Semiparametric regression analysis of clustered survival data with semi-competing risks

Peng, Mengjiao; Xiang, Liming; Wang, Shanshan

2018

Peng, M., Xiang, L., & Wang, S. (2018). Semiparametric regression analysis of clustered survival data with semi-competing risks. *Computational Statistics & Data Analysis*, 124, 53-70. doi:10.1016/j.csda.2018.02.003

<https://hdl.handle.net/10356/90231>

<https://doi.org/10.1016/j.csda.2018.02.003>

© 2018 Elsevier B.V. All rights reserved. This paper was published in *Computational Statistics & Data Analysis* and is made available with permission of Elsevier B.V.

Downloaded on 26 Apr 2023 20:50:30 SGT

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 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 α , 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 nonterminal 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 an nonterminal event, “event 2” be an 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

$$\lambda_k(t|b; X, Z) = \lambda_{0k}(t)e^{\beta'_k X + b'Z} \quad (1)$$

for $(k = 1, 2)$, where $\lambda_k(t|b; X, Z)$ is the conditional hazard function of $T^{(k)}$ given the event and covariate history before t , $\lambda_{0k}(\cdot)$ is the corresponding baseline hazard function, β_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 Σ , 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 α , which quantifies the association between $T^{(1)}$ and $T^{(2)}$:

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

for $t^{(1)} \geq 0, t^{(2)} \geq 0$, where $c\{u, v; \alpha\}$ is a copula function, α is a constant association parameter in one-to-one correspondence with Kendall's τ , 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 $i = 1, \dots, m$, where $N = \sum_{i=1}^m n_i$ denotes the total number of observations. To proceed with estimation, further notation is needed. Denote by $T_{ij}^{(1)}$ the nonterminal event time for subject j in cluster i and by $T_{ij}^{(2)}$ 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 $T_{ij}^{(k)}$ ($k = 1, 2$) are independent with the hazard functions following the proportional hazards model with random effects,

$$\lambda_k(t|X_{ij}, Z_{ij}, b_i) = \lambda_{0k}(t)e^{\beta_k' X_{ij} + b_i' Z_i}. \quad (3)$$

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

$$S(t_{ij}^{(1)}, t_{ij}^{(2)}) = \text{pr}(T_{ij}^{(1)} \geq t_{ij}^{(1)}, T_{ij}^{(2)} \geq t_{ij}^{(2)}) = c\{S_1(t_{ij}^{(1)}), S_2(t_{ij}^{(2)}); \alpha\}, t_{ij}^{(1)} \geq 0, t_{ij}^{(2)} \geq 0. \quad (4)$$

Let C^* denote the external censoring time that is independent of $(T^{(1)}, T^{(2)})$, and ς denote the duration of the study. Due to possible right-censoring, for subject j in cluster i we observe $\mathcal{O}_{ij} = \{T_{ij}, \tilde{T}_{ij}, X_{ij}, Z_{ij}, \delta_{ij}^{(1)}, \delta_{ij}^{(2)}, \tilde{\delta}_{ij}^{(2)}\}$, where $T_{ij} = \min(T_{ij}^{(1)}, T_{ij}^{(2)}, C_{ij})$ with $C_{ij} = \min(C_{ij}^*, \varsigma)$, $\tilde{T}_{ij} = \min(T_{ij}^{(2)}, C_{ij})$, $\delta_{ij}^{(k)} = I(T_{ij} = T_{ij}^{(k)})$ ($k = 1, 2$), $\tilde{\delta}_{ij}^{(2)} = I(\tilde{T}_{ij} = T_{ij}^{(2)})$, 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

$\mathcal{O} = \{\mathcal{O}_i\}_{i=1}^m$, where $\mathcal{O}_i = \{\mathcal{O}_{ij}\}_{j=1}^{m_i}$. 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 $\mathcal{O}$ is

$$L(\theta, \lambda, \Sigma | \mathcal{O}) = \prod_{i=1}^m \int_{b_i} L_i(\theta, \lambda | \mathcal{O}_i, b_i) \psi(b_i; \Sigma) db_i = \prod_{i=1}^m \int_{b_i} \prod_{j=1}^{n_i} L_{ij}(\theta, \lambda | \mathcal{O}_{ij}, b_i) \psi(b_i; \Sigma) db_i, \quad (5)$$

where

$$\begin{aligned} L_{ij}(\theta, \lambda | b_i, \mathcal{O}_{ij}) = & \Pr(T_{ij}^{(1)} = T_{ij}, T_{ij}^{(2)} = \tilde{T}_{ij} | b_i)^{\delta_{ij}^{(1)} \tilde{\delta}_{ij}^{(2)}} \cdot \Pr(T_{ij}^{(1)} = T_{ij}, T_{ij}^{(2)} > \tilde{T}_{ij} | b_i)^{\delta_{ij}^{(1)} (1 - \tilde{\delta}_{ij}^{(2)})} \\ & \cdot \Pr(C_{ij} \leq T_{ij}^{(1)}, C_{ij} \leq T_{ij}^{(2)} | b_i)^{(1 - \delta_{ij}^{(1)} - \delta_{ij}^{(2)})} \cdot \Pr(T_{ij}^{(2)} \leq C_{ij}, T_{ij}^{(2)} \leq T_{ij}^{(1)} | b_i)^{\delta_{ij}^{(2)}}. \end{aligned}$$

Plugging models (3) and (4) in the likelihood, (5) can be written as:

$$\begin{aligned}
L(\theta, \lambda, \Sigma | \mathcal{O}) = & \prod_{i=1}^m \int_{b_i} \prod_{j=1}^{n_i} \left[c_{12} \{S_1(T_{ij}), S_2(\tilde{T}_{ij}); \alpha\} S_1(T_{ij}) \lambda_{01}(T_{ij}) \right. \\
& \cdot e^{\beta'_1 x_{ij} + b'_i z_{ij}} S_2(\tilde{T}_{ij}) \lambda_{02}(\tilde{T}_{ij}) e^{\beta'_2 x_{ij} + b'_i z_{ij}} \left. \right]^{\delta_{ij}^{(1)} \tilde{\delta}_{ij}^{(2)}} \\
& \cdot \left[c_1 \{S_1(T_{ij}), S_2(\tilde{T}_{ij}); \alpha\} S_1(T_{ij}) \lambda_{01}(T_{ij}) e^{\beta'_1 x_{ij} + b'_i z_{ij}} \right]^{\delta_{ij}^{(1)} (1 - \tilde{\delta}_{ij}^{(2)})} \\
& \cdot \left[c \{S_1(T_{ij}), S_2(\tilde{T}_{ij})\} \right]^{(1 - \delta_{ij}^{(1)} - \delta_{ij}^{(2)})} \\
& \cdot \left[c_2 \{S_1(T_{ij}), S_2(\tilde{T}_{ij}); \alpha\} S_2(\tilde{T}_{ij}) \lambda_{02}(\tilde{T}_{ij}) e^{\beta'_2 x_{ij} + b'_i z_{ij}} \right]^{\delta_{ij}^{(2)}} \psi(b_i; \Sigma) db_i,
\end{aligned} \tag{6}$$

where $c_1\{u, v; \alpha\} = \frac{c\{u, v; \alpha\}}{\partial u}$, $c_2\{u, v; \alpha\} = \frac{c\{u, v; \alpha\}}{\partial v}$, $c_{12}\{u, v; \alpha\} = \frac{c\{u, v; \alpha\}}{\partial u \partial v}$.

Before deriving the maximum of $L(\theta, \lambda, \Sigma | \mathcal{O})$ 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 $\tilde{N}_{ij}(t) = \delta_{ij}^{(1)} \tilde{\delta}_{ij}^{(2)}(1 - \tilde{Y}_{ij}(t))$ 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 $i = 1, \dots, m$ and $j = 1, \dots, n_i$.

The maximum of $L(\theta, \lambda, \Sigma | \mathcal{O})$ 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 Λ_{01} and Λ_{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

$$\hat{\Lambda}_{01}\{T_{ij}\} = \frac{dN_{ij}^{(1)}(T_{ij})}{\sum_{(g,h) \in R(T_{ij})} e^{\beta'_1 X_{gh}} E\{w_1(T_{gh}) e^{b'_g Z_{gh}}\}}, \tag{7}$$

$$\hat{\Lambda}_{02}\{\tilde{T}_{ij}\} = \frac{dN_{ij}^{(2)}(\tilde{T}_{ij}) + d\tilde{N}_{ij}^{(2)}(\tilde{T}_{ij})}{\sum_{(g,h) \in \tilde{R}(\tilde{T}_{ij})} e^{\beta'_2 X_{gh}} E\{w_2(\tilde{T}_{gh}) e^{b'_g Z_{gh}}\}}, \tag{8}$$

where for $(g, h) \in R(T_{ij})$ or $(g, h) \in \tilde{R}(\tilde{T}_{ij})$, 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_g , and

$$\begin{aligned} w_1(T_{gh}) &= \delta_{gh}^{(1)} \tilde{\delta}_{gh}^{(2)} \left[\frac{D_{121}(T_{gh}, \tilde{T}_{gh})}{D_{12}(T_{gh}, \tilde{T}_{gh})} S_1(T_{gh}) + 1 \right] + \delta_{gh}^{(1)} (1 - \tilde{\delta}_{gh}^{(2)}) \left[\frac{D_{11}(T_{gh}, \tilde{T}_{gh})}{D_1(T_{gh}, \tilde{T}_{gh})} S_1(T_{gh}) + 1 \right] \\ &\quad + (1 - \delta_{gh}^{(1)} - \delta_{gh}^{(2)}) \left[\frac{D_1(T_{gh}, \tilde{T}_{gh}) \cdot S_1(T_{gh})}{c\{S_1(T_{gh}), S_2(\tilde{T}_{gh})\}} \right] + \delta_{gh}^{(2)} \left[\frac{D_{21}(T_{gh}, \tilde{T}_{gh})}{D_2(T_{gh}, \tilde{T}_{gh})} S_1(T_{gh}) \right], \\ w_2(\tilde{T}_{gh}) &= \delta_{gh}^{(1)} \tilde{\delta}_{gh}^{(2)} \left[\frac{D_{122}(T_{gh}, \tilde{T}_{gh})}{D_{12}(T_{gh}, \tilde{T}_{gh})} S_2(\tilde{T}_{gh}) + 1 \right] + \delta_{gh}^{(1)} (1 - \tilde{\delta}_{gh}^{(2)}) \left[\frac{D_{12}(T_{gh}, \tilde{T}_{gh})}{D_1(T_{gh}, \tilde{T}_{gh})} S_2(\tilde{T}_{gh}) \right] \\ &\quad + (1 - \delta_{gh}^{(1)} - \delta_{gh}^{(2)}) \left[\frac{D_2(T_{gh}, \tilde{T}_{gh}) \cdot S_2(\tilde{T}_{gh})}{c\{S_1(T_{gh}), S_2(\tilde{T}_{gh})\}} \right] + \delta_{gh}^{(2)} \left[\frac{D_{22}(T_{gh}, \tilde{T}_{gh})}{D_2(T_{gh}, \tilde{T}_{gh})} S_2(\tilde{T}_{gh}) + 1 \right], \end{aligned}$$

in which we resort to notation for derivatives:

$$\begin{aligned} D_1(t^{(1)}, t^{(2)}) &= \frac{\partial c\{S_1(t^{(1)}), S_2(t^{(2)})\}}{\partial S_1(t^{(1)})}, & D_2(t^{(1)}, t^{(2)}) &= \frac{\partial c\{S_1(t^{(1)}), S_2(t^{(2)})\}}{\partial S_2(t^{(2)})}, \\ D_{11}(t^{(1)}, t^{(2)}) &= \frac{\partial^2 c\{S_1(t^{(1)}), S_2(t^{(2)})\}}{\partial^2 S_1(t^{(1)})}, & D_{12}(t^{(1)}, t^{(2)}) &= \frac{\partial^2 c\{S_1(t^{(1)}), S_2(t^{(2)})\}}{\partial S_1(t^{(1)}) \partial S_2(t^{(2)})}, \\ D_{21}(t^{(1)}, t^{(2)}) &= \frac{\partial^2 c\{S_1(t^{(1)}), S_2(t^{(2)})\}}{\partial S_2(t^{(2)}) \partial S_1(t^{(1)})}, & D_{22}(t^{(1)}, t^{(2)}) &= \frac{\partial^2 c\{S_1(t^{(1)}), S_2(t^{(2)})\}}{\partial^2 S_2(t^{(2)})}, \\ D_{121}(t^{(1)}, t^{(2)}) &= \frac{\partial^3 c\{S_1(t^{(1)}), S_2(t^{(2)})\}}{\partial S_1(t^{(1)}) \partial S_2(t^{(2)}) \partial S_1(t^{(1)})}, & D_{122}(t^{(1)}, t^{(2)}) &= \frac{\partial^3 c\{S_1(t^{(1)}), S_2(t^{(2)})\}}{\partial S_1(t^{(1)}) \partial S_2(t^{(2)}) \partial S_2(t^{(2)})}. \end{aligned} \tag{9}$$

Thus the cumulative [baseline](#) hazard functions for the nonterminal and terminal events are estimated by the following piece-wise constant functions

$$\hat{\Lambda}_{01}(t) = \sum_{(i,j) \in R(t)} \hat{\Lambda}_{01}\{T_{ij}\}, \tag{10}$$

$$\hat{\Lambda}_{02}(t) = \sum_{(i,j) \in \tilde{R}(t)} \hat{\Lambda}_{02}\{\tilde{T}_{ij}\}, \tag{11}$$

respectively. Plugging $\lambda_{01}(t)$ and $\lambda_{02}(t)$ by jump sizes $\hat{\Lambda}_{01}\{T_{ij}\}$ and $\hat{\Lambda}_{01}\{\tilde{T}_{ij}\}$, respectively, into the modified likelihood (6), the updated value of θ 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 expect-

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:

$$L(\theta|\mathcal{O}, \hat{\Lambda}, \hat{\Sigma}) = \prod_{i=1}^m L_i(\theta|\mathcal{O}_i, \hat{\Lambda}, \hat{\Sigma}) = \prod_{i=1}^m \int_{b_i} L_i(\theta|\mathcal{O}_i, \hat{\Lambda}, b_i) \psi(b_i; \hat{\Sigma}) db_i, \quad (12)$$

where

$$\begin{aligned} L_i(\theta|\mathcal{O}_i, \hat{\Lambda}, b_i) = & \prod_{j=1}^{n_i} \left[\frac{\partial^2 c\{\hat{S}_1(T_{ij}), \hat{S}_2(\tilde{T}_{ij})\}}{\partial \hat{S}_1(T_{ij}) \partial \hat{S}_2(\tilde{T}_{ij})} \hat{S}_1(T_{ij}) \hat{\Lambda}_{01}\{T_{ij}\} e^{\beta'_1 X_{ij} + b'_i Z_{ij}} \hat{S}_2(\tilde{T}_{ij}) \hat{\Lambda}_{02}\{\tilde{T}_{ij}\} e^{\beta'_2 X_{ij} + b'_i Z_{ij}} \right]^{\delta_{ij}^{(1)} \tilde{\delta}_{ij}^{(2)}} \\ & \cdot \left[\frac{\partial c\{\hat{S}_1(T_{ij}), \hat{S}_2(\tilde{T}_{ij})\}}{\partial \hat{S}_1(T_{ij})} \hat{S}_1(T_{ij}) \hat{\Lambda}_{01}\{T_{ij}\} e^{\beta'_1 X_{ij} + b'_i Z_{ij}} \right]^{\delta_{ij}^{(1)} (1 - \tilde{\delta}_{ij}^{(2)})} \\ & \cdot \left[c\{\hat{S}_1(T_{ij}), \hat{S}_2(\tilde{T}_{ij})\} \right]^{(1 - \delta_{ij}^{(1)} - \tilde{\delta}_{ij}^{(2)})} \left[\frac{\partial c\{\hat{S}_1(T_{ij}), \hat{S}_2(\tilde{T}_{ij})\}}{\partial \hat{S}_2(\tilde{T}_{ij})} \hat{S}_2(\tilde{T}_{ij}) \hat{\Lambda}_{02}\{\tilde{T}_{ij}\} e^{\beta'_2 X_{ij} + b'_i Z_{ij}} \right]^{\delta_{ij}^{(2)}}. \end{aligned}$$

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.

$$E[L_i(\theta|\mathcal{O}_i, \hat{\Lambda}^{[k]})] = \int_{b_i} L_i(\theta|\mathcal{O}_i, \hat{\Lambda}^{[k]}, b_i) f(b_i|\mathcal{O}_i, \hat{\omega}^{[k]}) db_i. \quad (13)$$

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 $\{b_i^q, q = 1, \dots, Q\}$, from a distribution with density $f(b_i|\mathcal{O}_i, \hat{\omega}^{[k]})$ for $i = 1, \dots, m$ 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 $f(b_i|\mathcal{O}_i, \hat{\omega}^{[k]})$ with respect to b_i ,

$$f(b_i|\mathcal{O}_i, \hat{\omega}^{[k]}) = \frac{f(b_i, \hat{\omega}^{[k]}|\mathcal{O}_i)}{f(\hat{\omega}^{[k]}|\mathcal{O}_i)} = \frac{f(\hat{\omega}^{[k]}|\mathcal{O}_i, b_i) \cdot \psi(b_i; \hat{\Sigma}^{[k]})}{\int_{b_i} f(\hat{\theta}^{[k]}, \hat{\Lambda}^{[k]}|\mathcal{O}_i, b_i) \cdot \psi(b_i; \hat{\Sigma}^{[k]}) db_i}.$$

It is equivalent to maximizing $f(\hat{\omega}^{[k]}|\mathcal{O}_i, b_i) \cdot \psi(b_i; \hat{\Sigma}^{[k]})$ with respect to b_i , where $f(\hat{\omega}^{[k]}|\mathcal{O}_i, b_i) = L_i(\hat{\theta}^{[k]}|\mathcal{O}_i, \hat{\Lambda}^{[k]}, b_i)$. 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, $b_i^q \sim N(\tilde{b}_i, I^{-1}(\tilde{b}_i))$ with pdf $\phi(\tilde{b}_i; I^{-1}(\tilde{b}_i))$ for $i = 1, \dots, m$ and $q = 1, \dots, Q$. More specifically,

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

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

$$M_i = \max_l \left\{ f(b_i^q | \mathcal{O}_i, \hat{\omega}^{[k]}) / \phi(b_i^q; I^{-1}(\tilde{b}_i)) \right\}$$

in our simulation studies.

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

$$\hat{E}\{L_i(\theta | \mathcal{O}_i, \hat{\Lambda}^{[k]})\} = \frac{1}{Q} \sum_{q=1}^Q L_i(\theta | \mathcal{O}_i, \hat{\Lambda}^{[k]}, b_i^q). \quad (14)$$

MCMC-step 4: $\hat{\Sigma}^{[k+1]}$ can be computed by the sample variance or variance-covariance matrix of MCMC sample $\{b_i^q, q = 1, \dots, Q\}$ for $i = 1, \dots, m$.

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

$$\hat{E}\{L(\theta | \mathcal{O}_i, \hat{\Lambda}^{[k]})\} = \prod_{i=1}^n \hat{E}\{L_i(\theta | \mathcal{O}_i, \hat{\Lambda}^{[k]})\}$$

using numerical optimization algorithm. Given the estimates $\hat{\theta}^{[k+1]}$ we can consequently compute the updated estimate of Λ , $\hat{\Lambda}^{[k+1]}$, 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 (θ, Σ) are all less than or the score function of θ is less than a small quantity ϵ . 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 β and Λ 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. α 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 α . 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 $T_{ij}^{(k)}$ ($k = 1, 2$) and the random effects b_i .*

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

Condition 3 *The true value of θ and Σ are elements of the interior of a known compact set $\mathcal{K} = \{(\theta, \Sigma) : |\theta| \leq B \text{ for some constant } B, \text{ and } \Sigma \text{ is symmetric and positive definite with eigenvalues bounded away from } 0 \text{ and } \infty.\}$*

Condition 4 With probability 1, for $i = 1, \dots, m$ and $j = 1, \dots, n_i$, $\Pr(T_{ij} > \varsigma | X, Z) > 0$, $\Pr(\delta_{ij}^{(1)} = 0, \delta_{ij}^{(2)} = 0, T_{ij} = \varsigma | X, Z) > 0$, $\Pr(\tilde{T}_{ij} > \varsigma | X, Z) > 0$, $\Pr(\delta_{ij}^{(1)} = 0, \tilde{\delta}_{ij}^{(2)} = 0, \tilde{T}_{ij} = \varsigma | X, Z) > 0$.

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 $i = 1, \dots, m$ 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 $i = 1, \dots, (1 + p + q(q+1)/2)$ 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 1 is standard and required for identifiability in the presence of censoring. Conditions 2-5 ensures the boundedness of $\hat{\Lambda}$ and the compactness of θ . 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 Λ , 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 1-8 hold and random effects b_i follow a normal distribution $N(0, \Sigma_0)$ i.i.d.. Then $|\hat{\theta}_m - \theta_0| \rightarrow 0$, $\|\hat{\Lambda}_{mk} - \Lambda_{0k}\|_{l^\infty[0, \varsigma]} \rightarrow 0$ ($k = 1, 2$) and $|\hat{\Sigma}_m - \Sigma_0| \rightarrow 0$ as $m \rightarrow \infty$, where $(\theta_0, \Lambda_{0k}, \Sigma_0)$ are true values of parameters and $\|\cdot\|_{l^\infty[0, \varsigma]}$ is the supremum norm in the interval $[0, \varsigma]$.

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 η_0 for $i = 1, \dots, m$, $j = 1, \dots, n_i$. Then $|\hat{\theta}_m - \theta_0| \rightarrow 0$, $\|\hat{\Lambda}_{mk} - \Lambda_{0k}\|_{l^\infty[0, \varsigma]} \rightarrow 0$ ($k = 1, 2$) and $|\hat{\eta}_m - \eta_0| \rightarrow 0$ as $m \rightarrow \infty$, where $\theta_0, \Lambda_{0k}, \eta_0$ are true values of parameters.

Theorem 2 (Asymptotic Normality) 1) Assume that conditions 1-8 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 $m \rightarrow \infty$.

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 η_0 for $i = 1, \dots, m$, $j = 1, \dots, n_i$. 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 $m \rightarrow \infty$.

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 ± 2 . We consider two different distributions for random effects: b follows a normal distribution with mean 0 and variance η and $u = \exp(b)$ follows a gamma distribution $\Gamma(1/\eta, \eta)$ with mean 1 and variance η . The joint distribution of $(T^{(1)}, T^{(2)})$ is specified by a Gumbel copula with association parameter $\alpha = 1/(1 - \tau) - 1$, 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 $i = 1, \dots, m$ 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 τ , the reduction in relative bias (bias/true value) and sd tends to be more significant when the true value of τ 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 τ .

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.

Tables 1-4 about here

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 η and Kendall's τ than their approach in terms of both mean bias and sd, while estimates of regression coefficients β_1 and β_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 η and comparable estimates for the other parameters in comparison with the existing counterpart. As association τ increases, our estimates of η and τ significantly improve in terms of rmse.

Tables 5-6 about here

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.](#)

Figures 1-4 about here

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 η only especially when the value of τ 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 η , 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/>. 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= positive and 0=negative), tumor size in cm (*size*, the longest length of invasive tumor measured by pathologists or surgeons.), lymph nodal status (*node*, 1=present and 0=absent), age at diagnosis (*age*, 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 $\lambda_k(t|b; X, Z) = \lambda_{0k}(t)e^{\beta'_k X + b_k}$ 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 η 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-

tributed 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

- B. S. Caffo, W. Jank, and G. L. Jones. Ascent-based monte carlo expectation maximization. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 67(2):235–251, 2005. ISSN 1467-9868.
- Y. H. Chen. Maximum likelihood analysis of semicompeting risks data with semiparametric regression models. Lifetime Data Analysis, 18(1):36–57, 2012.
- N. J. Christian, I. D. Ha, and J.-H. Jeong. Hierarchical likelihood inference on clustered competing risks data. Statistics in Medicine, 35(2):251–267, 2016.
- T. Emura, M. Nakatochi, K. Murotani, and V. Rondeau. A joint frailty-copula model between tumour progression and death for meta-analysis. Statistical Methods in Medical Research, 2015. doi: 10.1177/0962280215604510.
- A. Gamst, M. Donohue, and R. Xu. Asymptotic properties and empirical evaluation of the npml in the proportional hazards mixed-effects model. Statistica Sinica, 19(3):997–1011, 2009.

- M. Gorfine and L. Hsu. Frailty-based competing risks model for multivariate survival data. Biometrics, 67(2):415–426, 2011.
- B. Haibe-Kains, C. Desmedt, S. Loi, A. C. Culhane, G. Bontempi, J. Quackenbush, and C. Sotiriou. A three-gene model to robustly identify breast cancer molecular subtypes. Journal of the National Cancer Institute, 104(4):311–325, 2012.
- J. J. Hsieh, W. J. Wang, and A. D. A. Regression analysis based on semicompeting risks data. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 70(1):3–20, 2008.
- F. Jiang and S. Haneuse. A semi-parametric transformation frailty model for semi-competing risks survival data. Scandinavian Journal of Statistics, 44(1):112–129, 2016.
- L. Lakhal, L. Rivest, and B. Abdous. Estimating survival and association in a semicompeting risks model. Biometrics, 64(1):180–188, 2008.
- K. H. Lee, F. Dominici, D. Schrag, and S. Haneuse. Hierarchical models for semicompeting risks data with application to quality of end-of-life care for pancreatic cancer. Journal of the American Statistical Association, 111(515):1075–1095, 2016.
- M. Lee, I. D. Ha, and Y. Lee. Frailty modeling for clustered competing risks data with missing cause of failure. Statistical Methods in Medical Research, 26(1):356–373, 2017. doi: 10.1177/0962280214545639. PMID: 25125452.
- R. A. Levine and G. Casella. Implementations of the monte carlo em algorithm. Journal of Computational and Graphical Statistics, 10(3):422–439, 2001.
- S. Murphy. Asymptotic theory for the frailty model. The Annals of Statistics, 23:182–198, 1995.
- R. Nelsen. An introduction to copulas, ser. Lecture Notes in Statistics. New York: Springer, 2006.
- L. M. Peng and J. P. Fine. Regression modeling of semicompeting risks data. Biometrics, 63(1):96–108, 2007.
- D. Pollard. Convergence of stochastic processes. Springer, New York, 1984.
- S. Ripatti and J. Palmgren. Estimation of multivariate frailty models using penalized partial likelihood. Biometrics, 56(4):1016–1022, 2000.
- S. Ripatti, K. Larsen, and J. Palmgren. Maximum likelihood inference for multivariate frailty models using an automated monte carlo em algorithm. Lifetime Data Analysis, 8(4):349–360, 2002.
- W. Rudin. Functional analysis. McGraw-Hill Series in Higher Mathematics, 1973.

- J. Shao et al. Impact of the bootstrap on sample surveys. Statistical Science, 18(2):191–198, 2003.
- F. Vaida and R. H. Xu. Proportional hazards model with random effects. Statistics in Medicine, 19(24):3309–3324, 2000.
- M. J. Van De Vijver, Y. D. He, L. J. van’t Veer, H. Dai, A. A. Hart, D. W. Voskuil, G. J. Schreiber, J. L. Peterse, C. Roberts, M. J. Marton, et al. A gene-expression signature as a predictor of survival in breast cancer. New England Journal of Medicine, 347(25):1999–2009, 2002.
- M. H. van Vliet, F. Reyat, H. M. Horlings, M. J. van de Vijver, M. J. Reinders, and L. F. Wessels. Pooling breast cancer datasets has a synergetic effect on classification performance and improves signature stability. BMC Genomics, 9(1):375, 2008.
- L. J. Van’t Veer, H. Dai, M. J. Van De Vijver, Y. D. He, A. A. Hart, M. Mao, H. L. Peterse, K. van der Kooy, M. J. Marton, A. T. Witteveen, et al. Gene expression profiling predicts clinical outcome of breast cancer. Nature, 415(6871):530–536, 2002.
- G. C. Wei and M. A. Tanner. A monte carlo implementation of the em algorithm and the poor man’s data augmentation algorithms. Journal of the American Statistical Association, 85(411):699–704, 1990.
- J. Xu, J. D. Kalbfleisch, and B. Tai. Statistical analysis of illnessdeath processes and semicompeting risks data. Biometrics, 66(3):716–725, 2010.
- D. Zeng, D. Lin, and G. Yin. Maximum likelihood estimation for the proportional odds model with random effects. Journal of the American Statistical Association, 100(470):470–483, 2005.

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 .

(N, m, n_i)		η	τ	β_1	β_2
$\tau = 0.3$					
(50, 5, 10)	bias	-0.0632	0.0196	0.0185	0.0402
	sd	0.1834	0.1063	0.2846	0.3062
	cp(%)	95.0	94.0	94.5	95.0
(100, 5, 20)	bias	-0.0406	-0.0100	0.0206	0.0269
	sd	0.1613	0.0854	0.1818	0.2125
	cp(%)	96.0	95.5	94.0	96.0
(100, 10, 10)	bias	-0.0054	0.0190	-0.0007	0.0200
	sd	0.1477	0.0799	0.1913	0.1846
	cp(%)	96.5	95.5	93.5	94.5
(200, 10, 20)	bias	-0.0207	0.0182	-0.0046	0.0084
	sd	0.1250	0.0577	0.1200	0.1295
	cp(%)	96.5	95.0	95.5	93.5
$\tau = 0.6$					
(50, 5, 10)	bias	-0.0102	-0.0046	0.0797	0.0498
	sd	0.2911	0.0963	0.6790	0.4672
	cp(%)	98.5	96.5	99.0	98.5
(100, 5, 20)	bias	-0.0533	0.0023	0.0427	-0.0198
	sd	0.1495	0.0563	0.1860	0.1646
	cp(%)	96.0	95.0	93.0	95.0
(100, 10, 10)	bias	-0.0270	-0.0016	0.0465	-0.0110
	sd	0.1607	0.0571	0.1908	0.1901
	cp(%)	97.0	94.5	93.0	93.0
(200, 10, 20)	bias	-0.0251	0.0006	0.0385	-0.0214
	sd	0.1177	0.0388	0.1347	0.1246
	cp(%)	96.0	94.0	94.0	94.5

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.

(N, m, n_i)		η	τ	β_1	β_2
$\tau = 0.3$					
(50, 5, 10)	bias	-0.1181	0.0112	0.0541	0.1012
	sd	0.3187	0.1108	0.2553	0.4445
	cp(%)	96.5	94.5	93.5	98.5
(100, 5, 20)	bias	-0.0751	0.0135	-0.0192	0.0319
	sd	0.2785	0.0846	0.1657	0.2039
	cp(%)	95.0	96.0	95.5	95.0
(100, 10, 10)	bias	-0.0540	0.0205	-0.0149	0.0183
	sd	0.2242	0.0777	0.1835	0.2161
	cp(%)	95.5	94.5	95.5	94.0
(200, 10, 20)	bias	-0.0334	0.0151	-0.0086	0.0114
	sd	0.2051	0.0594	0.1166	0.1344
	cp(%)	97.0	93.5	96.5	96.0
$\tau = 0.6$					
(50, 5, 10)	bias	-0.0998	-0.0128	0.0738	0.0379
	sd	0.3216	0.0841	0.3082	0.4408
	cp(%)	96.5	95.5	94.0	98.5
(100, 5, 20)	bias	-0.0997	-0.0012	0.0418	-0.0164
	sd	0.2264	0.0570	0.1860	0.2021
	cp(%)	98.0	95.0	94.5	94.0
(100, 10, 10)	bias	-0.0620	-0.0018	0.0411	-0.0153
	sd	0.2341	0.0565	0.1889	0.1787
	cp(%)	97.0	95.5	95.5	95.5
(200, 10, 20)	bias	-0.0763	-0.0062	0.0374	-0.0286
	sd	0.1796	0.0386	0.1304	0.1355
	cp(%)	95.5	93.5	94.5	94.5

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.

(N, m, n_i)		η	τ	β_1	β_2
$\tau = 0.3$					
(50, 5, 10)	bias	0.0243	0.0131	0.0117	0.0614
	sd	0.2091	0.1130	0.2506	0.3246
	cp(%)	94.5	95.5	95.5	97.5
(100, 5, 20)	bias	-0.0004	0.0227	0.0031	0.0251
	sd	0.1535	0.0732	0.1732	0.2044
	cp(%)	96.0	95.5	94.0	95.0
(100, 10, 10)	bias	0.0113	0.0104	0.0153	0.0337
	sd	0.1484	0.0883	0.1878	0.1989
	cp(%)	95.0	96.5	96.0	95.0
(200, 10, 20)	bias	0.0095	0.0051	-0.0076	0.0102
	sd	0.1132	0.0607	0.1275	0.1367
	cp(%)	95.0	96.5	94.0	96.0
$\tau = 0.6$					
(50, 5, 10)	bias	0.0223	-0.0248	0.0980	0.0250
	sd	0.2609	0.1052	0.2948	0.3034
	cp(%)	95.0	95.5	92.0	98.0
(100, 5, 20)	bias	0.0414	-0.0188	0.0720	0.0143
	sd	0.2703	0.0773	0.1971	0.2781
	cp(%)	98.0	97.0	93.5	97.5
(100, 10, 10)	bias	0.0185	-0.0089	0.0606	-0.0041
	sd	0.1811	0.0676	0.1774	0.1982
	cp(%)	95.5	98.0	94.5	95.5
(200, 10, 20)	bias	0.0012	0.0009	0.0624	0.0035
	sd	0.1148	0.0390	0.1213	0.1301
	cp(%)	96.5	94.5	93.5	95.0

Table 4: Simulation results based on 200 replications under the model with normally distributed random effects in the slope in Scenario 4, where parameters $\eta = 0.4$, $\beta_1 = 1$, $\beta_2 = 1$, $\lambda_{01}(t) = 1$, $\lambda_{02}(t) = 0.4$ and Kendall's $\tau = 0.3$ and 0.6.

$\tau = 0.3$					
(N, m, n_i)		η	τ	β_1	β_2
(100, 5, 20)	bias	0.0207	0.0134	0.0313	0.0788
	sd	0.3363	0.1120	0.2616	0.3429
	cp(%)	94.5	94.0	94.5	96.0
	bias	0.0096	0.0071	-0.0042	0.0359
	sd	0.3053	0.0855	0.1804	0.1787
	cp(%)	94.0	95.5	97.0	95.0
	bias	0.0095	0.0076	0.0163	0.0269
	sd	0.2546	0.0862	0.1935	0.1879
	cp(%)	92.0	95.5	95.0	94.5
(200, 10, 20)	bias	-0.0045	0.0081	-0.0093	-0.0051
	sd	0.2141	0.0621	0.1259	0.1368
	cp(%)	95.5	94.5	93.5	96.5
$\tau = 0.6$					
(50, 5, 10)	bias	0.0199	-0.0343	0.0914	0.0743
	sd	0.3347	0.1048	0.3806	0.5534
	cp(%)	94.5	96.5	97.0	99.5
(100, 5, 20)	bias	0.0202	-0.0049	0.0858	0.0208
	sd	0.2761	0.0589	0.2541	0.2088
	cp(%)	97.0	96.5	94.5	93.5
(100, 10, 10)	bias	0.0157	-0.0106	0.0453	-0.0099
	sd	0.2580	0.0722	0.2078	0.2064
	cp(%)	96.5	97.0	95.5	94.5
(200, 10, 20)	bias	0.0112	-0.0107	0.0496	0.0103
	sd	0.2161	0.0594	0.1881	0.1349
	cp(%)	95.0	98.0	96.5	96.5

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 $\beta_2 = 1$, $\lambda_{01}(t) = 1$, $\lambda_{02}(t) = 0.4$ and Kendall's $\tau = 0.3$ and 0.6.

(N, m, n_i)		Proposed method				Emura et al (2015)			
		η	τ	β_1	β_2	η	τ	β_1	β_2
$\tau = 0.3$									
(50, 5, 10)	bias	0.0030	0.0095	0.0301	-0.0034	-0.0494	0.1222	0.0374	0.0695
	sd	0.1918	0.1078	0.2938	0.3007	0.1482	0.1756	0.2490	0.2878
	rmse	0.1919	0.1082	0.2953	0.3007	0.1562	0.2139	0.2518	0.2961
	cp(%)	93.0	96.0	95.0	95.5	96.0	89.0	94.0	93.0
(100, 5, 20)	bias	-0.0267	0.0195	0.0353	0.0530	0.0396	0.0416	0.0308	0.0392
	sd	0.1600	0.0827	0.1680	0.1988	0.2943	0.1691	0.1868	0.1974
	rmse	0.1622	0.0850	0.1717	0.2058	0.2970	0.1741	0.1893	0.2012
	cp(%)	96.5	95.5	94.0	94.0	94.5	97.5	95.0	96.5
(100, 10, 10)	bias	0.0322	0.0114	-0.0027	0.0244	0.0183	0.0583	0.0363	0.0315
	sd	0.1611	0.0815	0.1776	0.2000	0.2045	0.1706	0.1636	0.1867
	rmse	0.1643	0.0823	0.1776	0.2015	0.2053	0.1803	0.1676	0.1894
	cp(%)	94.0	94.5	95.0	97.0	94.5	96.0	94.0	95.0
(200, 10, 20)	bias	0.0133	0.0074	0.0326	0.0384	0.0524	0.0180	-0.0008	0.0011
	sd	0.1517	0.0640	0.1273	0.1396	0.2950	0.1421	0.1485	0.1629
	rmse	0.1523	0.0644	0.1314	0.1448	0.2996	0.1433	0.1485	0.1629
	cp(%)	93.0	93.5	92.5	92.5	94.0	97.5	97.0	99.0
$\tau = 0.6$									
(50, 5, 10)	bias	-0.0407	0.0063	0.0669	-0.0172	-0.0644	-0.0035	0.0621	0.0687
	sd	0.0551	0.1007	0.3336	0.2644	0.1528	0.1664	0.2696	0.2907
	rmse	0.0685	0.1009	0.3402	0.2649	0.1659	0.1664	0.2767	0.2987
	cp(%)	93.0	95.0	94.0	94.5	96.5	95.0	94.5	94.0
(100, 5, 20)	bias	-0.0432	0.0161	0.0081	-0.0606	0.0510	-0.0274	0.0406	0.0476
	sd	0.0742	0.0596	0.1785	0.1816	0.2770	0.1373	0.1647	0.1948
	rmse	0.0859	0.0617	0.1787	0.1915	0.2816	0.1400	0.1696	0.2005
	cp(%)	97.5	93.5	96.5	94.0	92.0	93.5	94.0	95.5
(100, 10, 10)	bias	-0.0195	0.0149	-0.0013	-0.0432	0.0334	-0.0238	0.0663	0.0531
	sd	0.0512	0.0628	0.1843	0.1741	0.2354	0.1285	0.1864	0.2043
	rmse	0.0548	0.0646	0.1843	0.1794	0.2378	0.1307	0.1978	0.2111
	cp(%)	94.5	97.5	95.5	96.5	93.0	95.5	93.0	94.0
(200, 10, 20)	bias	-0.0275	0.0090	-0.0037	-0.0636	0.1210	-0.0361	0.0212	0.0266
	sd	0.0562	0.0387	0.1268	0.1117	0.3332	0.1167	0.1296	0.1464
	rmse	0.0626	0.0397	0.1269	0.1285	0.3545	0.1222	0.1313	0.1488
	cp(%)	94.5	94.5	95.5	93.5	93.5	94.5	94.5	96.5

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 $\beta_2 = 1$, $\lambda_{01}(t) = 1$, $\lambda_{02}(t) = 0.4$ and Kendall's $\tau = 0.3$ and 0.6.

(N, m, n_i)		Proposed method				Emura et al (2015)			
		η	τ	β_1	β_2	η	τ	β_1	β_2
$\tau = 0.3$									
(50, 5, 10)	bias	-0.0229	0.0203	0.0411	0.0234	-0.0320	0.1281	0.0341	0.0396
	sd	0.3617	0.1148	0.2719	0.3161	0.3354	0.1890	0.2842	0.2879
	rmse	0.3624	0.1166	0.2750	0.3170	0.3369	0.2283	0.2862	0.2906
	cp(%)	94.5	94.5	94.5	94.0	95.0	93.5	95.5	93.5
(100, 5, 20)	bias	0.0067	0.0157	0.0181	0.0298	-0.0684	0.0672	0.0181	0.0208
	sd	0.4070	0.0827	0.1878	0.2131	0.2573	0.1719	0.1550	0.1867
	rmse	0.4071	0.0842	0.1887	0.2152	0.2662	0.1846	0.1561	0.1878
	cp(%)	94.5	94.5	91.5	95.0	97.0	94.5	93.5	94.5
(100, 10, 10)	bias	0.0821	0.0246	0.0255	0.0388	-0.0156	0.0478	0.0325	0.0130
	sd	0.3455	0.0910	0.1863	0.2172	0.2350	0.1736	0.1823	0.1887
	rmse	0.3551	0.0942	0.1880	0.2207	0.2355	0.1800	0.1852	0.1892
	cp(%)	93.5	95.0	94.5	96.0	95.0	97.5	93.0	93.0
(200, 10, 20)	bias	0.0633	0.0146	0.0059	0.0131	-0.0155	-0.0001	0.0061	0.0122
	sd	0.3089	0.0660	0.1218	0.1377	0.2143	0.1406	0.1205	0.1233
	rmse	0.3154	0.0675	0.1219	0.1383	0.2149	0.1406	0.1206	0.1239
	cp(%)	94.0	96.0	96.0	95.0	97.0	95.5	97.5	92.5
$\tau = 0.6$									
(50, 5, 10)	bias	-0.0577	0.0257	-0.0059	-0.0644	-0.0212	-0.0019	0.0709	0.0826
	sd	0.1917	0.0772	0.2401	0.2645	0.3841	0.1563	0.3018	0.3248
	rmse	0.2002	0.0814	0.2402	0.2722	0.3846	0.1563	0.3100	0.3351
	cp(%)	95.5	95.5	97.0	95.5	96.0	95.0	92.5	95.0
(100, 5, 20)	bias	-0.0949	0.0163	-0.0196	-0.0661	-0.0258	-0.0129	0.0301	0.0404
	sd	0.2010	0.0506	0.1620	0.1952	0.3379	0.1294	0.1810	0.2054
	rmse	0.2222	0.0532	0.1632	0.2061	0.3389	0.1300	0.1835	0.2093
	cp(%)	97.0	94.0	96.0	94.5	95.0	94.0	93.0	95.5
(100, 10, 10)	bias	-0.0073	0.0245	-0.0509	-0.0890	-0.0050	-0.0086	0.0450	0.0624
	sd	0.1393	0.0525	0.1910	0.2091	0.2562	0.1351	0.1891	0.2028
	rmse	0.1395	0.0579	0.1977	0.2273	0.2563	0.1353	0.1944	0.2122
	cp(%)	96.5	92.0	95.0	93.5	96.0	95.0	96.0	94.0
(200, 10, 20)	bias	-0.0226	-0.0009	0.0254	-0.0313	-0.0111	-0.0172	0.0134	0.0213
	sd	0.1679	0.0391	0.1215	0.1331	0.2393	0.1029	0.1162	0.1196
	rmse	0.1695	0.0391	0.1241	0.1367	0.2395	0.1043	0.1170	0.1215
	cp(%)	97.5	95.0	94.5	94.0	96.0	94.5	94.5	96.0

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.

(N, m, n_i)		η	τ	β_1	β_2
$\tau = 0.3$					
(50, 5, 10)	bias	-0.0434	0.0233	0.0439	0.0456
	sd	0.2035	0.1174	0.3131	0.3053
	cp(%)	97.0	95.5	93.5	93.5
(100, 10, 10)	bias	-0.0467	0.0191	0.0065	0.0400
	sd	0.1755	0.0897	0.1750	0.1967
	cp(%)	97.5	94.0	93.5	94.5
(100, 5, 20)	bias	-0.0125	0.0158	0.0258	0.0431
	sd	0.1796	0.0921	0.1971	0.2055
	cp(%)	95.0	95.5	95.5	96.0
(200, 10, 20)	bias	-0.0080	0.0174	-0.0120	0.0020
	sd	0.1412	0.0578	0.1162	0.1397
	cp(%)	95.5	95.0	95.0	94.0
$\tau = 0.6$					
(50, 5, 10)	bias	-0.0490	0.0128	0.0673	0.0439
	sd	0.1758	0.0836	0.3322	0.2857
	cp(%)	96.5	97.0	95.5	94.0
(100, 5, 20)	bias	-0.0327	0.0011	0.0522	-0.0110
	sd	0.1935	0.0581	0.1783	0.1718
	cp(%)	95.5	96.5	91.5	96.0
(100, 10, 10)	bias	-0.0226	-0.0097	0.0638	-0.0133
	sd	0.1602	0.0607	0.1892	0.1827
	cp(%)	95.5	96.0	94.5	95.0
(200, 10, 20)	bias	-0.0254	0.0089	0.0310	-0.0321
	sd	0.1198	0.0374	0.1088	0.1229
	cp(%)	98.0	96.0	92.5	94.5

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 .

(N, m, n_i)		Proposed method				Emura et al (2015)			
		η	τ	β_1	β_2	η	τ	β_1	β_2
$\tau = 0.3$									
(50, 5, 10)	bias	0.0807	0.0062	0.0380	0.1176	3.2862	0.6091	10.1897	7.4917
	sd	0.3997	0.1122	0.2528	0.4747	4.0565	0.1645	88.6486	53.0448
	rmse	0.4078	0.1124	0.2556	0.4890	5.2206	0.6309	89.2323	53.5713
	cp(%)	94.5	94.5	94.5	98.0	93.5	5.0	100.0	98.5
(100, 5, 20)	bias	-0.0197	0.0191	0.0108	0.0574	6.2254	0.6489	7.7824	6.5380
	sd	0.2419	0.0771	0.1819	0.2071	8.9309	0.1044	46.6844	31.4339
	rmse	0.2427	0.0795	0.1822	0.2149	10.8865	0.6572	47.3287	32.1066
	cp(%)	94.5	95.5	95.0	94.0	95.0	0.0	99.0	99.0
(100, 10, 10)	bias	-0.0052	0.0123	0.0149	0.0367	4.5206	0.6618	5.0968	5.4785
	sd	0.2263	0.0804	0.1912	0.1997	4.6840	0.0626	60.8759	61.0783
	rmse	0.2264	0.0813	0.1918	0.2031	6.5097	0.6647	61.0889	61.3235
	cp(%)	96.0	96.5	95.0	93.5	95.0	0.0	97.0	97.0
(200, 10, 20)	bias	-0.0338	0.0120	0.0183	0.0110	6.0779	0.6819	10.1392	9.9668
	sd	0.1611	0.0622	0.1304	0.1486	4.8037	0.0276	36.2865	36.0600
	rmse	0.1646	0.0633	0.1316	0.1491	7.7470	0.6824	37.6764	37.4120
	cp(%)	97.0	95.5	93.0	95.5	87.0	0.0	98.0	98.0
$\tau = 0.6$									
(50, 5, 10)	bias	-0.0133	-0.0202	0.0965	0.0283	4.3264	0.3464	8.2619	10.5323
	sd	0.3059	0.1135	0.7807	0.3358	7.7003	0.1153	59.3232	73.3230
	rmse	0.3062	0.1153	0.7866	0.3370	8.8325	0.3651	59.8957	74.0756
	cp(%)	95.5	96.5	97.5	96.0	97.0	0.0	96.0	96.0
(100, 5, 20)	bias	-0.0413	-0.0041	0.0642	0.0192	5.0458	0.3418	13.7960	11.7703
	sd	0.2271	0.0584	0.1788	0.1913	6.2825	0.1269	108.7297	104.3235
	rmse	0.2309	0.0585	0.1900	0.1923	8.0580	0.3646	109.6015	104.9854
	cp(%)	96.5	96.0	93.0	94.5	95.0	0.0	97.0	98.0
(100, 10, 10)	bias	-0.0375	-0.0017	0.1011	-0.0074	4.1429	0.3773	30.7685	28.0709
	sd	0.2050	0.0737	0.5055	0.2066	3.5439	0.0362	123.1173	113.0520
	rmse	0.2084	0.0737	0.5155	0.2067	5.4519	0.3790	126.9038	116.4849
	cp(%)	96.0	98.5	99.0	96.5	79.0	0.0	96.5	96.5
(200, 10, 20)	bias	-0.0642	-0.0001	0.0444	-0.0195	6.5153	0.3880	17.8160	18.2816
	sd	0.1365	0.0391	0.1338	0.1292	4.7287	0.0197	79.8196	79.7025
	rmse	0.1509	0.0391	0.1409	0.1306	8.0504	0.3885	81.7837	81.7723
	cp(%)	96.0	94.5	93.5	95.5	89.0	0.0	98.0	98.0

Table 9: Analysis of the breast cancer data

parameter	proposed estimates ($\hat{se}$)	naive marginal estimate (sd)	cox estimate (sd)
regression for DMFS			
er	-0.2817(0.1388)*	-0.3299(0.1324)*	-0.0821(0.1275)
size	0.2081(0.0501)*	0.1980(0.0511)*	0.0836(0.0458)
node	0.1324(0.1055)	0.2831(0.1441)	0.5521(0.1237)*
age	-0.0111(0.0057)	-0.0150(0.0059)*	-0.0314(0.0053)*
grade 1	-0.4499(0.1866)*	-0.6317(0.2257)*	-0.4992(0.2153)*
grade 2	-0.1131(0.1190)	-0.0929(0.1319)	-0.1228(0.1276)
		$\eta = 0.5985$	
regression for OS			
er	-0.6003(0.1594)*	-0.6072(0.1263)*	-0.5533(0.1260)*
size	0.2510(0.0560)*	0.2348(0.0477)*	0.2134(0.0454)*
node	0.2090(0.1276)	0.2026(0.1351)	0.7601(0.1261)*
age	0.0082(0.0047)	0.0104(0.0052)*	0.0098(0.0046)*
grade 1	-0.6054(0.1968)*	-0.6219(0.2149)*	-0.5125(0.2147)*
grade 2	-0.2419(0.1548)	-0.2670(0.1302)*	-0.2615(0.1314)
		$\eta = 0.3074$	
τ	0.5974(0.0867)*	-	-
η	0.3787(0.0960)*	-	-

(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 1: Simulation results for estimated cumulative baseline hazard function $\Lambda_{01}(\cdot)$ by the proposed method based on 20 replications (colored red) with true cumulative baseline hazard function $\Lambda_{01}(t) = t$ (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 $\Lambda_{01}(\cdot)$ by Emura et al. (2015)'s method based on 20 replications (colored red) with true cumulative baseline hazard function $\Lambda_{01}(t) = t$ (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 $\tau = 0.3$.

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 ω_0 . Let

$$\begin{aligned}\bar{\Lambda}_{m1}(t) &= \sum_{i,j} \frac{\delta_{ij}^{(1)}(1 - Y_{ij}(t))}{\sum_{g,h} Y_{gh}(T_{ij}) e^{\beta'_{01} X_{gh}} E_{\omega} \{w_1(T_{gh}) e^{b'_g Z_{gh}}\}}, \\ \bar{\Lambda}_{m2}(t) &= \sum_{i,j} \frac{(\delta_{ij}^{(1)} \tilde{\delta}_{ij}^{(2)} + \delta_{ij}^{(2)})(1 - \tilde{Y}_{ij}(t))}{\sum_{g,h} \tilde{Y}_{gh}(\tilde{T}_{ij}) e^{\beta'_{02} X_{gh}} E_{\omega} \{w_2(\tilde{T}_{gh}) e^{b'_g Z_{gh}}\}}, \\ a_i(t) &= n_i^{-1} \sum_{j=1}^{n_i} \int_{u=0}^t \left\{ dN_{ij}^{(1)}(u) - Y_{ij}(u) e^{\beta'_{01} X_{ij}} E_{\omega} \{w_1(T_{ij}) e^{b'_i Z_{ij}}\} d\Lambda_{01}(u) \right\}, \\ b_i(t) &= n_i^{-1} \sum_{j=1}^{n_i} \int_{u=0}^t \left\{ dN_{ij}^{(1)}(u) d\tilde{N}_{ij}^{(2)}(u) + dN_{ij}^{(2)}(u) - \tilde{Y}_{ij}(u) e^{\beta'_{02} X_{ij}} E_{\omega} \{w_2(\tilde{T}_{ij}) e^{b'_i Z_{ij}}\} d\Lambda_{02}(u) \right\}, \\ f_{m1}(u) &= m^{-1} \sum_{i=1}^m n_i^{-1} \sum_{j=1}^{n_i} Y_{ij}(u) e^{\beta'_{01} X_{ij}} E_{\omega} \{w_1(T_{ij}) e^{b'_i Z_{ij}}\}, \\ f_{m2}(u) &= m^{-1} \sum_{i=1}^m n_i^{-1} \sum_{j=1}^{n_i} \tilde{Y}_{ij}(u) e^{\beta'_{02} X_{ij}} E_{\omega} \{w_2(\tilde{T}_{ij}) e^{b'_i Z_{ij}}\}.\end{aligned}$$

We show $\sup_{t \in [0, \varsigma]} |\bar{\Lambda}_{m1}(t) - \Lambda_{01}(t)| \rightarrow 0$ and $\sup_{t \in [0, \varsigma]} |\bar{\Lambda}_{m2}(t) - \Lambda_{02}(t)| \rightarrow 0$ almost surely. Note that $\{a_i(t) : i = 1, 2, \dots\}$ and $\{b_i(t) : i = 1, 2, \dots\}$ are two mean zero independent sequences for fixed t . By the strong law of large numbers (SLLN), $m^{-1} \sum_i a_i(t) \rightarrow 0$ almost surely (a.s.) and $m^{-1} \sum_i b_i(t) \rightarrow 0$ a.s., respectively. Given that $S_1(T_{ij})$, $S_2(\tilde{T}_{ij})$ and $c\{S_1(T_{ij}), S_2(\tilde{T}_{ij}); \alpha\}$ are bounded above 0, by the bounded assumption on X_{ij} and Z_{ij} we have $E_{\omega} \{w_1(T_{ij}) e^{b'_i Z_{ij}}\}$, $E_{\omega} \{w_2(\tilde{T}_{ij}) e^{b'_i Z_{ij}}\}$, $e^{\beta'_{01} X_{ij}}$ and $e^{\beta'_{02} X_{ij}}$ are all bounded. Applying the SLLN again, we have $f_{m1}(u) \rightarrow E_{\omega} [Y_{ij}(u) e^{\beta'_{01} X_{ij}} E_{\omega} \{w_1(T_{ij}) e^{b'_i Z_{ij}}\}]$ a.s. and

$f_{m2}(u) \rightarrow E_\omega \left[\tilde{Y}_{ij}(u) e^{\beta'_{02} X_{ij}} E_\omega \{ w_2(\tilde{T}_{ij}) e^{b'_i Z_{ij}} \} \right]$ a.s.. Note that

$$\begin{aligned} & \sum_{i,j} \int_{u=0}^t \left\{ dN_{ij}^{(1)}(u) - Y_{ij}(u) e^{\beta'_{01} X_{ij}} E_\omega \{ w_1(T_{ij}) e^{b'_i Z_{ij}} \} d\bar{\Lambda}_{n1}(u) \right\} \\ &= \sum_{i,j} \left\{ \delta_{ij}^{(1)} (1 - Y_{ij}(t)) - \sum_{g,h} \frac{Y_{ij}(T_{gh}) e^{\beta'_{01} X_{ij}} E_\omega \{ w_1(T_{ij}) e^{b'_i Z_{ij}} \} \delta_{gh}^{(1)} (1 - Y_{gh}(t))}{\sum_{s,r} Y_{sr}(T_{gh}) e^{\beta'_{01} X_{sr}} E_\omega \{ w_1(T_{sr}) e^{b'_s Z_{sr}} \}} \right\} \\ &= \sum_{i,j} \delta_{ij}^{(1)} (1 - Y_{ij}(t)) - \sum_{g,h} \left\{ \frac{\sum_{i,j} Y_{ij}(T_{gh}) e^{\beta'_{01} X_{ij}} E_\omega \{ w_1(T_{ij}) e^{b'_i Z_{ij}} \} \delta_{gh}^{(1)} (1 - Y_{gh}(t))}{\sum_{s,r} Y_{sr}(T_{gh}) e^{\beta'_{01} X_{sr}} E_\omega \{ w_1(T_{sr}) e^{b'_s Z_{sr}} \}} \right\} = 0. \end{aligned}$$

$$\begin{aligned} & \sum_{i,j} \int_{u=0}^t \left\{ dN_{ij}^{(1)}(u) d\tilde{N}_{ij}^{(2)}(u) + dN_{ij}^{(2)}(u) - \tilde{Y}_{ij}(u) e^{\beta'_{02} X_{ij}} E_\omega \{ w_2(\tilde{T}_{ij}) e^{b'_i Z_{ij}} \} d\bar{\Lambda}_{n2}(u) \right\} \\ &= \sum_{i,j} \left\{ (\delta_{ij}^{(1)} \tilde{\delta}_{ij}^{(2)} + \delta_{ij}^{(2)}) (1 - \tilde{Y}_{ij}(t)) - \sum_{g,h} \frac{\tilde{Y}_{ij}(\tilde{T}_{gh}) e^{\beta'_{02} X_{ij}} E_\omega \{ w_2(\tilde{T}_{ij}) e^{b'_i Z_{ij}} \} (\delta_{gh}^{(1)} \tilde{\delta}_{gh}^{(2)} + \delta_{gh}^{(2)}) (1 - \tilde{Y}_{gh}(t))}{\sum_{s,r} \tilde{Y}_{sr}(\tilde{T}_{gh}) e^{\beta'_{02} X_{sr}} E_\omega \{ w_2(\tilde{T}_{sr}) e^{b'_s Z_{sr}} \}} \right\} \\ &= \sum_{i,j} (\delta_{ij}^{(1)} \tilde{\delta}_{ij}^{(2)} + \delta_{ij}^{(2)}) (1 - \tilde{Y}_{ij}(t)) - \sum_{g,h} \left\{ \frac{\sum_{i,j} \tilde{Y}_{ij}(\tilde{T}_{gh}) e^{\beta'_{02} X_{ij}} E_\omega \{ w_2(\tilde{T}_{ij}) e^{b'_i Z_{ij}} \} (\delta_{gh}^{(1)} \tilde{\delta}_{gh}^{(2)} + \delta_{gh}^{(2)}) (1 - \tilde{Y}_{gh}(t))}{\sum_{s,r} \tilde{Y}_{sr}(\tilde{T}_{gh}) e^{\beta'_{02} X_{sr}} E_\omega \{ w_2(\tilde{T}_{sr}) e^{b'_s Z_{sr}} \}} \right\} \\ &= 0. \end{aligned}$$

For fixed t , by the *SLLN* we have

$$\begin{aligned} \int_{u=0}^t f_{m1}(u) d(\bar{\Lambda}_{m1} - \Lambda_{01})(u) &= m^{-1} \sum_{i,j} n_i^{-1} \int_{u=0}^t \left\{ dN_{ij}^{(1)}(u) - Y_{ij}(u) e^{\beta'_{01} X_{ij}} E_\omega \{ w_1(T_{ij}) e^{b'_i Z_{ij}} \} d\Lambda_{01}(u) \right\} \\ &\quad - m^{-1} \sum_{i,j} n_i^{-1} \int_{u=0}^t \left\{ dN_{ij}^{(1)}(u) - Y_{ij}(u) e^{\beta'_{01} X_{ij}} E_\omega \{ w_1(T_{ij}) e^{b'_i Z_{ij}} \} d\bar{\Lambda}_{m1}(u) \right\} \\ &= n^{-1} \sum_{i=1}^n a_i(t) \rightarrow 0 \text{ a.s. , and} \end{aligned}$$

$$\begin{aligned} & \int_{u=0}^t f_{m2}(u) d(\bar{\Lambda}_{m2} - \Lambda_{02})(u) \\ &= m^{-1} \sum_{i,j} n_i^{-1} \int_{u=0}^t \left\{ dN_{ij}^{(1)}(u) d\tilde{N}_{ij}^{(2)}(u) + dN_{ij}^{(2)}(u) - \tilde{Y}_{ij}(u) e^{\beta'_{02} X_{ij}} E_\omega \{ w_2(\tilde{T}_{ij}) e^{b'_i Z_{ij}} \} d\Lambda_{02}(u) \right\} \\ &\quad - m^{-1} \sum_{i,j} n_i^{-1} \int_{u=0}^t \left\{ dN_{ij}^{(1)}(u) d\tilde{N}_{ij}^{(2)}(u) + dN_{ij}^{(2)}(u) - \tilde{Y}_{ij}(u) e^{\beta'_{02} X_{ij}} E_\omega \{ w_2(\tilde{T}_{ij}) e^{b'_i Z_{ij}} \} d\bar{\Lambda}_{m2}(u) \right\} \\ &= m^{-1} \sum_{i=1}^n b_i(t) \rightarrow 0 \text{ a.s..} \end{aligned}$$

Since $\lim_{m \rightarrow \infty} f_{m1}(u) = E[Y_{ij}(u) e^{\beta'_{01} X_{ij}} E_\omega \{ w_1(T_{ij}) e^{b'_i Z_{ij}} \}]$, 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 $c_1 = \sup_{u \in [0, s]} c_1(u)$. For sufficiently

large n , we can write

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

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

Since $\hat{\alpha}_m, \hat{\beta}_{m1}, \hat{\beta}_{m2}, \hat{\Sigma}_m, X_{gh}, Z_{gh}$ are in compact sets and $w_1(T_{gh})$ and $w_2(\tilde{T}_{gh})$ are bounded, there exists some finite and possibly negative constants c_2 and c_3 such that

$$\hat{\beta}'_{m1} X_{gh} + \log E_{\hat{\omega}_m} \{w_1(T_{gh}) e^{b'_g Z_{gh}}\} \geq \hat{\beta}'_{01} X_{gh} + \log E_{\omega_0} \{w_1(T_{gh}) e^{b'_g Z_{gh}}\} + c_2,$$

$$\hat{\beta}'_{m2} X_{gh} + \log E_{\hat{\omega}_m} \{w_2(\tilde{T}_{gh}) e^{b'_g Z_{gh}}\} \geq \hat{\beta}'_{02} X_{gh} + \log E_{\omega_0} \{w_2(\tilde{T}_{gh}) e^{b'_g Z_{gh}}\} + c_3.$$

Therefore

$$\begin{aligned} \hat{\Lambda}_{m1}(\varsigma) &= \sum_{i,j} \frac{\delta_{ij}^{(1)}(1 - Y_{ij}(\varsigma))}{Y_{gh}(T_{ij}) \exp[\hat{\beta}'_{m1} X_{gh} + \log E_{\hat{\omega}_m} \{w_1(T_{gh}) e^{b'_g Z_{gh}}\}]} \\ &\leq \sum_{i,j} \frac{\delta_{ij}^{(1)}(1 - Y_{ij}(\varsigma))}{Y_{gh}(T_{ij}) \exp[\hat{\beta}'_{01} X_{gh} + \log E_{\omega_0} \{w_1(T_{gh}) e^{b'_g Z_{gh}}\} + c_2]} = e^{-c_2} \bar{\Lambda}_{m1}(\varsigma) \rightarrow e^{-c_2} \Lambda_{01}(\varsigma), \end{aligned}$$

$$\begin{aligned} \hat{\Lambda}_{m2}(\varsigma) &= \sum_{ij} \frac{(\delta_{ij}^{(1)} \tilde{\delta}_{ij}^{(2)} + \delta_{ij}^{(2)})(1 - \tilde{Y}_{ij}(\varsigma))}{\tilde{Y}_{gh}(\tilde{T}_{ij}) \exp[\hat{\beta}'_{m2} X_{gh} + \log E_{\hat{\omega}_m} \{w_2(\tilde{T}_{gh}) e^{b'_g Z_{gh}}\}]} \\ &\leq \sum_{ij} \frac{(\delta_{ij}^{(1)} \tilde{\delta}_{ij}^{(2)} + \delta_{ij}^{(2)})(1 - \tilde{Y}_{ij}(\varsigma))}{\tilde{Y}_{gh}(\tilde{T}_{ij}) \exp[\hat{\beta}'_{02} X_{gh} + \log E_{\omega_0} \{w_2(\tilde{T}_{gh}) e^{b'_g Z_{gh}}\} + c_3]} = e^{-c_3} \bar{\Lambda}_{m2}(\varsigma) \rightarrow e^{-c_3} \Lambda_{02}(\varsigma). \end{aligned}$$

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 $\hat{\Sigma}_m$ are in compact sets. Now we can apply Helly's selection theorem to infer the existence of a convergent subsequence of $\hat{\omega}_m = (\hat{\alpha}_m, \hat{\beta}_m, \hat{\Lambda}_m, \hat{\Sigma}_m)$ having limit ω^* , where $\hat{\beta}_m = (\hat{\beta}_{m1}, \hat{\beta}_{m2})$ and $\hat{\Lambda}_m = (\hat{\Lambda}_{m1}, \hat{\Lambda}_{m2})$.

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

$$\hat{\Lambda}_{m1}(t) = \int_0^t \frac{\sum_{g,h} Y_{gh}(u) \exp[\hat{\beta}'_{01} X_{gh} + \log E_{\omega_0} \{w_1(T_{gh}) e^{b'_g Z_{gh}}\}]}{\sum_{g,h} Y_{gh}(u) \exp[\hat{\beta}'_{m1} X_{gh} + \log E_{\hat{\omega}_m} \{w_1(T_{gh}) e^{b'_g Z_{gh}}\}]} d\bar{\Lambda}_{m1}(u) \rightarrow \Lambda_{01}^*(t), \quad (15)$$

$$\hat{\Lambda}_{m2}(t) = \int_0^t \frac{\sum_{g,h} \tilde{Y}_{gh}(u) \exp[\hat{\beta}'_{02} X_{gh} + \log E_{\omega_0} \{w_2(\tilde{T}_{gh}) e^{b'_g Z_{gh}}\}]}{\sum_{g,h} \tilde{Y}_{gh}(u) \exp[\hat{\beta}'_{m2} X_{gh} + \log E_{\hat{\omega}_m} \{w_2(\tilde{T}_{gh}) e^{b'_g Z_{gh}}\}]} d\bar{\Lambda}_{m2}(u) \rightarrow \Lambda_{02}^*(t), \quad (16)$$

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 $d\hat{\Lambda}_m(t)/d\bar{\Lambda}_m(t)$ 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 λ_{01} and λ_{02} are free to go to infinity at any T_{ij} and $\tilde{T}_{ij}$, respectively. We restrict Λ_{01} right continuous with jumps at T_{ij} and Λ_{02} to be right continuous with jumps at $\tilde{T}_{ij}$. We then have the log-likelihood based on the observed data

$$l_m(\alpha, \beta, \Lambda, \Sigma) = \sum_{i=1}^m \log \left\{ \int_{b_i} R_{1i}(\alpha, \beta, \Lambda, b_i) \psi(b_i, \Sigma) db_i \right\} + \sum_{i=1}^m \sum_{j=1}^{n_i} \delta_{ij}^{(1)} \log(\Lambda_1\{T_{ij}\}) + (\delta_{ij}^{(1)} \tilde{\delta}_{ij}^{(2)} + \delta_{ij}^{(2)}) \log(\Lambda_2\{\tilde{T}_{ij}\}). \quad (17)$$

where

$$\begin{aligned} R_{1i}(\alpha, \beta, \Lambda, b) = \exp \left[\sum_{j=1}^{n_i} \left\{ \delta_{ij}^{(1)} \tilde{\delta}_{ij}^{(2)} \log(D_{12}(T_{ij}, \tilde{T}_{ij})) - \delta_{ij}^{(1)} \Lambda_1(T_{ij}) e^{\beta'_1 X_{ij} + b'_i Z_{ij}} + \delta_{ij}^{(1)} (\beta'_1 X_{ij} + b'_i Z_{ij}) \right. \right. \\ \left. \left. + \delta_{ij}^{(1)} (1 - \tilde{\delta}_{ij}^{(2)}) \log(D_1(T_{ij}, \tilde{T}_{ij})) + \delta_{ij}^{(2)} \log(D_2(T_{ij}, \tilde{T}_{ij})) - (\delta_{ij}^{(1)} \tilde{\delta}_{ij}^{(2)} + \delta_{ij}^{(2)}) \Lambda_2(\tilde{T}_{ij}) e^{\beta'_2 X_{ij} + b'_i Z_{ij}} \right. \right. \\ \left. \left. + (\delta_{ij}^{(1)} \tilde{\delta}_{ij}^{(2)} + \delta_{ij}^{(2)}) (\beta'_2 X_{ij} + b'_i Z_{ij}) + (1 - \delta_{ij}^{(1)} - \delta_{ij}^{(2)}) \log \left[c \{S_1(T_{ij}), S_2(\tilde{T}_{ij})\} \right] \right\} \right]. \end{aligned}$$

Note that

$$\begin{aligned} 0 \leq m^{-1} \{l_m(\hat{\alpha}_m, \hat{\beta}_m, \hat{\Lambda}_m, \hat{\Sigma}_m) - l_m(\alpha_0, \beta_0, \bar{\Lambda}_m, \Sigma_0)\} \\ = m^{-1} \sum_{i=1}^m \log \left\{ \int_{b_i} R_{1i}(\hat{\alpha}_m, \hat{\beta}_m, \hat{\Lambda}_m, b_i) \psi(b_i, \hat{\Sigma}_m) db_i \right\} - m^{-1} \sum_{i=1}^m \log \left\{ \int_{b_i} R_{1i}(\alpha_0, \beta_0, \bar{\Lambda}_m, b_i) \psi(b_i, \Sigma_0) db_i \right\} \\ + m^{-1} \sum_{i=1}^m \sum_{j=1}^{n_i} \delta_{ij}^{(1)} \log(\hat{\Lambda}_m\{T_{ij}\}/\bar{\Lambda}_m\{T_{ij}\}) + m^{-1} \sum_{i=1}^m \sum_{j=1}^{n_i} (\delta_{ij}^{(1)} \tilde{\delta}_{ij}^{(2)} + \delta_{ij}^{(2)}) \log(\hat{\Lambda}_{02}\{\tilde{T}_{ij}\}/\bar{\Lambda}_{02}\{\tilde{T}_{ij}\}) \\ \rightarrow E \left[\log \left\{ \left(\int_{b_i} R_{1i}(\alpha^*, \beta^*, \Lambda^*, b_i) \psi(b_i, \Sigma^*) db_i \prod_{j=1}^{n_i} \left\{ \lambda_{01}^*(T_{ij})^{\delta_{ij}^{(1)}} \lambda_{02}^*(\tilde{T}_{ij})^{\delta_{ij}^{(1)} \tilde{\delta}_{ij}^{(2)} + \delta_{ij}^{(2)}} \right\} \right) \right. \right. \\ \left. \left. \cdot \left(\int_{b_i} R_{1i}(\alpha_0, \beta_0, \Lambda_0, b_i) \psi(b_i, \Sigma_0) db_i \prod_{j=1}^{n_i} \left\{ \lambda_{01}(T_{ij})^{\delta_{ij}^{(1)}} \lambda_{02}(\tilde{T}_{ij})^{\delta_{ij}^{(1)} \tilde{\delta}_{ij}^{(2)} + \delta_{ij}^{(2)}} \right\} \right)^{-1} \right\} \right]. \end{aligned}$$

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](#)

$$\begin{aligned} \int_{b_i} R_{1i}(\alpha^*, \beta^*, \Lambda^*, b_i) \psi(b_i, \Sigma^*) db_i \prod_{j=1}^{n_i} \left\{ \lambda_{01}^*(T_{ij})^{\delta_{ij}^{(1)}} \lambda_{02}^*(\tilde{T}_{ij})^{\delta_{ij}^{(1)} \tilde{\delta}_{ij}^{(2)} + \delta_{ij}^{(2)}} \right\} \\ = \int_{b_i} R_{1i}(\alpha_0, \beta_0, \Lambda_0, b_i) \psi(b_i, \Sigma_0) db_i \prod_{j=1}^{n_i} \left\{ \lambda_{01}(T_{ij})^{\delta_{ij}^{(1)}} \lambda_{02}(\tilde{T}_{ij})^{\delta_{ij}^{(1)} \tilde{\delta}_{ij}^{(2)} + \delta_{ij}^{(2)}} \right\}. \end{aligned} \quad (18)$$

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

$$\begin{aligned}
D_{1*}(t^{(1)}, t^{(2)}) &= \frac{\partial c\{S_1^*(t^{(1)}), S_2^*(t^{(2)})\}}{\partial S_1^*(t^{(1)})}, & D_{12*}(t^{(1)}, t^{(2)}) &= \frac{\partial^2 c\{S_1(t^{(1)}), S_2^*(t^{(2)})\}}{\partial S_1(t^{(1)})\partial S_2^*(t^{(2)})}, \\
D_{1*2*}(t^{(1)}, t^{(2)}) &= \frac{\partial^2 c\{S_1^*(t^{(1)}), S_2^*(t^{(2)})\}}{\partial S_1^*(t^{(1)})\partial S_2^*(t^{(2)})}, \\
D_\alpha(t^{(1)}, t^{(2)}) &= \frac{\partial c\{S_1(t^{(1)}), S_2(t^{(2)})\}}{\partial \alpha}, & D_{1\alpha}(t^{(1)}, t^{(2)}) &= \frac{\partial^2 c\{S_1(t^{(1)}), S_2(t^{(2)})\}}{\partial S_1(t^{(1)})\partial \alpha}, \\
D_{2\alpha}(t^{(1)}, t^{(2)}) &= \frac{\partial^2 c\{S_1(t^{(1)}), S_2(t^{(2)})\}}{\partial S_2(t^{(2)})\partial \alpha}, & D_{12\alpha}(t^{(1)}, t^{(2)}) &= \frac{\partial^3 c\{S_1(t^{(1)}), S_2(t^{(2)})\}}{\partial S_1(t^{(1)})\partial S_2(t^{(2)})\partial \alpha}.
\end{aligned} \tag{19}$$

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

$$\begin{aligned}
l_i(\theta, \lambda | \mathcal{O}_i, b_i) &= \prod_{j=1}^h \left\{ D_{12}(0, 0) \lambda_1(0) e^{\beta'_1 X_{ij} + b'_i Z_{ij}} \lambda_2(0) e^{\beta'_2 X_{ij} + b'_i Z_{ij}} \right\} \cdot \prod_{j=h+1}^{n_i} \left\{ D_1(0, \varsigma) \lambda_1(0) e^{\beta'_1 X_{ij} + b'_i Z_{ij}} \right\}^{1 - \tilde{\delta}_{ij}^{(2)}} \\
&\quad \cdot \prod_{j=h+1}^{n_i} \left\{ \int_0^\varsigma D_{12}(0, y) \lambda_{01}(0) e^{\beta'_1 X_{ij} + b'_i Z_{ij}} e^{-\Lambda_2(y) e^{\beta'_2 X_{ij} + b'_i Z_{ij}}} \lambda_2(y) e^{\beta'_2 X_{ij} + b'_i Z_{ij}} dy \right\}^{\tilde{\delta}_{ij}^{(2)}} \\
&= \prod_{j=1}^h \left\{ D_{12}(0, 0) \lambda_1(0) e^{\beta'_1 X_{ij} + b'_i Z_{ij}} \lambda_2(0) e^{\beta'_1 X_{ij} + b'_i Z_{ij}} \right\} \\
&\quad \cdot \prod_{j=h+1}^{n_i} \left[\left\{ (1 - \tilde{\delta}_{ij}^{(2)}) D_1(0, \varsigma) + \tilde{\delta}_{ij}^{(2)} (-D_1(0, \varsigma) + D_1(0, 0)) \right\} \cdot \lambda_1(0) e^{\beta'_1 X_{ij} + b'_i Z_{ij}} \right] \\
&= \{D_{12}(0, 0)\}^h \{D_1(0, 0)\}^{n_i - h} \prod_{j=1}^h \left\{ \lambda_2(0) e^{\beta'_2 X_{ij} + b'_i Z_{ij}} \right\} \cdot \prod_{j=1}^{n_i} \left\{ \lambda_1(0) e^{\beta'_1 X_{ij} + b'_i Z_{ij}} \right\}.
\end{aligned}$$

Equation (18) can be written as

$$\begin{aligned}
&\{D_{1*2*}(0, 0)\}^h \{D_{1*}(0, 0)\}^{n_i - h} \int_b \prod_{j=1}^h \left\{ \lambda_{02}^*(0) e^{\beta_{2*}' X_{ij} + b'_i Z_{ij}} \right\} \prod_{j=1}^{n_i} \left\{ \lambda_{01}^*(0) e^{\beta_{1*}' X_{ij} + b'_i Z_{ij}} \right\} \psi(b, \Sigma^*) db \\
&= \{D_{12}(0, 0)\}^h \{D_1(0, 0)\}^{n_i - h} \int_b \prod_{j=1}^h \left\{ \lambda_{02}(0) e^{\beta'_{02} X_{ij} + b'_i Z_{ij}} \right\} \prod_{j=1}^{n_i} \left\{ \lambda_{01}(0) e^{\beta'_{01} X_{ij} + b'_i Z_{ij}} \right\} \psi(b, \Sigma_0) db.
\end{aligned} \tag{20}$$

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

$$\begin{aligned}
&\{D_{1*}(0, 0)\}^{n_i} \int_{b_i} \prod_{j=1}^{n_i} \left\{ \lambda_{01}^*(0) e^{\beta_{1*}' X_{ij} + b'_i Z_{ij}} \right\} \psi(b_i, \Sigma^*) db_i \\
&= \{D_1(0, 0)\}^{n_i} \int_{b_i} \prod_{j=1}^{n_i} \left\{ \lambda_{01}(0) e^{\beta'_{01} X_{ij} + b'_i Z_{ij}} \right\} \psi(b_i, \Sigma_0) db_i.
\end{aligned} \tag{21}$$

That is equivalent to

$$\begin{aligned} & \{D_{1*}(0,0)\lambda_{01}^*(0)\}^{n_i} e^{\sum_{j=1}^{n_i} \beta_1^{*'} X_{ij}} E_{b_i}(e^{\sum_{j=1}^{n_i} b_i' Z_{ij}} | \Sigma = \Sigma^*) \\ &= \{D_1(0,0)\lambda_{01}(0)\}^{n_i} e^{\sum_{j=1}^{n_i} \beta_{01}' X_{ij}} E_{b_i}(e^{\sum_{j=1}^{n_i} b_i' Z_{ij}} | \Sigma = \Sigma_0). \end{aligned} \quad (22)$$

Since $S_1^*(0) = S_1(0) = S_2^*(0) = S_2(0) = 1$ and $D_1(0,0)$ is function only related to α , let $f(\alpha) = D_1(0,0)$, (8) can be written as

$$\begin{aligned} & n_i \left[(\log(f(\alpha^*)) - \log(f(\alpha_0))) + (\log \lambda_{01}^*(0) - \log \lambda_{01}(0)) \right] + \sum_{j=1}^{n_i} X_{ij}' (\beta_1^* - \beta_{01}) \\ &+ \left\{ E_{b_i}(e^{\sum_{j=1}^{n_i} b_i' Z_{ij}} | \Sigma = \Sigma^*) - E_{b_i}(e^{\sum_{j=1}^{n_i} b_i' Z_{ij}} | \Sigma = \Sigma_0) \right\} = 0, \end{aligned} \quad (23)$$

Specifically, when b_i is assumed to follow a normal distribution with mean 0 and variance Σ_0 , (23) can be written as

$$\begin{aligned} & n_i \left\{ (\log(f(\alpha^*)) - \log(f(\alpha_0))) + (\log \lambda_{01}^*(0) - \log \lambda_{01}(0)) \right\} \\ &+ \sum_{j=1}^{n_i} X_{ij}' (\beta_1^* - \beta_{01}) + \frac{(\sum_{j=1}^{n_i} Z_{ij})' (\Sigma^* - \Sigma_0) (\sum_{j=1}^{n_i} Z_{ij})}{2} = 0, \end{aligned} \quad (24)$$

for all $i = 1, \dots, m$.

By condition 8 in the paper, we have that $(\log(f(\alpha^*)) - \log(f(\alpha_0))) + (\log \lambda_{02}^*(0) - \log \lambda_{02}(0)) = 0$, $\beta_1^* = \beta_{01}$ 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 η_0 , (23) can be rewritten as

$$\{f(\alpha^*)\lambda_{01}^*(0)\}^{n_i} e^{\sum_{j=1}^{n_i} \beta_1^{*'} X_{ij}} g(\eta^*, n_i) = \{f(\alpha_0)\lambda_{01}(0)\}^{n_i} e^{\sum_{j=1}^{n_i} \beta_{01}' X_{ij}} g(\eta_0, n_i), \quad (25)$$

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 $\alpha^* = \alpha_0$, $\lambda_{01}^*(0) = \lambda_{01}(0)$, $\beta_1^* = \beta_{01}$ and $\eta^* = \eta_0$.

It therefore follows from (20) that

$$(D_{12*}(0,0))^h \prod_{j=1}^h \left\{ \lambda_{02}^*(0) e^{\beta_2^{*'} X_{ij} + b_i' Z_{ij}} \right\} = (D_{12}(0,0))^h \prod_{j=1}^h \left\{ \lambda_{02}(0) e^{\beta_{02}' X_{ij} + b_i' Z_{ij}} \right\}. \quad (26)$$

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

$$h \left(\log \lambda_{02}^*(0) - \log \lambda_{02}(0) \right) + (\beta_2^* - \beta_{02})' \sum_{j=1}^h X_{ij} = 0. \quad (27)$$

Setting $h = n_i$, we can obtain that

$$n_i \left(\log \lambda_{02}^*(0) - \log \lambda_{02}(0) \right) + (\beta_2^* - \beta_{02})' \sum_{j=1}^{n_i} X_{ij} = 0, \quad (28)$$

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

Next, we prove $\Lambda_{02}^*(t) = \Lambda_{02}(t)$. Let $\delta_{ij}^{(1)} = 1$ and $T_{ij} = 0$ for all $j = 1, \dots, n_i$. When $j = 1$ and $\tilde{\delta}_{i1}^{(2)} = 1$, we integrate $\tilde{T}_{i1}$ from 0 to t . For $j = 2, \dots, n_i$, we replace $\tilde{T}_{ij}$ with ς if $\tilde{\delta}_{ij}^{(2)} = 0$ and integrate $\tilde{T}_{ij}$ from 0 to ς if $\tilde{\delta}_{ij}^{(2)} = 1$. Summing the two sides of (18) over all possible values of $\tilde{\delta}_{ij}^{(2)}$ for $j > 1$, we obtain that

$$\begin{aligned} & \int_{b_i} \int_0^t D_{1*2*}(0, y) e^{-\Lambda_{02}^*(y) e^{\beta_{02}' X_{ij} + b_i' Z_{ij}}} \lambda_{02}^*(y) e^{\beta_{02}' X_{ij} + b_i' Z_{ij}} dy \cdot \psi(b_i, \Sigma_0) db_i \\ &= \int_{b_i} \int_0^t D_{12}(0, y) e^{-\Lambda_{02}(y) e^{\beta_{02}' X_{ij} + b_i' Z_{ij}}} \lambda_{02}(y) e^{\beta_{02}' X_{ij} + b_i' Z_{ij}} dy \cdot \psi(b_i, \Sigma_0) db_i, \end{aligned}$$

which is equivalent to

$$\int_{b_i} -c_1 \{S_1^*(0), S_2^*(t)\} \cdot \psi(b_i, \Sigma_0) db_i = \int_{b_i} -c_1 \{S_1^*(0), S_2(t)\} \cdot \psi(b_i, \Sigma_0) db_i.$$

Therefore,

$$c_1 \{S_1^*(0), S_2^*(t)\} = c_1 \{S_1(0)^*, S_2(t)\}. \quad (29)$$

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 $\Lambda_{02}^*(t) = \Lambda_{02}(t)$.

Similarly, we can show that $\Lambda_{01}^*(t) = \Lambda_{01}(t)$. Let $\delta_{ij}^{(2)} = \tilde{\delta}_{ij}^{(2)} = 0$ and $\tilde{T}_{ij} = \varsigma$ for all $j = 1, \dots, n_i$. Choose k from index set $\{1, \dots, n_i\}$. 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 ς if $\delta_{ij}^{(1)} = 0$, otherwise we integrate T_{ij} from 0 to ς .

We choose $k = 1$ and then obtain that

$$\begin{aligned} L_i(\alpha_0, \beta_0, \Lambda_0, \Sigma_0) &= \int_{b_i} \int_0^t D_1(y, \varsigma) e^{-\Lambda_{01}(y) e^{\beta_{01}' X_{ij} + b_i' Z_{ij}}} \lambda_{01}(y) e^{\beta_{01}' X_{ij} + b_i' Z_{ij}} dy \\ &\cdot \prod_{j=2}^{n_i} \left[\left\{ c \{S_1(\varsigma), S_2(\varsigma)\} \right\}^{1-\delta_{ij}^{(1)}} \cdot \left\{ \int_0^\varsigma D_1(y, \varsigma) e^{-\Lambda_{01}(y) e^{\beta_{01}' X_{ij} + b_i' Z_{ij}}} \lambda_{01}(y) e^{\beta_{01}' X_{ij} + b_i' Z_{ij}} dy \right\}^{\delta_{ij}^{(1)}} \right] \\ &= \int_{b_i} \left(\frac{c \{S_1(0), S_2(\varsigma)\} - c \{S_1(t), S_2(\varsigma)\}}{S_2(\varsigma)} \right) \cdot \prod_{j=1}^{n_i} S_2(\varsigma) \psi(b_i, \Sigma_0) db_i. \end{aligned} \quad (30)$$

Following from (18) that

$$c\{S_1^*(0), S_2(\varsigma)\} - c\{S_1^*(t), S_2(\varsigma)\} = c\{S_1(0), S_2(\varsigma)\} - c\{S_1(t), S_2(\varsigma)\}. \quad (31)$$

Since $\alpha^* = \alpha$ and $S_1^*(0) = S_1(0)$, $c\{S_1^*(t), S_2(\varsigma)\} = c\{S_1(t), S_2(\varsigma)\}$ for any copula c . This implies that $S_1^*(t) = S_1(t)$. Therefore, $\Lambda_{01}^*(t) = \Lambda_{01}(t)$ and the proof of Theorem 1 is complete.

Proof of Theorem 2

Let $\mathcal{H} = \{(h_1, h_2, \dots, h_6) : h_1 \in R^1, h_2 \in R^{d_1}, h_3 \in R^{d_2}, (h_4, h_5) \text{ are functions on } [0, \varsigma], h_6 \in R^{d_3(d_3+1)/2} \text{ and } \|h_1\|, \|h_2\|, \|h_3\|, \|h_4\|_v, \|h_5\|_v, \|h_6\| \leq 1\}$, 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 $(\alpha_0, \beta_{01}, \beta_{02}, \Lambda_{01}, \Lambda_{02}, \Sigma_0)$, to $l^\infty(\mathcal{H})$.

$$\begin{aligned} & S_m(\alpha, \beta_1, \beta_2, \Lambda_1, \Lambda_2, \Sigma)[h_1, h_2, h_3, h_4, h_5, h_6] \\ &= m^{-1} \frac{d}{d\epsilon} l_m(\alpha + \epsilon h_1, \beta_1 + \epsilon h_2, \beta_2 + \epsilon_3, \Lambda_1 + \epsilon h_4, \Lambda_2 + \epsilon h_5, \Sigma + \epsilon h_6) \Big|_{\epsilon=0} \\ &=: A_{m_1}[h_1] + A_{m_2}[h_2] + A_{m_3}[h_3] + A_{m_4}[h_4] + A_{m_5}[h_5] + A_{m_6}[h_6], \end{aligned} \quad (32)$$

where we treat Σ as an extended column vector consisting of the upper triangle elements of the covariance matrix. The terms A_{mp} for $p = 1, 2, \dots, 6$ 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 $A_{m_1}[h_1] = \mathcal{P}_m[h_1' l_\alpha]$, $A_{m_2}[h_2] = \mathcal{P}_m[h_2' l_{\beta_1}]$, $A_{m_3}[h_3] = \mathcal{P}_m[h_3' l_{\beta_2}]$, $A_{m_4}[h_4] = \mathcal{P}_m[h_4' l_{\Lambda_1}]$, $A_{m_5}[h_5] = \mathcal{P}_m[h_5' l_{\Lambda_2}]$ and $A_{m_6}[h_6] = \mathcal{P}_m[h_6' l_\Sigma]$, 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

$$d\mu_i(b_i) = \frac{R_{1i}(\omega, b_i) \psi(b_i, \Sigma) db_i}{\int_{b_i} R_{1i}(\omega, b_i) \psi(b_i, \Sigma) db_i}.$$

Then we have

$$\begin{aligned} A_{m_1}[h_1] &= m^{-1} \frac{d}{d\epsilon} l_m(\alpha + \epsilon h_1) \Big|_{\epsilon=0} \\ &= m^{-1} \sum_{i=1}^m \int_{b_i} \sum_{j=1}^{n_i} h_1 \left\{ \delta_{ij}^{(1)} \tilde{\delta}_{ij}^{(2)} \frac{D_{12\alpha}(T_{ij}, \tilde{T}_{ij})}{D_{12}(T_{ij}, \tilde{T}_{ij})} + \delta_{ij}^{(1)} (1 - \tilde{\delta}_{ij}^{(2)}) \frac{D_{1\alpha}(T_{ij}, \tilde{T}_{ij})}{D_1(T_{ij}, \tilde{T}_{ij})} \right. \\ &\quad \left. + \delta_{ij}^{(2)} \frac{D_{2\alpha}(T_{ij}, \tilde{T}_{ij})}{D_2(T_{ij}, \tilde{T}_{ij})} + (1 - \delta_{ij}^{(1)} - \delta_{ij}^{(2)}) \frac{D_\alpha(T_{ij}, \tilde{T}_{ij})}{c\{S_1(T_{ij}), S_2(\tilde{T}_{ij})\}} \right\} d\mu_i(b_i), \end{aligned}$$

$$\begin{aligned}
A_{m_2}[h_2] &= m^{-1} \frac{d}{d\epsilon} l_m(\beta_1 + \epsilon h_2) \Big|_{\epsilon=0} \\
&= m^{-1} \sum_{i=1}^m \int_{b_i} \sum_{j=1}^{n_i} h'_2 X_{ij} \left\{ \delta_{ij}^{(1)} - \Lambda_1(T_{ij}) e^{\beta'_1 X_{ij} + b'_i Z_{ij}} \left(\delta_{ij}^{(1)} \tilde{\delta}_{ij}^{(2)} \frac{D_{121}(T_{ij}, \tilde{T}_{ij})}{D_{12}(T_{ij}, \tilde{T}_{ij})} S_1(T_{ij}) + \delta_{ij}^{(1)} \right. \right. \\
&\quad + \delta_{ij}^{(1)} (1 - \tilde{\delta}_{ij}^{(2)}) \frac{D_{11}(T_{ij}, \tilde{T}_{ij})}{D_1(T_{ij}, \tilde{T}_{ij})} S_1(T_{ij}) + \delta_{ij}^{(2)} \frac{D_{21}(T_{ij}, \tilde{T}_{ij})}{D_2(T_{ij}, \tilde{T}_{ij})} S_1(T_{ij}) \\
&\quad \left. \left. + (1 - \delta_{ij}^{(1)} - \delta_{ij}^{(2)}) \frac{D_1(T_{ij}, \tilde{T}_{ij})}{c\{S_1(T_{ij}), S_2(\tilde{T}_{ij})\}} S_1(T_{ij}) \right) \right\} d\mu_i(b_i),
\end{aligned}$$

$$\begin{aligned}
A_{m_3}[h_3] &= m^{-1} \frac{d}{d\epsilon} l_m(\beta_2 + \epsilon h_3) \Big|_{\epsilon=0} \\
&= m^{-1} \sum_{i=1}^m \int_{b_i} \sum_{j=1}^{n_i} h'_3 X_{ij} \left\{ \delta_{ij}^{(1)} \tilde{\delta}_{ij}^{(2)} + \delta_{ij}^{(2)} - \Lambda_2(\tilde{T}_{ij}) e^{\beta'_2 X_{ij} + b'_i Z_{ij}} \left(\delta_{ij}^{(1)} \tilde{\delta}_{ij}^{(2)} \frac{D_{122}(T_{ij}, \tilde{T}_{ij})}{D_{12}(T_{ij}, \tilde{T}_{ij})} S_2(\tilde{T}_{ij}) \right. \right. \\
&\quad + \delta_{ij}^{(1)} \tilde{\delta}_{ij}^{(2)} + \delta_{ij}^{(2)} + \delta_{ij}^{(1)} (1 - \tilde{\delta}_{ij}^{(2)}) \frac{D_{12}(T_{ij}, \tilde{T}_{ij})}{D_1(T_{ij}, \tilde{T}_{ij})} S_2(\tilde{T}_{ij}) + \delta_{ij}^{(2)} \frac{D_{22}(T_{ij}, \tilde{T}_{ij})}{D_2(T_{ij}, \tilde{T}_{ij})} S_2(\tilde{T}_{ij}) \\
&\quad \left. \left. + (1 - \delta_{ij}^{(1)} - \delta_{ij}^{(2)}) \frac{D_2(T_{ij}, \tilde{T}_{ij})}{c\{S_1(T_{ij}), S_2(\tilde{T}_{ij})\}} S_2(\tilde{T}_{ij}) \right) \right\} d\mu_i(b_i),
\end{aligned}$$

$$\begin{aligned}
A_{m_4}[h_4] &= m^{-1} \frac{d}{d\epsilon} l_m(\Lambda_1 + \epsilon h_4) \Big|_{\epsilon=0} \\
&= m^{-1} \sum_{i=1}^m \sum_{j=1}^{n_i} \left\{ \delta_{ij}^{(1)} h_4(T_{ij}) - \int_0^{T_{ij}} h_4(s) d\Lambda_1(s) \int_{b_i} e^{\beta'_1 X_{ij} + b'_i Z_{ij}} \left(\delta_{ij}^{(1)} \tilde{\delta}_{ij}^{(2)} \frac{D_{121}(T_{ij}, \tilde{T}_{ij})}{D_{12}(T_{ij}, \tilde{T}_{ij})} S_1(T_{ij}) \right. \right. \\
&\quad + \delta_{ij}^{(1)} + \delta_{ij}^{(1)} (1 - \tilde{\delta}_{ij}^{(2)}) \frac{D_{11}(T_{ij}, \tilde{T}_{ij})}{D_1(T_{ij}, \tilde{T}_{ij})} S_1(T_{ij}) + \delta_{ij}^{(2)} \frac{D_{21}(T_{ij}, \tilde{T}_{ij})}{D_2(T_{ij}, \tilde{T}_{ij})} S_1(T_{ij}) \\
&\quad \left. \left. + (1 - \delta_{ij}^{(1)} - \delta_{ij}^{(2)}) \frac{D_1(T_{ij}, \tilde{T}_{ij})}{c\{S_1(T_{ij}), S_2(\tilde{T}_{ij})\}} S_1(T_{ij}) \right) d\mu_i(b_i) \right\},
\end{aligned}$$

$$\begin{aligned}
A_{m_5}[h_5] &= m^{-1} \frac{d}{d\epsilon} l_m(\Lambda_2 + \epsilon h_5) \Big|_{\epsilon=0} \\
&= m^{-1} \sum_{i=1}^m \sum_{j=1}^{n_i} \left\{ \left(\delta_{ij}^{(1)} \tilde{\delta}_{ij}^{(2)} + \delta_{ij}^{(2)} \right) h_5(\tilde{T}_{ij}) - \int_0^{\tilde{T}_{ij}} h_5(s) d\Lambda_2(s) \int_{b_i} e^{\beta'_2 X_{ij} + b'_i Z_{ij}} \left(\delta_{ij}^{(1)} \tilde{\delta}_{ij}^{(2)} \frac{D_{122}(T_{ij}, \tilde{T}_{ij})}{D_{12}(T_{ij}, \tilde{T}_{ij})} S_2(\tilde{T}_{ij}) \right. \right. \\
&\quad + \delta_{ij}^{(1)} \tilde{\delta}_{ij}^{(2)} + \delta_{ij}^{(2)} + \delta_{ij}^{(1)} (1 - \tilde{\delta}_{ij}^{(2)}) \frac{D_{12}(T_{ij}, \tilde{T}_{ij})}{D_1(T_{ij}, \tilde{T}_{ij})} S_2(\tilde{T}_{ij}) + \delta_{ij}^{(2)} \frac{D_{22}(T_{ij}, \tilde{T}_{ij})}{D_2(T_{ij}, \tilde{T}_{ij})} S_2(\tilde{T}_{ij}) \\
&\quad \left. \left. + (1 - \delta_{ij}^{(1)} - \delta_{ij}^{(2)}) \frac{D_2(T_{ij}, \tilde{T}_{ij})}{c\{S_1(T_{ij}), S_2(\tilde{T}_{ij})\}} S_2(\tilde{T}_{ij}) \right) d\mu_i(b_i) \right\},
\end{aligned}$$

$$A_{m_6}[h_6] = m^{-1} \frac{d}{d\epsilon} l_m(\Sigma + \epsilon h_6) \Big|_{\epsilon=0} = m^{-1} \sum_{i=1}^m \int_{b_i} \frac{d}{d\epsilon} \psi(b_i, \Sigma + \epsilon h_6) \Big|_{\epsilon=0} d\mu_i(b_i).$$

Specifically, when b_i is assumed to follow a normal distribution with mean 0 and variance Σ ,

$$A_{m_6}[h_6] = m^{-1} \sum_{i=1}^m \int_{b_i} \{b_i' \Sigma^{-1} H_6 \Sigma^{-1} b_i / 2 - \text{trace}(\Sigma^{-1} H_6) / 2\} d\mu_i(b_i).$$

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

$$A_{m_6}[h_6] = m^{-1} \sum_{i=1}^m \int_{u_i} \left\{ -\frac{1}{u_i} \left(\frac{1}{\eta} - 1 \right) - u_i - 1 + \log(\eta) + \varphi_0\left(\frac{1}{\eta}\right) \right\} \cdot \frac{h_6}{\eta^2} d\mu_i(u_i),$$

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 $[0, \varsigma]$, which has been justified in the proof of Theorem 1, we have A_{mp} for $p = 1, 2, \dots, 6$ all have bounded derivations on $[0, \varsigma]$. Therefore, A_{mp} are all bounded Lipschitz functions.

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

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

where the linear functionals A_p for $p = 1, \dots, 6$ are defined similar to A_{mp} , but replacing the empirical sum with the expectation. By definition, $S_m(\hat{\alpha}_m, \hat{\beta}_{m1}, \hat{\beta}_{m2}, \hat{\Lambda}_{m1}, \hat{\Lambda}_{m2}, \hat{\Sigma}_m) = 0$ and $S(\alpha_0, \beta_{01}, \beta_{02}, \Lambda_{01}, \Lambda_{02}, \Sigma_0) = 0$. Thus $\sqrt{m}(S_m(\alpha_0, \beta_{01}, \beta_{02}, \Lambda_{01}, \Lambda_{02}, \Sigma_0) - S(\alpha_0, \beta_{01}, \beta_{02}, \Lambda_{01}, \Lambda_{02}, \Sigma_0))$ 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

$$\begin{aligned} & \sup_{(h_1, \dots, h_6)} |(S_m - S)(\hat{\alpha}_m, \hat{\beta}_{m1}, \hat{\beta}_{m2}, \hat{\Lambda}_{m1}, \hat{\Lambda}_{m2}, \hat{\Sigma}_m) - (S_m - S)(\alpha_0, \beta_{01}, \beta_{02}, \Lambda_{01}, \Lambda_{02}, \Sigma_0)| \\ &= O_p\left(m^{-1/2} \vee \left\{ \|\hat{\alpha}_m - \alpha_0\| + \|\hat{\beta}_{m1} - \beta_{01}\| + \|\hat{\beta}_{m2} - \beta_{02}\| + \sup_{t \in [0, \varsigma]} |\hat{\Lambda}_{m1}(t) - \Lambda_{01}(t)| \right. \right. \\ & \quad \left. \left. + \sup_{t \in [0, \varsigma]} |\hat{\Lambda}_{m2}(t) - \Lambda_{02}(t)| + \|\hat{\Sigma}_m - \Sigma_0\| \right\}\right). \end{aligned}$$

based on the functional central limit theorem Gamst et al. (2009).

Next, we prove that the derivative, denoted by $\dot{S}(\alpha_0, \beta_{01}, \beta_{02}, \Lambda_{01}, \Lambda_{02}, \Sigma_0)$, is continuously invertible on its range. We consider $\dot{S}(\alpha_0, \beta_{01}, \beta_{02}, \Lambda_{01}, \Lambda_{02}, \Sigma_0)$ as a linear map, simply denoted by T , from the space

$\{(\alpha - \alpha_0, \beta_1 - \beta_{01}, \beta_2 - \beta_{02}, \Lambda_1 - \Lambda_{01}, \Lambda_2 - \Lambda_{02}, \Sigma - \Sigma_0) : (\alpha, \beta_1, \beta_2, \Lambda_1, \Lambda_2, \Sigma) \text{ is in the neighborhood } \mathcal{U} \text{ of } (\alpha_0, \beta_{01}, \beta_{02}, \Lambda_{01}, \Lambda_{02}, \Sigma_0)\}$ to $l^\infty(\mathcal{H})$. That is

$$\begin{aligned} T(\alpha - \alpha_0, \beta_1 - \beta_{01}, \beta_2 - \beta_{02}, \Lambda_1 - \Lambda_{01}, \Lambda_2 - \Lambda_{02}, \Sigma - \Sigma_0) \\ = (\alpha - \alpha_0)'Q_1(h_1, \dots, h_6) + (\beta_1 - \beta_{01})'Q_2(h_1, \dots, h_6) + (\beta_2 - \beta_{02})'Q_3(h_1, \dots, h_6) \\ + \int_0^\Sigma Q_4(h_1, \dots, h_6)d(\Lambda_1 - \Lambda_{01}) + \int_0^\Sigma Q_5(h_1, \dots, h_6)d(\Lambda_2 - \Lambda_{02}) + (\Sigma - \Sigma_0)'Q_6(h_1, \dots, h_6), \end{aligned}$$

where Q_i for $i = 1, \dots, 6$ are the respective partial derivatives of S with respect to $\alpha, \beta_1, \beta_2, \Lambda_1, \Lambda_2$ and Σ . Specifically, Q_i ' have the following forms:

$$\begin{aligned} Q_1(h_1, \dots, h_6) &= B_1(h_1, h_2, h_3, h_6)' + \int_0^\Sigma h_4(t)D_1(t) + \int_0^\Sigma h_5(t)D_2(t), \\ Q_2(h_1, \dots, h_6) &= B_2(h_1, h_2, h_3, h_6)' + \int_0^\Sigma h_4(t)D_3(t) + \int_0^\Sigma h_5(t)D_4(t), \\ Q_3(h_1, \dots, h_6) &= B_3(h_1, h_2, h_3, h_6)' + \int_0^\Sigma h_4(t)D_5(t) + \int_0^\Sigma h_5(t)D_6(t), \\ Q_4(h_1, \dots, h_6) &= B_4(h_1, h_2, h_3, h_6)' + b_4h_4(t) + \int_0^\Sigma h_4(t)D_7(t) + \int_0^\Sigma h_5(t)D_8(t), \\ Q_5(h_1, \dots, h_6) &= B_5(h_1, h_2, h_3, h_6)' + b_5h_5(t) + \int_0^\Sigma h_5(t)D_9(t) + \int_0^\Sigma h_5(t)D_{10}(t), \\ Q_6(h_1, \dots, h_6) &= B_6(h_1, h_2, h_3, h_6)' + \int_0^\Sigma h_4(t)D_{11}(t) + \int_0^\Sigma h_5(t)D_{12}(t), \end{aligned}$$

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

$$\begin{aligned} 0 &= T(\alpha - \alpha_0, \beta_1 - \beta_{01}, \beta_2 - \beta_{02}, \Lambda_1 - \Lambda_{01}, \Lambda_2 - \Lambda_{02}, \Sigma - \Sigma_0)[h_1, \dots, h_6] \\ &= \epsilon E\{(l_{\alpha_0}[h_1] + l_{\beta_{01}}[h_2] + l_{\beta_{02}}[h_3] + l_{\Lambda_{01}}[h_4] + l_{\Lambda_{02}}[h_5] + l_{\Sigma_0}[h_6])^2\}, \end{aligned}$$

where l_{α_0} denotes the score function of α evaluated at α_0 , $l_{\beta_{0k}}$ the score function of β_k evaluated at β_{0k} , $l_{\Lambda_{0k}}$ the score function of Λ_k evaluated at Λ_{0k} , for $k = 1, 2$, and l_{Σ_0} the score function of Σ evaluated at Σ_0 . Then we have

$$l_{\alpha_0}[h_1] + l_{\beta_{01}}[h_2] + l_{\beta_{02}}[h_3] + l_{\Lambda_{01}}[h_4] + l_{\Lambda_{02}}[h_5] + l_{\Sigma_0}[h_6] = 0 \quad (33)$$

almost surely, where

$$R_{2i}(\omega_0, b_i) = R_{1i}(\omega_0, b_i) \prod_{j=1}^{n_i} \{ \lambda_{01}(T_{ij})^{\delta_{ij}^{(1)}} \lambda_{02}(\tilde{T}_{ij})^{\delta_{ij}^{(1)} \tilde{\delta}_{ij}^{(2)} + \delta_{ij}^{(2)}} \}, \quad (34)$$

$$\begin{aligned} l_{\alpha_0}[h_1] &= \sum_{j=1}^{n_i} \int_{b_i} l_{\alpha_0, j}[h_1] \psi(b_i, \Sigma_0) db_i \\ &= \sum_{j=1}^{n_i} \int_{b_i} h_1 \left\{ \delta_{ij}^{(1)} \tilde{\delta}_{ij}^{(2)} \frac{D_{12\alpha}(T_{ij}, \tilde{T}_{ij})}{D_{12}(T_{ij}, \tilde{T}_{ij})} + \delta_{ij}^{(1)} (1 - \tilde{\delta}_{ij}^{(2)}) \frac{D_{1\alpha}(T_{ij}, \tilde{T}_{ij})}{D_1(T_{ij}, \tilde{T}_{ij})} \right. \\ &\quad \left. + \delta_{ij}^{(2)} \frac{D_{2\alpha}(T_{ij}, \tilde{T}_{ij})}{D_2(T_{ij}, \tilde{T}_{ij})} + (1 - \delta_{ij}^{(1)} - \delta_{ij}^{(2)}) \frac{D_{\alpha}(T_{ij}, \tilde{T}_{ij})}{c\{S_1(T_{ij}), S_2(\tilde{T}_{ij})\}} \right\} R_{2i}(\omega_0, b_i) \psi(b_i, \Sigma_0) db_i, \end{aligned} \quad (35)$$

$$\begin{aligned} l_{\beta_{01}}[h_2] &= \sum_{j=1}^{n_i} \int_{b_i} l_{\beta_{01}, j}[h_2] \psi(b_i, \Sigma_0) db_i \\ &= \sum_{j=1}^{n_i} \int_{b_i} h'_2 X_{ij} \left\{ \delta_{ij}^{(1)} - \Lambda_{01}(T_{ij}) e^{\beta'_{01} X_{ij} + b'_1 Z_{ij}} \left(\delta_{ij}^{(1)} \tilde{\delta}_{ij}^{(2)} \frac{D_{121}(T_{ij}, \tilde{T}_{ij})}{D_{12}(T_{ij}, \tilde{T}_{ij})} S_1(T_{ij}) + \delta_{ij}^{(1)} \right. \right. \\ &\quad \left. + \delta_{ij}^{(1)} (1 - \tilde{\delta}_{ij}^{(2)}) \frac{D_{11}(T_{ij}, \tilde{T}_{ij})}{D_1(T_{ij}, \tilde{T}_{ij})} S_1(T_{ij}) + \delta_{ij}^{(2)} \frac{D_{21}(T_{ij}, \tilde{T}_{ij})}{D_2(T_{ij}, \tilde{T}_{ij})} S_1(T_{ij}) \right. \\ &\quad \left. \left. + (1 - \delta_{ij}^{(1)} - \delta_{ij}^{(2)}) \frac{D_1(T_{ij}, \tilde{T}_{ij})}{c\{S_1(T_{ij}), S_2(\tilde{T}_{ij})\}} S_1(T_{ij}) \right) \right\} R_{2i}(\omega_0, b_i) \psi(b_i, \Sigma_0) db_i, \end{aligned} \quad (36)$$

$$\begin{aligned} l_{\beta_{02}}[h_3] &= \sum_{j=1}^{n_i} \int_{b_i} l_{\beta_{02}, j}[h_3] \psi(b_i, \Sigma_0) db_i \\ &= \sum_{j=1}^{n_i} \int_{b_i} h'_3 X_{ij} \left\{ \delta_{ij}^{(1)} \tilde{\delta}_{ij}^{(2)} + \delta_{ij}^{(2)} - \Lambda_{02}(\tilde{T}_{ij}) e^{\beta'_{02} X_{ij} + b'_2 Z_{ij}} \left(\delta_{ij}^{(1)} \tilde{\delta}_{ij}^{(2)} \frac{D_{122}(T_{ij}, \tilde{T}_{ij})}{D_{12}(T_{ij}, \tilde{T}_{ij})} S_2(\tilde{T}_{ij}) \right. \right. \\ &\quad \left. + \delta_{ij}^{(1)} \tilde{\delta}_{ij}^{(2)} + \delta_{ij}^{(2)} + \delta_{ij}^{(1)} (1 - \tilde{\delta}_{ij}^{(2)}) \frac{D_{12}(T_{ij}, \tilde{T}_{ij})}{D_1(T_{ij}, \tilde{T}_{ij})} S_2(\tilde{T}_{ij}) + \delta_{ij}^{(2)} \frac{D_{22}(T_{ij}, \tilde{T}_{ij})}{D_2(T_{ij}, \tilde{T}_{ij})} S_2(\tilde{T}_{ij}) \right. \\ &\quad \left. \left. + (1 - \delta_{ij}^{(1)} - \delta_{ij}^{(2)}) \frac{D_2(T_{ij}, \tilde{T}_{ij})}{c\{S_1(T_{ij}), S_2(\tilde{T}_{ij})\}} S_2(\tilde{T}_{ij}) \right) \right\} R_{2i}(\omega_0, b_i) \psi(b_i, \Sigma_0) db_i, \end{aligned} \quad (37)$$

$$\begin{aligned} l_{\Lambda_{01}}[h_4] &= \sum_{j=1}^{n_i} \int_{b_i} l_{\Lambda_{01}, j}[h_4] \psi(b_i, \Sigma_0) db_i \\ &= \sum_{j=1}^{n_i} \int_{b_i} \left\{ \delta_{ij}^{(1)} h_4(T_{ij}) - \int_0^{T_{ij}} h_4(s) d\Lambda_{01}(s) e^{\beta'_{01} X_{ij} + b'_1 Z_{ij}} \left(\delta_{ij}^{(1)} \tilde{\delta}_{ij}^{(2)} \frac{D_{121}(T_{ij}, \tilde{T}_{ij})}{D_{12}(T_{ij}, \tilde{T}_{ij})} S_1(T_{ij}) \right. \right. \\ &\quad \left. + \delta_{ij}^{(1)} + \delta_{ij}^{(1)} (1 - \tilde{\delta}_{ij}^{(2)}) \frac{D_{11}(T_{ij}, \tilde{T}_{ij})}{D_1(T_{ij}, \tilde{T}_{ij})} S_1(T_{ij}) + \delta_{ij}^{(2)} \frac{D_{21}(T_{ij}, \tilde{T}_{ij})}{D_2(T_{ij}, \tilde{T}_{ij})} S_1(T_{ij}) \right. \\ &\quad \left. \left. + (1 - \delta_{ij}^{(1)} - \delta_{ij}^{(2)}) \frac{D_1(T_{ij}, \tilde{T}_{ij})}{c\{S_1(T_{ij}), S_2(\tilde{T}_{ij})\}} S_1(T_{ij}) \right) \right\} R_{2i}(\omega_0, b_i) \psi(b_i, \Sigma_0) db_i, \end{aligned} \quad (38)$$

$$\begin{aligned}
l_{\Lambda_{02}}[h_5] &= \sum_{j=1}^{n_i} \int_{b_i} l_{\Lambda_{02},j}[h_5] \psi(b_i, \Sigma_0) db_i \\
&= \sum_{j=1}^{n_i} \int_{b_i} \left\{ (\delta_{ij}^{(1)} \tilde{\delta}_{ij}^{(2)} + \delta_{ij}^{(2)}) h_5(\tilde{T}_{ij}) - \int_0^{\tilde{T}_{ij}} h_5(s) d\Lambda_{02}(s) e^{\beta'_{02} X_{ij} + b'_i Z_{ij}} \left(\delta_{ij}^{(1)} \tilde{\delta}_{ij}^{(2)} \frac{D_{122}(T_{ij}, \tilde{T}_{ij})}{D_{12}(T_{ij}, \tilde{T}_{ij})} S_2(\tilde{T}_{ij}) \right. \right. \\
&\quad + \delta_{ij}^{(1)} \tilde{\delta}_{ij}^{(2)} + \delta_{ij}^{(2)} + \delta_{ij}^{(1)} (1 - \tilde{\delta}_{ij}^{(2)}) \frac{D_{12}(T_{ij}, \tilde{T}_{ij})}{D_1(T_{ij}, \tilde{T}_{ij})} S_2(\tilde{T}_{ij}) + \delta_{ij}^{(2)} \frac{D_{22}(T_{ij}, \tilde{T}_{ij})}{D_2(T_{ij}, \tilde{T}_{ij})} S_2(\tilde{T}_{ij}) \\
&\quad \left. \left. + (1 - \delta_{ij}^{(1)} - \delta_{ij}^{(2)}) \frac{D_2(T_{ij}, \tilde{T}_{ij})}{c\{S_1(T_{ij}), S_2(\tilde{T}_{ij})\}} S_2(\tilde{T}_{ij}) \right) \right\} R_{2i}(\omega_0, b_i) \psi(b_i, \Sigma_0) db_i,
\end{aligned} \tag{39}$$

$$l_{\Sigma_0}[h_6] = \int_{b_i} \frac{d}{d\epsilon} \psi(b_i, \Sigma_0 + \epsilon h_6) \Big|_{\epsilon=0} R_{2i}(\omega_0, b_i) \psi(b_i, \Sigma_0) db_i. \tag{40}$$

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 $j = 1, \dots, n_i$. For fixed integer h in $1, \dots, n_i$, let $\tilde{\delta}_{ij}^{(2)} = 1$, $\tilde{T}_{ij} = 0$ when $j \leq h$. When $j > h$, replace $\tilde{T}_{ij}$ with ς if $\tilde{\delta}_{ij}^{(2)} = 0$ and integrate $\tilde{T}_{ij}$ from 0 to ς otherwise. Then,

$$\begin{aligned}
R_{2i}(\omega_0, b) &= \prod_{j=1}^{n_i} \lambda_{01}(T_{ij}) e^{\beta'_{01} X_{ij} + b'_i Z_{ij}} D_{12}(T_{ij}, \tilde{T}_{ij})^{\tilde{\delta}_{ij}^{(2)}} D_1(T_{ij}, \tilde{T}_{ij})^{1 - \tilde{\delta}_{ij}^{(2)}} \\
&\quad \cdot \exp \left\{ \tilde{\delta}_{ij}^{(2)} (\beta'_{02} X_{ij} + b'_i Z_{ij}) - \tilde{\delta}_{ij}^{(2)} \Lambda_{02}(\tilde{T}_{ij}) e^{\beta'_{02} X_{ij} + b'_i Z_{ij}} \right\} \lambda_{02}(\tilde{T}_{ij})^{\tilde{\delta}_{ij}^{(2)}}.
\end{aligned} \tag{41}$$

For fixed integer h in $1, \dots, n_i$, we define measures $\mu_1, \dots, \mu_{n_i}$ on the set $\{0, 1\} \times [0, \varsigma]$ as follows:

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

and

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

where A is any Borel set in $[0, \varsigma]$.

We integrate both sides of equation (33) with respect to $\{(\tilde{\delta}_{i1}^{(2)}, \tilde{T}_{i1}), (\tilde{\delta}_{i2}^{(2)}, \tilde{T}_{i2}), \dots, (\tilde{\delta}_{in_i}^{(2)}, \tilde{T}_{in_i})\}$ and product

measure $d\mu_1, \dots, d\mu_{n_i}$, denoted by $d\mu$ for simplicity. Note that, for any b ,

$$\begin{aligned}
& \int R_{2i}(\omega_0, b) d\mu \\
&= \prod_{r \leq h} \left\{ \lambda_{01}(0) e^{\beta'_{01} X_{ir} + b'_i Z_{ir}} D_{12}(0, 0) e^{\beta'_{02} X_{ir} + b'_i Z_{ir}} \lambda_{02}(0) \right\} \prod_{\substack{\tilde{\delta}_{ir}^{(2)} \in \{0,1\} \\ r > h}} \prod_{r > h} \left\{ \lambda_{01}(0) e^{\beta'_{01} X_{ir} + b'_i Z_{ir}} D_1(0, \varsigma) \right\}^{1 - \tilde{\delta}_{ir}^{(2)}} \\
&\quad \cdot \left\{ \int_{y=0}^{\varsigma} \lambda_{01}(0) e^{\beta'_{01} X_{ir} + b'_i Z_{ir}} D_{12}(0, y) S_2(y) e^{\beta'_{02} X_{ir} + b'_i Z_{ir}} \lambda_{02}(y) dy \right\}^{\tilde{\delta}_{ir}^{(2)}} \\
&= \prod_{r \leq h} \left\{ \lambda_{01}(0) e^{\beta'_{01} X_{ir} + b'_i Z_{ir}} D_{12}(0, 0) e^{\beta'_{02} X_{ir} + b'_i Z_{ir}} \lambda_{02}(0) \right\} \prod_{\substack{\tilde{\delta}_{ir}^{(2)} \in \{0,1\} \\ r > h}} \prod_{r > h} \left\{ \lambda_{01}(0) e^{\beta'_{01} X_{ir} + b'_i Z_{ir}} D_1(0, \varsigma) \right\}^{1 - \tilde{\delta}_{ir}^{(2)}} \\
&\quad \cdot \left\{ \lambda_{01}(0) e^{\beta'_{01} X_{ir} + b'_i Z_{ir}} (-D_1(0, \varsigma) + D_1(0, 0)) \right\}^{\tilde{\delta}_{ir}^{(2)}} \\
&= \prod_{r \leq h} \left\{ \lambda_{01}(0) e^{\beta'_{01} X_{ir} + b'_i Z_{ir}} D_{12}(0, 0) e^{\beta'_{02} X_{ir} + b'_i Z_{ir}} \lambda_{02}(0) \right\} \prod_{r > h} \left\{ \lambda_{01}(0) e^{\beta'_{01} X_{ir} + b'_i Z_{ir}} D_1(0, 0) \right\}.
\end{aligned} \tag{42}$$

Consider the first term $l_{\alpha_0}[h_1]$ on the left side of equation (33). If $j \leq h$, from (35) we have

$$\begin{aligned}
l_{\alpha_0, j}[h_1] &= \int h_1 \frac{D_{12\alpha}(0, 0)}{D_{12}(0, 0)} R_{2i}(\omega_0, b) d\mu \\
&= h_1 \frac{D_{12\alpha}(0, 0)}{D_{12}(0, 0)} \prod_{r \leq h} \left\{ \lambda_{01}(0) e^{\beta'_{01} X_{ir} + b'_i Z_{ir}} D_{12}(0, 0) e^{\beta'_{02} X_{ir} + b'_i Z_{ir}} \lambda_{02}(0) \right\} \prod_{r > h} \left\{ \lambda_{01}(0) e^{\beta'_{01} X_{ir} + b'_i Z_{ir}} D_1(0, 0) \right\};
\end{aligned}$$

Otherwise if $j > h$,

$$\begin{aligned}
l_{\alpha_0, j}[h_1] &= h_1 \prod_{r \leq h} \left\{ \lambda_{01}(0) e^{\beta'_{01} X_{ir} + b'_i Z_{ir}} D_{12}(0, 0) e^{\beta'_{02} X_{ir} + b'_i Z_{ir}} \lambda_{02}(0) \right\} \prod_{r > h, r \neq j} \left\{ \lambda_{01}(0) e^{\beta'_{01} X_{ir} + b'_i Z_{ir}} D_1(0, 0) \right\} \\
&\quad \cdot \sum_{\substack{\tilde{\delta}_{ij}^{(2)} \in \{0,1\}}} \left\{ (1 - \tilde{\delta}_{ij}^{(2)}) D_{1\alpha}(0, \varsigma) \lambda_{01}(0) e^{\beta'_{01} X_{ij} + b'_i Z_{ij}} + \tilde{\delta}_{ij}^{(2)} \lambda_{01}(0) e^{\beta'_{01} X_{ij} + b'_i Z_{ij}} \int_0^{\varsigma} -D_{12\alpha}(0, y) dy \right\} \\
&= h_1 \frac{D_{1\alpha}(0, 0)}{D_1(0, 0)} \prod_{r \leq h} \left\{ \lambda_{01}(0) e^{\beta'_{01} X_{ir} + b'_i Z_{ir}} D_{12}(0, 0) e^{\beta'_{02} X_{ir} + b'_i Z_{ir}} \lambda_{02}(0) \right\} \prod_{r > h} \left\{ \lambda_{01}(0) e^{\beta'_{01} X_{ir} + b'_i Z_{ir}} D_1(0, 0) \right\}.
\end{aligned}$$

Then the contribution of $l_{\alpha_0}[h_1]$ in (33) reduces to

$$\begin{aligned}
& h_1 \left\{ \sum_{j \leq h} \frac{D_{12\alpha}(0, 0)}{D_{12}(0, 0)} + \sum_{j > h} \frac{D_{1\alpha}(0, 0)}{D_1(0, 0)} \right\} \prod_{r \leq h} \left\{ \lambda_{01}(0) e^{\beta'_{01} X_{ir}} D_{12}(0, 0) e^{\beta'_{02} X_{ir}} \lambda_{02}(0) \right\} \\
& \cdot \prod_{r > h} \left\{ \lambda_{01}(0) e^{\beta'_{01} X_{ir}} D_1(0, 0) \right\} \cdot E(e^{\sum_{j=1}^h b'_i Z_{ij} + \sum_{j=1}^{n_i} b'_i Z_{ij}}).
\end{aligned} \tag{43}$$

Consider the second term $l_{\beta_{01}}[h_2]$ on the left side of (33). If $j \leq h$, by (36) we have

$$\begin{aligned} l_{\beta_{01},j}[h_2] &= \int h'_2 X_{ij} (1-0) \times R_{2i}(\omega_0, b) d\mu \\ &= h'_2 X_{ij} \prod_{r \leq h} \left\{ \lambda_{01}(0) e^{\beta'_{01} X_{ir} + b'_i Z_{ir}} D_{12}(0, 0) e^{\beta'_{02} X_{ir} + b'_i Z_{ir}} \lambda_{02}(0) \right\} \prod_{r > h} \left\{ \lambda_{01}(0) e^{\beta'_{01} X_{ir} + b'_i Z_{ir}} D_1(0, 0) \right\}; \end{aligned}$$

Otherwise,

$$\begin{aligned} l_{\beta_{01},j}[h_2] &= h'_2 X_{ij} \prod_{r \leq h} \left\{ \lambda_{01}(0) e^{\beta'_{01} X_{ir} + b'_i Z_{ir}} D_{12}(0, 0) e^{\beta'_{02} X_{ir} + b'_i Z_{ir}} \lambda_{02}(0) \right\} \prod_{r > h, r \neq j} \left\{ \lambda_{01}(0) e^{\beta'_{01} X_{ir} + b'_i Z_{ir}} D_1(0, 0) \right\} \\ &\cdot \sum_{\tilde{\delta}_{ij}^{(2)} \in \{0,1\}} \left[(1 - \tilde{\delta}_{ij}^{(2)}) \lambda_{01}(0) e^{\beta'_{01} X_{ij} + b'_i Z_{ij}} D_1(0, \varsigma) + \tilde{\delta}_{ij}^{(2)} \int_{y=0}^{\varsigma} \lambda_{01}(0) e^{\beta'_{01} X_{ij} + b'_i Z_{ij}} D_{12}(0, y) S_2(y) e^{\beta'_{02} X_{ij} + b'_i Z_{ij}} \lambda_{02}(y) dy \right] \\ &= h'_2 X_{ij} \prod_{r \leq h} \left\{ \lambda_{01}(0) e^{\beta'_{01} X_{ir} + b'_i Z_{ir}} D_{12}(0, 0) e^{\beta'_{02} X_{ir} + b'_i Z_{ir}} \lambda_{02}(0) \right\} \prod_{r > h, r \neq j} \left\{ \lambda_{01}(0) e^{\beta'_{01} X_{ir} + b'_i Z_{ir}} D_1(0, 0) \right\} \\ &\cdot \sum_{\tilde{\delta}_{ij}^{(2)} \in \{0,1\}} \left[(1 - \tilde{\delta}_{ij}^{(2)}) \lambda_{01}(0) e^{\beta'_{01} X_{ij} + b'_i Z_{ij}} D_1(0, \varsigma) + \tilde{\delta}_{ij}^{(2)} \lambda_{01}(0) e^{\beta'_{01} X_{ij} + b'_i Z_{ij}} \left(-D_1(0, \varsigma) + D_1(0, 0) \right) \right] \\ &= h'_2 X_{ij} \prod_{r \leq h} \left\{ \lambda_{01}(0) e^{\beta'_{01} X_{ir} + b'_i Z_{ir}} D_{12}(0, 0) e^{\beta'_{02} X_{ir} + b'_i Z_{ir}} \lambda_{02}(0) \right\} \prod_{r > h} \left\{ \lambda_{01}(0) e^{\beta'_{01} X_{ir} + b'_i Z_{ir}} D_1(0, 0) \right\}. \end{aligned}$$

Then the contribution of $l_{\beta_{01}}[h_2]$ in (33) becomes

$$\sum_{j=1}^{n_i} h'_2 X_{ij} \prod_{r \leq h} \left\{ \lambda_{01}(0) e^{\beta'_{01} X_{ir} + b'_i Z_{ir}} D_{12}(0, 0) e^{\beta'_{02} X_{ir} + b'_i Z_{ir}} \lambda_{02}(0) \right\} \prod_{r > h} \left\{ \lambda_{01}(0) e^{\beta'_{01} X_{ir} + b'_i Z_{ir}} D_1(0, 0) \right\} E(e^{\sum_{j=1}^h b'_i Z_{ij} + \sum_{j=1}^{n_i} b'_i Z_{ij}}). \quad (44)$$

Consider the third term $l_{\beta_{02}}[h_3]$ on the left side of (33). If $j \leq h$, by (37) we have

$$\begin{aligned} l_{\beta_{02},j}[h_3] &= \int h'_3 X_{ij} (1-0) R_{2i}(\omega_0, b) d\mu \\ &= h'_3 X_{ij} \prod_{r \leq h} \left\{ \lambda_{01}(0) e^{\beta'_{01} X_{ir} + b'_i Z_{ir}} D_{12}(0, 0) e^{\beta'_{02} X_{ir} + b'_i Z_{ir}} \lambda_{02}(0) \right\} \prod_{r > h} \left\{ \lambda_{01}(0) e^{\beta'_{01} X_{ir} + b'_i Z_{ir}} D_1(0, 0) \right\}; \end{aligned}$$

otherwise,

$$\begin{aligned}
l_{\beta_{02},j}[h_3] &= h'_3 X_{ij} \prod_{r \leq h} \left\{ \lambda_{01}(0) e^{\beta'_{01} X_{ir} + b'_i Z_{ir}} D_{12}(0,0) e^{\beta'_{02} X_{ir} + b'_i Z_{ir}} \lambda_{02}(0) \right\} \prod_{r > h, r \neq j} \left\{ \lambda_{01}(0) e^{\beta'_{01} X_{ir} + b'_i Z_{ir}} D_1(0,0) \right\} \\
&\cdot \sum_{\tilde{\delta}_{ij}^{(2)} \in \{0,1\}} \left[(1 - \tilde{\delta}_{ij}^{(2)}) \left\{ -\Lambda_{02}(\varsigma) e^{\beta'_{02} X_{ij} + b'_i Z_{ij}} D_{12}(0,\varsigma) S_2(\varsigma) \lambda_{01}(0) e^{\beta'_{01} X_{ij} + b'_i Z_{ij}} \right\} \right. \\
&\quad \left. + \tilde{\delta}_{ij}^{(2)} \int_{y=0}^{\varsigma} \left\{ 1 - \Lambda_{02}(y) e^{\beta'_{02} X_{ij} + b'_i Z_{ij}} \frac{D_{122}(0,y)}{D_{12}(0,y)} S_2(y) \right\} D_{12}(0,y) \lambda_{01}(0) e^{\beta'_{01} X_{ij} + b'_i Z_{ij}} S_2(y) e^{\beta'_{02} X_{ij} + b'_i Z_{ij}} \lambda_{02}(y) dy \right] \\
&= h'_3 X_{ij} \prod_{r \leq h} \left\{ \lambda_{01}(0) e^{\beta'_{01} X_{ir} + b'_i Z_{ir}} D_{12}(0,0) e^{\beta'_{02} X_{ir} + b'_i Z_{ir}} \lambda_{02}(0) \right\} \prod_{r > h, r \neq j} \left\{ \lambda_{01}(0) e^{\beta'_{01} X_{ir} + b'_i Z_{ir}} D_1(0,0) \right\} \\
&\cdot \sum_{\tilde{\delta}_{ij}^{(2)} \in \{0,1\}} \left[(1 - \tilde{\delta}_{ij}^{(2)}) \left\{ -\Lambda_{02}(\varsigma) e^{\beta'_{02} X_{ij} + b'_i Z_{ij}} D_{12}(0,\varsigma) S_2(\varsigma) \lambda_{01}(0) e^{\beta'_{01} X_{ij} + b'_i Z_{ij}} \right\} \right. \\
&\quad \left. + \tilde{\delta}_{ij}^{(2)} \lambda_{01}(0) e^{\beta'_{01} X_{ij} + b'_i Z_{ij}} \int_{y=0}^{\varsigma} \frac{\partial \left\{ \frac{\partial D_1(0,y)}{\partial \lambda_{02}(y)} \Lambda_{02}(y) \right\}}{\partial y} dy \right] \\
&= h'_3 X_{ij} \prod_{r \leq h} \left\{ \lambda_{01}(0) e^{\beta'_{01} X_{ir} + b'_i Z_{ir}} D_{12}(0,0) e^{\beta'_{02} X_{ir} + b'_i Z_{ir}} \lambda_{02}(0) \right\} \prod_{r > h, r \neq j} \left\{ \lambda_{01}(0) e^{\beta'_{01} X_{ir} + b'_i Z_{ir}} D_1(0,0) \right\} \cdot 0 = 0.
\end{aligned}$$

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

$$\sum_{j \leq h} h'_3 X_{ij} \prod_{r \leq h} \left\{ \lambda_{01}(0) e^{\beta'_{01} X_{ir}} D_{12}(0,0) e^{\beta'_{02} X_{ir}} \lambda_{02}(0) \right\} \prod_{r > h} \left\{ \lambda_{01}(0) e^{\beta'_{01} X_{ir}} D_1(0,0) \right\} \cdot E(e^{\sum_{j=1}^h b'_i Z_{ij} + \sum_{j=1}^{n_i} b'_i Z_{ij}}). \quad (45)$$

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

$$l_{\Lambda_{01},j}[h_4] = h_4(0) \prod_{r \leq h} \left\{ \lambda_{01}(0) e^{\beta'_{01} X_{ir} + b'_i Z_{ir}} D_{12}(0,0) e^{\beta'_{02} X_{ir} + b'_i Z_{ir}} \lambda_{02}(0) \right\} \prod_{r > h} \left\{ \lambda_{01}(0) e^{\beta'_{01} X_{ir} + b'_i Z_{ir}} D_1(0,0) \right\};$$

Otherwise,

$$\begin{aligned}
l_{\Lambda_{01},j}[h_4] &= \prod_{r \leq h} \left\{ \lambda_{01}(0) e^{\beta'_{01} X_{ir} + b'_i Z_{ir}} D_{12}(0,0) e^{\beta'_{02} X_{ir} + b'_i Z_{ir}} \lambda_{02}(0) \right\} \prod_{r > h, r \neq j} \left\{ \lambda_{01}(0) e^{\beta'_{01} X_{ir} + b'_i Z_{ir}} D_1(0,0) \right\} \\
&\cdot \sum_{\tilde{\delta}_{ij}^{(2)} \in \{0,1\}} \left[(1 - \tilde{\delta}_{ij}^{(2)}) h_4(0) \lambda_{01}(0) e^{\beta'_{01} X_{ij} + b'_i Z_{ij}} D_1(0,\varsigma) \right. \\
&\quad \left. + \tilde{\delta}_{ij}^{(2)} h_4(0) \int_{y=0}^{\varsigma} \lambda_{01}(0) e^{\beta'_{01} X_{ij} + b'_i Z_{ij}} D_{12}(0,y) S_2(y) e^{\beta'_{02} X_{ij} + b'_i Z_{ij}} \lambda_{02}(y) dy \right] \\
&= \prod_{r \leq h} \left\{ \lambda_{01}(0) e^{\beta'_{01} X_{ir} + b'_i Z_{ir}} D_{12}(0,0) e^{\beta'_{02} X_{ir} + b'_i Z_{ir}} \lambda_{02}(0) \right\} \prod_{r > h, r \neq j} \left\{ \lambda_{01}(0) e^{\beta'_{01} X_{ir} + b'_i Z_{ir}} D_1(0,0) \right\} \\
&\cdot \sum_{\tilde{\delta}_{ij}^{(2)} \in \{0,1\}} \left[(1 - \tilde{\delta}_{ij}^{(2)}) h_4(0) \lambda_{01}(0) e^{\beta'_{01} X_{ij} + b'_i Z_{ij}} D_1(0,\varsigma) + \tilde{\delta}_{ij}^{(2)} h_4(0) \lambda_{01}(0) e^{\beta'_{01} X_{ij} + b'_i Z_{ij}} \left(-D_1(0,\varsigma) + D_1(0,0) \right) \right] \\
&= h_4(0) \prod_{r \leq h} \left\{ \lambda_{01}(0) e^{\beta'_{01} X_{ir} + b'_i Z_{ir}} D_{12}(0,0) e^{\beta'_{02} X_{ir} + b'_i Z_{ir}} \lambda_{02}(0) \right\} \prod_{r > h} \left\{ \lambda_{01}(0) e^{\beta'_{01} X_{ir} + b'_i Z_{ir}} D_1(0,0) \right\}.
\end{aligned}$$

Therefore, the contribution of $l_{\Lambda_{01}}[h_4]$ in (33) becomes

$$\sum_{j=1}^{n_i} h_4(0) \prod_{r \leq h} \left\{ \lambda_{01}(0) e^{\beta'_{01} X_{ir}} D_{12}(0, 0) e^{\beta'_{02} X_{ir}} \lambda_{02}(0) \right\} \prod_{r > h} \left\{ \lambda_{01}(0) e^{\beta'_{01} X_{ir}} D_1(0, 0) \right\} \cdot E \left(e^{\sum_{j=1}^h b'_i Z_{ij} + \sum_{j=1}^{n_i} b'_i Z_{ij}} \right). \quad (46)$$

For the fifth term $l_{\Lambda_{02}}[h_5]$ of (33), if $j \leq h$, it follows from (39) that for any b

$$l_{\Lambda_{02},j}[h_5] = h_5(0) \prod_{r \leq h} \left\{ \lambda_{01}(0) e^{\beta'_{01} X_{ir} + b'_i Z_{ir}} D_{12}(0, 0) e^{\beta'_{02} X_{ir} + b'_i Z_{ir}} \lambda_{02}(0) \right\} \prod_{r > h} \left\{ \lambda_{01}(0) e^{\beta'_{01} X_{ir} + b'_i Z_{ir}} D_1(0, 0) \right\}.$$

For the fifth term of (33), when $j > h$,

$$\begin{aligned} l_{\Lambda_{02},j}[h_5] &= \prod_{r \leq h} \left\{ \lambda_{01}(0) e^{\beta'_{01} X_{ir} + b'_i Z_{ir}} D_{12}(0, 0) e^{\beta'_{02} X_{ir} + b'_i Z_{ir}} \lambda_{02}(0) \right\} \prod_{r > h, r \neq j} \left\{ \lambda_{01}(0) e^{\beta'_{01} X_{ir} + b'_i Z_{ir}} D_1(0, 0) \right\} \\ &\quad \cdot \sum_{\tilde{\delta}_{ij}^{(2)} \in \{0,1\}} \left[- (1 - \tilde{\delta}_{ij}^{(2)}) \int_0^\varsigma h_5(y) d\Lambda_{02}(y) e^{\beta'_{02} X_{ij} + b'_i Z_{ij}} \lambda_{01}(0) e^{\beta'_{01} X_{ij} + b'_i Z_{ij}} D_{12}(0, \varsigma) S_2(\varsigma) \right. \\ &\quad + \tilde{\delta}_{ij}^{(2)} \int_{y=0}^\varsigma \left\{ h_5(y) - \int_{s=0}^y h_5(s) d\Lambda_{02}(s) e^{\beta'_{02} X_{ij} + b'_i Z_{ij}} \left(1 + \frac{D_{122}(0, y)}{D_{12}(0, y)} S_2(y) \right) \right\} \\ &\quad \cdot D_{12}(0, y) \lambda_{01}(0) e^{\beta'_{01} X_{ij} + b'_i Z_{ij}} S_2(y) e^{\beta'_{02} X_{ij} + b'_i Z_{ij}} \lambda_{02}(y) dy \Big] \\ &= \prod_{r \leq h} \left\{ \lambda_{01}(0) e^{\beta'_{01} X_{ir} + b'_i Z_{ir}} D_{12}(0, 0) e^{\beta'_{02} X_{ir} + b'_i Z_{ir}} \lambda_{02}(0) \right\} \prod_{r > h, r \neq j} \left\{ \lambda_{01}(0) e^{\beta'_{01} X_{ir} + b'_i Z_{ir}} D_1(0, 0) \right\} \\ &\quad \cdot \sum_{\tilde{\delta}_{ij}^{(2)} \in \{0,1\}} \left[- (1 - \tilde{\delta}_{ij}^{(2)}) \int_0^\varsigma h_5(y) d\Lambda_{02}(y) e^{\beta'_{02} X_{ij} + b'_i Z_{ij}} \lambda_{01}(0) e^{\beta'_{01} X_{ij} + b'_i Z_{ij}} D_{12}(0, \varsigma) S_2(\varsigma) \right. \\ &\quad + \tilde{\delta}_{ij}^{(2)} \lambda_{01}(0) e^{\beta'_{01} X_{ij} + b'_i Z_{ij}} \int_{y=0}^\varsigma \frac{\partial \left\{ \frac{\partial D_1(0, y)}{\partial \Lambda_{02}(y)} \right\}}{\partial y} dy \Big] \\ &= \prod_{r \leq h} \left\{ \lambda_{01}(0) e^{\beta'_{01} X_{ir} + b'_i Z_{ir}} D_{12}(0, 0) e^{\beta'_{02} X_{ir} + b'_i Z_{ir}} \lambda_{02}(0) \right\} \prod_{r > h, r \neq j} \left\{ \lambda_{01}(0) e^{\beta'_{01} X_{ir} + b'_i Z_{ir}} D_1(0, 0) \right\} \cdot 0 = 0. \end{aligned}$$

Thus the contribution of $l_{\Lambda_{02}}[h_5]$ in (33) becomes

$$\sum_{j \leq h} h_5(0) \prod_{r \leq h} \left\{ \lambda_{01}(0) e^{\beta'_{01} X_{ir}} D_{12}(0, 0) e^{\beta'_{02} X_{ir}} \lambda_{02}(0) \right\} \prod_{r > h} \left\{ \lambda_{01}(0) e^{\beta'_{01} X_{ir}} D_1(0, 0) \right\} E \left(e^{\sum_{j=1}^h b'_i Z_{ij} + \sum_{j=1}^{n_i} b'_i Z_{ij}} \right). \quad (47)$$

Through similar calculations, the sixth term $l_{\Sigma_0}[h_6]$ of (33) reduces to:

$$\begin{aligned} &\prod_{r \leq h} \left\{ \lambda_{01}(0) e^{\beta'_{01} X_{ir}} D_{12}(0, 0) e^{\beta'_{02} X_{ir}} \lambda_{02}(0) \right\} \prod_{r > h} \left\{ \lambda_{01}(0) e^{\beta'_{01} X_{ir}} D_1(0, 0) \right\} \\ &\quad \cdot E \left\{ \left[\frac{\partial \psi(b_i, \Sigma_0 + \epsilon h_6)}{\partial \epsilon} \right]_{\epsilon=0} e^{\sum_{j=1}^h b'_i Z_{ij} + \sum_{j=1}^{n_i} b'_i Z_{ij}} \right\}. \end{aligned} \quad (48)$$

Combining (43), (44), (45), (46), (47) and (48), equation (33) is rewritten as:

$$\left[h_1 \left\{ \sum_{j \leq h} \frac{D_{12\alpha}(0,0)}{D_{12}(0,0)} + \sum_{j > h} \frac{D_{1\alpha}(0,0)}{D_1(0,0)} \right\} + \sum_{j=1}^{n_i} h'_2 X_{ij} + \sum_{j \leq h} h'_3 X_{ij} + \sum_{j=1}^{n_i} h_4(0) + \sum_{j \leq h} h_5(0) \right] \cdot E \left(e^{\sum_{j=1}^h b'_i Z_{ij} + \sum_{j=1}^{n_i} b'_i Z_{ij}} \right) + E \left\{ \left[\frac{\partial \psi(b_i, \Sigma_0 + \epsilon h_6)}{\partial \epsilon} \right]_{\epsilon=0} e^{\sum_{j=1}^h b'_i Z_{ij} + \sum_{j=1}^{n_i} b'_i Z_{ij}} \right\} = 0. \quad (49)$$

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:

$$\begin{aligned} & h_1 \left\{ \sum_{j \leq h} \frac{D_{12\alpha}(0,0)}{D_{12}(0,0)} + \sum_{j > h} \frac{D_{1\alpha}(0,0)}{D_1(0,0)} \right\} + \sum_{j=1}^{n_i} h'_2 X_{ij} + \sum_{j \leq h} h'_3 X_{ij} \\ & + \sum_{j=1}^{n_i} h_4(0) + \sum_{j \leq h} h_5(0) + \frac{1}{2} \left(\sum_{j=1}^{n_i} Z_{ij} \right)' H_6 \left(\sum_{j=1}^{n_i} Z_{ij} \right) = 0. \end{aligned} \quad (50)$$

In (50), we can see that only the term multiplied with h_1 is a function of α , other terms are unrelated with α . (50) exists over different values of α , and thus $h_1 = 0$. 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 $1 \leq k_1 \leq k_2 \leq n_i$, leading to $h_3 = 0$ and $h_5(0) = 0$. Then (50) is simplified as

$$\sum_{j=1}^{n_i} h'_2 X_{ij} + \sum_{j=1}^{n_i} h_4(0) + \frac{1}{2} \left(\sum_{j=1}^{n_i} Z_{ij} \right)' H_6 \left(\sum_{j=1}^{n_i} Z_{ij} \right) = 0. \quad (51)$$

Applying condition 8 in the paper, we obtain $h_2 = 0$, $h_4(0) = 0$, $H_6 = 0$.

Setting $\delta_{ij}^{(1)} = \tilde{\delta}_{ij}^{(2)} = 1$ and $T_{ij} = 0$ for $j = 1, \dots, n_i$, $\tilde{T}_{ij} = 0$ for $j = 2, \dots, n_i$, and $c(u, v) = uv$ in (33) gives that

$$h_5(\tilde{T}_{i1}) = \left[\int_0^{\tilde{T}_{i1}} h_5(s) d\Lambda_{02}(s) \left\{ \int_b e^{\beta'_{02} X_{i1} + b'_i Z_{i1}} R_{2i}(\omega_0, b) \phi(0, \Sigma_0) db \right\} \right] / \left\{ \int_b R_{2i}(\omega_0, b) \phi(0, \Sigma_0) db \right\}.$$

Therefore, $g(y) \equiv \int_0^y h_5(t) d\Lambda_{02}(t)$ satisfies the homogeneous equation

$$\frac{g'(y)}{\lambda_{02}(y)} = \left[g(y) \left\{ \int_b e^{\beta'_{02} X_{i1} + b'_i Z_{i1}} R_{2i}(\omega_0, b) \phi(0, \Sigma_0) db \right\} \right] / \left\{ \int_b R_{2i}(\omega_0, b) \phi(0, \Sigma_0) db \right\}.$$

With boundary condition $g(0) = 0$, we conclude $g(y) = 0$, $h_5(y) = 0$ and thus $h_5(\cdot) = 0$. Similarly, setting $\delta_{ij}^{(1)} = 1$ for $j = 1, \dots, n_i$, $\tilde{\delta}_{ij}^{(2)} = 1$ and $T_{ij} = \tilde{T}_{ij} = 0$ for $j = 2, \dots, n_i$, $\tilde{\delta}_{i1}^{(2)} = 0$ and $\tilde{T}_{i1} = \varsigma$, $c(u, v) = uv$ in (33) finally yield that $h_4(\cdot) = 0$.

Under the gamma frailty assumption, (49) is rewritten as:

$$h_1 \left\{ \sum_{j \leq h} \frac{D_{12\alpha}(0,0)}{D_{12}(0,0)} + \sum_{j > h} \frac{D_{1\alpha}(0,0)}{D_1(0,0)} \right\} + \sum_{j=1}^{n_i} h'_2 X_{ij} + \sum_{j \leq h} h'_3 X_{ij} + \sum_{j=1}^{n_i} h_4(0) + \sum_{j \leq h} h_5(0) \\ + \left\{ -\left(\frac{1}{\eta_0} - 1\right) \frac{1}{(n_i + h - 2)\eta_0 + 1} - (n_i + h)\eta_0 - 2 + \log(\eta_0) + \varphi_0\left(\frac{1}{\eta_0}\right) \right\} \cdot \frac{h_6}{\eta_0^2} = 0, \quad (52)$$

where $\varphi(\frac{1}{\lambda})$ is the polygamma function.

Note that we can see that only the term multiplied with h_1 is a function of α_0 , other terms are unrelated with α_0 . (52) exists over different values of α_0 , and thus $h_1 = 0$. In (52), we can see only the term multiplied with h_6 is a function of η_0 , while the other terms are unrelated with η_0 . (52) exists over different values of η_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

$$\sum_{j=1}^{n_i} h'_2 X_{ij} + \sum_{j=1}^{n_i} h_4(0) = 0. \quad (53)$$

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

$$h_5(\tilde{T}_{i1}) = \left[\int_0^{\tilde{T}_{i1}} h_5(s) d\Lambda_{02}(s) \left\{ \int_u e^{\beta'_{02} X_{i1}} R_{2i}(\omega_0, b) d_u \Gamma(1/\eta_0, \eta_0) \right\} \right] / \left\{ \int_u R_{2i}(\omega_0, b) d_u \Gamma(1/\eta_0, \eta_0) \right\}.$$

Therefore $g(y) \equiv \int_0^y h_5(t) d\Lambda_{02}(t)$ satisfies the homogeneous equation

$$\frac{g'(y)}{\lambda_{02}(y)} = \left[g(y) \left\{ \int_u e^{\beta'_{02} X_{i1}} R_{2i}(\omega_0, b) d_u \Gamma(1/\eta_0, \eta_0) \right\} \right] / \left\{ \int_u R_{2i}(\omega_0, b) d_u \Gamma(1/\eta, \eta_0) \right\}.$$

With boundary condition $g(0) = 0$, we conclude $g(y) = 0$, $h_5(y) = 0$, thus $h_5(\cdot) = 0$. Similarly, set $\delta_{ij}^{(1)} = 1$ for $j = 1, \dots, n_i$, $\tilde{\delta}_{ij}^{(2)} = 1$ and $T_{ij} = \tilde{T}_{ij} = 0$ for $j = 2, \dots, n_i$, $\tilde{\delta}_{i1}^{(2)} = 0$ and $\tilde{T}_{i1} = \varsigma$, $c(u, v) = uv$ 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 $\dot{S}(\omega_0)$ 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.