows that partial derivative is,

$$\frac{\partial\widehat{\boldsymbol{\Omega}}}{\partial\theta} = 0.$$

Therefore,

$$\frac{\partial\widehat{\boldsymbol{H}}}{\partial\theta} = \begin{pmatrix} 0 & 0 \\ 0 & \boldsymbol{Q}' \end{pmatrix},$$

where and $\boldsymbol{Q}' = \frac{\partial\boldsymbol{Q}}{\partial\theta}$ is $G \times G$ diagonal matrix.

Similarly,

$$\frac{\partial^2\widehat{\boldsymbol{H}}}{\partial\theta^2} = \begin{pmatrix} 0 & 0 \\ 0 & \boldsymbol{Q}'' \end{pmatrix},$$

where $\boldsymbol{Q}'' = \frac{\partial^2\boldsymbol{Q}}{\partial\theta^2}$ is $G \times G$ diagonal matrix.

A1.3. Calculation of Jeffreys prior bias-reducing function

We have

$$\ell_{2i} = [\boldsymbol{v} - \exp(\boldsymbol{v})]\theta^{-1} - \log\Gamma\left(\frac{1}{\theta}\right) - \theta^{-1}\log\theta.$$

The first-order derivative is

$$\frac{\partial\ell_{2i}}{\partial\theta} = \theta^{-2}\left[-(\boldsymbol{v} - e^{\,\boldsymbol{v}}) + \psi^{(0)}(\frac{1}{\theta}) + (\log\theta - 1)\right],$$

where $\psi^{(0)} = \frac{\Gamma'\left(\frac{1}{\theta}\right)}{\Gamma\left(\frac{1}{\theta}\right)}$ is the digamma function.

The second-order derivative is

$$-\frac{\partial^2 \ell_{2i}}{\partial \theta^2} = 2\theta^{-3}\left[-(\boldsymbol{v} - e^{\boldsymbol{v}}) + \psi^{(0)}\left(\frac{1}{\theta}\right) + 0.5\theta^{-1}\psi^{(1)}\left(\frac{1}{\theta}\right) + \log\theta - 1.5\right].$$

where $\psi^{(1)} = \frac{\partial \psi^{(0)}}{\partial \theta}$ is the trigamma function.

The expected information matrix is

$$I(\theta) = E\left(-\frac{\partial^2 \ell_{2i}}{\partial \theta^2}\right).$$

Note that if V_i. is independent and identically distributed log-gamma with mean $\psi^{(0)}\left(\frac{1}{\theta}\right) + \log(\theta)$ then e^{V_i} follows gamma with mean 1.

$$E\left(-\frac{\partial^2 l_{2i}}{\partial \theta^2}\right) = 2\theta^{-3}\left\{-EV + Ee^V + \psi^{(0)}\left(\frac{1}{\theta}\right) + 0.5\theta^{-1}\psi^{(1)}\left(\frac{1}{\theta}\right) + \log\theta - 1.5\right\},$$

$$= 2\theta^{-3}\left\{-\left(\psi^{(0)}\left(\frac{1}{\theta}\right) + \log(\theta)\right) + 1 + \psi^{(0)}\left(\frac{1}{\theta}\right) + 0.5\theta^{-1}\psi^{(1)}\left(\frac{1}{\theta}\right) + \log\theta - 1.5\right\},$$

$$= 2\theta^{-3}\left\{0.5\theta^{-1}\psi^{(1)}\left(\frac{1}{\theta}\right) - 0.5\right\},$$

$$I(\theta) = \theta^{-4}\psi^{(1)}\left(\frac{1}{\theta}\right) - \theta^{-3}.$$

Therefore, Jeffreys's prior bias-reducing function for the gamma frailty is

$$M(\theta) = |I(\theta)|^{\frac{1}{2}} = \left|\theta^{-4}\psi^{(1)}\left(\frac{1}{\theta}\right) - \theta^{-3}\right|^{\frac{1}{2}}.$$

A1.4 Calculation of the conditional AIC (cAIC)

The conditional AIC is given by

$$cAIC = -2\ell_p + 2df_c,$$

where

$$\ell_p = \sum_{ij} \delta_{ij}\left(\boldsymbol{x}_{ij}^T\widehat{\boldsymbol{\beta}} + \hat{v}\right) - \sum_k d_k \log\left\{\sum_{(i,j)\in R_{(k)}} \exp\left(\boldsymbol{x}_{ij}^T\widehat{\boldsymbol{\beta}} + \hat{v}\right)\right\},$$

and

$$df_c = df_c(\boldsymbol{\beta}, \boldsymbol{v}, \theta) = trace\left(\boldsymbol{H}^{-1}\boldsymbol{H}_p'\right)$$

is an "effective degree of freedom adjustment" for estimating the fixed and random effects computed using the Hessian matrices, $\boldsymbol{H} = -\frac{\partial^2 h^*}{\partial \boldsymbol{\tau}^2}$ and $\boldsymbol{H}_p' = -\frac{\partial^2 \ell_p}{\partial \boldsymbol{\tau}^2}$ with $\boldsymbol{\tau} = (\boldsymbol{\beta}^T, \boldsymbol{v}^T)^T$.(Ha, Jeong et al. 2017) In the Cox PH model without frailty, a degree of freedom df_c becomes the number of the fixed effects, *p*, i.e., the dimension of $\boldsymbol{\beta}$.

Appendix 2 (total derivative approach)

A2.1. Total derivative approach estimation of Lognormal frailty parameter

The logarithm of the normal distribution is

$$\ell_{2i} = -\frac{\log(2\pi\theta)}{2} - \frac{v_i^2}{2\theta}.$$

The first partial derivative with respect to v_i

$$\frac{\partial}{\partial v_i}\ell_{2i}|_{\boldsymbol{\tau}=\hat{\boldsymbol{\tau}}} \quad = -\frac{\hat{v}_i}{\theta}.$$

The negative second partial derivation is

$$\mathrm{diag}\{\boldsymbol{Q}\} = -\frac{\partial^2}{\partial v_i^2}\ell_{2i}|_{\boldsymbol{\tau}=\hat{\boldsymbol{\tau}}} = \frac{1}{\theta}.$$

The total derivative of the first term in (3.17) is

$$\frac{\partial}{\partial\theta}h^*|_{\tau=\hat{\tau}} = \sum_{i=1}^{G}\frac{\partial}{\partial\theta}\left\{-\frac{\log(2\pi\theta)}{2} - \frac{v_i^2}{2\theta}\right\},$$

$$\frac{\partial h^*|_{\tau=\hat{\tau}}}{\partial\theta} = \sum_{i=1}^{G}\left(-\frac{1}{2\theta} + \frac{v_i^2}{2\theta^2}\right).$$

The derivation of $-\frac{\partial h^*|_{\tau=\hat{\tau}}}{\partial\theta^2}$ in (3.18), is

$$-\frac{\partial h^*|_{\tau=\hat{\tau}}}{\partial \theta^2} = -\sum_{i=1}^{G}\left\{\frac{\partial}{\partial \theta}\left(-\frac{1}{2\theta}+\frac{\hat{v}_i^2}{2\theta^2}\right)\right\} - \left(\frac{\partial^2 h^*|_{\boldsymbol{\tau}=\hat{\boldsymbol{\tau}}}}{\partial \boldsymbol{v}\partial \theta}\right)\left(\frac{\partial \hat{\boldsymbol{v}}}{\partial \theta}\right),$$

$$-\frac{\partial h^*|_{\tau=\hat{\tau}}}{\partial \theta^2} = \sum_{i=1}^{G}\left(-\frac{1}{2\theta^2}+\frac{v_i^2}{\theta^3}\right) - \left[\frac{\partial}{\partial v_i}\left(-\frac{1}{2\theta}+\frac{v_i^2}{2\theta^2}\right)\right]\left(\frac{\partial \hat{\boldsymbol{v}}}{\partial \theta}\right),$$

$$= \sum_{i=1}^{G}\left(-\frac{1}{2\theta^2}+\frac{\hat{v}_i^2}{\theta^3}\right) - \frac{\hat{v}_i}{\theta^2}\left(\frac{\partial \hat{\boldsymbol{v}}}{\partial \theta}\right),$$

where

$$\left(\frac{\partial \hat{\boldsymbol{v}}}{\partial \theta}\right) = \left(\boldsymbol{Z}^T\widehat{\boldsymbol{W}}\boldsymbol{Z}+\boldsymbol{Q}\right)^{-\mathbf{1}}\left(\left.\frac{\partial^2 h^*}{\partial \boldsymbol{v}\partial \theta}\right|_{\boldsymbol{\tau}=\hat{\boldsymbol{\tau}}}\right),$$

$$= \left(\boldsymbol{Z}^T\widehat{\boldsymbol{W}}\boldsymbol{Z}+\boldsymbol{Q}\right)^{-\mathbf{1}}\left(\frac{\hat{v}_i}{\theta^2}\right).$$

The second derivation of $\frac{\partial^2 \hat{\boldsymbol{v}}}{\partial \theta^2}$ is

$$\frac{\partial^2 \hat{\boldsymbol{v}}}{\partial \theta^2} = -\left(\boldsymbol{Z}^T\widehat{\boldsymbol{W}}\boldsymbol{Z}+\widehat{\boldsymbol{Q}}\right)^{-1}\left(\boldsymbol{Z}^T\widehat{\boldsymbol{W}}'\boldsymbol{Z}+\widehat{\boldsymbol{Q}}'\right)\left(\boldsymbol{Z}^T\widehat{\boldsymbol{W}}\boldsymbol{Z}+\widehat{\boldsymbol{Q}}\right)^{-1}\left(\frac{\hat{v}_i}{\theta^2}\right) - 2\left(\boldsymbol{Z}^T\widehat{\boldsymbol{W}}\boldsymbol{Z}+\widehat{\boldsymbol{Q}}\right)^{-\mathbf{1}}\left(\frac{\hat{v}_i}{\theta^3}\right).$$

Since $\boldsymbol{Q}$ does not depends on $\boldsymbol{v}$ no total derivative should be calculated. We have

$$\boldsymbol{Q}' = \operatorname{diag}\left\{-\frac{1}{\theta^2}\right\}_{G\times G},$$

and

$$\boldsymbol{Q}'' = \operatorname{diag}\{2\theta^{-3}\}_{G\times G}.$$

A2.2. Total derivative approach estimation of Gamma frailty parameter

$$\ell_{2_i} = (v_i - e^{v_i})\theta^{-1} - \log\Gamma\left(\frac{1}{\theta}\right) - \theta^{-1}\log\theta.$$

The first partial derivative with respect to v_i

$$\left.\frac{\partial}{\partial v_i}\ell_{2_i}\right|_{\boldsymbol{\tau}=\hat{\boldsymbol{\tau}}} = \left(1-e^{\hat{v}_i}\right)\theta^{-1}.$$

The negative second partial derivation is a diagonal matrix with ith element

$$\operatorname{diag}\{\boldsymbol{Q}\} = -\frac{\partial^2}{\partial v_i^2}h^*|_{\boldsymbol{\tau}=\hat{\boldsymbol{\tau}}} = -\frac{\partial}{\partial v_i}\left\{\left(1-e^{\hat{v}_i}\right)\theta^{-1}\right\},$$

$$\boldsymbol{Q} = \theta^{-1}e^{\hat{v}_i}.$$

The total derivative of the first term in (3.17) is

$$\frac{\partial h^*|_{\boldsymbol{\tau}=\hat{\boldsymbol{\tau}}}}{\partial\theta} = \sum_{i=1}\frac{\partial}{\partial\theta}\left\{\left(\hat{v}_i - e^{\hat{v}_i}\right)\theta^{-1} - \log\Gamma\left(\frac{1}{\theta}\right) - \theta^{-1}\log\theta\right\},$$

$$= -\left(\hat{v}_i - e^{\hat{v}_i}\right)\theta^{-2} + \theta^{-2}\frac{\Gamma'\left(\frac{1}{\theta}\right)}{\Gamma\left(\frac{1}{\theta}\right)} + \theta^{-2}\log\theta - \theta^{-2},$$

$$\frac{\partial h^*|_{\boldsymbol{\tau}=\hat{\boldsymbol{\tau}}}}{\partial\theta} = -\left(\hat{v}_i - e^{\hat{v}_i}\right)\theta^{-2} + \theta^{-2}\psi^{(0)}\left(\frac{1}{\theta}\right) + \theta^{-2}\log\theta - \theta^{-2},$$

where $\psi^{(0)} = \frac{\Gamma'\left(\frac{1}{\theta}\right)}{\Gamma\left(\frac{1}{\theta}\right)}$ is the digamma function.

The second total derivation of $-\frac{\partial h^*|_{\boldsymbol{\tau}=\hat{\boldsymbol{\tau}}}}{\partial\theta^2}$ is

$$-\frac{\partial h^*|_{\boldsymbol{\tau}=\hat{\boldsymbol{\tau}}}}{\partial\theta^2} = \sum_{i=1} -\frac{\partial^2\ell_{2i}}{\partial\theta^2} - \left(\frac{\partial^2 h^*|_{\boldsymbol{\tau}=\hat{\boldsymbol{\tau}}}}{\partial\boldsymbol{v}\partial\theta}\right)\left(\frac{\partial\hat{\boldsymbol{v}}}{\partial\theta}\right),$$

$$= \sum_{i=1}\left[-\frac{\partial}{\partial\theta}\left\{-\left(\hat{v}_i - e^{\hat{v}_i}\right)\theta^{-2} + \theta^{-2}\psi^{(0)}\left(\frac{1}{\theta}\right) + \theta^{-2}(\log\theta - 1)\right\}\right]$$

$$-\left[\frac{\partial}{\partial v_i}\left\{-\left(\hat{v}_i - e^{\hat{v}_i}\right)\theta^{-2} + \theta^{-2}\psi^{(0)}\left(\frac{1}{\theta}\right) + \theta^{-2}(\log\theta - 1)\right\}\left(\frac{\partial\hat{\boldsymbol{v}}}{\partial\theta}\right)\right].$$

$$= \sum_{i=1}\left[2\theta^{-3}\left\{-\left(\hat{v}_i - e^{\hat{v}_i}\right) + \psi^{(0)}\left(\frac{1}{\theta}\right) + 0.5\theta^{-1}\psi^{(1)}\left(\frac{1}{\theta}\right) + \log(\theta) - 1.5\right\}\right] + \left(1 - e^{\hat{\boldsymbol{v}}}\right)\theta^{-2}\left(\frac{\partial\hat{\boldsymbol{v}}}{\partial\theta}\right),$$

where $\psi^{(1)} = \frac{\partial\psi^{(0)}}{\partial\theta}$ is the trigamma function and

$$\left(\frac{\partial\hat{\boldsymbol{v}}}{\partial\theta}\right) = \left(\boldsymbol{Z}^T\widehat{\boldsymbol{W}}\boldsymbol{Z} + \widehat{\boldsymbol{Q}}\right)^{-\mathbf{1}}\left(\left.\frac{\partial^2 h^*}{\partial\boldsymbol{v}\partial\theta}\right|_{\boldsymbol{\tau}=\hat{\boldsymbol{\tau}}}\right),$$

$$= -\left(\boldsymbol{Z}^T\widehat{\boldsymbol{W}}\boldsymbol{Z} + \widehat{\boldsymbol{Q}}\right)^{-\mathbf{1}}\left(\left(1 - e^{\hat{\boldsymbol{v}}}\right)\theta^{-2}\right).$$

The second derivation of $\frac{\partial^2\hat{\boldsymbol{v}}}{\partial\theta^2}$

$$\frac{\partial^2\hat{\boldsymbol{v}}}{\partial\theta^2} = -\left(\boldsymbol{Z}^T\widehat{\boldsymbol{W}}\boldsymbol{Z} + \widehat{\boldsymbol{Q}}\right)^{-1}\left(\boldsymbol{Z}^T\widehat{\boldsymbol{W}}'\boldsymbol{Z} + \widehat{\boldsymbol{Q}}'\right)\left(\boldsymbol{Z}^T\widehat{\boldsymbol{W}}\boldsymbol{Z} + \widehat{\boldsymbol{Q}}\right)^{-1}\left(\left.\frac{\partial^2 h^*}{\partial\boldsymbol{v}\partial\theta}\right|_{\boldsymbol{\tau}=\hat{\boldsymbol{\tau}}}\right)$$

$$+ \left(\boldsymbol{Z}^T\widehat{\boldsymbol{W}}\boldsymbol{Z} + \widehat{\boldsymbol{Q}}\right)^{-\mathbf{1}}\left(\left.\frac{\partial^3 h^*}{\partial\boldsymbol{v}\partial\theta^2}\right|_{\boldsymbol{\tau}=\hat{\boldsymbol{\tau}}}\right),$$

$$= -\left(\boldsymbol{Z}^T\widehat{\boldsymbol{W}}\boldsymbol{Z} + \widehat{\boldsymbol{Q}}\right)^{-1}\left(\boldsymbol{Z}^T\widehat{\boldsymbol{W}}'\boldsymbol{Z} + \widehat{\boldsymbol{Q}}'\right)\left(\boldsymbol{Z}^T\widehat{\boldsymbol{W}}\boldsymbol{Z} + \widehat{\boldsymbol{Q}}\right)^{-1}\left(\left(1 - e^{\widehat{\boldsymbol{v}}}\right)\theta^{-2}\right)$$

$$+ \left(\boldsymbol{Z}^T\widehat{\boldsymbol{W}}\boldsymbol{Z} + \widehat{\boldsymbol{Q}}\right)^{-\mathbf{1}}\left(-2\left(1 - e^{\widehat{\boldsymbol{v}}}\right)\theta^{-3}\right).$$

The next step is to calculate the terms in the observed information in (3.18). First, we compute the total derivative of the first term in (3.18).

$$\left.\frac{\partial^3 h^*}{\partial \boldsymbol{v} \partial \theta^2}\right|_{\boldsymbol{\tau}=\widehat{\boldsymbol{\tau}}} = 2\left(1 - e^{\widehat{\boldsymbol{v}}}\right)\theta^{-3}$$

Finally, we calculate the total derivative for $\boldsymbol{Q}$ since it depends on $\boldsymbol{v}$**.** Thus, from

$$\widehat{\boldsymbol{Q}}' = \frac{\partial}{\partial \theta}\boldsymbol{Q}|_{\boldsymbol{\tau}=\widehat{\boldsymbol{\tau}}} + \frac{\partial}{\partial \boldsymbol{v}}\boldsymbol{Q}|_{\boldsymbol{\tau}=\widehat{\boldsymbol{\tau}}}\left(\frac{\partial \widehat{\boldsymbol{v}}}{\partial \theta}\right).$$

$$\widehat{\boldsymbol{Q}}' = \widehat{\boldsymbol{Q}}'_{\theta} + \widehat{\boldsymbol{Q}}'_{\widehat{v}}\left(\frac{\partial \widehat{\boldsymbol{v}}}{\partial \theta}\right).$$

Since $\boldsymbol{Q} = \theta^{-1}e^{\widehat{\boldsymbol{v}}}$, it follows that $\widehat{\boldsymbol{Q}}''_{\widehat{\boldsymbol{v}}} = \boldsymbol{Q}'_{\widehat{\boldsymbol{v}}} = \widehat{\boldsymbol{Q}} = \theta^{-1}e^{\widehat{\boldsymbol{v}}}$ and $\widehat{\boldsymbol{Q}}'_{\theta} = -\theta^{-2}e^{\widehat{\boldsymbol{v}}}$.

The second total derivative of $\frac{\partial^2}{\partial \theta^2}\boldsymbol{Q}|_{\boldsymbol{\tau}=\widehat{\boldsymbol{\tau}}}$ is

$$\widehat{\boldsymbol{Q}}'' = \widehat{\boldsymbol{Q}}''_{\theta} + 2 * \widehat{\boldsymbol{Q}}''_{\widehat{v}\theta}\left(\frac{\partial \widehat{\boldsymbol{v}}}{\partial \theta}\right) + \widehat{\boldsymbol{Q}}'_{\widehat{v}}\left(\frac{\partial^2 \widehat{\boldsymbol{v}}}{\partial \theta^2}\right) + \widehat{\boldsymbol{Q}}''_{\widehat{v}}\left(\frac{\partial \widehat{\boldsymbol{v}}}{\partial \theta}\right)^2.$$

A2.3. Second-order Laplace approximation of Gamma frailty

To reduce the bias further in estimating the dispersion parameters, the second-order approximation $S_\tau(h^*)$ needs to be used. According toHa and Lee (2003), the second-order method is needed when estimating the frailty parameters. The second-order approximation equation $S_\tau(h^*)$ given by

$$S_\tau(h^*) = h_A^* - \left\{\frac{F(h^*)}{24}\right\},$$

where $F = trace(S)|_{(\boldsymbol{\tau}=\widehat{\boldsymbol{\tau}})}$. The ith element of $G \times G$ diagonal matrix S is given by

$$S = 3\frac{\tilde{h}^{(4)}}{\left[\tilde{h}^{(2)}\right]^2} - 5\frac{\left[\tilde{h}^{(3)}\right]^2}{\left[\tilde{h}^{(2)}\right]^3},$$

where $\tilde{h}^{(k)} = \tilde{h}^{(k)}(\tilde{v})$ and $\tilde{h}^{(k)} = -\frac{\partial^k h}{\partial v^k}$. The estimation of the frailty parameter involves solving the socre function $\frac{\partial}{\partial \theta}S_\tau(h^*) = 0$, which involves too many complicated terms.

Appendix 3 (mathematical derivation of the log skew-normal frailty)

A3.1. Joint score functions

The score functions from section 4.3 are

$$\frac{\partial h^*}{\partial \boldsymbol{\beta}} = \boldsymbol{X}^T(\boldsymbol{\delta} - \boldsymbol{\mu}),$$

$$\frac{\partial h^*}{\partial \boldsymbol{v}} = \boldsymbol{Z}^T(\boldsymbol{\delta} - \boldsymbol{\mu}) + \frac{\partial}{\partial \boldsymbol{v}}\ell_{2i}(\boldsymbol{\theta}; v_i).$$

From equation (4.5), we have

$$\frac{\partial \ell_{2i}(\boldsymbol{\theta}; v_i)}{\partial v_i} = \frac{\partial}{\partial v_i}\left\{-\frac{{z_i}^2}{2} + \log[\Phi(az_i)]\right\},$$

since $z_i = \frac{v_i}{\sigma}\sqrt{Var(\mathcal{Z})}$, then $\frac{\partial}{\partial v_i}\left(\frac{{z_i}^2}{2}\right) = z_i \frac{\partial z_i}{\partial v_i} = \frac{z_i}{\sigma}\sqrt{Var(\mathcal{Z})}$ and

$$\frac{\partial}{\partial v_i}\log[\Phi(az_i)] = \frac{\Phi'(az_i) * (az_i)'}{\Phi(az_i)},$$

$$= \frac{a}{\sigma}\sqrt{Var(\mathcal{Z})}\frac{\phi(az_i)}{\Phi(az_i)},$$

on letting

$$M = \frac{\phi(az_i)}{\Phi(az_i)},$$

We have

$$\frac{\partial}{\partial v_i}\log[\Phi(az_i)] = M * \frac{a}{\sigma}\sqrt{Var(\mathcal{Z})}.$$

Therefore,

$$R = -\frac{z_i}{\sigma}\sqrt{Var(\mathcal{Z})} + M * \frac{a}{\sigma}\sqrt{Var(\mathcal{Z})},$$

$$R = \frac{1}{\sigma}\sqrt{Var(\mathcal{Z})}\{-z_i + aM\}.$$

A3.2. Information matrix

Next, we show the direct calculation of $\boldsymbol{Q}$ in the observed information matrix $\boldsymbol{H}$ of $\boldsymbol{\beta}$ and $\boldsymbol{v} = (v_1, \dots, v_G)^T$.

$$\boldsymbol{H} = \boldsymbol{H}(\boldsymbol{\beta}, \boldsymbol{v}) = \begin{pmatrix} \boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X} & \boldsymbol{X}^T\boldsymbol{W}\boldsymbol{Z} \\ \boldsymbol{Z}^T\boldsymbol{W}\boldsymbol{X} & \boldsymbol{Z}^T\boldsymbol{W}\boldsymbol{Z} + \boldsymbol{Q} \end{pmatrix}$$

where $\boldsymbol{Q}$ is a $G \times G$ diagonal matrix with the ith element given by the negative second derivative of the log of the joint density function for all random effects with respect to the vector v_i,

$$-\frac{\partial^2}{\partial v_i^2}\ell_{i2}(\theta; v_i) = -\frac{\partial}{\partial v_i}R,$$

$$= -\frac{\partial}{\partial v_i}\left\{\frac{1}{\sigma}\sqrt{Var(\mathcal{Z})}\{-z_i + aM\}\right\},$$

$$= -\left\{\frac{1}{\sigma}\sqrt{Var(\mathcal{Z})}\left\{-\frac{1}{\sigma}\sqrt{Var(\mathcal{Z})} + a * M_v'\right\}\right\}.$$

where

$$M_v' = \frac{\partial}{\partial v}M = \left[\frac{\phi(az_i)}{\Phi(az_i)}\right]' = \frac{\Phi(az_i) \times [\phi(az_i)]' - \phi(az_i)[\Phi(az_i)]'}{[\Phi(az_i)]^2},$$

But

$$[\phi(az_i)]' = \frac{\partial}{\partial v}\phi(az_i) = \frac{1}{\sqrt{2\pi}}e^{-\frac{(az_i)^2}{2}} = -\frac{a^2 z_i}{\sigma}\sqrt{Var(\mathcal{Z})}\phi(az_i).$$

Therefore

$$M_v' = -\frac{a^2 z_i}{\sigma}\sqrt{Var(\mathcal{Z})}\frac{\phi(az_i)}{\Phi(az_i)} - \frac{a}{\sigma}\sqrt{Var(\mathcal{Z})}\left[\frac{\phi(az_i)}{\Phi(az_i)}\right]^2,$$

$$M_v' = -\frac{a}{\sigma}\sqrt{Var(\mathcal{Z})}\{az_iM + M^2\}.$$

$$= \frac{\partial^2}{\partial v_i^2}\ell_{i2}(\theta; v_i) = -\left\{\frac{1}{\sigma}\sqrt{Var(\mathcal{Z})}\left\{-\frac{1}{\sigma}\sqrt{Var(\mathcal{Z})} + a * \left[-\frac{a}{\sigma}\sqrt{Var(\mathcal{Z})}(az_iM + M^2)\right]\right\}\right\},$$

The ith element of the diagonal matrix $\boldsymbol{Q}$ is given by

$$\mathrm{diag}\{\boldsymbol{Q}\} = \frac{Var(\mathcal{Z})}{\sigma^2}[1 + a^2(az_iM + M^2)].$$

A3.3. Mathematical derivations of frailty and shape parameters

From equation (4.9) the first and second derivative of the adjusted profile likelihood are given by

$$\frac{\partial h_A^*}{\partial \boldsymbol{\theta}} = \frac{\partial h^*|_{\tau=\hat{\tau}}}{\partial \boldsymbol{\theta}} - \frac{1}{2} trace\left(\widehat{\boldsymbol{H}}^{-1}\frac{\partial \widehat{\boldsymbol{H}}}{\partial \boldsymbol{\theta}}\right), \qquad (A3.1)$$

and

$$-\frac{\partial^2 h_A^*}{\partial \boldsymbol{\theta}^2} = -\frac{\partial h^*|_{\tau=\hat{\tau}}}{\partial \boldsymbol{\theta}^2} + \frac{1}{2} trace\left(-\widehat{\boldsymbol{H}}^{-1}\frac{\partial \widehat{\boldsymbol{H}}}{\partial \boldsymbol{\theta}}\widehat{\boldsymbol{H}}^{-1}\frac{\partial \widehat{\boldsymbol{H}}}{\partial \boldsymbol{\theta}} + \widehat{\boldsymbol{H}}^{-1}\frac{\partial^2 \widehat{\boldsymbol{H}}}{\partial \boldsymbol{\theta}^2}\right), \qquad (A3.2)$$

respectively. The parameters $\boldsymbol{\theta} = (\sigma, a)^T$ can be estimated via the Newton-Raphson algorithm given by

$$\begin{pmatrix}\hat{\sigma}^{(k+1)}\\ \hat{a}^{(k+1)}\end{pmatrix} = \begin{pmatrix}\hat{\sigma}^{(k)}\\ \hat{a}^{(k)}\end{pmatrix} + \left(\mathcal{J}(\theta)^{-1}\mathcal{S}(\theta)\right)\Big|_{(\sigma,a)=\left(\hat{\sigma}^{(k)},\hat{a}^{(k)}\right)},$$

where $\boldsymbol{\mathcal{S}}(\boldsymbol{\theta}) = \frac{\partial h_A^*}{\partial \boldsymbol{\theta}} = \left(\frac{\partial h_A^*}{\partial \sigma}, \frac{\partial h_A^*}{\partial a}\right)^T$ and $\boldsymbol{\mathcal{J}}(\boldsymbol{\theta})^{-1} = -\frac{\partial^2 h_A^*}{\partial \boldsymbol{\theta}^2} = \begin{bmatrix} -\frac{\partial^2 h_A^*}{\partial \sigma^2} & -\frac{\partial h_A^*}{\partial \sigma \partial a} \\ -\frac{\partial h_A^*}{\partial a \partial \sigma} & -\frac{\partial^2 h_A^*}{\partial a^2}\end{bmatrix}$.

The partial derivation of first term $\frac{\partial}{\partial \theta} h^*|_{\tau=\hat{\tau}}$ in $(A3.1)$ is

$$\frac{\partial h^*|_{\tau=\hat{\tau}}}{\partial \boldsymbol{\theta}} = \sum_{i=1}^{G} \frac{\partial \ell_{i2}(\boldsymbol{\theta}; \hat{v}_i)}{\partial \boldsymbol{\theta}}, \qquad (A3.3)$$

since $\sum_{ij} \ell_{1ij}$ does not involve $\boldsymbol{\theta}$.

The partial derivative of the second term $\frac{\partial \hat{H}}{\partial \theta}$ in $(A3.1)$ is,

$$\frac{\partial \widehat{\boldsymbol{H}}}{\partial \boldsymbol{\theta}} = \begin{pmatrix} \boldsymbol{X}^T\widehat{\boldsymbol{W}}'\boldsymbol{X} & \boldsymbol{X}^T\widehat{\boldsymbol{W}}'\boldsymbol{Z} \\ \boldsymbol{Z}^T\widehat{\boldsymbol{W}}'\boldsymbol{X} & \boldsymbol{Z}^T\widehat{\boldsymbol{W}}'\boldsymbol{Z} + \boldsymbol{Q}'\end{pmatrix}, \qquad (A3.4)$$

where $\widehat{\boldsymbol{W}}' = \frac{\partial \widehat{\boldsymbol{W}}}{\partial \boldsymbol{\theta}}$ and $\boldsymbol{Q}' = \frac{\partial \boldsymbol{Q}}{\partial \boldsymbol{\theta}}$. Since $\widehat{\boldsymbol{W}}$ does not depend on $\boldsymbol{\theta}$, it follows that the partial derivative is,

$$\frac{\partial \widehat{\boldsymbol{H}}}{\partial \boldsymbol{\theta}} = \begin{pmatrix} \mathbf{0}_{p\times p} & \mathbf{0}_{p\times G} \\ \mathbf{0}_{G\times p} & \boldsymbol{Q}'\end{pmatrix} \qquad (A3.5)$$

Now, we calculate the second derivatives. That is the term in the observed information matrix in $(A3.2)$. We have that

$$-\frac{\partial h^*|_{\tau=\hat{\tau}}}{\partial \boldsymbol{\theta}^2} = -\sum_{i=1}^{G} \frac{\partial^2 \ell_{i2}(\boldsymbol{\theta}; \hat{v}_i)}{\partial \boldsymbol{\theta}^2}. \qquad (A3.6)$$

The last term needed to calculate in $(A3.2)$ is

$$\frac{\partial^2 \widehat{\boldsymbol{H}}}{\partial \boldsymbol{\theta}^2} = \begin{pmatrix} \boldsymbol{X}^T \widehat{\boldsymbol{W}}'' \boldsymbol{X} & \boldsymbol{X}^T \widehat{\boldsymbol{W}}'' \boldsymbol{Z} \\ \boldsymbol{Z}^T \widehat{\boldsymbol{W}}'' \boldsymbol{X} & \boldsymbol{Z}^T \widehat{\boldsymbol{W}}'' \boldsymbol{Z} + \widehat{\boldsymbol{Q}}'' \end{pmatrix} = \begin{pmatrix} \mathbf{0}_{p\times p} & \mathbf{0}_{p\times G} \\ \mathbf{0}_{G\times p} & \widehat{\boldsymbol{Q}}'' \end{pmatrix} \qquad (A3.7)$$

where $\widehat{\boldsymbol{W}}'' = \frac{\partial^2}{\partial \boldsymbol{\theta}^2} \boldsymbol{W}|_{\tau=\hat{\tau}} = \mathbf{0}$ and $\widehat{\boldsymbol{Q}}'' = \frac{\partial^2}{\partial \boldsymbol{\theta}^2} \boldsymbol{Q}|_{\tau=\hat{\tau}}$.

A3.3.1 Mathematical derivations of terms $\left[\frac{\partial h_A^*}{\partial \sigma}\right]$ and $\left[-\frac{\partial h_A^*}{\partial \sigma^2}\right]$

From $(A3.3)$, we have that $\frac{\partial h_A^*}{\partial \sigma}$ is

$$\frac{\partial h^*|_{\tau=\hat{\tau}}}{\partial \sigma} = \sum_{i=1}^{G} \frac{\partial \ell_{i2}(\theta; \hat{v})}{\partial \sigma}$$

But

$$\frac{\partial}{\partial \sigma} \ell_{i2}(\boldsymbol{\theta}; v)|_{\tau=\hat{\tau}} = \frac{\partial}{\partial \sigma} \left\{ \left(\log 2 - \frac{1}{2} \log(2\pi) - \frac{z_i^2}{2} + \log\big(\Phi(a z_i)\big) + \log\left\{ \frac{1}{\sigma} \sqrt{Var(\mathcal{Z})} \right\} \bigg|_{\tau=\hat{\tau}} \right) \right\},$$

Note:

$$\hat{z}_i = \frac{\hat{v}_i}{\sigma} \sqrt{Var(\mathcal{Z})}$$

Thus

$$\frac{\partial}{\partial \sigma} z_i|_{\tau=\hat{\tau}} = \frac{\partial}{\partial \sigma} \left\{ \left(\frac{v_i}{\sigma} \sqrt{Var(\mathcal{Z})} \bigg|_{\tau=\hat{\tau}} \right) \right\},$$

$$= -\frac{\hat{v}_i}{\sigma^2} \sqrt{Var(\mathcal{Z})},$$

$$\frac{\partial}{\partial \sigma} z_i|_{\tau=\hat{\tau}} = -\frac{\hat{z}_i}{\sigma},$$

and

$$\frac{\partial}{\partial\sigma} z_i^2\Big|_{\tau=\hat{\tau}} = 2\hat{z}_i\left(-\frac{\hat{z}_i}{\sigma}\right),$$

$$\frac{\partial}{\partial\sigma}\hat{z}_i^2 = -\frac{2\hat{z}_i^2}{\sigma}.$$

Also

$$\frac{\partial}{\partial\sigma}\left(\log\big(\Phi(az_i)\big)\Big|_{\tau=\hat{\tau}}\right) = \left(\frac{\Phi'(az_i) * (az_i)'}{\Phi(az_i)}\bigg|_{\tau=\hat{\tau}}\right),$$

$$= \frac{\phi(a\hat{z}_i) * \left(-\frac{a\hat{z}_i}{\sigma}\right)}{\Phi(a\hat{z}_i)},$$

$$\frac{\partial}{\partial\sigma}\left(\log\big(\Phi(az_i)\big)\Big|_{\tau=\hat{\tau}}\right) = -\widehat{\mathrm{M}} * \left(\frac{a\hat{z}_i}{\sigma}\right),$$

where $\widehat{M} = \frac{\phi(a\hat{z}_i)}{\Phi(a\hat{z}_i)}$.

Hence

$$\frac{\partial}{\partial\sigma}\ell_{i2}(\theta; v)|_{\tau=\hat{\tau}} = \frac{\hat{z}_i^2}{\sigma} - \widehat{\mathrm{M}} * \left(\frac{a\hat{z}_i}{\sigma}\right) - \frac{1}{\sigma}.$$

and

$$\frac{\partial h^*|_{\tau=\hat{\tau}}}{\partial\sigma} = \sum_{i=1}^{G}\left(\frac{\hat{z}_i^2}{\sigma} - \widehat{\mathrm{M}} * \left(\frac{a\hat{z}_i}{\sigma}\right) - \frac{1}{\sigma}\right).$$

The second partial derivation from $(A3.6)$ is

$$-\frac{\partial h^*|_{\tau=\hat{\tau}}}{\partial\sigma^2} = -\sum_{i=1}^{G}\frac{\partial^2\ell_{i2}(\theta; \hat{v}_i)}{\partial\sigma^2},$$

$$= -\left\{\sum_{i=1}^{G}\left(- \frac{3\hat{z}_i^2}{\sigma^2} - \widehat{M}'_\sigma * \left(\frac{a\hat{z}_i}{\sigma}\right) - \widehat{M} * \left(\frac{-2a\hat{z}_i}{\sigma^2}\right) + \frac{1}{\sigma^2}\right)\right\},$$

$$-\frac{\partial h^*|_{\tau=\hat{\tau}}}{\partial \sigma^2} = \sum_{i=1}^{G}\left(\frac{3\hat{z}_i^2}{\sigma^2} + \widehat{M}'_\sigma * \left(\frac{a\hat{z}_i}{\sigma}\right) - \widehat{M} * \left(\frac{2a\hat{z}_i}{\sigma^2}\right) - \frac{1}{\sigma^2}\right),$$

where

$$\widehat{M}'_\sigma = \frac{\partial}{\partial \sigma} M|_{\tau=\hat{\tau}} = \left(\frac{[\phi(az_i)]' \times \Phi(az_i) - [\Phi(az_i)]' \times \phi(az_i)}{[\Phi(az_i)]^2}\bigg|_{\tau=\hat{\tau}}\right).$$

First consider

$$[\phi(a\hat{z}_i)]' = \frac{\partial}{\partial \sigma}\phi(az_i)\,|_{\tau=\hat{\tau}},$$

$$= \frac{\partial}{\partial \sigma}\left\{\text{constant} * \exp\left(-\frac{a^2 z_i^2}{2}\right)\bigg|_{\tau=\hat{\tau}}\right\},$$

$$= \text{contant} * \frac{a^2\hat{z}_i^2}{\sigma}\exp\left(-\frac{a^2\hat{z}_i^2}{2}\right),$$

$$[\phi(a\hat{z}_i)]' = \frac{a^2\hat{z}_i^2}{\sigma}\phi(a\hat{z}_i).$$

Therefore,

$$\widehat{M}'_\sigma = \frac{\frac{a^2\hat{z}_i^2}{\sigma}\phi(a\hat{z}_i)\Phi(a\hat{z}_i) + \frac{a\hat{z}_i}{\sigma} \times [\phi(a\hat{z}_i)]^2}{[\Phi(a\hat{z}_i)]^2},$$

$$= \frac{\frac{a^2\hat{z}_i^2}{\sigma}\phi(a\hat{z}_i)}{\Phi(a\hat{z}_i)} + \frac{\frac{a\hat{z}_i}{\sigma} \times [\phi(a\hat{z}_i)]^2}{[\Phi(a\hat{z}_i)]^2},$$

$$\widehat{M}'_\sigma = \frac{a\hat{z}_i}{\sigma}\left\{a\hat{z}_i\widehat{M} + \widehat{M}^2\right\}.$$

Next, we compute the $\widehat{\boldsymbol{Q}}'_\sigma = \frac{\partial \boldsymbol{Q}}{\partial \sigma}\Big|_{\tau=\hat{\tau}}$ term in $(A3.5)$

$$\frac{\partial \boldsymbol{Q}}{\partial \sigma}\bigg|_{\tau=\hat{\tau}} = \frac{\partial}{\partial \sigma}\left\{\left(\frac{Var(\mathcal{Z})}{\sigma^2}\{1 + a^2(az_iM + M^2)\}\bigg|_{\tau=\hat{\tau}}\right)\right\}$$

$$\widehat{\boldsymbol{Q}}'_\sigma = -\frac{2}{\sigma^3}Var(\mathcal{Z})\left\{1 + a^2\left(a\hat{z}_i\widehat{M} + \widehat{M}^2\right)\right\} + \frac{a^2Var(\mathcal{Z})}{\sigma^2}\left\{\frac{-a\hat{z}_i}{\sigma}\widehat{M} + a\hat{z}_i\widehat{M}'_\sigma + 2\widehat{M} * \widehat{M}'_\sigma\right\}$$

Finally, we compute $\widehat{\boldsymbol{Q}}''_\sigma = \frac{\partial^2\boldsymbol{Q}}{\partial\sigma^2}\Big|_{\tau=\hat{\tau}}$

$$\frac{\partial^2\boldsymbol{Q}}{\partial\sigma^2} = \frac{\partial}{\partial\sigma}\left[-\frac{2}{\sigma^3}Var(\mathcal{Z})\{1 + a^2(az_iM + M^2)\}\right.$$

$$\left. + \frac{a^2Var(\mathcal{Z})}{\sigma^2}\left\{\frac{-az_i}{\sigma}M + az_iM'_\sigma + 2\widehat{M} * M'_\sigma\right\}\Big|_{\tau=\hat{\tau}}\right],$$

$$= \frac{6}{\sigma^4}Var(\mathcal{Z})\left\{1 + a^2\left(a\hat{z}_i\widehat{M} + \widehat{M}^2\right)\right\} - \frac{2a^2}{\sigma^3}Var(\mathcal{Z})\left\{\frac{-a\hat{z}_i}{\sigma}\widehat{M} + a\hat{z}_i\widehat{M}'_\sigma + 2\widehat{M} * \widehat{M}'_\sigma\right\}$$

$$- \frac{2a^2Var(\mathcal{Z})}{\sigma^3}\left\{\frac{-a\hat{z}_i}{\sigma}\widehat{M} + a\hat{z}_i\widehat{M}'_\sigma + 2\widehat{M} * \widehat{M}'_\sigma\right\}$$

$$+ \frac{a^2Var(\mathcal{Z})}{\sigma^2}\left\{\frac{2a\hat{z}_i}{\sigma^2}\widehat{M} - \frac{a\hat{z}_i}{\sigma}\widehat{M}'_\sigma - \frac{a\hat{z}_i}{\sigma}\widehat{M}'_\sigma + a\hat{z}_i\widehat{M}''_\sigma + 2\left(\widehat{M}'_\sigma\right)^2 + 2\widehat{M} * \widehat{M}''_\sigma\right\},$$

$$\widehat{\boldsymbol{Q}}''_\sigma = \frac{6}{\sigma^4}Var(\mathcal{Z})\left\{1 + a^2\left(a\hat{z}_i\widehat{M} + \widehat{M}^2\right)\right\} - \frac{4a^2}{\sigma^3}Var(\mathcal{Z})\left\{\frac{-a\hat{z}_i}{\sigma}\widehat{M} + a\hat{z}_i\widehat{M}'_\sigma + 2\widehat{M} * \widehat{M}'_\sigma\right\}$$

$$+ \frac{a^2Var(\mathcal{Z})}{\sigma^2}\left\{\frac{2a\hat{z}_i}{\sigma^2}\widehat{M} - 2\frac{a\hat{z}_i}{\sigma}\widehat{M}'_\sigma + a\hat{z}_i\widehat{M}''_\sigma + 2\left(\left(\widehat{M}'_\sigma\right)^2 + \widehat{M} * \widehat{M}''_\sigma\right)\right\},$$

where

$$\widehat{M}''_\sigma = \frac{\partial}{\partial\sigma}M'_\sigma|_{\tau=\hat{\tau}},$$

$$= \frac{\partial}{\partial\sigma}\left[\frac{a\hat{z}_i}{\sigma}\left(a\hat{z}_i\widehat{M} + \widehat{M}^2\right)\Big|_{\tau=\hat{\tau}}\right],$$

$$\widehat{M}''_\sigma = -\frac{2a\hat{z}_i}{\sigma^2}\left(a\hat{z}_i\widehat{M} + \widehat{M}^2\right) + \frac{a\hat{z}_i}{\sigma}\left\{-\frac{a\hat{z}_i}{\sigma}\widehat{M} + a\hat{z}_i\widehat{M}'_\sigma + 2\widehat{M} * \widehat{M}'_\sigma\right\}.$$

A3.3.2 mathematical derivations of terms $\left[\frac{\partial \boldsymbol{h}_A^*}{\partial \boldsymbol{a}}\right]$ and $\left[-\frac{\partial^2 \boldsymbol{h}_A^*}{\partial \boldsymbol{a}^2}\right]$

From $(A3.3)$, we have that $\frac{\partial h_A^*}{\partial a}$ is

$$\frac{\partial h^*|_{\tau=\hat{\tau}}}{\partial a} = \sum_{i=1}^{G}\frac{\partial \ell_{i2}(\theta;\hat{v}_i)}{\partial a}.$$

But

$$\frac{\partial}{\partial a}\ell_{i2}(\boldsymbol{\theta};v_i)|_{\tau=\hat{\tau}} = \frac{\partial}{\partial a}\left\{\left(\log 2 - \frac{1}{2}\log(2\pi) - \frac{\hat{z}_i^2}{2} + \log\big(\Phi(a\hat{z}_i)\big) + \log\left\{\frac{1}{\sigma}\sqrt{Var(\mathcal{Z})}\right\}\Bigg|_{\tau=\hat{\tau}}\right)\right\},$$

$$= \frac{\partial}{\partial a}\left\{\left(-\frac{\hat{z}_i^2}{2} + \log\big(\Phi(a\hat{z}_i)\big) + \frac{1}{2}\log\big(Var(\mathcal{Z})\big)\Bigg|_{\tau=\hat{\tau}}\right)\right\},$$

where

$$\hat{z}_i = \frac{\hat{v}_i}{\sigma}\sqrt{Var(\mathcal{Z})} = \frac{\hat{v}_i}{\sigma}\left(1 - \frac{2a^2}{\pi(1+a^2)}\right)^{\frac{1}{2}}.$$

Then on letting $A = \frac{\partial}{\partial a} z_i|_{\tau=\hat{\tau}}$,

$$A = \frac{1}{2}\frac{\hat{v}_i}{\sigma}\left(1 - \frac{2a^2}{\pi(1+a^2)}\right)^{-\frac{1}{2}}\left(\frac{-4a}{\pi(1+a^2)^2}\right),$$

$$A = -\frac{\hat{v}_i}{\sigma}\left(1 - \frac{2a^2}{\pi(1+a^2)}\right)^{-\frac{1}{2}}\left(\frac{2a}{\pi(1+a^2)^2}\right).$$

and

$$\frac{\partial}{\partial a} z_i^2\Big|_{\tau=\hat{\tau}} = 2z_i\frac{\partial}{\partial a}z_i\Bigg|_{\tau=\hat{\tau}} = 2\hat{z}_i A.$$

Also,

$$\frac{\partial}{\partial a}\log\big(\Phi(az_i)\big)\Big|_{\tau=\hat{\tau}} = \frac{\Phi'(az_i) * (az_i)'}{\Phi(az_i)}\Bigg|_{\tau=\hat{\tau}}$$

$$= \frac{\phi(a\hat{z}_i) * (\hat{z}_i + a * A)}{\Phi(a\hat{z}_i)}$$

$$\frac{\partial}{\partial a}\left[\log\big(\Phi(a\hat{z}_i)\big)\right] = \widehat{M}(\hat{z}_i + a * A)$$

and

$$\frac{\partial}{\partial a}\left[\log\big(Var(\mathcal{Z})\big)\right] = \frac{\partial}{\partial a}\left(1 - \frac{2a^2}{\pi(1+a^2)}\right)$$

$$= \frac{\left(\frac{-4a}{\pi(1+a^2)^2}\right)}{Var(\mathcal{Z})}$$

we have

$$\frac{\partial}{\partial a}\ell_{i2}(\boldsymbol{\theta}; v_i)|_{\tau=\hat{\tau}} = -\hat{z}_i A + \widehat{M}[\hat{z}_i + a * A] - \frac{1}{2}\left[\frac{4a}{\pi(1+a^2)^2 Var(\mathcal{Z})}\right].$$

and

$$\frac{\partial h^*|_{\tau=\hat{\tau}}}{\partial a} = \sum_{i=1}^{G}\left\{-\hat{z}_i A + \widehat{M}(\hat{z}_i + a * A) - \frac{2a}{\pi(1+a^2)^2 Var(\mathcal{Z})}\right\}.$$

Next, we calculate

$$-\frac{\partial^2}{\partial a^2}\ell_{i2}(\boldsymbol{\theta}; v_i)|_{\tau=\hat{\tau}} = -\frac{\partial}{\partial a}\left\{-\hat{z}_i\mathcal{A}(a,\sigma) + \widehat{\mathcal{M}}(\hat{z};\theta)[\hat{z}_i + a\mathcal{A}(a,\sigma)] - \frac{2a}{\pi(1+a^2)^2 Var(\mathcal{Z})}\right\},$$

$$= A^2 + \hat{z}_i A' - \widehat{M}'_a(\hat{z}_i + a * A) - \widehat{M}(2A + aA') + C'$$

$$-\frac{\partial^2 h^*|_{\tau=\hat{\tau}}}{\partial a^2} = \sum_{i=1}^{G}\left[A^2 + \hat{z}_i A' - \widehat{M}'_a(\hat{z}_i + a * A) - \widehat{M}(2A + aA') + C'\right]$$

where

$$A' = \frac{\partial}{\partial a}A,$$

$$= -\frac{\hat{v}_i}{\sigma}\frac{\partial}{\partial a}\left\{\left(1 - \frac{2a^2}{\pi(1+a^2)}\right)^{-\frac{1}{2}}\left(\frac{2a}{\pi(1+a^2)^2}\right)\right\},$$

$$A' = -\frac{\hat{v}_i}{\sigma}\left\{\left(1 - \frac{2a^2}{\pi(1+a^2)}\right)^{-\frac{1}{2}}\left(\frac{2-6a^2}{\pi(1+a^2)^3}\right) + \left(1 - \frac{2a^2}{\pi(1+a^2)}\right)^{-\frac{3}{2}}\left(\frac{2a}{\pi(1+a^2)^2}\right)^2\right\},$$

$$C' = \frac{\partial}{\partial a}\left\{\frac{2a}{\pi(1+a^2)^2 Var(\mathcal{Z})}\right\},$$

$$= \frac{2[\pi(1+a^2)^2 Var(\mathcal{Z})] - 2a\left[\pi(1+a^2)^2\left(\frac{-4a}{\pi(1+a^2)^2}\right) + Var(\mathcal{Z})(2\pi)(1+a^2)(2a)\right]}{[\pi(1+a^2)^2 Var(\mathcal{Z})]^2},$$

$$= \frac{2[\pi(1+a^2)^2 Var(\mathcal{Z})] - 2a[-4a + 4a\pi Var(\mathcal{Z})(1+a^2)]}{[\pi(1+a^2)^2 Var(\mathcal{Z})]^2},$$

$$C' = \frac{2[(1+a^2)^2 Var(\mathcal{Z})] - 8a^2[-\pi^{-1} + (1+a^2)Var(\mathcal{Z})]}{\pi[(1+a^2)^2 Var(\mathcal{Z})]^2},$$

$$\widehat{M}'_a = \frac{\partial}{\partial a}\widehat{M}\Big|_{\tau=\hat{\tau}} = \frac{[\phi(az_i)]' \times \Phi(az_i) - [\Phi(az_i)]' \times \phi(az_i)}{[\Phi(az_i)]^2}\Bigg|_{\tau=\hat{\tau}},$$

$$= \frac{\left(-az_i^2 - a^2\hat{z}_i A\right)\phi(a\hat{z}_i)\Phi(a\hat{z}_i) - (\hat{z}_i + a * A)\phi(a\hat{z}_i)\phi(a\hat{z}_i)}{[\Phi(a\hat{z}_i)]^2},$$

$$= -a\hat{z}_i(\hat{z}_i + a * A)\frac{\phi(a\hat{z}_i)}{\Phi(a\hat{z}_i)} - (\hat{z}_i + a * A)\left[\frac{\phi(a\hat{z}_i)}{\Phi(a\hat{z}_i)}\right]^2,$$

$$= -a\hat{z}_i(\hat{z}_i + a * A)\widehat{M} - (\hat{z}_i + a * A)M^2,$$

$$\widehat{M}'_a = -(\hat{z}_i + a * A)\left(a\hat{z}_i\widehat{M} + \widehat{M}^2\right).$$

Next, we compute the term $\widehat{\boldsymbol{Q}}'_a = \frac{\partial \boldsymbol{Q}}{\partial a}\Big|_{\tau=\hat{\tau}}$ in $(A3.5)$

$$\widehat{\boldsymbol{Q}}'_a = \frac{\partial}{\partial a}\left\{\frac{Var(\mathcal{Z})}{\sigma^2}\left(1 + a^2(az_iM + M^2)\right)\Bigg|_{\tau=\hat{\tau}}\right\}$$

$$= -\frac{1}{\sigma^2}\left(\frac{4a}{\pi(1+a^2)^2}\right)\left[1 + a^2\left(a\hat{z}_i\widehat{M} + \widehat{M}^2\right)\right]$$

$$+ \left(\frac{Var(\mathcal{Z})}{\sigma^2}\right)\left\{2a\left(a\hat{z}_i\widehat{M} + \widehat{M}^2\right) + a^2\left(\hat{z}_i\widehat{M} + aA\widehat{M} + a\hat{z}_i\widehat{M}'_a + 2M * \widehat{M}'_a\right)\right\}$$

Finally, we compute term $\boldsymbol{Q}''_a = \frac{\partial^2 Q}{\partial a^2}\Big|_{\tau=\hat{\tau}}$ in $(A3.7)$

$$\boldsymbol{Q}''_a = \frac{\partial}{\partial a}\Bigg\{-\frac{1}{\sigma^2}\left(\frac{4a}{\pi(1+a^2)^2}\right)\{1 + a^2(az_iM + M^2)\}$$

$$+ \left(\frac{Var(\mathcal{Z})}{\sigma^2}\right)\{2a(az_iM + M^2) + a^2(z_iM + aAM + az_iM'_a + 2M * M'_a)\}\Bigg|_{\tau=\hat{\tau}}\Bigg\}$$

$$= -\frac{4(1-3a^2)}{\sigma^2\pi(1+a^2)^3}\{1 + a^2(a\hat{z}_i M + M^2)\}$$

$$-\frac{1}{\sigma^2}\left(\frac{4a}{\pi(1+a^2)^2}\right)\{2a[a\hat{z}_i\widehat{M} + \widehat{M}^2] + a^2(\hat{z}_i M + aA\widehat{M} + a\hat{z}_i\widehat{M}'_a + 2\widehat{M} * \widehat{M}'_a)\}$$

$$-\frac{1}{\sigma^2}\left(\frac{4a}{\pi(1+a^2)^2}\right)\{2a(a\hat{z}_i\widehat{M} + \widehat{M}^2)$$

$$+ a^2(\hat{z}_i\widehat{M} + aA\widehat{M} + a\hat{z}_i\widehat{M}'_a + 2\widehat{M} * \widehat{M}'_a)\}$$

$$+\left(\frac{Var(\mathcal{Z})}{\sigma^2}\right)\Big\{2[a\hat{z}_i\widehat{M} + \widehat{M}^2] + 2a[\hat{z}_i\widehat{M} + aA\widehat{M} + a\hat{z}_i\widehat{M}'_a + 2\widehat{M} * \widehat{M}'_a]$$

$$+ 2a[\hat{z}_i\widehat{M} + aA\widehat{M} + a\hat{z}_i\widehat{M}'_a + 2\widehat{M} * \widehat{\mathcal{M}}'_a]$$

$$+ a^2\Big[A\widehat{M} + \hat{z}_i\widehat{M}'_a + A\widehat{M} + aA'\widehat{M} + aA\widehat{M}'_a + \hat{z}_i\widehat{M}'_a + aA\widehat{M}'_a + aA\widehat{M}''_a$$

$$+ 2\left(\left(\widehat{M}'_a\right)^2 + \widehat{M} * \widehat{M}''_a\right)\Big]\Big\}$$

$$\widehat{\boldsymbol{Q}}''_a = -\frac{4(1-3a^2)}{\sigma^2\pi(1+a^2)^3}\{1 + a^2(a\hat{z}_i\widehat{M} + \widehat{M}^2)\}$$

$$-\frac{1}{\sigma^2}\left(\frac{8a}{\pi(1+a^2)^2}\right)\{2a[a\hat{z}_i\widehat{M} + \widehat{M}^2] + a^2(\hat{z}_i\widehat{M} + aA\widehat{M} + a\hat{z}_i\widehat{M}'_a + 2\widehat{M} * \widehat{M}'_a)\}$$

$$+\left(\frac{Var(\mathcal{Z})}{\sigma^2}\right)\Big\{2[a\hat{z}_i\widehat{M} + \widehat{M}^2] + 4a[\hat{z}_i\widehat{M} + aA\widehat{M} + a\hat{z}_i\widehat{M}'_a + 2\widehat{M} * \widehat{M}'_a]$$

$$+ a^2\Big[2A\widehat{M} + 2\hat{z}_i\widehat{M}'_a + aA'\widehat{M} + 2aA\widehat{M}'_a + aA\widehat{M}''_a + 2\left(\left(\widehat{M}'_a\right)^2 + \widehat{M} * \widehat{M}''_a\right)\Big]\Big\},$$

where $\widehat{M}''_a = \frac{\partial}{\partial a}M'_a|_{\tau=\hat{\tau}}$. That is

$$\widehat{M}''_a = -\frac{\partial}{\partial a}\{((z_i + aA)(az_iM + M^2)|_{\tau=\hat{\tau}})\}$$

$$= -(2A + aA')[a\hat{z}_i\widehat{M} + \widehat{M}^2] - (\hat{z}_i + aA)[\hat{z}_i\widehat{M} + aA\widehat{M} + a\hat{z}_i\widehat{M}'_a + 2\widehat{M} * \widehat{M}'_a]$$

$$\widehat{\mathcal{M}}''_a = -\{(2A + aA')[a\hat{z}_i\widehat{M} + \widehat{M}^2] + (\hat{z}_i + aA)[\hat{z}_i\widehat{M} + aA\widehat{M} + a\hat{z}_i\widehat{M}'_a + 2\widehat{M} * \widehat{M}'_a]\}.$$

A3.3.3 mathematical derivations of terms $\left[-\frac{\partial^2 h_A^*}{\partial\sigma\partial a}\right]$ and $\left[-\frac{\partial^2 h_A^*}{\partial a\partial\sigma}\right]$

The direct calculation of the term $-\left[\frac{\partial^2}{\partial a\partial\sigma}\ell_{i2}(\boldsymbol{\theta}; v_i)|_{\tau=\hat{\tau}}\right]$ is

$$-\frac{\partial^2}{\partial a\partial\sigma}\ell_{i2}(\boldsymbol{\theta};v_i)|_{\tau=\hat{\tau}} = -\frac{\partial}{\partial a}\left\{\left(\frac{z_i^2}{\sigma} - \widehat{M} * \frac{az_i}{\sigma} - \frac{1}{\sigma}\Big|_{\tau=\hat{\tau}}\right)\right\},$$

$$= -\frac{2\hat{z}_i}{\sigma}A + \widehat{M}'_a * \frac{a\hat{z}_i}{\sigma} + \frac{\widehat{M}}{\sigma}[\hat{z}_i + aA].$$

Therefore,

$$-\frac{\partial h^*|_{\tau=\hat{\tau}}}{\partial a\partial\sigma} = \sum_{i=1}^{G}\left\{-\frac{2\hat{z}_i}{\sigma}A + \widehat{M}'_a * \frac{a\hat{z}_i}{\sigma} + \frac{\widehat{M}}{\sigma}[\hat{z}_i + aA]\right\}.$$

Also

$$-\frac{\partial^2}{\partial\sigma\partial a}\ell_{i2}(\boldsymbol{\theta};v_i)|_{\tau=\hat{\tau}} = -\frac{\partial}{\partial\sigma}\left\{\left(-\hat{z}_iA + \widehat{M}[\hat{z}_i + aA] - \frac{2a}{\pi(1+a^2)^2Var(\mathcal{Z})}\Big|_{\tau=\hat{\tau}}\right)\right\},$$

$$= \frac{\partial}{\partial\sigma}\left\{\left(\hat{z}_iA - \widehat{M}[\hat{z}_i + aA]\big|_{\tau=\hat{\tau}}\right)\right\},$$

$$= -\frac{\hat{z}_i}{\sigma}A + \hat{z}_iA'_\sigma - \widehat{M}'_\sigma[\hat{z}_i + aA] - \widehat{M}\left[-\frac{\hat{z}_i}{\sigma} + aA'_\sigma\right],$$

where

$$A'_\sigma = \frac{\partial}{\partial\sigma}A = \frac{\partial}{\partial\sigma}\left\{-\frac{v_i}{\sigma}\left(1 - \frac{2a^2}{\pi(1+a^2)}\right)^{-\frac{1}{2}}\left(\frac{2a}{\pi(1+a^2)^2}\right)\Bigg|_{\tau=\hat{\tau}}\right\}$$

$$A'_\sigma = \frac{\hat{v}_i}{\sigma^2}\left(1 - \frac{2a^2}{\pi(1+a^2)}\right)^{-\frac{1}{2}}\left(\frac{2a}{\pi(1+a^2)^2}\right) = -\frac{A}{\sigma}$$

Therefore,

$$-\frac{\partial^2\ell_{i2}(\boldsymbol{\theta};\hat{v}_i)}{\partial\sigma\partial a} = -\frac{\hat{z}_i}{\sigma}A - \frac{\hat{z}_i}{\sigma}A - \widehat{M}'_\sigma[\hat{z}_i + aA] - \widehat{M}\left[-\frac{\hat{z}_i}{\sigma} - a\frac{A}{\sigma}\right]$$

$$-\frac{\partial^2\ell_{i2}(\boldsymbol{\theta};\hat{v}_i)}{\partial\sigma\partial a} = -\frac{2\hat{z}_i}{\sigma}A - \widehat{M}'_\sigma[\hat{z}_i + aA] + \frac{\widehat{M}}{\sigma}[\hat{z}_i + aA]$$

$$-\frac{\partial h^*|_{\tau=\hat{\tau}}}{\partial\sigma\partial a} = \sum_{i=1}^{G}\left\{-\frac{2\hat{z}_i}{\sigma}A - \widehat{M}'_\sigma[\hat{z}_i + aA] + \frac{\widehat{M}}{\sigma}[\hat{z}_i + aA]\right\}$$

Note: $-\frac{\partial h^*|_{\tau=\hat{\tau}}}{\partial\sigma\partial a} = -\frac{\partial h^*|_{\tau=\hat{\tau}}}{\partial a\partial\sigma}$, since $\frac{a\hat{z}_i}{\sigma}\widehat{M}'_a = -\widehat{M}'_\sigma[\hat{z}_i + aA]$.

Next, we calculate the term $\widehat{\boldsymbol{Q}}''_{a\sigma} = \frac{\partial \boldsymbol{Q}}{\partial a\partial\sigma}\Big|_{\tau=\hat{\tau}}$ in $(A3.7)$ is

$$\widehat{\boldsymbol{Q}}''_{a\sigma} = \frac{\partial}{\partial a}\Bigg\{-\frac{2}{\sigma^3}Var(\mathcal{Z})\{1 + a^2(az_iM + M^2)\}$$

$$+\frac{a^2Var(\mathcal{Z})}{\sigma^2}\left\{\frac{-az_i}{\sigma}M + az_iM'_\sigma + 2M * M'_\sigma\right\}\Bigg|_{\tau=\hat{\tau}}\Bigg\}$$

$$\widehat{\boldsymbol{Q}}''_{a\sigma} = \frac{8a}{\sigma^3\pi(1+a^2)^2}\left\{1 + a^2\left(a\hat{z}_i\widehat{M} + \widehat{M}^2\right)\right\}$$

$$-\frac{2}{\sigma^3}Var(\mathcal{Z})\left\{2a\left[a\hat{z}_i\widehat{M} + \widehat{M}^2\right] + a^2\left[\hat{z}_i\widehat{M} + aA\widehat{M} + a\hat{z}_i\widehat{M}'_a + 2\widehat{M} * \widehat{M}'_a\right]\right\}$$

$$+\left(\frac{2aVar(\mathcal{Z})}{\sigma^2} - \frac{4a^3}{\sigma^2\pi(1+a^2)^2}\right)\left\{\frac{-a\hat{z}_i}{\sigma}\widehat{M} + a\hat{z}_i\widehat{M}'_\sigma + 2\widehat{M} * \widehat{M}'_\sigma\right\}$$

$$+\frac{a^2Var(\mathcal{Z})}{\sigma^2}\left\{\frac{-\hat{z}_i}{\sigma}\widehat{M} - \frac{aA}{\sigma}\widehat{M} - \frac{a\hat{z}_i}{\sigma}\widehat{M}'_a + \hat{z}_i\widehat{M}'_\sigma + aA\widehat{M}'_\sigma + a\hat{z}_i\widehat{M}''_{a\sigma}\right.$$

$$\left. + 2\left[\widehat{M}'_a * \widehat{M}'_\sigma + \widehat{M} * \widehat{M}''_{a\sigma}\right]\right\},$$

where $\widehat{M}''_{a\sigma} = \frac{\partial}{\partial a}M'_\sigma|_{\tau=\hat{\tau}}$. That is

$$\widehat{M}''_{a\sigma} = \frac{\partial}{\partial a}\left\{\frac{az_i}{\sigma}[az_iM + M^2]\Big|_{\tau=\hat{\tau}}\right\}$$

$$\widehat{M}''_{a\sigma} = \frac{1}{\sigma}(\hat{z}_i + aA)\left\{a\hat{z}_i\widehat{M} + \widehat{M}^2\right\} + \frac{a\hat{z}_i}{\sigma}\left\{\hat{z}_i\widehat{M} + aA\widehat{M} + a\hat{z}_i\widehat{M}'_a + 2\widehat{M} * \widehat{M}'_a\right\}$$

Finally, we calculate the term $\widehat{\boldsymbol{Q}}''_{\sigma a} = \frac{\partial \boldsymbol{Q}}{\partial\sigma\partial a}\Big|_{\tau=\hat{\tau}}$.

$$\widehat{\boldsymbol{Q}}''_{\sigma a} = \frac{\partial}{\partial \sigma}\Bigg\{-\frac{1}{\sigma^2}\left(\frac{4a}{\pi(1+a^2)^2}\right)\{1 + a^2(az_iM + M^2)\}$$

$$+\left(\frac{Var(\mathcal{Z})}{\sigma^2}\right)\{2a(az_iM + M^2) + a^2(z_iM + aAM + az_iM'_a + 2M * M'_a)\}\Bigg|_{\tau=\hat{\tau}}\Bigg\}$$

$$\widehat{\boldsymbol{Q}}''_{\sigma a} = \left(\frac{8a}{\pi\sigma^3(1+a^2)^2}\right)\left\{1 + a^2\left(az_i\widehat{M} + \widehat{M}^2\right)\right\}$$

$$- \frac{4a}{\sigma^2\pi(1+a^2)^2}\left\{a^2\left(-\frac{a\hat{z}_i}{\sigma}\widehat{M} + a\hat{z}_i\widehat{M}'_\sigma + 2\widehat{M} * \widehat{M}'_\sigma\right)\right\}$$

$$- \left(\frac{2Var(\mathcal{Z})}{\sigma^3}\right)\left\{2a\left(a\hat{z}_i\widehat{M} + \widehat{M}^2\right) + a^2\left(\hat{z}_i\widehat{M} + aA\widehat{M} + a\hat{z}_i\widehat{M}'_a + 2\widehat{M} * \widehat{M}'_a\right)\right\}$$

$$+ \left(\frac{Var(\mathcal{Z})}{\sigma^2}\right)\left\{2a\left(-\frac{a\hat{z}_i}{\sigma}\widehat{M} + a\hat{z}_i\widehat{M}'_\sigma + 2\widehat{M} * \widehat{M}'_\sigma\right)\right.$$

$$+ a^2\left(-\frac{\hat{z}_i}{\sigma}\widehat{M} + \hat{z}_i\widehat{M}'_\sigma + aA'_\sigma\widehat{M} + aA\widehat{M}'_\sigma - \frac{a\hat{z}_i}{\sigma}\widehat{M}'_a + a\hat{z}_i\widehat{M}''_{\sigma a}\right.$$

$$\left.\left. + 2\left[\widehat{M}'_\sigma * \widehat{M}'_a + \widehat{M} * \widehat{M}''_{\sigma a}\right]\right)\right\},$$

where $\widehat{M}''_{\sigma a} = \frac{\partial}{\partial\sigma}\left(M'_a|_{\tau=\hat{\tau}}\right)$. That is

$$\widehat{M}''_{\sigma a} = -\frac{\partial}{\partial\sigma}\{(z_i + aA)[az_iM + M^2]|_{\tau=\hat{\tau}}\}$$

$$\widehat{M}''_{\sigma a} = -\left\{\left(-\frac{\hat{z}_i^2}{\sigma} + aA'_\sigma\right)\left[a\hat{z}_i\widehat{M} + \widehat{M}^2\right] + (\hat{z}_i + aA)\left[-\frac{a\hat{z}_i}{\sigma}\widehat{M} + a\hat{z}_i\widehat{M}'_\sigma + 2 * \widehat{M} * \widehat{M}'_\sigma\right]\right\}$$