<table>
<thead>
<tr>
<th>Title</th>
<th>Semiparametric regression analysis of clustered survival data with semi-competing risks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Author(s)</td>
<td>Peng, Mengjiao; Xiang, Liming; Wang, Shanshan</td>
</tr>
<tr>
<td>Date</td>
<td>2018</td>
</tr>
<tr>
<td>URL</td>
<td><a href="http://hdl.handle.net/10220/48475">http://hdl.handle.net/10220/48475</a></td>
</tr>
<tr>
<td>Rights</td>
<td>© 2018 Elsevier B.V. All rights reserved. This paper was published in Computational Statistics &amp; Data Analysis and is made available with permission of Elsevier B.V.</td>
</tr>
</tbody>
</table>
Semiparametric Regression Analysis of Clustered Survival Data with Semi-competing Risks

Mengjiao Peng, Liming Xiang*, Shanshan Wang

School of Physical and Mathematical Sciences, Nanyang Technological University, Singapore

Abstract

Analysis of semi-competing risks data is becoming increasingly important in medical research in which a subject may experience both nonterminal and terminal events, and the time to the intermediate nonterminal event (e.g. onset of a disease) is subject to dependent censoring by the terminal event (e.g. death) but not vice versa. Typically, both two types of events are dependent. In many applications, subjects may also be nested within clusters, such as patients in a multi-center study, leading to possible association among event times due to unobserved shared factors across subjects. To incorporate dependency within clusters and association between two types of event times, we propose a new flexible semiparametric modeling framework where a copula model is employed for the joint distribution of the nonterminal and terminal events, and their marginal distributions are modeled by Cox proportional hazards models with random effects. A nonparametric maximum likelihood estimation procedure is developed and implemented through a Monte Carlo EM algorithm. The proposed estimator is also shown to enjoy desirable asymptotic properties. Results from extensive simulation studies indicate that the proposed method performs very well in finite samples and is especially robust against misspecification of the random effects distribution. We further illustrate the practical utility of the method by analyzing data from a multi-institutional study of breast cancer.

Key Words: Copula, Clustered data, Monte Carlo EM algorithm, Proportional hazards model, Random effects, Semi-competing risks.

* Corresponding author. Tel.: +65 65137451 ; Fax: +65 65158213.
E-mail address: lmxiang@ntu.edu.sg.
1 Introduction

In many biomedical studies, survival data arises with a cluster structure, such as the survival/failure times of subjects observed in multicenter clinical trials, genetic and family studies, or group-randomized trials. Usually, the subjects within the same cluster share the common cluster effects, leading to dependencies among survival/failure times within each cluster. The inference can be misleading when ignoring such cluster effects. A common approach to accommodate the intra-cluster dependence is to incorporate an unobserved frailty (or random effects) in a survival model, where the frailty acts multiplicatively to the hazard function in order to model the correlation within a cluster. Ripatti and Palmgren (2000), Vaida and Xu (2000), and Ripatti et al. (2002) generalized the frailty model to the proportional hazards model with random effects which allows random effects on arbitrary covariates. However, these works considered only one survival endpoint of interest.

When two or several different types of events are of interest, most of the existing works for clustered data deal with competing risks by modeling one type of event at a time and regarding the occurrence of other types of events as censoring based on independent censoring assumption. See for example Gorfine and Hsu (2011); Christian et al. (2016); Lee et al. (2017) and references therein. This assumption of independent censoring is problematic in many practical situations. For example, in an international multi-institutional study of breast cancer (Haibe-Kains et al., 2012), there were two clinical endpoints of interest. The primary endpoint was overall survival (OS), which was defined as the time from diagnosis to the date of death. The distant metastasis-free survival (DMFS), defined as the time from diagnosis to the time of distant metastasis, was considered as as a meaningful intermediate endpoint for overall survival in patients with breast cancer. Yet, some patients died without experiencing DMFS. In contrast, the occurrence of DMFS did not prevent the observation of death after the DMFS endpoint. Such a dependent censoring scenario falls into the paradigm of “semi-competing risks”. Evidence from the work of van Vliet et al. (2008) showed that data collected by different clinical institutions from the underlying breast cancer population produced different results in predicting disease outcome, implying potential cluster effects in the data. This motivates the need for analyzing clustered semi-competing risks data by incorporating both the dependence between the two endpoints and the association within clusters.

“Semi-competing risks” consist of a terminal event (e.g. death), and an intermediate nonterminal event (e.g. onset of a distant metastasis). The terminal event can still be observed if the nonterminal event occurs earlier. Yet, the terminal event will censor the nonterminal event when it
occurs before the nonterminal event. Both two events contribute “semi-competing risks”. As the
times to the terminal and nonterminal events of an individual are usually correlated, the censoring
of the nonterminal event by the terminal event forms a type of dependent/informative censoring.
In this case, the conventional analysis of clustered data based on independent censoring assumption
could be biased.

When analyzing semi-competing risks data, one is often interested in the effects of covariates
on the nonterminal and terminal event times incorporating association between them. To this end,
regression analysis of such data has been studied by many researchers. Peng and Fine (2007),
Hsieh et al. (2008), Lakhal et al. (2008) and Chen (2012) considered copula models by imposing
separate marginal regression models on the two event times under dependent censoring scenario.
Alternatively, Xu et al. (2010) proposed a three-state illness-death model with shared gamma
frailty for modeling the semi-competing risks data, where the common frailty contributed to the
dependence of both nonterminal and terminal events. Their model avoids a modeling problem of the
restriction of triangle range to some extent. Their model was extended by Jiang and Haneuse (2016)
to a class of transformation frailty models, allowing a wider range of possible frailty distributions.
However, all these aforementioned models did not consider the dependencies among event times
within clusters. The application of these methods to clustered semi-competing risk data may lead
to a loss of efficiency.

Despite of the large literature on analysis of semi-competing risks data, relatively few works are
available for such data with clustered structure. Emura et al. (2015) introduced frailty to a joint
copula model for semi-competing risks data arisen in a meta-analysis including several existing
studies. They approximated the baseline hazard functions using splines, and adopted a penalized
maximum likelihood method to estimate model parameters. Their method requires the gamma
distribution assumption for frailty so that estimation through maximizing a penalized log-likelihood
can be mathematically tractable. However, it is not necessary for one to assume the gamma frailty in
practice. The method would fail to produce consistent estimators if the gamma frailty distribution
is misspecified. In addition, there is a lack of theoretical justification for the asymptotic properties
of the existing estimator. Using Bayesian hierarchical modeling approach, Lee et al. (2016) further
extended Xu et al. (2010)’s multistate illness-death model to the clustered semi-competing risks
data setting, allowing a parametric or nonparametric specification for the cluster-specific random
effects distribution. In this paper, we consider a flexible copula-based method for analysis of
clustered semi-competing risks data from a frequentist perspective. Compared with the existing
Bayesian hierarchical multistate modeling approach, the advantages of the proposed method are three folded. First, it enables us to simultaneously and explicitly formulate the marginal behaviors of both non-terminal and terminal event times through semiparametric survival models and the dependency structure inherent within clusters by random effects. Second, it measures the degree of dependency between the two event time variables using only a single parameter $\alpha$, which can be easily interpreted in terms of Kendall’s tau. Third, it allows us to establish a desired asymptotic theory for the resulting estimator.

We first formulate marginal distribution for each event time using the Cox proportional hazards model with shared random effects to account for association among event times within clusters. Through a copula model, it further allows us to connect the two marginal models together to form a joint model with an unspecified parameter for association between the non-terminal and terminal events. Particularly, our method does not restrict random effects following a gamma distribution, providing a more flexible modeling framework.

We propose a nonparametric maximum likelihood (NPML) approach to estimate the cumulative baseline hazard function by approximating it using piecewise constants, and develop the maximum likelihood estimation for the marginal regression parameters, variance-covariance of random effects and copula association parameter. The estimation procedure is implemented via an EM algorithm, which avoids the integral problem caused by random effects and approximates the required conditional expectation using the Markov Chain Monte Carlo (MCMC) method. Our numerical results demonstrate that the proposed estimation of parameters is robust against misspecification of the random effects distribution and the estimates of baseline hazard functions are more accurate than those obtained from the existing method.

We also establish asymptotic properties for the proposed estimators. Our estimators are shown to be consistent and asymptotically normal using techniques including martingale, the theory of counting process and martingale central limit theorem developed by Murphy (1995).

The rest of the paper is organized as follows. Section 2 introduces the model, while Section 3 and Section 4 describe the proposed estimation method and a detailed algorithm for its implementation, respectively. Asymptotic properties of the proposed estimator are provided in Section 5. Results of simulation studies and of the real data applications are provided in Section 6 and Section 7, respectively. Some concluding remarks are given in the last section. All technical details are relegated to the Appendix.
2 Model

Let “event 1” be a nonterminal event, “event 2” be a terminal event, and \( T^{(1)} \) and \( T^{(2)} \) be the corresponding event times. We are interested in assessing the effects of covariates on the marginal distribution of \( T^{(k)} \), which is modeled as

\[
\lambda_k(t; X, Z) = \lambda_0(t)e^{\beta_k'X + b'Z}
\]  

for \((k = 1, 2)\), where \( \lambda_k(t; X, Z) \) is the conditional hazard function of \( T^{(k)} \) given the event and covariate history before \( t \), \( \lambda_0(\cdot) \) is the corresponding baseline hazard function, \( \beta_k \) is a vector of fixed effects, \( b \) is a vector of random effects with mean zero and density function \( \psi(b; \Sigma) \) indexed by variance-covariance matrix \( \Sigma \), and \( X \) and \( Z \) are vectors of covariates associated with the fixed and random effects, respectively.

We further assume that the joint survival function of \( T^{(1)} \) and \( T^{(2)} \), given covariates \( X \) and \( Z \), can be specified by a known copula subject to an unknown parameter \( \alpha \), which quantifies the association between \( T^{(1)} \) and \( T^{(2)} \):

\[
S(t^{(1)}, t^{(2)}) = \text{pr}(T^{(1)} \geq t^{(1)}, T^{(2)} \geq t^{(2)}) = c\{S_1(t^{(1)}), S_2(t^{(2)}); \alpha\},
\]

for \( t^{(1)} \geq 0, t^{(2)} \geq 0 \), where \( c\{u, v; \alpha\} \) is a copula function, \( \alpha \) is a constant association parameter in one-to-one correspondence with Kendall’s \( \tau \), and \( S_k(\cdot) \) is the marginal survival function of \( T^{(k)}(k = 1, 2) \). The key advantage of the copula model is that it allows us to describe the joint distribution of \( T^{(1)} \) and \( T^{(2)} \) through an explicit function of their marginal distributions and the corresponding association parameter. Below we give two common examples of the copula \( c \) accounting for positive dependence only: the Clayton copula

\[
c\{u, v; \alpha\} = (u^{-\alpha} + v^{-\alpha} - 1)^{-1/\alpha}, \quad \alpha > 0,
\]

and the Gumbel copula

\[
c\{u, v; \alpha\} = \exp\left[-\left\{(-\log u)^{\alpha+1} + (-\log v)^{\alpha+1}\right\}^{1/(\alpha+1)}\right], \quad \alpha > 0.
\]

The Clayton and Gumbel copulas above are two of the most important Archimedean copulas and are popular in practice because of their simple explicit expressions with only a single parameter.
controlling the strength of dependence. They can be used to capture asymmetry between the lower and upper tail dependence. The Clayton copula exhibits a higher degree of dependence in the lower tail than in the upper tail, while the Gumbel copula has strong upper-tail dependence with relatively weak lower-tail dependence. The Gumbel copula would be an appropriate choice when $T^{(1)}$ and $T^{(2)}$ in the semi-competing risks setting are strongly dependent at high values but less dependent at low values. Our numerical studies in Sections 6-7 focus on the Gumbel copula for illustration purpose though other choices of copulas could be possible. For more information on different types of copulas and their properties, we refer the reader to the book by Nelsen (2006) and the references therein.

3 Estimation

In this paper, we assume that the data consists of possibly right-censored nonterminal and terminal event time observations from $m$ clusters, with $n_i$ observations in the cluster $i$ for $i = 1, \ldots, m$, where $N = \sum_{i=1}^{m} n_i$ denotes the total number of observations. To proceed with estimation, further notation is needed. Denote by $T^{(1)}_{ij}$ the nonterminal event time for subject $j$ in cluster $i$ and by $T^{(2)}_{ij}$ the corresponding terminal event time. The observations within a cluster are assumed to be dependent, but conditional on the cluster-specific random effect $b_i$ the nonterminal and terminal event times $T^{(k)}_{ij}$ ($k = 1, 2$) are independent with the hazard functions following the proportional hazards model with random effects,

$$
\lambda_k(t|X_{ij}, Z_{ij}, b_i) = \lambda_0(t) e^{\beta_k'X_{ij} + b_i'Z_{ij}}. \tag{3}
$$

The corresponding joint survival function of nonterminal and terminal event times for subject $j$ in cluster $i$ is specified as:

$$
S(t^{(1)}_{ij}, t^{(2)}_{ij}) = \text{pr}(T^{(1)}_{ij} \geq t^{(1)}_{ij}, T^{(2)}_{ij} \geq t^{(2)}_{ij}) = c\{S_1(t^{(1)}_{ij}), S_2(t^{(2)}_{ij}); \alpha\}, t^{(1)}_{ij} \geq 0, t^{(2)}_{ij} \geq 0. \tag{4}
$$

Let $C^*$ denote the external censoring time that is independent of $(T^{(1)}, T^{(2)})$, and $c$ denote the duration of the study. Due to possible right-censoring, for subject $j$ in cluster $i$ we observe $O_{ij} = \{T_{ij}, \tilde{T}_{ij}, X_{ij}, Z_{ij}, \delta^{(1)}_{ij}, \delta^{(2)}_{ij}, \tilde{\delta}^{(2)}_{ij}\}$, where $T_{ij} = \min(T^{(1)}_{ij}, T^{(2)}_{ij}, C_{ij})$ with $C_{ij} = \min(C^*_{ij}, \zeta)$, $\tilde{T}_{ij} = \min(T^{(2)}_{ij}, C_{ij})$, $\delta^{(k)}_{ij} = I(T_{ij} = T^{(k)}_{ij})(k = 1, 2)$, $\tilde{\delta}^{(2)}_{ij} = I(\tilde{T}_{ij} = T^{(2)}_{ij})$, and $I(\cdot)$ is the indicator function. Thus for a random sample of $m$ clusters, each of size $n_i$, the observed data consist of
\( O = \{ O_i \}_{i=1}^m \), where \( O_i = \{ O_{ij} \}_{j=1}^{m_i} \). For each subject \((i,j)\), there are four possible cases that we can observe.

Case 1: \( T_{ij}^{(1)} \leq C_{ij} \), \( T_{ij}^{(2)} \leq C_{ij} \) \( \iff \) \( T_{ij}^{(1)} = T_{ij}, T_{ij}^{(2)} = \tilde{T}_{ij} \) \( \iff \) \( \delta_{ij} (1) \tilde{\delta}_{ij} (2) = 1 \); which indicates that the nonterminal event occurs firstly and the terminal event occurs secondly before censoring.

Case 2: \( T_{ij}^{(1)} \leq C_{ij} \leq T_{ij}^{(2)} \) \( \iff \) \( T_{ij}^{(1)} = T_{ij}, T_{ij}^{(2)} > \tilde{T}_{ij} \) \( \iff \) \( \delta_{ij} (1) (1 - \tilde{\delta}_{ij} (2) ) = 1 \); which indicates that the nonterminal event occurs before censoring and the terminal event is right-censored.

Case 3: \( C_{ij} \leq T_{ij}^{(1)}, C_{ij} \leq T_{ij}^{(2)} \) \( \iff \) \( T_{ij}^{(1)} > T_{ij}, T_{ij}^{(2)} > \tilde{T}_{ij} \) \( \iff \) \( 1 - \delta_{ij} (1) - \delta_{ij} (2) = 1 \); which indicates that both the nonterminal and terminal events are right-censored.

Case 4: \( T_{ij}^{(2)} \leq C_{ij} \), \( T_{ij}^{(2)} \leq T_{ij}^{(1)} \) \( \iff \) \( T_{ij}^{(1)} > T_{ij}, T_{ij}^{(2)} = T_{ij} \) \( \iff \) \( \delta_{ij} (2) = 1 \); which indicates that the terminal events occurs firstly before censoring, and the nonterminal event is right-censored by the terminal event.

Therefore, let \( \beta = (\beta_1, \beta_2), \theta = (\alpha, \beta), \lambda = (\lambda_{01}(\cdot), \lambda_{02}(\cdot)) \), the likelihood function concerning the parameters \( \omega = (\theta, \lambda, \Sigma) \) given observed data \( O \) is

\[
L(\theta, \lambda, \Sigma|O) = \prod_{i=1}^{m} \int_{b_i} L_i(\theta, \lambda|O_i, b_i) \psi(b_i; \Sigma) db_i = \prod_{i=1}^{m} \int_{b_i} \prod_{j=1}^{n_i} L_{ij}(\theta, \lambda|O_{ij}, b_i) \psi(b_i; \Sigma) db_i, \tag{5}
\]

where

\[
L_{ij}(\theta, \lambda|b_i, O_{ij}) = \Pr(T_{ij}^{(1)} = T_{ij}, T_{ij}^{(2)} = \tilde{T}_{ij}|b_i) \delta_{ij} (1) \tilde{\delta}_{ij} (2) \cdot \Pr(T_{ij}^{(1)} = T_{ij}, T_{ij}^{(2)} > \tilde{T}_{ij}|b_i) \delta_{ij} (1) (1 - \tilde{\delta}_{ij} (2)) \cdot \Pr(C_{ij} \leq T_{ij}^{(1)}, C_{ij} \leq T_{ij}^{(2)}|b_i) (1 - \delta_{ij} (1) - \delta_{ij} (2)) \cdot \Pr(T_{ij}^{(2)} \leq C_{ij}, T_{ij}^{(2)} \leq T_{ij}^{(1)}|b_i) \delta_{ij} (2).
\]
Plugging models (3) and (4) in the likelihood, (5) can be written as:

\[
L(\theta, \lambda, \Sigma|O) = \prod_{i=1}^{m} \prod_{t=1}^{n_i} \left[ c_{12}\{S_1(T_{ij}), S_2(\tilde{T}_{ij}); \alpha\} \times \lambda_{01}(T_{ij}) \right. \\
\Times \left. e^{\beta_1 x_{ij} + \beta_2 z_{ij}} S_2(\tilde{T}_{ij}) \lambda_{02}(\tilde{T}_{ij}) e^{\beta_2 x_{ij} + \beta_1 z_{ij}} \right]^{\delta_{ij}^{(1)}} \left[ c_1\{S_1(T_{ij}), S_2(\tilde{T}_{ij}); \alpha\} S_1(T_{ij}) \lambda_{01}(T_{ij}) e^{\beta_1 x_{ij} + \beta_2 z_{ij}} \right]^{\delta_{ij}^{(2)}} \\
\Times \left[ c\{S_1(T_{ij}), S_2(\tilde{T}_{ij})\} \right]^{(1-\delta_{ij}^{(1)}-\delta_{ij}^{(2)})} \\
\Times \left[ c_{2}\{S_1(T_{ij}), S_2(\tilde{T}_{ij}); \alpha\} S_2(\tilde{T}_{ij}) \lambda_{02}(\tilde{T}_{ij}) e^{\beta_2 x_{ij} + \beta_1 z_{ij}} \right]^{\delta_{ij}^{(2)}} \psi(b_i; \Sigma)db_i,
\]

where \( c_1\{u, v; \alpha\} = e_{\alpha}^{u,v} \), \( c_2\{u, v; \alpha\} = e_{\alpha}^{u,v} \), and \( c_{12}\{u, v; \alpha\} = e_{\alpha}^{u.v.\alpha} \).

Before deriving the maximum of \( L(\theta, \lambda, \Sigma|O) \) in (6), we introduce some notations about counting processes here. Let \( Y_{ij}(t) = I(T_{ij} > t), \tilde{Y}_{ij}(t) = I(\tilde{T}_{ij} > t), N_{ij}^{(1)}(t) = \delta_{ij}^{(1)}(1 - Y_{ij}(t)) \) be the counting process of event 1 that is the first-occurring event for subject \( i \), \( N_{ij}^{(2)}(t) = \delta_{ij}^{(2)}(1 - \tilde{Y}_{ij}(t)) \) be the counting process of event 2 that is the first-occurring event for subject \( i \), and \( \tilde{N}_{ij}(t) = \delta_{ij}^{(1)} \delta_{ij}^{(2)}(1 - \tilde{Y}_{ij}(t)) \) be the counting process of event 2 that is observed subsequently to event 1. Also we assume that \( \text{pr}(T_{ij}^{(1)} = T_{ij}^{(2)}) = 0 \) for \( i = 1, \ldots, m \) and \( j = 1, \ldots, n_i \).

The maximum of \( L(\theta, \lambda, \Sigma|O) \) in (6) does not exist if \( \Lambda_{01}(\cdot) \) and \( \Lambda_{02}(\cdot) \) are restricted to be absolutely continuous. By allowing \( \Lambda_{01}(\cdot) \) and \( \Lambda_{02}(\cdot) \) to be increasing right-continuous functions with the forms \( \Lambda_{01}(t) = \int_{0}^{t} \lambda_{01}(t) dt \) and \( \Lambda_{02}(t) = \int_{0}^{t} \lambda_{02}(t) dt \), we replace \( \lambda_{01}(t) \) and \( \lambda_{02}(t) \) in (6) with the jump sizes of \( \Lambda_{01} \) and \( \Lambda_{02} \) at time \( t \) denoted by \( \Lambda_{01}(t) \) and \( \Lambda_{02}(t) \), respectively. By letting the derivative of the modified complete likelihood with respect to \( \Lambda_{01}\{T_{ij}\} \) and \( \Lambda_{02}\{\tilde{T}_{ij}\} \) to 0, we obtain the nonparametric maximum likelihood estimates

\[
\hat{\Lambda}_{01}\{T_{ij}\} = \frac{dN_{ij}^{(1)}(T_{ij})}{\sum_{(g,h) \in R(T_{ij})} e^{\beta_1 x_{gh} + \beta_2 z_{gh}} \{w_1(T_{gh})e^{\beta_1 x_{gh}}\}}, \quad \text{(7)}
\]

\[
\hat{\Lambda}_{02}\{\tilde{T}_{ij}\} = \frac{dN_{ij}^{(2)}(\tilde{T}_{ij}) + d\tilde{N}_{ij}^{(2)}(\tilde{T}_{ij})}{\sum_{(g,h) \in R(\tilde{T}_{ij})} e^{\beta_2 x_{gh} + \beta_1 z_{gh}} \{w_2(\tilde{T}_{gh})e^{\beta_2 x_{gh}}\}}, \quad \text{(8)}
\]

where for \( (g,h) \in R(T_{ij}) \) or \( (g,h) \in R(\tilde{T}_{ij}) \), i.e., \( T_{gh} \leq T_{ij} \) or \( \tilde{T}_{gh} \leq \tilde{T}_{ij} \), \( R(t) \) is the at-risk set at
time \( t \) for the nonterminal event, \( \tilde{R}(t) \) is the at-risk set at time \( t \) for the terminal event, \( E[\cdot] \) is the expectation with respect to \( b \), and

\[
w_1(T_{gh}) = \delta_{gh} \frac{D_{11}(T_{gh}, \tilde{T}_{gh})}{D_{12}(T_{gh}, \tilde{T}_{gh})} S_1(T_{gh}) + 1 + \delta_{gh} \left(1 - \tilde{\delta}_{gh} \right) \frac{D_{11}(T_{gh}, \tilde{T}_{gh})}{D_{12}(T_{gh}, \tilde{T}_{gh})} S_1(T_{gh}) + 1 \]

\[
w_2(\tilde{T}_{gh}) = \delta_{gh} \frac{D_{12}(T_{gh}, \tilde{T}_{gh})}{D_{12}(T_{gh}, \tilde{T}_{gh})} S_2(\tilde{T}_{gh}) + 1 + \delta_{gh} \left(1 - \tilde{\delta}_{gh} \right) \frac{D_{12}(T_{gh}, \tilde{T}_{gh})}{D_{12}(T_{gh}, \tilde{T}_{gh})} S_2(\tilde{T}_{gh}) + 1,
\]

in which we resort to notation for derivatives:

\[
D_1(t^{(1)}, t^{(2)}) = \frac{\partial c\{S_1(t^{(1)}), S_2(t^{(2)})\}}{\partial S_1(t^{(1)})}, \quad D_2(t^{(1)}, t^{(2)}) = \frac{\partial c\{S_1(t^{(1)}), S_2(t^{(2)})\}}{\partial S_2(t^{(2)})},
\]

\[
D_{11}(t^{(1)}, t^{(2)}) = \frac{\partial^2 c\{S_1(t^{(1)}), S_2(t^{(2)})\}}{\partial S_1(t^{(1)})^2}, \quad D_{12}(t^{(1)}, t^{(2)}) = \frac{\partial^2 c\{S_1(t^{(1)}), S_2(t^{(2)})\}}{\partial S_1(t^{(1)}) \partial S_2(t^{(2)})},
\]

\[
D_{21}(t^{(1)}, t^{(2)}) = \frac{\partial^2 c\{S_1(t^{(1)}), S_2(t^{(2)})\}}{\partial S_2(t^{(2)})^2}, \quad D_{22}(t^{(1)}, t^{(2)}) = \frac{\partial^3 c\{S_1(t^{(1)}), S_2(t^{(2)})\}}{\partial S_2(t^{(2)})^3},
\]

Thus the cumulative baseline hazard functions for the nonterminal and terminal events are estimated by the following piece-wise constant functions

\[
\hat{\Lambda}_{01}(t) = \sum_{(i,j) \in \tilde{R}(t)} \hat{\Lambda}_{01}\{T_{ij}\},
\]

\[
\hat{\Lambda}_{02}(t) = \sum_{(i,j) \in \tilde{R}(t)} \hat{\Lambda}_{02}\{\tilde{T}_{ij}\},
\]

respectively. Plugging \( \lambda_{01}(t) \) and \( \lambda_{02}(t) \) by jump sizes \( \hat{\Lambda}_{01}\{T_{ij}\} \) and \( \hat{\Lambda}_{01}\{T_{ij}\} \), respectively, into the modified likelihood (6), the updated value of \( \theta \) can be obtained through maximizing the likelihood.

### 4 Implementation

To implement the proposed NPML estimation, we rely on the Monte Carlo Expectation Maximization (MCEM) algorithm [Wei and Tanner 1990; Levine and Casella 2001], in which random effects are treated as missing values to avoid complicated numerical integrations. Particularly, the expec-
tation in the E-step of the algorithm is computed numerically through the MCMC method, such as the acceptance-rejection sampling, while the M-step maximizes the approximated expectation to update parameter estimates.

Let \( \hat{\Lambda} = (\hat{\Lambda}_{01}, \hat{\Lambda}_{02}) \). The complete likelihood corresponding to equation (6) is written as:

\[
L(\theta|O_i, \hat{\Lambda}) = \prod_{i=1}^{m} L_i(\theta|O_i, \hat{\Lambda}) = \prod_{i=1}^{m} \int_{b_i} \psi(b_i; \hat{\Sigma}) \, db_i, 
\]

where

\[
L_i(\theta|O_i, \hat{\Lambda}) = \prod_{j=1}^{n_i} \left[ \frac{\partial^2 e\{\hat{S}_1(T_{ij}), \hat{S}_2(T_{ij})\}}{\partial \hat{S}_1(T_{ij}) \partial \hat{S}_2(T_{ij})} \hat{S}_1(T_{ij}) \hat{\Lambda}_{01} \{T_{ij}\} e^{\hat{\beta}_1 X_{ij} + b_i' Z_{ij}} \hat{S}_2(T_{ij}) \hat{\Lambda}_{02} \{T_{ij}\} e^{\hat{\beta}_2 X_{ij} + b_i' Z_{ij}} \right]^{\hat{\delta}_{ij}^{(1)}} \frac{\partial e\{\hat{S}_1(T_{ij}), \hat{S}_2(T_{ij})\}}{\partial \hat{S}_1(T_{ij})} \hat{S}_1(T_{ij}) \hat{\Lambda}_{01} \{T_{ij}\} e^{\hat{\beta}_1 X_{ij} + b_i' Z_{ij}} \right]^{\hat{\delta}_{ij}^{(2)}} \frac{\partial e\{\hat{S}_1(T_{ij}), \hat{S}_2(T_{ij})\}}{\partial \hat{S}_2(T_{ij})} \hat{S}_2(T_{ij}) \hat{\Lambda}_{02} \{T_{ij}\} e^{\hat{\beta}_2 X_{ij} + b_i' Z_{ij}} \right]^{\hat{\delta}_{ij}^{(3)}}. 
\]

In the E-step, we compute the expectation of the complete likelihood (12) given the observed data and current estimates \( \hat{\omega}^{[k]} = (\hat{\theta}^{[k]}, \hat{\Lambda}^{[k]}, \hat{\Sigma}^{[k]}) \), i.e.

\[
E[L_i(\theta|O_i, \hat{\Lambda})] = \int_{b_i} L_i(\theta|O_i, \hat{\Lambda}^{[k]}, b_i) f(b_i|O_i, \hat{\omega}^{[k]}) \, db_i. 
\]

The calculation is challenging as the expectation has no closed form expression in the semi-competing risks setting. We use the MCMC method to approximate the expectation. We first generate a MCMC sample of size \( Q \), denoted by \( \{b_i^q, q = 1, \cdots, Q\} \), from a distribution with density \( f(b_i|O_i, \hat{\omega}^{[k]}) \) for \( i = 1, \cdots, m \) using the acceptance-rejection sampler, and then estimate the expectation (13) by empirical (sample) mean. Details are given as follows:

**MCMC-step 1:** Compute point \( \tilde{b}_i \) through maximizing \( f(b_i|O_i, \hat{\omega}^{[k]}) \) with respect to \( b_i \),

\[
f(b_i|O_i, \hat{\omega}^{[k]}) = \frac{f(b_i, \hat{\omega}^{[k]}|O_i)}{f(\hat{\omega}^{[k]}|O_i)} = \frac{f(\hat{\omega}^{[k]}|O_i, b_i) \cdot \psi(b_i; \hat{\Sigma}^{[k]})}{\int_{b_i} f(\hat{\omega}^{[k]}|O_i, b_i) \cdot \psi(b_i; \hat{\Sigma}^{[k]}) \, db_i}.
\]

It is equivalent to maximizing \( f(\hat{\omega}^{[k]}|O_i, b_i) \cdot \psi(b_i; \hat{\Sigma}^{[k]}) \) with respect to \( b_i \), where \( f(\hat{\omega}^{[k]}|O_i, b_i) = L_i(\hat{\theta}^{[k]}|O_i, \hat{\Lambda}^{[k]}, b_i) \). By the Newton-Raphson method, we can obtain the value of \( \tilde{b}_i \) and its inverse hessian matrix \( I^{-1}(\tilde{b}_i) \).
**MCMC-step 2**: Use the Acceptance-Rejection method to generate random numbers from a normal distribution with mean \( \tilde{b}_i \) and variance \( I^{-1}(\tilde{b}_i) \). That is, \( b^q_i \sim N(\tilde{b}_i, I^{-1}(\tilde{b}_i)) \) for \( i = 1, \cdots, m \) and \( q = 1, \cdots, Q \). More specifically,

\[
\text{if } u_i \leq \frac{f(b^q_i|O_i, \hat{\omega}^k)}{M_i \phi(b^q_i; I^{-1}(\tilde{b}_i))}, \text{ then accept } b^q_i; \text{ else reject } b^q_i.
\]

where \( u_i \) follows a uniform distribution on \([0, 1]\) and \( M_i \) is a constant specified. We use

\[
M_i = \max_l \left\{ \frac{f(b^q_i|O_i, \hat{\omega}^k)}{\phi(b^q_i; I^{-1}(\tilde{b}_i))} \right\}
\]

in our simulation studies.

**MCMC-step 3**: Based on the MCMC sample, the expectation can be then approximated by

\[
\hat{E}\{L_i(\theta|O_i, \hat{\Lambda}^k)\} = \frac{1}{Q} \sum_{q=1}^{Q} L_i(\theta|O_i, \hat{\Lambda}^k, b^q_i).
\]

**MCMC-step 4**: \( \hat{\Sigma}^{[k+1]} \) can be computed by the sample variance or variance-covariance matrix of MCMC sample \{\( b^q_i, q = 1, \cdots, Q \)\} for \( i = 1, \cdots, m \).

In the M-step, we obtain updated parameter estimates \( \hat{\theta}^{[k+1]} \) through maximizing the expected complete likelihood

\[
\hat{E}\{L(\theta|O_i, \hat{\Lambda}^k)\} = \prod_{i=1}^{n_i} \hat{E}\{L_i(\theta|O_i, \hat{\Lambda}^k)\}
\]

using numerical optimization algorithm. Given the estimates \( \hat{\theta}^{[k+1]} \) we can consequently compute the updated estimate of \( \Lambda, \hat{\Lambda}^{[k+1]} \), based on the NPML estimates in (10)-(11).

Starting with an initial set of estimates \( \hat{\omega}^0 = (\hat{\theta}^0, \hat{\Lambda}^0, \hat{\Sigma}^0) \), the E-step and M-step of the algorithm are iterated until convergence is achieved. That is, the differences between subsequent estimates of \( (\theta, \Sigma) \) are all less than or the score function of \( \theta \) is less than a small quantity \( \epsilon \). The resulting estimators at convergence are denoted by \( \hat{\omega} = (\hat{\theta}, \hat{\Lambda}, \hat{\Sigma}) \). We implement the proposed algorithm in R and use R function optim to perform numerical optimizations involved. Our simulation results show that the algorithm can find good solutions in reasonable computational time. In the case of \( N = 200, m = 10 \) and \( n_i = 20 \), for example, it takes 30 seconds for the calculation of all estimates.

It is worth mentioning that the performance of the proposed MCEM algorithm depends on how well the MCMC sample approximates to the target distribution. To ensure a good approximation,
one can do the following: 1) choose accurate starting values; 2) check if multiple MCMC samples initialized from different initial values give similar results; 3) make sure to run MCMC samples with large size. In this paper, we take initial values of parameters $\beta$ and $\Lambda$ to be the corresponding estimates obtained in the two marginal models using R function `phmm`, while initial values of the variance components are set to be the average of the estimated frailty variances in both marginal models. $\alpha$ is initially set to be 1. Our experience in numerical studies shows that these initial values work well and the convergence of the proposed MCEM algorithm is not sensitive to the choice of the starting value of $\alpha$. In addition, we use the MCMC sample size as 100, which produces simulation results similar to those obtained with bigger MCMC samples of size, say, 200.

Another possible problem of using MCMC sampling to approximate the conditional expectation is that the likelihood based on the MCMC sample may not keep increasing in the EM algorithm. To overcome this problem, a data-driven strategy proposed by Caffo et al. (2005) can be adopted for recovering EM’s ascent (i.e. likelihood increasing) property with high probability. For more details on the data-driven automated MCEM algorithm, we refer the interested reader to Caffo et al. (2005).

5 Theoretical Results

We study the large sample properties of the proposed estimators in the setting where the random effects are assumed to follow a normal or log-gamma distribution with mean 0 and some variance. The normal and log-gamma distributions are two most popular assumptions for the random effects distribution in the literature though other parametric distribution assumptions about random effects are possible.

Before presenting the asymptotic properties of the proposed estimators $\hat{\omega} = (\hat{\theta}, \hat{\Lambda}, \hat{\Sigma})$, we first list the following conditions.

**Condition 1** Conditional on the covariates $X_{ij}$ and $Z_{ij}$, the latent censoring times $C_{ij}$ are independent of the event times $T^{(k)}_{ij} (k = 1, 2)$ and the random effects $b_i$.

**Condition 2** The covariates $X$ and $Z$ are bounded.

**Condition 3** The true value of $\theta$ and $\Sigma$ are elements of the interior of a known compact set $\mathcal{K} = \{ (\theta, \Sigma) : |\theta| \leq B \text{ for some constant } B, \text{ and } \Sigma \text{ is symmetric and positive definite with eigenvalues bounded away from 0 and } \infty. \}$
Condition 4 With probability 1, for \( i = 1, \cdots, m \) and \( j = 1, \cdots, n_i \), \( \Pr(T_{ij} > \varsigma|X,Z) > 0 \), \( \Pr(\delta_{ij}^{(1)} = 0, \delta_{ij}^{(2)} = 0, T_{ij} = \varsigma|X,Z) > 0 \), \( \Pr(\delta_{ij}^{(1)} = 0, \delta_{ij}^{(2)} = 0, T_{ij} = \varsigma|X,Z) > 0 \).

Condition 5 The copula \( c \) has bounded partial derivatives in interval \([0,1]\).

Condition 6 The function \( \Lambda_{0k}(\cdot) \) \((k = 1, 2)\) is strictly increasing and continuously differentiable.

Condition 7 If \( (n_i, \sum_{j=1}^{n_i} X_{ij}')a = 0 \) for \( i = 1, \cdots, m \) and any \((p+1) \times 1\) vector \( a \), then \( a = 0 \).

Condition 8 The number of cluster is larger than \( 1 + p + q(q+1)/2 \), and from all clusters, we can select \( 1 + p + q(q+1)/2 \) clusters denoted by \( i = 1, \cdots, (1 + p + q(q+1)/2) \) such that the determinant of matrix \( M_{(1+p+q(q+1)/2) \times (1+p+q(q+1)/2)} \) is not zero, where the \( i \)th row of \( M \) is \((n_i, \sum_{j=1}^{n_i} X_{ij}, \text{the triangle elements of matrix } (\sum_{j=1}^{n_i} Z_{ij}) / (\sum_{j=1}^{n_i} Z_{ij}))\).

Condition 2 is standard and required for identifiability in the presence of censoring. Conditions 1, 3 are used in Gamst et al. (2009). Condition 4 is used in Chen (2012). Condition 5 is an extra condition to ensure the boundedness of \( \hat{\Lambda} \) in the presence of copula. Condition 6 is required to conduct nonparametric maximum likelihood estimation for \( \Lambda \), which is used in Gamst et al. (2009) and Chen (2012). Conditions 7, 8 are needed for identifiability in the proposed model, which involves two sets of regression parameters. Overall, conditions 1, 8 are standard and realizable.

The consistency and the asymptotic normality of estimators \( \hat{\theta}, \hat{\Lambda} \) and \( \hat{\Sigma} \) are established in the following two theorems.

**Theorem 1 (Consistency) 1)** Assume that conditions 7, 8 hold and random effects \( b_i \) follow a normal distribution \( N(0, \Sigma_0) \) i.i.d. Then \( |\hat{\theta}_m - \theta_0| \to 0 \), \( \|\hat{\Lambda}_{mk} - \Lambda_{0k}\|_{\infty[0,\varsigma]} \to 0 \) \((k = 1, 2)\) and \( |\hat{\Sigma}_m - \Sigma_0| \to 0 \) as \( m \to \infty \), where \((\theta_0, \Lambda_{0k}, \Sigma_0)\) are true values of parameters and \( \| \cdot \|_{\infty[0,\varsigma]} \) is the supremum norm in the interval \([0,\varsigma]\).

2) Assume that conditions 7, 7 hold, \( Z_{ij} = 1 \) and \( u_i = \exp(b_i) \) follow a gamma distribution \( \Gamma(1/\eta_0, \eta_0) \) i.i.d. with mean 1 and variance \( \eta_0 \) for \( i = 1, \cdots, m \), \( j = 1, \cdots, n_i \). Then \( |\hat{\theta}_m - \theta_0| \to 0 \), \( \|\hat{\Lambda}_{mk} - \Lambda_{0k}\|_{\infty[0,\varsigma]} \to 0 \) \((k = 1, 2)\) and \( |\hat{\eta}_m - \eta_0| \to 0 \) as \( m \to \infty \), where \( \theta_0, \Lambda_{0k}, \eta_0 \) are true values of parameters.

**Theorem 2 (Asymptotic Normality) 1)** Assume that conditions 7, 8 hold and random effects \( b_i \) follow a normal distribution \( N(0, \Sigma_0) \) i.i.d.. Then \( \sqrt{m} \times (\hat{\theta}_m - \theta_0, \hat{\Lambda}_m - \Lambda_0, \hat{\Sigma}_m - \Sigma_0) \) converges weakly to a mean-0 Gaussian process as \( m \to \infty \).
2) Assume that conditions [1-7 hold, \( Z_{ij} = 1 \) and \( u_i = \exp(b_i) \) follow a gamma distribution \( \Gamma(1/\eta_0, \eta_0) \) i.i.d. with mean 1 and variance \( \eta_0 \) for \( i = 1, \ldots, m, j = 1, \ldots, n_i \). Then \( \sqrt{m} \times (\hat{\theta}_m - \theta_0, \hat{\Lambda}_m - \Lambda_0, \hat{\eta}_m - \eta_0) \) converges weakly to a mean-0 Gaussian process as \( m \to \infty \).

The detailed proofs of both theorems are provided in the Appendix. In particular, we apply the techniques of counting process and the strong law of large numbers to prove Theorem 1. We first invoke the compactness of the parameter space and Helly’s selection theorem (Murphy, 1995) to conclude the existence of a convergent subsequence, which converges to \( \omega^* = (\theta^*, \Lambda^*, \Sigma^*) \), of our estimators. By Glivenko-Cantelli Theorem (Gamst et al., 2009) and the property of Kullback-Leibler information, we have \( L(\omega^*) = L(\omega_0) \). We then prove the identifiability, that is, \( \omega^* = \omega_0 \). Since our model involves two sets of regression parameters and three counting processes, the identifiability issue is much more challenging than the case with only one counting process. Given that the equality \( L(\omega^*) = L(\omega_0) \) is satisfied for all the cases, it suffices to prove \( \omega^* = \omega_0 \) under some special cases. Therefore, we impose new conditions [7-8] to conclude our estimators converge to true values.

To prove the asymptotic normality in Theorem 2, we first apply the martingale central limit theorem (Pollard, 1984, chap. 8) to show that the derivative of \( L_m(\omega_0) \), denoted by \( \sqrt{m}S_m(\omega_0) \), converges in distribution to a mean-zero Gaussian process. Fréchet differentiability of score function \( S(\omega) \) holds by its smoothness. Then we prove the derivative of the score function, denoted by \( \dot{S}(\omega_0) \), is continuously invertible on its range under conditions [7-8]. By Theorem 2 in Murphy (1995), the asymptotic normality of our estimators follows from the verified four conditions on the score function: convergence to a tight Gaussian process, Fréchet differentiability, invertibility, and the approximation condition.

It is noted that in the proofs of Theorems 1-2, we have provided a general way to characterize the distribution of random effects \( b_i \) and expressed the normal and log-gamma random effects as two special cases. In this way, the theoretical properties can be extended to incorporate other parametric distribution assumptions for random effects under conditions that their corresponding moments required in the proof procedures can be explicitly calculated and identifiability conditions [7-8] need to be updated accordingly.
6 Simulation Studies

In this section, we assess the performance of our proposed method via finite sample simulation studies and compare our proposed method with existing method.

We generate data for the two event times $T^{(1)}$ and $T^{(2)}$ from the proportional hazards models with random effects $\lambda_1(t) = \lambda_{01}(t)\exp\{\beta_1 X + bZ\}$ and $\lambda_2(t) = \lambda_{02}(t)\exp\{\beta_2 X + bZ\}$, respectively, where $\beta_1 = 1$, $\beta_2 = 1$, $\lambda_{01}(t) = 1, \lambda_{02}(t) = 0.4$, which correspond to two exponential marginal distributions. Covariate $X$ is generated from a standard normal distribution truncated at $\pm2$. We consider two different distributions for random effects: $b$ follows a normal distribution with mean 0 and variance $\eta$ and $u = \exp(b)$ follows a gamma distribution $\Gamma(1/\eta, \eta)$ with mean 1 and variance $\eta$. The joint distribution of $(T^{(1)}, T^{(2)})$ is specified by a Gumbel copula with association parameter $\alpha = \frac{1}{(1 - \tau)} - 1$, where Kendall’s tau is set to be $\tau = 0.3, 0.6$. We consider the Gumbel copula model only in the simulation studies based on the fact that the choice of copula does not affect the behavior of the NPML estimation as shown in the numerical analysis by Chen (2012). We generate censoring times $C$ from a uniform distribution $U[0, \varsigma]$, where the duration of the study is chosen to be $\varsigma = 4$. The observed nonterminal and terminal event times are then obtained by $\min(T^{(1)}, T^{(2)}, C)$ and $\min(T^{(2)}, C)$, respectively. This results in $30\% - 60\%$ censoring for $T^{(1)}$ and $20\% - 50\%$ censoring for $T^{(2)}$.

In order to evaluate the finite sample performance of the proposed estimators with different variance component levels and show that it can handle the random effect in the intercept or slope, we consider the following scenarios:

**Scenario 1**: $\eta = 0.2$, $Z = 1$

**Scenario 2**: $\eta = 0.4$, $Z = 1$

**Scenario 3**: $\eta = 0.2$, $Z \sim U(-2, 2)$

**Scenario 4**: $\eta = 0.4$, $Z \sim U(-2, 2)$

Scenarios 1 and 2 present the random intercept cases, while Scenarios 3 and 4 allow the random effect on slope, where the variance component changes from 0.2 to 0.4.

For each scenario, we conduct 200 simulation runs and summarize parameter estimates in terms of the mean bias, the sample standard deviation of estimates (sd) and the coverage percentage for the 95% confidence intervals (cp). Simulation results are reported in Tables 1-4 for the model with normally distributed random effects, i.e. $b_i \sim N(0, \eta)$ i.i.d. in scenarios 1-4, respectively.
In each scenario, the mean biases of the proposed estimators are close to zero with small sd’s and corresponding cp nearly 95% for all settings. In general, when the cluster size \( n_i \) for \( i = 1, \ldots, m \) or the number of clusters \( m \) increases, both the mean bias and sd of the proposed estimators reduce, indicating the improvements in accuracy and efficiency, respectively. For the estimation of \( \tau \), the reduction in relative bias (bias/true value) and sd tends to be more significant when the true value of \( \tau \) becomes larger or association between the two events becomes stronger. This implies that the proposed copula-based modeling approach can better capture the association between both events, which are highly dependent, and thus provides more accurate estimates of \( \tau \).

Comparing the performance between Table 1 and Table 2 or Table 3 and Table 4 for two different levels of the variance of random effects, we observe that with a greater variance component the aforementioned findings still hold, but resulting in relatively larger mean biases and sd’s, as expected. However, when the sample size increases, for example from \( N = 50 \) to \( N = 200 \) in Table 2 and Table 4, the mean biases and sd’s tend to be very small again and thus the differences in estimation induced by the change of the variance component can be ignored.

Next we consider the model with gamma frailty \( u_i = \exp(b_i) \sim \Gamma(1/\eta, \eta) \) i.i.d.. For brief illustration, we here report simulation results in Table 5 and Table 6 for scenarios 1-2, respectively. Findings are similar to those with normal random effects shown in Tables 1-2. Corresponding results for scenarios 3-4 are presented in the web-based Supplementary Material, exhibiting similar trends to those observed in Tables 3-4. Since Emura et al. (2015)’s approach can be applied to the case with gamma frailty, we also present the corresponding results obtained by their approach using R package joint.Cox in Table 5 and Table 6. The proposed method yields better estimates of variance component \( \eta \) and Kendall’s \( \tau \) than their approach in terms of both mean bias and sd, while estimates of regression coefficients \( \beta_1 \) and \( \beta_2 \) obtained from both methods are comparable in terms of root mean square error (rmse) for some small sample sizes. When the total sample size increases to \( N = 200 \), our method outperforms theirs in terms of rmse for estimating all parameters in scenario 1, regardless of association level. With a large variance component in scenario 2, we note from Table 6 that the proposed method leads to slightly worse estimates of \( \eta \) and comparable estimates for the other parameters in comparison with the existing counterpart. As association \( \tau \) increases, our estimates of \( \eta \) and \( \tau \) significantly improve in terms of rmse.
To visualize the performance of estimated cumulative baseline hazard functions, we show curves of $\hat{\Lambda}_01(\cdot)$ and $\hat{\Lambda}_02(\cdot)$ obtained by the proposed method in Figures 1-2, respectively, based on the first 20 simulation runs in scenario 1. The straight black lines in figures represent the true cumulative baseline hazard function and the red curves are the estimated cumulative baseline hazard functions. The corresponding curves estimated by the spline-based method of Emura et al. (2015) are also shown in Figures 3-4 for comparison purpose. Across all the settings, the proposed method provides the estimated cumulative baseline hazard functions close to the corresponding true function and the biases reduce as the number of clusters and/or the cluster size increases. Compared to the corresponding curves in Figures 3-4, the proposed method yields the estimates of the cumulative baseline hazards similar to those smooth estimates. When the sample size becomes large, our method slightly improves the performance of the cumulative baseline hazard function estimation to some extent.

Moreover, we also examine the performance of the proposed estimators under a misspecified distribution of random effects. Table 7 reports estimation results under the proposed model with normally distributed random effects in Scenario 1 when the true distribution of random effects is log-gamma, that is $\exp(b_i)$ follows a gamma distribution. Comparing to the corresponding results in Table 5 with the correctly specified distribution of random effects, there is limited impact of such a misspecification on estimation efficiency of variance component $\eta$ only especially when the value of $\tau$ is large, while estimation for the other parameters seems not sensitive to the misspecification. In addition, when the true distribution of random effects is a mixture of normal distributions $0.5N(0, 0.2) + 0.5N(0, 0.4)$, we provide estimation results in Table 8 obtained by both the proposed and existing methods with a misspecified log-gamma distribution of random effects. It is found that the existing method fails to provide valid estimates of all parameters, with over 30% divergent cases and unreasonably big biases and sd’s for those convergent cases in the presence of some outliers caused by the mixture normal structure of random effects. On the other hand, except for having slightly greater sd’s for estimates of $\eta$, our method still works well in estimating parameters. These indicate that the proposed method is more robust against outliers and random effects distribution misspecification than the existing approach.
Tables 7-8 about here

We have also considered the Weibull baseline hazard functions for the two marginal models. The corresponding simulation results are provided in the web-based supplementary due to space limit, and exhibit similar trends to those seen in the exponential marginal model design above.

7 Real Data Analysis

For illustration purpose, we now apply the proposed method to analysis of data from the multi-institutional clinical studies of breast cancer, conducted by 36 cancer centers with total 5715 breast cancer patients in US and Europe ([Haibe-Kains et al., 2012]). The databases are available at [http://compbio.dfci.harvard.edu/pubs/sbtpaper/](http://compbio.dfci.harvard.edu/pubs/sbtpaper/). The aims are twofold: to evaluate the classification concordance based on genetic features and to assess the prognostic values of both clinical and genetic characteristics in predicting the disease outcome of patients. In the current analysis, we are interested in the later one. To assist treatment decisions for breast cancer, both clinical features and gene expression profiles of patients have also been used to predict patient survival outcome ([Van De Vijver et al., 2002] [Van’t Veer et al., 2002]).

Data from 6 different patient cohorts, which contains valid observations (either censored or completed) of the nonterminal and terminal event times, are extracted from the available databases provided by 6 institutions including CAL dataset of breast cancer patients from the university of California, San Francisco, and the California Pacific Medical Center (United States); NKI dataset from National Kanker Institute (the Netherlands); STNO dataset from Stanford/Norway (United States and Norway); TRANSBIG dataset collected by the TransBIG consortium (Europe); UCSF dataset from University of California, San Francisco (United States); UNC dataset from University of North Carolina(United States). The survival outcomes of our interest are the time for distant metastasis-free survival (DMFS) defined as the duration from the initial diagnosis to the time at which a distant metastasis was detected, and the time for overall survival (OS) defined as the duration from initial diagnosis to death. DMFS and OS are considered under semi-competing risks, where DMFS is the nonterminal event and OS is the terminal event. We are interested in how the clinical information of breast cancer patients affects both two events and the correlation between these two event times. The clinical features include estrogen receptor status (\(er, 1=\text{positive and } 0=\text{negative}\)), tumor size in cm (\(size, \text{the longest length of invasive tumor measured by pathologists or surgeons.}\)), lymph nodal status (\(node, 1=\text{present and } 0=\text{absent}\)), age at diagnosis (\(age, \text{in years}\)),
histological grade (grade, 1=invasive ductal carcinoma, 2=other invasive carcinoma, 3=noninvasive ductal carcinoma in situ and other noninvasive carcinoma; or 1=low, 2=intermediate, 3=high). There are only \( n = 784 \) patients who had complete information in all five clinical features. They were aged 24 to 89. Among them, 103 were from CAL dataset, 122 from NKI dataset, 88 from STNO dataset, 190 from TRANSBIG dataset, 93 from UCSF dataset and 188 from UNC dataset.

We analyze the data using the proposed method and two naive methods, marginal proportional hazards models with/without frailty for nonterminal and terminal events separately. Table 9 reports the estimates and their corresponding standard errors. The standard deviation is estimated by 500 randomly selected bootstrap samples using the stratified sampling method given by Shao et al. (2003). For each cluster, the sample of original cluster size is first generated by with-replacement bootstrap, independently across clusters. These independent samples together form a bootstrap population sample. For comparison purpose, two naive methods are employed to fit the nonterminal DMFS and terminal OS separately: the usual Cox PH model with corresponding estimates obtained by R function `coxph` and the PH mixed-effects model \( \lambda_k(t|b; X, Z) = \lambda_0(t)e^{\beta_kX + b_k} \) for \( k = 1, 2 \) using R function `phmm`. All these three methods indicate that prognostic factors, estrogen receptor status and histological grade, are statistically significant (at the 5% level) in association with the DMFS and OS. However, after accounting for the dependence between the DMFS and OS, we find that covariate age appears insignificant in the proposed model, unlike it is in both naive models. The estimated value of \( \eta \) is quite large and \( \hat{\tau} = 0.5974 \) with strong significance. Both results show strong evidence of significant cluster effects and relatively high correlation between the DMFS and OS in this study. Thus, the proposed method is a more suitable than the naive marginal methods for analysis of this data set.

Table 9 about here

8 Discussion

We have proposed a copula regression model with random effects for joint analysis of both non-terminal and terminal event times under semi-competing risks and clustered structure. The maximum likelihood estimation (MLE) is developed for regression parameters, while the baseline hazard functions are estimated using the NPML estimation method. The proposed method extends the joint regression analysis studied in Chen (2012) to the clustered semicompeting risks data setting. A key feature of our method is that it does not require the random effects to be gamma dis-
distributed as is in the work of Emura et al. (2015), and thus allows for a flexible modeling framework and wider range of applications. We provide rigorous justification of the estimators through the counting process and martingale central limit theory. Our theoretical results show that the estimators have desired asymptotic properties, which facilitate further inference. The proposed estimation procedure is implemented through an MCEM algorithm, which generally results in stable estimates in our experiment with both simulated and real data sets. Furthermore, our simulation results show that the proposed method not only performs very well in terms of the bias and standard error for either gamma or normally distributed random effects, but also exhibits robustness against misspecification of the random effects distribution. In addition, under the model with random effects being gamma distributed, our method provides more favorable estimates of the cumulative baseline hazard functions than and comparable estimates of model parameters with the existing approach.

Acknowledgement

This research was supported by the Singapore Ministry of Education Academic Research Fund Tier 2 Grant (MOE2013-T2-2-118) and Tier 1 Grant (RG30/12). The authors would like to thank the two anonymous referees for their many constructive comments and suggestions on the earlier version of this paper.

References


Table 1: Simulation results based on 200 replications under the model with normally distributed random effects in the intercept in Scenario 1, where parameters $\eta = 0.2$, $\beta_1 = 1$, $\beta_2 = 1$, $\lambda_{01}(t) = 1$, $\lambda_{02}(t) = 0.4$, and Kendall’s $\tau = 0.3$ and 0.6.

<table>
<thead>
<tr>
<th>$(N,m,n_i)$</th>
<th>$\eta$</th>
<th>$\tau$</th>
<th>$\beta_1$</th>
<th>$\beta_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\tau = 0.3$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$(50,5,10)$</td>
<td>bias</td>
<td>$-0.0632$</td>
<td>$0.0196$</td>
<td>$0.0185$</td>
</tr>
<tr>
<td></td>
<td>sd</td>
<td>$0.1834$</td>
<td>$0.1063$</td>
<td>$0.2846$</td>
</tr>
<tr>
<td></td>
<td>cp(%)</td>
<td>95.0</td>
<td>94.0</td>
<td>94.5</td>
</tr>
<tr>
<td>$(100,5,20)$</td>
<td>bias</td>
<td>$-0.0406$</td>
<td>$-0.0100$</td>
<td>$0.0206$</td>
</tr>
<tr>
<td></td>
<td>sd</td>
<td>$0.1613$</td>
<td>$0.0854$</td>
<td>$0.1818$</td>
</tr>
<tr>
<td></td>
<td>cp(%)</td>
<td>96.0</td>
<td>95.5</td>
<td>94.0</td>
</tr>
<tr>
<td>$(100,10,10)$</td>
<td>bias</td>
<td>$-0.0054$</td>
<td>$0.0190$</td>
<td>$-0.0007$</td>
</tr>
<tr>
<td></td>
<td>sd</td>
<td>$0.1477$</td>
<td>$0.0799$</td>
<td>$0.1913$</td>
</tr>
<tr>
<td></td>
<td>cp(%)</td>
<td>96.5</td>
<td>95.5</td>
<td>93.5</td>
</tr>
<tr>
<td>$(200,10,20)$</td>
<td>bias</td>
<td>$-0.0207$</td>
<td>$0.0182$</td>
<td>$-0.0046$</td>
</tr>
<tr>
<td></td>
<td>sd</td>
<td>$0.1250$</td>
<td>$0.0577$</td>
<td>$0.1200$</td>
</tr>
<tr>
<td></td>
<td>cp(%)</td>
<td>96.5</td>
<td>95.0</td>
<td>95.5</td>
</tr>
<tr>
<td>$\tau = 0.6$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$(50,5,10)$</td>
<td>bias</td>
<td>$-0.0102$</td>
<td>$-0.0046$</td>
<td>$0.0797$</td>
</tr>
<tr>
<td></td>
<td>sd</td>
<td>$0.2911$</td>
<td>$0.0963$</td>
<td>$0.6790$</td>
</tr>
<tr>
<td></td>
<td>cp(%)</td>
<td>98.5</td>
<td>96.5</td>
<td>99.0</td>
</tr>
<tr>
<td>$(100,5,20)$</td>
<td>bias</td>
<td>$-0.0533$</td>
<td>$0.0023$</td>
<td>$0.0427$</td>
</tr>
<tr>
<td></td>
<td>sd</td>
<td>$0.1495$</td>
<td>$0.0563$</td>
<td>$0.1860$</td>
</tr>
<tr>
<td></td>
<td>cp(%)</td>
<td>96.0</td>
<td>95.0</td>
<td>93.0</td>
</tr>
<tr>
<td>$(100,10,10)$</td>
<td>bias</td>
<td>$-0.0270$</td>
<td>$-0.0016$</td>
<td>$0.0465$</td>
</tr>
<tr>
<td></td>
<td>sd</td>
<td>$0.1607$</td>
<td>$0.0571$</td>
<td>$0.1908$</td>
</tr>
<tr>
<td></td>
<td>cp(%)</td>
<td>97.0</td>
<td>94.5</td>
<td>93.0</td>
</tr>
<tr>
<td>$(200,10,20)$</td>
<td>bias</td>
<td>$-0.0251$</td>
<td>$0.0006$</td>
<td>$0.0385$</td>
</tr>
<tr>
<td></td>
<td>sd</td>
<td>$0.1177$</td>
<td>$0.0388$</td>
<td>$0.1347$</td>
</tr>
<tr>
<td></td>
<td>cp(%)</td>
<td>96.0</td>
<td>94.0</td>
<td>94.0</td>
</tr>
</tbody>
</table>
Table 2: Simulation results based on 200 replications under the model with normally distributed random effects in the intercept in Scenario 2, where parameters $\eta = 0.4$, $\beta_1 = 1$, $\beta_2 = 1$, $\lambda_{01}(t) = 1$, $\lambda_{02}(t) = 0.4$ and Kendall’s $\tau = 0.3$ and $0.6$.

<table>
<thead>
<tr>
<th>$(N, m, n_i)$</th>
<th>$\eta$</th>
<th>$\tau$</th>
<th>$\beta_1$</th>
<th>$\beta_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\tau = 0.3$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(50, 5, 10)</td>
<td>bias</td>
<td>-0.1181</td>
<td>0.0112</td>
<td>0.0541</td>
</tr>
<tr>
<td></td>
<td>sd</td>
<td>0.3187</td>
<td>0.1108</td>
<td>0.2553</td>
</tr>
<tr>
<td></td>
<td>cp(%)</td>
<td>96.5</td>
<td>94.5</td>
<td>93.5</td>
</tr>
<tr>
<td>(100, 5, 20)</td>
<td>bias</td>
<td>-0.0751</td>
<td>0.0135</td>
<td>-0.0192</td>
</tr>
<tr>
<td></td>
<td>sd</td>
<td>0.2785</td>
<td>0.0846</td>
<td>0.1657</td>
</tr>
<tr>
<td></td>
<td>cp(%)</td>
<td>95.0</td>
<td>96.0</td>
<td>95.5</td>
</tr>
<tr>
<td>(100, 10, 10)</td>
<td>bias</td>
<td>-0.0540</td>
<td>0.0205</td>
<td>-0.0149</td>
</tr>
<tr>
<td></td>
<td>sd</td>
<td>0.2242</td>
<td>0.0777</td>
<td>0.1835</td>
</tr>
<tr>
<td></td>
<td>cp(%)</td>
<td>95.5</td>
<td>94.5</td>
<td>95.5</td>
</tr>
<tr>
<td>(200, 10, 20)</td>
<td>bias</td>
<td>-0.0334</td>
<td>0.0151</td>
<td>-0.0086</td>
</tr>
<tr>
<td></td>
<td>sd</td>
<td>0.2051</td>
<td>0.0594</td>
<td>0.1166</td>
</tr>
<tr>
<td></td>
<td>cp(%)</td>
<td>97.0</td>
<td>93.5</td>
<td>96.5</td>
</tr>
<tr>
<td>$\tau = 0.6$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(50, 5, 10)</td>
<td>bias</td>
<td>-0.0998</td>
<td>-0.0128</td>
<td>0.0738</td>
</tr>
<tr>
<td></td>
<td>sd</td>
<td>0.3216</td>
<td>0.0841</td>
<td>0.3082</td>
</tr>
<tr>
<td></td>
<td>cp(%)</td>
<td>96.5</td>
<td>95.5</td>
<td>94.0</td>
</tr>
<tr>
<td>(100, 5, 20)</td>
<td>bias</td>
<td>-0.0997</td>
<td>-0.0012</td>
<td>0.0418</td>
</tr>
<tr>
<td></td>
<td>sd</td>
<td>0.2264</td>
<td>0.0570</td>
<td>0.1860</td>
</tr>
<tr>
<td></td>
<td>cp(%)</td>
<td>98.0</td>
<td>95.0</td>
<td>94.5</td>
</tr>
<tr>
<td>(100, 10, 10)</td>
<td>bias</td>
<td>-0.0620</td>
<td>-0.0018</td>
<td>0.0411</td>
</tr>
<tr>
<td></td>
<td>sd</td>
<td>0.2341</td>
<td>0.0565</td>
<td>0.1889</td>
</tr>
<tr>
<td></td>
<td>cp(%)</td>
<td>97.0</td>
<td>95.5</td>
<td>95.5</td>
</tr>
<tr>
<td>(200, 10, 20)</td>
<td>bias</td>
<td>-0.0763</td>
<td>-0.0062</td>
<td>0.0374</td>
</tr>
<tr>
<td></td>
<td>sd</td>
<td>0.1796</td>
<td>0.0386</td>
<td>0.1304</td>
</tr>
<tr>
<td></td>
<td>cp(%)</td>
<td>95.5</td>
<td>93.5</td>
<td>94.5</td>
</tr>
</tbody>
</table>
Table 3: Simulation results based on 200 replications under the model with normally distributed random effects in the slope in Scenario 3, where parameters $\eta = 0.2$, $\beta_1 = 1$, $\beta_2 = 1$, $\lambda_{01}(t) = 1$, $\lambda_{02}(t) = 0.4$ and Kendall’s $\tau = 0.3$ and 0.6.

<table>
<thead>
<tr>
<th>$(N,m,n_i)$</th>
<th>$\tau = 0.3$</th>
<th>$\tau = 0.6$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\eta$</td>
<td>$\tau$</td>
</tr>
<tr>
<td>(50, 5, 10)</td>
<td>bias</td>
<td>0.0243</td>
</tr>
<tr>
<td></td>
<td>sd</td>
<td>0.2091</td>
</tr>
<tr>
<td></td>
<td>cp(%)</td>
<td>94.5</td>
</tr>
<tr>
<td>(100, 5, 20)</td>
<td>bias</td>
<td>−0.0004</td>
</tr>
<tr>
<td></td>
<td>sd</td>
<td>0.1535</td>
</tr>
<tr>
<td></td>
<td>cp(%)</td>
<td>96.0</td>
</tr>
<tr>
<td>(100, 10, 10)</td>
<td>bias</td>
<td>0.0113</td>
</tr>
<tr>
<td></td>
<td>sd</td>
<td>0.1484</td>
</tr>
<tr>
<td></td>
<td>cp(%)</td>
<td>95.0</td>
</tr>
<tr>
<td>(200, 10, 20)</td>
<td>bias</td>
<td>0.0095</td>
</tr>
<tr>
<td></td>
<td>sd</td>
<td>0.1132</td>
</tr>
<tr>
<td></td>
<td>cp(%)</td>
<td>95.0</td>
</tr>
</tbody>
</table>
Table 4: Simulation results based on 200 replications under the model with normally distributed random effects in the slope in Scenario 4, where parameters $\eta = 0.4$, $\beta_1 = 1$, $\beta_2 = 1$, $\lambda_0(t) = 1$, $\lambda_0(t) = 0.4$ and Kendall’s $\tau = 0.3$ and $0.6$.

<table>
<thead>
<tr>
<th>$(N, m, n_i)$</th>
<th>$\tau = 0.3$</th>
<th>$\tau = 0.6$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$(N, m, n_i)$</td>
<td>bias</td>
<td>$\eta$</td>
</tr>
<tr>
<td>$\text{bias}$</td>
<td>0.0207</td>
<td>0.0134</td>
</tr>
<tr>
<td>$\text{sd}$</td>
<td>0.3363</td>
<td>0.1120</td>
</tr>
<tr>
<td>$\text{cp}(%)$</td>
<td>94.5</td>
<td>94.0</td>
</tr>
<tr>
<td>$(100, 5, 20)$</td>
<td>bias</td>
<td>0.0096</td>
</tr>
<tr>
<td>$\text{sd}$</td>
<td>0.3053</td>
<td>0.0855</td>
</tr>
<tr>
<td>$\text{cp}(%)$</td>
<td>94.0</td>
<td>95.5</td>
</tr>
<tr>
<td>$(100, 10, 10)$</td>
<td>bias</td>
<td>0.0095</td>
</tr>
<tr>
<td>$\text{sd}$</td>
<td>0.2546</td>
<td>0.0862</td>
</tr>
<tr>
<td>$\text{cp}(%)$</td>
<td>92.0</td>
<td>95.5</td>
</tr>
<tr>
<td>$(200, 10, 20)$</td>
<td>bias</td>
<td>-0.0045</td>
</tr>
<tr>
<td>$\text{sd}$</td>
<td>0.2141</td>
<td>0.0621</td>
</tr>
<tr>
<td>$\text{cp}(%)$</td>
<td>95.5</td>
<td>94.5</td>
</tr>
<tr>
<td>$(50, 5, 10)$</td>
<td>bias</td>
<td>0.0199</td>
</tr>
<tr>
<td>$\text{sd}$</td>
<td>0.3347</td>
<td>0.1048</td>
</tr>
<tr>
<td>$\text{cp}(%)$</td>
<td>94.5</td>
<td>96.5</td>
</tr>
<tr>
<td>$(100, 5, 20)$</td>
<td>bias</td>
<td>0.0202</td>
</tr>
<tr>
<td>$\text{sd}$</td>
<td>0.2761</td>
<td>0.0589</td>
</tr>
<tr>
<td>$\text{cp}(%)$</td>
<td>97.0</td>
<td>96.5</td>
</tr>
<tr>
<td>$(100, 10, 10)$</td>
<td>bias</td>
<td>0.0157</td>
</tr>
<tr>
<td>$\text{sd}$</td>
<td>0.2580</td>
<td>0.0722</td>
</tr>
<tr>
<td>$\text{cp}(%)$</td>
<td>96.5</td>
<td>97.0</td>
</tr>
<tr>
<td>$(200, 10, 20)$</td>
<td>bias</td>
<td>0.0112</td>
</tr>
<tr>
<td>$\text{sd}$</td>
<td>0.2161</td>
<td>0.0594</td>
</tr>
<tr>
<td>$\text{cp}(%)$</td>
<td>95.0</td>
<td>98.0</td>
</tr>
</tbody>
</table>
Table 5: Simulation results for the model with gamma frailty obtained by the proposed and Emura et al. (2015)’s methods based on 200 replications in Scenario 1, with parameters $\eta = 0.2$, $\beta_1 = 1$ and $\beta_2 = 1$, $\lambda_0(t) = 1$, $\lambda_0(t) = 0.4$ and Kendall’s $\tau = 0.3$ and 0.6.

<table>
<thead>
<tr>
<th>$(N, m, n_i)$</th>
<th>$\tau = 0.3$</th>
<th>Proposed method</th>
<th>Emura et al (2015)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\eta$</td>
<td>$\beta_1$</td>
<td>$\beta_2$</td>
</tr>
<tr>
<td>(50, 5, 10)</td>
<td>bias</td>
<td>0.0030</td>
<td>0.0095</td>
</tr>
<tr>
<td></td>
<td>sd</td>
<td>0.1918</td>
<td>0.1078</td>
</tr>
<tr>
<td></td>
<td>rmse</td>
<td>0.1919</td>
<td>0.1082</td>
</tr>
<tr>
<td></td>
<td>cp(%)</td>
<td>93.0</td>
<td>96.0</td>
</tr>
<tr>
<td>(100, 5, 20)</td>
<td>bias</td>
<td>-0.0267</td>
<td>0.0195</td>
</tr>
<tr>
<td></td>
<td>sd</td>
<td>0.1600</td>
<td>0.0827</td>
</tr>
<tr>
<td></td>
<td>rmse</td>
<td>0.1622</td>
<td>0.0850</td>
</tr>
<tr>
<td></td>
<td>cp(%)</td>
<td>96.5</td>
<td>95.0</td>
</tr>
<tr>
<td>(100, 10, 10)</td>
<td>bias</td>
<td>0.0322</td>
<td>0.0114</td>
</tr>
<tr>
<td></td>
<td>sd</td>
<td>0.1611</td>
<td>0.0815</td>
</tr>
<tr>
<td></td>
<td>rmse</td>
<td>0.1643</td>
<td>0.0823</td>
</tr>
<tr>
<td></td>
<td>cp(%)</td>
<td>94.0</td>
<td>94.5</td>
</tr>
<tr>
<td>(200, 10, 20)</td>
<td>bias</td>
<td>0.0133</td>
<td>0.0074</td>
</tr>
<tr>
<td></td>
<td>sd</td>
<td>0.1517</td>
<td>0.0640</td>
</tr>
<tr>
<td></td>
<td>rmse</td>
<td>0.1523</td>
<td>0.0644</td>
</tr>
<tr>
<td></td>
<td>cp(%)</td>
<td>93.0</td>
<td>93.5</td>
</tr>
<tr>
<td>$\tau = 0.6$</td>
<td>bias</td>
<td>-0.0407</td>
<td>0.0063</td>
</tr>
<tr>
<td>(50, 5, 10)</td>
<td>sd</td>
<td>0.0551</td>
<td>0.1007</td>
</tr>
<tr>
<td></td>
<td>rmse</td>
<td>0.0685</td>
<td>0.1009</td>
</tr>
<tr>
<td></td>
<td>cp(%)</td>
<td>93.0</td>
<td>95.0</td>
</tr>
<tr>
<td>(100, 5, 20)</td>
<td>bias</td>
<td>-0.0432</td>
<td>0.0161</td>
</tr>
<tr>
<td></td>
<td>sd</td>
<td>0.0742</td>
<td>0.0596</td>
</tr>
<tr>
<td></td>
<td>rmse</td>
<td>0.0859</td>
<td>0.0617</td>
</tr>
<tr>
<td></td>
<td>cp(%)</td>
<td>97.5</td>
<td>95.0</td>
</tr>
<tr>
<td>(100, 10, 10)</td>
<td>bias</td>
<td>-0.0195</td>
<td>0.0149</td>
</tr>
<tr>
<td></td>
<td>sd</td>
<td>0.0512</td>
<td>0.0626</td>
</tr>
<tr>
<td></td>
<td>rmse</td>
<td>0.0548</td>
<td>0.0646</td>
</tr>
<tr>
<td></td>
<td>cp(%)</td>
<td>94.5</td>
<td>97.5</td>
</tr>
<tr>
<td>(200, 10, 20)</td>
<td>bias</td>
<td>-0.0275</td>
<td>0.0090</td>
</tr>
<tr>
<td></td>
<td>sd</td>
<td>0.0562</td>
<td>0.0387</td>
</tr>
<tr>
<td></td>
<td>rmse</td>
<td>0.0626</td>
<td>0.0397</td>
</tr>
<tr>
<td></td>
<td>cp(%)</td>
<td>94.5</td>
<td>94.5</td>
</tr>
</tbody>
</table>
Table 6: Simulation results for the model with gamma frailty obtained by the proposed and Emura et al. (2015)'s methods based on 200 replications in Scenario 2, with parameters $\eta = 0.4$, $\beta_1 = 1$ and $\beta_2 = 1$, $\lambda_0(t) = 1$, $\lambda_0(t) = 0.4$ and Kendall’s $\tau = 0.3$ and 0.6.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\eta$ $\tau$ $\beta_1$ $\beta_2$</td>
<td>$\eta$ $\tau$ $\beta_1$ $\beta_2$</td>
</tr>
<tr>
<td>$\tau = 0.3$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(50, 5, 10)</td>
<td>bias -0.0229 0.0203 0.0411 0.0234</td>
<td>-0.0320 0.1281 0.0341 0.0396</td>
</tr>
<tr>
<td></td>
<td>sd 0.3617 0.1148 0.2719 0.3161</td>
<td>0.3354 0.1890 0.2842 0.2879</td>
</tr>
<tr>
<td></td>
<td>rmse 0.3624 0.1166 0.2750 0.3170</td>
<td>0.3369 0.2283 0.2862 0.2906</td>
</tr>
<tr>
<td></td>
<td>cp(%) 94.5 94.5 94.5 94.0</td>
<td>95.0 93.5 95.5 93.5</td>
</tr>
<tr>
<td>(100, 5, 20)</td>
<td>bias 0.0067 0.0157 0.0181 0.0298</td>
<td>-0.0684 0.0672 0.0181 0.0208</td>
</tr>
<tr>
<td></td>
<td>sd 0.4071 0.0827 0.1878 0.2131</td>
<td>0.2573 0.1719 0.1550 0.1867</td>
</tr>
<tr>
<td></td>
<td>rmse 0.4071 0.0842 0.1887 0.2152</td>
<td>0.2662 0.1846 0.1561 0.1878</td>
</tr>
<tr>
<td></td>
<td>cp(%) 94.5 94.5 91.5 95.0</td>
<td>97.0 94.5 93.5 94.5</td>
</tr>
<tr>
<td>(100, 10, 10)</td>
<td>bias 0.0821 0.0246 0.0255 0.0388</td>
<td>-0.0156 0.0478 0.0325 0.0130</td>
</tr>
<tr>
<td></td>
<td>sd 0.3551 0.0910 0.1863 0.2172</td>
<td>0.2350 0.1736 0.1823 0.1887</td>
</tr>
<tr>
<td></td>
<td>rmse 0.3551 0.0942 0.1880 0.2207</td>
<td>0.2355 0.1800 0.1852 0.1892</td>
</tr>
<tr>
<td></td>
<td>cp(%) 93.5 95.0 94.5 96.0</td>
<td>95.0 97.5 93.0 93.0</td>
</tr>
<tr>
<td>(200, 10, 20)</td>
<td>bias 0.0633 0.0146 0.0059 0.0131</td>
<td>-0.0155 -0.0001 0.0061 0.0122</td>
</tr>
<tr>
<td></td>
<td>sd 0.3089 0.0660 0.1218 0.1377</td>
<td>0.2143 0.1406 0.1205 0.1233</td>
</tr>
<tr>
<td></td>
<td>rmse 0.3184 0.0675 0.1219 0.1383</td>
<td>0.2149 0.1406 0.1206 0.1239</td>
</tr>
<tr>
<td></td>
<td>cp(%) 94.0 96.0 96.0 95.0</td>
<td>97.0 95.5 97.5 92.5</td>
</tr>
<tr>
<td>$\tau = 0.6$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(50, 5, 10)</td>
<td>bias -0.0577 0.0257 -0.0059 -0.0644</td>
<td>-0.0212 -0.0019 0.0709 0.0826</td>
</tr>
<tr>
<td></td>
<td>sd 0.1917 0.0772 0.2401 0.2645</td>
<td>0.3841 0.1563 0.3018 0.3248</td>
</tr>
<tr>
<td></td>
<td>rmse 0.2002 0.0814 0.2402 0.2722</td>
<td>0.3846 0.1563 0.3100 0.3351</td>
</tr>
<tr>
<td></td>
<td>cp(%) 95.5 95.5 97.0 95.5</td>
<td>96.0 95.0 92.5 95.0</td>
</tr>
<tr>
<td>(100, 5, 20)</td>
<td>bias -0.0949 0.0163 -0.0196 -0.0661</td>
<td>-0.0258 -0.0129 0.0301 0.0404</td>
</tr>
<tr>
<td></td>
<td>sd 0.2010 0.0506 0.1620 0.1952</td>
<td>0.3379 0.1294 0.1810 0.2054</td>
</tr>
<tr>
<td></td>
<td>rmse 0.2222 0.0532 0.1632 0.2061</td>
<td>0.3389 0.1300 0.1835 0.2093</td>
</tr>
<tr>
<td></td>
<td>cp(%) 97.0 94.0 96.0 94.5</td>
<td>95.0 94.0 93.0 95.5</td>
</tr>
<tr>
<td>(100, 10, 10)</td>
<td>bias -0.0073 0.0245 -0.0509 -0.0890</td>
<td>-0.0050 -0.0086 0.0450 0.0624</td>
</tr>
<tr>
<td></td>
<td>sd 0.1393 0.0525 0.1910 0.2091</td>
<td>0.2562 0.1351 0.1891 0.2028</td>
</tr>
<tr>
<td></td>
<td>rmse 0.1395 0.0579 0.1977 0.2273</td>
<td>0.2563 0.1353 0.1944 0.2122</td>
</tr>
<tr>
<td></td>
<td>cp(%) 96.5 92.0 95.0 93.5</td>
<td>96.0 95.0 96.0 94.0</td>
</tr>
<tr>
<td>(200, 10, 20)</td>
<td>bias -0.0226 -0.0099 0.0254 -0.0313</td>
<td>-0.0111 -0.0172 0.0134 0.0213</td>
</tr>
<tr>
<td></td>
<td>sd 0.1679 0.0391 0.1215 0.1331</td>
<td>0.2393 0.1029 0.1162 0.1196</td>
</tr>
<tr>
<td></td>
<td>rmse 0.1695 0.0391 0.1241 0.1367</td>
<td>0.2395 0.1043 0.1170 0.1215</td>
</tr>
<tr>
<td></td>
<td>cp(%) 97.5 95.0 94.5 94.0</td>
<td>96.0 94.5 94.5 96.0</td>
</tr>
</tbody>
</table>
Table 7: Simulation results obtained by the proposed methods using a misspecified normal random effect model in Scenario 1 based on 200 replications when the true random effects distribution is gamma. The true parameters $\eta = 0.2$, $\beta_1 = 1$ and $\beta_2 = 1$, $\lambda_{01}(t) = 1$, $\lambda_{02}(t) = 0.4$ and Kendall’s $\tau = 0.3$ and 0.6.

<table>
<thead>
<tr>
<th>$(N, m, n_i)$</th>
<th>$\eta$</th>
<th>$\tau$</th>
<th>$\beta_1$</th>
<th>$\beta_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\tau = 0.3$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(50, 5, 10)</td>
<td>bias</td>
<td>-0.0434</td>
<td>0.0233</td>
<td>0.0439</td>
</tr>
<tr>
<td></td>
<td>sd</td>
<td>0.2035</td>
<td>0.1174</td>
<td>0.3131</td>
</tr>
<tr>
<td></td>
<td>cp(%)</td>
<td>97.0</td>
<td>95.5</td>
<td>93.5</td>
</tr>
<tr>
<td>(100, 10, 10)</td>
<td>bias</td>
<td>-0.0467</td>
<td>0.0191</td>
<td>0.0065</td>
</tr>
<tr>
<td></td>
<td>sd</td>
<td>0.1755</td>
<td>0.0897</td>
<td>0.1750</td>
</tr>
<tr>
<td></td>
<td>cp(%)</td>
<td>97.5</td>
<td>94.0</td>
<td>93.5</td>
</tr>
<tr>
<td>(100, 5, 20)</td>
<td>bias</td>
<td>-0.0125</td>
<td>0.0158</td>
<td>0.0258</td>
</tr>
<tr>
<td></td>
<td>sd</td>
<td>0.1796</td>
<td>0.0921</td>
<td>0.1971</td>
</tr>
<tr>
<td></td>
<td>cp(%)</td>
<td>95.0</td>
<td>95.5</td>
<td>95.5</td>
</tr>
<tr>
<td>(200, 10, 20)</td>
<td>bias</td>
<td>-0.0080</td>
<td>0.0174</td>
<td>-0.0120</td>
</tr>
<tr>
<td></td>
<td>sd</td>
<td>0.1412</td>
<td>0.0578</td>
<td>0.1162</td>
</tr>
<tr>
<td></td>
<td>cp(%)</td>
<td>95.5</td>
<td>95.0</td>
<td>95.0</td>
</tr>
<tr>
<td></td>
<td>$\tau = 0.6$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(50, 5, 10)</td>
<td>bias</td>
<td>-0.0490</td>
<td>0.0128</td>
<td>0.0673</td>
</tr>
<tr>
<td></td>
<td>sd</td>
<td>0.1758</td>
<td>0.0836</td>
<td>0.3322</td>
</tr>
<tr>
<td></td>
<td>cp(%)</td>
<td>96.5</td>
<td>97.0</td>
<td>95.5</td>
</tr>
<tr>
<td>(100, 5, 20)</td>
<td>bias</td>
<td>-0.0327</td>
<td>0.0011</td>
<td>0.0522</td>
</tr>
<tr>
<td></td>
<td>sd</td>
<td>0.1935</td>
<td>0.0581</td>
<td>0.1783</td>
</tr>
<tr>
<td></td>
<td>cp(%)</td>
<td>95.5</td>
<td>96.5</td>
<td>91.5</td>
</tr>
<tr>
<td>(100, 10, 10)</td>
<td>bias</td>
<td>-0.0226</td>
<td>-0.0097</td>
<td>0.0638</td>
</tr>
<tr>
<td></td>
<td>sd</td>
<td>0.1602</td>
<td>0.0607</td>
<td>0.1892</td>
</tr>
<tr>
<td></td>
<td>cp(%)</td>
<td>95.5</td>
<td>96.0</td>
<td>94.5</td>
</tr>
<tr>
<td>(200, 10, 20)</td>
<td>bias</td>
<td>-0.0254</td>
<td>0.0089</td>
<td>0.0310</td>
</tr>
<tr>
<td></td>
<td>sd</td>
<td>0.1198</td>
<td>0.0374</td>
<td>0.1088</td>
</tr>
<tr>
<td></td>
<td>cp(%)</td>
<td>98.0</td>
<td>96.0</td>
<td>92.5</td>
</tr>
</tbody>
</table>
Table 8: Simulation results obtained by both the proposed and Emura et al. (2015)’s methods using a misspecified gamma frailty model in scenario 1 based on 200 replications when the true random effects distribution is a mixture of normals $0.5N(0,0.2) + 0.5N(0,0.4)$. The true parameters $\beta_1 = 1$ and $\beta_2 = 1$, $\lambda_{01}(t) = 1$, $\lambda_{02}(t) = 0.4$ and Kendall’s $\tau = 0.3$ and 0.6.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\eta$</td>
<td>$\tau$</td>
</tr>
<tr>
<td>$(50, 10)$</td>
<td>bias</td>
<td>0.0807</td>
</tr>
<tr>
<td></td>
<td>sd</td>
<td>0.3997</td>
</tr>
<tr>
<td></td>
<td>rmse</td>
<td>0.4078</td>
</tr>
<tr>
<td></td>
<td>cp(%)</td>
<td>94.5</td>
</tr>
<tr>
<td>$(50, 10)$</td>
<td>bias</td>
<td>-0.0197</td>
</tr>
<tr>
<td></td>
<td>sd</td>
<td>0.2419</td>
</tr>
<tr>
<td></td>
<td>rmse</td>
<td>0.2427</td>
</tr>
<tr>
<td></td>
<td>cp(%)</td>
<td>94.5</td>
</tr>
<tr>
<td>$(100, 5, 10)$</td>
<td>bias</td>
<td>-0.0052</td>
</tr>
<tr>
<td></td>
<td>sd</td>
<td>0.2263</td>
</tr>
<tr>
<td></td>
<td>rmse</td>
<td>0.2264</td>
</tr>
<tr>
<td></td>
<td>cp(%)</td>
<td>96.0</td>
</tr>
<tr>
<td>$(100, 5, 10)$</td>
<td>bias</td>
<td>-0.0338</td>
</tr>
<tr>
<td></td>
<td>sd</td>
<td>0.1611</td>
</tr>
<tr>
<td></td>
<td>rmse</td>
<td>0.1646</td>
</tr>
<tr>
<td></td>
<td>cp(%)</td>
<td>97.0</td>
</tr>
<tr>
<td>$(200, 10, 20)$</td>
<td>bias</td>
<td>-0.0133</td>
</tr>
<tr>
<td></td>
<td>sd</td>
<td>0.3059</td>
</tr>
<tr>
<td></td>
<td>rmse</td>
<td>0.3062</td>
</tr>
<tr>
<td></td>
<td>cp(%)</td>
<td>95.5</td>
</tr>
<tr>
<td>$(200, 10, 20)$</td>
<td>bias</td>
<td>-0.0413</td>
</tr>
<tr>
<td></td>
<td>sd</td>
<td>0.2271</td>
</tr>
<tr>
<td></td>
<td>rmse</td>
<td>0.2309</td>
</tr>
<tr>
<td></td>
<td>cp(%)</td>
<td>96.5</td>
</tr>
<tr>
<td>$(200, 10, 20)$</td>
<td>bias</td>
<td>-0.0375</td>
</tr>
<tr>
<td></td>
<td>sd</td>
<td>0.2050</td>
</tr>
<tr>
<td></td>
<td>rmse</td>
<td>0.2084</td>
</tr>
<tr>
<td></td>
<td>cp(%)</td>
<td>96.0</td>
</tr>
<tr>
<td>$(200, 10, 20)$</td>
<td>bias</td>
<td>-0.0642</td>
</tr>
<tr>
<td></td>
<td>sd</td>
<td>0.1365</td>
</tr>
<tr>
<td></td>
<td>rmse</td>
<td>0.1509</td>
</tr>
<tr>
<td></td>
<td>cp(%)</td>
<td>96.0</td>
</tr>
</tbody>
</table>
Table 9: Analysis of the breast cancer data

<table>
<thead>
<tr>
<th>parameter</th>
<th>proposed estimates ($\hat{se}$)</th>
<th>naive marginal estimate (sd)</th>
<th>cox estimate (sd)</th>
</tr>
</thead>
<tbody>
<tr>
<td>regression for DMFS</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>er</td>
<td>-0.2817(0.1388)*</td>
<td>-0.3299(0.1324)*</td>
<td>-0.0821(0.1275)</td>
</tr>
<tr>
<td>size</td>
<td>0.2081(0.0501)*</td>
<td>0.1980(0.0511)*</td>
<td>0.0836(0.0458)</td>
</tr>
<tr>
<td>node</td>
<td>0.1324(0.1055)</td>
<td>0.2831(0.1441)</td>
<td>0.5521(0.1237)*</td>
</tr>
<tr>
<td>age</td>
<td>-0.0111(0.0057)</td>
<td>-0.0150(0.0059)*</td>
<td>-0.0314(0.0053)*</td>
</tr>
<tr>
<td>grade 1</td>
<td>-0.4499(0.1866)*</td>
<td>-0.6317(0.2257)*</td>
<td>-0.4992(0.2153)*</td>
</tr>
<tr>
<td>grade 2</td>
<td>-0.1131(0.1190)</td>
<td>-0.0929(0.1319)</td>
<td>-0.1228(0.1276)</td>
</tr>
<tr>
<td>$\eta = 0.5985$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>regression for OS</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>er</td>
<td>-0.6003(0.1594)*</td>
<td>-0.6072(0.1263)*</td>
<td>-0.5533(0.1260)*</td>
</tr>
<tr>
<td>size</td>
<td>0.2510(0.0560)</td>
<td>0.2348(0.0477)*</td>
<td>0.2134(0.0454)*</td>
</tr>
<tr>
<td>node</td>
<td>0.2090(0.1276)</td>
<td>0.2026(0.1351)</td>
<td>0.7601(0.1261)*</td>
</tr>
<tr>
<td>age</td>
<td>0.0082(0.0047)</td>
<td>0.0104(0.0052)*</td>
<td>0.0098(0.0046)*</td>
</tr>
<tr>
<td>grade 1</td>
<td>-0.6054(0.1968)*</td>
<td>-0.6219(0.2149)*</td>
<td>-0.5125(0.2147)*</td>
</tr>
<tr>
<td>grade 2</td>
<td>-0.2419(0.1548)</td>
<td>-0.2670(0.1302)*</td>
<td>-0.2615(0.1314)</td>
</tr>
<tr>
<td>$\eta = 0.3074$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\tau$</td>
<td>0.5974(0.0867)*</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>$\eta$</td>
<td>0.3787(0.0960)*</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
Figure 1: Simulation results for estimated cumulative baseline hazard function $\Lambda_0(t)$ by the proposed method based on 20 replications (colored red) with true cumulative baseline hazard function $\Lambda_0(t) = t$ (colored black) under the model with gamma frailty in Scenario 1 and $\tau = 0.3$. 

(a) $N = 50$, $m = 5$, $n_i = 10$

(b) $N = 100$, $m = 5$, $n_i = 20$

(c) $N = 100$, $m = 10$, $n_i = 10$

(d) $N = 200$, $m = 10$, $n_i = 20$
Figure 2: Simulation results for estimated cumulative baseline hazard function $\Lambda_{02}(\cdot)$ by the proposed method based on 20 replications (colored red) with true cumulative baseline hazard function $\Lambda_{02}(t) = 0.4t$ (colored black) under the model with gamma frailty in Scenario 1 and $\tau = 0.3$. 

(a) $N = 50$, $m = 5$, $n_i = 10$
(b) $N = 100$, $m = 5$, $n_i = 20$
(c) $N = 100$, $m = 10$, $n_i = 10$
(d) $N = 200$, $m = 10$, $n_i = 20$
Figure 3: Simulation results for estimated cumulative baseline hazard function $\Lambda_0(t)$ by Emura et al. (2015)’s method based on 20 replications (colored red) with true cumulative baseline hazard function $\Lambda_0(t) = t$ (colored black) under the model with gamma frailty in Scenario 1 and $\tau = 0.3$. 

(a) $N = 50$, $m = 5$, $n_i = 10$
(b) $N = 100$, $m = 5$, $n_i = 20$
(c) $N = 100$, $m = 10$, $n_i = 10$
(d) $N = 200$, $m = 10$, $n_i = 20$
Figure 4: Simulation results for estimated cumulative baseline hazard function $\Lambda_{02}(\cdot)$ by Emura et al. (2015)’s method based on 20 replications (colored red) with true cumulative baseline hazard function $\Lambda_{02}(t) = 0.4t$ (colored black) under the model with gamma frailty in Scenario 1 and $\tau = 0.3$. 

(a) $N = 50, m = 5, n_i = 10$

(b) $N = 100, m = 5, n_i = 20$

(c) $N = 100, m = 10, n_i = 10$

(d) $N = 200, m = 10, n_i = 20$
Appendix

Proof of Theorem 1

We establish the result in Theorem 1 through the following three steps. We first prove \( \hat{\Lambda}_m(\cdot) = (\hat{\Lambda}_{m1}(\cdot), \hat{\Lambda}_{m2}(\cdot)) \) is bounded on \([0, \varsigma] \). Then we show the existence of a convergent subsequence of \( \hat{\omega}_m = (\hat{\alpha}_m, \hat{\beta}_m, \hat{\Lambda}_m, \hat{\Sigma}_m) \) by the compactness of the parameter space and Helly’s selection theorem. Finally we prove that the limit of this subsequence must be \( \omega_0 \). Let

\[
\hat{\Lambda}_{m1}(t) = \sum_{i,j} \frac{\delta_{ij}^{(1)} (1 - Y_{ij}(t))}{\gamma_{gh}(T_{ij}) e^{\beta_{01} X_{ij}} E_{\omega}(w_1(T_{gh}) e^{\beta_{02} Z_{gh}})},
\]

\[
\hat{\Lambda}_{m2}(t) = \sum_{i,j} \frac{(\delta_{ij}^{(1)} \tilde{Y}_{ij}^{(2)} + \delta_{ij}^{(2)})(1 - Y_{ij}(t))}{\gamma_{gh}(T_{ij}) e^{\beta_{02} X_{ij}} E_{\omega}(w_2(T_{gh}) e^{\beta_{02} Z_{gh}})},
\]

\[
a_i(t) = n_i^{-1} \sum_{j=1}^{n_i} \int_0^t \left\{ dN_{ij}(1) - Y_{ij}(u) e^{\delta_{ij} X_{ij}} e^{\beta_{01} X_{ij}} E_{\omega}(w_1(T_{ij}) e^{\beta_{02} Z_{ij}}) \right\} d\Lambda_{01}(u),
\]

\[
b_i(t) = n_i^{-1} \sum_{j=1}^{n_i} \int_0^t \left\{ dN_{ij}(2) - \tilde{Y}_{ij}(u) e^{\delta_{ij} X_{ij}} e^{\beta_{02} X_{ij}} E_{\omega}(w_2(T_{ij}) e^{\beta_{02} Z_{ij}}) \right\} d\Lambda_{02}(u),
\]

\[
f_{m1}(u) = m^{-1} \sum_{i=1}^m n_i^{-1} \sum_{j=1}^{n_i} Y_{ij}(u) e^{\delta_{ij} X_{ij}} E_{\omega}(w_1(T_{ij}) e^{\beta_{02} Z_{ij}}),
\]

\[
f_{m2}(u) = m^{-1} \sum_{i=1}^m n_i^{-1} \sum_{j=1}^{n_i} \tilde{Y}_{ij}(u) e^{\delta_{ij} X_{ij}} E_{\omega}(w_2(T_{ij}) e^{\beta_{02} Z_{ij}}).
\]

We show \( \sup_{t \in [0, \varsigma]} |\hat{\Lambda}_{m1}(t) - \Lambda_{01}(t)| \to 0 \) and \( \sup_{t \in [0, \varsigma]} |\hat{\Lambda}_{m2}(t) - \Lambda_{02}(t)| \to 0 \) almost surely. Note that \( \{a_i(t) : i = 1, 2, \ldots \} \) and \( \{b_i(t) : i = 1, 2, \ldots \} \) are two mean zero independent sequences for fixed \( t \). By the strong law of large numbers (SLLN), \( m^{-1} \sum_i a_i(t) \to 0 \) almost surely (a.s.) and \( m^{-1} \sum_i b_i(t) \to 0 \) a.s., respectively. Given that \( S_1(T_{ij}), S_2(T_{ij}) \) and \( c \{S_1(T_{ij}), S_2(T_{ij}); \alpha \} \) are bounded above 0, by the bounded assumption on \( X_{ij} \) and \( Z_{ij} \) we have \( E_{\omega}(w_1(T_{ij}) e^{\beta_{02} Z_{ij}}), E_{\omega}(w_2(T_{ij}) e^{\beta_{02} Z_{ij}}), e^{\delta_{ij} X_{ij}} \) and \( e^{\delta_{ij} X_{ij}} \) are all bounded. Applying the SLLN again, we have \( f_{m1}(u) \to E_\omega \left[ Y_{ij}(u) e^{\delta_{ij} X_{ij}} E_{\omega}(w_1(T_{ij}) e^{\beta_{02} Z_{ij}}) \right] \) a.s. and
\[
\int_{u=0}^{t} f_{m1}(u)d(\bar{\Lambda}_{m1} - \Lambda_{01})(u) = m^{-1} \sum_{i,j} n_i^{-1} \int_{u=0}^{t} \left\{ dN_{ij}^{(1)}(u) - Y_{ij}(u)e^{\beta_{ij} X_{ij}} E_w\{w_1(T_{ij})e^{b_{ij} Z_{ij}}\} \right\} d\Lambda_{01}(u) \\
- m^{-1} \sum_{i,j} n_i^{-1} \int_{u=0}^{t} \left\{ dN_{ij}^{(1)}(u) - Y_{ij}(u)e^{\beta_{ij} X_{ij}} E_w\{w_1(T_{ij})e^{b_{ij} Z_{ij}}\} \right\} d\bar{\Lambda}_{m1}(u) \\
= -n^{-1} \sum_{i=1}^{n} a_i(t) \to 0 \text{ a.s.}, \quad \text{and}
\]

\[
\int_{u=0}^{t} f_{m2}(u)d(\bar{\Lambda}_{m2} - \Lambda_{02})(u) = m^{-1} \sum_{i,j} n_i^{-1} \int_{u=0}^{t} \left\{ dN_{ij}^{(1)}(u)d\bar{\Lambda}_{m2}(u) - dN_{ij}^{(2)}(u) - Y_{ij}(u)e^{\beta_{ij} X_{ij}} E_w\{w_2(T_{ij})e^{b_{ij} Z_{ij}}\} d\Lambda_{02}(u) \right\} \\
- m^{-1} \sum_{i,j} n_i^{-1} \int_{u=0}^{t} \left\{ dN_{ij}^{(1)}(u)d\bar{\Lambda}_{m2}(u) - dN_{ij}^{(2)}(u) - Y_{ij}(u)e^{\beta_{ij} X_{ij}} E_w\{w_2(T_{ij})e^{b_{ij} Z_{ij}}\} d\bar{\Lambda}_{m2}(u) \right\} \\
= -m^{-1} \sum_{i=1}^{n} b_i(t) \to 0 \text{ a.s.}
\]

Since \( \lim_{m \to \infty} f_{m1}(u) = E[Y_{ij}(u)e^{\beta_{ij} X_{ij}} E_w\{w_1(T_{ij})e^{b_{ij} Z_{ij}}\}] \), which is bounded away from zero, there exists some \( c_1(u) > 0 \) such that eventually \( f_{m1}(u) > c_1(u) \) almost surely. Let \( c_1 = \sup_{u \in [0, 1]} c_1(u) \). For sufficiently
large $n$, we can write

$$0 \leq c_1 \int_{u=0}^{t} d(\hat{\Lambda}_{m1} - \Lambda_{01})(u) \leq \int_{u=0}^{t} f_{m1}(u)d(\hat{\Lambda}_{m1} - \Lambda_{01})(u) \to 0 \text{ a.s.}$$

Using the squeeze theorem, we have $\int_{u=0}^{t} d(\hat{\Lambda}_{m1} - \Lambda_{01})(u) \to 0$ a.s. It follows that $\hat{\Lambda}_{m1}(t) \to \Lambda_{01}(t)$ a.s. and $\hat{\Lambda}_{m2}(t) \to \Lambda_{02}(t)$ a.s. for all $t \in [0, \varsigma)$. Since $\hat{\Lambda}_{m}(t)$ are non-decreasing functions converging pointwise to the continuous functions $\Lambda_{0i}(t)$ for $i = 1, 2$, hence $\hat{\Lambda}_{m}(t)$ are locally (on $[0, \varsigma]$ in particular) uniformly convergent.

Since $\hat{\alpha}_m$, $\hat{\beta}_m$, $\bar{\Sigma}_m$, $X_{gh}$, $Z_{gh}$ are in compact sets and $w_1(T_{gh})$ and $w_2(T_{gh})$ are bounded, there exists some finite and possibly negative constants $c_2$ and $c_3$ such that

$$\hat{\beta}_m' X_{gh} + \log E_{\omega_m}\{w_1(T_{gh})e^{b_2' Z_{gh}}\} \geq \hat{\beta}_{01}' X_{gh} + \log E_{\omega_0}\{w_1(T_{gh})e^{b_2' Z_{gh}}\} + c_2,$$

$$\hat{\beta}_m' X_{gh} + \log E_{\omega_m}\{w_2(T_{gh})e^{b_2' Z_{gh}}\} \geq \hat{\beta}_{02}' X_{gh} + \log E_{\omega_0}\{w_2(T_{gh})e^{b_2' Z_{gh}}\} + c_3.$$  

Therefore

$$\hat{\Lambda}_{m1}(s) = \sum_{i,j} \frac{\delta_{ij}^{(1)}(1 - Y_{ij}(s))}{Y_{gh}(T_{ij}) \exp[\hat{\beta}_{m}' X_{gh} + \log E_{\omega_m}\{w_1(T_{gh})e^{b_2' Z_{gh}}\}]} \leq \sum_{i,j} \frac{\delta_{ij}^{(1)}(1 - Y_{ij}(s))}{Y_{gh}(T_{ij}) \exp[\hat{\beta}_{01}' X_{gh} + \log E_{\omega_0}\{w_1(T_{gh})e^{b_2' Z_{gh}}\} + c_2]} = e^{-c_2} \hat{\Lambda}_{m1}(s) \to e^{-c_2} \Lambda_{01}(s),$$

$$\hat{\Lambda}_{m2}(s) = \sum_{i,j} \frac{(\delta_{ij}^{(1)} \delta_{ij}^{(2)} + \delta_{ij}^{(2)})(1 - Y_{ij}(s))}{Y_{gh}(T_{ij}) \exp[\hat{\beta}_{m}' X_{gh} + \log E_{\omega_m}\{w_2(T_{gh})e^{b_2' Z_{gh}}\}]} \leq \sum_{i,j} \frac{(\delta_{ij}^{(1)} \delta_{ij}^{(2)} + \delta_{ij}^{(2)})(1 - Y_{ij}(s))}{Y_{gh}(T_{ij}) \exp[\hat{\beta}_{02}' X_{gh} + \log E_{\omega_0}\{w_2(T_{gh})e^{b_2' Z_{gh}}\} + c_3]} = e^{-c_3} \hat{\Lambda}_{m2}(s) \to e^{-c_3} \Lambda_{02}(s).$$

So far we have established that $\hat{\Lambda}_{m1}$ and $\hat{\Lambda}_{m2}$ have upper bounds almost surely, and $\hat{\alpha}_m$, $\hat{\beta}_m$ and $\bar{\Sigma}_m$ are in compact sets. Now we can apply Helly’s selection theorem to infer the existence of a convergent subsequence of $\hat{\omega}_m = (\hat{\alpha}_m, \hat{\beta}_m, \hat{\Lambda}_m, \bar{\Sigma}_m)$ having limit $\omega^*$, where $\hat{\beta}_m = (\hat{\beta}_m, \hat{\beta}_m)$ and $\hat{\Lambda}_m = (\hat{\Lambda}_m, \hat{\Lambda}_m)$.

Next we show $\omega^* = \omega_0$. Since

$$\hat{\Lambda}_{m1}(t) = \int_{0}^{t} \sum_{g,h} Y_{gh}(u) \exp[\hat{\beta}_{01}' X_{gh} + \log E_{\omega_0}\{w_1(T_{gh})e^{b_2' Z_{gh}}\}] \frac{d\Lambda_{m1}(u)}{\Lambda_{01}(t)},$$

$$\hat{\Lambda}_{m2}(t) = \int_{0}^{t} \sum_{g,h} \tilde{Y}_{gh}(u) \exp[\hat{\beta}_{02}' X_{gh} + \log E_{\omega_0}\{w_2(T_{gh})e^{b_2' Z_{gh}}\}] \frac{d\Lambda_{m2}(u)}{\Lambda_{02}(t)},$$

we can see that $\Lambda^* = (\Lambda_{01}, \Lambda_{02})$ is absolutely continuous with respect to $\Lambda_0 = (\Lambda_{01}, \Lambda_{02})$. In addition,
\( \Lambda^*(t) = (\Lambda^*_{01}(t), \Lambda^*_{02}(t)) \) is differentiable with respect to \( t \) and \( d\Lambda_m(t)/d\tilde{\Lambda}_m(t) \) converges to \( d\Lambda^*(t)/d\Lambda_0(t) \).

Note that the finite sample likelihood as expressed in Equation (6) of the paper has no finite maximum since \( \lambda_{01} \) and \( \lambda_{02} \) are free to go to infinity at any \( T_{ij} \) and \( \tilde{T}_{ij} \), respectively. We restrict \( \Lambda_{01} \) right continuous with jumps at \( T_{ij} \) and \( \Lambda_{02} \) to be right continuous with jumps at \( \tilde{T}_{ij} \). We then have the log-likelihood based on the observed data

\[
l_m(\alpha, \beta, \Lambda, \Sigma) = \sum_{i=1}^{m} \log \left\{ \int_{b_i} R_{1i}(\alpha, \beta, \Lambda, b_i) \psi(b_i, \Sigma) db_i \right\} + \sum_{i=1}^{m} \sum_{j=1}^{n_i} \delta_{ij}^{(1)} \log(\Lambda_1(T_{ij})) + (\delta_{ij}^{(1)} \delta_{ij}^{(2)} + \delta_{ij}^{(2)}) \log(\Lambda_2(\tilde{T}_{ij})).
\]

(17)

where

\[
R_{1i}(\alpha, \beta, \Lambda, b) = \exp \left[ \sum_{j=1}^{n_i} \left( \delta_{ij}^{(1)} \delta_{ij}^{(2)} \log(D_{12}(T_{ij}, \tilde{T}_{ij})) - \delta_{ij}^{(1)} \Lambda_1(T_{ij}) e^{\delta_{ij}^{(1)} \Lambda_1(T_{ij})} + \delta_{ij}^{(2)} (\delta_{ij}^{(1)} \beta_{ij} X_{ij} + \beta_{ij} Z_{ij}) \right. \right.
\]

\[
+ \left. \left. \delta_{ij}^{(2)} (1 - \delta_{ij}^{(2)}) \log(D_2(T_{ij}, \tilde{T}_{ij})) + \delta_{ij}^{(2)} \log(D_2(T_{ij}, \tilde{T}_{ij})) - \delta_{ij}^{(1)} \delta_{ij}^{(2)} \Lambda_2(\tilde{T}_{ij}) e^{\delta_{ij}^{(1)} \Lambda_2(\tilde{T}_{ij})} + \delta_{ij}^{(2)} \right) \log(\hat{\Lambda}(T_{ij}, S_2(\tilde{T}_{ij}))) - \lambda_{02}(\tilde{T}_{ij})/\lambda_{02}(\tilde{T}_{ij}) \right] \right].
\]

Note that

\[
0 \leq m^{-1}\{l_m(\alpha_0, \beta_0, \Lambda_m, \Sigma_m) - l_m(\alpha, \beta, \Lambda, \Sigma)\}
\]

\[
= m^{-1} \sum_{i=1}^{m} \log \left\{ \int_{b_i} R_{1i}(\alpha_0, \beta_0, \Lambda_m, b_i) \psi(b_i, \Sigma_m) db_i \right\} - m^{-1} \sum_{i=1}^{m} \log \left\{ \int_{b_i} R_{1i}(\alpha_0, \beta_0, \Lambda_m, b_i) \psi(b_i, \Sigma) db_i \right\}
\]

\[
+ m^{-1} \sum_{i=1}^{m} \sum_{j=1}^{n_i} \delta_{ij}^{(1)} \log(\Lambda_m(T_{ij})/\Lambda_m(T_{ij})) + m^{-1} \sum_{i=1}^{m} \sum_{j=1}^{n_i} (\delta_{ij}^{(1)} \delta_{ij}^{(2)} + \delta_{ij}^{(2)} \log(\Lambda_2(T_{ij})/\Lambda_2(T_{ij}))
\]

\[
- E \left[ \log \left\{ \left( \int_{b_i} R_{1i}(\alpha_0, \beta_0, \Lambda_0, b_i) \psi(b_i, \Sigma) db_i \prod_{j=1}^{n_i} \left[ \lambda_{01}(T_{ij}) \lambda_{02}(\tilde{T}_{ij}) \delta_{ij}^{(1)} \delta_{ij}^{(2)} + \delta_{ij}^{(2)} \right] \right) \right\} \right].
\]

Because the Kullback-Leibler information between the density indexed by \( (\alpha^*, \beta^*, \Lambda^*, \Sigma^*) \) and the true density is negative, we have that \( l(\alpha^*, \beta^*, \Lambda^*, \Sigma^*) = l(\alpha_0, \beta_0, \Lambda_0, \Sigma_0) \) almost surely. That is almost surely

\[
\int_{b_i} R_{1i}(\alpha^*, \beta^*, \Lambda^*, b_i) \psi(b_i, \Sigma^*) db_i \prod_{j=1}^{n_i} \left[ \lambda_{01}(T_{ij}) \lambda_{02}(\tilde{T}_{ij}) \delta_{ij}^{(1)} \delta_{ij}^{(2)} + \delta_{ij}^{(2)} \right]
\]

\[
= \int_{b_i} R_{1i}(\alpha_0, \beta_0, \Lambda_0, b_i) \psi(b_i, \Sigma_0) db_i \prod_{j=1}^{n_i} \left[ \lambda_{01}(T_{ij}) \lambda_{02}(\tilde{T}_{ij}) \delta_{ij}^{(1)} \delta_{ij}^{(2)} + \delta_{ij}^{(2)} \right].
\]

(18)

Next we adapt the techniques used in the identifiability step of [Zeng et al. (2005)] and [Gamst et al. (2009)] to show that \( (\alpha^*, \beta^*, \Lambda^*, \Sigma^*) = (\alpha_0, \beta_0, \Lambda_0, \Sigma_0) \). For notational convenience, we recall the notation in Equa-
tion (9) of the paper and denote that

\[ D_{1\alpha}(t^{(1)}, t^{(2)}) = \frac{\partial^2 c(S_1(t^{(1)}), S_2(t^{(2)}))}{\partial S_1(t^{(1)}) \partial S_2(t^{(2)})}, \]

\[ D_{12\alpha}(t^{(1)}, t^{(2)}) = \frac{\partial^2 c(S_1(t^{(1)}), S_2(t^{(2)}))}{\partial S_1(t^{(1)}) \partial S_2(t^{(2)})}. \]

(19)

Let \( \delta_{ij}^{(1)} = 1 \) and \( T_{ij} = 0 \), for all \( j = 1, \cdots, n_i \). With some fixed \( h, 1 \leq h \leq n_i \), set \( \delta_{ij}^{(2)} = 1 \) and \( \tilde{T}_{ij} = 0 \) for \( j \leq h \). For \( j > h \), we replace \( \tilde{T}_{ij} \) with \( \varsigma \) when \( \delta_{ij}^{(2)} = 0 \), while we integrate \( \tilde{T}_{ij} \) from 0 to \( \varsigma \) in (19) when \( \delta_{ij}^{(2)} = 1 \). Thus, we obtain that

\[ l_t(\theta, \lambda|\mathcal{O}_t, b_t) = \prod_{j=1}^{h} \left\{ D_{12}(0, 0) \lambda_1(0) e^{\beta_i \cdot X_{ij} + b_i' \cdot Z_{ij}} \lambda_2(0) e^{\beta_j' \cdot X_{ij} + b_j' \cdot Z_{ij}} \right\}^{1-\delta_{ij}^{(2)}} \cdot \prod_{j=h+1}^{n_i} \left\{ D_1(0, 0) \lambda_1(0) e^{\beta_i \cdot X_{ij} + b_i' \cdot Z_{ij}} \lambda_2(0) e^{\beta_j' \cdot X_{ij} + b_j' \cdot Z_{ij}} \right\}^{\delta_{ij}^{(2)}} \]

\[ = \prod_{j=1}^{h} \left\{ D_{12}(0, 0) \lambda_1(0) e^{\beta_i \cdot X_{ij} + b_i' \cdot Z_{ij}} \lambda_2(0) e^{\beta_j' \cdot X_{ij} + b_j' \cdot Z_{ij}} \right\} \cdot \prod_{j=h+1}^{n_i} \left\{ (1 - \tilde{T}_{ij}^{(2)}) D_1(0, 0) + \delta_{ij}^{(2)} ( - D_1(0, 0) + D_1(0, 0) \right\} \cdot \left\{ \lambda_1(0) e^{\beta_i \cdot X_{ij} + b_i' \cdot Z_{ij}} \right\} \]

\[ = \prod_{j=1}^{h} \left\{ D_{12}(0, 0) \right\} \cdot \prod_{j=1}^{h} \left\{ D_1(0, 0) \right\} \cdot \prod_{j=1}^{h} \left\{ \lambda_2(0) e^{\beta_j' \cdot X_{ij} + b_j' \cdot Z_{ij}} \right\} \cdot \prod_{j=h+1}^{n_i} \left\{ \lambda_1(0) e^{\beta_i \cdot X_{ij} + b_i' \cdot Z_{ij}} \right\}. \]

Equation (18) can be written as

\[ \left\{ D_{12}(0, 0) \right\}^h \left\{ D_1(0, 0) \right\}^{n_i-h} \int_{b_1}^{h} \prod_{j=1}^{h} \left\{ \lambda_2(0) e^{\beta_j' \cdot X_{ij} + b_j' \cdot Z_{ij}} \right\} \prod_{j=1}^{h} \left\{ \lambda_1(0) e^{\beta_i \cdot X_{ij} + b_i' \cdot Z_{ij}} \right\} \psi(b, \Sigma^*) \text{d}b \]

\[ = \left\{ D_{12}(0, 0) \right\}^h \left\{ D_1(0, 0) \right\}^{n_i-h} \int_{b_1}^{h} \prod_{j=1}^{h} \left\{ \lambda_2(0) e^{\beta_j' \cdot X_{ij} + b_j' \cdot Z_{ij}} \right\} \prod_{j=1}^{h} \left\{ \lambda_1(0) e^{\beta_i \cdot X_{ij} + b_i' \cdot Z_{ij}} \right\} \psi(b, \Sigma_0) \text{d}b. \]

(20)

Let \( h = 0 \), it follows from (20) that

\[ \left\{ D_{12}(0, 0) \right\}^{n_i} \int_{b_1}^{h} \prod_{j=1}^{n_i} \left\{ \lambda_2(0) e^{\beta_j' \cdot X_{ij} + b_j' \cdot Z_{ij}} \right\} \psi(b_1, \Sigma^*) \text{d}b_1 \]

\[ = \left\{ D_1(0, 0) \right\}^{n_i} \int_{b_1}^{h} \prod_{j=1}^{n_i} \left\{ \lambda_1(0) e^{\beta_i \cdot X_{ij} + b_i' \cdot Z_{ij}} \right\} \psi(b_1, \Sigma_0) \text{d}b_1. \]

(21)
That is equivalent to
\[ \{D_{11}(0,0)\lambda_{01}(0)\}^{n_i}e^{\sum_{j=1}^{n_i} \beta_{i}^{*} X_{ij}} E_{b_i}(e^{\sum_{j=1}^{n_i} b_{ij}^{*} Z_{ij}} | \Sigma = \Sigma^*) = \{D_{1}(0,0)\lambda_{01}(0)\}^{n_i}e^{\sum_{j=1}^{n_i} \beta_{01} X_{ij}} E_{b_i}(e^{\sum_{j=1}^{n_i} b_{ij}^{*} Z_{ij}} | \Sigma = \Sigma_0). \] 
\[ (22) \]

Since $S_1^*(0) = S_1(0) = S_2^*(0) = S_2(0) = 1$ and $D_{11}(0,0)$ is function only related to $\alpha$, let $f(\alpha) = D_{11}(0,0)$, \[ \text{can be written as} \]
\[ n_i \left[ (\log(f(\alpha^*)) - \log(f(\alpha_0))) + (\log \lambda_{01}(0) - \log \lambda_{01}(0)) \right] + \sum_{j=1}^{n_i} X_{ij}(\beta_{i}^{*} - \beta_{01}) \]
\[ + \left\{ E_{b_i}(e^{\sum_{j=1}^{n_i} b_{ij}^{*} Z_{ij}} | \Sigma = \Sigma^*) - E_{b_i}(e^{\sum_{j=1}^{n_i} b_{ij}^{*} Z_{ij}} | \Sigma = \Sigma_0) \right\} = 0, \]
\[ (23) \]

Specifically, when $b_i$ is assumed to follow a normal distribution with mean 0 and variance $\Sigma_0$, \[ (23) \] can be written as
\[ n_i \left[ (\log(f(\alpha^*)) - \log(f(\alpha_0))) + (\log \lambda_{01}(0) - \log \lambda_{01}(0)) \right] + \sum_{j=1}^{n_i} X_{ij}(\beta_{i}^{*} - \beta_{01}) \]
\[ + \frac{1}{2} \left( \sum_{j=1}^{n_i} Z_{ij} \right) (\Sigma^* - \Sigma_0) (\sum_{j=1}^{n_i} Z_{ij}) = 0, \]
\[ (24) \]
for all $i = 1, \ldots, m$.

By condition 8 in the paper, we have that \( (\log(f(\alpha^*)) - \log(f(\alpha_0))) + (\log \lambda_{02}(0) - \log \lambda_{02}(0)) = 0, \beta_{i}^{*} = \beta_{01} \) and $\Sigma^* = \Sigma_0$. Since the above equations applied for different copulas with different forms of function $f(\cdot)$, it follows that $\alpha^* = \alpha_0$ and $\lambda_{01}(0) = \lambda_{01}(0)$.

When it is assumed that $Z_{ij} = 1$ and $u_i = \exp(b_i)$ follows a gamma distribution with mean 1 and variance $\eta_0$, \[ (23) \] can be rewritten as
\[ \{f(\alpha^*)\lambda_{01}(0)\}^{n_i}e^{\sum_{j=1}^{n_i} \beta_{i}^{*} X_{ij}} g(\eta^*, n_i) = \{f(\alpha_0)\lambda_{01}(0)\}^{n_i}e^{\sum_{j=1}^{n_i} \beta_{01} X_{ij}} g(\eta_0, n_i), \]
\[ (25) \]
where $g(\eta, n) = (\eta + 1)(2\eta + 1) \cdots ((n-1)\eta + 1)$. By condition 7 in the paper and the above equations applied for different sample sizes, different types of $\lambda(\cdot)$ and copulas with different forms of function $f(\cdot)$, we have that $\alpha^* = \alpha_0$, $\lambda_{01}(0) = \lambda_{01}(0)$, $\beta_{i}^{*} = \beta_{01}$ and $\eta^* = \eta_0$.

It therefore follows from \[ (20) \] that
\[ (D_{12*}(0,0))^h \prod_{j=1}^{b} \left\{ \lambda_{01}^{*}(0)e^{\sum_{j=1}^{n_i} \beta_{01} X_{ij} + b_{ij}^{*} Z_{ij}} \right\} = (D_{12}(0,0))^h \prod_{j=1}^{b} \left\{ \lambda_{02}(0)e^{\sum_{j=1}^{n_i} b_{ij}^{*} Z_{ij}} \right\}. \]
\[ (26) \]
As $S_1^*(0) = S_1(0) = S_2^*(0) = S_2(0) = 1$ and $\alpha^* = \alpha_0$, we have that $D_{12*}(0,0) = D_{12}(0,0)$. The above
equation can be written as
\[
h \left( \log \lambda_{02}^{\alpha}(0) - \log \lambda_{02}(0) \right) + (\beta_2^* - \beta_{02})' \sum_{j=1}^{n} X_{ij} = 0. \quad (27)
\]

Setting \( h = n_i \), we can obtain that
\[
n_i \left( \log \lambda_{02}^{\alpha}(0) - \log \lambda_{02}(0) \right) + (\beta_2^* - \beta_{02})' \sum_{j=1}^{n} X_{ij} = 0,
\]
(28)

for all \( i = 1, \ldots, m \). By condition 7 in the paper, we have \( \lambda_{02}^{\alpha}(0) = \lambda_{02}(0) \) and \( \beta_2^* = \beta_{02} \).

Next, we prove \( \Lambda_{02}^{\alpha}(t) = \Lambda_{02}(t) \). Let \( \delta_{ij}^{(1)} = 1 \) and \( T_{ij} = 0 \) for all \( j = 1, \ldots, n_i \). When \( j = 1 \) and \( \delta_{i1}^{(2)} = 1 \), we integrate \( \tilde{T}_{i1} \) from 0 to \( t \). For \( j = 2, \ldots, n_i \), we replace \( \tilde{T}_{ij} \) with \( \zeta \) if \( \delta_{ij}^{(2)} = 0 \) and integrate \( \tilde{T}_{ij} \) from 0 to \( \zeta \) if \( \delta_{ij}^{(2)} = 1 \). Summing the two sides of (18) over all possible values of \( \delta_{ij}^{(2)} \) for \( j > 1 \), we obtain that
\[
\int_{b_i}^{t} \int_{0}^{t} D_{12}(0, y) e^{-\Lambda_{02}(y)} e^{\beta_{02}X_{ij} + \beta_{ij}Z_{ij}} \lambda_{02}(y) e^{\beta_{02}X_{ij} + \beta_{ij}Z_{ij}} dy \cdot \psi(h, \Sigma_0) db_i = \int_{b_i}^{t} \left( \prod_{j=1}^{n_i} S_{ij}(0) \right) \cdot \psi(h, \Sigma_0) db_i,
\]
which is equivalent to
\[
\int_{b_i}^{t} \left( \prod_{j=1}^{n_i} S_{ij}(0) \right) \cdot \psi(h, \Sigma_0) db_i = \int_{b_i}^{t} \left( \prod_{j=1}^{n_i} S_{ij}(0) \right) \cdot \psi(h, \Sigma_0) db_i.
\]

Therefore,
\[
c_1 \left( S_{1}^*(0), S_{2}^*(0) \right) = c_1 \left( S_{1}(0)^*, S_{2}(0) \right). \quad (29)
\]

As \( \alpha^* = \alpha_0 \), \( S_{1}^*(0) = S_{1}(0) \) and the above equation holds for all kinds of copulas, we obtain that \( S_{2}^*(t) = S_{2}(t) \) and \( \Lambda_{02}^{\alpha}(t) = \Lambda_{02}(t) \).

Similarly, we can show that \( \Lambda_{01}^{\alpha}(t) = \Lambda_{01}(t) \). Let \( \delta_{ij}^{(2)} = \delta_{ij}^{(2)} = 0 \) and \( \tilde{T}_{ij} = \zeta \) for all \( j = 1, \ldots, n_i \).

Choose \( k \) from index set \( \{1, \ldots, n_i\} \). For \( j \leq k \), we integrate \( T_{ij} \) from 0 to \( t \) if \( \delta_{ij}^{(1)} = 1 \). On the other hand, for \( j > k \), we replace \( T_{ij} \) with \( \zeta \) if \( \delta_{ij}^{(1)} = 0 \), otherwise we integrate \( T_{ij} \) from 0 to \( \zeta \).

We choose \( k = 1 \) and then obtain that
\[
L_i(\alpha_0, \beta_0, \Lambda_0, \Sigma_0) = \int_{b_i}^{t} \int_{0}^{t} D_{1}(y, z) e^{-\Lambda_{01}(y)} e^{\beta_{01}X_{ij} + \beta_{ij}Z_{ij}} \lambda_{01}(y) e^{\beta_{01}X_{ij} + \beta_{ij}Z_{ij}} dy \left( \prod_{j=2}^{n_i} \left( \frac{c(S_1(\zeta), S_2(\zeta))}{S_2(\zeta)} \right) \right) \cdot \prod_{j=1}^{n_i} S_2(\zeta) \psi(h, \Sigma_0) db_i.
\]

42
Following from (18) that
\[ c\{S_1^*(0), S_2^*(c)\} - c\{S_1^*(t), S_2^*(c)\} = c\{S_1(0), S_2(c)\} - c\{S_1(t), S_2(c)\}. \]  
(31)

Since \( \alpha^* = \alpha \) and \( S_1^*(0) = S_1(0) \), \( c\{S_1^*(t), S_2^*(c)\} = c\{S_1(t), S_2(c)\} \) for any copula \( c \). This implies that \( S_1^*(t) = S_1(t) \). Therefore, \( \Lambda_0^\alpha(t) = \Lambda_0(t) \) and the proof of Theorem 1 is complete.

**Proof of Theorem 2**

Let \( \mathcal{H} = \{(h_1, h_2, \cdots, h_6) : h_1 \in R^1, h_2 \in R^{d_1}, h_3 \in R^{d_2}, (h_4, h_5) \) are functions on \([0, \varsigma]\), \( h_6 \in R^{d_3(d_3+1)/2} \) and \( \|h_1\|, \|h_2\|, \|h_3\|, \|h_4\|, \|h_5\|, \|h_6\| \leq 1 \), where \( \|h_4\|_v \) and \( \|h_5\|_v \) denote the total variations of \( h_4(\cdot) \) and \( h_5(\cdot) \) on \([0, \varsigma]\), respectively. Let \( S_m \) be a sequence of maps from \( \mathcal{U} \), a neighborhood of \((\alpha_0, \beta_0, \alpha_1, \alpha_2, \Sigma)\), to \( l^\infty(\mathcal{H}) \).

\[
S_m(\alpha, \beta_1, \beta_2, \Lambda_1, \Lambda_2, \Sigma)[h_1, h_2, h_3, h_4, h_5, h_6] = m^{-1} \left. \frac{d}{d\epsilon} l_m(\alpha + \epsilon h_1, \beta_1 + \epsilon h_2, \beta_2 + \epsilon \varsigma, \Lambda_1 + \epsilon h_4, \Lambda_2 + \epsilon h_5, \Sigma + \epsilon h_6) \right|_{\epsilon=0} = A_{m}[h_1] + A_{m}[h_2] + A_{m}[h_3] + A_{m}[h_4] + A_{m}[h_5] + A_{m}[h_6],
\]

where we treat \( \Sigma \) as an extended column vector consisting of the upper triangle elements of the covariance matrix. The terms \( A_{mp} \) for \( p = 1, 2, \cdots, 6 \) are linear functionals on \( R^1, R^{d_1}, R^{d_2}, BV[0, \varsigma], BV[0, \varsigma] \) and \( R^{d_3(d_3+1)/2} \), respectively. Here \( BV[0, \varsigma] \) is the space of functions with finite variation on \([0, \varsigma]\). For a single cluster, let \( l_\alpha, l_{\beta_1}, l_{\beta_2}, l_{\Lambda_1}, l_{\Lambda_2}, l_{\Sigma} \) be the score functions for \( \alpha, \beta_1, \beta_2, \Lambda_1, \Lambda_2, \Sigma \), respectively. Then \( A_{m_1}[h_1] = \mathcal{P}_m[h_1] \alpha], A_{m_2}[h_2] = \mathcal{P}_m[h_2] \beta_1], A_{m_3}[h_3] = \mathcal{P}_m[h_3] \beta_2], A_{m_4}[h_4] = \mathcal{P}_m[h_4] \Lambda_1], A_{m_5}[h_5] = \mathcal{P}_m[h_5] \Lambda_2] \) and \( A_{m_6}[h_6] = \mathcal{P}_m[h_6] \Sigma] \), where \( \mathcal{P}_m \) denotes the empirical measure based on \( m \) independent clusters. We now seek for explicit expressions for \( A_{mp} \).

Recall the log-likelihood in (17) and the notation defined in [19] in the proof of Theorem 1. For simplicity of notation, we further denote
\[
d\mu_i(b_i) = \frac{R_{i1}(\omega, b_i)\psi(b_i, \Sigma)db_i}{\int_{b_i} R_{i1}(\omega, b_i)\psi(b_i, \Sigma)db_i}.
\]

Then we have
\[
A_{m_1}[h_1] = m^{-1} \left. \frac{d}{d\epsilon} l_m(\alpha + \epsilon h_1) \right|_{\epsilon=0} = m^{-1} \sum_{i=1}^m \int_{b_i} \sum_{j=1}^n h_1 \left\{ \delta^{(1)}_{ij} \delta^{(2)}_{ij} \frac{D_{12}(T_{ij}, \tilde{T}_{ij})}{D_{12}(T_{ij}, \tilde{T}_{ij})} + \delta^{(1)}_{ij} (1 - \delta^{(2)}_{ij}) \frac{D_{1}(T_{ij}, \tilde{T}_{ij})}{D_{1}(T_{ij}, \tilde{T}_{ij})} 
+ \delta^{(2)}_{ij} D_{12}(T_{ij}, \tilde{T}_{ij}) \right\} d\mu_i(b_i),
\]

43
\[ A_{m2}[h_2] = m^{-1} \frac{d}{dc} l_m(\alpha_1 + ch_2) \bigg|_{c=0} = m^{-1} \sum_{i=1}^{m} \sum_{j=1}^{n_1} h_{ij} Y_{ij} \left\{ \delta_{ij} - \Lambda_1(T_{ij}) e^{\beta_{ij} X_{ij}} \right\} - \int_{0}^{T_{ij}} h_{ij}(s) d\Lambda_1(s) \int_{b_i} e^{\beta_{ij} X_{ij}} \left( \delta_{ij} \delta_{ij} + \delta_{ij} \right) \right\} d\mu_i(b_i), \]

\[ A_{m3}[h_3] = m^{-1} \frac{d}{dc} l_m(\alpha_2 + ch_3) \bigg|_{c=0} = m^{-1} \sum_{i=1}^{m} \sum_{j=1}^{n_1} h_{ij} Y_{ij} \left\{ \delta_{ij} - \Lambda_2(T_{ij}) e^{\beta_{ij} X_{ij}} \right\} - \int_{0}^{T_{ij}} h_{ij}(s) d\Lambda_2(s) \int_{b_i} e^{\beta_{ij} X_{ij}} \left( \delta_{ij} \delta_{ij} + \delta_{ij} \right) \right\} d\mu_i(b_i), \]

\[ A_{m4}[h_4] = m^{-1} \frac{d}{dc} l_m(\alpha_1 + ch_2) \bigg|_{c=0} = m^{-1} \sum_{i=1}^{m} \sum_{j=1}^{n_1} \left\{ \delta_{ij} h_{ij}(T_{ij}) - \int_{0}^{T_{ij}} h_{ij}(s) d\alpha_1(s) \int_{b_i} e^{\beta_{ij} X_{ij}} \left( \delta_{ij} \delta_{ij} + \delta_{ij} \right) \right\} d\mu_i(b_i), \]

\[ A_{m5}[h_5] = m^{-1} \frac{d}{dc} l_m(\alpha_2 + ch_3) \bigg|_{c=0} = m^{-1} \sum_{i=1}^{m} \sum_{j=1}^{n_1} \left\{ \delta_{ij} h_{ij}(T_{ij}) - \int_{0}^{T_{ij}} h_{ij}(s) d\alpha_2(s) \int_{b_i} e^{\beta_{ij} X_{ij}} \left( \delta_{ij} \delta_{ij} + \delta_{ij} \right) \right\} d\mu_i(b_i), \]
\[ A_{m_0}[h_0] = m^{-1} \frac{d}{d\epsilon} I_m(\Sigma + \epsilon h_0) \bigg|_{\epsilon = 0} = m^{-1} \sum_{i=1}^{m} \int_{b_i} \frac{d}{d\epsilon} \psi(b_i, \Sigma + \epsilon h_0) \bigg|_{\epsilon = 0} \, d\mu_i(b_i). \]

Specifically, when \( b_i \) is assumed to follow a normal distribution with mean 0 and variance \( \Sigma \),

\[ A_{m_0}[h_0] = m^{-1} \sum_{i=1}^{m} \int_{b_i} \left( u_i^2 \Sigma^{-1} H_0 \Sigma^{-1} b_i/2 - \text{trace}(\Sigma^{-1} H_0)/2 \right) \, d\mu_i(b_i). \]

Or when it is assumed that \( Z_{ij} = 1 \) and \( u_i = \exp(b_i) \) follows a gamma distribution with mean 1 and variance \( \eta \),

\[ A_{m_0}[h_0] = m^{-1} \sum_{i=1}^{m} \int_{b_i} \left\{ -\frac{1}{u_i} (\frac{1}{\eta} - 1) - u_i - 1 + \log(\eta) + \varphi_0(\frac{1}{\eta}) \right\} \frac{h_0}{\eta^2} \, d\mu_i(u_i), \]

where \( \varphi_0(\cdot) \) is the polygamma function. By conditions 2 and 5, and the result that \( \Lambda_1(\cdot) \) and \( \Lambda_2(\cdot) \) are bounded on \([0, c]\), which has been justified in the proof of Theorem 1, we have \( A_{mp} \) for \( p = 1, 2, \cdots, 6 \) all have bounded derivations on \([0, c]\). Therefore, \( A_{mp} \) are all bounded Lipschitz functions.

Define the limit map \( S : (\alpha, \beta_1, \beta_2, \Lambda_1, \Lambda_2, \Sigma)[h_1, \cdots, h_6] \rightarrow l^\infty(\mathcal{H}) \) by

\[ S(\alpha, \beta_1, \beta_2, \Lambda_1, \Lambda_2, \Sigma)[h_1, \cdots, h_6] = A_1[h_1] + A_2[h_2] + \cdots + A_6[h_6], \]

where the linear functionals \( A_p \) for \( p = 1, \cdots, 6 \) are defined similar to \( A_{mp} \), but replacing the empirical sum with the expectation. By definition, \( S_m(\hat{\alpha}_m, \hat{\beta}_m, \hat{\beta}_{mn2}, \hat{\Lambda}_0, \hat{\Sigma}_m) = 0 \) and \( S(\alpha_0, \beta_{01}, \beta_{02}, \Lambda_0, \Lambda_0, \Sigma_0) = 0 \). Thus \( \sqrt{m}(S_m(\alpha_0, \beta_{01}, \beta_{02}, \Lambda_0, \Lambda_0, \Sigma_0) - S(\alpha_0, \beta_{01}, \beta_{02}, \Lambda_0, \Lambda_0, \Sigma_0)) \) weakly converges to a tight Gaussian process on \( l^\infty(\mathcal{H}) \) since \( \mathcal{H} \) is a Donsker class and the functionals \( A_{mp} \) are bounded Lipschitz functionals with respect to \( \mathcal{H} \) [Pollard 1984].

In fact, it is easy to show that \( S(\alpha, \beta_1, \beta_2, \Lambda_1, \Lambda_2, \Sigma) \) is Frechet differentiable because of its smoothness. In a manner similar to Lemma 1 in the appendix of [Murphy 1995], we can show an approximation condition that

\[
\sup_{(h_1, \cdots, h_6)} |(S_m - S)(\hat{\alpha}_m, \hat{\beta}_{m1}, \hat{\beta}_{mn2}, \hat{\Lambda}_0, \hat{\Sigma}_m) - (S_m - S)(\alpha_0, \beta_{01}, \beta_{02}, \Lambda_0, \Lambda_0, \Sigma_0)| = O_p\left( m^{-1/2} \sqrt{\left\{ \| \hat{\alpha}_m - \alpha_0 \| + \| \hat{\beta}_{m1} - \beta_{01} \| + \| \hat{\beta}_{mn2} - \beta_{02} \| + \sup_{t \in [0, c]} |\hat{\Lambda}_{m1}(t) - \Lambda_{01}(t)| \\
+ \sup_{t \in [0, c]} |\hat{\Lambda}_{m2}(t) - \Lambda_{02}(t)| + \| \hat{\Sigma}_m - \Sigma_0 \| \right\} \right).
\]

based on the functional central limit theorem [Gamst et al. 2009].

Next, we prove that the derivative, denoted by \( \hat{S}(\alpha_0, \beta_{01}, \beta_{02}, \Lambda_0, \Lambda_0, \Sigma_0) \), is continuously invertible on its range. We consider \( \hat{S}(\alpha_0, \beta_{01}, \beta_{02}, \Lambda_0, \Lambda_0, \Sigma_0) \) as a linear map, simply denoted by \( T \), from the space
\{ (\alpha - \alpha_0, \beta_1 - \beta_{01}, \beta_2 - \beta_{02}, \Lambda_1 - \Lambda_{01}, \Lambda_2 - \Lambda_{02}, \Sigma - \Sigma_0) : (\alpha, \beta_1, \beta_2, \Lambda_1, \Lambda_2, \Sigma) \text{ is in the neighborhood} \mathcal{U} \text{ of} (\alpha_0, \beta_{01}, \beta_{02}, \Lambda_{01}, \Lambda_{02}, \Sigma_0) \} \text{ to} l^\infty(\mathcal{H}). \text{ That is }

T(\alpha - \alpha_0, \beta_1 - \beta_{01}, \beta_2 - \beta_{02}, \Lambda_1 - \Lambda_{01}, \Lambda_2 - \Lambda_{02}, \Sigma - \Sigma_0)

= (\alpha - \alpha_0)'Q_1(h_1, \ldots, h_6) + (\beta_1 - \beta_{01})'Q_2(h_1, \ldots, h_6) + (\beta_2 - \beta_{02})'Q_3(h_1, \ldots, h_6)

+ \int_0^\infty Q_4(h_1, \ldots, h_6)d(\Lambda_1 - \Lambda_{01}) + \int_0^\infty Q_5(h_1, \ldots, h_6)d(\Lambda_2 - \Lambda_{02}) + (\Sigma - \Sigma_0)'Q_6(h_1, \ldots, h_6),

where \( Q_i \) for \( i = 1, \ldots, 6 \) are the respective partial derivatives of \( S \) with respect to \( \alpha, \beta_1, \beta_2, \Lambda_1, \Lambda_2 \) and \( \Sigma \). Specifically, \( Q_i \) have the following forms:

\[
Q_1(h_1, \ldots, h_6) = B_1(h_1, h_2, h_3, h_6)' + \int_0^\infty h_4(t)D_1(t) + \int_0^\infty h_5(t)D_2(t),
\]

\[
Q_2(h_1, \ldots, h_6) = B_2(h_1, h_2, h_3, h_6)' + \int_0^\infty h_4(t)D_3(t) + \int_0^\infty h_5(t)D_4(t),
\]

\[
Q_3(h_1, \ldots, h_6) = B_3(h_1, h_2, h_3, h_6)' + \int_0^\infty h_4(t)D_5(t) + \int_0^\infty h_5(t)D_6(t),
\]

\[
Q_4(h_1, \ldots, h_6) = B_4(h_1, h_2, h_3, h_6)' + b_4h_4(t) + \int_0^\infty h_4(t)D_7(t) + \int_0^\infty h_5(t)D_8(t),
\]

\[
Q_5(h_1, \ldots, h_6) = B_5(h_1, h_2, h_3, h_6)' + b_5h_5(t) + \int_0^\infty h_5(t)D_9(t) + \int_0^\infty h_5(t)D_{10}(t),
\]

\[
Q_6(h_1, \ldots, h_6) = B_6(h_1, h_2, h_3, h_6)' + \int_0^\infty h_4(t)D_{11}(t) + \int_0^\infty h_5(t)D_{12}(t),
\]

where \( B_1, \ldots, B_6 \) are constant matrices, \( D_1(t), \ldots, D_{12}(t) \) are continuously differentiable functions and \( b_4 > 0, b_5 > 0 \). \( Q = (Q_1, \ldots, Q_6) \). To prove \( T \) is invertible, it suffices to show the invertibility of the linear operator \( Q(h_1, \ldots, h_6) \) (Rudin [1973]). Suppose \( Q(h_1, \ldots, h_6) = 0 \), then \( T(\alpha - \alpha_0, \beta_1 - \beta_{01}, \beta_2 - \beta_{02}, \Lambda_1 - \Lambda_{01}, \Lambda_2 - \Lambda_{02}, \Sigma - \Sigma_0)[h_1, \ldots, h_6] = 0 \) for any \( (\alpha, \beta_1, \beta_2, \Lambda_1, \Lambda_2, \Sigma) \) in the neighborhood \( \mathcal{U} \). For a small constant \( \epsilon \), let \( \alpha = \alpha_0 + \epsilon h_1, \beta_1 = \beta_{01} + \epsilon h_2, \beta_2 = \beta_{02} + \epsilon h_3, \Lambda_1(t) = \Lambda_{01}(t) + \epsilon \int_0^t h_4(s)d\Lambda_{01}(s), \Lambda_2(t) = \Lambda_{02}(t) + \int_0^t h_5(s)d\Lambda_{02}(s), \Sigma = \Sigma_0 + \epsilon h_6 \). It follows, by the definition of \( T \), that

\[
0 = T(\alpha - \alpha_0, \beta_1 - \beta_{01}, \beta_2 - \beta_{02}, \Lambda_1 - \Lambda_{01}, \Lambda_2 - \Lambda_{02}, \Sigma - \Sigma_0)[h_1, \ldots, h_6]
\]

\[
= \epsilon E \{ l_{\alpha_0}[h_1] + l_{\beta_{01}}[h_2] + l_{\beta_{02}}[h_3] + l_{\Lambda_{01}}[h_4] + l_{\Lambda_{02}}[h_5] + l_{\Sigma_0}[h_6] \}^2,
\]

where \( l_{\alpha_0} \) denotes the score function of \( \alpha \) evaluated at \( \alpha_0 \), \( l_{\beta_{0k}} \) the score function of \( \beta_k \) evaluated at \( \beta_{0k} \), \( l_{\Lambda_{0k}} \) the score function of \( \Lambda_k \) evaluated at \( \Lambda_{0k} \), for \( k = 1, 2 \), and \( l_{\Sigma_0} \) the score function of \( \Sigma \) evaluated at \( \Sigma_0 \). Then we have

\[
l_{\alpha_0}[h_1] + l_{\beta_{01}}[h_2] + l_{\beta_{02}}[h_3] + l_{\Lambda_{01}}[h_4] + l_{\Lambda_{02}}[h_5] + l_{\Sigma_0}[h_6] = 0
\]

(33)

46
almost surely, where

\[
R_{2i}(\omega, b_i) = R_{1i}(\omega, b_i) \prod_{j=1}^{n_i} \left\{ \lambda_{01}(T_{ij})^{\delta_{ij}^{(1)}} \lambda_{02}(\tilde{T}_{ij})^{\delta_{ij}^{(2)}} \right\},
\]  \tag{34}

\[
l_{\alpha[h1]} = \sum_{j=1}^{n_i} \int_{b_i} l_{\alpha, j}[h1] \psi(b_i, \Sigma_0) db_i
\]

\[
= \sum_{j=1}^{n_i} \int_{b_i} h_j^* X_{ij} \left\{ \delta_{ij}^{(1)} - \Lambda_{01}(T_{ij}) e^{\delta_{n1} x_{ij} + v_i^T z_i} \left( \delta_{ij}^{(1)} \tilde{\beta}_{ij}^{(2)} D_{121}(T_{ij}, \tilde{T}_{ij}) S_1(T_{ij}) + \delta_{ij}^{(1)} \right) + \delta_{ij}^{(2)} (1 - \delta_{ij}^{(2)}) D_{121}(T_{ij}, \tilde{T}_{ij}) S_1(T_{ij}) \right\} R_{2i}(\omega, b_i) \psi(b_i, \Sigma_0) db_i,
\]  \tag{35}

\[
l_{\beta[h2]} = \sum_{j=1}^{n_i} \int_{b_i} l_{\beta, j}[h2] \psi(b_i, \Sigma_0) db_i
\]

\[
= \sum_{j=1}^{n_i} \int_{b_i} h_j^* X_{ij} \left\{ \delta_{ij}^{(1)} \tilde{\beta}_{ij}^{(2)} D_{121}(T_{ij}, \tilde{T}_{ij}) S_1(T_{ij}) + \delta_{ij}^{(1)} (1 - \delta_{ij}^{(2)}) D_{121}(T_{ij}, \tilde{T}_{ij}) S_1(T_{ij}) + \delta_{ij}^{(2)} D_{212}(T_{ij}, \tilde{T}_{ij}) S_2(T_{ij}) \right\} R_{2i}(\omega, b_i) \psi(b_i, \Sigma_0) db_i,
\]  \tag{36}

\[
l_{\lambda[h3]} = \sum_{j=1}^{n_i} \int_{b_i} l_{\lambda, j}[h3] \psi(b_i, \Sigma_0) db_i
\]

\[
= \sum_{j=1}^{n_i} \int_{b_i} h_j^* X_{ij} \left\{ \delta_{ij}^{(1)} \tilde{\beta}_{ij}^{(2)} D_{121}(T_{ij}, \tilde{T}_{ij}) S_1(T_{ij}) + \delta_{ij}^{(1)} (1 - \delta_{ij}^{(2)}) D_{121}(T_{ij}, \tilde{T}_{ij}) S_1(T_{ij}) + \delta_{ij}^{(2)} D_{222}(T_{ij}, \tilde{T}_{ij}) S_2(T_{ij}) \right\} R_{2i}(\omega, b_i) \psi(b_i, \Sigma_0) db_i
\]  \tag{37}

\[
l_{\alpha[h4]} = \sum_{j=1}^{n_i} \int_{b_i} l_{\alpha, j}[h4] \psi(b_i, \Sigma_0) db_i
\]

\[
= \sum_{j=1}^{n_i} \int_{b_i} h_j^* X_{ij} \left\{ \delta_{ij}^{(1)} \tilde{\beta}_{ij}^{(2)} D_{121}(T_{ij}, \tilde{T}_{ij}) S_1(T_{ij}) + \delta_{ij}^{(1)} (1 - \delta_{ij}^{(2)}) D_{121}(T_{ij}, \tilde{T}_{ij}) S_1(T_{ij}) + \delta_{ij}^{(2)} D_{222}(T_{ij}, \tilde{T}_{ij}) S_2(T_{ij}) \right\} R_{2i}(\omega, b_i) \psi(b_i, \Sigma_0) db_i
\]  \tag{38}

47
\[ l_{\Lambda_0}[h_5] = \sum_{j=1}^{n_i} \int_{b_i} l_{\Lambda_{0,j}}[h_5] \psi(b_i, \Sigma_0) db_i \]

\[ = \sum_{j=1}^{n_i} \int_{b_i} \left\{ \left( \delta_{ij}^{(1)} \bar{\gamma}^{(2)} + \delta_{ij}^{(2)} \right) h_5(\tilde{T}_{ij}) - \int_{0}^{\tilde{T}_{ij}} h_5(s) d\Lambda_{02}(s) e^{\beta_{02} X_{ij} + b'_j Z_{ij}} \left( \delta_{ij}^{(1)} \bar{\gamma}^{(2)} \right) \frac{D_{12}(T_{ij}, \tilde{T}_{ij})}{D_{12}(T_{ij}, \tilde{T}_{ij})} S_2(\tilde{T}_{ij}) \right. \]

\[ + \delta_{ij}^{(1)} \bar{\gamma}^{(2)} + \delta_{ij}^{(2)}(1 - \bar{\gamma}^{(2)}) \frac{D_{12}(T_{ij}, \tilde{T}_{ij})}{D_{12}(T_{ij}, \tilde{T}_{ij})} S_2(\tilde{T}_{ij}) + \delta_{ij}^{(2)} \frac{D_{12}(T_{ij}, \tilde{T}_{ij})}{D_{12}(T_{ij}, \tilde{T}_{ij})} S_2(\tilde{T}_{ij}) \]

\[ + \left( 1 - \delta_{ij}^{(1)} - \delta_{ij}^{(2)} \right) \frac{D_{2}(T_{ij}, \tilde{T}_{ij})}{c \{ S_1(T_{ij}), S_2(T_{ij}) \}} S_2(\tilde{T}_{ij}) \right\} R_{2i}(\omega_0, b_i) \psi(b_i, \Sigma_0) db_i, \]

(39)

\[ l_{\Sigma_0}[h_6] = \int_{b_i} \frac{d}{de} \psi(b_i, \Sigma_0 + e h_6) \bigg|_{e=0} R_{2i}(\omega_0, b_i) \psi(b_i, \Sigma_0) db_i. \]

(40)

In the following, we show that equation (33) implies that \( h_1 = 0, h_2 = 0, h_3 = 0, h_4(\cdot) = 0, h_5(\cdot) = 0, h_6 = 0 \). Let \( X_{ij} \) and \( Z_{ij} \) be fixed. Set \( \delta_{ij}^{(1)} = 1, \delta_{ij}^{(2)} = 0 \) and \( T_{ij} = 0 \) for all \( j = 1, \cdots, n_i \). For fixed integer \( h \) in \( 1, \cdots, n_i \), let \( \delta_{ij}^{(2)} = 1, \tilde{T}_{ij} = 0 \) when \( j < h \). When \( j > h \), replace \( \tilde{T}_{ij} \) with \( \tilde{T}_{ij} \) if \( \delta_{ij}^{(2)} = 0 \) and integrate \( \tilde{T}_{ij} \) from 0 to \( \tilde{T}_{ij} \) otherwise. Then,

\[ R_{2i}(\omega_0, b) = \prod_{j=1}^{n_i} \lambda_{01}(T_{ij}) e^{\beta_{01} X_{ij} + b'_j Z_{ij}} D_{12}(T_{ij}, \tilde{T}_{ij})^{\delta_{ij}^{(1)}} D_{12}(T_{ij}, \tilde{T}_{ij})^{1-\delta_{ij}^{(2)}} \]

\[ \cdot \exp \left\{ \delta_{ij}^{(2)} \left( \beta_{02} X_{ij} + b'_j Z_{ij} \right) - \delta_{ij}^{(2)} \Lambda_{02}(T_{ij}) e^{\beta_{02} X_{ij} + b'_j Z_{ij}} \right\} \lambda_{02}(\tilde{T}_{ij})^{\delta_{ij}^{(2)}}. \]

(41)

For fixed integer \( h \) in \( 1, \cdots, n_i \), we define measures \( \mu_1, \cdots, \mu_{n_i} \) on the set \( \{0, 1\} \times [0, \zeta] \) as follows:

\[ \mu_r(\{0\} \times A) = 0, \mu_r(\{1\} \times A) = I(0 \in A), r \leq h \]

and

\[ \mu_r(\{0\} \times A) = I(\zeta \in A), \mu_r(\{1\} \times A) = \int I_A dt, r > h, \]

where \( A \) is any Borel set in \([0, \zeta]\.\)

We integrate both sides of equation (33) with respect to \( \{ (\delta_{1i}^{(2)}, \tilde{T}_{1i1}), (\delta_{1i}^{(2)}, \tilde{T}_{1i2}), \cdots, (\delta_{ni}^{(2)}, \tilde{T}_{ni}) \} \) and product
measure \, d\mu_1, \ldots, d\mu_n$, denoted by \( d\mu \) for simplicity. Note that, for any \( h \),

\[
\int R_{2i}(\omega_0, b) d\mu = \prod_{r \leq h} \left\{ e^{\beta_{0i}X_{ci} + \beta_{1i}'Z_{ci}} \right\} \prod_{r > h} \left\{ e^{\beta_{0i}X_{ci} + \beta_{1i}'Z_{ci}} \right\} 1^{\delta_{2i}}
\]

\[
= \prod_{r \leq h} \left\{ \lambda_{0i}(0) e^{\beta_{0i}X_{ci} + \beta_{1i}'Z_{ci}} \right\} \prod_{r > h} \left\{ \lambda_{0i}(0) e^{\beta_{0i}X_{ci} + \beta_{1i}'Z_{ci}} \right\} 1^{\delta_{2i}}
\]

Consider the first term \( l_{0j} [h_1] \) on the left side of equation \[ \ref{eq:33} \]. If \( j \leq h \), from \[ \ref{eq:35} \] we have

\[
l_{0j} [h_1] = \int h_1 \frac{D_{12a}(0,0) \ D_{12}(0,0)}{R_{2i}(\omega_0, b) d\mu} = h_1 \frac{D_{12a}(0,0) \ D_{12}(0,0)}{R_{2i}(\omega_0, b) d\mu}
\]

Otherwise if \( j > h \),

\[
l_{0j} [h_1] = h_1 \frac{D_{12a}(0,0) \ D_{12}(0,0)}{R_{2i}(\omega_0, b) d\mu}
\]

Then the contribution of \( l_{0j} [h_1] \) in \[ \ref{eq:33} \] reduces to

\[
h_1 \left\{ \sum_{j \leq h} \frac{D_{12a}(0,0)}{D_{12}(0,0)} + \sum_{j > h} \frac{D_{12a}(0,0)}{D_{12}(0,0)} \right\} \prod_{r \leq h} \left\{ e^{\beta_{0i}X_{ci} + \beta_{1i}'Z_{ci}} \right\} \prod_{r > h} \left\{ \lambda_{0i}(0) e^{\beta_{0i}X_{ci} + \beta_{1i}'Z_{ci}} \right\} 1^{\delta_{2i}}
\]

\[
\cdot \prod_{r > h} \left\{ \lambda_{0i}(0) e^{\beta_{0i}X_{ci} + \beta_{1i}'Z_{ci}} \right\} \cdot E \left( e^{\sum_{j=1}^{n} \beta_{1ij}Z_{ij} + \sum_{j=1}^{n} \beta_{1ij}Z_{ij}} \right).
\]
Consider the second term $l_{\beta_0} [h_2]$ on the left side of (33). If $j \leq h$, by (36) we have

$$l_{\beta_0,j} [h_2] = \int h_2 X_{ij} (1 - 0) \times R_{2i}(\omega_0, b) d\mu$$

$$= h_2 X_{ij} \prod_{r \leq h} \left\{ \lambda_{01}(0)e^{\beta_{01}X_{ir}+b'_rZ_{ir}}D_{12}(0, 0)e^{\beta_{02}X_{ir}+b'_rZ_{ir}}\lambda_{02}(0) \right\} \prod_{r > h} \left\{ \lambda_{01}(0)e^{\beta_{01}X_{ir}+b'_rZ_{ir}}D_{1}(0, 0) \right\};$$

Otherwise,

$$l_{\beta_0,j} [h_2] = h_2 X_{ij} \prod_{r \leq h} \left\{ \lambda_{01}(0)e^{\beta_{01}X_{ir}+b'_rZ_{ir}}D_{12}(0, 0)e^{\beta_{02}X_{ir}+b'_rZ_{ir}}\lambda_{02}(0) \right\} \prod_{r > h, r \neq j} \left\{ \lambda_{01}(0)e^{\beta_{01}X_{ir}+b'_rZ_{ir}}D_{1}(0, 0) \right\}$$

$$\cdot \sum_{\tilde{\delta}_{ij}^{(2)} \in \{0, 1\}} \left[ (1 - \tilde{\delta}_{ij}^{(2)})\lambda_{01}(0)e^{\beta_{01}X_{ij}+b'_rZ_{ij}}D_{1}(0, c) + \tilde{\delta}_{ij}^{(2)} \int_{\mu=0}^{\infty} \lambda_{01}(0)e^{\beta_{01}X_{ij}+b'_rZ_{ij}}D_{12}(0, y)S_2(y)e^{\beta_{02}X_{ij}+\beta'_rZ_{ij}}\lambda_{02}(y)dy \right]$$

$$= h_2 X_{ij} \prod_{r \leq h} \left\{ \lambda_{01}(0)e^{\beta_{01}X_{ir}+b'_rZ_{ir}}D_{12}(0, 0)e^{\beta_{02}X_{ir}+b'_rZ_{ir}}\lambda_{02}(0) \right\} \prod_{r > h, r \neq j} \left\{ \lambda_{01}(0)e^{\beta_{01}X_{ir}+b'_rZ_{ir}}D_{1}(0, 0) \right\}$$

$$\cdot \sum_{\tilde{\delta}_{ij}^{(2)} \in \{0, 1\}} \left[ (1 - \tilde{\delta}_{ij}^{(2)})\lambda_{01}(0)e^{\beta_{01}X_{ij}+b'_rZ_{ij}}D_{1}(0, c) + \tilde{\delta}_{ij}^{(2)} \lambda_{01}(0)e^{\beta_{01}X_{ij}+b'_rZ_{ij}} \left( - D_1(0, c) + D_{1}(0, 0) \right) \right]$$

$$= h_2 X_{ij} \prod_{r \leq h} \left\{ \lambda_{01}(0)e^{\beta_{01}X_{ir}+b'_rZ_{ir}}D_{12}(0, 0)e^{\beta_{02}X_{ir}+b'_rZ_{ir}}\lambda_{02}(0) \right\} \prod_{r > h} \left\{ \lambda_{01}(0)e^{\beta_{01}X_{ir}+b'_rZ_{ir}}D_{1}(0, 0) \right\}.$$
otherwise,

\[
\begin{align*}
    l_{\beta_{02}, [h_3]} &= h_3'X_{ij} \prod_{r \leq h} \left\{ \lambda_{01}(0)e^{\beta_{01}'X_{ir}+\beta_{02}'Z_{ir}} D_{12}(0, 0)e^{\beta_{02}'X_{ir}+\beta_{02}'Z_{ir}} \lambda_{02}(0) \right\} \prod_{r > h, r \neq j} \left\{ \lambda_{01}(0)e^{\beta_{01}'X_{ir}+\beta_{02}'Z_{ir}} D_{1}(0, 0) \right\} \\
    & \cdot \sum_{\delta_{ij}^{(2)} \in (0, 1)} \left\{ 1 - \delta_{ij}^{(2)} \right\} \left\{ -\Lambda_{02}(0) e^{\beta_{01}'X_{ij}+\beta_{02}'Z_{ij}} D_{12}(0, 0) S_{2}(0) \lambda_{01}(0) e^{\beta_{01}'X_{ij}+\beta_{02}'Z_{ij}} \lambda_{02}(0) \right\} \\
    & + \delta_{ij}^{(2)} \int_{y=0}^{\infty} \left\{ 1 - \Lambda_{02}(y) e^{\beta_{01}'X_{ij}+\beta_{02}'Z_{ij}} D_{12}(0, y) S_{2}(y) \lambda_{01}(0) e^{\beta_{01}'X_{ij}+\beta_{02}'Z_{ij}} \lambda_{02}(0) \right\} dy \\
    &= h_3'X_{ij} \prod_{r \leq h} \left\{ \lambda_{01}(0)e^{\beta_{01}'X_{ir}+\beta_{02}'Z_{ir}} D_{12}(0, 0)e^{\beta_{02}'X_{ir}+\beta_{02}'Z_{ir}} \lambda_{02}(0) \right\} \prod_{r > h, r \neq j} \left\{ \lambda_{01}(0)e^{\beta_{01}'X_{ir}+\beta_{02}'Z_{ir}} D_{1}(0, 0) \right\} \\
    & \cdot \sum_{\delta_{ij}^{(2)} \in (0, 1)} \left\{ 1 - \delta_{ij}^{(2)} \right\} \left\{ -\Lambda_{02}(0) e^{\beta_{01}'X_{ij}+\beta_{02}'Z_{ij}} D_{12}(0, 0) S_{2}(0) \lambda_{01}(0) e^{\beta_{01}'X_{ij}+\beta_{02}'Z_{ij}} \lambda_{02}(0) \right\} \\
    & + \delta_{ij}^{(2)} \lambda_{01}(0) e^{\beta_{01}'X_{ij}+\beta_{02}'Z_{ij}} \int_{y=0}^{\infty} \frac{\partial \Lambda_{02}(y)}{\partial y} dy \\
    &= h_3'X_{ij} \prod_{r \leq h} \left\{ \lambda_{01}(0)e^{\beta_{01}'X_{ir}+\beta_{02}'Z_{ir}} D_{12}(0, 0)e^{\beta_{02}'X_{ir}+\beta_{02}'Z_{ir}} \lambda_{02}(0) \right\} \prod_{r > h, r \neq j} \left\{ \lambda_{01}(0)e^{\beta_{01}'X_{ir}+\beta_{02}'Z_{ir}} D_{1}(0, 0) \right\} \cdot 0 = 0.
\end{align*}
\]

Therefore, the contribution of \(l_{\beta_{02}, [h_3]}\) in \(\text{(33)}\) turns to be

\[
\sum_{j \leq h} h_3'X_{ij} \prod_{r \leq h} \left\{ \lambda_{01}(0)e^{\beta_{01}'X_{ir}+\beta_{02}'Z_{ir}} D_{12}(0, 0)e^{\beta_{02}'X_{ir}+\beta_{02}'Z_{ir}} \lambda_{02}(0) \right\} \prod_{r > h} \left\{ \lambda_{01}(0)e^{\beta_{01}'X_{ir}+\beta_{02}'Z_{ir}} D_{1}(0, 0) \right\} \cdot E \left( \sum_{j=1}^{h} \beta_{02}'Z_{ij} + \sum_{j=1}^{h} \beta_{01}'X_{ij} \right).
\]

We consider the fourth term \(l_{\Lambda_{01}, [h_4]}\) of \(\text{(33)}\). If \(j \leq h\), by \(\text{(38)}\)

\[
\begin{align*}
    l_{\Lambda_{01}, [h_4]} &= h_4(0) \prod_{r \leq h} \left\{ \lambda_{01}(0)e^{\beta_{01}'X_{ir}+\beta_{02}'Z_{ir}} D_{12}(0, 0)e^{\beta_{02}'X_{ir}+\beta_{02}'Z_{ir}} \lambda_{02}(0) \right\} \prod_{r > h} \left\{ \lambda_{01}(0)e^{\beta_{01}'X_{ir}+\beta_{02}'Z_{ir}} D_{1}(0, 0) \right\}.
\end{align*}
\]

Otherwise,

\[
\begin{align*}
    l_{\Lambda_{01}, [h_4]} &= \prod_{r \leq h} \left\{ \lambda_{01}(0)e^{\beta_{01}'X_{ir}+\beta_{02}'Z_{ir}} D_{12}(0, 0)e^{\beta_{02}'X_{ir}+\beta_{02}'Z_{ir}} \lambda_{02}(0) \right\} \prod_{r > h} \left\{ \lambda_{01}(0)e^{\beta_{01}'X_{ir}+\beta_{02}'Z_{ir}} D_{1}(0, 0) \right\} \\
    & \cdot \sum_{\delta_{ij}^{(2)} \in (0, 1)} \left\{ 1 - \delta_{ij}^{(2)} \right\} h_4(0) \lambda_{01}(0) e^{\beta_{01}'X_{ij}+\beta_{02}'Z_{ij}} D_{1}(0, \varsigma) \\
    & + \delta_{ij}^{(2)} h_4(0) \int_{y=0}^{\infty} \lambda_{01}(0) e^{\beta_{01}'X_{ij}+\beta_{02}'Z_{ij}} D_{12}(0, y) S_{2}(y) e^{\beta_{02}'X_{ij}+\beta_{02}'Z_{ij}} \lambda_{02}(0) dy \\
    &= \prod_{r \leq h} \left\{ \lambda_{01}(0)e^{\beta_{01}'X_{ir}+\beta_{02}'Z_{ir}} D_{12}(0, 0)e^{\beta_{02}'X_{ir}+\beta_{02}'Z_{ir}} \lambda_{02}(0) \right\} \prod_{r > h, r \neq j} \left\{ \lambda_{01}(0)e^{\beta_{01}'X_{ir}+\beta_{02}'Z_{ir}} D_{1}(0, 0) \right\} \\
    & \cdot \sum_{\delta_{ij}^{(2)} \in (0, 1)} \left\{ 1 - \delta_{ij}^{(2)} \right\} h_4(0) \lambda_{01}(0) e^{\beta_{01}'X_{ij}+\beta_{02}'Z_{ij}} D_{1}(0, \varsigma) + \delta_{ij}^{(2)} h_4(0) \lambda_{01}(0) e^{\beta_{01}'X_{ij}+\beta_{02}'Z_{ij}} \left( -D_{1}(0, \varsigma) + D_{1}(0, 0) \right) \\
    &= h_4(0) \prod_{r \leq h} \left\{ \lambda_{01}(0)e^{\beta_{01}'X_{ir}+\beta_{02}'Z_{ir}} D_{12}(0, 0)e^{\beta_{02}'X_{ir}+\beta_{02}'Z_{ir}} \lambda_{02}(0) \right\} \prod_{r > h, r \neq j} \left\{ \lambda_{01}(0)e^{\beta_{01}'X_{ir}+\beta_{02}'Z_{ir}} D_{1}(0, 0) \right\}.
\end{align*}
\]
Therefore, the contribution of \( l_{\Lambda_01} [h_4] \) in (43) becomes

\[
\sum_{j=1}^{n_4} h_4(0) \prod_{r \leq h} \left\{ \lambda_0(0) e^{\beta_{01} X_{ir}} D_{12}(0, 0) e^{\beta_{02} X_{ir} \lambda_0(0)} \right\} \prod_{r > h} \left\{ \lambda_0(0) e^{\beta_{01} X_{ir} D_1(0, 0)} \right\} \cdot E \left( e^{\sum_{j=1}^{n_4} b_r Z_{ij} + \sum_{j=1}^{n_4} b_r' Z_{ij} \lambda_0(0)} \right).
\]  

(46)

For the fifth term \( l_{\Lambda_02} [h_5] \) of (43), if \( j \leq h \), it follows from (39) that for any \( b \)

\[
l_{\Lambda_02, j} [h_5] = h_5(0) \prod_{r \leq h} \left\{ \lambda_0(0) e^{\beta_{01} X_{ir} + b_r' Z_{ir} D_{12}(0, 0) e^{\beta_{02} X_{ir} + b_r' Z_{ir} \lambda_0(0)}} \right\} \prod_{r > h} \left\{ \lambda_0(0) e^{\beta_{01} X_{ir} + b_r' Z_{ir} D_1(0, 0)} \right\}.
\]

For the fifth term of (43), when \( j > h \),

\[
l_{\Lambda_02, j} [h_5] = \prod_{r \leq h} \left\{ \lambda_0(0) e^{\beta_{01} X_{ir} + b_r' Z_{ir} D_{12}(0, 0) e^{\beta_{02} X_{ir} + b_r' Z_{ir} \lambda_0(0)}} \right\} \prod_{r > h, r \neq j} \left\{ \lambda_0(0) e^{\beta_{01} X_{ir} + b_r' Z_{ir} D_1(0, 0)} \right\}
\]

\[
\cdot \sum_{\delta_{ij}^{(2)} \in \{0, 1\}} \left[ - (1 - \delta_{ij}^{(2)}) \int_0^\infty h_5(y) d\Lambda_02(y) e^{\beta_{02} X_{ij} + b_r' Z_{ij} \lambda_0(0)} e^{\beta_{01} X_{ij} + b_r' Z_{ij}} D_{12}(0, 0) S_2(c)
\]

\[
+ \delta_{ij}^{(2)} \int_0^\infty \left( h_5(y) - \int_0^\infty h_5(s) d\Lambda_02(s) e^{\beta_{02} X_{ij} + b_r' Z_{ij}} \left( 1 + \frac{D_{122}(0, y)}{D_{12}(0, y)} S_2(y) \right) \right) dy
\]

\[
\cdot D_{12}(0, y) \lambda_0(0) e^{\beta_{01} X_{ir} + b_r' Z_{ir} S_2(y) e^{\beta_{02} X_{ir} + b_r' Z_{ir} \lambda_0(0)} dy
\]

\[
= \prod_{r \leq h} \left\{ \lambda_0(0) e^{\beta_{01} X_{ir} + b_r' Z_{ir} D_{12}(0, 0) e^{\beta_{02} X_{ir} + b_r' Z_{ir} \lambda_0(0)}} \right\} \prod_{r > h, r \neq j} \left\{ \lambda_0(0) e^{\beta_{01} X_{ir} + b_r' Z_{ir} D_1(0, 0)} \right\}
\]

\[
\cdot \sum_{\delta_{ij}^{(2)} \in \{0, 1\}} \left[ - (1 - \delta_{ij}^{(2)}) \int_0^\infty h_5(y) d\Lambda_02(y) e^{\beta_{02} X_{ij} + b_r' Z_{ij} \lambda_0(0)} e^{\beta_{01} X_{ij} + b_r' Z_{ij}} D_{12}(0, 0) S_2(c)
\]

\[
+ \delta_{ij}^{(2)} \int_0^\infty \left( h_5(y) - \int_0^\infty h_5(s) d\Lambda_02(s) e^{\beta_{02} X_{ij} + b_r' Z_{ij}} \left( 1 + \frac{D_{122}(0, y)}{D_{12}(0, y)} S_2(y) \right) \right) dy
\]

\[
\cdot \int_0^\infty \lambda_0(0) e^{\beta_{01} X_{ir} + b_r' Z_{ir} D_{12}(0, 0) e^{\beta_{02} X_{ir} + b_r' Z_{ir} \lambda_0(0)} dy
\]

\[
= \prod_{r \leq h} \left\{ \lambda_0(0) e^{\beta_{01} X_{ir} + b_r' Z_{ir} D_{12}(0, 0) e^{\beta_{02} X_{ir} + b_r' Z_{ir} \lambda_0(0)}} \right\} \prod_{r > h, r \neq j} \left\{ \lambda_0(0) e^{\beta_{01} X_{ir} + b_r' Z_{ir} D_1(0, 0)} \right\} 
\cdot 0 = 0.
\]

Thus the contribution of \( l_{\Lambda_02} [h_5] \) in (43) becomes

\[
\sum_{j \leq h} h_5(0) \prod_{r \leq h} \left\{ \lambda_0(0) e^{\beta_{01} X_{ir} D_{12}(0, 0) e^{\beta_{02} X_{ir} \lambda_0(0)}} \right\} \prod_{r > h} \left\{ \lambda_0(0) e^{\beta_{01} X_{ir} D_1(0, 0)} \right\} E \left( e^{\sum_{j=1}^{n_4} b_r Z_{ij} + \sum_{j=1}^{n_4} b_r' Z_{ij} \lambda_0(0)} \right).
\]  

(47)

Through similar calculations, the sixth term \( l_{\Sigma_0} [h_6] \) of (43) reduces to:

\[
\prod_{r \leq h} \left\{ \lambda_0(0) e^{\beta_{01} X_{ir} D_{12}(0, 0) e^{\beta_{02} X_{ir} \lambda_0(0)}} \right\} \prod_{r > h} \left\{ \lambda_0(0) e^{\beta_{01} X_{ir} D_1(0, 0)} \right\}
\]

\[
\cdot E \left[ \frac{\partial \psi(b_1, \Sigma_0 + c h_6)}{\partial \epsilon} \right] e^{\sum_{j=1}^{n_4} b_r Z_{ij} + \sum_{j=1}^{n_4} b_r' Z_{ij} \lambda_0(0)}.
\]  

(48)

\[
\text{52}
\]
Applying condition 8 in the paper, we obtain
\[
\sum_{j=1}^{n_1} h_2 X_{ij} + \sum_{j=1}^{n_1} h_3 X_{ij} + \sum_{j=1}^{n_1} h_4(0) + \sum_{j=1}^{n_1} h_5(0) + \sum_{j=1}^{n_1} h_6(0)
\]
\[
\cdot E\left( e^{\sum_{j=1}^{n_1} b'_j Z_{ij} + \sum_{j=1}^{n_1} b'_j Z_{ij}} \right) + E\left( \left. \frac{\partial \theta(b_1, \Sigma_0 + \Theta b_0)}{\partial \epsilon} \right|_{\epsilon=0} \right) E\left( e^{\sum_{j=1}^{n_1} b'_j Z_{ij} + \sum_{j=1}^{n_1} b'_j Z_{ij}} \right) = 0.
\]

The following is to prove \([49]\) implies that \(h_1 = h_2 = h_3 = 0, h_4(\cdot) = 0, h_5(\cdot) = 0, h_6 = 0\).

Under the normal random-effects assumption, \([49]\) is rewritten as:
\[
\sum_{j=1}^{n_1} h_2 X_{ij} + \sum_{j=1}^{n_1} h_3 X_{ij} + \sum_{j=1}^{n_1} h_4(0) + \sum_{j=1}^{n_1} h_5(0) + \sum_{j=1}^{n_1} h_6(0)
\]
\[
= \frac{1}{2} \left( \sum_{j=1}^{n_1} \sum_{j=1}^{n_1} Z_{ij} \right) \cdot H_0 \left( \sum_{j=1}^{n_1} Z_{ij} \right) = 0.
\]

In \([50]\), we can see that only the term multiplied with \(h_1\) is a function of \(\alpha\), other terms are unrelated with \(\alpha\). \([50]\) exists over different values of \(\alpha\), and thus \(h_1 = 0\). Since the choice of \(h\) is arbitrary, we conclude that \(\sum_{j=1}^{n_1} h_2 X_{ij} + (k_2 - k_1) h_5(0) = 0\) for any \(1 \leq k_1 \leq k_2 \leq n_i\), leading to \(h_3 = 0\) and \(h_5(0) = 0\). Then \([50]\) is simplified as
\[
\sum_{j=1}^{n_1} h_2 X_{ij} + \sum_{j=1}^{n_1} h_3 X_{ij} + \sum_{j=1}^{n_1} h_4(0) + \sum_{j=1}^{n_1} h_5(0) + \sum_{j=1}^{n_1} h_6(0)
\]
\[
= \frac{1}{2} \left( \sum_{j=1}^{n_1} \sum_{j=1}^{n_1} Z_{ij} \right) \cdot H_0 \left( \sum_{j=1}^{n_1} Z_{ij} \right) = 0.
\]

Applying condition 8 in the paper, we obtain \(h_2 = 0, h_4(0) = 0, H_0 = 0\).

Setting \(\delta_{ij}^{(1)} = \delta_{ij}^{(2)} = 1\) and \(T_{ij} = 0\) for \(j = 1, \cdots, n_i, \bar{T}_{ij} = 0\) for \(j = 2, \cdots, n_i\), and \(c(u, v) = uv\) in \([38]\) gives that
\[
\tilde{h}_5(\bar{T}_{i1}) = \int_0^{\bar{T}_{i1}} h_5(s) \, d\Lambda_{02}(s) \left\{ \int_b e^{\sum_{j=1}^{n_1} b'_j Z_{ij} R_{21}(\omega_0, b) \phi(0, \Sigma_0) \, db} \right\} \leq \left\{ \int_b R_{21}(\omega_0, b) \phi(0, \Sigma_0) \, db \right\}.
\]

Therefore, \(g(y) = \int_0^y h_5(t) \, d\Lambda_{02}(t)\) satisfies the homogeneous equation
\[
\frac{g'(y)}{\lambda_{02}(y)} = \left\{ \frac{g(y)}{\int_b R_{21}(\omega_0, b) \phi(0, \Sigma_0) \, db} \right\} \leq \left\{ \int_b R_{21}(\omega_0, b) \phi(0, \Sigma_0) \, db \right\}.
\]

With boundary condition \(g(0) = 0\), we conclude \(g(y) = 0, h_5(y) = 0\) and thus \(h_5(\cdot) = 0\). Similarly, setting \(\tilde{\delta}_{ij}^{(1)} = 1\) for \(j = 1, \cdots, n_i, \tilde{\delta}_{ij}^{(2)} = 1\) and \(T_{ij} = \bar{T}_{ij} = 0\) for \(j = 2, \cdots, n_i, \tilde{\delta}_{i1}^{(2)} = 0\) and \(\bar{T}_{i1} = \zeta, c(u, v) = uv\) in \([38]\) finally yield that \(h_4(\cdot) = 0\).
Under the gamma frailty assumption, (49) is rewritten as:

\[
\begin{align*}
&h_1 \left\{ \sum_{j \leq h} \frac{D_{1(j)}}{D_{1(0,0)}} + \sum_{j > h} \frac{D_{1(h)}}{D_{1(0,0)}} \right\} + \sum_{j = 1}^{n_i} h_2 X_{ij} + \sum_{j = 1}^{n_i} h_3 X_{ij} + \sum_{j = 1}^{n_i} h_4(0) + \sum_{j = 1}^{n_i} h_5(0) \\
&+ \left\{ - \left( \frac{1}{\eta_0} - 1 \right) \frac{1}{(n_i + h - 2)\eta_0 + 1} - (n_i + h)\eta_0 - 2 + \log(\eta_0) + \phi(\frac{1}{\eta_0}) \right\} \cdot \frac{h_6}{\eta_0} = 0,
\end{align*}
\]

where \(\phi(\cdot)\) is the polygamma function.

Note that we can see that only the term multiplied with \(h_1\) is a function of \(\alpha_0\), other terms are unrelated with \(\alpha_0\). (52) exists over different values of \(\alpha_0\), and thus \(h_1 = 0\). In (52), we can see only the term multiplied with \(h_6\) is a function of \(\eta_0\), while the other terms are unrelated with \(\eta_0\). (52) exists over different values of \(\eta_0\), thus \(h_6 = 0\). Since the choice of \(h\) is arbitrary, we conclude that \(\sum_{j = k_1}^{k_2} h_3 X_{ij} + (k_2 - k_1)h_5(0) = 0\) for any \(1 \leq k_1 \leq k_2 \leq n_i\), yielding \(h_3 = 0\) and \(h_5(0) = 0\). Now (52) is changed to

\[
\sum_{j = 1}^{n_i} h_2 X_{ij} + \sum_{j = 1}^{n_i} h_4(0) = 0.
\]

Applying condition 7 in the paper, we have \(h_2 = 0\), \(h_4(0) = 0\). Set \(\delta^{(1)}_{ij} = \delta^{(2)}_{ij} = 1\) and \(T_{ij} = 0\) for \(j = 1, \ldots, n_i\), \(\tilde{T}_{ij} = 0\) for \(j = 2, \ldots, n_i\) and \(c(u, v) = uv\) in (33). It follows that

\[
h_5(\tilde{T}_{11}) = \left[ \int_0^{\tilde{T}_{11}} h_5(s) d\Lambda_2(s) \left\{ \int_u e^{\delta^{(2)}_{11} X_{11}} R_{21}(\omega_0, b) d\alpha \Gamma(1/\eta_0, \eta_0) \right\} \right] / \left\{ \int_u R_{21}(\omega_0, b) d\alpha \Gamma(1/\eta_0, \eta_0) \right\}.
\]

Therefore \(g(y) \equiv \int_0^y h_5(t) d\Lambda_2(t)\) satisfies the homogeneous equation

\[
\frac{g'(y)}{\lambda_{20}(y)} = \left[ g(y) \left\{ \int_u e^{\delta^{(2)}_{11} X_{11}} R_{21}(\omega_0, b) d\alpha \Gamma(1/\eta_0, \eta_0) \right\} \right] / \left\{ \int_u R_{21}(\omega_0, b) d\alpha \Gamma(1/\eta_0, \eta_0) \right\}.
\]

With boundary condition \(g(0) = 0\), we conclude \(g(y) = 0\), \(h_5(\cdot) = 0\), thus \(h_5(\cdot) = 0\). Similarly, set \(\delta^{(1)}_{ij} = 1\) for \(j = 1, \ldots, n_i\), \(\delta^{(2)}_{ij} = 1\) and \(T_{ij} = \tilde{T}_{ij} = 0\) for \(j = 2, \ldots, n_i\), \(\tilde{\delta}^{(2)}_{11} = 0\) and \(\tilde{T}_{11} = \zeta\), \(c(u, v) = uv\) in (33), it follows that \(h_4(\cdot) = 0\).

Now we have proved that \(h_1 = h_2 = h_3 = 0\), \(h_4(\cdot) = 0\), \(h_5(\cdot) = 0\), \(h_6 = 0\). This implies that \(Q\) is one-to-one and thus \(\hat{S}(\omega_0)\) is invertible. Thus the asymptotic normality of \(\hat{\alpha}, \hat{\beta}_1, \hat{\beta}_2, \hat{\Lambda}_1, \hat{\Lambda}_2\) and \(\hat{\Sigma}\) follows from the martingale central limit theorem in Theorem 2 of Murphy (1995) under the normal random-effects or gamma frailty assumption.