Annals of the Institute of Statistical Mathematics (2021) 73:979–1010
https://doi.org/10.1007/s10463-020-00774-z

Generalized inverse-Gaussian frailty models with application to TARGET neuroblastoma data

Luiza S. C. Piancastelli[1,3] · Wagner Barreto-Souza[2,3] · Vinícius D. Mayrink[3]

Received: 11 April 2020 / Revised: 3 October 2020 / Accepted: 13 October 2020 /
Published online: 24 November 2020

Abstract
A new class of survival frailty models based on the generalized inverse-Gaussian (GIG) distributions is proposed. We show that the GIG frailty models are flexible and mathematically convenient like the popular gamma frailty model. A piecewise-exponential baseline hazard function is employed, yielding flexibility for the proposed class. Although a closed-form observed log-likelihood function is available, simulation studies show that employing an EM-algorithm is advantageous concerning the direct maximization of this function. Further simulated results address the comparison of different methods for obtaining standard errors of the estimates and confidence intervals for the parameters. Additionally, the finite-sample behavior of the EM-estimators is investigated and the performance of the GIG models under misspecification assessed. We apply our methodology to a TARGET (*Therapeutically Applicable Research to Generate Effective Treatments*) data about the survival time of patients with neuroblastoma cancer and show some advantages of the GIG frailties over existing models in the literature.

Keywords EM-algorithm · Frailty · Generalized inverse-Gaussian models · Neuroblastoma · Robustness

1 Introduction

When dealing with time to event data, the most popular statistical approach is the proportional hazards model by Cox (1972). This model is based on the hazard function and accommodates well censored and truncated data, which are key elements in survival analysis.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s10463-020-00774-z) contains supplementary material, which is available to authorized users.

✉ Wagner Barreto-Souza
wagner.barretosouza@kaust.edu.sa

Extended author information available on the last page of the article

One situation in which the proportional hazards model can be deficient occurs when unobserved sources of heterogeneity are present in the data. This might be explained by the lack of important covariates in the study, which are difficult to measure or were not collected because the researcher did not know its importance in the first place. In this case, the deficiency of the proportional hazards model is the assumption of a homogeneous population. Another common situation in which the proportional hazards model is problematic occurs when there is correlated survival data. The correlation arises, for example, when repeated measures are collected for each individual or when some common traits such as biological or environmental factors are shared.

The described situations are usually treated by assuming a frailty model configured as a natural extension to the Cox model. This approach introduces a latent random component that acts multiplicatively in the hazard function for an individual or a group of individuals. The univariate frailty modeling can handle unobserved sources of heterogeneity, and independence is assumed among individual frailties. One of the multivariate versions of this methodology known as the shared frailty model was introduced by Clayton (1978), which was motivated by the analysis of familial tendency in disease incidence. The author assumed that individuals in the same group share the frailty term and a positive dependence between those individuals is created.

In both univariate and multivariate versions, fitting the unobserved risk components is of most importance to properly evaluate the covariate effects. Some distributions in $\mathbb{R}^+$ are commonly assumed for the frailty term having certain desired mathematical properties since the inferential aspects of the frailty models pose additional difficulties in comparison with the usual mixed models due to censoring and truncation. The gamma distribution is the most common choice for this task. This is a model that became very popular due to its mathematical convenience, being explored by several authors such as Vaupel et al. (1979), Oakes (1982, (1986), Klein (1992) and Yashin et al. (1995). Other well-known options in this field are the parametric inverse-Gaussian (Hougaard 1984), positive stable (Hougaard 1986), log-normal (McGilchrist and Aisbett 1991) and power variance family (Crowder 1989; Hougaard et al. 1992) frailty models. Inference methods for shared frailty models are discussed for instance by Duchateau et al. (2002) and Vu and Knuiman (2002). For an account on computational tools with important practical information for users when dealing with shared frailty models, we refer to Hirsch and Wienke (2012).

Semiparametric versions of a frailty model are often preferred in the literature since it allows the estimation of regression effects without the need to explicitly impose a particular form for the baseline hazard function. Assuming a parametric formulation for this function can be restrictive since the analysis is bounded by the possible shapes of the chosen function. In addition, the parametric formulation can be difficult to identify or to test for adequacy. There are different methods to develop semiparametric versions of frailty models. For instance, the approach introduced by Klein (1992) is based on a modified EM-algorithm involving the Cox proportional hazards model. Other possible strategies are the penalized partial likelihood functions introduced by Therneau et al. (2003), piecewise constant hazards (Kim and Proschan 1991) with raising number of pieces and splines, among others.

We will focus on the piecewise constant hazards approach. Although this is a parametric model at all, it is quite flexible since we do not have to assume any specific form for the baseline hazard function. For a discussion about the advantages of this choice, we refer Lawless and Zhan (1998). Other contributions on frailty modeling are due to Balakrishnan and Peng (2006), Wang and Klein (2012), Callegaro and Iacobelli (2012), Chen et al. (2013), Putter and van Houwelingen (2015), Balakrishnan and Pal (2016), Leão et al. (2017), Barreto-Souza and Mayrink (2019), and Schneider et al. (2019), just to name a few. For a good account on frailty models, we recommend the books by Hougaard (2000), Wienke (2011) and Hanagal (2019).

Our chief goal in this paper is to introduce a new class of frailty models based on the generalized inverse-Gaussian (GIG) distributions, which has some advantages over existing models in the literature, as it will be shown along with the paper. Our proposal praises for efficiency in terms of computation and flexibility without compromising mathematical tractability, since all the main expressions have closed forms, like in the gamma model.

It is worth mentioning that both GIG and power variance function (PVF) families are infinitely divisible [for instance, see, respectively, Barndorff-Nielsen and Halgreen (1977) and Hanagal (2019)] so that a natural question can arise: Is there some connection between them? In fact, they share the inverse-Gaussian model as a common member (in our case this model is obtained by taking $\lambda = -1/2$; see Remark 1), but the other cases are different; the Laplace transform of the GIG distributions given in Eq. (2) of this paper is different from the PVF Laplace transform given in Section 5.5 from Hanagal (2019).

We now highlight some important contributions of this present paper as follows. (1) *Robustness.* In simulation studies, we show that the GIG distribution ensures the flexibility we seek, as the model can return optimal results when data are generated with different frailty distributions. (2) *Mathematical tractability.* We provide closed forms for the unconditional density, survival, and hazard functions related to the generalized inverse-Gaussian frailty model. Further, the conditional distribution of the frailty given the data is again GIG distributed. This kind of conjugacy, also true for the gamma case, is attractive, and it allows us to provide an EM-algorithm with closed E-step. Therefore, our proposed frailty models enjoy the same mathematical and analytical tractability of the gamma model. We are unaware of other existing frailty models having all these features. (3) *Cluster survival data.* Although the neuroblastoma data considered here is not clustered, we introduce our model in a more general setting allowing clustered data analysis. (4) *Computational tractability.* Some numerical problems are experienced in the data analysis through the gamma frailty model, in contrast with our proposed class of GIG models.

This paper is organized as follows. In Sect. 2, we introduce the class of GIG frailty models and obtain some basic results. In Sect. 3, we discuss maximum likelihood estimation and propose an EM-algorithm (Dempster et al. 1977) for estimating the parameters with a focus in the piecewise exponential baseline assumption, where the key expressions and detailed description are provided. In Sect. 4, we present comprehensive Monte Carlo simulation studies aiming at (a) motivation for the proposed EM-algorithm; (b) discussion and comparison of strategies for getting standard errors. Statistical analysis based on GIG frailty models to a TARGET

(*Therapeutically Applicable Research to Generate Effective Treatments*) data about the survival time of patients with neuroblastoma cancer is addressed in Sect. 5. Discussion about our findings and some points to be attacked in future works are presented in Sect. 6. This paper contains Supplementary Material, which include additional technical results and Monte Carlo simulations about the robustness of the GIG frailty model under misspecification.

2 Model specification and basic results

This section introduces the GIG frailty models. We begin by discussing the case without covariates. The inclusion of covariates will be addressed in next section. The generalized inverse-Gaussian distribution with parameters $\lambda \in \mathbb{R}$, $a > 0$ and $b > 0$ has the following density function:

$$g(x) = \frac{(a/b)^{\lambda/2}}{2K_\lambda(\sqrt{ab})} x^{\lambda-1} \exp\{-(ax + bx^{-1})/2\}, \quad \text{for} \quad x > 0, \tag{1}$$

where $K_\lambda(t) = \frac{1}{2}\int_0^\infty u^{\lambda-1} \exp\{-\frac{t}{2}(u + u^{-1})\}\mathrm{d}u$ denotes the third kind modified Bessel function, which is available in base `R` program (R Core Team 2020), more specifically via `besselK` function. A random variable X with density function (1) is denoted by $X \sim$ GIG (a, b, λ).

Some basic properties of the GIG distribution are presented next. The Laplace transform associated to (1) is given by

$$L(t) = \frac{K_\lambda(\sqrt{(a+2t)b})}{K_\lambda(\sqrt{ab})} \Big(\frac{a}{a+2t}\Big)^{\lambda/2}, \quad \text{for} \quad t > -a/2. \tag{2}$$

The moments of a GIG distributed random variable X are given by

$$E(X^k) = \frac{K_{\lambda+k}(\sqrt{ab})}{K_\lambda(\sqrt{ab})}(b/a)^{k/2}, \quad k \in \mathbb{R}. \tag{3}$$

Having a closed form for the Laplace transform is an advantage in the construction of frailty models since the marginal survival and density functions of the frailty depend on this function. The fact that the Laplace transform of the gamma distribution has a closed form is one of the reasons for the popularity of such a model. The main structure of our proposed class of frailty models is given in what follows.

Definition 1 Assume Z is a random variable following a GIG distribution with parameters $a = \alpha^{-1}$, $b = \alpha^{-1}$ and $\lambda > 0$, where $\alpha > 0$; we denote $Z \sim$ GIG $(\alpha^{-1}, \alpha^{-1}, \lambda)$. Consider the time to event of an individual to be denoted by a random variable T, where the latent effect Z acts multiplicatively in the baseline hazard function. Conditional on Z, the hazard function is given by $h(t|Z) = Z\, h_0(t)$. Here $h_0(t)$ is the baseline hazard function and Z is the frailty variable.

Remark 1 A motivation for assuming the GIG distribution for the frailty term is that, for a fixed λ, there are several known special cases of this distribution. The inverse-Gaussian (IG), reciprocal inverse-Gaussian (RIG), hyperbolic (HYP) and positive hyperbolic (PHYP) frailty models are obtained by choosing $\lambda = -1/2$, $\lambda = 1/2$, $\lambda = 0$ and $\lambda = 1$, respectively. Hence, by changing the value of λ, one can fit different frailty distributions to the data.

Remark 2 The parameterization considered for our frailty class is such that the IG case ($\lambda = -1/2$) satisfies $E(Z) = 1$ and $\text{Var}(Z) = \alpha$. This is the parameterization considered for IG frailty model in the literature; see Wienke (2011).

Through known results on frailty modeling, we can obtain the marginal survival function $S(\cdot)$ and density $f(\cdot)$ of T. In line with Definition 1, we denote by $h_0(t)$ the baseline hazard function and $H_0(t) = \int_0^t h_0(u)du$ the cumulative baseline hazard function. The marginal survival function is

$$S(t) = L(H_0(t)) = \left(\frac{1}{1+2\alpha H_0(t)}\right)^{\lambda/2} \frac{K_\lambda\left(\sqrt{\alpha^{-1}\left[\alpha^{-1}+2H_0(t)\right]}\right)}{K_\lambda(\alpha^{-1})}, \quad \text{for} \quad t > 0. \tag{4}$$

From (4) and by using the recurrence identity of the Bessel function $K_\nu(z) = \frac{z}{2\nu}[K_{\nu+1}(z) - K_{\nu-1}(z)]$, we obtain that the marginal density function assumes the form

$$f(t) = \frac{h_0(t)}{(1+2\alpha H_0(t))^{\frac{\lambda+1}{2}}} \frac{K_{\lambda+1}\left(\sqrt{\alpha^{-1}[\alpha^{-1}+2H_0(t)]}\right)}{K_\lambda(\alpha^{-1})}, \quad \text{for} \quad t > 0.$$

Next, we discuss another way of characterizing the frailty distribution that was introduced by Farrington et al. (2012), which is the relative frailty variance (hereafter RFV). The RFV is a measure of how the heterogeneity of the population evolves over time. This function can be used as a way to compare patterns of dependence among different frailty models and is obtained through the Laplace transform of the frailty distribution. Let $J(s) = \log L(-s)$, where $L(\cdot)$ is the Laplace transform and μ is the expected value of the frailty. Then, RFV $(s) = J''(-s/\mu)/J'(-s/\mu)^2$; see Eq. (3) from Farrington et al. (2012). The expressions required to calculate the RFV for the GIG frailty model are given in the Supplementary Material.

In Fig. 1, we present the evolution of the RFV(s) function for the IG, RIG, HYP, and PHYP frailties. For the four distributions under comparison, we find for each of them the root of the equation RFV $(0) - 0.7 = 0$. That is the value of α for which RFV(0) is equal to 0.7. In this way, we can better compare how their evolutions differ, since they have the same starting point.

The results of Fig. 1 evidence similarity within the GIG family regarding the RVF evolution. All special cases have shown decreasing RFV over time, which means that individuals who survive tend to be less heterogeneous (or less dependent) over

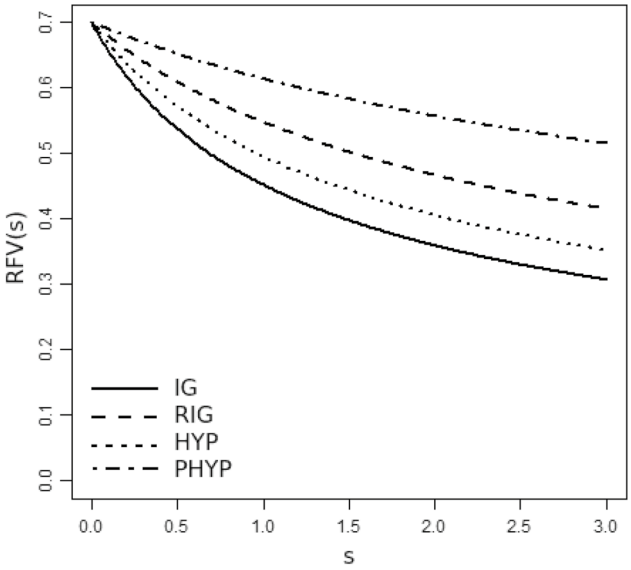

Fig. 1 Relative frailty variance evolution for IG, RIG, HYP and PHYP frailties with α chosen such that RFV(0) = 0.7 in each case

time. What differentiates among the distributions under comparison is the speed in which this happens. This can be observed in the curvature of the trajectory RFV (s) versus s. The most strong curvature is related to the inverse-Gaussian model. As a consequence, in this frailty distribution, the decrease in heterogeneity (or dependence) occurs more quickly. The PHYP frailty shows a slower decay. The HYP and RIG models have intermediate behaviors for the evolution of the relative frailty variance.

We end this section by calling attention to the fact that the proposed methodology is not restricted to the four options of λ discussed in Remark 1. These special cases are considered in the simulation studies developed ahead, and we believe that they provide enough material to understand the main aspects related to the GIG class. Another strategy indicated for practical applications, as we will consider for the neuroblastoma data analysis, is testing a grid of λ values and choosing the one that maximizes the log-likelihood function. In other words, the strategy is to perform a profile likelihood approach for estimating λ.

3 GIG frailty model for clustered survival data

This section deals with GIG frailty models allowing for a regression structure and also considers a framework extending the analysis to the multivariate case. In the univariate approach, each individual has its own frailty. This version is applicable, for example, when important covariates are missing from the analysis or when individual heterogeneity is naturally present. On the other hand, in shared frailty

models, individuals in a group share the frailty, which creates a positive dependence among them. With that, we are able to model correlated or clustered survival data. Evidently, this can be reduced to the univariate approach when groups are formed by one observation each. As mentioned previously, although the data considered here is univariate, we introduce our class of frailty models in a more general setting and, therefore, they can be applied for broad situations.

Assume m clusters with the ith cluster having n_i individuals, for $i = 1, \ldots, m$. Here T_{ij}^0 and C_{ij} denote the failure and censoring times for the individual $j = 1, \ldots, n_i$ in the ith cluster. The total sample size is $n = \sum_{i=1}^{m} n_i$. In addition, let $T_{ij} = \min\{T_{ij}^0, C_{ij}\}$ be the observable response variables and $\delta_{ij} = I\{T_{ij}^0 \leq C_{ij}\}$ the failure indicator. Naturally, the frailty Z_i is associated to the ith cluster. In order to complete the model specification, we make the assumptions that given Z_i, $\{(T_{ij}^0, C_{ij}); j = 1, \ldots, n_i\}$ are conditionally independent and that T_{ij}^0 and C_{ij} are independent for all j. Another assumption we rely on is that the censoring times within a cluster $\{C_{ij}; j = 1, \ldots, n_i\}$ are non-informative with respect to Z_i for all i.

Conditional on the frailty, the model has a structure similar to the proportional hazards model by Cox (1972). The frailty term Z_i is inserted as follows: $h(t_{ij}|Z_i) = Z_i h_0(t_{ij}) \exp(x_{ij}^\top \beta)$, for $t_{ij} > 0$, where x_{ij} is the vector of covariates of the jth individual from the ith cluster.

In the forthcoming discussion, we present maximum likelihood estimation for the parametric case, where a specific form for the baseline hazard function needs to be assumed. Further, we propose an EM-algorithm for estimating the model parameters with a focus on the piecewise exponential baseline hazard function, which provides model flexibility since we do not impose any specific shape for that function.

3.1 Maximum likelihood estimation

In this section, we present the unconditional survival and density functions to construct the likelihood function and perform maximum likelihood estimation under the parametric approach. The joint survival function of a cluster can be found through the Laplace transform of the frailty variable in (2). The joint survival function associated with the ith cluster is

$$\begin{aligned} S(t_{i1}, \ldots, t_{in_i}) &= L\left(\sum_{j=1}^{n_i} H_0(t_{ij}) e^{x_{ij}^\top \beta}\right) \\ &= \left(\frac{1}{1 + 2\alpha \sum_{j=1}^{n_i} H_0(t_{ij}) e^{x_{ij}^\top \beta}}\right)^{\lambda/2} \\ &\quad \frac{K_\lambda\left(\sqrt{\alpha^{-1}\left[\alpha^{-1} + 2\sum_{j=1}^{n_i} H_0(t_{ij}) e^{x_{ij}^\top \beta}\right]}\right)}{K_\lambda(\alpha^{-1})}, \end{aligned} \tag{5}$$

for $t_{ij} > 0 \ \forall i, j$. By using Lemma 1 from Supplementary Material, we obtain that the joint density associated to this survival function is

$$f(t_{i1}, \ldots, t_{in_i}) = \frac{\alpha^{-\lambda}}{K_\lambda(\alpha^{-1})}(2/\alpha)^{n_i}\chi^{(n_i)}\left(\alpha^{-1}\left[\alpha^{-1} + 2\sum_{j=1}^{n_i} H_0(t_{ij})e^{x_{ij}^\top\beta}\right]\right)\prod_{j=1}^{n_i} h_0(t_{ij})\, e^{x_{ij}^\top\beta}, \tag{6}$$

where $\chi^{(k)}(x) = (-1)^k \dfrac{\partial}{\partial x^k}\dfrac{K_\lambda(\sqrt{x})}{x^{\lambda/2}}$, for $k \in \mathbb{N}$.

Let $\theta = (\beta, \phi, \alpha)^\top$ be the parameter vector, where ϕ denotes the parameter vector associated to the baseline hazard function H_0, and $\ell(\theta)$ denotes the log-likelihood function. Here, λ is assumed to be fixed. Using the expressions in (5) and (6) together with Lemma 1 from the Supplementary Material, we can write the log-likelihood function as

$$\ell(\theta) = \log\left\{\prod_{i=1}^{m}\frac{\alpha^{-(\sum_{j=1}^{n_i}\delta_{ij}+\lambda)}}{K_\lambda(\alpha^{-1})}\Psi_{\lambda+\sum_{j=1}^{n_i}\delta_{ij}}\left(\alpha^{-1}\left[\alpha^{-1} + 2\sum_{j=1}^{n_i} H_0(t_{ij})e^{x_{ij}^\top\beta}\right]\right)\right.$$
$$\left.\prod_{j=1}^{n_i}\left[h_0(t_{ij})e^{x_{ij}^\top\beta}\right]^{\delta_{ij}}\right\}, \tag{7}$$

where $\Psi_\lambda(x) = K_\lambda(\sqrt{x})/x^{\lambda/2}$. In order to fit a GIG frailty model, we specify the baseline hazard function $h_0(\cdot)$ and obtain the parameter estimates by maximizing the log-likelihood function through numerical optimization methods. Some of them are available in `R` (R Core Team 2020) through the function `optim`; this includes the BFGS and Nelder-Mead methods (Fletcher 2000). Another possible strategy for this optimization is to use the Newton–Raphson algorithm (via the `R` function `nlm`) as discussed by Duchateau and Janssen (2008) and Emura et al. (2019).

A GIG frailty model can be fitted after specifying the form of the baseline risk function. For instance, a Weibull baseline hazard leads to the following formulation: $h_0(t) = \sigma\gamma t^{\gamma-1}$ and $H_0(t) = \sigma t^\gamma$, for $t > 0$, and $\sigma, \gamma > 0$. However, often is the case where one does not have enough information to be able to designate the form of $h_0(\cdot)$. In this case, we adopt a piecewise-constant hazard function (also known as piecewise exponential hazard function), a general framework that serves the purpose of approximating baseline-hazard functions of any shape, providing flexibility. This approach is commonly adopted and well accepted in the frailty models literature. For instance, Lawless and Zhan (1998) argue that the use of the piecewise-constant hazard function avoids many problems associated with nonparametric and semiparametric methods for incomplete survival data, but still provides a high degree of robustness. We will denote this approach by the PE-GIG frailty model, where PE stands for piecewise exponential baseline hazard function.

The baseline hazard function of a piecewise exponential distribution is given by

$$h_0(t) = \eta_l, \quad \text{for} \quad t^{(l-1)} \le t < t^{(l)} \quad \text{and} \quad l = 1, \ldots, k+1\,, \tag{8}$$

where k cut points (or knots) are specified satisfying $0 \equiv t^{(0)} < \min\{t_1, \ldots, t_n\} < t^{(1)} < \ldots < t^{(k)} < \max\{t_1, \ldots, t_n\} < t^{(k+1)} \equiv \infty, \quad \eta_l > 0$

for $l = 1, \ldots, k+1$, and $\eta = (\eta_1, \ldots, \eta_{k+1})^\top$. Hence, the cumulative baseline function is

$$H_0(t) = \sum_{j=0}^{i} \eta_{j+1}[\min\{t, t^{(j+1)}\} - t^{(j)}], \quad \text{for} \quad t^{(i)} \leq t < t^{(i+1)}.$$

Kim and Proschan (1991) introduced the estimation of the survival function through the piecewise exponential estimator; they considered as many partitions as the number of observed failures. However, we will follow the steps in Lawless and Zhan (1998) that showed, through simulation studies, that frailty models based on the piecewise constant hazards often provide satisfactory results for estimating parameters when 8-10 pieces are admitted. In this paper, our choice for the cut points/knots is data-driven through empirical quantiles of the uncensored survival times. More specifically, $t^{(i)}$ is the $\left(\frac{i}{k+1}\right)$-quantile of the uncensored times, for $i = 1, \ldots, k$. For instance, by taking $k = 3$, we have $t^{(1)}$, $t^{(2)}$, and $t^{(3)}$ being the first quartile, the median, and the third quartile, respectively.

3.2 EM-algorithm

In this section, we propose an Expectation-Maximization (EM) algorithm (Dempster et al. 1977) for estimating the parameters by assuming a piecewise exponential baseline hazard function $h_0(\cdot)$ as defined in (8). Although we can integrate out the latent frailties and obtain the parameter estimates by maximizing the observed log-likelihood in (7), our simulation studies have shown there are advantages in considering an EM approach instead. These studies will be presented in Sect. 4, illustrating how an EM estimation procedure provides better or equivalent parameter estimates in contrast with the direct optimization procedure, where the greatest distinction among the methods lies on the estimation of the frailty variance parameter.

An EM type of algorithm is appropriate to handle the presence of missing/latent data in the likelihood function and has been used in the literature of frailty models by Klein (1992), Wang and Klein (2012), Callegaro and Iacobelli (2012) and more recently by Barreto-Souza and Mayrink (2019) and Schneider et al. (2019), just to name a few. In the Expectation (E) step, we compute the conditional expectation of the complete log-likelihood function given the observed data, evaluated at the current parameter estimates; this is called Q-function. In the Maximization (M) step, we maximize the Q-function. The new estimates obtained in the M step are used to update the Q-function. We iterate between these two steps until a pre-specified convergence criterion is reached.

The complete data set is denoted by $(t_{ij}, \delta_{ij}, Z_i)$, for $j = 1, \ldots, n_i$ and $i = 1, \ldots, m$. We observe the pairs (t_{ij}, δ_{ij}) and the Z_i's are the latent random effects. The complete log-likelihood function can be written as $\ell_c(\theta) = \ell_1(\beta, \eta) + \ell_2(\alpha)$, where

$$\ell_1(\beta, \eta) \propto \sum_{i=1}^{m} \sum_{j=1}^{n_i} \delta_{ij}\left(x_{ij}^\top \beta + \log h_0(t_{ij})\right) - \sum_{i=1}^{m} \sum_{j=1}^{n_i} Z_i H_0(t_{ij}) e^{x_{ij}^\top \beta}$$

and

$$\ell_2(\alpha) \propto -m \log K_\lambda(\alpha^{-1}) - \frac{1}{2\alpha} \sum_{i=1}^{m} \left(Z_i + Z_i^{-1}\right).$$

In order to determine the E-step, we need to find the conditional distribution of Z_i given the observed data $\left\{(t_{ij}, \delta_{ij})\right\}_{j=1}^{n_i}$. By using basic properties of conditional densities, it can be shown that this conditional distribution is a GIG law. More specifically, we have that

$$Z_i | \left\{(t_{ij}, \delta_{ij})\right\}_{j=1}^{n_i} \sim \text{GIG}\left(\alpha^{-1} + 2\sum_{j=1}^{n_i} H_0(t_{ij}) e^{x_{ij}^\top \beta}, \alpha^{-1}, \lambda + \sum_{j=1}^{n_i} \delta_{ij}\right). \tag{9}$$

The Q-function is the conditional expected value of the complete log-likelihood function given the observed data at the current estimated parameter values. In other words, we can write

$$\begin{aligned} Q(\theta, \theta^{(r)}) \equiv & E_{\theta^{(r)}}\left(\ell_c(\theta) \middle| \{(t_{ij}, \delta_{ij})\}, j = 1, \ldots, n_i,\ i = 1, \ldots, m\right) \\ & \propto Q_1(\beta, \eta; \theta^{(r)}) + Q_2(\alpha; \theta^{(r)}), \end{aligned}$$

where $\theta^{(r)}$ denotes the estimated vector of parameters of the EM-algorithm in the rth step,

$$Q_1(\beta, \eta; \theta^{(r)}) = \sum_{i=1}^{m} \sum_{j=1}^{n_i} \delta_{ij} \left(x_{ij}^\top \beta + \log h_0(t_{ij})\right) - \sum_{i=1}^{m} \sum_{j=1}^{n_i} \omega_i(\theta^{(r)}) H_0(t_{ij}) e^{x_{ij}^\top \beta}$$

and

$$Q_2(\alpha; \theta^{(r)}) = -m \log K_\lambda(\alpha^{-1}) - \frac{1}{2\alpha} \sum_{i=1}^{m} \left(\omega_i(\theta^{(r)}) + \kappa_i(\theta^{(r)})\right),$$

with $\omega_i(\theta) \equiv E\left(Z_i | \{(t_{ij}, \delta_{ij})\}_{j=1}^{n_i}\right)$ and $\kappa_i(\theta) \equiv E\left(Z_i^{-1} | \{(t_{ij}, \delta_{ij})\}_{j=1}^{n_i}\right)$, for $i = 1, \ldots, m$. Expressions for these conditional expectations are presented in next proposition.

Proposition 1 (E-step of the EM-algorithm) *For $i = 1, \ldots, m$, we have that*

$$\omega_i(\theta) = \frac{K_{\lambda + \sum_{j=1}^{n_i} \delta_{ij} + 1}\left(\sqrt{\alpha^{-1}[\alpha^{-1} + 2\sum_{j=1}^{n_i} H_0(t_{ij}) e^{x_{ij}^\top \beta}]}\right)}{K_{\lambda + \sum_{j=1}^{n_i} \delta_{ij}}\left(\sqrt{\alpha^{-1}[\alpha^{-1} + 2\sum_{j=1}^{n_i} H_0(t_{ij}) e^{x_{ij}^\top \beta}]}\right)} \left(1 + 2\alpha \sum_{j=1}^{n_i} H_0(t_{ij}) e^{x_{ij}^\top \beta}\right)^{-1/2}$$

and

$$\kappa_i(\theta) = \frac{K_{\lambda+\sum_{j=1}^{n_i}\delta_{ij}-1}\left(\sqrt{\alpha^{-1}[\alpha^{-1}+2\sum_{j=1}^{n_i}H_0(t_{ij})e^{x_{ij}^\top\beta}]}\right)}{K_{\lambda+\sum_{j=1}^{n_i}\delta_{ij}}\left(\sqrt{\alpha^{-1}[\alpha^{-1}+2\sum_{j=1}^{n_i}H_0(t_{ij})e^{x_{ij}^\top\beta}]}\right)}\left(1+2\alpha\sum_{j=1}^{n_i}H_0(t_{ij})e^{x_{ij}^\top\beta}\right)^{1/2}.$$

Proof It follows immediately by using (3) and (9). □

Using the previous expressions and providing a fixed set of cut points for the piecewise exponential hazard function, we have fully defined our Q-function. The associated score function to the Q-function $\partial Q(\theta;\theta^{(r)})/\partial\theta$ has its components given by

$$\frac{\partial Q_1(\theta;\theta^{(r)})}{\partial\beta_s} = \sum_{i=1}^{m}\sum_{j=1}^{n_i}\delta_{ij}x_{ijs} - \sum_{i=1}^{m}\sum_{j=1}^{n_i}\omega_i(\theta^{(r)})H_0(t_{ij})e^{x_{ij}^\top\beta}x_{ijs}, \quad s=1,\ldots,p,$$

$$\frac{\partial Q_1(\theta;\theta^{(r)})}{\partial\eta_s} = \sum_{i=1}^{m}\sum_{j=1}^{n_i}\delta_{ij}\frac{h_0^{(s)}(t_{ij})}{h_0(t_{ij})} - \sum_{i=1}^{m}\sum_{j=1}^{n_i}\omega_i(\theta^{(r)})H_0^{(s)}(t_{ij})e^{x_{ij}^\top\beta}, \quad s=1,\ldots,k+1,$$

and

$$\frac{\partial Q_2(\theta;\theta^{(r)})}{\partial\alpha} = \frac{m}{\alpha^2}\frac{K'_\lambda(\alpha^{-1})}{K_\lambda(\alpha^{-1})} + \frac{1}{2\alpha^2}\sum_{i=1}^{m}\left(\omega_i(\theta^{(r)})+\kappa_i(\theta^{(r)})\right),$$

where

$$h_0^{(s)}(t) \equiv \frac{\partial h_0(t)}{\partial\eta_s} = \begin{cases} 1, & \text{for } t^{(s-1)} \le t < t^{(s)}, \\ 0, & \text{otherwise}, \end{cases}$$

$$H_0^{(s)}(t) \equiv \frac{\partial H_0(t)}{\partial\eta_s} = \begin{cases} 0, & \text{for } t \le t^{(s-1)}, \\ \min\{t, t^{(s)}\} - t^{(s-1)}, & \text{for } t > t^{(s-1)}, \end{cases}$$

for $s = 1,\ldots,k+1$.

The EM-algorithm establishes simpler expressions to be maximized rather than running a direct maximization procedure based on the observed likelihood function in (7). More specifically, in the EM approach, the Bessel function is only involved in the maximization regarding α, which relies on a problem of finding the root of a nonlinear equation. On the other hand, the maximization of the observed log-likelihood function (7) is more cumbersome since the argument of the Bessel function involves all the parameters. In addition, applying the EM approach allows us to do the optimization separately for (β,η) and α.

A complete description of the EM-algorithm for the GIG frailty models is given in Algorithm 1 including initial guesses for initialization of the procedure.

Algorithm 1 EM-algorithm for the class of PE-GIG frailty models

1: Provide initial guesses for the parameters. Denoting $\theta^{(r)}$ as the estimate of the set of parameters at step r, $\beta^{(0)}$ will be the coefficients obtained by fitting the proportional hazards model through the `R` package `survival` (Therneau 2015). Set $\alpha^{(0)} = 1$ and $\eta_k^{(0)} = \sum_{i=1}^{m}\sum_{j=1}^{n_i}\delta_{ij}/\sum_{i=1}^{m}\sum_{j=1}^{n_i}t_{ij}$, if t_{ij} is a failure time for $t^{(l)} \leq t_{ij} < t^{(l+1)}$. The term $t^{(l)}$ represents the l-th change point, for $l = 1, \ldots, k$. The expression for $\eta_k^{(0)}$ is the maximum likelihood estimator of the rate of an exponential distribution in each given interval using only the failure times within that interval.

2: Update the Q-function using the expressions found for $\omega_i(\theta)$ and $\kappa_i(\theta)$ with $\theta = \widehat{\theta}^{(r)}$.

3: Use the expressions of $h_0(t)$ and $H_0(t)$, based on the piecewise exponential distribution, and numerically obtain the maximum likelihood estimates of β and η by maximizing Q_1. For this task, we chose to work with the `optim` function in `R` under the BFGS maximization routine (Fletcher 2000). The new estimates of the parameters are denoted by $\widehat{\beta}^{(r+1)}$ and $\widehat{\eta}^{(r+1)}$.

4: Find the maximum likelihood estimate for α by maximizing Q_2 numerically. Again, we use `optim` with BFGS. The new estimate is denoted by $\widehat{\alpha}^{(r+1)}$.

5: Verify a convergence criterion. For example, $||\theta^{(r+1)} - \theta^{(r)}|| < \epsilon$ for some pre-established $\epsilon > 0$. If the convergence criterion is satisfied, set $\theta^{(r+1)}$ as the final parameter estimates. Otherwise, update $\theta^{(r)}$ with $\theta^{(r+1)}$ and return to Step 2.

We now discuss some strategies of how to obtain standard errors of the parameter estimates. Let $\theta_* = (\beta, \alpha)^\top$. Following the procedure discussed in Klein (1992), we get the standard error through the partial observed information matrix $J_n(\theta_*) \equiv -\partial^2\ell(\theta)/\partial\theta_*\partial\theta_*^\top$, where η is treated as a nuisance parameter vector. The explicit elements of this matrix are provided in Appendix A.1. The strategy to use partial information instead of the total information is common in frailty models and aims at avoiding numerical issues due to possible high-dimensional matrices to be inverted. After getting convergence of the EM-algorithm, we replace (β, η, α) by their EM-estimates in $J_n(\theta_*)$ to assess the standard errors. We also consider the partial information matrix $I_n(\theta_*)$ given in (10) due to the EM-algorithm (Louis 1982), where its elements are provided in Appendix 1.

Another important strategy to obtain the standard errors of the parameter estimates is the nonparametric bootstrap (Efron 1979). The bootstrap replicas are obtained by resampling the survival data with replacement. When data are clustered and the number of clusters is moderate or large, we suggest resampling the clusters. The case where the number of clusters is small, one way is to perform the two-step bootstrap method discussed by Xiao and Abrahamowicz (2010), which involves both resamplings from the clusters and from individuals within each selected cluster. We call the attention that the knots, to be considered when fitting the PE-GIG frailty model for the bootstrap samples, will be probably different among the replications. This is due to our data-driven choice, which was discussed at the end of Sect. 3.1.

The methods for getting the standard errors discussed here will be compared and contrasted via Monte Carlo simulation in Sect. 4.2. As we will see, the bootstrap method provides much better results than the other two approaches based on the observed information matrices. Recommendations for the construction of bootstrap-based confidence intervals for the parameters are also provided and evaluated in Sect. 4.2.

4 Simulation studies

In this section, we present Monte Carlo studies aiming at (a) motivation for the proposed EM-algorithm and (b) discussion and comparison of strategies for getting standard errors, in Sects. 4.1 and 4.2, respectively. Simulated results about the robustness of the GIG frailty models under misspecification are addressed in the Supplementary Material.

4.1 Comparison between direct maximization and EM-algorithm

In this section, we illustrate through simulation studies that employing the proposed EM-algorithm for finding the maximum likelihood estimates of GIG frailty model parameters is advantageous in contrast with the direct maximization of the observed log-likelihood function, hereby denoted by DMLE. To do this, we simulate 1000 synthetic data sets under the four GIG special cases and, under the correctly specified frailty distribution, estimate the parameters using both methods. In this experiment, the setting without clustering structure is investigated, that is, $n_i = 1\,\forall i$, and the sample sizes are set to $m = 200$ and $m = 400$. For the scenarios where $m = 200$, 5 cutoff points are stipulated for the piecewise constant hazards function, based on data quantiles. As the number of observations increases to 400, we also increase the number of cutoff points to $k = 10$.

The data is simulated as follows. Given the frailties, the failure times T_{ij}^0 were simulated from a Weibull distribution with $\gamma = 2$ and $\sigma = 0.25$. The censoring times C_{ij} were generated, independently of T_{ij}^0, following a Weibull distribution with parameters $\gamma = 2$ and $\sigma = 0.05$. The observed data are $T_{ij} = \min\{T_{ij}^0, C_{ij}\}$ and $\delta_{ij} = I\{T_{ij}^0 \leq C_{ij}\}$. With this configuration, we have a percentage of approximately 30% of censored observations. We employ two covariates that were generated from the Bernoulli(0.5) and Uniform$(-1, 1)$ distributions with the true values of their effects being $(\beta_1, \beta_2) = (1.5, -1)$. The true value of α was set to 0.5.

Table 1 displays the results of this comparison study. Under each frailty model, mean and standard errors (in parenthesis) of the parameter estimates are given for the EM and DMLE methods. On the left side, the results of the study with sample size equal to 200 are shown, while the right side columns contain the results for the same simulation study performed with $m = 400$. It can be seen that the mean parameter estimates obtained by applying the EM-algorithm are considerably closer to the true values, something that holds for all frailty distributions in consideration. When m is 200, the frailty variance parameter is underestimated by both methods. However, the estimation bias is much larger under the direct maximization procedure. As expected, such bias decreases as the sample size grows, improving the average estimation in both cases. Nonetheless, the distinction between the two estimation procedures is evident. The mean estimates of α obtained using the EM-algorithm approaches to the true value for all frailty distributions, contrarily to the DMLE case in which this parameter is still largely underestimated. Although the most striking difference among the methods lies in the estimation of α, it is seen that the EM

Table 1 Mean and standard error (in parenthesis) of parameter estimates based on the EM-algorithm and direct maximization of the log-likelihood function (DMLE) among GIG frailty models

Frailty	Method	$\mathbf{m} = 200 (k = 5)$			$\mathbf{m} = 400 (k = 10)$		
		$\alpha = 0.5$	$\beta_1 = 1.5$	$\beta_2 = -1$	$\alpha = 0.5$	$\beta_1 = 1.5$	$\beta_2 = -1$
IG	EM	0.361 (0.213)	1.423 (0.216)	−0.989 (0.337)	0.471 (0.214)	1.477 (0.167)	−0.989 (0.243)
	DMLE	0.117 (0.187)	1.336 (0.199)	−0.785 (0.319)	0.237 (0.281)	1.374 (0.178)	−0.866 (0.233)
HYP	EM	0.305 (0.177)	1.403 (0.218)	−0.986 (0.316)	0.432 (0.224)	1.466 (0.165)	−0.997 (0.256)
	DMLE	0.128 (0.226)	1.336 (0.221)	−0.812 (0.336)	0.256 (0.323)	1.375 (0.185)	−0.910 (0.258)
RIG	EM	0.288 (0.162)	1.395 (0.209)	−0.968 (0.329)	0.433 (0.366)	1.461 (0.170)	−0.975 (0.245)
	DMLE	0.134 (0.241)	1.313 (0.216)	−0.855 (0.355)	0.281 (0.273)	1.379 (0.194)	−0.907 (0.256)
PHYP	EM	0.303 (0.198)	1.406 (0.211)	−0.957 (0.321)	0.430 (0.216)	1.470 (0.167)	−0.989 (0.232)
	DMLE	0.145 (0.277)	1.325 (0.221)	−0.872 (0.331)	0.264 (0.325)	1.377 (0.190)	−0.923 (0.233)

procedure is also preferred for the regression parameters. It yields parameter estimates that are closer to the true values, in comparison to the DMLE procedure.

Figure 2 illustrates the results of the simulation study when $m = 400$ through boxplots of the parameter estimates. From this figure, we can observe the superior performance of the EM-algorithm for estimating all the parameters over the direct maximization approach, which does not produce satisfactory results even for $m = 400$, under the four frailty models considered.

We conclude that although we can integrate out the unobserved frailty values and estimate the parameters by maximizing directly the observed log-likelihood

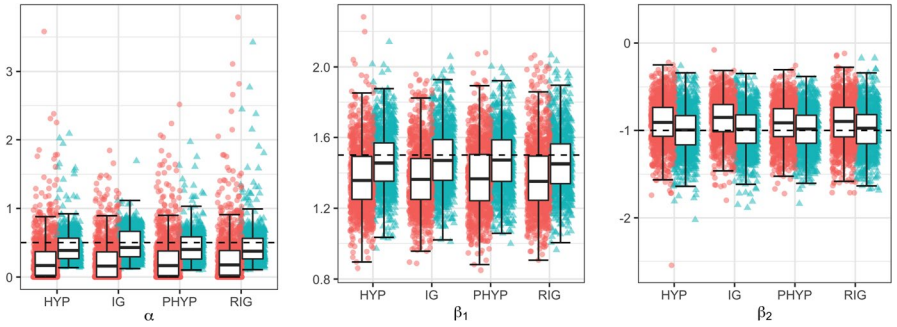

Fig. 2 Boxplots of the maximum likelihood estimates obtained with direct maximization (∘) and EM-algorithm (△) under different GIG frailty models with $m = 400$. Dashed line corresponds to the true value of the parameter

function, better results are achieved by employing an EM-algorithm. Through Monte Carlo simulation studies, we illustrated how this procedure yields a superior estimation of both the frailty variance parameter and the regression coefficients and hence should be preferred for the class of GIG frailty models.

4.2 Standard error studies

Having found that the EM-algorithm is favored for the estimation of the GIG frailty model parameters, we now investigate the derivation of their standard errors and confidence intervals. This is a problem on which there is no agreement in the frailty models literature, and the best solution seems to depend on the problem at hand. For instance, Wang and Klein (2012) and Barreto-Souza and Mayrink (2019), who also estimate the parameters using an EM-algorithm, utilize the observed information matrix from the observed log-likelihood function for this task. Contrastingly, in the model proposed by Callegaro and Iacobelli (2012), an EM approach is used for estimation, but the strategy for obtaining the standard error differs. The authors consider a reduced version of the EM information matrix (Louis 1982), a strategy which is indicated by Abrahantes and Burzykowski (2005). Finally, in the work by Balakrishnan and Pal (2016), parameters are estimated by maximization of a numerical approximation of the observed log-likelihood function, and their standard errors computed via bootstrap. According to the authors, a bootstrap method is preferred to the Hessian approach, as the second underestimates the variability of the parameters.

We propose to assess this for the class of GIG frailty models via a small simulation study where different approaches are compared and contrasted. In the method to be referred to as Hessian, standard errors are computed according to the Hessian of the observed log-likelihood function, with the parameters of the PE baseline hazard function taken as a nuisance. This is calculated analytically, with the expressions provided in Appendix A.1. A second approach works similarly, utilizing the information matrix of the EM-algorithm by Louis (1982). We provide the expressions required to obtain this matrix in Appendix A.2. Then, confidence intervals for the Hessian and EM procedures are constructed under asymptotic normality. A third and last approach is based on nonparametric bootstrap. In this case, we sample the observations with replacement and fit the model to a certain number of resampled data sets. The bootstrap sample size is taken as the number of observations in the original data set. After this procedure is repeated B times, we obtain bootstrap estimates of the standard errors. In the bootstrap case, we will assess two types of confidence intervals. We denote by *percentile bootstrap* the confidence intervals which are constructed based on the quantiles of the B bootstrap parameter estimates. Alternatively, a confidence interval based on asymptotic normality employs the bootstrap estimates of the standard errors. We refer to the former method as *normal-approximation bootstrap*. In the proposed study, 500 Monte Carlo replicas are considered, and the data are generated under the IG and HYP frailties. The synthetic data are drawn with the same configurations of Sect. 4.1, that is, $\beta_1 = 1.5$, $\beta_2 = -1$, and $\alpha = 0.5$, with sample size $m = 400$ ($n_i = 1 \ \ \forall i$), and $k = 10$. Additionally, bootstrap-based procedures use $B = 200$ replicas.

We first focus on the standard errors of the parameter estimates, where the mean estimates of these quantities under the Hessian, EM, and bootstrap methodologies are contrasted with the Monte Carlo standard deviations. For each synthetic data set, we fit the correctly specified GIG frailty model via the proposed EM-algorithm and calculate their standard errors via the three competing approaches. In Table 2, the empirical standard errors obtained from the bootstrap, Hessian, and EM approaches are reported for each parameter. Ideally, these would resemble the Monte Carlo standard deviations in column MC, which stands for Monte Carlo.

It is observed that the Hessian and EM approaches significantly underestimate the standard deviation (SD) of the estimated regression coefficients, something that holds for both the IG and HYP scenarios. This is evidence by looking at the results related to the parameter β_2, with associated continuous covariate, for which there is a striking difference among the MC and Hessian/EM methods. On the other hand, the empirical bootstrap standard error is very close to the Monte Carlo standard deviation for both estimates of β_1 and β_2, as desired. Concerning the frailty variance parameter α, the EM and bootstrap approaches are competitive and advantageous over the Hessian method. Through this simulation study, we can conclude that the bootstrap approach is preferred for calculating the standard errors of the GIG frailty model parameter estimates. Contrarily to the other methods under comparison, the bootstrap approach produced satisfactory results for the estimated regression coefficients standard errors, which is of great interest in practical applications. Additionally, a reasonable result was achieved for the standard errors related to the frailty variance parameter estimates.

We now focus our attention on the construction of confidence intervals for the parameters. Three intervals based on asymptotic normality are evaluated, utilizing the EM, Hessian, and bootstrap approaches. Additionally, we include the percentile-based bootstrap confidence interval. Table 3 displays the results of the proposed study for data generated with frailty IG distributed. For the significance levels of 1%, 5%, and 10%, the percentage of times the confidence interval derived under each methodology contains the true parameter value is reported. It is observed that the coverages due to the normal-approximation bootstrap method are near the nominal ones for the three significance levels under consideration. The percentile bootstrap method behaves well for β_1, related to the binary covariate, but has a worse performance for β_2, related to the continuous regressor. However, it is still superior to the Hessian and EM methods in both cases. Notably,

Table 2 Standard deviations of the estimates via Monte Carlo (MC) and empirical standard errors obtained from the bootstrap, Hessian, and EM methods under the HYP and IG frailty models

Model	Par.	MC	Bootstrap	Hessian	EM
HYP	β_1	0.164	0.170	0.118	0.118
	β_2	0.254	0.249	0.152	0.152
	α	0.226	0.193	0.170	0.181
IG	β_1	0.171	0.173	0.118	0.119
	β_2	0.254	0.253	0.154	0.155
	α	0.227	0.183	0.174	0.189

Table 3 Empirical confidence interval coverages based on different methods under the IG frailty model parameters with significance level (sig.) at 1%, 5%, and 10%

Sig.	Par.	Percentile bootstrap	Normal-approx. bootstrap	Hessian	EM
1%	β_1	98.6	98.2	91.4	91.4
	β_2	95.0	99.4	88.0	88.2
	α	98.2	99.4	99.0	99.2
5%	β_1	95.6	93.2	81.2	81.6
	β_2	87.4	94.4	76.0	76.0
	α	94.6	91.8	83.0	97.2
10%	β_1	91.8	89.0	74.6	75.4
	β_2	79.0	88.8	71.0	71.0
	α	89.0	79.6	74.4	86.2

these two approaches underestimate the standard errors of the parameters, creating narrow confidence intervals whose coverages are far from the desired nominal level in general. Good results are yielded by the Hessian and EM methods for the frailty variance parameter when the significance level is 1%. However, this behavior is not persistent, as evidenced by the results of both methods when the nominal level is 10%. The best performance for α is due to the percentile bootstrap method. Likely, a normality approximation is reasonable for the regression parameter estimates, but not for estimates of α, causing the percentile-based approach to be favored. We will explore this further, but first, we analyze the results of this study when the HYP frailty is considered, reported in Table 4.

A similar pattern is found in the second scenario of our study. For all nominal levels, the bootstrap confidence intervals are preferred to the Hessian and EM approaches. Moreover, the normal-approximation bootstrap confidence intervals yield coverages that are very close to the desired levels for both parameters β_1 and β_2 associated with the regressors. As before, the only method producing consistent and satisfactory results for α among all confidence levels is the percentile-based bootstrap.

Table 4 Empirical confidence interval coverages based on different methods under the HYP frailty model parameters with significance level (sig.) at 1%, 5%, and 10%

Sig.	Par.	Percentile bootstrap	Normal-approx. bootstrap	Hessian	EM
1%	β_1	99.0	99.0	92.0	92.0
	β_2	95.0	99.4	88.4	88.4
	α	98.0	98.6	97.6	99.4
5%	β_1	96.4	93.6	84.4	84.4
	β_2	88.6	94.8	78.8	79.0
	α	93.4	84.2	78.4	94.6
10%	β_1	92.6	90.0	77.6	78.0
	β_2	81.4	89.6	69.6	69.8
	α	88.8	71.8	67.6	75.8

Figure 3 displays histograms of the standardized (by the mean and standard deviation) parameter estimates overlayed with the standard normal density curve. The top and bottom rows correspond to the standardized $\widehat{\beta}_1$, $\widehat{\beta}_2$, and $\widehat{\alpha}$ estimates obtained in the IG and HYP scenarios, respectively. This figure supports our claim that the estimates of α are further from the normality, causing the worse performance of the confidence intervals based on normality. Contrastingly, Gaussianity approximation seems to be a reasonable assumption for the regression parameter estimators.

In this subsection, different methods to gather standard errors and construct confidence intervals for the GIG frailty model parameters were compared. Monte Carlo simulation studies evidenced how the standard errors obtained via the Hessian and EM approaches tend to be underestimated. Confidence intervals for regression coefficients are of high interest in practical applications, but these commonly used methods produce liberal confidence intervals for the GIG frailty model parameters. Our study showed that the bootstrap options are superior, of which the normal-based yielded the best results in terms of the regression parameters. Additionally, we identified that the frailty variance parameter estimates are further from the asymptotic normality assumption, causing a percentile-based interval to be preferred.

Following these arguments, we recommend that standard errors and confidence intervals for the parameters are derived using bootstrap. More specifically, we advise that normality-based intervals are employed for the regression effects, and a percentile-based procedure is taken for the frailty variance parameter. We highlight that the same bootstrap replicas are used for the construction of these two types of intervals,

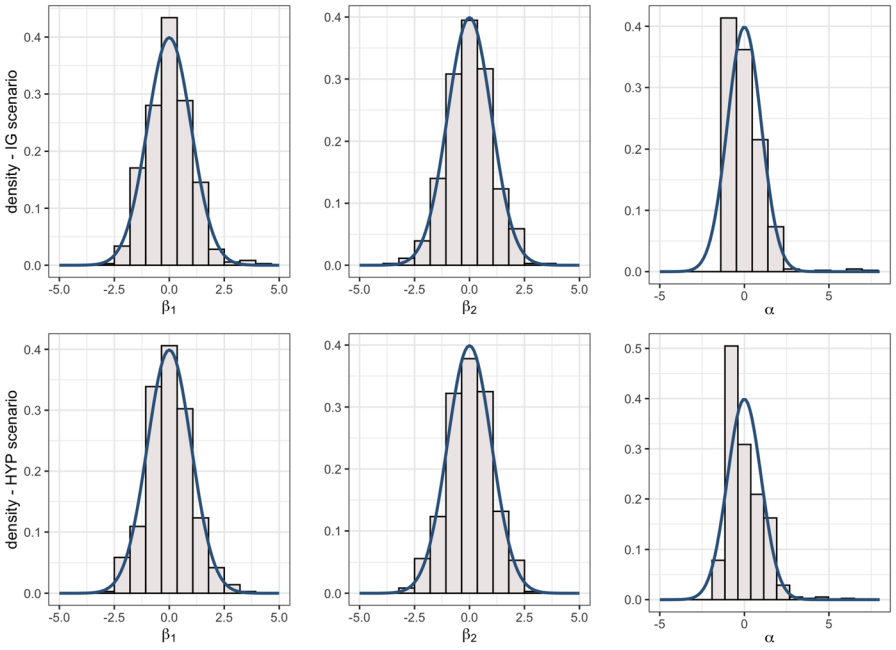

Fig. 3 Histogram of standardized parameter estimates under the IG (top row) and HYP (bottom row) scenarios with standard normal density curve

where we simply utilize those to obtain estimates of the standard deviation in the first case and take percentiles for the former.

5 TARGET neuroblastoma data analysis

To illustrate the proposed methodology in this paper, we explore clinical data of children diagnosed with neuroblastoma cancer collected by the *Therapeutically Applicable Research to Generate Effective Treatments* (TARGET) initiative. Neuroblastoma is a type of cancer that originates in primitive forms of the nerve cells of the sympathetic nervous system. We aim to evaluate the effect of MYCN gene amplification and the presence of hyperdiploid chromosomes in the DNA content of the tumor cell on the survival time of patients. Although the amplification of the MYCN gene is commonly associated with a worse prognosis, the role of hyperdiploid chromosomes in the evolution of tumors still remains an ongoing question in cancer research. The prognosis of the presence of an abnormal number of chromosomes is not well understood as different conclusions have been drawn with respect to different types of cancer.

For instance, Dastugue et al. (2013) associated higher ploidy to a better the prognosis for childhood B-acute lymphoblastic leukemia. In the same vein, Carroll (2013) added that hyperdiploidy seems to favor a beneficial impact in leukemogenesis, possible being a consequence rather than a driver of malignancy. However, Donovan et al. (2014) recognized glioblastoma hyperdiploid tumor cells as a potential contributor to tumor evolution and disease recurrence in adult brain cancer patients. Following this argument, one of our aims in this data application is to reliably assess the effect of hyperdiploid chromosomes on the prognosis of children with neuroblastoma cancer by the usage of appropriate statistical methodology.

5.1 Neuroblastoma data description and preliminary analysis

The TARGET program is intended to study the molecular changes causing childhood cancers. The main goal of this initiative is to collect data allowing researchers to work toward the development of effective and less toxic treatments. The program makes available online data files of multiple cancer types that can be found in the TARGET data matrix section of their Web site https://ocg.cancer.gov/programs/target/data-matrix; it is necessary to request permission from the National Cancer Institute to use this data, as we did. In this paper, we explore the clinical data of the neuroblastoma cancer containing patient survival data and multiple covariates. As described in their related homepage, the neuroblastoma cancer arises in immature nerve cells of the sympathetic nervous system, primarily affecting infants and children. It accounts for 12% of childhood cancer mortality, those between 18 months and 5 years of age being the most affected.

We work with the subset of patients classified as high risk by the Children's Oncology Group (COG), categorization that determines the type of treatment to be received. Of the total of 533 high-risk patients with a known vital status, we selected

315 observations that had full information recorded so to test the influence of multiple variables. Among these 315 individuals, 178 of them deceased until the end of the study and the others were right censored. In a preliminary analysis, we assessed the significance of several covariates in the survival time of the patients. Some of them are: The International Neuroblastoma Staging System (INSS stage) that is a clinically and surgically based staging system used to categorize tumor extent; the International Neuroblastoma Pathology Classification Mitosis Karyorrhexis Index Category (MKI); the patient's age at diagnosis; the patient's gender; the International Neuroblastoma Pathology Classification Tumor Cell Differentiation Degree Category, among others. At this stage, we selected covariates using the Cox model and the log-rank test. From this, we kept the variables MYCN gene amplification (categorized and amplified or not amplified) and Ploidy, that is DNA ploidy analysis by flow cytometry result category, described as diploid (DI=1) or hyperdiploid (DI>1), with no interaction among them.

An important and well-known tool to explore time to event data is the Kaplan–Meier estimate of the survival function, introduced in the seminal paper by Kaplan and Meier (1958). In Fig. 4, we report this graph for the two covariates selected. We observe that the distance between the curves does not remain proportional over time, indicating that the Cox model is not appropriate for this data application.

Our modeling strategy is to apply frailty models where the proportionality of hazards is not suitable. At the same time, such approach allows to account for individual heterogeneity not explained by the covariates. For instance, it is reasonable to expect that biological distinctions (e.g., genetic features) exist among the patients.

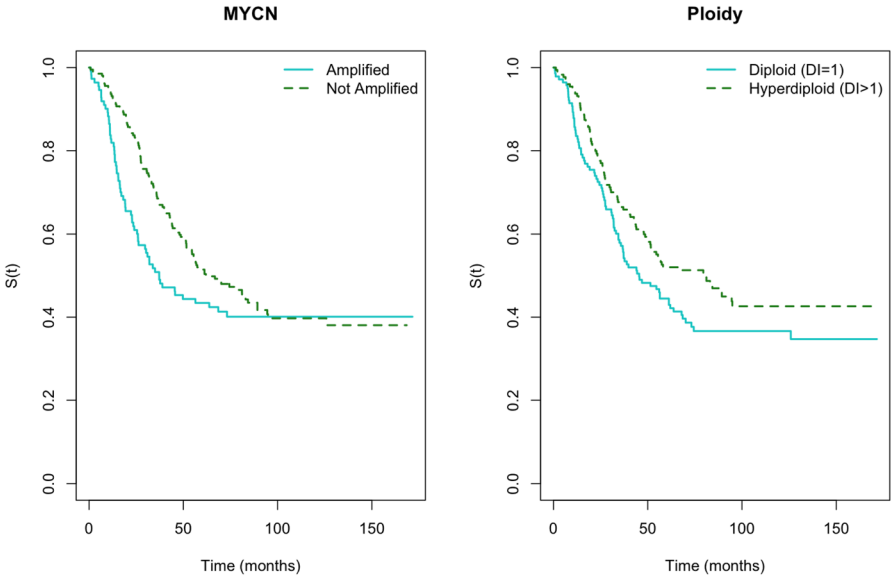

Fig. 4 Descriptive analysis of the TARGET neuroblastoma data set. Kaplan–Meier estimates of the survival function for the variables MYCN and Ploidy

In addition to the proposed GIG class, we employ well-established frailty models with implementations available in `R` program, as would most commonly be done in practice. We seek among the competitive models a robust estimation of the covariate effects and frailty variance. A reliable estimation of this quantities is crucial for our data application. The hazard ratios of MYCN and Ploidy will tell us about the influence of these variables in the patients' survival times, facilitating a prognosis assessment and contributing to the discussion about the role of hyperdiploid chromosomes in cancer cells. In addition, the frailty variance is also of interest as it informs how much heterogeneity to expect among the survival times of high risk neuroblastoma patients.

5.2 Statistical analysis of neuroblastoma data

In this section, we analyze the TARGET neuroblastoma data using the proposed class of GIG frailties and also well-established frailty models. Our goal is to compare how our models with those commonly used in practice. Our response variable is the time (in months) from diagnosis to the last follow-up or death of the patient.

We investigate the effects of two covariates on the survival time of the patients. They were selected after the preliminary analysis described in Sect. 5.1. The covariates are (1) MYCN Status, it corresponds to MYCN gene amplification status and is categorized as amplified or non-amplified; and (2) Ploidy (DNA ploidy analysis by flow cytometry result category), categorizes the DNA content of the tumor cell as diploid or hyperdiploid.

We would like to compare the estimation of the hazard ratios and frailty variance among the four special cases of the GIG class as well as the gamma, positive frailty variance (PVF), and log-normal frailties. By using univariate frailty models, we account for non-observed individual risk factors. We focus on semiparametric and PE baseline hazard implementations of the aforementioned models since no information about the parametric form of the baseline risk function is available. Besides, testing the adequacy of parametric forms of this function is not an easy task. We fit the semiparametric implementations of the gamma, PVF and log-normal frailty models available in the `R` packages `frailtyEM` (Balan and Putter 2019) and `frailtySurv` (Monaco et al. 2018).

According to Barreto-Souza and Mayrink (2019), the popular gamma frailty model can present problems in the estimation of the frailty variance depending on the form of the likelihood to be maximized. Trying to circumvent this, two different implementations of this model are investigated here. We fit the versions in `frailtyEM` and `frailtySurv` packages. In the first, a non-penalized log-likelihood function is maximized while in the former a pseudo full likelihood approach is employed.

We choose the number of cut points (named k) for the GIG class by fitting each special case for a range of k from 3 to 30. We notice that for the IG, HYP, RIG, and PHYP frailties, the smallest AIC (Akaike 1974) values were among k from 10 to 15, with small variation in this range. Hence, we choose $k = 10$ and present the results with this configuration.

Table 5 contains the results of fitting the aforementioned frailty models to the TARGET neuroblastoma data. In addition to the parameter estimates, we report the exponential of β_{mycn} and β_{ploidy}, i.e., the hazard ratio, which gives us the practical interpretation of the regression coefficients. See Wienke (2011) for a proper interpretation of these hazard ratios for frailty models.

It is immediate that the parameter estimates vary considerably among the fits of the different models. There is great variability in the estimation of the frailty variance, but the regression coefficients are also not immune to that. Furthermore, frailty variance-related parameter, indicated in the column α, displays huge standard errors under the gamma (`frailtySurv`) and log-normal models. We highlight that in the `frailtySurv` package standard errors are computed through bootstrap replication, as are the ones in the GIG class. To provide a fair comparison, the same number of replicas (500) was used to produce the results in both cases.

Having found a significant variability of results among different frailty models and its implementations, we seek to investigate the robustness of the estimates produced in each case. We propose to examine this through a bootstrap study where, in each replication, we sample 200 out of the 315 observations and fit the models under comparison. After a certain number of replications, we can assess the variation of the parameter estimates. More specifically, we examine how these vary around the maximum likelihood estimates gathered from the fit to the entire data set. Naturally, the less variability, more consistent the model is and more reliable are the estimates obtained from the complete data. This study is also motivated by the fact that model selection tools are scarce in the frailty models literature. It is specially hard in this case where the competing models employ different forms of the likelihood and distinct forms of the baseline risk functions.

The study is performed using 1000 bootstrap replicas. We report the results in Figs. 5 and 6 where each box corresponds to one of the models and the colors to the frailty distribution. Figure 5 contains the boxplots of the hazard ratio estimates for the covariates amplified MYCN and diploid ploidy. We center those around the values obtained from fit to the complete data set, that we refer as the "true" value.

In the left-side plot, we can see that there is a greater variability on the estimation of the hazard ratio of MYCN under the fits of the gamma and PVF frailties in

Table 5 Estimates of the parameters and standard errors (in parentheses) for the neuroblastoma data under the GIG, gamma, log-normal, and PVF models

Model	β_{mycn}	β_{ploidy}	Var	α	$e^{\beta_{mycn}}$	$e^{\beta_{ploidy}}$
Gamma (`frailtyEM`)	1.312 (0.366)	0.586 (0.341)	3.866	0.259 (0.100)	3.714	1.797
Gamma (`frailtySurv`)	1.203 (0.907)	0.568 (0.224)	3.117	3.117 (20.996)	3.329	1.765
PVF (`frailtyEM`)	1.429 (0.317)	0.275 (0.281)	2.021	0.495 (0.303)	4.177	1.317
LN (`frailtySurv`)	0.561 (0.417)	0.421 (0.213)	40.845	1.933 (16.307)	1.753	1.524
PE-IG ($k = 10$)	0.463 (0.278)	0.426 (0.250)	5.939	5.939 (0.919)	1.588	1.530
PE-RIG ($k = 10$)	0.430 (0.272)	0.347 (0.241)	0.948	1.467 (0.093)	1.537	1.415
PE-HYP ($k = 10$)	0.542 (0.309)	0.491 (0.292)	3.550	7.817 (0.732)	1.721	1.634
PE-PHYP ($k = 10$)	0.361 (0.249)	0.327 (0.212)	0.509	0.702 (0.029)	1.435	1.388

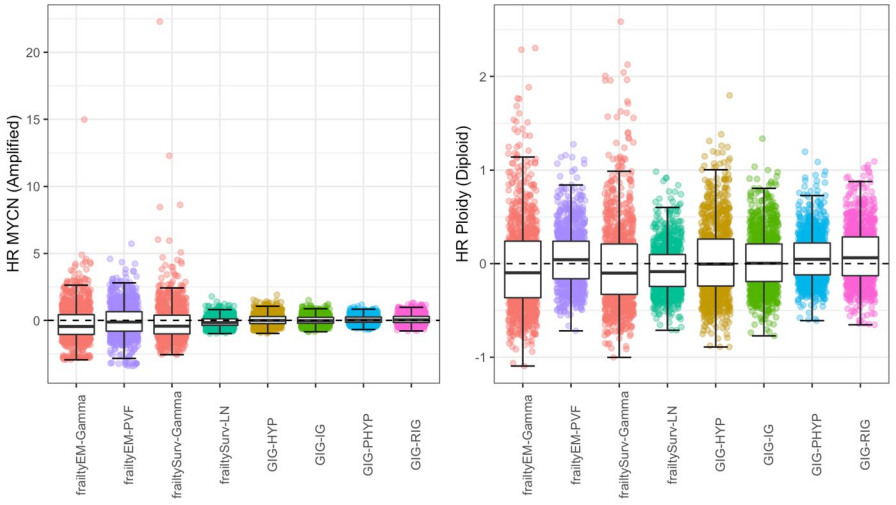

Fig. 5 Results of bootstrap study for centered hazard ratios of MYCN (left) and Ploidy (right)

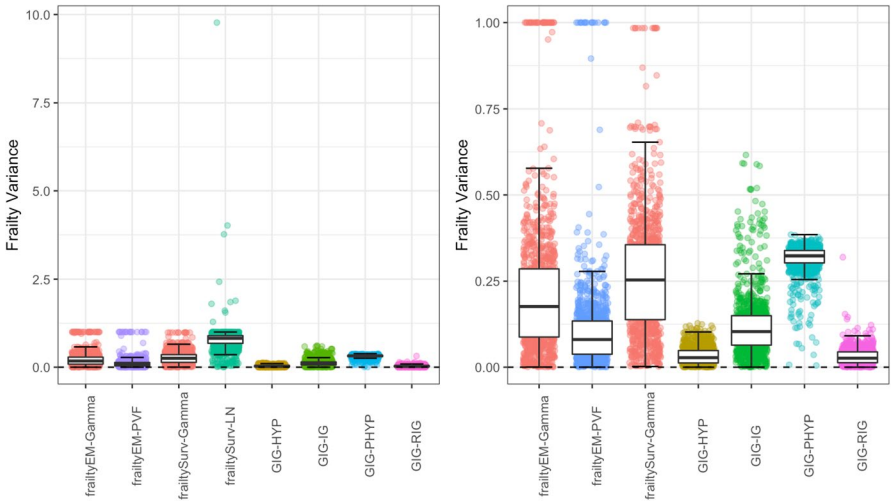

Fig. 6 Results of bootstrap study for centered and scaled frailty variances for all models to the left. Plot to the right present the same results by excluding the log-normal model

comparison to the competing models. Meanwhile, the log-normal and GIG frailties behave well, which is evidenced by their narrower boxes that are concentrated close to the "true" value. As it concerns Ploidy, the right-hand side plot shows that there is a less striking difference among the models considered. Although both gamma fits seem to produce more outliers, in general, all the models are reasonably robust when estimating the effect of the Ploidy variable.

Figure 6 contains the result of this study for the frailty variance. In this case, we center and scale the frailty variance estimates around the "true" values to avoid any interference of the different scales. On the left-hand side, all the models under comparison are included, where the log-normal frailty clearly does not perform like the other ones. It displays high outlier values and is located further from zero than the other models. We filter out the log-normal boxplot from the right-side plot to compare among the other cases. Now, we identify the gamma fits to be less consistent and can point out advantages toward the HYP and RIG models whose boxplots are the narrowest and concentrated more closely to zero.

We conclude that the HYP and RIG special cases of the GIG family were the most robust models for estimating the three quantities of interest in the application to the TARGET neuroblastoma data set. They return consistent estimates of both the hazard ratios and frailty variance, something that is not achieved by the competing models. Having shown advantages toward the GIG class in this data application, we propose to make the model selection among a broader range of GIG models by considering several values of λ. Hence, we present the selection of λ based on a grid of values. The grid of λ is set between -5 and 5 with 0.1 spacing. We calculate the value of the maximized log-likelihood function (profile likelihood) obtained under each different λ.

The result is reported in Fig. 7 where it is evident that the value $\lambda = 0$ achieves the highest log-likelihood value. The irregularity of the log-likelihood function on the right is most likely due to the numerical instability of the Bessel function, but this does not compromise our choice for $\lambda = 0$ since the function is well behaved around this point. This result goes in accordance with the conclusions drawn from our bootstrap simulation study, where we pointed out the HYP($\lambda = 0$) or RIG($\lambda = 0.5$) cases to be preferable. Furthermore, by combining these results, we can conclude that the

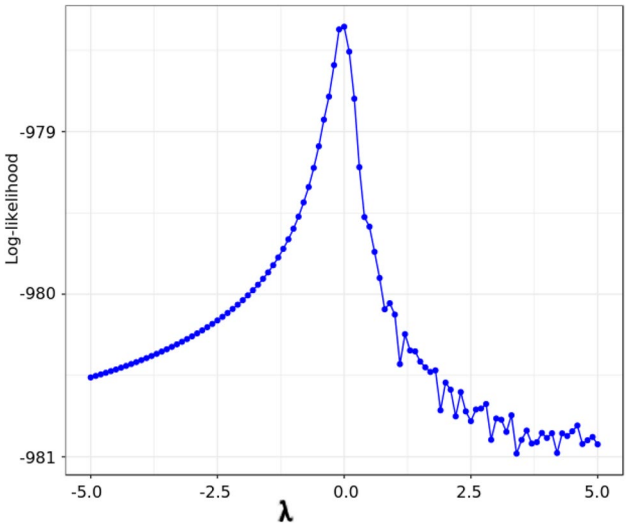

Fig. 7 Profile log-likelihood function of the parameter λ for the neuroblastoma data

most suitable model to the data in terms of model selection is also a robust choice, as desired.

To summarize, our main goal in the application of frailty models to the TARGET neuroblastoma data set was to conclude about the effects of the two genetic covariates and patient heterogeneity. We fitted different frailty models where individual heterogeneity not explained by the covariates was captured by a random effect. A considerable variability in the results when using different frailty distributions was highlighted. We proposed to tackle this by checking the robustness of the estimates through a bootstrap study where we fit each model several times to samples of the original data set. By doing this, it is evidenced that two special cases of the GIG class produce consistent estimates with advantages over the others. Further, we use a profile-likelihood approach to select the most suitable λ value. In this analysis, we used an extensive grid of λ values to take into consideration a range of GIG models and not only its special cases. The result indicates $\lambda = 0$ to be preferred, which corresponds to the HYP frailty. At the end, we can show that the GIG frailty chosen through model selection via profile likelihood is also robust, as ideal.

Hence, we draw our conclusions using parameter estimates of the HYP frailty model. We have that the failure rate of patients with amplified MYCN is 1.721 times that of patients with the unamplified status, conditional on the same frailty. Also, the failure rate of patients with diploid chromosomes is 1.634 times that of patients with hyperdiploid chromosomes, given the frailty value. Further, there is large effect of the patient individual heterogeneity, reflected by the frailty variance of 3.550. The statistical result on the effect of an amplified MYCN gene agrees with what is mentioned in Yoshimoto et al. (1999). According to the authors, the MYCN amplification has proven to be an independent prognostic factor for identifying rapid tumor progression and predicting poor prognosis independent of age and clinical stage. As for the presence of hyperdiploid chromosomes in the DNA content of tumor cells, this is a factor in which there is no consensus on its prognostic influence and seems to differ between different types of cancer. Through the analysis of the TARGET neuroblastoma data, we can see that there is evidence that the presence of diploid cells is a protective factor in the lifetime of children with high-risk neuroblastoma cancer, whose biological reasons have yet to be investigated.

6 Discussion

In this paper, we proposed a flexible class of frailty models based on the generalized inverse-Gaussian (GIG) distributions, which enjoy mathematical tractability like the gamma frailty model. Simulation studies were carried to investigate the best methodologies to obtain parameter estimates, standard errors, and confidence intervals for the parameters of the proposed class.

The proposed model proved to be robust under misspecification of the frailty distribution as it was able to yield satisfactory estimates of the covariate effects in all scenarios. Such results were presented in the Supplementary Material. Under gamma and generalized exponential artificial data, there was a particular case of our class that performed better than the correct specified model. It was also possible

to obtain a particular case that performed properly when the true frailty was log-normal distributed; this is the scenario in which the competing models returned high biased estimates. In addition, we explored the fit of the GIG special cases under misspecification of λ. Our findings indicate that, although it affects the estimation of the frailty variance, misspecifying λ does not influence largely the estimation of the fixed effects.

We highlight that, as expected, there is no single model that is appropriate in all situations. The frailties obtained by fitting the special cases of the GIG distribution showed different behaviors in the simulation studies. This evidences that they can capture distinct dependence structures providing the desired flexibility. We emphasize the advantage of having a robust and flexible model such as the GIG frailty in hand because identifying the true frailty distribution is not a simple task in practical problems and using selection and diagnostic methods are still scarce in this field.

A complete statistical analysis was conducted comprising the most commonly used frailty models and the class of GIG frailties with the goal of producing reliable conclusions about the effects of the amplification of the MYCN gene and the presence of hyperdiploid chromosomes on the survival times of children with neuroblastoma cancer. After identifying that the fitted models yielded considerably different results, we compared the robustness of the parameter estimates among them. This was done via a bootstrap study which showed the HYP and RIG special cases to be the most consistent. A profile likelihood approach was used to select λ, indicating the choice of the HYP frailty. Having agreement between the results from the robustness study and the profile likelihood approach, we use the fitted HYP frailty model to draw conclusions about the covariate effects and the population heterogeneity. We highlight that this real data application illustrates how our class can be advantageous in practical situations.

Future research includes applying GIG frailty models in other contexts such as current status data. Another point of interest to be attacked is to propose a time-varying GIG frailty model. One possibility of fulfilling this task is to assume a time-varying frailty given by $Z(t) \equiv w(t, Z)$, where $w(t, Z) = Z^{h_0(t)}$, for $t > 0$, Z follows a GIG distribution as described in this paper and $h_0(\cdot)$ denotes the baseline hazard function. This structure was introduced by Enki et al. (2014), with Z following a generalized gamma distribution. As argued by these authors, time-varying frailty models can be applied to study infectious diseases. As suggested by one referee, other possible points to be explored are the proposal of GIG frailty models for dealing with clustered dependent censored data and clustered semi-competing risks data as alternatives to the works by Schneider et al. (2019) and Emura et al. (2017) (see also Peng et al. (2018)), respectively. We feel that our proposed models and associated EM-algorithm can be adapted for these cases. Regarding the approach over the baseline hazard function, one alternative strategy suggested by one referee is to consider $H_0(\cdot)$ being a non-decreasing step function with jumps only at the observed event times and then use counting process notation to rewrite the complete log-likelihood function like done by Zeng and Lin (2006) and Zeng and Lin (2007). Hence, the Q-function can be obtained, and we do not need to deal with the knots selection problem since the number of estimates η_s is the number of observed event times. We hope to explore some of these points in future works.

Acknowledgements We thank the Associate Editor and two anonymous Referees for their insightful comments and suggestions that lead to a great improvement of the paper. We also thank the National Cancer Institute (Office of Cancer Genomics) for granting us permission to use the TARGET Neuroblastoma Clinical data for publication. W. Barreto-Souza would also like to acknowledge support for his research from the KAUST Research Fund, NIH 1R01EB028753-01, and the *Conselho Nacional de Desenvolvimento Científico e Tecnológico* (CNPq-Brazil, Grant number 305543/2018-0). Part of this work is from the Master's Thesis of Luiza S.C. Piancastelli realized at the Department of Statistics of the Universidade Federal de Minas Gerais, Brazil.

Appendix

A.1 Observed information matrix

Define $c_i(\theta) = \sum_{j=1}^{n_i} H_0(t_{ij}) e^{x_{ij}^\top \beta}$ and $a_{irs}(\theta) = \sum_{j=1}^{n_i} H_0(t_{ij}) e^{x_{ij}^\top \beta} x_{ijr} x_{ijs}$, for $r, s = 1, \ldots, p$, $\Delta_i \equiv \Delta_i(\theta) = \left\{ \alpha^{-1} \left(\alpha^{-1} + 2 \sum_{j=1}^{n_i} H_0(t_{ij}) e^{x_{ij}^\top \beta} \right) \right\}^{1/2}$, and $\lambda_i^* = \lambda + \sum_{j=1}^{n_i} \delta_{ij}$, for $i = 1, \ldots, m$. We have that the elements of the observed information matrix $J_n(\theta_*) = -\partial^2 \ell(\theta) / \partial \theta_* \partial \theta_*^\top$ are given by

$$\begin{aligned}
\frac{\partial^2 \ell(\theta)}{\partial \beta_r \beta_s} =& \frac{1}{\alpha^2} \sum_{i=1}^{m} \frac{K''_{\lambda_i^*}(\Delta_i) K_{\lambda_i^*}(\Delta_i) \Delta_i - K'_{\lambda_i^*}(\Delta_i) [K'_{\lambda_i^*}(\Delta_i) \Delta_i + K_{\lambda_i^*}(\Delta_i)]}{K_{\lambda_i^*}(\Delta_i)^2 \Delta_i^3} a_{ir}(\theta) a_{is}(\theta) \\
&+ \frac{1}{\alpha} \sum_{i=1}^{m} \frac{K'_{\lambda_i^*}(\Delta_i)}{K_{\lambda_i^*}(\Delta_i) \Delta_i} a_{irs}(\theta) - \sum_{i=1}^{m} \frac{\lambda_i^* a_{irs}(\theta)}{\alpha^{-1} + 2 c_i(\theta)} + 2 \sum_{i=1}^{m} \frac{\lambda_i^* a_{ir}(\theta) a_{is}(\theta)}{(\alpha^{-1} + 2 c_i(\theta))^2}, \\
\frac{\partial^2 \ell(\theta)}{\partial \beta_r \partial \alpha} =& -\frac{1}{\alpha^2} \sum_{i=1}^{m} \frac{K'_{\lambda_i^*}(\Delta_i)}{K_{\lambda_i^*}(\Delta_i) \Delta_i} a_{ir}(\theta) - \frac{1}{\alpha^2} \sum_{i=1}^{m} \frac{\lambda_i^* a_{ir}(\theta)}{(\alpha^{-1} + 2 c_i(\theta))^2} \\
&- \frac{1}{\alpha^3} \sum_{i=1}^{m} \frac{K''_{\lambda_i^*}(\Delta_i) K_{\lambda_i^*}(\Delta_i) \Delta_i - K'_{\lambda_i^*}(\Delta_i) [K'_{\lambda_i^*}(\Delta_i) \Delta_i + K_{\lambda_i^*}(\Delta_i)]}{K_{\lambda_i^*}(\Delta_i)^2 \Delta_i^3} \\
&(\alpha^{-1} + c_i(\theta)) a_{ir}(\theta),
\end{aligned}$$

for $r, s = 1, \ldots, p$, and

$$\begin{aligned}\frac{\partial^2 \ell(\theta)}{\partial \alpha^2} =& \frac{1}{2\alpha^2}\sum_{i=1}^{m}\lambda_i^* - \frac{2m}{\alpha^3}\frac{K'_\lambda(\alpha^{-1})}{K_\lambda(\alpha^{-1})} - \frac{m}{\alpha^4}\frac{K''_\lambda(\alpha^{-1})K_\lambda(\alpha^{-1}) - K'_\lambda(\alpha^{-1})^2}{K_\lambda(\alpha^{-1})^2} \\ &+ \frac{2}{\alpha^3}\sum_{i=1}^{m}\frac{K'_{\lambda_i^*}(\Delta_i)}{K_{\lambda_i^*}(\Delta_i)\Delta_i}(\alpha^{-1}+c_i(\theta)) + \frac{1}{\alpha^4}\sum_{i=1}^{m}\frac{K'_{\lambda_i^*}(\Delta_i)}{K_{\lambda_i^*}(\Delta_i)\Delta_i} \\ &- \frac{1}{\alpha^3}\sum_{i=1}^{m}\frac{\lambda_i^*}{\alpha^{-1}+2c_i(\theta)} + \frac{1}{2\alpha^4}\sum_{i=1}^{m}\frac{\lambda_i^*}{(\alpha^{-1}+2c_i(\theta))^2} \\ &+ \frac{1}{\alpha^4}\sum_{i=1}^{m}\frac{K''_{\lambda_i^*}(\Delta_i)K_{\lambda_i^*}(\Delta_i)\Delta_i - K'_{\lambda_i^*}(\Delta_i)[K'_{\lambda_i^*}(\Delta_i)\Delta_i + K_{\lambda_i^*}(\Delta_i)]}{K_{\lambda_i^*}(\Delta_i)^2\Delta_i^3}(\alpha^{-1}+c_i(\theta))^2.\end{aligned}$$

A.2 Louis information matrix

From Louis (1982), we have that the information matrix obtained from the EM-algorithm, say $\mathbf{I}_n(\theta_*)$, is given by

$$I_n(\theta_*) = E\left(-\frac{\partial^2 \ell_c(\theta)}{\partial\theta_*\partial\theta_*^\top}\Big|Y^{obs}\right) - E\left(\frac{\partial \ell_c(\theta)}{\partial\theta_*}\frac{\partial \ell_c(\theta)}{\partial\theta_*}^\top\Big|Y^{obs}\right), \tag{10}$$

where we have defined $Y^{obs} = \{(t_{ij}, \delta_{ij}), j = 1, \ldots, n_i, i = 1, \ldots, m\}$.

Let $\tau_i(\theta) = E(Z_i^2|Y^{obs})$ and $\nu_i(\theta) = E(Z_i^{-2}|Y^{obs})$ for $i = 1, \ldots, m$, where explicit expressions are directly available by using (3) and (9). The elements of the information matrix (10) are given by

$$E\left(-\frac{\partial^2 \ell_c(\theta)}{\partial\beta_r\partial\beta_s}\middle|Y^{obs}\right)=\sum_{i=1}^{m}\sum_{j=1}^{n_i}\omega_i(\theta)H_0(t_{ij})e^{x_{ij}^\top\beta}x_{ijr}x_{ijs},\quad r,s=1,\ldots,p,$$

$$\begin{aligned}E\left(-\frac{\partial^2 \ell_c(\theta)}{\partial\alpha^2}\middle|Y^{obs}\right)=&\frac{m}{\alpha^4}\frac{K''_\lambda(\alpha^{-1})K_\lambda(\alpha^{-1})-K'_\lambda(\alpha^{-1})^2}{K_\lambda(\alpha^{-1})^2}\\&+\frac{2m}{\alpha^3}\frac{K'_\lambda(\alpha^{-1})}{K_\lambda(\alpha^{-1})}+\frac{1}{\alpha^3}\sum_{i=1}^{m}\big(\omega_i(\theta)+\kappa_i(\theta)\big),\end{aligned}$$

$$\begin{aligned}E\left(\frac{\partial \ell_c(\theta)}{\partial\beta_r}\frac{\partial \ell_c(\theta)}{\partial\beta_s}\middle|Y^{obs}\right)=&\left(\sum_{i=1}^{m}\sum_{j=1}^{n_i}\delta_{ij}x_{ijr}\right)\left(\sum_{i=1}^{m}\sum_{j=1}^{n_i}\delta_{ij}x_{ijs}\right)\\&-\left(\sum_{i=1}^{m}\sum_{j=1}^{n_i}\delta_{ij}x_{ijs}\right)\left(\sum_{i=1}^{m}\omega_i(\theta)a_{ir}(\theta)\right)\\&-\left(\sum_{i=1}^{m}\sum_{j=1}^{n_i}\delta_{ij}x_{ijr}\right)\left(\sum_{i=1}^{m}\omega_i(\theta)a_{is}(\theta)\right)+\sum_{i=1}^{m}\tau_i(\theta)a_{ir}(\theta)a_{is}(\theta)\\&+\sum_{i\neq i'}^{m}\omega_i(\theta)\omega_{i'}(\theta)a_{ir}(\theta)a_{i's}(\theta),\quad r,s=1,\ldots,p,\end{aligned}$$

$$\begin{aligned}E\left(\left(\frac{\partial \ell_c(\theta)}{\partial\alpha}\right)^2\middle|Y^{obs}\right)=&\frac{m^2}{\alpha^4}\left(\frac{K'_\lambda(\alpha^{-1})}{K_\lambda(\alpha^{-1})}\right)^2+\frac{m}{\alpha^4}\frac{K'_\lambda(\alpha^{-1})}{K_\lambda(\alpha^{-1})}\sum_{i=1}^{m}\big(\omega_i(\theta)+\kappa_i(\theta)\big)\\&+\frac{1}{4\alpha^4}\sum_{i=1}^{m}\big(\tau_i(\theta)+2+\nu_i(\theta)\big)\\&+\frac{1}{4\alpha^4}\sum_{i\neq i'}^{m}\big(\omega_i(\theta)+\kappa_i(\theta)\big)\big(\omega_{i'}(\theta)+\kappa_{i'}(\theta)\big),\end{aligned}$$

$$\begin{aligned}E\left(\frac{\partial \ell_c(\theta)}{\partial\beta_r}\frac{\partial \ell_c(\theta)}{\partial\alpha}\middle|Y^{obs}\right)=&\frac{m}{\alpha^2}\frac{K'_\lambda(\alpha^{-1})}{K_\lambda(\alpha^{-1})}\left(\sum_{i=1}^{m}\sum_{j=1}^{n_i}\delta_{ij}x_{ijr}-\sum_{i=1}^{m}\omega_i(\theta)a_{ir}(\theta)\right)\\&+\frac{1}{2\alpha^2}\left(\sum_{i=1}^{m}\sum_{j=1}^{n_i}\delta_{ij}x_{ijr}\right)\\&\times\left(\sum_{i=1}^{m}\big(\omega_i(\theta)+\kappa_i(\theta)\big)\right)-\frac{1}{2\alpha^2}\sum_{i=1}^{m}\big(\tau_i(\theta)+1\big)a_{ir}(\theta)\\&-\frac{1}{2\alpha^2}\sum_{i\neq i'}^{m}\big(\omega_i(\theta)+\kappa_i(\theta)\big)\omega_{i'}(\theta)a_{i'r}(\theta),\ r=1,\ldots,p,\end{aligned}$$

and $E\left(-\frac{\partial^2 \ell_c(\theta)}{\partial\beta_r\partial\alpha}\middle|Y^{obs}\right)=0$, for $r=1,\ldots,p$, where we have defined $a_{ir}(\theta)=\sum_{j=1}^{n_i}H_0(t_{ij})e^{x_{ij}^\top\beta}x_{ijr}$, $b_{is}(\theta)=\sum_{j=1}^{n_i}H_0^{(s)}(t_{ij})e^{x_{ij}^\top\beta}$, for $i=1,\ldots,m$, $r=1,\ldots,p$, and $s=1,\ldots,k+1$.

References

Abrahantes, J. C., Burzykowski, T. (2005). A version of the EM algorithm for proportional hazard model with random effects. *Biometrical Journal*, *47*, 847–862.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*, 716–723.

Balakrishnan, N., Pal, S. (2016). Expectation maximization-based likelihood inference for flexible cure rate models with Weibull lifetimes. *Statistical Methods in Medical Research*, *25*, 1535–1563.

Balakrishnan, N., Peng, Y. (2006). Generalized gamma frailty model. *Statistics in Medicine*, *25*, 2797–2816.

Balan, T., Putter, H. (2019). frailtyEM: An R package for estimating semiparametric shared frailty models. *Journal of Statistical Software*, *90*, 1–29.

Barndorff-Nielsen, O. E., Halgreen, C. (1977). Infinite divisibility of the Hyperbolic and generalized inverse Gaussian distribution. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, *38*, 309–312.

Barreto-Souza, W., Mayrink, V. D. (2019). Semiparametric generalized exponential frailty model for clustered survival data. *Annals of the Institute of Statistical Mathematics*, *71*, 679–701.

Callegaro, A., Iacobelli, S. (2012). The Cox shared frailty model with log-skew-normal frailties. *Statistical Modelling*, *12*, 399–418.

Carroll, W. L. (2013). Safety in numbers: Hyperdiploidy and prognosis. *Blood*, *121*, 2374–237.

Chen, P., Zhang, J., Zhang, R. (2013). Estimation of the accelerated failure time frailty model under generalized gamma frailty. *Computational Statistics and Data Analysis*, *62*, 171–180.

Clayton, D. (1978). A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika*, *65*, 141–151.

Cox, D. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society - Series B*, *34*, 187–220.

Crowder, M. (1989). A multivariate distribution with Weibull connections. *Journal of the Royal Statistical Society - Series B*, *51*, 93–107.

Dastugue, N., Suciu, S., Plat, G., Speleman, F., Cave, H., Girard, S., Bakkus, M., Pages, M. P., Yakouben, K., Nelken, B., Uyttebroeck, A., Gervais, C., Lutz, P., Teixeira, M. R., Heimann, P., Ferster, A., Rohrlich, P., Collonge, M. A., Munzer, M., Luquet, I., Boutard, P., Sirvent, N., Karrasch, M., Bertrand, Y., Benoit, Y. (2013). Hyperdiploidy with 58–66 chromosomes in childhood B-acute lymphoblastic leukemia is highly curable: 58951 CLG-EORTC results. *Blood*, *121*, 2415–2423.

Dempster, A. P., Laird, N. M., Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm (with discussion). *Journal of the Royal Statistical Society - Series B*, *39*, 1–38.

Donovan, P., Cato, K., Legaie, R., Jayalath, R., Olsson, G., Hall, B., Olson, S., Boros, S., Reynolds, B., Harding, A. (2014). Hyperdiploid tumor cells increase phenotypic heterogeneity within Glioblastoma tumors. *Molecular bioSystems*, *10*, 741–758.

Duchateau, L., Janssen, P. (2008). *The Frailty Model*. New York: Springer.

Duchateau, L., Janssen, P., Lindsey, P., Legrand, C., Nguti, R., Silvester, R. (2002). The shared frailty model and the power for heterogeneity tests in multicenter trials. *Computational Statistics and Data Analysis*, *40*(3), 603–620.

Efron, B. (1979). Bootstrap methods: Another look at the Jackknife. *The Annals of Statistics*, *7*, 1–26.

Emura, T., Nakatochi, M., Murotani, K., Rondeau, V. (2017). A joint frailty-copula model between tumour progression and death for meta-analysis. *Statistical Methods in Medical Research*, *26*(6), 2649–2666.

Emura, T., Matsui, S., Rondeau, V. (2019). *Survival Analysis with Correlated Endpoints*: *Joint Frailty-Copula Models. JSS Research Series in Statistics*. Singapore: Springer.

Enki, D. G., Noufaily, A., Farrington, P. (2014). A time-varying shared frailty model with application to infectious diseases. *Annals of Applied Statistics*, *8*, 430–447.

Farrington, C., Unkel, S., Anaya-Izquierdo, K. (2012). The relative frailty variance and shared frailty models. *Journal of the Royal Statistical Society - Series B*, *74*, 673–696.

Fletcher, R. (2000). *Practical Methods of Optimization*, 2nd ed., New York: Wiley.

Hanagal, D. D. (2019). *Modeling Survival Data Using Frailty Models*. Singapore: Springer.

Hirsch, K., Wienke, A. (2012). Software for semiparametric shared gamma and log-normal frailty models: an overview. *Computer Methods and Programs in Biomedicine*, *107*(3), 582–597.

Hougaard, P. (1984). Life table methods for heterogeneous populations: Distributions describing the heterogeneity. *Biometrika*, *71*, 75–83.

Hougaard, P. (1986). A class of multivariate failure time distributions. *Biometrika*, *73*, 671–678.

Hougaard, P. (2000). *Analysis of Multivariate Survival Data*. New York: Springer.

Hougaard, P., Harvald, B., Holm, N. V. (1992). Measuring the similarities between the lifetimes of adult danish twins born between 1881–1930. *Biometrika*, *87*, 17–24.
Kaplan, E., Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, *53*, 457–481.
Kim, J. S., Proschan, F. (1991). Piecewise exponential estimation of the survival function. *IEEE Translations on Reliability*, *40*, 134–139.
Klein, J. P. (1992). Semiparametric estimation of random effects using the Cox model based on the EM algorithm. *Biometrics*, *48*, 795–806.
Lawless, J. F., Zhan, M. (1998). Analysis of interval-grouped recurrent event data using piecewise constant rate function. *Canadian Journal of Statistics*, *26*, 549–565.
Leão, J., Leiva, V., Saulo, H., Tomazella, V. (2017). Birnbaum-Saunders frailty regression models: Diagnostics and application to medical data. *Biometrical Journal*, *59*, 291–314.
Louis, T. A. (1982). Finding the observed information matrix when using the EM algorithm. *Journal of the Royal Statistical Society - Series B*, *44*, 226–233.
McGilchrist, C. A., Aisbett, C. W. (1991). Regression with frailty in survival analysis. *Biometrics*, *47*, 461–466.
Monaco, V., Gorfine, M., Hsu, L. (2018). General semiparametric shared frailty model estimation and simulation with frailtySurv. *Journal of Statistical Software*, *86*, 1–42.
Oakes, D. (1982). A model for association in bivariate survival data. *Journal of the Royal Statistical Society - Series B*, *44*, 414–422.
Oakes, D. (1986). Semiparametric inference in a model for association in bivariate survival data. *Biometrika*, *73*, 353–361.
Peng, M., Xiang, L., Wang, S. (2018). Semiparametric regression analysis of clustered survival data with semi-competing risks. *Computational Statistics and Data Analysis*, *124*, 53–70.
Putter, H., van Houwelingen, H. C. (2015). Dynamic frailty models based on compound birth-death processes. *Biostatistics*, *16*, 550–564.
R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/. Accessed Aug 2020.
Schneider, S., Demarqui, F. N., Colosimo, E. A., Mayrink, V. D. (2019). An approach to model clustered survival data with dependent censoring. *Biometrical Journal*, *62*(1), 157–174.
Therneau, T. (2015). A package for survival analysis in S. R package version 2.43.3. https://CRAN.R-project.org/package=survival. Accessed Aug 2019.
Therneau, T. M., Grambsch, P. M., Pankratz, V. S. (2003). Penalized survival models. *Journal of Computational and Graphical Statistics*, *12*, 156–175.
Vaupel, J. W., Manton, K. G., Stallard, E. (1979). The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography*, *16*, 439–454.
Vu, H. T. V., Knuiman, M. W. (2002). A hybrid ML-EM algorithm for calculation of maximum likelihood estimates in semiparametric shared frailty models. *Computational Statistics and Data Analysis*, *40*(1), 173–187.
Wang, H., Klein, J. P. (2012). Semiparametric estimation for the additive inverse Gaussian frailty model. *Communications in Statistics - Theory and Methods*, *41*, 2269–2278.
Wienke, A. (2011). *Frailty models in survival analysis*. New York: Chapman and Hall/CRC.
Xiao, Y., Abrahamowicz, M. (2010). Bootstrap-based methods for estimating standard error Cox's regression analyses of clustered event times. *Statistics in Medicine*, *29*, 915–923.
Yashin, A., Vaupel, J. W., Iachine, I. (1995). Correlated individual frailty: an advantageous approach to survival analysis of bivariate data. *Mathematical Population Studies*, *5*, 145–159.
Yoshimoto, M., Toledo, S., Caran, E., Seixas, M., Lee, M., Abib, S., Vianna, S., Schettini, S., Andrade, J. (1999). MYCN gene amplification: Identification of cell populations containing double minutes and homogeneously staining regions in neuroblastoma tumors. *The American Journal of Pathology*, *155*, 1439–1443.
Zeng, D., Lin, D. Y. (2006). Efficient estimation of semiparametric transformation models for counting processes. *Biometrika*, *93*, 627–640.
Zeng, D., Lin, D. Y. (2007). Maximum likelihood estimation in semiparametric regression models with censored data. *Journal of the Royal Statistical Society - Series B*, *69*, 507–564.

Affiliations

Luiza S. C. Piancastelli[1,3] · Wagner Barreto-Souza[2,3] · Vinícius D. Mayrink[3]

Luiza S. C. Piancastelli
luiza.piancastelli@ucdconnect.ie

Vinícius D. Mayrink
vdm@est.ufmg.br

[1] School of Mathematics and Statistics, University College Dublin, D04 V1W8 Belfield, Dublin 4, Ireland

[2] Statistics Program, King Abdullah University of Science and Technology, Thuwal 23955-6900, Saudi Arabia

[3] Departamento de Estatística, Universidade Federal de Minas Gerais, Belo Horizonte 31270-901, Brazil