COMMUNICATIONS IN STATISTICS—THEORY AND METHODS
2019, VOL. 48, NO. 5, 1193–1220
https://doi.org/10.1080/03610926.2018.1425450

Fitting competing risks data to bivariate Pareto models

Jia-Han Shih, Wei Lee, Li-Hsien Sun, and Takeshi Emura

Graduate Institute of Statistics, National Central University, Taiwan

ABSTRACT
This paper revisits two bivariate Pareto models for fitting competing risks data. The first model is the Frank copula model, and the second one is a bivariate Pareto model introduced by Sankaran and Nair (1993). We discuss the identifiability issues of these models and develop the maximum likelihood estimation procedures including their computational algorithms and model-diagnostic procedures. Simulations are conducted to examine the performance of the maximum likelihood estimation. Real data are analyzed for illustration.

ARTICLE HISTORY
Received 22 June 2017
Accepted 2 January 2018

KEYWORDS
Bivariate Pareto distribution; Frank copula; Kendall's tau; Newton-Raphson algorithm; Survival analysis.

1. Introduction

Vilfredo Pareto proposed a distribution for fitting income data, later known as the Pareto distribution. The probability density function (pdf) and the survival function of the Pareto distribution are $f(x) = \alpha\gamma(\alpha x)^{-\gamma-1}$ and $S(x) = (\alpha x)^{-\gamma}$, $x \geq 1/\alpha$, where $\alpha > 0$ is a scale parameter and $\gamma > 0$ is a shape parameter which represents the inequality of income distribution. According to the review by Arnold (2014), there are five different types of the Pareto distribution. The original Pareto distribution is referred to the *Pareto type I distribution*. The unnatural range, $x \geq 1/\alpha$, may yield inconvenience for the analysis of failure time data since the origin of failure time is usually zero.

The *Pareto type II distribution* (also known as the Lomax distribution) has the range $x \geq 0$. In analysis of bivariate failure time data, the Pareto type II seems to be more popular than the type I. Lindley and Singpurwalla (1986) introduced a bivariate Pareto model for life lengths of system components which shall be called the Lindley-Singpurwalla bivariate Pareto (LSBP) model. Sankaran and Nair (1993) extended the LSBP model for applications to reliability which shall be called the Sankaran and Nair bivariate Pareto (SNBP) model. These models are based on the Pareto type II or the Lomax distribution. The bivariate model based on the Pareto type I is referred to p.91 of Mardia (1970).

The Pareto type II distribution has been used for fitting competing risks data. Escarela and Carrière (2003) proposed to fit the Frank copula model with the Pareto margins for the prostate cancer data. Sankaran and Kundu (2014) proposed to fit the SNBP model for the life test data on appliances. While both papers demonstrated the usefulness of their Pareto

CONTACT T. Emura takeshiemura@gmail.com Graduate Institute of Statistics, National Central University, No. 300, Zhongda Rd., Zhongli District, Taoyuan City 32001, Taiwan.
Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/lsta.
Supplemental data for this article can be accessed on the publisher's website

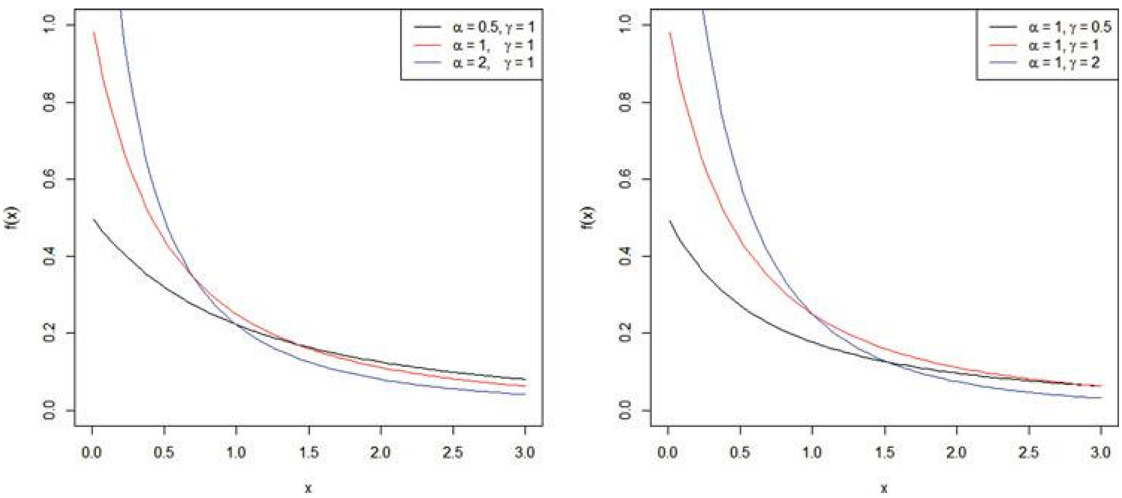

Figure 1. The pdf of the Pareto type II distribution.

models, they neither developed computational algorithms nor conducted simulation studies. In addition, the issue of model identifiability (Tsiatis 1975) was not discussed.

In this context, we revisit the aforementioned Pareto models, namely the Frank model and the SNBP models, for their ability to fit competing risks data. We first clarify the identifiability issue under these two models. We then develop computational and model-diagnostic tools to perform likelihood-based inference.

The paper is organized as follows. Section 2 gives the preliminary materials. Section 3 reviews the bivariate Pareto models. Section 4 develops maximum likelihood estimation procedures. Section 5 discusses model selection and diagnostic tools. Section 6 performs simulations and Section 7 analyzes real data. Section 8 concludes with future works.

2. Preliminary

This section reviews the definitions for the Pareto distribution and copulas.

2.1. Univariate Pareto model

Suppose that a random variable X follows the Pareto type II distribution. Its pdf and survival function are $f_X(x) = \alpha\gamma(1+\alpha x)^{-\gamma-1}$ and $S_X(x) = (1+\alpha x)^{-\gamma}$, $x \geq 0$, where $\alpha > 0$ is a scale parameter and $\gamma > 0$ is a shape parameter. One can show that $X + 1/\alpha$ follows the Pareto type I distribution. Figure 1 reveals that $f_X(x)$ is decreasing in x. The formula for the k-th moment is

$$E(X^k) = \frac{\Gamma(k+1)\Gamma(\gamma-k)}{\alpha^k\Gamma(\gamma)},\ \gamma > k.$$

According to the above formula, the k-th moment does not exist if $\gamma \leq k$. Thus, the mean does not exist for $\gamma \leq 1$. It is often more convenient to use the median

$$M_X = S_X^{-1}(0.5) = \frac{2^{1/\gamma}-1}{\alpha}.$$

2.2. Copula function

A bivariate copula is a bivariate distribution function $C : [0, 1]^2 \mapsto [0, 1]$ with the $unif(0, 1)$ margin (Nelsen 2006). Therefore, any bivariate copula satisfies the uniformity conditions $C(u, 1) = u, 0 \le u \le 1$, and $C(1, v) = v, 0 \le v \le 1$.

Let X and Y be continuous failure times with a joint survival function

$$S(x, y) = \Pr(X > x, Y > y).$$

Let $S_1(x) = S(x, 0)$ and $S_2(y) = S(0, y)$ be its marginal survival functions. By Sklar's theorem (Sklar 1959), there exists a unique copula C such that

$$S(x, y) = C\{S_1(x), S_2(y)\}.$$

Conversely, one can obtain a bivariate survival function $C\{S_1(x), S_2(y)\}$ by specifying a copula C and two marginal survival functions S_1 and S_2. The copula C is especially called "survival copula" due to its relevance to survival functions (Nelsen 2006). For instance, one can use the Frank copula (Frank 1979)

$$C_\theta(u, v) = -\frac{1}{\theta}\log\left\{1 + \frac{(e^{-\theta u} - 1)(e^{-\theta v} - 1)}{e^{-\theta} - 1}\right\}, \ \theta \ne 0$$

to create a survival function $C_\theta\{S_1(x), S_2(y)\}$ having arbitrary marginal distributions.

3. Bivariate Pareto model

This section reviews two Pareto models (the Frank model and SNBP model) that have been considered for fitting bivariate competing risks data.

3.1. The bivariate Pareto model with the Frank copula

We define two survival functions of the Pareto distribution

$$S_1(x) = (1 + \alpha_1 x)^{-\gamma_1}, \ x \ge 0, \ S_2(y) = (1 + \alpha_2 y)^{-\gamma_2}, \ y \ge 0,$$

where $\alpha_1 > 0, \alpha_2 > 0, \gamma_1 > 0$, and $\gamma_2 > 0$. Escarela and Carrière (2003) considered the bivariate Pareto model with the Frank copula, defined as

$$S(x, y) = \Pr(X > x, Y > y) = -\frac{1}{\theta}\log\left\{1 + \frac{(e^{-\theta S_1(x)} - 1)(e^{-\theta S_2(y)} - 1)}{e^{-\theta} - 1}\right\}, \ \theta \ne 0. \quad (1)$$

The copula parameter θ describes the dependence between X and Y. The positive dependence corresponds to $\theta > 0$ while the negative dependence corresponds to $\theta < 0$. Kendall's tau for association between X and Y is

$$\tau_\theta = 1 - \frac{4}{\theta}\left(1 - \frac{1}{\theta}\int_0^\theta \frac{t}{e^t - 1}dt\right), \ \theta \ne 0.$$

The Clayton copula would have been the most popular choice for fitting bivariate competing risks data. For instance, Emura et al. (2017a, b; 2018) fitted the Clayton copula for dependence between time-to-cancer relapse and time-to-death. For a case like this, the Clayton copula is useful for describing their positive dependence. The main advantage of the Frank copula over the Clayton is the better ability for describing negative dependence. We shall show a real data analysis where negative dependence arises.

3.2. The SNBP distribution

Sankaran and Nair (1993) introduced a bivariate Pareto distribution

$$S(x, y) = (1 + \alpha_1 x + \alpha_2 y + \alpha_0 xy)^{-\gamma},\ x \geq 0,\ y \geq 0, \tag{2}$$

where $\alpha_1 > 0$, $\alpha_2 > 0$, $\gamma > 0$ and $0 \leq \alpha_0 \leq (\gamma + 1)\alpha_1\alpha_2$. We shall call it the Sankaran and Nair bivariate Pareto (SNBP) distribution. The marginal survival functions are

$$S_1(x) = (1 + \alpha_1 x)^{-\gamma},\ x \geq 0,\ S_2(y) = (1 + \alpha_2 y)^{-\gamma},\ y \geq 0.$$

Unlike the Frank model, the common shape parameter $\gamma > 0$ for two margins is assumed in the SNBP model. One can show that the survival copula for the SNBP model is

$$C_{\delta,\gamma}(u, v) = \{u^{-1/\gamma} + v^{-1/\gamma} - 1 + \delta(u^{-1/\gamma} - 1)(v^{-1/\gamma} - 1)\}^{-\gamma},\ \delta = \alpha_0/\alpha_1\alpha_2$$

which resembles the Clayton copula with parameter γ. However, since γ is determined by the marginal distribution, the copula parameter can be regarded as α_0 or δ.

The dependence properties of the SNBP model were discussed in Sankaran and Kundu (2014). The correlation coefficient ρ is positive for $\alpha_0 < \alpha_1\alpha_2$ while ρ is negative for $\alpha_0 > \alpha_1\alpha_2$. The case $\alpha_0 = \alpha_1\alpha_2$ corresponds to independence. However, the analytical formula of ρ has not been obtained.

The case of $\alpha_0 = 0$ gives the LSBP model (Lindley and Singpurwalla 1986),

$$S(x, y) = (1 + \alpha_1 x + \alpha_2 y)^{-\gamma},\ x \geq 0,\ y \geq 0.$$

The survival copula for the LSBP model is $C(u, v) = (u^{-1/\gamma} + v^{-1/\gamma} - 1)^{-\gamma}$ which is the Clayton copula (Nelsen 2006). Thus, the LSBP model can only produces positive dependence. The main advantage of the SNBP model over the LSBP model is the ability to describe negative dependence.

3.3. Example on the gross income data

To explain the two bivariate Pareto models, we analyze data on the gross income for the professors at the department of statistics at UC (University of California) system "http://ucpay.globl.org/". The data consist of professors, associate professors, and assistant professors who worked for more than 5 years at UC ($n = 77$). The data is available in Supplementary Material, which includes the income of many famous professors such as David Hinkley.

Let X be the 2005 income and Y be the 2010 income. We fit the samples (X_i, Y_i), $i = 1, 2, \ldots, 77$, to the Frank and SNBP models by using the maximum likelihood estimator (MLE) as detailed in Supplementary Material.

Under the Frank copula model, we obtain MLEs $\hat{\alpha}_1 = 1.2 \times 10^{-12}$, $\hat{\alpha}_2 = 2.8 \times 10^{-13}$, $\hat{\gamma}_1 = 1.3 \times 10^7$, $\hat{\gamma}_2 = 3.3 \times 10^7$, $\hat{\theta} = 8.52$, $\hat{M}_1 = 104725$ (US\$) and $\hat{M}_2 = 134410$ (US\$). Kendall's tau is $\hat{\tau}_\theta = 0.62$ which is computed from $\hat{\theta} = 8.52$. This indicates the positive dependence between the 2005 and 2010 incomes.

Under the SNBP model, we obtain MLEs by following the two-step procedure described in Sankaran and Kundu (2014). The estimation results are $\hat{\alpha}_1 = 1.5 \times 10^{-13}$, $\hat{\alpha}_2 = 9.0 \times 10^{-14}$, $\hat{\gamma} = 1.1 \times 10^8$, $\hat{\alpha}_0 = 3.1 \times 10^{-27}$, $\hat{M}_1 = 104700$ (US\$) and $\hat{M}_2 = 129546$ (US\$). The correlation coefficient is positive since $\hat{\alpha}_0 < \hat{\alpha}_1\hat{\alpha}_2$. However, we do not have a simple form of ρ as mentioned before.

In this data analysis, one can easily estimate the dependence parameter since the pair (X_i, Y_i) are observable for all samples. However, in the presence of competing risks, one can only observe X_i or Y_i. Special developments are necessary for the competing risks data, especially for the problem of identifiability of the dependence parameter.

4. Competing risks analysis

This section considers maximum likelihood estimation based on competing risks data.

4.1. Identifiability, data, and likelihood

Let X and Y be failure times due to two different causes (say, Cause 1 and Cause 2). Under competing risks, one can observe the first occurring failure time $T = \min(X, Y)$ and the cause indicators $\delta = \mathbf{I}(T = X)$ and $\delta^* = \mathbf{I}(T = Y)$. Since X and Y cannot be observed simultaneously, the dependence between X and Y is difficult to estimate. This phenomenon is known as nonidentifiability (Tsiatis 1975).

One can remove the nonidentifiability by imposing restrictive conditions on the model of (X, Y). For instance, model parameters are theoretically identifiable in some bivariate parametric classes with one- or two-parameter margins (David and Moeschberger 1978; Basu and Ghosh 1978). In the subsequent discussions, we shall consider restrictive conditions on the bivariate Pareto models to avoid the nonidentifiability.

Let (X_i, Y_i, C_i) $i = 1, 2, \ldots, n$ be i.i.d. random variables, where (X_i, Y_i) follows the model (1) or (2), and C_i is the independent censoring time. Let $T_i = \min(X_i, Y_i, C_i)$ be the observed failure time, $\delta_i = \mathbf{I}(T_i = X_i)$ be the indicator of Cause 1, $\delta_i^* = \mathbf{I}(T_i = Y_i)$ as the indicator of Cause 2. The data consist of $(T_i, \delta_i, \delta_i^*)$ for $i = 1, 2, \ldots, n$ (Table 1).

Based on Table 1, the log-likelihood function is

$$\ell_n(\varphi) = \sum_{i=1}^{n} \{\delta_i \log f^{(1)}(T_i) + \delta_i^* \log f^{(2)}(T_i) + (1 - \delta_i - \delta_i^*) \log S_T(T_i)\},$$

where φ is a vector of parameters, $S_T(t) = S(t, t)$ is the overall survival function, $f^{(1)}(t) = -\partial S(x, y)/\partial x|_{x=y=t}$ and $f^{(2)}(t) = -\partial S(x, y)/\partial y|_{x=y=t}$ are the sub-density functions. The MLE is defined as

$$\hat{\varphi} = \arg\max_{\varphi}\{\ell_n(\varphi)\}.$$

The nonidentifiability of the competing risks data may yield deleterious effects on the MLE. In particular, under the Frank model (1), we found that the profile likelihood

$$\tilde{\ell}_n(\theta) = \arg\max_{(\alpha_1, \alpha_2, \gamma_1, \gamma_2)} \{\ell_n(\alpha_1, \alpha_2, \gamma_1, \gamma_2, \theta)\}$$

Table 1. Three observation patterns under competing risks.

Observation	δ_i	δ_i^*	Likelihood contribution
Cause 1 failure	1	0	$\Pr(X_i = T_i, Y_i > T_i) = f^{(1)}(T_i)$
Cause 2 failure	0	1	$\Pr(X_i > T_i, Y_i = T_i) = f^{(2)}(T_i)$
Censoring	0	0	$\Pr(X_i > T_i, Y_i > T_i) = S_T(T_i)$

can be either monotone increasing or decreasing with respect to the copula parameter θ. This implies that the MLE can have a peak at an extreme $\theta \approx -\infty$ or $\theta \approx +\infty$. The profile likelihood is useful only when it has a peak within a reasonable range of parameters. We also found that whether the profile likelihood has a reasonable peak or not is determined by chance (determined by data).

This phenomenon implies that, the estimation of θ should be more regarded as a model selection process. A similar phenomenon occurs under the SNBP model (2) with respect to the parameter α_0. Hence, the dependence parameter θ or α_0 should be regarded as a fixed parameter after selection by the profile likelihood (if its reasonable peak exists). If one truly wishes to estimate θ or α_0, one needs to impose strong model restrictions that shall be discussed in Section 4.2. After our numerical experiments, we reached three different versions of the bivariate Pareto models where the parameter spaces are restricted to be identifiable. Below, we introduce such models.

4.2. *Frank model with common margins*

Under the Frank model, we consider the common marginal survival function $S_1(t) = S_2(t) = S(t) = (1+\alpha t)^{-\gamma}$. This model may be suitable when two causes are exchangeable (e.g., Navarro, Ruiz, Sandoval 2008). Then, the log-likelihood function is

$$\ell_n(\varphi) = \sum_{i=1}^{n} (\delta_i + \delta_i^*)\{\log h(T_i) + \log S(T_i) - \theta S(T_i) + \log(e^{-\theta S(T_i)} - 1) - \log(e^{-\theta} - 1)$$
$$+ \theta S_T(T_i)\} + \sum_{i=1}^{n} (1 - \delta_i - \delta_i^*)\left(-\log\theta + \log\left[-\log\left\{1 + \frac{(e^{-\theta S(T_i)} - 1)^2}{e^{-\theta} - 1}\right\}\right]\right),$$

where $\varphi = (\theta, \alpha, \gamma)$, $S(t) = (1+\alpha t)^{-\gamma}$, and $h(t) = \alpha\gamma/(1+\alpha t)$.

We apply the randomized Newton-Raphson (RNR) algorithm (Hu and Emura 2015) to obtain the MLE. Appendix 1 provides the concrete algorithm of the RNR and the explicit expressions of the partial derivatives of $\ell_n(\varphi)$. The algorithm does not depend on a specific platform (e.g., R) and is fully reproducible by any computational environment. The standard error (SE) is obtained from the second derivatives of $\ell_n(\varphi)$ which are available from the last step of the RNR algorithm.

The MLE for the median of X is $\hat{M}_1 = \hat{S}_1^{-1}(0.5) = (2^{1/\hat{\gamma}} - 1)/\hat{\alpha}$. The SE is

$$SE(\hat{M}_1) = \sqrt{\left\{0, \frac{\partial g(\varphi)}{\partial \alpha}, \frac{\partial g(\varphi)}{\partial \gamma}\right\}^{\mathrm{T}} \left\{-\frac{\partial^2 \ell_n(\varphi)}{\partial\varphi\partial\varphi^{\mathrm{T}}}\right\}^{-1} \left\{0, \frac{\partial g(\varphi)}{\partial \alpha}, \frac{\partial g(\varphi)}{\partial \gamma}\right\}}\Bigg|_{\varphi=\hat{\varphi}},$$

where

$$\frac{\partial g(\varphi)}{\partial \alpha} = \frac{1 - 2^{1/\gamma}}{\alpha^2}, \quad \frac{\partial g(\varphi)}{\partial \gamma} = \frac{-2^{1/\gamma}\log 2}{\alpha\gamma^2}.$$

The confidence interval (CI) is obtained based on the normal approximation.

Our simulations shall reveal that the above SE yields under-estimation and hence the CI becomes liberal even for very large samples. Therefore, we also consider the parametric bootstrap method (Efron and Tibshirani 1993) to compute the SE and CI. The concrete algorithm is given in Appendix 2.

4.3. Frank model with fixed θ

Given an assume value of θ, we estimate other four parameters $\varphi = (\alpha_1, \alpha_2, \gamma_1, \gamma_2)$. The idea of the assumed value of θ is not new. This has been used for sensitivity analysis in semi-parametric approaches (Braekers and Veraverbeke 2005; Chen 2010; de Uña-Álvarez and Veraverbeke 2013; 2017; Moradian, Denis Larocque, Bellavance 2017).

Under this setting, the log-likelihood function is

$$\begin{aligned}\ell_n(\varphi) = &\sum_{i=1}^{n} \delta_i \{\log h_1(T_i) + \log S_1(T_i) - \theta S_1(T_i) + \log(e^{-\theta S_1(T_i)} - 1)\\ &- \log(e^{-\theta} - 1) + \theta S_T(T_i)\}\\ &+ \sum_{i=1}^{n} \delta_i^* \{\log h_2(T_i) + \log S_2(T_i) - \theta S_2(T_i) + \log(e^{-\theta S_2(T_i)} - 1)\\ &- \log(e^{-\theta} - 1) + \theta S_T(T_i)\}\\ &+ \sum_{i=1}^{n} (1 - \delta_i - \delta_i^*) \left(-\log\theta + \log\left[-\log\left\{ 1 + \frac{(e^{-\theta S_1(T_i)} - 1)(e^{-\theta S_2(T_i)} - 1)}{e^{-\theta} - 1} \right\}\right]\right),\end{aligned}$$

where $S_j(t) = (1 + \alpha_j t)^{-\gamma_j}$ and $h_j(t) = \alpha_j \gamma_j / (1 + \alpha_j t)$, $j = 1, 2$.

Appendix 3 provides the RNR algorithm to obtain the MLE, along with all the explicit expressions of the partial derivatives of $\ell_n(\varphi)$. The SE is obtained from these expressions which arise from the last step of the RNR algorithm.

The MLE for the median of X is $\hat{M}_1 = \hat{S}_1^{-1}(0.5) = (2^{1/\hat{\gamma}_1} - 1)/\hat{\alpha}_1$. The SE is

$$SE(\hat{M}_1) = \sqrt{\left\{ \frac{\partial g(\varphi)}{\partial \alpha_1}, 0, \frac{\partial g(\varphi)}{\partial \gamma_1}, 0 \right\}^{\mathrm{T}} \left\{ -\frac{\partial^2 \ell_n(\varphi)}{\partial \varphi \partial \varphi^{\mathrm{T}}} \right\}^{-1} \left\{ \frac{\partial g(\varphi)}{\partial \alpha_1}, 0, \frac{\partial g(\varphi)}{\partial \gamma_1}, 0 \right\}\Bigg|_{\varphi=\hat{\varphi}}},$$

where

$$\frac{\partial g(\varphi)}{\partial \alpha_1} = \frac{1 - 2^{1/\gamma_1}}{\alpha_1^2}, \quad \frac{\partial g(\varphi)}{\partial \gamma_1} = \frac{-2^{1/\gamma_1} \log 2}{\alpha_1 \gamma_1^2}.$$

The CI is obtained by the normal approximation.

4.4. The SNBP model with fixed α_0

Considering the SNBP model (2), we assume that the parameter α_0 is known and estimate the parameter $\varphi = (\alpha_1, \alpha_2, \gamma)$. Under the SNBP model, the cause-specific hazards are defined as

$$h^{(1)}(t) = \frac{f^{(1)}(t)}{S_T(t)} = \frac{\gamma(\alpha_1 + \alpha_0 t)}{1 + \alpha_1 t + \alpha_2 t + \alpha_0 t^2}, \quad h^{(2)}(t) = \frac{f^{(2)}(t)}{S_T(t)} = \frac{\gamma(\alpha_2 + \alpha_0 t)}{1 + \alpha_1 t + \alpha_2 t + \alpha_0 t^2}.$$

Then, the log-likelihood function is

$$\begin{aligned}\ell_n(\varphi) &= \sum_{i=1}^{n} [\delta_i \log\{h^{(1)}(T_i)\} + \delta_i^* \log\{h^{(2)}(T_i)\} + \log S_T(T_i)]\\ &= (m + m^*)\log\gamma + \sum_{i=1}^{n} \delta_i \{\log(\alpha_1 + \alpha_0 T_i) - \log(1 + \alpha_1 T_i + \alpha_2 T_i + \alpha_0 T_i^2)\}\end{aligned}$$

$$+\sum_{i=1}^{n}\delta_i^*\{\log(\alpha_2+\alpha_0 T_i)-\log(1+\alpha_1 T_i+\alpha_2 T_i+\alpha_0 T_i^2)\}$$

$$-\gamma\sum_{i=1}^{n}\log(1+\alpha_1 T_i+\alpha_2 T_i+\alpha_0 T_i^2),$$

where $m=\Sigma_{i=1}^n\delta_i$ and $m^*=\Sigma_{i=1}^n\delta_i^*$.

Appendix 4 provides the RNR algorithm to maximize $\ell_n(\varphi)$, along with the explicit expressions of the partial derivatives of $\ell_n(\varphi)$. The SE is obtained from these expressions which arise from the last step of the RNR algorithm.

The MLE for the median of X is $\hat{M}_1=\hat{S}_1^{-1}(0.5)=(2^{1/\hat{\gamma}}-1)/\hat{\alpha}_1$. The SE is

$$SE(\hat{M}_1)=\sqrt{\left\{\frac{\partial g(\varphi)}{\partial\alpha_1},0,\frac{\partial g(\varphi)}{\partial\gamma}\right\}^{\mathrm{T}}\left\{-\frac{\partial^2\ell_n(\varphi)}{\partial\varphi\partial\varphi^{\mathrm{T}}}\right\}^{-1}\left\{\frac{\partial g(\varphi)}{\partial\alpha_1},0,\frac{\partial g(\varphi)}{\partial\gamma}\right\}}\Bigg|_{\varphi=\hat{\varphi}},$$

where

$$\frac{\partial g(\varphi)}{\partial\alpha_1}=\frac{1-2^{1/\gamma}}{\alpha_1^2},\quad\frac{\partial g(\varphi)}{\partial\gamma}=\frac{-2^{1/\gamma}\log 2}{\alpha_1\gamma^2}.$$

The CI is obtained by the normal approximation.

5. Model selection and diagnostic

We have introduced three different models (Frank model with common margins, Frank model with fixed θ, and the SNBP model with fixed α_0). For a given dataset, one needs to select a suitable model. Since the three models have different number of parameters, the information theoretic criterion is useful for model comparison.

Let k be the number of unknown parameters in the model. The Akaike information criterion (AIC) and the Bayesian information criterion (BIC) are defined as

$$\mathrm{AIC}=2k-2\ell_n(\hat{\varphi}),\ \ \mathrm{BIC}=k\log n-2\ell_n(\hat{\varphi}).$$

The preferred model is the one with the minimum AIC (or BIC) value. Note that AIC (or BIC) is insufficient to validate the model as it simply tells the best model among the candidates. If all the candidate models fit poorly, one may still select the poor model.

Therefore, we consider a model-diagnostic procedure that was previously employed by Escarela and Carrière (2003), Shih (2016), and Shih and Emura (2018). The sub-distribution functions for Cause 1 and Cause 2 are defined as

$$F^{(1)}(t)=\int_0^t f^{(1)}(z)dz,\ F^{(2)}(t)=\int_0^t f^{(2)}(z)dz.$$

The parametric forms of $F^{(j)}(t)$, $j=1,2$ are

$$F^{(j)}(t)=\int_0^t h_j(z)S_j(z)\frac{e^{-\theta\{S_j(z)-S_T(z)\}}(e^{-\theta S_{3-j}(z)}-1)}{e^{-\theta}-1}dz,\ (Frank)$$

$$F^{(j)}(t)=\int_0^t \gamma(\alpha_j+\alpha_0 z)(1+\alpha_1 z+\alpha_2 z+\alpha_0 z^2)^{-\gamma-1}dz.\ (SNBP)$$

These two sub-distribution functions do not have closed-form. Therefore, one needs to compute the sub-distribution functions numerically (e.g., by using R *integrate* function). The parametric estimators for $F^{(j)}(t)$, $j = 1, 2$ are obtained by plugging the MLE. If the model fits to the data well, the parametric estimator and the nonparametric estimator of $F^{(j)}(t)$ (Lawless 2003, p.437) shall be close to each other. One can use the R *cmprsk* package to compute the nonparametric estimator.

6. Simulation

We conduct extensive simulation studies to examine the performance of the proposed likelihood-based methods.

We generate data (X_i, Y_i) for $i = 1, 2, \ldots, n$ from the Frank copula model with the common margins or the SNBP model with fixed α_0. We also generate independent censoring time $C_i \sim U(0, w)$, where $w > 0$ is a constant to control censoring percentages. Then, we obtain the data $(T_i, \delta_i, \delta_i^*)$ by letting $T_i = \min(X_i, Y_i, C_i)$, $\delta_i = \mathbf{I}(T_i = X_i)$ and $\delta_i^* = \mathbf{I}(T_i = Y_i)$ for $i = 1, 2, \ldots, n$. Based on the generated data, we compute the MLE, SE, and 95% CI. We also count the number of iterations to assess the convergence speed of the RNR algorithm. Our simulation results are based on 200 repetitions.

6.1. Results under the Frank model with common margins

Table 2 shows the performance of the MLE. When the censoring is present, estimates are somewhat biased even for large samples. In addition, the MSEs for $\hat{\theta}$ do not properly decrease when the sample size increases. This represents the difficulty of estimating parameters based on competing risks data, even if one assumes the common margins. When censoring is absent (0% censoring), all estimates are nearly unbiased, and the MSEs decrease with the increased sample sizes. The unpleasant results on the biases and MSEs are, in fact, due to a few outlying values. Hence, we shall see the coverage rate of the CIs, which are less affected by outliers.

Table 2 also shows the convergence speed of the RNR algorithm. The average number of the NR iterations is 10, and the average number of the randomizations varies from 1 (0% censoring) to 10 (40% censoring). Hence, in the worst case, $10 \times 10 = 100$ iterations are required until convergence. This is not the drawback of the algorithm itself, but is due to the flat shape

Table 2. Simulation results under the Frank model with common margins.

			$\hat{\theta}$		$\hat{\alpha}$		$\hat{\gamma}$			
Par.	Prop. (%)	n	Mean	MSE	Mean	MSE	Mean	MSE	AI	AR
$\alpha = 1.0$	$X_i = T_i(50\%)$	1000	5.710	17.03	1.081	0.239	0.992	0.021	8.540	2.105
$\gamma = 1.0$	$Y_i = T_i(50\%)$	1500	5.348	16.07	1.103	0.313	0.977	0.021	8.465	0.505
$\theta = 5.0$	$C_i = T_i(0\%)$	2000	5.677	9.847	1.057	0.124	1.000	0.010	8.665	0.130
	$X_i = T_i(40\%)$	1000	2.435	46.94	1.702	2.034	0.821	0.157	9.610	9.068
	$Y_i = T_i(40\%)$	1500	2.125	56.35	1.856	2.474	0.783	0.179	10.08	6.455
	$C_i = T_i(20\%)$	2000	1.599	52.60	1.891	2.668	0.773	0.183	9.275	7.965
$\tau_\theta = 0.5$	$X_i = T_i(30\%)$	1000	− 2.136	419.7	2.046	3.012	0.717	0.239	11.18	10.92
$M_1 = 1.0$	$Y_i = T_i(30\%)$	1500	0.265	191.0	2.048	3.113	0.754	0.225	11.22	10.28
	$C_i = T_i(40\%)$	2000	0.308	137.5	2.023	3.052	0.740	0.228	10.84	11.44

AI = the average iteration number until convergence, AR = the average randomization number until convergence, Par. = parameter, Prop. = proportion of events. $MSE(\hat{\alpha}) = E(\hat{\alpha} - \alpha)^2$, $MSE(\hat{\gamma}) = E(\hat{\gamma} - \gamma)^2$, $MSE(\hat{\theta}) = E(\hat{\theta} - \theta)^2$.

Table 3. Simulation results under the Frank model with common margins.

(i) Hessian method			$\hat{\theta}$			$\hat{\alpha}$			$\hat{\gamma}$		
Par.	Prop. (%)	n	SD	SE	CP	SD	SE	CP	SD	SE	CP
$\alpha = 1.0$	$X_i = T_i(50\%)$	1000	4.075	3.290	0.900	0.483	0.153	0.915	0.146	0.095	0.935
$\gamma = 1.0$	$Y_i = T_i(50\%)$	1500	4.004	2.489	0.890	0.551	0.126	0.900	0.144	0.075	0.930
$\theta = 5.0$	$C_i = T_i(0\%)$	2000	3.072	2.288	0.950	0.348	0.107	0.930	0.102	0.066	0.945
	$X_i = T_i(40\%)$	1000	6.369	2.933	0.725	1.245	0.324	0.725	0.354	0.121	0.745
	$Y_i = T_i(40\%)$	1500	6.952	2.746	0.670	1.323	0.289	0.675	0.364	0.091	0.665
	$C_i = T_i(20\%)$	2000	6.422	2.065	0.625	1.372	0.252	0.660	0.363	0.079	0.680
$\tau_\theta = 0.5$	$X_i = T_i(30\%)$	1000	19.25	149.7	0.735	1.388	0.615	0.710	0.399	0.199	0.630
$M_1 = 1.0$	$Y_i = T_i(30\%)$	1500	13.02	16.77	0.735	1.423	0.452	0.660	0.407	0.146	0.630
	$C_i = T_i(40\%)$	2000	10.77	3.500	0.685	1.420	0.402	0.645	0.401	0.137	0.620

(ii) Bootstrap method			$\hat{\theta}$			$\hat{\alpha}$			$\hat{\gamma}$		
Par.	Prop. (%)	n	SD	SE	CP	SD	SE	CP	SD	SE	CP
$\alpha = 1.0$	$X_i = T_i(50\%)$	1000	4.075	4.596	0.975	0.483	0.640	0.970	0.146	0.184	0.990
$\gamma = 1.0$	$Y_i = T_i(50\%)$	1500	4.004	3.887	0.960	0.551	0.491	0.955	0.144	0.146	0.985
$\theta = 5.0$	$C_i = T_i(0\%)$	2000	3.072	3.633	0.985	0.348	0.409	0.940	0.102	0.126	0.985
	$X_i = T_i(40\%)$	1000	6.369	6.459	0.950	1.245	1.240	0.950	0.354	0.347	0.955
	$Y_i = T_i(40\%)$	1500	6.952	6.347	0.950	1.323	1.255	0.940	0.364	0.339	0.950
	$C_i = T_i(20\%)$	2000	6.422	6.055	0.940	1.372	1.251	0.960	0.363	0.334	0.970
$\tau_\theta = 0.5$	$X_i = T_i(30\%)$	1000	19.25	14.50	0.945	1.388	1.289	0.960	0.399	0.396	0.945
$M_1 = 1.0$	$Y_i = T_i(30\%)$	1500	13.02	13.58	0.950	1.423	1.358	0.930	0.407	0.383	0.955
	$C_i = T_i(40\%)$	2000	10.77	12.62	0.920	1.420	1.323	0.905	0.401	0.367	0.900

SD = sample standard deviation, SE = standard error, CP = coverage probability of the 95% CI, Par. = parameter, Prop. = proportion of events.

of the log-likelihood function in the presence of censoring. The RNR algorithm always converged for all the 200 repetitions.

Table 3 shows the performance of the SEs and 95% CIs. The SEs based on the Hessian tends to underestimate the SDs. Consequently, the CIs show under-coverage (coverage rate less than the nominal 0.95). On the other hand, the SEs based on the bootstrap give good agreements with the SDs, leading to correct coverage rates of the CIs. Overall, the bootstrap shows much better performance than the Hessian. However, we must notice that the bootstrap requires the assumption of the uniform censoring distribution (Appendix 2). Such an assumption is not always true in real data analyses.

6.2. Results under the SNBP model with fixed α_0

Tables 4 and 5 show the results of estimation. Under the SNBP model, the average number of the NR iterations varies from 3 to 5, and the average number of the randomizations is less than 1. Thus, the convergence speed is quick. All the estimates are almost unbiased and the MSEs decrease when the sample sizes increase. In addition, the SEs are very close to the SDs in all configurations. In most cases, the coverage probabilities of CIs are all close to the nominal 95% level. Similar simulation results may be obtained under the Frank model with fixed θ.

Our estimation procedure in Section 4.4 does not always guarantee the constraint $0 \le \alpha_0 \le (\gamma + 1)\alpha_1\alpha_2$. Table 4 reports the proportion of violating the constraint in the simulations. Fortunately, the constraint is met in most cases or can be met by large samples.

Table 4. Simulation results under the SNBP model given α_0.

			$\hat{\alpha}_1$		$\hat{\alpha}_2$		$\hat{\gamma}$		$\hat{M}_1$				
Parameter	Prop. (%)	n	Mean	MSE	Mean	MSE	Mean	MSE	Mean	MSE	AI	AR	Con.
$\alpha_0 = 0.5$	$X_i = T_i$(52%)	100	1.014	0.172	0.507	0.060	2.179	0.412	0.429	0.007	4.5	0.1	16%
$\alpha_1 = 1.0$	$Y_i = T_i$(28%)	200	0.998	0.099	0.501	0.035	2.114	0.245	0.426	0.003	4.1	0.0	8%
$\alpha_2 = 0.5$	$C_i = T_i$(20%)	300	0.993	0.062	0.494	0.021	2.059	0.127	0.426	0.002	4.0	0.0	5%
$\gamma = 2.0$	$X_i = T_i$(45%)	100	1.056	0.310	0.525	0.083	2.241	0.710	0.433	0.008	4.6	0.2	24%
	$Y_i = T_i$(25%)	200	1.025	0.153	0.510	0.046	2.127	0.347	0.427	0.004	4.2	0.0	12%
	$C_i = T_i$(30%)	300	1.002	0.085	0.499	0.026	2.075	0.208	0.428	0.003	4.1	0.1	7%
	$X_i = T_i$(39%)	100	1.134	0.546	0.585	0.179	2.278	1.167	0.443	0.010	4.8	0.3	23%
	$Y_i = T_i$(21%)	200	1.066	0.226	0.534	0.076	2.149	0.553	0.427	0.004	4.3	<0.1	15%
	$C_i = T_i$(40%)	300	1.033	0.133	0.511	0.038	2.091	0.344	0.427	0.003	4.2	<0.1	10%
$\alpha_0 = 0.0$	$X_i = T_i$(54%)	100	2.171	0.853	1.077	0.239	0.510	0.011	1.573	0.137	4.1	0.0	0%
$\alpha_1 = 2.0$	$Y_i = T_i$(26%)	200	2.076	0.436	1.016	0.107	0.505	0.006	1.557	0.067	3.9	0.0	0%
$\alpha_2 = 1.0$	$C_i = T_i$(20%)	300	2.018	0.226	0.995	0.063	0.503	0.003	1.553	0.043	3.8	0.0	0%
$\gamma = 0.5$	$X_i = T_i$(47%)	100	2.138	1.063	1.067	0.331	0.532	0.021	1.591	0.228	4.4	0.3	0%
	$Y_i = T_i$(23%)	200	2.067	0.381	1.007	0.096	0.509	0.007	1.548	0.078	4.0	0.1	0%
	$C_i = T_i$(30%)	300	2.016	0.234	1.009	0.070	0.505	0.004	1.548	0.051	3.9	0.1	0%
	$X_i = T_i$(40%)	100	2.223	1.556	1.080	0.338	0.548	0.043	1.593	0.218	4.8	1.0	0%
	$Y_i = T_i$(20%)	200	2.103	0.564	1.033	0.157	0.515	0.012	1.550	0.087	4.2	0.1	0%
	$C_i = T_i$(40%)	300	2.005	0.319	1.009	0.095	0.511	0.006	1.557	0.058	4.0	0.1	0%
$\alpha_0 = 2.0$	$X_i = T_i$(36%)	100	4.169	2.868	5.108	3.586	1.043	0.053	0.261	0.004	4.1	<0.1	0%
$\alpha_1 = 4.0$	$Y_i = T_i$(44%)	200	4.077	1.463	5.105	2.245	1.020	0.034	0.260	0.002	3.9	<0.1	0%
$\alpha_2 = 5.0$	$C_i = T_i$(20%)	300	3.918	0.914	4.884	1.581	1.034	0.025	0.258	0.001	3.8	<0.1	0%
$\gamma = 1.0$	$X_i = T_i$(31%)	100	4.323	4.277	5.296	6.299	1.070	0.120	0.262	0.004	4.2	<0.1	0%
	$Y_i = T_i$(39%)	200	4.102	1.896	5.187	3.189	1.030	0.053	0.262	0.002	4.0	<0.1	0%
	$C_i = T_i$(30%)	300	3.906	1.426	4.823	2.059	1.061	0.050	0.257	0.001	4.0	<0.1	0%
	$X_i = T_i$(27%)	100	4.474	7.728	5.542	10.98	1.116	0.207	0.266	0.006	4.7	0.2	1%
	$Y_i = T_i$(33%)	200	4.211	3.636	5.313	5.149	1.081	0.148	0.262	0.003	4.4	0.1	1%
	$C_i = T_i$(40%)	300	3.963	1.566	4.922	2.362	1.053	0.066	0.261	0.002	4.2	0.0	0%

AI = the average iteration number until convergence, AR = the average randomization number until convergence, Con. = the proportion of violating the constraint $0 \le \alpha_0 \le (\gamma+1)\alpha_1\alpha_2$, $MSE(\hat{\alpha}_1) = E(\hat{\alpha}_1 - \alpha_1)^2$, $MSE(\hat{\alpha}_2) = E(\hat{\alpha}_2 - \alpha_2)^2$, $MSE(\hat{\gamma}) = E(\hat{\gamma} - \gamma)^2$.

7. Data analysis

We analyze the data on time-to-death from 483 prostate cancer patients (Andrews and Herzberg 1985). There are 125 deaths from prostate cancer (Cause 1), 219 deaths from other diseases (Cause 2) and 139 censorings (survived until the study end). Let X_i be time-to-death by prostate cancer, Y_i be time-to-death by other diseases, and C_i be censoring time. The observed data are $T_i = \min(X_i, Y_i, C_i)$, $\delta_i = \mathbf{I}(T_i = X_i)$ and $\delta_i^* = \mathbf{I}(T_i = Y_i)$ for $i = 1, 2, \ldots, 483$.

Under the Frank copula model, we selected the copula parameter $\theta = -5$ ($\tau_\theta = -0.46$) which gives the largest profile log-likelihood value (Figure 2). Under $\theta = -5$, the RNR algorithm converges at the 8th iteration step after 23 randomization steps on initial values. Figure 3 demonstrates that the MLE $(\hat{\alpha}_1, \hat{\alpha}_2, \hat{\gamma}_1, \hat{\gamma}_2)$ attains the maximum of the log-likelihood function.

Under the SNBP model, we selected the dependence parameter $\alpha_0 = 0.00007$ which gives the largest profile log-likelihood value (Figure 2). Kendall's tau is $\tau_{\delta,\hat{\gamma}} = -0.22$ which is obtained by numerically evaluating

$$\tau_{\delta,\hat{\gamma}} = 4\int_0^1\int_0^1 C_{\delta,\hat{\gamma}}(u,v)\,dC_{\delta,\hat{\gamma}}(u,v) - 1,\ \delta = \alpha_0/\hat{\alpha}_1\hat{\alpha}_2.$$

Table 5. Simulation results under the SNBP model given α_0.

			$\hat{\alpha}_1$			$\hat{\alpha}_2$			$\hat{\gamma}$			$\hat{M}_1$		
Parameter	Prop. (%)	n	SD	SE	CP	SD	SE	CP	SD	SE	CP	SD	SE	CP
$\alpha_0 = 0.5$	$X_i = T_i$(52%)	100	0.42	0.43	0.970	0.24	0.25	0.935	0.62	0.62	0.940	0.08	0.08	0.960
$\alpha_1 = 1.0$	$Y_i = T_i$(28%)	200	0.32	0.30	0.940	0.19	0.17	0.935	0.48	0.43	0.935	0.06	0.06	0.960
$\alpha_2 = 0.5$	$C_i = T_i$(20%)	300	0.25	0.24	0.940	0.15	0.14	0.960	0.35	0.34	0.945	0.05	0.05	0.965
$\gamma = 2.0$	$X_i = T_i$(45%)	100	0.56	0.53	0.955	0.29	0.30	0.970	0.81	0.79	0.940	0.09	0.09	0.970
$M_1 = .41$	$Y_i = T_i$(25%)	200	0.39	0.37	0.930	0.21	0.20	0.930	0.58	0.54	0.925	0.06	0.06	0.940
	$C_i = T_i$(30%)	300	0.29	0.29	0.925	0.16	0.16	0.955	0.45	0.43	0.935	0.05	0.05	0.950
	$X_i = T_i$(39%)	100	0.73	0.69	0.945	0.42	0.39	0.970	1.04	1.00	0.915	0.10	0.10	0.965
	$Y_i = T_i$(21%)	200	0.47	0.46	0.940	0.27	0.25	0.930	0.73	0.68	0.910	0.06	0.06	0.960
	$C_i = T_i$(40%)	300	0.36	0.37	0.940	0.20	0.20	0.955	0.58	0.54	0.955	0.05	0.05	0.955
$\alpha_0 = 0.0$	$X_i = T_i$(54%)	100	0.91	0.86	0.950	0.48	0.45	0.945	0.11	0.10	0.945	0.36	0.39	0.960
$\alpha_1 = 2.0$	$Y_i = T_i$(26%)	200	0.66	0.58	0.915	0.33	0.30	0.915	0.08	0.07	0.945	0.25	0.27	0.955
$\alpha_2 = 1.0$	$C_i = T_i$(20%)	300	0.48	0.46	0.935	0.25	0.24	0.950	0.06	0.06	0.955	0.20	0.22	0.970
$\gamma = 0.5$	$X_i = T_i$(47%)	100	1.02	0.95	0.935	0.57	0.50	0.930	0.14	0.13	0.940	0.47	0.41	0.950
$M_1 = 1.5$	$Y_i = T_i$(23%)	200	0.61	0.64	0.965	0.31	0.33	0.960	0.08	0.08	0.955	0.28	0.27	0.970
	$C_i = T_i$(30%)	300	0.48	0.51	0.965	0.26	0.27	0.970	0.06	0.07	0.965	0.22	0.22	0.955
	$X_i = T_i$(40%)	100	1.23	1.12	0.960	0.58	0.57	0.950	0.20	0.17	0.960	0.46	0.43	0.970
	$Y_i = T_i$(20%)	200	0.75	0.75	0.965	0.40	0.39	0.960	0.11	0.11	0.955	0.29	0.29	0.955
	$C_i = T_i$(40%)	300	0.57	0.59	0.955	0.31	0.31	0.955	0.08	0.08	0.970	0.23	0.24	0.935
$\alpha_0 = 2.0$	$X_i = T_i$(36%)	100	1.69	1.74	0.965	1.90	2.09	0.970	0.23	0.26	0.960	0.06	0.06	0.945
$\alpha_1 = 4.0$	$Y_i = T_i$(44%)	200	1.21	1.21	0.950	1.50	1.47	0.960	0.18	0.18	0.945	0.04	0.04	0.955
$\alpha_2 = 5.0$	$C_i = T_i$(20%)	300	0.95	0.95	0.945	1.25	1.15	0.960	0.15	0.15	0.940	0.03	0.03	0.945
$\gamma = 1.0$	$X_i = T_i$(31%)	100	2.05	2.11	0.960	2.50	2.54	0.955	0.34	0.33	0.955	0.06	0.06	0.955
$M_1 = .25$	$Y_i = T_i$(39%)	200	1.38	1.43	0.940	1.78	1.76	0.945	0.23	0.22	0.955	0.05	0.04	0.950
	$C_i = T_i$(30%)	300	1.19	1.11	0.945	1.43	1.34	0.945	0.22	0.19	0.935	0.03	0.03	0.945
	$X_i = T_i$(27%)	100	2.75	2.55	0.945	3.28	3.11	0.940	0.44	0.45	0.945	0.08	0.07	0.960
	$Y_i = T_i$(33%)	200	1.90	1.74	0.915	2.25	2.16	0.935	0.38	0.31	0.925	0.05	0.05	0.960
	$C_i = T_i$(40%)	300	1.25	1.36	0.960	1.54	1.66	0.975	0.25	0.25	0.960	0.04	0.04	0.955

SD = sample standard deviation, SE = average standard error, CP = coverage probability of the 95% CI, $\hat{M}_1 = (2^{1/\hat{\gamma}} - 1)/\hat{\alpha}_1$.

Under $\alpha_0 = 0.00007$, the RNR algorithm converges at the 10th iteration step with 6 randomization steps. Figure 4 reveals that the MLE $(\hat{\alpha}_1, \hat{\alpha}_2, \hat{\gamma})$ attains the maximum of the log-likelihood function.

We summarize the fitted results for the Frank and SNBP models in Table 6. The mean failure times of prostate cancer and other causes are not available under the Frank model since $\hat{\gamma}_1, \hat{\gamma}_2 \leq 1$. In contrast, they are available under the SNBP model since $\hat{\gamma} > 1$. The Frank model produces a slightly larger log-likelihood value than the SNBP model. However, the SNBP model gives a better value of AIC (BIC) due to the smaller degrees of freedom.

Figure 5 displays the model-diagnostic plots. The model-based sub-distribution functions of the two models are very close to the nonparametric sub-distribution functions, indicating good fits for both the Frank and SNBP models.

The same data have been analyzed by Escarela and Carrière (2003) using the Frank model with the Pareto margins that account for covariates. They derived the median lifetimes for the average patients (under average covariates). Under $\theta = -5$, their median times to death are calibrated about 200 for prostate cancer and about 70 for other diseases (Figure 3 of their

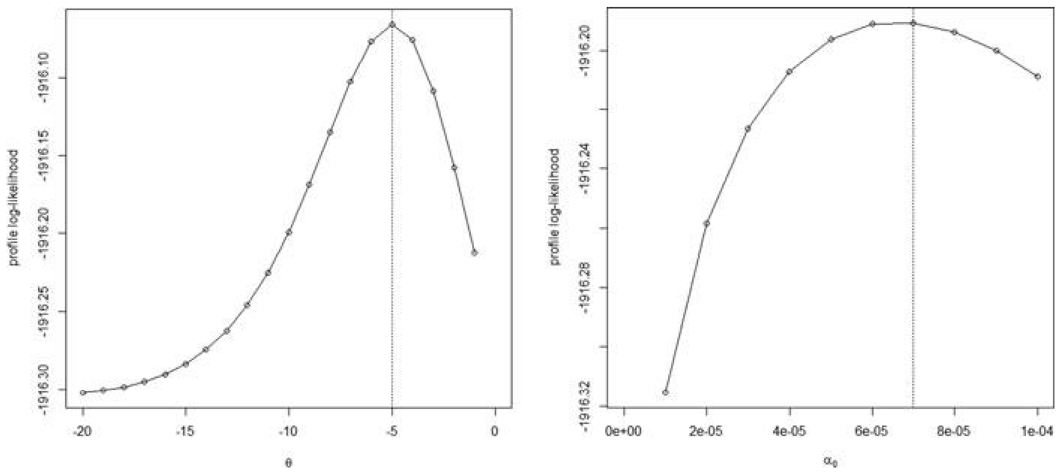

Figure 2. Profile log-likelihood plots for the Frank (left) and SNBP (right) models.

paper). These values are very similar to our results under the Frank model. Therefore, the following discussions focus on the SNBP model.

We use the median failure time to examine how long prostate cancer patients can survival. Under the SNBP model, the median failure time due to prostate cancer is 128.3 months (about 11 years), and the median failure time due to other diseases is 61.8 months (about 5 years).

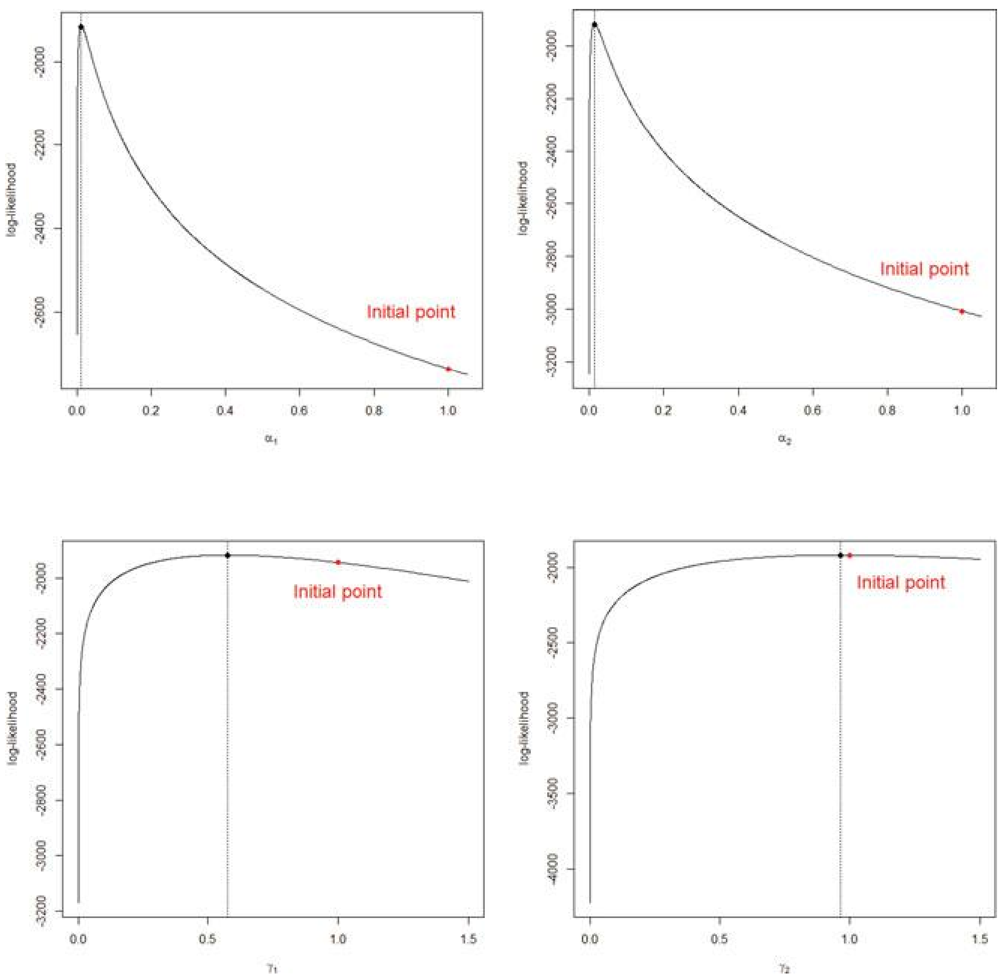

Figure 3. Log-likelihood functions under the Frank model based on the prostate cancer data. The vertical lines are drawn at $\hat{\alpha}_1 = 0.0119$, $\hat{\alpha}_2 = 0.0149$, $\hat{\gamma}_1 = 0.5758$ and $\hat{\gamma}_2 = 0.9647$.

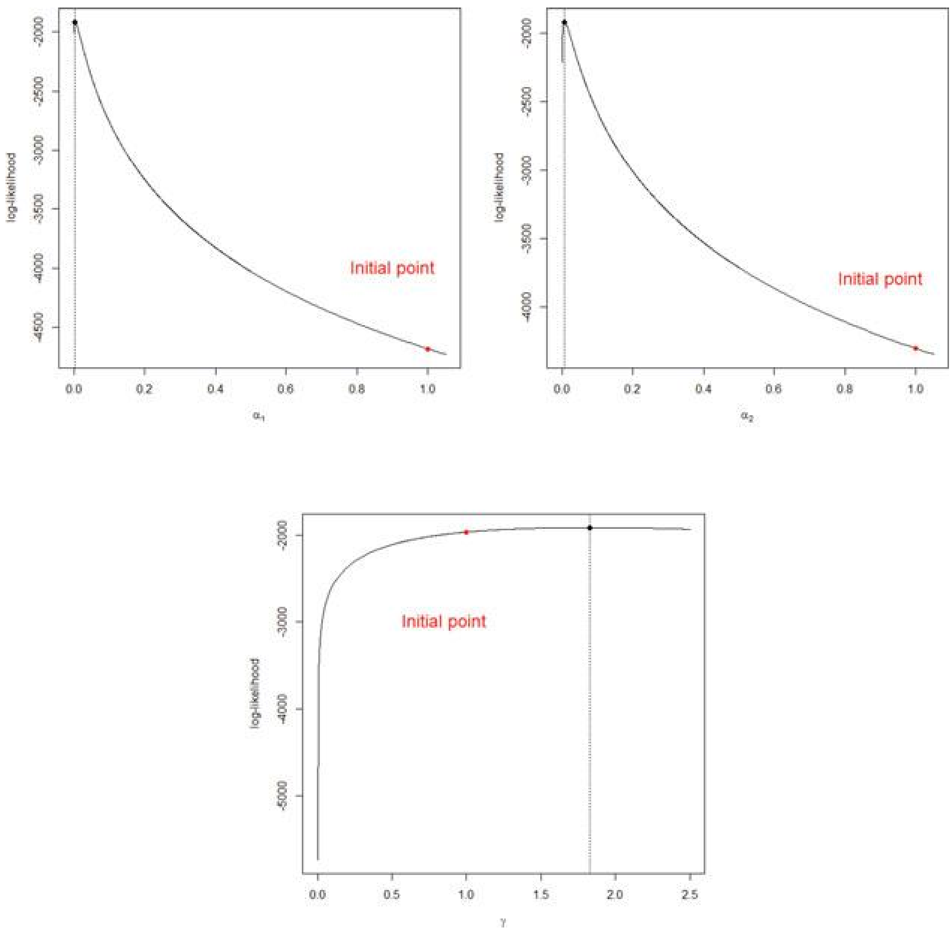

Figure 4. Log-likelihood functions under the SNBP model based on the prostate cancer data. The vertical lines are drawn at $\hat{\alpha}_1 = 0.0036$, $\hat{\alpha}_2 = 0.0075$ and $\hat{\gamma} = 1.8277$.

It is known that patients diagnosed with prostate cancer can still have long life expectancy since prostate cancer is not a fatal disease. So the death is not likely due to the prostate cancer itself.

We provide the fitted density $f_j(t)$ and survival function $S_j(t)$ for $j = 1, 2$ under the SNBP model (Figure 6). In this figure, the locations of median failure times may be useful to compare the survival difference between $j = 1$ and $j = 2$.

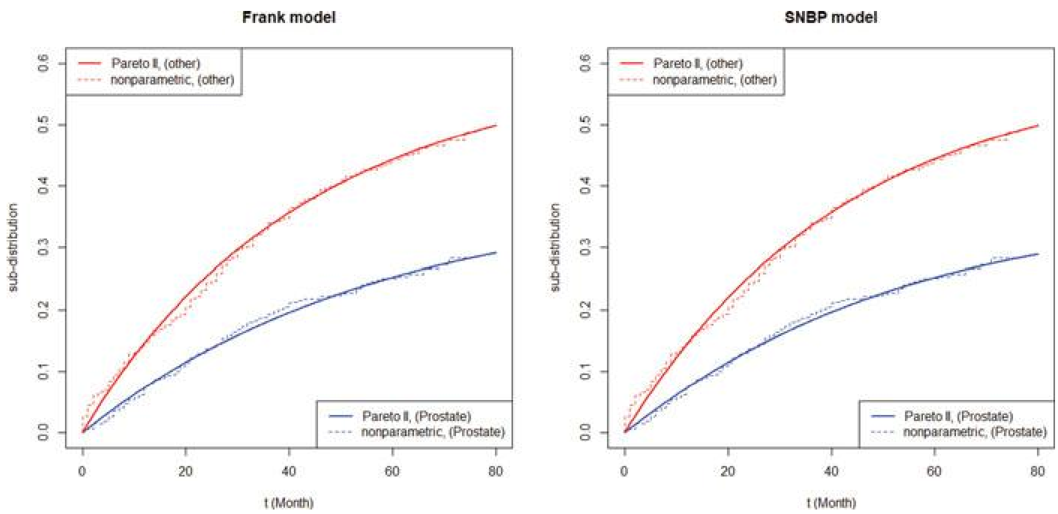

Figure 5. Model-diagnostic plots for the Frank model (left) and the SNBP model (Right).

Table 6. Results of fitting two models based on the prostate cancer data.

(i) **Frank model with** $\theta = -5\ (\tau_\theta = -0.46)$										
	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\gamma}_1$	$\hat{\gamma}_2$	$\hat{M}_1$	$\hat{M}_2$	—	—	$\ell_n(\hat{\varphi})$	AIC	BIC
Est. (CI)	0.0119 (0.0031, 0.0452)	0.0149 (0.0055, 0.0403)	0.5758 (0.1983, 1.6722)	0.9647 (0.4444, 2.0942)	196.2 (111.7, 344.7)	70.66 (57.53, 86.79)	—	—	−1916.1	3840.1 (df = 4)	3856.9 (df = 4)
(ii) SNBP model with $\alpha_0 = 0.00007\ (\tau_{\delta,\gamma} = -0.22)$											
	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\gamma}$	—	$\hat{M}_1$	$\hat{M}_2$	$E(X)$	$E(Y)$	$\ell_n(\hat{\varphi})$	AIC	BIC
Est. (CI)	0.0036 (0.0021, 0.0062)	0.0075 (0.0048, 0.0117)	1.8277 (1.3632, 2.4506)	—	128.3 (95.8, 171.9)	61.79 (51.01, 74.85)	336.2 (243.8, 463.5)	161.9 (119.6, 219.0)	−1916.2	3838.4 (df = 3)	3850.9 (df = 3)

Remark: CI denotes the 95% confidence interval, $\hat{M}_1$ and $\hat{M}_2$ are the estimates of medians of X and Y, respectively. Smaller AIC and BIC correspond to a better fit.

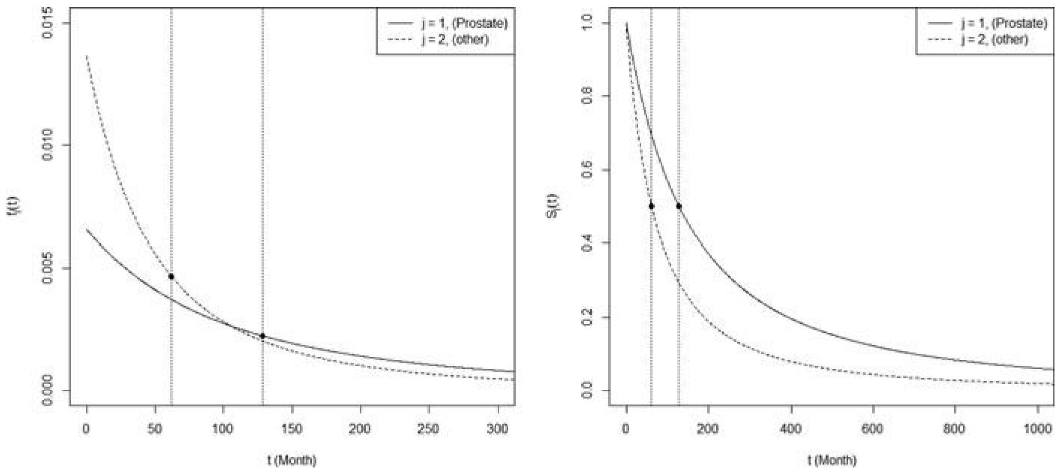

Figure 6. The Pareto type II marginal densities ($f_j(t)$, $j = 1, 2$) and survival functions ($S_j(t)$, $j = 1, 2$) under $(\alpha_1, \alpha_2, \gamma) = (0.0036, 0.0075, 1.8277)$. Vertical lines are drawn at their median values.

8. Conclusion and future works

Under the suggested bivariate Pareto models, we provide the computational algorithms, numerical performances, and model-diagnostic procedures for analyzing competing risks data. Through the real data analysis on the prostate cancer data, we have shown that both the Frank and SNBP models fit equally well. Although the SNBP model is restricted to the common shape parameter, it still gives a decent fit. This peculiar phenomenon may be due to the dependence structure of SNBP model suitable for the prostate cancer data. The Frank model with the common margins is even more restrictive, but it is the only model that allows us to estimate the dependence parameter with the SE and CI.

It remains a challenging problem to estimate the dependence parameter (copula parameter) in both the Frank model and SNBP model. As we claimed, the profile likelihood often does not have a peak under competing risks unless the restriction of common margins is imposed. Even if the common margin assumption is made, one needs large samples and the absence of censoring to correctly estimate a dependence parameter. If covariates are available, it becomes easier to estimate dependence (Heckman and Honore 1989). Escarela and Carrière (2003) estimated a dependence parameter in the presence of covariates. For high-dimensional covariates, Emura and Chen (2016) proposed an algorithm to estimate the dependence parameter by maximizing a concordance index between the survival time and covariates.

The important issue related to covariates is how to formulate the covariate effects. Escarela and Carrière (2003) suggested a regression model for scale parameters given by

$$\alpha_1 = \exp(\beta_{10} + \beta_1^T \mathbf{Z}_1), \ \alpha_2 = \exp(\beta_{20} + \beta_2^T \mathbf{Z}_2).$$

If the dimensions of covariates are high, one may use compound covariates to reduce their dimensions via univariate regression (Emura et al. 2017b). That is, one can form regression models on compound covariates, defined as

$$\alpha_1 = \exp\{\beta_{10} + \beta_1(\hat{\mathbf{c}}_1^T \mathbf{Z}_1)\}, \ \alpha_2 = \exp\{\beta_{20} + \beta_2(\hat{\mathbf{c}}_2^T \mathbf{Z}_2)\},$$

where the weight vectors $\hat{\mathbf{c}}_1$ and $\hat{\mathbf{c}}_2$ are determined by univariate regression.

This paper focuses on the estimation of parameters in the bivariate Pareto models. However, it may be interesting to consider a prediction analysis using the models. With the bivariate Pareto models, one lifetime can be predicted by the other. Recent works related to prediction analysis (Noughabi and Kayid 2017) briefly mentioned the potentials of using the bivariate Pareto type I model for prediction.

Acknowledgements

The authors thank two anonymous reviewers for their helpful comments that improved the paper. This work is supported by the research grants funded by the government of Taiwan (MOST 105-2118-M-008-003-MY2).

Funding

Ministry of Science and Technology, Taiwan MOST 105-2118-M-008-003-MY2.

References

Andrews, D. F., and A. M. Herzberg. 1985. *Data: A collection of problems from many fields for the student and research worker*. New York: Springer.

Arnold, B. C. 2014. Pareto distribution. *Wiley StatsRef: Statistics Reference Online*, doi:10.1002/9781118445112.

Basu, A. P., and J. K. Ghosh. 1978. Identifiability of the multinormal and other distributions under competing risks model. *Journal of Multivariate Analysis* 8:413–429. doi:10.1016/0047-259X(78)90064-7.

Braekers, R., and N. Veraverbeke. 2005. A copula-graphic estimator for the conditional survival function under dependent censoring. *The Canadian Journal of Statistics* 33:429–447. doi:10.1002/cjs.5540330308.

Chen, Y. H. 2010. Semiparametric marginal regression analysis for dependent competing risks under an assumed copula. *Journal of the Royal Statistical Society, Ser. B* 72:235–251. doi:10.1111/j.1467-9868.2009.00734.x.

David, H. A., and M. L. Moeschberger. 1978. *The theory of competing risks (No. 39)*. London: Griffin.

de Uña-Álvarez, J., and N. Veraverbeke. 2013. Generalized copula-graphic estimator. *Test* 22 (2):343–360. doi:10.1007/s11749-012-0314-2.

de Uña-Álvarez, J., and N. Veraverbeke. 2017. Copula-graphic estimation with left-truncated and right-censored data. *Statistics* 51 (2):387–403. doi:10.1080/02331888.2016.1274898.

Efron, B., and R. J. Tibshirani. 1993. An Introduction to the Bootstrap. Boca Raton, New York.

Emura, T., and Y. H. Chen. 2016. Gene selection for survival data under dependent censoring: A copula based approach. *Statistical Methods in Medical Research* 25 (6):2840–2857. doi:10.1177/0962280214533378.

Emura, T. and Y. H. Chen. 2018. Survival analysis with dependent censoring, copula-based approaches. *JSS Research Series in Statistics*, Berlin: Springer.

Emura, T., M. Nakatochi, S. Matsui, H. Michimae, and V. Rondeau. 2017b. Personalized dynamic prediction of death according to tumour progression and high-dimensional genetic factors: Meta-analysis with a joint model. *Statistical Methods in Medical Research* doi: 10.1177/0962280216688032. doi:10.1177/0962280216688032.

Emura, T., M. Nakatochi, K. Murotani, and V. Rondeau. 2017a. A joint frailty-copula model between tumour progression and death for meta-analysis. *Statistical Methods in Medical Research* 26 (6):2649–66. doi:10.1177/0962280215604510.

Escarela, G., and J. F. Carrière. 2003 Fitting competing risks with an assumed copula. *Statistical Methods in Medical Research* 12 (4):333–349. doi:10.1191/0962280203sm335ra.

Frank, M. J. 1979. On the simultaneous associativity of $f(x,y)$ and $x+y–f(x,y)$. *Aequationes Matbematicae*, 19, 194–226. doi:10.1007/BF02189866.

Heckman, J. J., and B. E. Honore. 1989. The identifiability of the competing risks models. *Biometrika* 76:325–330. doi:10.1093/biomet/76.2.325.

Hu, Y. H., and T. Emura. 2015. Maximum likelihood estimation for a special exponential family under random double-truncation. *Computational Statistics* 30 (4):1199–1229. doi:10.1007/s00180-015-0564-z.

Lawless, J. F. 2003. *Statistical models and methods for lifetime data (2nd ed.)*. Hoboken, New Jersey: A John WILEY & SONS.

Lindley, D. V., and N. D. Singpurwalla. 1986. Multivariate Distributions for the life lengths of components of a system sharing a common environment. *Journal of Applied Probability* 23 (2):418–431. doi:10.1017/S0021900200029715.

Mardia, K. V. 1970. Families of bivariate distributions. In *Griffin's statistical monographs and courses*. London.

Moradian, H., D. Denis Larocque, and F. Bellavance. 2017. Survival forests for data with dependent censoring. *Statistical Methods in Medical Research*. doi:10.1177/0962280217727314 doi:10.1177/0962280217727314.

Navarro, J., J. M. Ruiz, and C. J. Sandoval. 2008. Properties of systems with two exchangeable Pareto components. *Statistical Papers* 49 (2):177–190. doi:10.1007/s00362-006-0003-x.

Nelsen, R. B. 2006 *An introduction to copulas* (2nd ed.).New York: Springer.

Noughabi, M. S., and M. Kayid. 2017. Bivariate quantile residual life: A characterization theorem and statistical properties. *Statistical Papers* doi: 10.1007/s00362-017-0905-9. doi:10.1007/s00362-017-0905-9.

Sankaran, P. G., and D. Kundu. 2014. A bivariate pareto model. *Statistics* 48 (2):241–255. doi:10.1080/02331888.2012.719521.

Sankaran, P. G., and N. U. Nair. 1993. A bivariate Pareto model and its applications to reliability. *Naval Research Logistics* 40 (7):1013–1020. doi:10.1002/1520-6750(199312)40:7%3c1013::AID-NAV3220400711%3e3.0.CO;2-7.

Shih, J. H. 2016. Dependence measures and competing risks models under the generalized Farlie-Gumbel-Morgenstern copula. *National Central University Electronic Theses & Dissertations*.

Shih, J. H., and T. Emura. 2018. Likelihood-based inference for bivariate latent failure time models with competing risks under the generalized FGM copula (submitted for publication)

Sklar, A. 1959. Fonctions de répartition à n dimensions et leurs marges. *Publications de l'Institut de Statistique de L'Université de Paris* 8:229–231.

Tsiatis, A. 1975 A nonidentifiability aspect of the problem of competing risks. PNAS USA 72:20–22. doi:10.1073/pnas.72.1.20.

Appendix 1: The RNR algorithm under the Frank copula with common margins

Step 1. Set initial value $(\theta^{(0)}, \alpha^{(0)}, \gamma^{(0)})$.
Step 2. Repeat the Newton-Raphson iterations:

$$\begin{bmatrix}\theta^{(k+1)}\\ \alpha^{(k+1)}\\ \gamma^{(k+1)}\end{bmatrix} = \begin{bmatrix}\theta^{(k)}\\ \alpha^{(k)}\\ \gamma^{(k)}\end{bmatrix} - \left.\begin{bmatrix}\frac{\partial^2\ell_n(\varphi)}{\partial\theta^2} & \frac{\partial^2\ell_n(\varphi)}{\partial\theta\partial\alpha} & \frac{\partial^2\ell_n(\varphi)}{\partial\theta\partial\gamma}\\ \frac{\partial^2\ell_n(\varphi)}{\partial\theta\partial\alpha} & \frac{\partial^2\ell_n(\varphi)}{\partial\alpha^2} & \frac{\partial^2\ell_n(\varphi)}{\partial\alpha\partial\gamma}\\ \frac{\partial^2\ell_n(\varphi)}{\partial\theta\partial\gamma} & \frac{\partial^2\ell_n(\varphi)}{\partial\alpha\partial\gamma} & \frac{\partial^2\ell_n(\varphi)}{\partial\gamma^2}\end{bmatrix}^{-1} \begin{bmatrix}\frac{\partial\ell_n(\varphi)}{\partial\theta}\\ \frac{\partial\ell_n(\varphi)}{\partial\alpha}\\ \frac{\partial\ell_n(\varphi)}{\partial\gamma}\end{bmatrix}\right|_{\theta=\theta^{(k)},\alpha=\alpha^{(k)},\gamma=\gamma^{(k)}}.$$

- If $\max\{|\theta^{(k+1)}-\theta^{(k)}|, |\alpha^{(k+1)}-\alpha^{(k)}|, |\gamma^{(k+1)}-\gamma^{(k)}|\} < \varepsilon$, stop the algorithm then set the MLE $\hat{\varphi} = (\theta^{(k+1)}, \alpha^{(k+1)}, \gamma^{(k+1)})$.
- If the Hessian matrix is singular, or its determinant is positive, or iteration number is greater than 20, stop the algorithm and then return to Step 1 with the initial value replaced by $(\theta^{(0)} + u_1, \alpha^{(0)} \times e^{u_2}, \gamma^{(0)} \times e^{u_3})$, where $u_i \sim U(-r_i, r_i), i = 1, 2, 3$ are independent uniform random variables.

Remark: We set $\varepsilon = 10^{-5}, \theta^{(0)} = 1, \alpha^{(0)} = 1, \gamma^{(0)} = 1, r_1 = 13, r_2 = 3$, and $r_3 = 3$ for simulations. If the NR steps give wrong results, we restart the NR steps by adding uniform random noises to the initial values. This algorithm is called the "Randomized NR Algorithm" (Hu and Emura 2015).

All the derivative expressions are given below.

$$
\begin{aligned}
\frac{\partial \ell_n(\varphi)}{\partial \theta} &= \sum_{i=1}^{n}(\delta_i+\delta_i^*)\left\{-S(t_i)-\frac{S(t_i)e^{-\theta S(t_i)}}{e^{-\theta S(t_i)}-1}+\frac{e^{-\theta}}{e^{-\theta}-1}+S_T(t_i)+\theta\frac{\partial S_T(t_i)}{\partial\theta}\right\}\\
&+\sum_{i=1}^{n}(1-\delta_i-\delta_i^*)\left\{\frac{-1}{\theta}+\frac{(e^{-\theta}-1)[-2e^{-\theta S(t_i)}S(t_i)(e^{-\theta S(t_i)}-1)]}{[e^{-\theta}-1+(e^{-\theta S(t_i)}-1)^2](e^{-\theta}-1)}\log\left[1+\frac{(e^{-\theta S(t_i)}-1)^2}{(e^{-\theta}-1)}\right]^{-1}\right\}\\
&+\sum_{i=1}^{n}(1-\delta_i-\delta_i^*)\left\{\frac{e^{-\theta}(e^{-\theta S(t_i)}-1)^2}{[e^{-\theta}-1+(e^{-\theta S(t_i)}-1)^2](e^{-\theta}-1)}\right\}\log\left[1+\frac{(e^{-\theta S(t_i)}-1)^2}{(e^{-\theta}-1)}\right]^{-1},\\
\frac{\partial \ell_n(\varphi)}{\partial \alpha} &= \sum_{i=1}^{n}(\delta_i+\delta_i^*)\left\{\frac{1}{h(t_i)}\frac{\partial h(t_i)}{\partial\alpha}+\frac{1}{S(t_i)}\frac{\partial S(t_i)}{\partial\alpha}-\theta\frac{\partial S(t_i)}{\partial\alpha}-\frac{\theta e^{-\theta S(t_i)}}{e^{-\theta S(t_i)}-1}\frac{\partial S(t_i)}{\partial\alpha}+\theta\frac{\partial S_T(t_i)}{\partial\alpha}\right\}\\
&+\sum_{i=1}^{n}(1-\delta_i-\delta_i^*)\left\{\frac{-2\theta e^{-\theta S(t_i)}(e^{-\theta S(t_i)}-1)}{[e^{-\theta}-1+(e^{-\theta S(t_i)}-1)^2]}\frac{\partial S(t_i)}{\partial\alpha}\log\left[1+\frac{(e^{-\theta S(t_i)}-1)^2}{(e^{-\theta}-1)}\right]^{-1}\right\},\\
\frac{\partial \ell_n(\varphi)}{\partial \gamma} &= \sum_{i=1}^{n}(\delta_i+\delta_i^*)\left\{\frac{1}{h(t_i)}\frac{\partial h(t_i)}{\partial\gamma}+\frac{1}{S(t_i)}\frac{\partial S(t_i)}{\partial\gamma}-\theta\frac{\partial S(t_i)}{\partial\gamma}-\frac{\theta e^{-\theta S(t_i)}}{e^{-\theta S(t_i)}-1}\frac{\partial S(t_i)}{\partial\gamma}+\theta\frac{\partial S_T(t_i)}{\partial\gamma}\right\}\\
&+\sum_{i=1}^{n}(1-\delta_i-\delta_i^*)\left\{\frac{-2\theta e^{-\theta S(t_i)}(e^{-\theta S(t_i)}-1)}{[e^{-\theta}-1+(e^{-\theta S(t_i)}-1)^2]}\frac{\partial S(t_i)}{\partial\gamma}\log\left[1+\frac{(e^{-\theta S(t_i)}-1)^2}{(e^{-\theta}-1)}\right]^{-1}\right\},
\end{aligned}
$$

where

$$
\begin{aligned}
\frac{\partial h(t_i)}{\partial\alpha} &= \frac{\gamma}{(1+\alpha t_i)^2},\ \frac{\partial S(t_i)}{\partial\alpha}=\frac{-\gamma t_i}{(1+\alpha t_i)^{\gamma+1}},\ \frac{\partial h(t_i)}{\partial\gamma}=\frac{\alpha}{(1+\alpha t_i)},\ \frac{\partial S(t_i)}{\partial\gamma}=\frac{-\log(1+\alpha t_i)}{(1+\alpha t_i)^{\gamma}},\\
\frac{\partial S_T(t_i)}{\partial\theta} &= \frac{1}{\theta^2}\log\left[1+\frac{(e^{-\theta S(t_i)}-1)^2}{(e^{-\theta}-1)}\right]-\frac{-2S(t_i)e^{-\theta S(t_i)}(e^{-\theta S(t_i)}-1)(e^{-\theta}-1)+e^{-\theta}(e^{-\theta S(t_i)}-1)^2}{\theta[e^{-\theta}-1+(e^{-\theta S(t_i)}-1)^2](e^{-\theta}-1)},\\
\frac{\partial S_T(t_i)}{\partial\alpha} &= \frac{2e^{-\theta S(t_i)}(e^{-\theta S(t_i)}-1)}{e^{-\theta}-1+(e^{-\theta S(t_i)}-1)^2}\times\frac{\partial S(t_i)}{\partial\alpha},\ \frac{\partial S_T(t_i)}{\partial\gamma}=\frac{2e^{-\theta S(t_i)}(e^{-\theta S(t_i)}-1)}{e^{-\theta}-1+(e^{-\theta S(t_i)}-1)^2}\times\frac{\partial S(t_i)}{\partial\gamma},\\
\frac{\partial \ell_n^2(\varphi)}{\partial\theta^2} &= \sum_{i=1}^{n}(\delta_i+\delta_i^*)\left\{\frac{-S(t_i)^2e^{-\theta S(t_i)}}{(e^{-\theta S(t_i)}-1)^2}+\frac{e^{-\theta}}{(e^{-\theta}-1)^2}+2\frac{\partial S_T(t_i,t_i)}{\partial\theta}+\theta\frac{\partial^2 S_T(t_i)}{\partial\theta^2}\right\}\\
&+\sum_{i=1}^{n}(1-\delta_i-\delta_i^*)\left\{\frac{1}{\theta^2}+\frac{2S(t_i)^2e^{-\theta S(t_i)}(e^{-\theta}-1)(2e^{-\theta S(t_i)}-1)}{[e^{-\theta}-1+(e^{-\theta S(t_i)}-1)^2](e^{-\theta}-1)}\log\left[1+\frac{(e^{-\theta S(t_i)}-1)^2}{(e^{-\theta}-1)}\right]^{-1}\right\}\\
&+\sum_{i=1}^{n}\frac{-(1-\delta_i-\delta_i^*)\left\{2\theta\frac{\partial S(t_i)}{\partial\alpha}e^{-\theta S(t_i)}(e^{-\theta S(t_i)}-1)\right\}^2\left\{\log\left[1+\frac{(e^{-\theta S(t_i)}-1)^2}{(e^{-\theta}-1)}\right]+1\right\}}{\left\{\log\left[1+\frac{(e^{-\theta S(t_i)}-1)^2}{(e^{-\theta}-1)}\right][e^{-\theta}-1+(e^{-\theta S(t_i)}-1)^2](e^{-\theta}-1)\right\}^2}\\
&+\sum_{i=1}^{n}\frac{-(1-\delta_i-\delta_i^*)e^{-\theta}(e^{-\theta S(t_i)}-1)^2}{[e^{-\theta}-1+(e^{-\theta S(t_i)}-1)^2](e^{-\theta}-1)}\log\left[1+\frac{(e^{-\theta S(t_i)}-1)^2}{(e^{-\theta}-1)}\right]^{-1},\\
\frac{\partial \ell_n^2(\varphi)}{\partial\alpha^2} &= \sum_{i=1}^{n}(\delta_i+\delta_i^*)\left\{\frac{1}{h(t_i)^2}\left(h(t_i)\frac{\partial^2 h(t_i)}{\partial\alpha^2}-\left[\frac{\partial h(t_i)}{\partial\alpha}\right]^2\right)+\frac{1}{S(t_i)^2}\left(S(t_i)\frac{\partial^2 S(t_i)}{\partial\alpha^2}-\left[\frac{\partial S(t_i)}{\partial\alpha}\right]^2\right)\right\}\\
&+\sum_{i=1}^{n}(\delta_i+\delta_i^*)\left\{-\theta\frac{\partial^2 S(t_i)}{\partial\alpha^2}-\frac{\theta e^{-2\theta S(t_i)}}{e^{-\theta S(t_i)}-1}\left[\frac{\partial^2 S(t_i)}{\partial\alpha^2}\right]-\frac{\theta^2e^{-\theta S(t_i)}}{(e^{-\theta S(t_i)}-1)^2}\left[\frac{\partial S(t_i)}{\partial\alpha}\right]^2+\theta\frac{\partial^2 S_T(t_i)}{\partial\alpha^2}\right\}
\end{aligned}
$$

$$+\sum_{i=1}^{n}\frac{-(1-\delta_i-\delta_i^*)\left\{2\theta\frac{\partial S(t_i)}{\partial\alpha}e^{-\theta S(t_i)}(e^{-\theta S(t_i)}-1)\right\}^2\left\{\log\left[1+\frac{(e^{-\theta S(t_i)}-1)^2}{(e^{-\theta}-1)}\right]+1\right\}}{\left\{\log\left[1+\frac{(e^{-\theta S(t_i)}-1)^2}{(e^{-\theta}-1)}\right][e^{-\theta}-1+(e^{-\theta S(t_i)}-1)^2]\right\}^2}$$

$$+\sum_{i=1}^{n}\frac{-(1-\delta_i-\delta_i^*)2\theta e^{-\theta S(t_i)}\frac{\partial^2 S(t_i)}{\partial\alpha^2}(e^{-\theta S(t_i)}-1)}{\log\left[1+\frac{(e^{-\theta S(t_i)}-1)^2}{(e^{-\theta}-1)}\right][e^{-\theta}-1+(e^{-\theta S(t_i)}-1)^2]}$$

$$+\sum_{i=1}^{n}\frac{(1-\delta_i-\delta_i^*)2\theta^2 e^{-\theta S(t_i)}\left[\frac{\partial S(t_i)}{\partial\alpha}\right]^2(2e^{-\theta S(t_i)}-1)}{\log\left[1+\frac{(e^{-\theta S(t_i)}-1)^2}{(e^{-\theta}-1)}\right][e^{-\theta}-1+(e^{-\theta S(t_i)}-1)^2]},$$

$$\frac{\partial\ell_n^2(\varphi)}{\partial\gamma^2}=\sum_{i=1}^{n}(\delta_i+\delta_i^*)\left\{\frac{1}{h(t_i)^2}\left(h(t_i)\frac{\partial^2 h(t_i)}{\partial\gamma^2}-\left[\frac{\partial h(t_i)}{\partial\gamma}\right]^2\right)\right\}$$

$$+\sum_{i=1}^{n}(\delta_i+\delta_i^*)\left\{\frac{1}{S(t_i)^2}\left(S(t_i)\frac{\partial^2 S(t_i)}{\partial\gamma^2}-\left[\frac{\partial S(t_i)}{\partial\gamma}\right]^2\right)+\theta\frac{\partial^2 S_T(t_i)}{\partial\gamma^2}\right\}$$

$$+\sum_{i=1}^{n}(\delta_i+\delta_i^*)\left\{-\theta\frac{\partial^2 S(t_i)}{\partial\gamma^2}-\frac{\theta e^{-2\theta S(t_i)}}{e^{-\theta S(t_i)}-1}\left[\frac{\partial^2 S(t_i)}{\partial\gamma^2}\right]-\frac{\theta^2 e^{-\theta S(t_i)}}{(e^{-\theta S(t_i)}-1)^2}\left[\frac{\partial S(t_i)}{\partial\gamma}\right]^2\right\}$$

$$+\sum_{i=1}^{n}\frac{-(1-\delta_i-\delta_i^*)\left(2\theta\frac{\partial S(t_i)}{\partial\gamma}e^{-\theta S(t_i)}(e^{-\theta S(t_i)}-1)\right)^2\left(\log\left[1+\frac{(e^{-\theta S(t_i)}-1)^2}{(e^{-\theta}-1)}\right]+1\right)}{\left(\log\left[1+\frac{(e^{-\theta S(t_i)}-1)^2}{(e^{-\theta}-1)}\right][e^{-\theta}-1+(e^{-\theta S(t_i)}-1)^2]\right)^2}$$

$$+\sum_{i=1}^{n}\frac{-(1-\delta_i-\delta_i^*)2\theta e^{-\theta S(t_i)}\frac{\partial^2 S(t_i)}{\partial\gamma^2}(e^{-\theta S(t_i)}-1)}{\log\left[1+\frac{(e^{-\theta S(t_i)}-1)^2}{(e^{-\theta}-1)}\right][e^{-\theta}-1+(e^{-\theta S(t_i)}-1)^2]}$$

$$+\sum_{i=1}^{n}\frac{(1-\delta_i-\delta_i^*)2\theta^2 e^{-\theta S(t_i)}\left[\frac{\partial S(t_i)}{\partial\gamma}\right]^2(2e^{-\theta S(t_i)}-1)}{\log\left[1+\frac{(e^{-\theta S(t_i)}-1)^2}{(e^{-\theta}-1)}\right][e^{-\theta}-1+(e^{-\theta S(t_i)}-1)^2]},$$

$$\frac{\partial\ell_n^2(\varphi)}{\partial\theta\partial\alpha}=\sum_{i=1}^{n}(\delta_i+\delta_i^*)\left\{-\frac{\partial S(t_i)}{\partial\alpha}-\frac{\partial S(t_i)}{\partial\alpha}\frac{e^{-\theta S(t_i)}[e^{-\theta S(t_i)}+\theta S(t_i)-1]}{(e^{-\theta S(t_i)}-1)^2}+\frac{\partial S_T(t_i,t_i)}{\partial\alpha}+\theta\frac{\partial^2 S_T(t_i)}{\partial\theta\partial\alpha}\right\}$$

$$+\sum_{i=1}^{n}(1-\delta_i-\delta_i^*)\left\{\frac{2\frac{\partial S(t_i)}{\partial\alpha}e^{-\theta S(t_i)}[e^{-\theta S(t_i)}-1]^2\left(e^{-\theta}\frac{(e^{-\theta S(t_i)}-1)}{(e^{-\theta}-1)}-2S(t_i)e^{-\theta S(t_i)}\right)}{\left(\log\left[1+\frac{(e^{-\theta S(t_i)}-1)^2}{(e^{-\theta}-1)}\right][e^{-\theta}-1+(e^{-\theta S(t_i)}-1)^2]\right)^2}\right\}$$

$$+\sum_{i=1}^{n}(1-\delta_i-\delta_i^*)\left\{\frac{\log\left[1+\frac{(e^{-\theta S(t_i)}-1)^2}{(e^{-\theta}-1)}\right][-e^{-\theta}-2S(t_i)e^{-\theta S(t_i)}(e^{-\theta S(t_i)}-1)]}{\left(\log\left[1+\frac{(e^{-\theta S(t_i)}-1)^2}{(e^{-\theta}-1)}\right][e^{-\theta}-1+(e^{-\theta S(t_i)}-1)^2]\right)^2}\right\}$$

$$+\sum_{i=1}^{n}(1-\delta_i-\delta_i^*)\left\{\frac{-2\frac{\partial S(t_i)}{\partial\alpha}e^{-\theta S(t_i)}[e^{-\theta S(t_i)}-\theta S(t_i)(1-2e^{-\theta S(t_i)})]}{\log\left[1+\frac{(e^{-\theta S(t_i)}-1)^2}{(e^{-\theta}-1)}\right][e^{-\theta}-1+(e^{-\theta S(t_i)}-1)^2]}\right\},$$

$$\frac{\partial\ell_n^2(\varphi)}{\partial\theta\partial\gamma}=\sum_{i=1}^{n}(\delta_i+\delta_i^*)\left\{-\frac{\partial S(t_i)}{\partial\gamma}-\frac{\partial S(t_i)}{\partial\gamma}\frac{e^{-\theta S(t_i)}\left(e^{-\theta S(t_i)}+\theta S(t_i)-1\right)}{(e^{-\theta S(t_i)}-1)^2}+\frac{\partial S_T(t_i,t_i)}{\partial\gamma}+\theta\frac{\partial^2 S_T(t_i)}{\partial\theta\partial\gamma}\right\}$$

$$+\sum_{i=1}^{n}(1-\delta_i-\delta_i^*)\left\{\frac{2\frac{\partial S(t_i)}{\partial\gamma}e^{-\theta S(t_i)}[e^{-\theta S(t_i)}-1]^2\left(e^{-\theta}\frac{(e^{-\theta S(t_i)}-1)}{(e^{-\theta}-1)}-2S(t_i)e^{-\theta S(t_i)}\right)}{\left(\log\left[1+\frac{(e^{-\theta S(t_i)}-1)^2}{(e^{-\theta}-1)}\right][e^{-\theta}-1+(e^{-\theta S(t_i)}-1)^2]\right)^2}\right\}$$

$$+\sum_{i=1}^{n}(1-\delta_i-\delta_i^*)\left\{\frac{\log[1+\frac{(e^{-\theta S(t_i)}-1)^2}{(e^{-\theta}-1)}][-e^{-\theta}-2S(t_i)e^{-\theta S(t_i)}(e^{-\theta S(t_i)}-1)]}{\left(\log\left[1+\frac{(e^{-\theta S(t_i)}-1)^2}{(e^{-\theta}-1)}\right][e^{-\theta}-1+(e^{-\theta S(t_i)}-1)^2]\right)^2}\right\}$$

$$+\sum_{i=1}^{n}(1-\delta_i-\delta_i^*)\left\{\frac{-2\frac{\partial S(t_i)}{\partial\gamma}e^{-\theta S(t_i)}[e^{-\theta S(t_i)}-\theta S(t_i)(1-2e^{-\theta S(t_i)})]}{\log\left[1+\frac{(e^{-\theta S(t_i)}-1)^2}{(e^{-\theta}-1)}\right][e^{-\theta}-1+(e^{-\theta S(t_i)}-1)^2]}\right\},$$

$$\frac{\partial\ell_n^2(\varphi)}{\partial\alpha\partial\gamma}=\sum_{i=1}^{n}(\delta_i+\delta_i^*)\left\{\frac{1}{h(t_i)^2}\left(h(t_i)\frac{\partial^2 h(t_i)}{\partial\alpha\partial\gamma}-\frac{\partial h(t_i)}{\partial\alpha}\frac{\partial h(t_i)}{\partial\gamma}\right)\right\}$$

$$+\sum_{i=1}^{n}(\delta_i+\delta_i^*)\left\{\frac{1}{S(t_i)^2}\left(S(t_i)\frac{\partial^2 S(t_i)}{\partial\alpha\partial\gamma}-\frac{\partial S(t_i)}{\partial\alpha}\frac{\partial S(t_i)}{\partial\gamma}\right)+\theta\frac{\partial^2 S_T(t_i)}{\partial\alpha\partial\gamma}\right\}$$

$$+\sum_{i=1}^{n}(\delta_i+\delta_i^*)\left\{\frac{\theta e^{-\theta S(t_i)}}{(e^{-\theta S(t_i)}-1)^2}\left(\frac{\partial^2 S(t_i)}{\partial\alpha\partial\gamma}[e^{-\theta S(t_i)}-1]+\theta\frac{\partial S(t_i)}{\partial\alpha}\frac{\partial S(t_i)}{\partial\gamma}\right)\right\}$$

$$+\sum_{i=1}^{n}(1-\delta_i-\delta_i^*)\left\{\frac{4\theta^2 e^{-2\theta S(t_i)}(e^{-\theta S(t_i)}-1)^2\frac{\partial S(t_i)}{\partial\alpha}\frac{\partial S(t_i)}{\partial\gamma}\left(\log\left[1+\frac{(e^{-\theta S(t_i)}-1)^2}{(e^{-\theta}-1)}\right]+1\right)}{\left(\log\left[1+\frac{(e^{-\theta S(t_i)}-1)^2}{(e^{-\theta}-1)}\right][e^{-\theta}-1+(e^{-\theta S(t_i)}-1)^2]\right)^2}\right\}$$

$$+\sum_{i=1}^{n}(1-\delta_i-\delta_i^*)\left\{\frac{-2\theta^2 e^{-\theta S(t_i)}(e^{-\theta S(t_i)}-1)(1-2e^{-\theta S(t_i)})\frac{\partial S(t_i)}{\partial\alpha}\frac{\partial S(t_i)}{\partial\gamma}}{\log\left[1+\frac{(e^{-\theta S(t_i)}-1)^2}{(e^{-\theta}-1)}\right][e^{-\theta}-1+(e^{-\theta S(t_i)}-1)^2]}\right\}$$

$$+\sum_{i=1}^{n}(1-\delta_i-\delta_i^*)\left\{\frac{-2\theta e^{-\theta S(t_i)}(e^{-\theta S(t_i)}-1)\frac{\partial^2 S(t_i)}{\partial\alpha\partial\gamma}}{\log\left[1+\frac{(e^{-\theta S(t_i)}-1)^2}{(e^{-\theta}-1)}\right][e^{-\theta}-1+(e^{-\theta S(t_i)}-1)^2]}\right\},$$

where

$$\frac{\partial^2 h(t_i)}{\partial\alpha^2}=\frac{-2\gamma t_i}{(1+\alpha t_i)^3},\ \frac{\partial^2 S(t_i)}{\partial\alpha^2}=\gamma(\gamma+1)t_i^2(1+\alpha t_i)^{-(\gamma+2)},\ \frac{\partial^2 h(t_i)}{\partial\gamma^2}=0,$$

$$\frac{\partial^2 S(t_i)}{\partial\gamma^2}=(1+\alpha t_i)^{-\gamma}[\log(1+\alpha t_i)]^2,\ \frac{\partial^2 h(t_i)}{\partial\alpha\partial\gamma}=\frac{1}{(1+\alpha t_i)^2},\ \frac{\partial^2 S(t_i)}{\partial\alpha\partial\gamma}=\frac{t_i[\gamma\log(1+\alpha t_i)-1]}{(1+\alpha t_i)^{\gamma+1}},$$

$$\frac{\partial^2 S_T(t_i)}{\partial\theta^2}=\frac{e^{-\theta}(e^{-\theta S(t_i)}-1)^2-2e^{-\theta S(t_i)}S(t_i)(e^{-\theta S(t_i)}-1)(e^{-\theta}-1)}{\theta^2(e^{-\theta}-1)[e^{-\theta}-1+(e^{-\theta S(t_i)}-1)^2]}$$

$$+\frac{[2S(t_i)e^{-\theta S(t_i)}(e^{-\theta S(t_i)}-1)(e^{-\theta}-1)-e^{-\theta}(e^{-\theta S(t_i)}-1)^2][e^{-\theta}+2e^{-\theta S(t_i)}S(t_i)(e^{-\theta S(t_i)}-1)]}{\theta(e^{-\theta}-1)[e^{-\theta}-1+(e^{-\theta S(t_i)}-1)^2]}$$

$$+\frac{e^{-\theta}(e^{-\theta S(t_i)}-1)[2e^{-\theta S(t_i)}S(t_i)(e^{-\theta}-1)-e^{-\theta}(e^{-\theta S(t_i)}-1)][e^{-\theta}-1+(e^{-\theta S(t_i)}-1)^2]}{\theta(e^{-\theta}-1)[e^{-\theta}-1+(e^{-\theta S(t_i)}-1)^2]}$$

$$+\frac{2S(t_i)^2 e^{-\theta S(t_i)}(e^{-\theta}-1)(1-2e^{-\theta S(t_i)})+e^{-\theta}(e^{-\theta S(t_i)}-1)^2}{\theta(e^{-\theta}-1)[e^{-\theta}-1+(e^{-\theta S(t_i)}-1)^2]}-\frac{2}{\theta^3}\log\left[1+\frac{(e^{-\theta S(t_i)}-1)^2}{(e^{-\theta}-1)}\right]$$

$$+\frac{[-2S(t_i)e^{-\theta S(t_i)}(e^{-\theta S(t_i)}-1)(e^{-\theta}-1)+e^{-\theta}(e^{-\theta S(t_i)}-1)^2]}{\theta^2(e^{-\theta}-1)[e^{-\theta}-1+(e^{-\theta S(t_i)}-1)^2]},$$

$$\frac{\partial^2 S_T(t_i)}{\partial\alpha^2}=\frac{\left(2\frac{\partial^2 S(t_i)}{\partial\alpha^2}e^{-\theta S(t_i)}(e^{-\theta S(t_i)}-1)+2\theta e^{-\theta S(t_i)}\left[\frac{\partial S(t_i)}{\partial\alpha}\right]^2[1-2e^{-\theta S(t_i)}]\right)}{e^{-\theta}-1+(e^{-\theta S(t_i)}-1)^2}$$

$$+\frac{4\theta\left[\frac{\partial S(t_i)}{\partial\alpha}\right]^2 e^{-2\theta S(t_i)}(e^{-\theta S(t_i)}-1)^2}{[e^{-\theta}-1+(e^{-\theta S(t_i)}-1)^2]^2},$$

$$\frac{\partial^2 S_T(t_i)}{\partial \gamma^2} = \frac{\left(2\frac{\partial^2 S(t_i)}{\partial \gamma^2} e^{-\theta S(t_i)}(e^{-\theta S(t_i)}-1) + 2\theta e^{-\theta S(t_i)}\left[\frac{\partial S(t_i)}{\partial \gamma}\right]^2 [1-2e^{-\theta S(t_i)}]\right)}{e^{-\theta}-1+(e^{-\theta S(t_i)}-1)^2} + \frac{4\theta\left[\frac{\partial S(t_i)}{\partial \gamma}\right]^2 e^{-2\theta S(t_i)}(e^{-\theta S(t_i)}-1)^2}{[e^{-\theta}-1+(e^{-\theta S(t_i)}-1)^2]^2},$$

$$\frac{\partial^2 S_T(t_i)}{\partial \theta \partial \alpha} = 2e^{-\theta S(t_i)}\frac{\partial S(t_i)}{\partial \alpha}\left\{\frac{S(t_i)(1-2e^{-\theta S(t_i)}) + (e^{-\theta S(t_i)}-1)[e^{-\theta}+2S(t_i)e^{-\theta S(t_i)}(e^{-\theta S(t_i)}-1)]}{e^{-\theta}-1+(e^{-\theta S(t_i)}-1)^2}\right\}$$

$$\frac{\partial^2 S_T(t_i)}{\partial \theta \partial \gamma} = 2e^{-\theta S(t_i)}\frac{\partial S(t_i)}{\partial \gamma}\left\{\frac{S(t_i)(1-2e^{-\theta S(t_i)}) + (e^{-\theta S(t_i)}-1)[e^{-\theta}+2S(t_i)e^{-\theta S(t_i)}(e^{-\theta S(t_i)}-1)]}{e^{-\theta}-1+(e^{-\theta S(t_i)}-1)^2}\right\}$$

$$\frac{\partial^2 S_T(t_i)}{\partial \alpha \partial \gamma} = \frac{2e^{-\theta S(t_i)}\left[\frac{\partial^2 S(t_i)}{\partial \alpha \partial \gamma}(e^{-\theta S(t_i)}-1) + \theta\frac{\partial S(t_i)}{\partial \alpha}\frac{\partial S(t_i)}{\partial \gamma}(1-2e^{-\theta S(t_i)})\right]}{e^{-\theta}-1+(e^{-\theta S(t_i)}-1)^2} + \frac{4\theta\frac{\partial S(t_i)}{\partial \alpha}\frac{\partial S(t_i)}{\partial \gamma}e^{-2\theta S(t_i)}(e^{-\theta S(t_i)}-1)^2}{[e^{-\theta}-1+(e^{-\theta S(t_i)}-1)^2]^2}.$$

Appendix 2: A bootstrap method under the Frank model with the common margin

Step 1. Generate B bootstrap samples $(T_i^{(b)}, \delta_i^{(b)}, \delta_i^{*(b)})$, $i = 1, \dots, n$, $b = 1, 2, \dots, B$, based on the Frank copula model with estimated parameters $\hat{\varphi} = (\hat{\theta}, \hat{\alpha}, \hat{\gamma})$ and independent censoring time $C \sim U(0, w)$, where $w = \max\{T_j : \delta_j = \delta_j^* = 0\}$.
Step 2. Compute the bootstrap MLEs, $\hat{\varphi}^{(b)} = (\hat{\theta}^{(b)}, \hat{\alpha}^{(b)}, \hat{\gamma}^{(b)})$, $b = 1, \dots, B$.
Step 3. The bootstrap SE is computed, for example, as the standard error

$$SE(\hat{\theta}) = \left(\frac{1}{B-1}\sum_{b=1}^{B}(\hat{\theta}^{(b)} - \bar{\hat{\theta}}^{(\cdot)})^2\right)^{1/2}, \quad \bar{\hat{\theta}}^{(\cdot)} = \frac{1}{B}\sum_{b=1}^{B}\hat{\theta}^{(b)}.$$

Step 4. The bootstrap $(1-\varepsilon) \times 100\%$ CI is computed, for example, as the $(\varepsilon/2) \times 100\%$ and $(1-\varepsilon/2) \times 100\%$ points of $\hat{\theta}^{(b)}, b = 1, \dots, B$.

Remark: We set $B = 200$ for simulations.

Appendix 3: The RNR algorithm under the Frank copula model with fixed θ

Step 1. Set initial value $(\alpha_1^{(0)}, \alpha_2^{(0)}, \gamma_1^{(0)}, \gamma_2^{(0)})$.
Step 2. Repeat the Newton-Raphson iterations:

$$\begin{bmatrix}\alpha_1^{(k+1)}\\ \alpha_2^{(k+1)}\\ \gamma_1^{(k+1)}\\ \gamma_2^{(k+1)}\end{bmatrix} = \begin{bmatrix}\alpha_1^{(k)}\\ \alpha_2^{(k)}\\ \gamma_1^{(k)}\\ \gamma_2^{(k)}\end{bmatrix} - \begin{bmatrix}\frac{\partial^2 \ell_n(\varphi)}{\partial \alpha_1^2} & \frac{\partial^2 \ell_n(\varphi)}{\partial \alpha_1 \partial \alpha_2} & \frac{\partial^2 \ell_n(\varphi)}{\partial \alpha_1 \partial \gamma_1} & \frac{\partial^2 \ell_n(\varphi)}{\partial \alpha_1 \partial \gamma_2}\\ \frac{\partial^2 \ell_n(\varphi)}{\partial \alpha_1 \partial \alpha_2} & \frac{\partial^2 \ell_n(\varphi)}{\partial \alpha_2^2} & \frac{\partial^2 \ell_n(\varphi)}{\partial \alpha_2 \partial \gamma_1} & \frac{\partial^2 \ell_n(\varphi)}{\partial \alpha_2 \partial \gamma_2}\\ \frac{\partial^2 \ell_n(\varphi)}{\partial \alpha_1 \partial \gamma_1} & \frac{\partial^2 \ell_n(\varphi)}{\partial \alpha_2 \partial \gamma_1} & \frac{\partial^2 \ell_n(\varphi)}{\partial \gamma_1^2} & \frac{\partial^2 \ell_n(\varphi)}{\partial \gamma_1 \partial \gamma_2}\\ \frac{\partial^2 \ell_n(\varphi)}{\partial \alpha_1 \partial \gamma_2} & \frac{\partial^2 \ell_n(\varphi)}{\partial \alpha_2 \partial \gamma_2} & \frac{\partial^2 \ell_n(\varphi)}{\partial \gamma_1 \partial \gamma_2} & \frac{\partial^2 \ell_n(\varphi)}{\partial \gamma_2^2}\end{bmatrix}^{-1} \begin{bmatrix}\frac{\partial \ell_n(\varphi)}{\partial \alpha_1}\\ \frac{\partial \ell_n(\varphi)}{\partial \alpha_2}\\ \frac{\partial \ell_n(\varphi)}{\partial \gamma_1}\\ \frac{\partial \ell_n(\varphi)}{\partial \gamma_2}\end{bmatrix}\Bigg|_{\substack{\alpha_1=\alpha_1^{(k)}, \gamma_1=\gamma_1^{(k)}\\ \alpha_2=\alpha_2^{(k)}, \gamma_2=\gamma_2^{(k)}}}.$$

- If $\max\{|\alpha_1^{(k+1)} - \alpha_1^{(k)}|, |\alpha_2^{(k+1)} - \alpha_2^{(k)}|, |\gamma_1^{(k+1)} - \gamma_1^{(k)}|, |\gamma_2^{(k+1)} - \gamma_2^{(k)}|\} < \varepsilon$, stop the algorithm then set the MLE $\hat{\varphi} = (\alpha_1^{(k+1)}, \alpha_2^{(k+1)}, \gamma_1^{(k+1)}, \gamma_2^{(k+1)})$.
- If the Hessian matrix is singular or $\max\{\alpha_1^{(k+1)}, \alpha_2^{(k+1)}, \gamma_1^{(k+1)}, \gamma_2^{(k+1)}\} > d$, stop the algorithm then return to Step 1 with the initial value replaced by $(\alpha_1^{(0)} \times e^{u_1}, \alpha_2^{(0)} \times e^{u_2}, \gamma_1^{(0)} \times e^{u_3}, \gamma_2^{(0)} \times e^{u_4})$, where $u_i \sim U(-r_i, r_i)$, $i = 1, 2, 3, 4$ are independent uniform random variables.

Remark: We set $\varepsilon = 10^{-5}, \alpha_1^{(0)} = 1, \alpha_2^{(0)} = 1, \gamma_1^{(0)} = 1, \gamma_2^{(0)} = 1, r_1 = 6, r_2 = 6, r_3 = 6, r_4 = 6$ and $d = e^{10}$ for data analysis.

All partial derivatives of the log-likelihood are given below.

$$\frac{\partial \ell_n(\varphi)}{\partial \alpha_1} = \sum_{i=1}^n \delta_i \left\{ \frac{1}{h_1(t_i)} \frac{\partial h_1(t_i)}{\partial \alpha_1} + \frac{1}{S_1(t_i)} \frac{\partial S_1(t_i)}{\partial \alpha_1} - \theta \frac{\partial S_1(t_i)}{\partial \alpha_1} + \theta \frac{\partial S_T(t_i)}{\partial \alpha_1} \right\}$$
$$+ \sum_{i=1}^n \delta_i^* \left\{ \frac{-\theta e^{-\theta S_1(t_i)}}{(e^{-\theta S_1(t_i)} - 1)} \frac{\partial S_1(t_i)}{\partial \alpha_1} + \theta \frac{\partial S_T(t_i)}{\partial \alpha_1} \right\} + \sum_{i=1}^n (1 - \delta_i - \delta_i^*) \frac{\partial}{\partial \alpha_1} \log[S_T(t_i)],$$

$$\frac{\partial \ell_n(\varphi)}{\partial \alpha_2} = \sum_{i=1}^n \delta_i \left\{ \frac{-\theta e^{-\theta S_2(t_i)}}{(e^{-\theta S_2(t_i)} - 1)} \frac{\partial S_2(t_i)}{\partial \alpha_2} + \theta \frac{\partial S_T(t_i)}{\partial \alpha_2} \right\}$$
$$+ \sum_{i=1}^n \delta_i^* \left\{ \frac{1}{h_2(t_i)} \frac{\partial h_2(t_i)}{\partial \alpha_2} + \frac{1}{S_2(t_i)} \frac{\partial S_2(t_i)}{\partial \alpha_2} - \theta \frac{\partial S_2(t_i)}{\partial \alpha_2} + \theta \frac{\partial S_T(t_i)}{\partial \alpha_2} \right\}$$
$$+ \sum_{i=1}^n (1 - \delta_i - \delta_i^*) \frac{\partial}{\partial \alpha_2} \log[S_T(t_i)],$$

$$\frac{\partial \ell_n(\varphi)}{\partial \gamma_1} = \sum_{i=1}^n \delta_i \left\{ \frac{1}{h_1(t_i)} \frac{\partial h_1(t_i)}{\partial \gamma_1} + \frac{1}{S_1(t_i)} \frac{\partial S_1(t_i)}{\partial \gamma_1} - \theta \frac{\partial S_1(t_i)}{\partial \gamma_1} + \theta \frac{\partial S_T(t_i)}{\partial \gamma_1} \right\}$$
$$+ \sum_{i=1}^n \delta_i^* \left\{ \frac{-\theta e^{-\theta S_1(t_i)}}{(e^{-\theta S_1(t_i)} - 1)} \frac{\partial S_1(t_i)}{\partial \gamma_1} + \theta \frac{\partial S_T(t_i)}{\partial \gamma_1} \right\} + \sum_{i=1}^n (1 - \delta_i - \delta_i^*) \frac{\partial}{\partial \gamma_1} \log[S_T(t_i)],$$

$$\frac{\partial \ell_n(\varphi)}{\partial \gamma_2} = \sum_{i=1}^n \delta_i \left\{ \frac{-\theta e^{-\theta S_2(t_i)}}{(e^{-\theta S_2(t_i)} - 1)} \frac{\partial S_2(t_i)}{\partial \gamma_2} + \theta \frac{\partial S_T(t_i)}{\partial \gamma_2} \right\}$$
$$+ \sum_{i=1}^n \delta_i^* \left\{ \frac{1}{h_2(t_i)} \frac{\partial h_2(t_i)}{\partial \gamma_2} + \frac{1}{S_2(t_i)} \frac{\partial S_2(t_i)}{\partial \gamma_2} - \theta \frac{\partial S_2(t_i)}{\partial \gamma_2} + \theta \frac{\partial S_T(t_i)}{\partial \gamma_2} \right\}$$
$$+ \sum_{i=1}^n (1 - \delta_i - \delta_i^*) \frac{\partial}{\partial \gamma_2} \log[S_T(t_i)],$$

where $j = 1, 2$,

$$\frac{\partial h_j(t_i)}{\partial \alpha_j} = \frac{\gamma_j}{(1 + \alpha_j t_i)^2}, \quad \frac{\partial S_j(t_i)}{\partial \alpha_j} = \frac{-\gamma_j t_i}{(1 + \alpha_j t_i)^{\gamma + 1}}, \quad \frac{\partial h_j(t_i)}{\partial \gamma_j} = \frac{\alpha_j}{(1 + \alpha_j t_i)},$$
$$\frac{\partial S_j(t_i)}{\partial \gamma_j} = \frac{-\log(1 + \alpha_j t_i)}{(1 + \alpha_j t_i)^{\gamma_j}}, \quad \frac{\partial S_T(t_i)}{\partial \alpha_j} = \frac{\partial S_j(t_i)}{\partial \alpha_j} \frac{e^{-\theta S_j(t_i)} (e^{-\theta S_{3-j}(t_i)} - 1)}{e^{-\theta} - 1 + (e^{-\theta S_1(t_i)} - 1)(e^{-\theta S_2(t_i)} - 1)},$$
$$\frac{\partial S_T(t_i)}{\partial \gamma_j} = \frac{\partial S_j(t_i)}{\partial \gamma_j} \frac{e^{-\theta S_j(t_i)} (e^{-\theta S_{3-j}(t_i)} - 1)}{e^{-\theta} - 1 + (e^{-\theta S_1(t_i)} - 1)(e^{-\theta S_2(t_i)} - 1)},$$

$$\frac{\partial \ell_n^2(\varphi)}{\partial \alpha_1^2} = \sum_{i=1}^n \delta_i \left\{ \frac{1}{h_1(t_i)^2} \left(h_1(t_i) \frac{\partial^2 h_1(t_i)}{\partial \alpha_1^2} - \left[\frac{\partial h_1(t_i)}{\partial \alpha_1} \right]^2 \right) \right\}$$
$$+ \sum_{i=1}^n \delta_i \left\{ \frac{1}{S_1(t_i)^2} \left(S_1(t_i) \frac{\partial^2 S_1(t_i)}{\partial \alpha_1^2} - \left[\frac{\partial S_1(t_i)}{\partial \alpha_1} \right]^2 \right) \right\} + \sum_{i=1}^n \delta_i \left\{ -\theta \frac{\partial^2 S_1(t_i)}{\partial \alpha_1^2} + \theta \frac{\partial^2 S_T(t_i)}{\partial \alpha_1^2} \right\}$$
$$+ \sum_{i=1}^n \delta_i^* \left\{ -\frac{\theta (e^{-\theta S_1(t_i)} - 1) e^{-\theta S_1(t_i)}}{(e^{-\theta S_1(t_i)} - 1)} \left[\frac{\partial^2 S_1(t_i)}{\partial \alpha_1^2} \right] \right\}$$

$$+\sum_{i=1}^{n}\delta_i^*\left\{-\frac{\theta^2 e^{-\theta S_1(t_i)}}{(e^{-\theta S_1(t_i)}-1)^2}\left[\frac{\partial S_1(t_i)}{\partial \alpha_1}\right]^2+\theta\frac{\partial^2 S_T(t_i)}{\partial \alpha_1^2}\right\}$$

$$+\sum_{i=1}^{n}(1-\delta_i-\delta_i^*)\left\{\frac{-2\theta e^{-\theta S_1(t_i)}\frac{\partial^2 S_1(t_i)}{\partial \alpha_1^2}(e^{-\theta S(t_i)}-1)}{\log\left[1+\frac{(e^{-\theta S_1(t_i)}-1)^2}{(e^{-\theta}-1)}\right][e^{-\theta}-1+(e^{-\theta S_1(t_i)}-1)^2]}\right\}$$

$$+\sum_{i=1}^{n}(1-\delta_i-\delta_i^*)\left\{\frac{-2\theta^2 e^{-\theta S_1(t_i)}\left[\frac{\partial S_1(t_i)}{\partial \alpha_1}\right]^2(-2e^{-\theta S_1(t_i)}+1)}{\log\left[1+\frac{(e^{-\theta S_1(t_i)}-1)^2}{(e^{-\theta}-1)}\right][e^{-\theta}-1+(e^{-\theta S_1(t_i)}-1)^2]}\right\}$$

$$+\sum_{i=1}^{n}(1-\delta_i-\delta_i^*)\frac{\partial^2}{\partial \alpha_1^2}\log[S_T(t_i)],$$

$$\frac{\partial \ell_n^2(\varphi)}{\partial \alpha_2^2}=\sum_{i=1}^{n}\delta_i\left\{-\frac{\theta(e^{-\theta S_2(t_i)}-1)e^{-\theta S_2(t_i)}}{(e^{-\theta S_2(t_i)}-1)}\left[\frac{\partial^2 S_2(t_i)}{\partial \alpha_2^2}\right]-\frac{\theta^2 e^{-\theta S_2(t_i)}}{(e^{-\theta S_2(t_i)}-1)^2}\left[\frac{\partial S_2(t_i)}{\partial \alpha_2}\right]^2+\theta\frac{\partial^2 S_T(t_i)}{\partial \alpha_2^2}\right\}$$

$$+\sum_{i=1}^{n}\delta_i^*\left\{\frac{1}{h_2(t_i)^2}\left(h_2(t_i)\frac{\partial^2 h_2(t_i)}{\partial \alpha_2^2}-\left[\frac{\partial h_2(t_i)}{\partial \alpha_2}\right]^2\right)\right\}$$

$$+\sum_{i=1}^{n}\delta_i^*\left\{\frac{1}{S_2(t_i)^2}\left(S_2(t_i)\frac{\partial^2 S_2(t_i)}{\partial \alpha_2^2}-\left[\frac{\partial S_2(t_i)}{\partial \alpha_2}\right]^2\right)\right\}+\sum_{i=1}^{n}\delta_i^*\left\{-\theta\frac{\partial^2 S_2(t_i)}{\partial \alpha_2^2}+\theta\frac{\partial^2 S_T(t_i)}{\partial \alpha_2^2}\right\}$$

$$+\sum_{i=1}^{n}(1-\delta_i-\delta_i^*)\left\{\frac{-2\theta e^{-\theta S_2(t_i)}\frac{\partial^2 S_2(t_i)}{\partial \alpha_2^2}(e^{-\theta S_2(t_i)}-1)}{\log\left[1+\frac{(e^{-\theta S_2(t_i)}-1)^2}{(e^{-\theta}-1)}\right][e^{-\theta}-1+(e^{-\theta S_2(t_i)}-1)^2]}\right\}$$

$$+\sum_{i=1}^{n}(1-\delta_i-\delta_i^*)\left\{\frac{2\theta^2 e^{-\theta S_2(t_i)}\left[\frac{\partial S_2(t_i)}{\partial \alpha_2}\right]^2(2e^{-\theta S_2(t_i)}-1)}{\log\left[1+\frac{(e^{-\theta S_2(t_i)}-1)^2}{(e^{-\theta}-1)}\right][e^{-\theta}-1+(e^{-\theta S_2(t_i)}-1)^2]}\right\}$$

$$+\sum_{i=1}^{n}(1-\delta_i-\delta_i^*)\frac{\partial^2}{\partial \alpha_2^2}\log[S_T(t_i)],$$

$$\frac{\partial \ell_n^2(\varphi)}{\partial \gamma_1^2}=\sum_{i=1}^{n}\delta_i\left\{\frac{1}{h_1(t_i)^2}\left(h_1(t_i)\frac{\partial^2 h_1(t_i)}{\partial \gamma_1^2}-\left[\frac{\partial h_1(t_i)}{\partial \gamma_1}\right]^2\right)\right\}$$

$$+\sum_{i=1}^{n}\delta_i\left\{\frac{1}{S_1(t_i)^2}\left(S_1(t_i)\frac{\partial^2 S_1(t_i)}{\partial \gamma_1^2}-\left[\frac{\partial S_1(t_i)}{\partial \gamma_1}\right]^2\right)\right\}$$

$$+\sum_{i=1}^{n}\delta_i\left\{-\theta\frac{\partial^2 S_1(t_i)}{\partial \gamma_1^2}+\theta\frac{\partial^2 S_T(t_i)}{\partial \gamma_1^2}\right\}+\sum_{i=1}^{n}\delta_i^*\left\{-\frac{\theta(e^{-\theta S_1(t_i)}-1)e^{-\theta S_1(t_i)}}{(e^{-\theta S_1(t_i)}-1)}\left[\frac{\partial^2 S_1(t_i)}{\partial \gamma_1^2}\right]\right\}$$

$$+\sum_{i=1}^{n}\delta_i^*\left\{-\frac{\theta^2 e^{-\theta S_1(t_i)}}{(e^{-\theta S_1(t_i)}-1)^2}\left[\frac{\partial S_1(t_i)}{\partial \gamma_1}\right]^2+\theta\frac{\partial^2 S_T(t_i)}{\partial \gamma_1^2}\right\}$$

$$+\sum_{i=1}^{n}(1-\delta_i-\delta_i^*)\left\{\frac{-2\theta e^{-\theta S_1(t_i)}\frac{\partial^2 S_1(t_i)}{\partial \gamma_1^2}(e^{-\theta S_1(t_i)}-1)}{\log\left[1+\frac{(e^{-\theta S_1(t_i)}-1)^2}{(e^{-\theta}-1)}\right][e^{-\theta}-1+(e^{-\theta S_1(t_i)}-1)^2]}\right\}$$

$$+\sum_{i=1}^{n}(1-\delta_i-\delta_i^*)\left\{\frac{2\theta^2 e^{-\theta S_1(t_i)}\left[\frac{\partial S(t_i)}{\partial \gamma_1}\right]^2(2e^{-\theta S_1(t_i)}-1)}{\log\left[1+\frac{(e^{-\theta S_1(t_i)}-1)^2}{(e^{-\theta}-1)}\right][e^{-\theta}-1+(e^{-\theta S_1(t_i)}-1)^2]}\right\}$$

$$+\sum_{i=1}^{n}(1-\delta_i-\delta_i^*)\frac{\partial^2}{\partial\alpha_1^2}\log[S_T(t_i)],$$

$$\frac{\partial\ell_n^2(\varphi)}{\partial\gamma_2^2}=\sum_{i=1}^{n}\delta_i\left\{-\frac{\theta(e^{-\theta S_2(t_i)}-1)e^{-\theta S_2(t_i)}}{(e^{-\theta S_2(t_i)}-1)}\left[\frac{\partial^2 S_2(t_i)}{\partial\gamma_2^2}\right]-\frac{\theta^2 e^{-\theta S_2(t_i)}}{(e^{-\theta S_2(t_i)}-1)^2}\left[\frac{\partial S_2(t_i)}{\partial\gamma_2}\right]^2+\theta\frac{\partial^2 S_T(t_i)}{\partial\gamma_2^2}\right\}$$

$$+\sum_{i=1}^{n}\delta_i\left\{-\frac{\theta^2 e^{-\theta S_2(t_i)}}{(e^{-\theta S_2(t_i)}-1)^2}\left[\frac{\partial S_2(t_i)}{\partial\gamma_2}\right]^2+\theta\frac{\partial^2 S_T(t_i)}{\partial\gamma_2^2}\right\}$$

$$+\sum_{i=1}^{n}\delta_i^*\left\{\frac{1}{h_2(t_i)^2}\left(h_2(t_i)\frac{\partial^2 h_2(t_i)}{\partial\gamma_2^2}-\left[\frac{\partial h_2(t_i)}{\partial\alpha_2}\right]^2\right)\right\}$$

$$+\sum_{i=1}^{n}\delta_i^*\left\{\frac{1}{S_2(t_i)^2}\left(S_2(t_i)\frac{\partial^2 S_2(t_i)}{\partial\gamma_2^2}-\left[\frac{\partial S_2(t_i)}{\partial\gamma_2}\right]^2\right)\right\}+\sum_{i=1}^{n}\delta_i^*\left\{-\theta\frac{\partial^2 S_2(t_i)}{\partial\gamma_2^2}+\theta\frac{\partial^2 S_T(t_i)}{\partial\gamma_2^2}\right\}$$

$$+\sum_{i=1}^{n}(1-\delta_i-\delta_i^*)\frac{-2\theta e^{-\theta S_2(t_i)}\frac{\partial^2 S_2(t_i)}{\partial\gamma_2^2}(e^{-\theta S_2(t_i)}-1)}{\log\left[1+\frac{(e^{-\theta S_2(t_i)}-1)^2}{(e^{-\theta}-1)}\right][e^{-\theta}-1+(e^{-\theta S_2(t_i)}-1)^2]}$$

$$+\sum_{i=1}^{n}(1-\delta_i-\delta_i^*)\frac{2\theta^2 e^{-\theta S_2(t_i)}\left[\frac{\partial S_2(t_i)}{\partial\gamma}\right]^2(2e^{-\theta S_2(t_i)}-1)}{\log\left[1+\frac{(e^{-\theta S_2(t_i)}-1)^2}{(e^{-\theta}-1)}\right][e^{-\theta}-1+(e^{-\theta S_2(t_i)}-1)^2]}$$

$$+\sum_{i=1}^{n}(1-\delta_i-\delta_i^*)\frac{\partial^2}{\partial\gamma_2^2}\log[S_T(t_i)],$$

$$\frac{\partial\ell_n^2(\varphi)}{\partial\alpha_1\partial\gamma_1}=\sum_{i=1}^{n}\delta_i\left\{\frac{1}{h_1(t_i)^2}\left(h_1(t_i)\frac{\partial^2 h_1(t_i)}{\partial\alpha_1\partial\gamma_1}-\frac{\partial h_1(t_i)}{\partial\alpha_1}\frac{\partial h_1(t_i)}{\partial\gamma_1}\right)\right\}$$

$$+\sum_{i=1}^{n}\delta_i\left\{\frac{1}{S_1(t_i)^2}\left(S_1(t_i)\frac{\partial^2 S_1(t_i)}{\partial\alpha_1\partial\gamma_1}-\frac{\partial S_1(t_i)}{\partial\alpha_1}\frac{\partial S_1(t_i)}{\partial\gamma_1}\right)\right\}+\sum_{i=1}^{n}\delta_i\left\{-\theta\frac{\partial^2 S_1(t_i)}{\partial\alpha_1\partial\gamma_1}+\theta\frac{\partial^2 S_T(t_i)}{\partial\alpha_1\partial\gamma_1}\right\}$$

$$+\sum_{i=1}^{n}\delta_i^*\left\{-\frac{\theta(e^{-\theta S_1(t_i)}-1)e^{-\theta S_1(t_i)}}{(e^{-\theta S_1(t_i)}-1)}\left[\frac{\partial^2 S_1(t_i)}{\partial\alpha_1\partial\gamma_1}\right]-\frac{\theta^2 e^{-\theta S_1(t_i)}}{(e^{-\theta S_1(t_i)}-1)^2}\frac{\partial S_1(t_i)}{\partial\alpha_1}\frac{\partial S_1(t_i)}{\partial\gamma_1}+\theta\frac{\partial^2 S_T(t_i)}{\partial\alpha_1\partial\gamma_1}\right\}$$

$$+\sum_{i=1}^{n}\delta_i^*\left\{\frac{2\theta e^{-\theta S(t_i)}\left[\theta\frac{\partial S_1(t_i)}{\partial\alpha_1}\frac{\partial S_1(t_i)}{\partial\gamma_1}(2e^{-\theta S_1(t_i)}-1)-\frac{\partial^2 S_1(t_i)}{\partial\alpha_1\partial\gamma_1}(e^{-\theta S_1(t_i)}-1)\right]}{\log\left[1+\frac{(e^{-\theta S_1(t_i)}-1)^2}{(e^{-\theta}-1)}\right][e^{-\theta}-1+(e^{-\theta S_1(t_i)}-1)^2]}\right\}$$

$$+\sum_{i=1}^{n}(1-\delta_i-\delta_i^*)\frac{\partial^2}{\partial\alpha_1^2}\log[S_T(t_i)],$$

$$\frac{\partial\ell_n^2(\varphi)}{\partial\alpha_2\partial\gamma_2}=\sum_{i=1}^{n}\delta_i\left\{-\frac{\theta(e^{-\theta S_2(t_i)}-1)e^{-\theta S_2(t_i)}}{(e^{-\theta S_2(t_i)}-1)}\left[\frac{\partial^2 S_2(t_i)}{\partial\alpha_2\partial\gamma_2}\right]-\frac{\theta^2 e^{-\theta S_2(t_i)}}{(e^{-\theta S_2(t_i)}-1)^2}\frac{\partial S_2(t_i)}{\partial\alpha_2}\frac{\partial S_2(t_i)}{\partial\gamma_2}+\theta\frac{\partial^2 S_T(t_i)}{\partial\alpha_2\partial\gamma_2}\right\}$$

$$+\sum_{i=1}^{n}\delta_i\left\{-\frac{\theta^2 e^{-\theta S_2(t_i)}}{(e^{-\theta S_2(t_i)}-1)^2}\frac{\partial S_2(t_i)}{\partial\alpha_2}\frac{\partial S_2(t_i)}{\partial\gamma_2}+\theta\frac{\partial^2 S_T(t_i)}{\partial\alpha_2\partial\gamma_2}\right\}$$

$$+\sum_{i=1}^{n}\delta_i\left\{\frac{-2\theta e^{-\theta S_2(t_i)}\left(\frac{\partial^2 S_2(t_i)}{\partial\alpha_2\partial\gamma_2}(e^{-\theta S_2(t_i)}-1)+\theta\frac{\partial S_2(t_i)}{\partial\alpha_2}\frac{\partial S_2(t_i)}{\partial\gamma_2}(-2e^{-\theta S_2(t_i)}+1)\right)}{\log\left[1+\frac{(e^{-\theta S_2(t_i)}-1)^2}{(e^{-\theta}-1)}\right][e^{-\theta}-1+(e^{-\theta S_2(t_i)}-1)^2]}\right\}$$

$$+\sum_{i=1}^{n}\delta_i^*\left\{\frac{1}{h_2(t_i)^2}\left(h_2(t_i)\frac{\partial^2 h_2(t_i)}{\partial\alpha_2\partial\gamma_2}-\frac{\partial h_2(t_i)}{\partial\alpha_2}\frac{\partial h_2(t_i)}{\partial\gamma_2}\right)\right\}$$

$$+\sum_{i=1}^{n}\delta_i^*\left\{\frac{1}{S_2(t_i)^2}\left(S_1(t_i)\frac{\partial^2 S_2(t_i)}{\partial\alpha_2\partial\gamma_2}-\frac{\partial S_2(t_i)}{\partial\alpha_2}\frac{\partial S_2(t_i)}{\partial\gamma_2}\right)\right\}$$

$$+\sum_{i=1}^{n}\delta_i^*\left\{-\theta\frac{\partial^2 S_2(t_i)}{\partial\alpha_2\partial\gamma_2}+\theta\frac{\partial^2 S_T(t_i)}{\partial\alpha_2\partial\gamma_2}\right\}+\sum_{i=1}^{n}(1-\delta_i-\delta_i^*)\frac{\partial^2}{\partial\alpha_1^2}\log[S_T(t_i)],$$

$$\frac{\partial\ell_n^2(\varphi)}{\partial\gamma_1\partial\gamma_2}=\sum_{i=1}^{n}(\delta_i+\delta_i^*)\theta\frac{\partial S_T(t_i)}{\partial\gamma_1\partial\gamma_2}+\sum_{i=1}^{n}(1-\delta_i-\delta_i^*)\frac{\partial^2}{\partial\gamma_1\partial\gamma_2}\log[S_T(t_i)],$$

$$\frac{\partial\ell_n^2(\varphi)}{\partial\alpha_1\partial\alpha_2}=\sum_{i=1}^{n}(\delta_i+\delta_i^*)\theta\frac{\partial S_T(t_i)}{\partial\alpha_1\partial\alpha_2}+\sum_{i=1}^{n}(1-\delta_i-\delta_i^*)\frac{\partial^2}{\partial\alpha_1\partial\alpha_2}\log[S_T(t_i)],$$

$$\frac{\partial\ell_n^2(\varphi)}{\partial\alpha_1\partial\gamma_2}=\sum_{i=1}^{n}(\delta_i+\delta_i^*)\theta\frac{\partial S_T(t_i)}{\partial\alpha_1\partial\gamma_2}+\sum_{i=1}^{n}(1-\delta_i-\delta_i^*)\frac{\partial^2}{\partial\alpha_1\partial\gamma_2}\log[S_T(t_i)],$$

$$\frac{\partial\ell_n^2(\varphi)}{\partial\gamma_1\partial\alpha_2}=\sum_{i=1}^{n}(\delta_i+\delta_i^*)\theta\frac{\partial S_T(t_i)}{\partial\gamma_1\partial\alpha_2}+\sum_{i=1}^{n}(1-\delta_i-\delta_i^*)\frac{\partial^2}{\partial\gamma_1\partial\alpha_2}\log[S_T(t_i)],$$

where

$$\frac{\partial^2 h_j(t_i)}{\partial\alpha_j^2}=\frac{-2\gamma_j t_i}{(1+\alpha_j t_i)^3},\ \frac{\partial^2 S_j(t_i)}{\partial\alpha_j^2}=\gamma_j(\gamma_j+1)t_i^2(1+\alpha_j t_i)^{-(\gamma_j+2)},$$

$$\frac{\partial^2 S_j(t_i)}{\partial\gamma_j^2}=(1+\alpha_j t_i)^{-\gamma_j}[\log(1+\alpha_j t_i)]^2,\ \frac{\partial^2 h_j(t_i)}{\partial\gamma_j^2}=0,$$

$$\frac{\partial^2 h_j(t_i)}{\partial\alpha_j\partial\gamma_j}=\frac{1}{(1+\alpha_j t_i)^2},\ \frac{\partial^2 S_j(t_i)}{\partial\alpha_j\partial\gamma_j}=\frac{t_i[\gamma_j\log(1+\alpha_j t_i)-1]}{(1+\alpha_j t_i)^{\gamma_j+1}},$$

$$\frac{\partial^2 S_T(t_i)}{\partial\alpha_j^2}=\frac{[e^{-\theta}-1+(e^{-\theta S_j(t_i)}-1)(e^{-\theta S_{3-j}(t_i)}-1)]e^{-\theta S_j(t_i)}(e^{-\theta S_{3-j}(t_i)}-1)}{[e^{-\theta}-1+(e^{-\theta S_j(t_i)}-1)(e^{-\theta S_{3-j}(t_i)}-1)]^2}\left\{\frac{\partial^2 S_j(t_i)}{\partial\alpha_j^2}-\theta\left[\frac{\partial S_j(t_i)}{\partial\alpha_j}\right]^2\right\}$$

$$+\frac{\theta e^{-2\theta S_j(t_i)}(e^{-\theta S_{3-j}(t_i)}-1)^2}{[e^{-\theta}-1+(e^{-\theta S_j(t_i)}-1)(e^{-\theta S_{3-j}(t_i)}-1)]^2}\left[\frac{\partial S_j(t_i)}{\partial\alpha_j}\right]^2,$$

$$\frac{\partial^2 S_T(t_i)}{\partial\alpha_1\partial\alpha_2}=\frac{[e^{-\theta}-1+(e^{-\theta S_1(t_i)}-1)(e^{-\theta S_2(t_i)}-1)]e^{-\theta S_1(t_i)}e^{-\theta S_2(t_i)}}{[e^{-\theta}-1+(e^{-\theta S_1(t_i)}-1)(e^{-\theta S_2(t_i)}-1)]^2}\left\{-\theta\frac{\partial S_1}{\partial\alpha_1}\frac{\partial S_1}{\partial\alpha_2}\right\}$$

$$+\frac{\theta e^{-\theta S_1(t_i)}e^{-\theta S_2(t_i)}(e^{-\theta S_1(t_i)}-1)(e^{-\theta S_2(t_i)}-1)}{[e^{-\theta}-1+(e^{-\theta S_1(t_i)}-1)(e^{-\theta S_2(t_i)}-1)]^2}\left[\frac{\partial S_1}{\partial\alpha_1}\frac{\partial S_1}{\partial\alpha_2}\right],$$

$$\frac{\partial^2 S_T(t_i)}{\partial\gamma_j^2}=\frac{[e^{-\theta}-1+(e^{-\theta S_j(t_i)}-1)(e^{-\theta S_{3-j}(t_i)}-1)]e^{-\theta S_j(t_i)}(e^{-\theta S_{3-j}(t_i)}-1)}{[e^{-\theta}-1+(e^{-\theta S_j(t_i)}-1)(e^{-\theta S_{3-j}(t_i)}-1)]^2}\left\{\frac{\partial^2 S_j(t_i)}{\partial\gamma_j^2}-\theta\left[\frac{\partial S_j(t_i)}{\partial\gamma_j}\right]^2\right\}$$

$$+\frac{\theta e^{-2\theta S_j(t_i)}(e^{-\theta S_{3-j}(t_i)}-1)^2}{[e^{-\theta}-1+(e^{-\theta S_j(t_i)}-1)(e^{-\theta S_{3-j}(t_i)}-1)]^2}\left[\frac{\partial S_j(t_i)}{\partial\gamma_j}\right]^2,$$

$$\frac{\partial^2 S_T(t_i)}{\partial\gamma_1\partial\gamma_2}=\frac{[e^{-\theta}-1+(e^{-\theta S_1(t_i)}-1)(e^{-\theta S_2(t_i)}-1)]e^{-\theta S_1(t_i)}e^{-\theta S_2(t_i)}}{[e^{-\theta}-1+(e^{-\theta S_1(t_i)}-1)(e^{-\theta S_2(t_i)}-1)]^2}\left\{-\theta\frac{\partial S_1}{\partial\gamma_1}\frac{\partial S_1}{\partial\gamma_2}\right\}$$

$$+\frac{\theta e^{-\theta S_1(t_i)}e^{-\theta S_2(t_i)}(e^{-\theta S_1(t_i)}-1)(e^{-\theta S_2(t_i)}-1)}{[e^{-\theta}-1+(e^{-\theta S_1(t_i)}-1)(e^{-\theta S_2(t_i)}-1)]^2}\left[\frac{\partial S_1}{\partial\gamma_1}\frac{\partial S_1}{\partial\gamma_2}\right],$$

$$\frac{\partial^2 S_T(t_i)}{\partial\alpha_1\partial\gamma_1}=\frac{[e^{-\theta}-1+(e^{-\theta S_1(t_i)}-1)(e^{-\theta S_2(t_i)}-1)]e^{-\theta S_1(t_i)}(e^{-\theta S_2(t_i)}-1)}{[e^{-\theta}-1+(e^{-\theta S_1(t_i)}-1)(e^{-\theta S_2(t_i)}-1)]^2}\left\{\frac{\partial^2 S_1}{\partial\alpha_1\partial\gamma_1}-\theta\frac{\partial S_1}{\partial\alpha_1}\frac{\partial S_1}{\partial\gamma_1}\right\}$$

$$+\frac{\theta e^{-2\theta S_1(t_i)}(e^{-\theta S_2(t_i)}-1)^2}{[e^{-\theta}-1+(e^{-\theta S_1(t_i)}-1)(e^{-\theta S_2(t_i)}-1)]^2}\left[\frac{\partial S_1(t_i)}{\partial\alpha_1}\frac{\partial S_1(t_i)}{\partial\gamma_1}\right],$$

$$\frac{\partial^2 S_T(t_i)}{\partial\alpha_2\partial\gamma_2} = \frac{[e^{-\theta}-1+(e^{-\theta S_1(t_i)}-1)(e^{-\theta S_2(t_i)}-1)]e^{-\theta S_2(t_i)}(e^{-\theta S_1(t_i)}-1)}{[e^{-\theta}-1+(e^{-\theta S_1(t_i)}-1)(e^{-\theta S_2(t_i)}-1)]^2}\left\{\frac{\partial^2 S_2}{\partial\alpha_2\partial\gamma_2}-\theta\frac{\partial S_2}{\partial\alpha_2}\frac{\partial S_2}{\partial\gamma_2}\right\}$$
$$+\frac{\theta e^{-2\theta S_2(t_i)}(e^{-\theta S_1(t_i)}-1)^2}{[e^{-\theta}-1+(e^{-\theta S_1(t_i)}-1)(e^{-\theta S_2(t_i)}-1)]^2}\left[\frac{\partial S_2(t_i)}{\partial\alpha_2}\frac{\partial S_2(t_i)}{\partial\gamma_2}\right],$$

$$\frac{\partial^2 S_T(t_i)}{\partial\alpha_1\partial\gamma_2} = \frac{[e^{-\theta}-1+(e^{-\theta S_1(t_i)}-1)(e^{-\theta S_2(t_i)}-1)]e^{-\theta S_1(t_i)}e^{-\theta S_2(t_i)}}{[e^{-\theta}-1+(e^{-\theta S_1(t_i)}-1)(e^{-\theta S_2(t_i)}-1)]^2}\left\{-\theta\frac{\partial S_1}{\partial\alpha_1}\frac{\partial S_2}{\partial\gamma_2}\right\}$$
$$+\frac{\theta e^{-\theta S_1(t_i)}e^{-\theta S_2(t_i)}(e^{-\theta S_1(t_i)}-1)(e^{-\theta S_2(t_i)}-1)}{[e^{-\theta}-1+(e^{-\theta S_1(t_i)}-1)(e^{-\theta S_2(t_i)}-1)]^2}\left[\frac{\partial S_1}{\partial\alpha_1}\frac{\partial S_2}{\partial\gamma_2}\right],$$

$$\frac{\partial^2 S_T(t_i)}{\partial\alpha_2\partial\gamma_1} = \frac{[e^{-\theta}-1+(e^{-\theta S_1(t_i)}-1)(e^{-\theta S_2(t_i)}-1)]e^{-\theta S_1(t_i)}e^{-\theta S_2(t_i)}}{[e^{-\theta}-1+(e^{-\theta S_1(t_i)}-1)(e^{-\theta S_2(t_i)}-1)]^2}\left\{-\theta\frac{\partial S_2}{\partial\alpha_2}\frac{\partial S_1}{\partial\gamma_1}\right\}$$
$$+\frac{\theta e^{-\theta S_1(t_i)}e^{-\theta S_2(t_i)}(e^{-\theta S_1(t_i)}-1)(e^{-\theta S_2(t_i)}-1)}{[e^{-\theta}-1+(e^{-\theta S_1(t_i)}-1)(e^{-\theta S_2(t_i)}-1)]^2}\left[\frac{\partial S_2}{\partial\alpha_2}\frac{\partial S_1}{\partial\gamma_1}\right].$$

Appendix 4: The RNR algorithm under the SNBP model with fixed α_0

Step 1. Set initial value $(\alpha_1^{(0)},\alpha_2^{(0)},\gamma^{(0)})$.
Step 2. Repeat the Newton-Raphson iterations:

$$\begin{bmatrix}\alpha_1^{(k+1)}\\ \alpha_2^{(k+1)}\\ \gamma^{(k+1)}\end{bmatrix} = \begin{bmatrix}\alpha_1^{(k)}\\ \alpha_2^{(k)}\\ \gamma^{(k)}\end{bmatrix} - \begin{bmatrix}\frac{\partial^2\ell_n(\varphi)}{\partial\alpha_1^2} & \frac{\partial^2\ell_n(\varphi)}{\partial\alpha_1\partial\alpha_2} & \frac{\partial^2\ell_n(\varphi)}{\partial\alpha_1\partial\gamma}\\ \frac{\partial^2\ell_n(\varphi)}{\partial\alpha_1\partial\alpha_2} & \frac{\partial^2\ell_n(\varphi)}{\partial\alpha_2^2} & \frac{\partial^2\ell_n(\varphi)}{\partial\alpha_2\partial\gamma}\\ \frac{\partial^2\ell_n(\varphi)}{\partial\alpha_1\partial\gamma} & \frac{\partial^2\ell_n(\varphi)}{\partial\alpha_2\partial\gamma} & \frac{\partial^2\ell_n(\varphi)}{\partial\gamma^2}\end{bmatrix}^{-1}\begin{bmatrix}\frac{\partial\ell_n(\varphi)}{\partial\alpha_1}\\ \frac{\partial\ell_n(\varphi)}{\partial\alpha_2}\\ \frac{\partial\ell_n(\varphi)}{\partial\gamma}\end{bmatrix}\Bigg|_{\alpha_1=\alpha_1^{(k)},\alpha_2=\alpha_2^{(k)},\gamma=\gamma^{(k)}}$$

- If $\max\{|\alpha_1^{(k+1)}-\alpha_1^{(k)}|, |\alpha_2^{(k+1)}-\alpha_2^{(k)}|, |\gamma^{(k+1)}-\gamma^{(k)}|\}<\varepsilon$, stop the algorithm then set the MLE $\hat{\varphi}=(\alpha_1^{(k+1)},\alpha_2^{(k+1)},\gamma^{(k+1)})$.
- If the Hessian matrix is singular or $\max\{\alpha_1^{(k+1)}, \alpha_2^{(k+1)}, \gamma^{(k+1)}\}>d$, stop the algorithm then return to Step 1 with the initial value replaced by $(\alpha_1^{(0)}\times e^{u_1},\alpha_2^{(0)}\times e^{u_2},\gamma^{(0)}\times e^{u_3})$, where $u_i\sim U(-r_i,r_i)$, $i=1,2,3$ are independent uniform random variables.

Remark: We set $\varepsilon=10^{-5}$, $\alpha_1^{(0)}=1$, $\alpha_2^{(0)}=1$, $\gamma^{(0)}=1$, $r_1=6$, $r_2=6$,$r_3=6$ and $d=e^{10}$ for simulations and data analysis.

All the first- and second- derivatives of the log-likelihood are given below.

$$\frac{\partial\ell_n(\varphi)}{\partial\alpha_1}=\sum_{i=1}^n\left\{\frac{\delta_i}{\alpha_1+\alpha_0t_i}-\frac{(\delta_i+\delta_i^*+\gamma)t_i}{1+\alpha_1t_i+\alpha_2t_i+\alpha_0t_i^2}\right\},$$

$$\frac{\partial\ell_n(\varphi)}{\partial\alpha_2}=\sum_{i=1}^n\left\{\frac{\delta_i^*}{\alpha_2+\alpha_0t_i}-\frac{(\delta_i+\delta_i^*+\gamma)t_i}{1+\alpha_1t_i+\alpha_2t_i+\alpha_0t_i^2}\right\},$$

$$\frac{\partial\ell_n(\varphi)}{\partial\gamma}=\frac{m+m^*}{\gamma}-\sum_{i=1}^n\log(1+\alpha_1t_i+\alpha_2t_i+\alpha_0t_i^2),$$

$$\frac{\partial\ell_n^2(\varphi)}{\partial\alpha_1^2}=\sum_{i=1}^n\left\{\frac{-\delta_i}{(\alpha_1+\alpha_0t_i)^2}+\frac{(\delta_i+\delta_i^*+\gamma)t_i^2}{(1+\alpha_1t_i+\alpha_2t_i+\alpha_0t_i^2)^2}\right\},$$

$$\frac{\partial\ell_n^2(\varphi)}{\partial\alpha_2^2}=\sum_{i=1}^n\left\{\frac{-\delta_i^*}{(\alpha_2+\alpha_0t_i)^2}+\frac{(\delta_i+\delta_i^*+\gamma)t_i^2}{(1+\alpha_1t_i+\alpha_2t_i+\alpha_0t_i^2)^2}\right\},$$

$$\frac{\partial \ell_n^2(\varphi)}{\partial \gamma^2} = -\frac{m+m^*}{\gamma^2},\ \frac{\partial \ell_n^2(\varphi)}{\partial \alpha_1 \partial \alpha_2} = \sum_{i=1}^{n} \left\{ \frac{(\delta_i + \delta_i^* + \gamma) t_i^2}{(1+\alpha_1 t_i + \alpha_2 t_i + \alpha_0 t_i^2)^2} \right\},$$

$$\frac{\partial \ell_n^2(\varphi)}{\partial \alpha_1 \partial \gamma} = \sum_{i=1}^{n} \left\{ \frac{-t_i}{1+\alpha_1 t_i + \alpha_2 t_i + \alpha_0 t_i^2} \right\},\ \frac{\partial \ell_n^2(\varphi)}{\partial \alpha_2 \partial \gamma} = \sum_{i=1}^{n} \left\{ \frac{-t_i}{1+\alpha_1 t_i + \alpha_2 t_i + \alpha_0 t_i^2} \right\},$$

where $m = \sum_{i=1}^{n} \delta_i$ and $m^* = \sum_{i=1}^{n} \delta_i^*$.