A Bayesian semiparametric partially PH model for clustered time-to-event data

Bernardo Nipoti
School of Computer Science and Statistics, Trinity College, Dublin, Ireland
Alejandro Jara
Department of Statistics, Pontificia Universidad Católica de Chile, Santiago, Chile
Michele Guindani
Department of Statistics, The University of California, Irvine, CA, USA

Abstract

A standard approach for dealing with unobserved heterogeneity and clustered time-to-event data within the proportional hazards (PH) context has been the introduction of a cluster-specific random effect (frailty) that is common to subjects within the same cluster. PH models with shared random effects have been widely employed because they provide useful summary information in the absence of estimates of a baseline survival distribution and may be formulated in a semiparametric fashion. However, the conditional PH assumption could be too strong for some applications. For example, the marginal association of survival functions within a cluster does not depend on the subject-specific covariates. We propose an alternative partially PH modelling approach based on the introduction of cluster-dependent random hazard functions and on the use of mixture models induced by completely random measures. We show that the proposed approach accommodates for different degrees of association within a cluster, which varies as a function of cluster level and individual covariates. We also show that a particular specification of the proposed model has the appealing property of preserving marginally the PH structure. We illustrate the performances of the proposed modelling approach on simulated and real datasets.

Keywords: Completely random measures; Partially proportional hazard model; Frailty model; Hazard rate; Kendall's τ; Survival ratio.

1 Introduction

Cox's proportional hazards (PH) model (Cox, 1972) has been widely used in the analysis of time-to-event data. In the standard PH regression model, the hazard function for an individual with covariate vector $\mathbf{z}_i$, $h(\cdot \mid \mathbf{z}_i)$, is given by $h(\cdot \mid \mathbf{z}_i) = h(\cdot)\exp\{\mathbf{z}_i'\boldsymbol{\beta}\}$, where $\boldsymbol{\beta}$ is a vector of regression coefficients and $h(\cdot)$ is a baseline hazard function, corresponding to $\mathbf{z}_i = \mathbf{0}$. The PH assumption has several advantages: (i) the effect of a factor can be easily summarized by the relative risk, and (ii) mathematical simplicity, which has been exploited to produce a simple semiparametric approach via the partial likelihood. However, in the standard version of the model, the responses are assumed to be independent, which is not appropriate for data containing heterogeneous groups of subjects.

A common approach to take into account this heterogeneity is to extend the PH model by including a group-specific random effect term in the hazard function, usually called frailty. Under this model the conditional hazard function for the ith individual in the jth cluster, with covariate vector $\mathbf{z}_{i,j}$, is given by $h_{i,j}(\cdot) = h_j(\cdot)\exp\{\mathbf{z}_{i,j}'\boldsymbol{\beta}\}$, where the group-specific baseline hazard is given by $h_j(\cdot) = V_j h_0(\cdot)$, with V_j being a positive random variable and $h_0(\cdot)$ a common baseline hazard function (see, e.g. Hougaard, 2000). To complete the model specification, the random effect components V_j's are often assumed to be independent and identically distributed (iid). The resulting shared frailty PH model provides a convenient tool for modelling unobserved heterogeneity and for taking into account the correlation of the data since, conditionally on the random effect term, the interpretation of the hazard rates and other relevant quantities is similar to that of a standard PH model (see, e.g., Clayton, 1978, Hougaard, 2000). However, the random effect PH model presents some potentially important drawbacks too. For instance, the choice of the baseline hazard and the random effect's distribution can have a crucial impact on the inference (see, e.g., Hougaard, 2000, Chapter 7). In addition, for some applications the simple marginal association structure implied by the model may not be appropriate, and the conditional PH assumption could be too strong (see, e.g., De Iorio et al., 2009).

There is a rich Bayesian nonparametric (BNP) literature robustifying the frailty distribution (see, e.g., Müller et al., 2015), the baseline hazard function (Kalbfleisch, 1978, Hjort, 1990, Ibrahim et al., 2001, Sinha and Dey, 1997, Gelfand and Mallick, 1995, Carlin and Hodges, 1999), and relaxing the iid nature of the cluster-specific frailty terms to allow for more general covariate-dependent marginal association structures (see, e.g., Pennell and Dunson, 2006, Zhou et al., 2015). BNP approaches for baseline hazards include gamma processes (Kalbfleisch, 1978),

beta processes (Hjort, 1990), piecewise exponential priors (Ibrahim et al., 2001), correlated increments priors (Sinha and Dey, 1997), Bernstein polynomials (Gelfand and Mallick, 1995, Carlin and Hodges, 1999), and penalized B-splines (Kneib and Fahrmeir, 2007). Pennell and Dunson (2006) proposed a Bayesian semiparametric approach that allows for subject-specific frailties to change dynamically with time and pice-wise constant time dependent regression coefficients in the analysis of recurrent time-to-event data. Zhou et al. (2015) proposed a covariate-adjusted frailty PH model, where the frailty is modelled through a linear dependent tailfree process so that its complete distributional shape can change with both continuous and categorical cluster-level covariates. The approach proposed by Zhou et al. (2015) still assumes a conditional PH assumption within and across clusters, and the approach proposed by Pennell and Dunson (2006) can be expressed as a conditional Cox model with time-dependent covariates.

We propose an alternative partially PH modelling approach that generalizes the shared random effect PH models. In our proposal the cluster-specific baseline hazard functions are treated as functional random parameters and a BNP framework is used to provide a flexible and robust model for them. Specifically, the cluster-specific baseline hazard functions are modelled as a mixture model induced by a completely random measure (CRM), in the class of generalized-gamma CRMs (Dykstra and Laud, 1981, Lo and Weng, 1989, Nieto-Barajas and Walker, 2004, James, 2005, Lijoi and Nipoti, 2014, Arbel et al., 2016). The resulting framework allows for a convenient modelling of heterogenous intra-cluster associations, by accommodating covariate-dependent marginal associations between the survival functions of subjects within a cluster. We illustrate the marginal association structure induced by the proposed model by studying the behavior of Kendall's τ and survival ratio, two quantities that have been widely used in the literature to assess the intra-cluster dependence (Anderson et al., 1992). We also show that a particular version of the proposed model retains the PH structure marginally and that it is the only model with such property in the class of the generalized-gamma CRMs.

2 The modelling approach

2.1 The conditional partially PH model

Let $T_{i,j} \in \mathbb{R}_+$ be the time-to-event for the ith individual in the jth cluster, with $j = 1, \ldots, r$ and $i = 1, \ldots, n_j$. Let $\mathbf{z}_{i,j} \in \mathcal{Z} \subseteq \mathbb{R}^p$ be a p–dimensional vector of explanatory covariates associated with the ith individual in the jth cluster. We assume that, given the cluster-specific baseline hazard function h_j, $j = 1, \ldots, r$, the time-to-event variables $T_{i,j}$ are independent, following a

clustered PH model with conditional density $p(\cdot \mid \mathbf{z}_{i,j}, \boldsymbol{\beta}, h_j)$, that is,

$$T_{i,j} \mid \mathbf{z}_{i,j}, \boldsymbol{\beta}, h_j \overset{ind.}{\sim} p(\cdot \mid \mathbf{z}_{i,j}, \boldsymbol{\beta}, h_j), \tag{1}$$

where

$$p(t \mid \mathbf{z}_{i,j}, \boldsymbol{\beta}, h_j) = \exp\{\mathbf{z}'_{i,j}\boldsymbol{\beta}\} h_j(t) \exp\left[-\exp\{\mathbf{z}'_{i,j}\boldsymbol{\beta}\} \int_0^t h_j(u)du\right], \tag{2}$$

which implies that the conditional hazard function for the ith individual in the jth cluster is given by $h_{i,j}(\cdot) = h_j(\cdot)\exp\{\mathbf{z}'_{i,j}\boldsymbol{\beta}\}$. We complete the model specification by assuming that the cluster-specific hazard functions are random functional parameters with common distribution. Specifically, we extend the ideas proposed by Dykstra and Laud (1981) in the context of non-clustered data and assume that the conditional hazard functions can be expressed as a mixture model induced by iid cluster-specific random distributions,

$$h_j(t) = \int_{\mathbb{Y}} k(t \mid y)\mu_j(\mathrm{d}y) \tag{3}$$

and

$$\mu_j \mid G \overset{iid}{\sim} G,$$

$j = 1, \ldots, r$, where $\mathbb{Y}$ is an appropriate measurable space, $k(\cdot \mid \cdot)$ is a suitable kernel (i.e., a jointly measurable mapping from $\mathbb{R}^+ \times \mathbb{Y}$ to $\mathbb{R}^+$), μ_j is a random probability measure defined on $\mathbb{Y}$, such that $\lim_{t\to\infty} \int_0^t h_j(s)\mathrm{d}s = +\infty$ a.s., and G is the common probability law for the mixing distributions. Here we assume that G is the law of a CRM, parameterized by a finite-dimensional parameter ϑ, for which we introduce the notation CRM(ϑ). The specific class of CRM employed in the definition of the model is given in Section 2.2, along with a brief discussion of its basic properties.

The modelling approach given by expressions (1) – (3) generalizes the class of shared frailty PH models, which can be recovered as a special case of our framework. Specifically, if we assume $k(t \mid y) = h_0(t)m(y)$ in (3), for some suitable functions h_0 and m, then the resulting conditional hazard function is given by $h_j(t) = h_0(t)V_j$, where $V_j \overset{D}{=} \int_{\mathbb{Y}} m(y)\mu_j(\mathrm{d}y)$. As in the shared frailty PH model, the proposed model assumes that the conditional PH assumption holds for subjects within the same cluster. However, our modelling approach also allows for non-proportional conditional hazard functions across clusters.

2.2 Completely random measures

The notion of CRM was first introduced by Kingman (1967) and can be thought of as the foundation of many commonly used BNP priors (see, e.g., Lijoi and Prünster, 2010). A CRM

μ_j is defined as a boundedly finite random measure on a measurable space $\mathbb{Y}$ such that, for any collection $A_1, \ldots, A_n$ of pairwise disjoint measurable subsets of $\mathbb{Y}$, the random variables $\mu_j(A_1), \ldots, \mu_j(A_n)$ are independent. We assume that μ_j, $j = 1, \ldots, r$, has only random atoms, which implies that it is fully characterized by its Lévy intensity ν_ϑ (a measure on $\mathbb{R}^+ \times \mathbb{Y}$, parametrized by ϑ), through the Lévy-Khintchine representation,

$$\mathbb{E}_\vartheta\left[\mathrm{e}^{-\mu_j(f)}\right] = \exp\left\{-\int_{\mathbb{R}^+\times\mathbb{Y}}\left(1-\mathrm{e}^{-sf(y)}\right)\nu_\vartheta(\mathrm{d}s,\mathrm{d}y)\right\}, \tag{4}$$

for any $f : \mathbb{Y} \to \mathbb{R}$ such that $\int |f|\mathrm{d}\mu_j < \infty$, where $\mathbb{E}_\vartheta(\cdot) = \mathbb{E}_{\mu_j|\vartheta}(\cdot)$. Thus (4) provides an explicit expression for the Laplace functional transform $\mathbb{E}_\vartheta\left[\mathrm{e}^{-\mu_j(\cdot)}\right]$ of the CRM μ_j. For future convenience we introduce also the Laplace exponent ψ_j of μ_j, defined as

$$\psi_j(f) = -\log\left(\mathbb{E}_\vartheta\left[\mathrm{e}^{-\mu_j(f)}\right]\right). \tag{5}$$

It is often useful to write ν_ϑ as the product of a transition kernel ρ_y, and a diffuse measure α on $\mathbb{Y}$,

$$\nu_\vartheta(\mathrm{d}s,\mathrm{d}y) = \rho_y(s)\mathrm{d}s\,\alpha(\mathrm{d}y).$$

Because it includes important special cases, we focus here on the class of generalized-gamma (gg) CRMs (Brix, 1999). A gg-CRM is a homogeneous CRM, i.e. $\rho_y(s) = \rho(s)$ for every y, characterized by

$$\rho(s) = \frac{\sigma}{\Gamma(1-\sigma)}s^{-1-\sigma}\mathrm{e}^{-\theta s},$$

for some parameters $\sigma \in (0,1)$ and $\theta \geq 0$. A gamma-CRM corresponds to $\sigma \to 0$ and $\theta = 1$, and has transition kernel given by $\rho(s) = \exp\{-s\}s^{-1}$, whereas a σ-stable CRM sets $\theta = 0$. We further consider $\alpha = cP_0$, where $c > 0$ and P_0 is a probability measure defined on $\mathbb{Y}$. In summary, we complete the model specification by assuming that

$$\mu_1, \ldots, \mu_k \mid \sigma, c, P_0 \overset{iid}{\sim} \text{gg-CRM}(\sigma, \theta, c, P_0). \tag{6}$$

The properties of the random hazard functions h_j, induced by model (3) and (6), depend on the choice of the kernel function. The effect of the kernel has been extensively investigated (see, e.g., Lo and Weng, 1989, James, 2005, De Blasi et al., 2009). However, to the best of our knowledge, only De Blasi et al. (2009) provides insights on the effect of the choice of the mixing CRM, by investigating the asymptotic properties of functionals of the hazard mixture, both a priori and a posteriori.

With reference to Section 2.1, we can retrieve popular specifications of shared frailty PH models by assuming $m(y) = \mathbb{1}_{\mathbb{Y}}(y)$ and suitably specifying the distribution of the CRMs, where

$\mathbb{1}_{\mathbb{Y}}(y)$ is the indicator function, taking the value 1 if $y \in \mathbb{Y}$ and 0 otherwise. Specifically, if we assume that μ_j is a gamma CRM, then $V_j \sim \mu_j(\mathbb{Y})$ is a gamma random variable with shape parameter 1 and scale parameter c. Similarly, if μ_j is a σ-stable CRM and $c = 1$, then V_j has positive stable distribution with parameter $\sigma \in (0, 1)$.

3 Distributional and association properties

3.1 The marginal distribution properties

Let $\tilde{h}_{i,j}$ and $\tilde{S}_{i,j}$ be the marginal hazard and survival function for the ith subject in the $j$the cluster under the proposed model. Therefore, $\tilde{S}_{i,j}(t) = \mathbb{E}_\vartheta[S_{i,j}(t)]$, where $S_{i,j}(t) = \exp\left\{-\int_0^t h_{i,j}(s)\mathrm{d}s\right\}$ is the corresponding conditional survival function, and $\tilde{h}_{i,j}(t) = -\frac{\partial \tilde{S}_{i,j}/\partial t}{\tilde{S}_{i,j}}$. Based on results by James (2005), it is straightforward to show that

$$\mathbb{E}_\vartheta[h_{i,j}(t) \,|\, T_{i,j} \geq t] = -\frac{\partial \mathbb{E}_\vartheta[S_{i,j}(t)]/\partial t}{\mathbb{E}_\vartheta[S_{i,j}(t)]} = -\frac{\partial \tilde{S}_{i,j}/\partial t}{\tilde{S}_{i,j}} = \tilde{h}_{i,j}(t),$$

i.e. $\mathbb{E}_\vartheta\left[h_{i,j}(t) \,|\, T_{i,j} \geq t\right]$ coincides with the population average hazard function. By means of the Lévy-Khintchine representation, the following expressions are obtained for the marginal survival and hazard function under any class of CRMs, respectively,

$$\tilde{S}_{i,j}(t) = \exp\left\{-\int_{\mathbb{R}^+\times\mathbb{Y}} \left(1 - \exp\left\{-sK_t(y)\,\mathrm{e}^{\mathbf{z}'_{i,j}\boldsymbol{\beta}}\right\}\right) \nu_\vartheta(\mathrm{d}s, \mathrm{d}y)\right\},$$

and

$$\tilde{h}_{i,j}(t) = \int_{\mathbb{R}^+\times\mathbb{Y}} \exp\left\{-sK_t(y)\,\mathrm{e}^{\mathbf{z}'_{i,j}\boldsymbol{\beta}}\right\} k(t \mid y)\, \nu_\vartheta(\mathrm{d}s, \mathrm{d}y)\mathrm{e}^{\mathbf{z}'_{i,j}\boldsymbol{\beta}}, \tag{7}$$

where $K_t(y) = \int_0^t k(s \mid y)\,\mathrm{d}s$. Under model (1) – (6), with $\theta = 0$, the previous expressions reduce to

$$\tilde{S}_{i,j}(t) = S'_0(t)^{\exp\{\sigma \mathbf{z}'_{i,j}\boldsymbol{\beta}\}},$$

and

$$\tilde{h}_{i,j}(t) = h'_0(t)\mathrm{e}^{\sigma \mathbf{z}'_{i,j}\boldsymbol{\beta}}, \tag{8}$$

respectively, where $S'_0(t) = \exp\left\{-\int_{\mathbb{Y}} K_t(y)^\sigma \alpha(\mathrm{d}y)\right\}$ and $h'_0(t) = \sigma \int_{\mathbb{Y}} k(t \mid y) K_t(y)^{\sigma-1}\alpha(\mathrm{d}y)$. Therefore, if we assume a σ-stable CRM, our model remarkably retains the PH assumption marginally. As a matter of fact, this property characterizes the σ-stable version of our model also with respect to all other alternatives based on any $\mu_j(\cdot)$ in the class of gg-CRMs. The following result is proved in the Appendix A of the online supplementary material.

Proposition 1. *Consider the conditional partially PH model given by expressions* (1) – (3)*, where* $\mu_1, \ldots, \mu_r$ *are independent and identically distributed gg-CRMs with parameters* $\sigma \in (0, 1)$ *and* $\theta \geq 0$*. Then, the PH structure is retained marginally if and only if* $\theta = 0$*.*

To illustrate how the marginal structure can vary when different members of the class of gg-CRMs are considered, we also illustrate the case of the commonly used gamma CRM. Under the conditional partially PH model given by expressions (1) – (3), where

$$\mu_1, \ldots, \mu_r \mid c, P_0 \overset{iid}{\sim} \text{gamma CRM}(c, P_0),$$

the corresponding expressions for the marginal survival and hazard functions are given, respectively, by

$$\tilde{S}_{i,j}(t) = \exp\left\{ - \int_{\mathbb{Y}} \log\left(1 + K_t(y) \mathrm{e}^{\mathbf{z}'_{i,j}\boldsymbol{\beta}}\right) \alpha(\mathrm{d}y) \right\},$$

and

$$\tilde{h}_{i,j}(t) = \int_{\mathbb{Y}} \frac{k(t \mid y) \mathrm{e}^{\mathbf{z}'_{i,j}\boldsymbol{\beta}}}{1 + K_t(y) \mathrm{e}^{\mathbf{z}'_{i,j}\boldsymbol{\beta}}} \alpha(\mathrm{d}y).$$

The different marginal structures implied by the gamma and σ-stable versions of the model emphasize the interesting correspondence between the BNP formulation considered here and the commonly used parametric shared frailty PH models. In the recent literature, there has been a renewed attention on positive stable frailty models (see, e.g, Choi and Huang, 2012). The positive stable distribution serves as a "bridge distribution" for clustered time-to-event data under a shared frailty PH model since the regression parameter in the marginal model can be obtained as a product of the conditional regression parameters and the frailty parameter σ (Liu et al., 2011), similarly to what we obtained under the proposed σ-stable model. In both frameworks, the parameter σ has a tempering effect, since the value of σ attenuates the effect of the covariates in the marginal model. This leads to a straightforward interpretation of the model parameters σ and $\boldsymbol{\beta}$. We will further clarify the interpretation of σ for the modelling of the association structure within each cluster in Section 3.2.

An important advantage of the proposed model over other BNP-based approaches is that the marginal likelihood of the clustered time-to-event data can be analytically computed for some specific kernel functions, speeding up posterior computations dramatically. Specifically, we consider a generalization of the kernel proposed by Dykstra and Laud (1981), given by $k(t \mid y) = d(t)\mathbb{1}_{(0,t^a]}(y)$, where $d(t)$ is a positive real-valued function and $a > 0$. In the following, we explicitly obtain the expression of the marginal distribution under a generic CRM, assuming that $d(t) = \mathrm{e}^b$, where $b \in \mathbb{R}$. This choice ensures that the support of the resulting random hazard

functions covers the space of the increasing hazard functions, where the model is known to hold consistency properties in a non-clustered data scenario (De Blasi et al., 2009). The resulting version of the model is, therefore, suitable for the wide range of applications, e.g. biomedical, in which risk increases with age. For simplicity, we limit our discussion to the case of no censoring. However, the results can be extended to the case of any uninformative censoring scheme at the expense of an increased notational burden. The following proposition is proved in the Appendix B of the online supplementary material.

Proposition 2. *Consider the conditional partially PH model given by expressions* (1) – (3)*, where* $\mu_j \mid \vartheta \overset{iid}{\sim} CRM(\vartheta)$ *and* $k(t \mid y) = \mathrm{e}^b \mathbb{1}_{(0,t^a]}(y)$*. Set* $\mathbf{T}_j = (T_{1,j}, \ldots, T_{n_j,j})$ *and* $\mathbf{z}_j = (\mathbf{z}'_{1,j}, \ldots, \mathbf{z}'_{n_j,j})'$*. Then, for any* $j = 1, \ldots, r$*, it follows that the marginal distribution of the time-to-event data is given by*

$$p(\mathbf{T}_j \mid \boldsymbol{\beta}, \mathbf{z}_j) = \exp\left\{ n_j b - \sum_{i=1}^{n_j} \left(\int_{T^a_{(i-1),j}}^{T^a_{(i),j}} \int_{\mathbb{R}^+} \left(1 - \exp\left\{-sF_{i,j}(y)\right\}\right) \nu_\vartheta(\mathrm{d}s, \mathrm{d}y) - \mathbf{z}'_{i,j}\boldsymbol{\beta} \right) \right\} \times (-1)^{n_j} \sum_{\ell \in \mathcal{L}_j} \prod_{i=1}^{n_j} \mathcal{B}_{n^{(\ell)}_{i,j}}\left(\zeta^{(j)}_{\bullet,i}\right),$$

where $\{(T_{(1),j}, \mathbf{z}_{(1),j}), \ldots, (T_{(n_j),j}, \mathbf{z}_{(n_j),j})\}$ *are the ordered pairs* $((T_{i,j}, \mathbf{z}_{i,j}))_{i=1}^{n_j}$*, with the order being taken with respect to the increasing values of the elements in* $\mathbf{T}_j$*,*

$$F_{i,j}(y) = \sum_{r=i}^{n_j} (T_{(r),j} - y^{1/a}) \mathrm{e}^{\mathbf{z}'_{(r),j}\boldsymbol{\beta} + b},$$

$$\mathcal{L}_j = \{\ell = (\ell_1, \ldots, \ell_{n_j}) \text{ s.t. } \ell_i \in \{1, \ldots, i\} \text{ for every } i = 1, \ldots, n_j\},$$

$\mathcal{B}_{n^{(\ell)}_{i,j}}\left(\zeta^{(j)}_{\bullet,i}\right) = \mathcal{B}_{n^{(\ell)}_{i,j}}\left(\zeta^{(j)}_{1,i}, \ldots, \zeta^{(j)}_{n^{(\ell)}_{i,j},i}\right)$ *is the n-th complete Bell polynomial, with* $n^{(\ell)}_{i,j} = \#\{r \in \{1, \ldots, n_j\} \text{ s.t. } \ell_r = i\}$*, and, for each* ℓ *and* i *such that* $n^{(\ell)}_{i,j} \geq 1$*,*

$$\zeta^{(j)}_{r,i} = (-1)^r \int_{T^a_{(i-1),j}}^{T^a_{(i),j}} \int_{\mathbb{R}^+} \exp\left\{-sF_{i,j}(y)\right\} s^r \nu_\vartheta(\mathrm{d}s, \mathrm{d}y),$$

for $r = 1, \ldots, n^{(\ell)}_{i,j}$*. We agree that* $T_{(0),j} = 0$*.*

Proposition 2 is valid for any family of CRMs. When the Laplace transform of μ_j is known, the marginal distribution of the responses assigned to the same cluster can be written in closed form. Furthermore, simplifications are possible if n_j is small, which is a common situation in many applications of frailty models. For example, when the cluster size is $n_j = 2$ and $\alpha = U(0, T]$

for some $T > 0$, then the marginal distribution under the σ-stable version of the proposed model is given by

$$p(\mathbf{T}_j \mid \boldsymbol{\beta}, \mathbf{z}_j) = \exp\left\{-\left(\mathcal{I}_1^{(j)} + \mathcal{I}_2^{(j)}\right)\right\} A_{1,j} A_{2,j} \left(\zeta_{1,1}^{(j)}\zeta_{1,2}^{(j)} + \left(\zeta_{1,1}^{(j)}\right)^2 + \zeta_{2,1}^{(j)}\right),$$

where $A_{i,j} = \mathrm{e}^{\mathbf{z}'_{(i),j}\boldsymbol{\beta}+b}$, for $i = 1, 2$, and, if we agree on the notation $C_j := A_{1,j}T_{(1),j} + A_{2,j}T_{(2),j}$,

$$\mathcal{I}_1^{(j)} = C_j^{\sigma}\, {}_2F_1\left(a, -\sigma; 1+a; \frac{(A_{1,j}+A_{2,j})\min(T, T_{(1),j}^a)^{1/a}}{C_j}\right)\frac{\min(T, T_{(1),j}^a)}{T},$$

$$\begin{aligned}\mathcal{I}_2^{(j)} = (A_{2,j}T_{(2),j})^{\sigma}\Bigg({}_2F_1\left(a, -\sigma; 1+a; \frac{\min(T, T_{(2),j}^a)^{1/a}}{T_{(2),j}}\right)\frac{\min(T, T_{(2),j}^a)}{T} \\ - {}_2F_1\left(a, -\sigma; 1+a; \frac{\min(T, T_{(1),j}^a)^{1/a}}{T_{(1),j}}\right)\frac{\min(T, T_{(1),j}^a)}{T}\Bigg),\end{aligned}$$

$$\zeta_{1,1}^{(j)} = C_j^{\sigma-1}\sigma\, {}_2F_1\left(a, 1-\sigma; 1+a; \frac{(A_{1,j}+A_{2,j})\min(T, T_{(1),j}^a)^{1/a}}{C_j}\right)\frac{\min(T, T_{(1),j}^a)}{T},$$

$$\begin{aligned}\zeta_{1,2}^{(j)} := (A_{2,j}T_{(2),j})^{\sigma-1}\sigma\Bigg({}_2F_1\left(a, 1-\sigma; 1+a; \frac{\min(T, T_{(2),j}^a)^{1/a}}{T_{(2),j}}\right)\frac{\min(T, T_{(2),j}^a)}{T} \\ - {}_2F_1\left(a, 1-\sigma; 1+a; \frac{\min(T, T_{(1),j}^a)^{1/a}}{T_{(1),j}}\right)\frac{\min(T, T_{(1),j}^a)}{T}\Bigg),\end{aligned}$$

$$\zeta_{2,1}^{(j)} = -C_j^{\sigma-2}(\sigma-1)\, {}_2F_1\left(a, 2-\sigma; 1+a; \frac{(A_{1,j}+A_{2,j})\min(T, T_{(1),j}^a)^{1/a}}{C_j}\right)\frac{\min(T, T_{(1),j}^a)}{T},$$

where ${}_2F_1$ denotes the Gaussian hypergeometric function. $\{(T_{(1),j}, \mathbf{z}_{(1),j}), (T_{(2),j}, \mathbf{z}_{(2),j})\}$, as in Proposition 2, are the ordered pairs $((T_{i,j}, \mathbf{z}_{i,j}))_{i=1}^2$ where the order is taken with respect to increasing values of the elements in $\mathbf{T}_j$.

3.2 The marginal association structure properties

We derive the expressions for the Kendall's τ and survival ratio under the proposed model. The former describes the association between two survival functions and, therefore, is considered as a global measure of dependence. The latter evaluates the degree of dependence at a single time point, thus capturing changes over time and local dependence (see, e.g., Anderson et al., 1992). Without loss of generality, let us consider the time-to-event data for the first two units in a cluster j, $T_{1,j}$ and $T_{2,j}$. The joint survival function for $(T_{1,j}, T_{2,j})$, $S_{1,2,j}$, is given by

$$S_{1,2,j}(t_1, t_2) = \exp\left\{-\int_0^{t_1} h_{1,j}(s)\mathrm{d}s - \int_0^{t_2} h_{2,j}(s)\mathrm{d}s\right\}.$$

Let $\tau_{1,2,j}$ and $\Sigma_{1,2,j}(t_1,t_2)$ denote the Kendall's τ and the survival ratio associated with the marginal distribution implied by the conditional PH model, respectively. We provide explicit expressions for these association parameters, by assuming that $\mu_j \mid \vartheta \overset{iid}{\sim} \text{CRM}(\vartheta)$. The following proposition is proved in the Appendix C of the online supplementary material.

Proposition 3. *Consider the conditional PH model given by expressions* (1) – (3)*, where* $\mu_j \mid \vartheta \overset{iid}{\sim} CRM(\vartheta)$*. The Kendall's* τ *and survival ratio for the first two experimental units in the* j*th cluster is given by*

$$\begin{aligned}
\tau_{1,2,j} &= \int_0^\infty \int_0^\infty \mathbb{E}_\vartheta[S_{1,2,j}(t_1,t_2)] \frac{\partial^2}{\partial t_1 \partial t_2} \left\{\mathbb{E}_\vartheta[S_{1,2,j}(t_1,t_2)]\right\} \mathrm{d}t_1 \mathrm{d}t_2, \\
&= 4\mathrm{e}^{(\mathbf{z}_{1,j}+\mathbf{z}_{2,j})'\boldsymbol{\beta}} \int_0^\infty \int_0^\infty \mathrm{e}^{-2\psi\left(K^{(1,2)}_{t_1,t_2}\right)} \left[\int_{\mathbb{Y}} \xi_2\left(K^{(1,2)}_{t_1,t_2}(y)\right) k_1(t_1 \mid y)\, k_2(t_2 \mid y)\, \alpha(\mathrm{d}y) \right. \\
&\quad \left. + \int_{\mathbb{Y}} \xi_1\left(K^{(1,2)}_{t_1,t_2}(y)\right) k_1(t_1 \mid y)\, \alpha(\mathrm{d}y) \int_{\mathbb{Y}} \xi_1\left(K^{(1,2)}_{t_1,t_2}(y)\right) k_2(t_2 \mid y)\, \alpha(\mathrm{d}y)\right] \mathrm{d}t_1\, \mathrm{d}t_2 - 1,
\end{aligned} \tag{9}$$

and

$$\begin{aligned}
\Sigma_{1,2,j}(t_1,t_2) &= \frac{\mathbb{E}_\vartheta[S_{1,2,j}(t_1,t_2)]}{\mathbb{E}_\vartheta[S_{1,j}(t_1)]\mathbb{E}_\vartheta[S_{2,j}(t_2)]}, \\
&= \exp\left\{-\int_{\mathbb{Y}\times\mathbb{R}^+} \left(\mathrm{e}^{-sK^{(1)}_{t_1}(y)} + \mathrm{e}^{-sK^{(2)}_{t_2}(y)} - \mathrm{e}^{-sK^{(1,2)}_{t_1,t_2}(y)} - 1\right) \nu_\vartheta(\mathrm{d}s, \mathrm{d}y)\right\},
\end{aligned} \tag{10}$$

respectively, where, for notational simplicity, we denote $k_i(t \mid y) = k(t \mid y)\, \exp\{\mathbf{z}'_{i,j}\boldsymbol{\beta}\}$*,* $K^{(i)}_t(y) = \int_0^t k_i(s \mid y)\, \mathrm{d}s$*,* $i = 1, 2$*,* $K^{(1,2)}_{t_1,t_2}(y) = K^{(1)}_{t_1}(y) + K^{(2)}_{t_2}(y)$*,* $\xi_n(q) = \int_0^\infty s^n \exp\{-qs\}\rho_y(s)\mathrm{d}s$*, and* ψ *is the Laplace exponent of* μ_j*, for any* $j = 1, \ldots, r$*, as defined in* (5)*.*

From Proposition 3 it follows that, under the σ-stable version of the proposed model, the Kendall's τ and survival ratio for the first two experimental units in the jth cluster is given by

$$\begin{aligned}
\tau_{1,2,j} &= 4\sigma c\, \mathrm{e}^{(\mathbf{z}_{1,j}+\mathbf{z}_{2,j})'\boldsymbol{\beta}} \int_0^\infty \int_0^\infty \mathrm{e}^{-2c\int_{\mathbb{Y}}\left(K^{(1,2)}_{t_1,t_2}(y)\right)^\sigma P_0(\mathrm{d}y)} \\
&\quad \times \left\{(1-\sigma)\int_{\mathbb{Y}} \left(K^{(1,2)}_{t_1,t_2}(y)\right)^{\sigma-2} k_1(t_1 \mid y) k_2(t_2 \mid y) P_0(\mathrm{d}y) \right. \\
&\quad \left. + \sigma c \int_{\mathbb{Y}} \left(K^{(1,2)}_{t_1,t_2}(y)\right)^{\sigma-1} k_1(t_1 \mid y) P_0(\mathrm{d}y) \int_{\mathbb{Y}} \left(K^{(1,2)}_{t_1,t_2}(y)\right)^{\sigma-1} k_2(t_2 \mid y) P_0(\mathrm{d}y)\right\} \\
&\quad \mathrm{d}t_1\, \mathrm{d}t_2 - 1,
\end{aligned}$$

and

$$\Sigma_{1,2,j}(t_1,t_2) = \exp\left\{-c\int_{\mathbb{Y}} \left(\left(K^{(1,2)}_{t_1,t_2}(y)\right)^\sigma - K^{(1)}_{t_1}(y)^\sigma - K^{(2)}_{t_2}(y)^\sigma\right) P_0(\mathrm{d}y)\right\},$$

respectively. From Proposition 3 it also follows that by assuming

$$\mu_1, \ldots, \mu_r \mid c, P_0 \overset{iid}{\sim} \text{gamma CRM}(c, P_0),$$

then,

$$\begin{aligned}\tau_{1,2,j} &= 4c\mathrm{e}^{(\mathbf{z}_{1,j}+\mathbf{z}_{2,j})'\boldsymbol{\beta}} \int_0^\infty \int_0^\infty \mathrm{e}^{-2c\int_{\mathbb{Y}} \log\left(1+K^{(1,2)}_{t_1,t_2}(y)\right)P_0(\mathrm{d}y)} \left[\int_{\mathbb{Y}} \frac{k_1(t_1 \mid y)k_2(t_2 \mid y)}{(1+K^{(1,2)}_{t_1,t_2}(y))^2} P_0(\mathrm{d}y)\right. \\ &\quad \left. + c\int_{\mathbb{Y}} \frac{k_1(t_1 \mid y)}{1+K^{(1,2)}_{t_1,t_2}(y)} P_0(\mathrm{d}y) \int_{\mathbb{Y}} \frac{k_2(t_2 \mid y)}{1+K^{(1,2)}_{t_1,t_2}(y)} P_0(\mathrm{d}y)\right] \mathrm{d}t_1\, \mathrm{d}t_2 - 1,\end{aligned}$$

and

$$\Sigma_{1,2,j}(t_1, t_2) = \exp\left\{-c\int_{\mathbb{Y}} \log \frac{1+K^{(1,2)}_{t_1,t_2}(y)}{(1+K^{(1)}_{t_1}(y))(1+K^{(1)}_{t_2}(y))} P_0(\mathrm{d}y)\right\}.$$

Despite their apparent complexity, the previous expressions can be explicitly evaluated once the kernel $k(\cdot \mid \cdot)$ and the normalized base measure P_0 are specified. We can also observe that $\Sigma_{1,2,j}(t_1, t_2) = 1$ corresponds to the case of intra-cluster independence, whereas $\Sigma_{1,2,j}(t_1, t_2) > 1$ (resp. $\Sigma_{1,2,j}(t_1, t_2) < 1$) corresponds to the case of positive (resp. negative) intra-cluster association. The previous expressions also show that both the Kendall's τ and the survival ratio may depend on the values of the individual specific covariates and thus they are able to accommodate different degrees of association within each cluster, in contrast to standard frailty models. It is important to stress that predictor-dependent association structures can be obtained also by adopting other models. For instance, this can be achieved by considering time-dependent frailty terms or by relaxing the iid assumption of the frailties and introducing the dependence on predictors of the distribution of the frailty terms. Unfortunately, to the best of our knowledge, the existing literature on these models have not explored the induced marginal association structure of the model. This can therefore be considered as an advantage of the proposed conditional partially PH model, for which explicit expressions for common marginal association measures can be found. Another way of inducing a predictor dependent association structure and, at the same time, avoid assumptions on the relationship between predictors and the distribution of the time-to-event, is by considering dependent mixture models. Such an alternative approach will be considered in Section 5.2 where the fit of our model is compared with a linear dependent Dirichlet process (LDDP) mixture (see, e.g. De Iorio et al., 2004, 2009, Jara et al., 2010). For the sake of illustration, we assume now that $c = 1$ in the σ-stable version of the proposed model. This allows us to draw a comparison of the two proposed versions of our model, each one being characterized by only one parameter, that is $\sigma \in (0, 1)$ for the σ-stable

version and $c > 0$ and the gamma version. As discussed previously, shared frailty models are obtained as a special case of the proposed model by setting $k(t \mid y) = h_0(t)m(y)$. Indeed, by further assuming that $m(y) = \mathbb{1}_{\mathbb{Y}}(y)$, we can obtain the well-known expressions for the Kendall's τ under the gamma and stable shared frailty models (see, e.g., Duchateau and Janssen, 2008). More specifically, we obtain $\tau_c = 1/(1+2c)$ and $\tau_\sigma = 1 - \sigma$ for the gamma and the σ-stable version of the frailty PH model, respectively.

We now illustrate how Kendall's τ changes as a function of subject-specific covariates in each cluster. In this illustration we set $k(t \mid y) = \mathbb{1}_{(0,t]}(y)$. Furthermore, we set P_0 to be a uniform distribution on the finite interval $(0, 1]$. We consider a single covariate per each individual and focus on 4 possible combinations of values, namely $\{(0,0), (3,3), (0,3), (-3,3)\}$.

Figure 1 shows the values of the Kendall's τ under the two versions of proposed model, as a function of the parameter values. In both cases, we note that the Kendall's τ is positive, indicating a positive overall dependence. If we compare the results with the Kendall's τ obtained under shared frailty PH models, i.e. τ_c and τ_σ reported above, a few similar features can be recognized. In the gamma case, when $c \to 0$, τ approaches 1, whereas as c increases, $\tau \to 0$. Similarly, in the σ-stable case, when $\sigma \to 0$, $\tau \to 1$, whereas if $\sigma \to 1$, then $\tau \to 0$. However, Figure 1 also shows that, for any fixed value of the parameters c and σ, different combinations of the covariate values generate significantly different association structures within a cluster, which is a particular feature of our modelling framework only.

[Figure 1 about here.]

We note that the pattern of dependence is quite different between the two versions of the proposed model. The exploratory analyses show that under the σ-stable version of the model, the Kendall's τ is affected more by the relative differences observed between the cluster covariates $|Z_1 - Z_2|$ than their individual magnitudes. This behaviour is exemplified in Figure 1(a). This shows that when $Z_1 = Z_2 = 0$ and $Z_1 = Z_2 = 3$, the curves are very close to each other. On the other hand, when $(Z_1, Z_2) = (0, 3)$ and $(Z_1, Z_2) = (-3, 3)$ significantly different Kendall's τ functions are obtained. A similar interpretation does not appear to hold under the gamma version of the model.

The two versions of the conditional partially PH model also show different local dependence structures. In order to avoid confounding of global and covariate effects, we illustrate the local dependence structure assuming null covariate values. We further set c and σ so that the overall Kendall's τ is 0.25 for both the σ-stable and the gamma version of the model. This is achieved

when $c \approx 1.905$ and $\sigma \approx 0.736$. Figure 2 displays the contour curves for the survival ratio, $\Sigma_{1,2,j}(t_1, t_2)$, as a function of the survival functions. Both cases show positive local intra-class correlation. Furthermore, higher values of the survival ratio correspond to smaller values of the marginal survival functions, or, equivalently, larger t_1 and t_2. However, by comparing Figure 2(a) and 2(b), it is evident that the gamma and σ-stable CRM based models are characterized by late and early intra–cluster dependence, respectively. More specifically, contour lines indicating positive correlation appear earlier for the σ-stable CRMs, corresponding to higher values of the survival functions. For example, compare the contour lines for the value $\Sigma = 1.05$ in Figure 2. On the other hand, the rate of increase of the survival ratio is slower in the σ-stable case. Thus, high values of the survival ratio appear earlier when gamma CRMs are considered. As a matter of fact, in the σ-stable case, two units of the same cluster tend to be relatively weakly correlated in the long term. Those patterns of failures are often observed in familial associations of onset ages for diseases with low penetrance (Fine et al., 2003). Therefore, the parameter σ can be thought of as a dependence parameter. As suggested by our numerical study, if $\sigma \to 1$, then $\tau \to 0$ and $\Sigma \to 1$, capturing both local and global independence between survival times. On the other hand, a value $\sigma < 1$ reflects positive correlation between observations within and between clusters. The interpretation of σ as a parameter capturing the dependence in a cluster is supported also by the marginalization properties of the proposed σ-stable model, since in the marginal model the parameter σ affects multiplicatively the regression coefficients $\boldsymbol{\beta}$. Hence, the stronger is the association between survival times in the same cluster (the smaller the value of σ), the weaker should be the effect of the individual covariates of the subjects in the cluster. Once again, the previous discussion shows how our modelling framework preserves and extends well-known results for the shared frailty PH models with gamma and positive stable distributions (Duchateau and Janssen, 2008).

[Figure 2 about here.]

4 A simulation study

We first illustrate the performances of the proposed modelling approach on simulated datasets.

4.1 The simulation settings

The clustered time-to-event data were simulated under two different simulation scenarios. In the first case we consider a σ-stable shared frailty PH model, with a Weibull baseline hazard

function $h_0(t) = \kappa t^{\kappa-1}/\lambda^{\kappa}$ (Manatunga and Oakes, 1999), where $\kappa = 1.1$ and $\lambda = 1$, a model specification which corresponds to increasing hazard functions. In the second case we consider a conditional partially PH model given by expressions (1) – (3), with $k(t \mid y) = \mathrm{e}^b \mathbb{1}_{(0,t^a]}(y)$ and where $\mu_1, \ldots, \mu_r$ are iid σ-stable CRMs with base measure $\alpha(\cdot) = \mathbb{1}_{(0,T]}(\cdot)$, where $a = 2$, $T = 1$ and $b = 0$. Under both simulation scenarios, we consider three data-generating mechanisms, obtained by setting $\sigma \in \{0.25, 0.5, 0.75\}$. It is important to note that with these specifications, the standard σ-stable shared frailty PH model is a limiting case of our proposed model for $T \to 0$.

For each data-generating mechanism, we consider three different sample sizes by setting $r = 100$, 200 and 500. In all cases we set $n_j = 2$, and consider a single binary predictor, $z_{i,j} \in \{0, 1\}$, randomly generated from a Bernoulli of parameter $p = 0.5$, and set $\beta = 4$. For each of the 18 simulation settings, we generate 100 Monte Carlo replicates. The CRMs μ_j for the proposed model were generated via the Ferguson & Klass algorithm (see, e.g., Orbanz and Williamson, 2011).

For each simulated dataset, the σ-stable version of the proposed conditional partially PH model was fit, by assuming $k(t \mid y) = \mathrm{e}^b \mathbb{1}_{(0,t^a]}(y)$, $\alpha(\cdot) = \mathbb{1}_{(0,T]}(\cdot)$, and using the Markov chain Monte Carlo (MCMC) algorithm described in the Appendix D of the online supplementary material. The model specification was completed by assuming $\beta \sim \mathrm{N}(0, 1000)$, $\sigma \sim \mathrm{U}(0, 1)$, $b \sim \mathrm{N}(0, 10)$, $\log(T) \sim \mathrm{N}(0, 10)$ and $\log(a) \sim \mathrm{N}(0, 10)$. For comparison purposes, the σ-stable shared frailty PH model, with a Weibull baseline hazard function, was also fit to each simulated dataset. In this case, the model was completed by assuming $\beta \sim \mathrm{N}(0, 1000)$, $\sigma \sim \mathrm{U}(0, 1)$, $\kappa \sim \mathrm{N}(1, 10)\mathbb{1}_{\{\kappa>1\}}$ and $\log(\lambda) \sim \mathrm{N}(0, 10)$. The choice of a prior for κ with support $(1, \infty)$, makes the comparison between the two models fair as both have support limited to increasing hazard functions.

The standard shared frailty model was fit using the MCMC algorithm described in the Appendix E of the online supplementary material. For each simulation scenario, generated sample and considered model, we ran the MCMC algorithm for 5,000 iterations after a burn-in period of 1,000 iterations. Standard MCMC tests (not shown), suggested convergence of the chains. The results obtained in the simulation study and reported in the next section were robust to a sensitivity analysis with different prior specifications.

The performance of the models was evaluated by computing the mean squared error (MSE) of the posterior mean of the corresponding parameters. In the case of the estimation of survival ratio, the models where compared by means of the L_∞ and L_1 distance between the posterior

mean and the true value. The competing models were also compared by means of the log pseudo marginal likelihood method (LPML) developed by Geisser and Eddy (1979). A larger value of the LPML indicates that the corresponding model has better predictive ability. The computation of the LPML is given in Appendix F of the supplementary material.

4.2 The results

Tables 1 and 2 show the simulation results for the regression coefficient and Kendall's τ when the data are generated under a σ-stable shared frailty PH model and under the σ-stable conditional partially PH model, respectively. The results suggest that when the assumptions of the shared frailty PH model apply and we fit the proposed model, the parameter σ, the regression coefficients and the association structure, measured via Kendall's τ, are well estimated by their posterior means. Specifically, the estimates of Kendall's τ are not affected by the values of the predictors in a given pair. Furthermore, the posterior mean of Kendall's τ for every combination of the covariates has similar bias and MSE than the estimator arising from the corresponding σ-stable shared frailty PH model. The results also show that the MSE reduces as the number of clusters r increases, suggesting that the posterior mean for σ, β and Kendall's τ for any combination of the covariates, is a consistent estimator of the corresponding parameter.

[Table 1 about here.]

On the other hand, if the conditional partially PH model assumption is the true model, the results suggest that the shared frailty PH model leads to strongly biased estimators of the association structure. Also, the MSE does not get smaller when the sample size increases. As expected, the results under the proposed model suggest that the posterior mean is an unbiased estimator of the association structure for every sample size and that it is consistent.

[Table 2 about here.]

Similar results are observed for the survival ratio. Table S.1, given in Appendix F of the online supplementary material, shows the L_1 and L_∞ distance between the reciprocal of the true survival ratio and the posterior mean of this association functional parameter. When the data are generated from the proposed model, the average difference in L_∞ (L_1) distance between the models, across simulations, combinations of covariate values and simulation settings, was 0.2057 (0.0519). In this case, the estimates of the survival ratio under the proposed model are closer to the true function in all simulation settings. On the other hand, when the shared

frailty PH model is the correct data generating mechanism, the two models behaved in a similar way and the average difference in L_∞ (L_1) distance between the models was -0.0048 (-0.0026). In this scenario, in 24 settings (out of 27) the shared frailty PH model performs better than the proposed model with respect to the L_∞ distance. In terms of L_1 distance, in 2 simulation settings the proposed model performs better and in 6 cases the two models perform equally well. Table 3 displays the results on the behaviour of the model selection criterion. This table shows the mean of the LPML for each model and the percentage of times across simulations in which the LPML selects the true time-to-event model assumption. In agreement with the results discussed for the model parameters, the LPML suggests that there are no differences between the fit of the models when the data are generated from a shared frailty PH model. Also, the LPML correctly selects the proposed model when the shared frailty assumption is not valid, implying an association structure that varies with the predictors. Therefore, the results show that the LPML is an adequate model selection criteria and that the power for selecting the correct regression model assumption is high even for group sample sizes as small as $r = 100$.

[Table 3 about here.]

Finally, plots of the estimated marginal hazard functions for different combinations of the predictors are given in the Appendix F of the online supplementary material. The results show that the posterior mean of the marginal survival function under the proposed model can correctly estimate the true, with a behavior similar to the described for the regression coefficients and association parameters.

5 Application to insurance data

5.1 The data

In recent years, last survivor policies have become quite popular in the actuarial industry. These policies are issued to couples and are structured so that the payoff is due only at the time of death of the second member. In order to fairly price such policies, it is important to adequately take into account the joint survival times of the members of the couples and how these are associated. As a matter of fact, the analysis of several datasets of married couples has identified a significant positive correlation between the survival of the spouses. The possible sources of association have been described by the common lifestyle, the involvement in a common disaster and also the so-called broken-heart syndrome (Youn and Shemyakin, 1999).

We consider a dataset of joint survival times from a Canadian insurance company, described in detail by Luciano et al. (2008). The dataset contains information on $r = 197$ policy contracts, in a 5-year period, from December 29, 1988, to December 31, 1993. Each contract was stipulated by two people for which we know gender and date of birth. Here the response $T_{i,j}$ is defined as the time-to-death, calculated starting from the signing date of the contract, of the ith individual in the jth couple ($n_j = 2$). For each partner, the vector of covariates $\mathbf{z}_{i,j}$ consists of their gender ($z_{i,j,1}$) and age at the moment the contract was signed ($z_{i,j,2}$). As for the gender, we set $z_{i,j,1} = 0$ or 1 to indicate if the individual is a female or a male, respectively.

5.2 Model fit and results

We fit the σ-stable version of the conditional partially PH model by assuming $k(t \mid y) = \mathrm{e}^b \mathbb{1}_{(0,t^a]}(y)$, $P_0(\cdot) = \mathbb{1}_{(0,T]}(\cdot)/T$ and the MCMC algorithm described in the Appendix D of the online supplementary material. We ran the MCMC algorithm for 10,000 iterations after a burn-in period of 2,000 iterations. Standard MCMC tests suggested convergence of the chains (see, Appendix G.2). The model specification was completed with the same hyper-priors described in Section 4. The σ-stable shared frailty PH model, with a Weibull baseline hazard function, was also fit to the data, under the same prior specification described in Section 4. The values for LPML under the proposed model and the σ-stable shared frailty PH model was 609 and 444, respectively, showing that the proposed model outperforms a natural competitor from a predictive point of view. To evaluate the assumption on the relationship of predictors and the time-to-event distribution and the implied association structure we also fit a BNP dependent mixture model, arguably one of the most general models for regression data (see, e.g. Barrientos et al., 2012). Specifically, we considered a linear dependent Dirichlet process (LDDP) mixture of bivariate lognormals model (see, e.g. De Iorio et al., 2004, 2009, Jara et al., 2010). The details of the implemented LDDP model are given in Appendix H of the supplementary material. The LPML under the LDDP model was 534 suggesting that, from a predictive point of view, the additional generality provided by the LDDP model regarding the mean and association structure are not needed for the insurance data.

The results suggest that there is no significant effect of gender on the conditional or marginal risk of death. Further, within a cluster the log relative risk of death (70 years old vs 40 years old) is equal to 1.9570. On the contrary, if we randomly select a 70 years old person and a 40 years old person from the population (i.e., not from the same cluster), then the estimated

log-relative risk of death is 0.4885. The posterior inference on Kendall's τ coefficient shows the existence of different degrees of association depending on the subjects' covariates. Table 4 illustrates the results for Kendall's τ as a function of the age of each member in a male/female couple. The results suggest that the association of survival times is greater for couples of the same age than for couples of different age. For instance, if we assume that the first member is a male and the second a female, the posterior mean (95% credible interval) of Kendall's τ was 0.7029 $(0.6681, 0.7465)$ and 0.7030 $(0.6694, 0.7380)$ if both individuals were 40 and 70 years old, respectively. On the other hand, if the couple consisted of a 40 years old male and a 70 years old female, the posterior mean (95% credible interval) for Kendall's τ was only 0.5822 $(0.4582, 0.6708)$. Similarly, for a 70 years old male and a 40 years old female, the posterior mean (95% credible interval) for the Kendall's τ was 0.6059 $(0.5051, 0.6835)$. The heterogeneity in the association structure as a function of the different covariate values observed in this dataset explains the better predictive performance of our model with respect to the alternative shared frailty PH model.

[Table 4 about here.]

Figure 3 shows estimated marginal survival curves for different combinations of the predictors under the σ-stable conditional partially PH model. The posterior mean of the survival curve is displayed, along with a point-wise 95% credible region. The figure also displays the empirical survival function obtained by aggregating the data for the predictor age at the moment of contract. The results suggest that the fit of the proposed model is adequate.

[Figure 3 about here.]

6 Discussion

We have proposed a class of conditional partially PH models and showed that it has appealing properties in terms of the implied marginal distribution and association structure for the analysis of clustered time-to-event data. More specifically, our proposal accommodates for the presence of different degrees of association depending on subjects' specific covariates, allowing for both a straightforward interpretation of the parameters of the survival model, and a convenient modelling of heterogeneous intra-cluster associations. We have illustrated the performance of the proposed model with a simulation study and an application to the analysis of last survivor

policies, where the model is shown to outperform parametric competitors by virtue of its flexible covariate-dependent modelling of the association structure in a cluster.

While in our illustrations we worked with a model specification corresponding to an assumption of increasing hazards, the proposed model is more general and, according to the modelling goals, other choices can be made. For instance, by reversing the inequality in the kernel proposed by Dykstra and Laud (1981), that is by choosing $k(t \mid y) = \mathrm{e}^b \mathbb{1}_{[t^a,\infty)}(y)$ in (3), we obtain a family of decreasing random hazards. Other tractable examples are discussed in Lo and Weng (1989): $k(t \mid y) = \mathrm{e}^b \mathbb{1}_{\{|t-t_0|\geq y\}}$ and $k(t \mid y) = \mathrm{e}^b \mathbb{1}_{\{|t-t_0|\leq y\}}$, for some $t_0 > 0$, lead respectively to a class of U-shaped symmetric hazards with minimum in t_0 and a class of unimodal symmetric hazards centered at t_0. The class of completely monotone hazards is instead recovered by choosing $k(t \mid y) = \mathrm{e}^{b-ty}$. If the goal is to estimate hazard functions without shape restriction, then one might consider more flexible kernels, such as the log-normal, although at the cost of losing some of the analytical tractability of the examples considered above. Based on our experiments, the choice of the normalized base measure P_0 has not a major impact on the produced inference. For instance, Appendix G.3 shows comparable inferences are drawn under the proposed model if an exponential base measure is considered for fitting the insurance data. Thus, it might be convenient to choose the base measure so to favour analytical tractability. We pursued this purpose by working with a uniform base measure on some interval $[0, T]$, but other specifications can be adopted. For example Lo and Weng (1989) suggest to choose P_0 depending on $k(\cdot \mid \cdot)$, with an argument analogous to the one of conjugate priors.

Future extensions of the proposed modelling framework might allow for borrowing of strength across clusters by enabling shared components in the modelling of the CRMs μ_j. For example, dependent priors for hazard functions could be introduced as in Lijoi and Nipoti (2014), so that the cluster-specific random hazard functions μ_j could be modelled as a mixture, $\mu_j = \tilde{\mu}_0 + \tilde{\mu}_j$, where $\tilde{\mu}_0$ denotes a common CRM, shared by all clusters, and $\tilde{\mu}_j$ is a cluster-specific idiosyncratic CRM. The two measures $\tilde{\mu}_0$ and $\tilde{\mu}_j$ are characterized, respectively, by Lévy intensities $\nu_{\vartheta,0}(\mathrm{d}s, \mathrm{d}x) = \epsilon\, \nu_\vartheta(\mathrm{d}s, \mathrm{d}x)$ and $\nu_{\vartheta,j}(\mathrm{d}s, \mathrm{d}x) = (1-\epsilon)\, \nu_\vartheta(\mathrm{d}s, \mathrm{d}x)$ with $\epsilon \in [0, 1]$ a parameter governing the amount of borrowing allowed across clusters. Since the marginal CRMs μ_j's are identically distributed with Lévy intensity $\nu_\vartheta(\mathrm{d}s, \mathrm{d}x)$, the resulting shared CRM model retains the marginal properties and the parameters' interpretation of the class of conditional partially PH models we have described in this manuscript. An alternative way to induce borrowing of information across clusters can be obtained by adopting a hierarchical approach, that is modelling the base measure α with an almost surely discrete nonparametric prior, in the spirit of

Camerlenghi et al. (2018).

7 Supporting information

A web Appendix which contains proofs, details on posterior sampling and complementary results on the study of simulated and real data, is available with this paper.

Aknowledgment

We wish to thank the Society of Actuaries, through the courtesy of Edward (Jed) Frees and Emiliano Valdez, for allowing us to use the Insurance dataset in this paper. A. Jara's research is supported by grant 1141193 from the Fondo Nacional de Desarrollo Científico y Tecnológico (FONDECYT), from the Chilean Government. Part of this work was performed during a visit of B. Nipoti to Pontificia Universidad Católica de Chile, in the framework of FONDECYT 1141193 grant.

References

Anderson, J. E., Louis, T. A., Holm, N. V., and Harvald, B. (1992). Time-dependent association measures for bivariate survival distributions. *Journal of the American Statistical Association*, 87(419):641–650.

Arbel, J., Lijoi, A., and Nipoti, B. (2016). Full bayesian inference with hazard mixture models. *Computational Statistics & Data Analysis*, 93:359–372.

Barrientos, A. F., Jara, A., and Quintana, F. A. (2012). On the support of MacEachern's dependent Drichlet processes and extensions. *Bayesian Analysis*, 7:277– 310.

Brix, A. (1999). Generalized gamma measures and shot-noise Cox processes. *Advances in Applied Probability*, 31:929–953.

Camerlenghi, F., Lijoi, A., and Prünster, I. (2018). Bayesian survival analysis with hierarchies of nonparametric priors. *Tech. report*.

Carlin, B. P. and Hodges, J. S. (1999). Hierarchical proportional hazards regression models for highly stratified data. *Biometrics*, 55:1162–1170.

Choi, S. and Huang, X. (2012). A general class of semiparametric transformation frailty models for nonproportional hazards survival data. *Biometrics*, 68(4):1126–1135.

Clayton, D. G. (1978). A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika*, 65(1):141–151.

Cox, D. (1972). Regression models and life tables (with discussion). *J. Roy. Statist. Soc. Ser. A*, 34:187–202.

De Blasi, P., Peccati, G., and Prünster, I. (2009). Asymptotics for posterior hazards. *The Annals of Statistics*, 37(4):1906–1945.

De Iorio, M., Johnson, W. O., Müller, P., and Rosner, G. L. (2009). Bayesian nonparametric non-proportional hazards survival modelling. *Biometrics*, 65:762–771.

De Iorio, M., Müller, P., Rosner, G. L., and MacEachern, S. N. (2004). An ANOVA model for dependent random measures. *Journal of the American Statistical Association*, 99:205–215.

Duchateau, L. and Janssen, P. (2008). *The frailty model*. Springer.

Dykstra, R. L. and Laud, P. (1981). A Bayesian nonparametric approach to reliability. *The Annals of Statistics*, 9:356–367.

Fine, J. P., Glidden, D. V., and Lee, K. E. (2003). A simple estimator for a shared frailty regression model. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 65(1):317–329.

Geisser, S. and Eddy, W. F. (1979). A predictive approach to model selection. *Journal of the American Statistical Association*, 74(365):153–160.

Gelfand, A. E. and Mallick, B. K. (1995). Bayesian analysis of proportional hazards models built from monotone functions. *Biometrics*, 51:843–852.

Hjort, N. L. (1990). Nonparametric Bayes estimators based on beta processes in models for life history data. *The Annals of Statistics*, 18:1259–1294.

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

Ibrahim, J. G., Chen, M. H., and Sinha, D. (2001). *Bayesian Survival Analysis*. Springer-Verlag.

James, L. (2005). Bayesian Poisson process partition calculus with an application to Bayesian Lévy moving averages. *Ann. Statist.*, 33:1771–1799.

Jara, A., Lesaffre, E., De Iorio, M., and Quintana, F. A. (2010). Bayesian semiparametric inference for multivariate doubly-interval-censored data. *The Annals of Applied Statistics*, 4:2126–2149.

Kalbfleisch, J. D. (1978). Nonparametric Bayesian analysis of survival time data. *Journal of the Royal Statistical Society, Series B: Methodological*, 40:214–221.

Kingman, J. (1967). Completely random measures. *Pacific J. Math.*, 21:59–78.

Kneib, T. and Fahrmeir, L. (2007). A mixed model approach for geoadditive hazard regression. *Scandinavian Journal of Statistics*, 34(1):207–228.

Lijoi, A. and Nipoti, B. (2014). A class of hazard rate mixtures for combining survival data from different experiments. *Journal of the American Statistical Association*, 109(506):802–814.

Lijoi, A. and Prünster, I. (2010). Models beyond the Dirichlet process. In Hjort, N., Holmes, C., Müller, P., and Walker, S., editors, *Bayesian Nonparametrics*, pages 80–136. Cambridge University Press, Cambridge.

Liu, D., Kalbfleisch, J. D., and Schaubel, D. E. (2011). A positive stable frailty model for clustered failure time data with covariate-dependent frailty. *Biometrics*, 67(1):8–17.

Lo, A. and Weng, C. (1989). On a class of Bayesian nonparametric estimates. II. Hazard rate estimates. *Ann. Inst. Statist. Math.*, 41:227–245.

Luciano, E., Spreeuw, J., and Vigna, E. (2008). Modelling stochastic mortality for dependent lives. *Insurance: Mathematics and Economics*, 43(2):234–244.

Manatunga, A. K. and Oakes, D. (1999). Parametric analysis for matched pair survival data. *Lifetime Data Analysis*, 5(4):371–387.

Müller, P., Quintana, F. A., Jara, A., and Hanson, T. E. (2015). *Bayesian Nonparametric Data Analysis*. Springer, New York, USA.

Nieto-Barajas, L. E. and Walker, S. G. (2004). Bayesian nonparametric survival analysis via Lévy driven markov processes. *Statistica Sinica*, 14(4):1127–1146.

Orbanz, P. and Williamson, S. (2011). Unit–rate poisson representations of completely random measures. Technical report, Technical report.

Pennell, M. L. and Dunson, D. B. (2006). Bayesian semiparametric dynamic frailty models for multiple event time data. *Biometrics*, 62(4):1044–1052.

Sinha, D. and Dey, D. K. (1997). Semiparametric Bayesian analysis of survival data. *Journal of the American Statistical Association*, 92:1195–1212.

Youn, H. and Shemyakin, A. (1999). Statistical aspects of joint life insurance pricing. *1999 Proceedings of the Business and Statistics Section of the American Statistical Association*, 34:38.

Zhou, H., Hanson, T., Jara, A., and Zhang, J. (2015). Modelling county level breast cancer survival data using a covariate-adjusted frailty proportional hazards model. *The annals of applied statistics*, 9(1):43.

Address of the corresponding author

Bernardo Nipoti

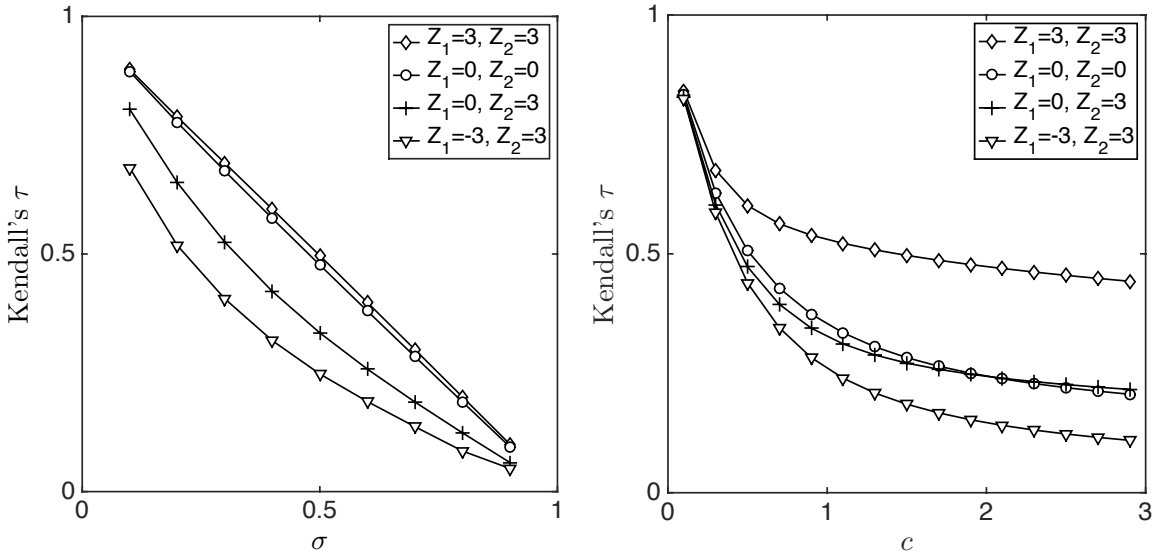
Lloyd Institute

College Green, Dublin 2

Dublin

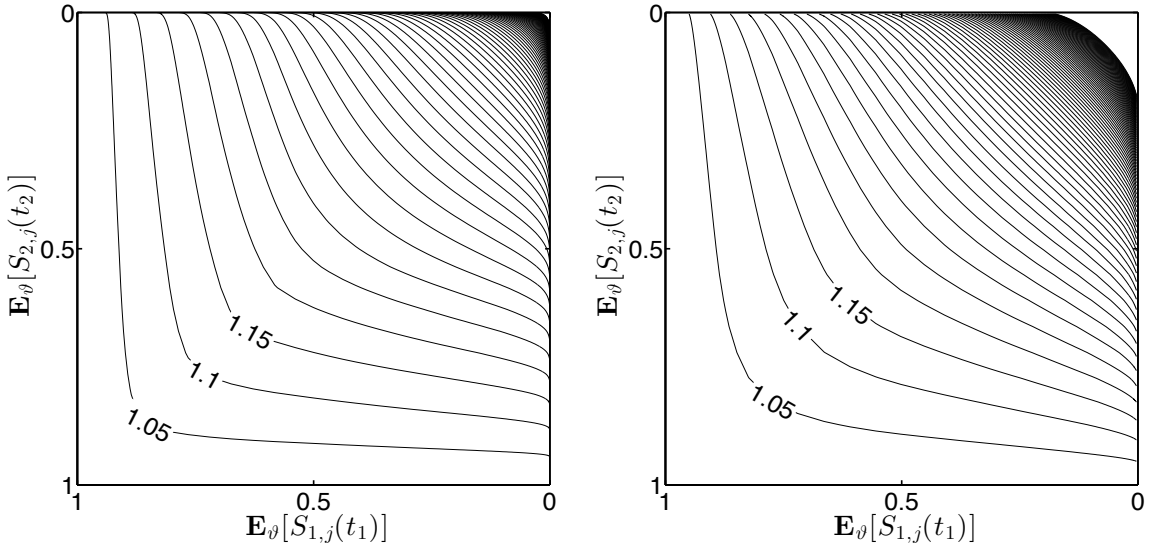
Ireland

nipotib@tcd.ie



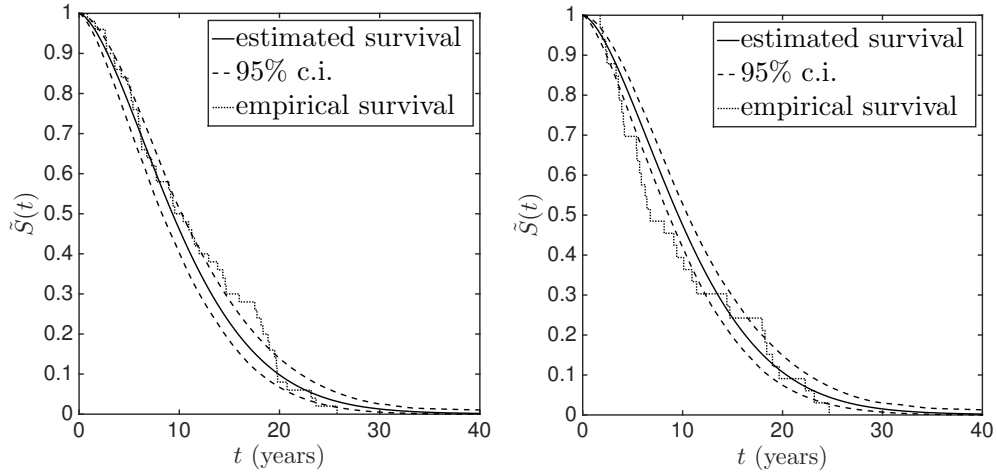
(a) σ-stable (b) gamma

Figure 1: Kendall's τ plotted as a function of the parameters σ and c for the σ-stable and gamma conditional PH model, respectively, for different sets of covariates values. See Section 3.2 for details.



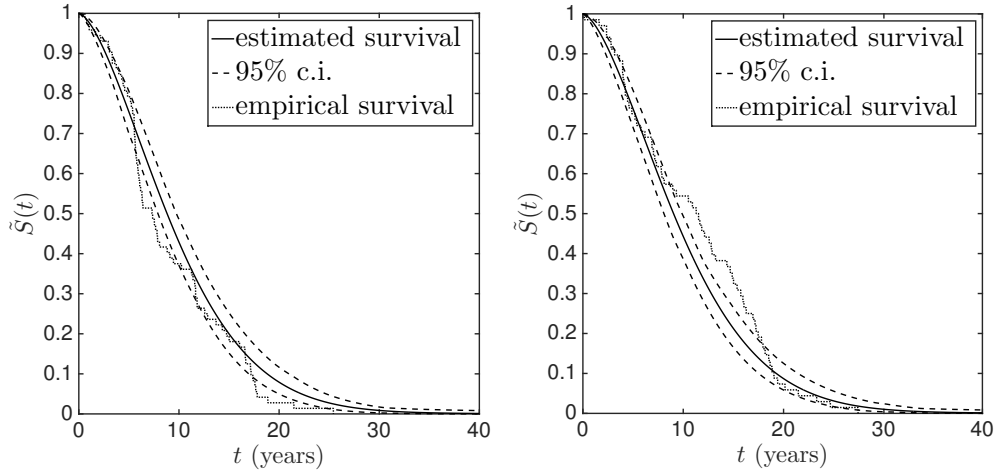
(a) σ-stable (b) gamma

Figure 2: Contour plots of the survival ratio $\Sigma_{1,2,j}(t_1, t_2)$ under the σ-stable and gamma version of the model.



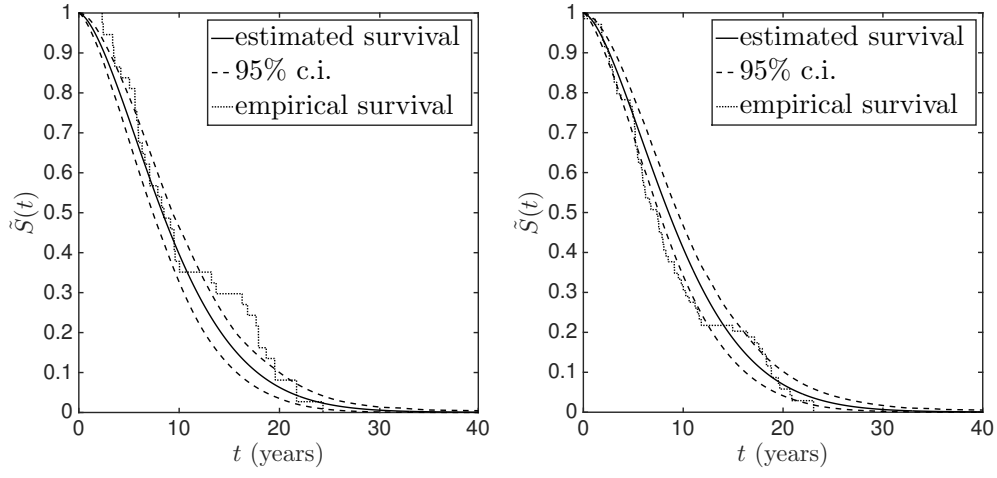
(a) 60-64 years old women

(b) 60-64 years old men



(c) 65-69 years old women

(d) 65-69 years old men



(e) 70-74 years old women

(f) 70-74 years old men

Figure 3: Insurance data: Posterior mean of the survival function for different combinations of the predictors (continuous line). A point-wise 95% credible region is also displayed as dotted lines. In each case, the empirical survival function obtained by aggregating the data with respect to the predictor age at the moment of contract is also displayed; the intervals considered for the aggregation of the data are indicated in each case. The result under the proposed model corresponds to the midpoint of the corresponding age interval.

Table 1: Simulated Data: Data generated from a σ-stable shared frailty PH model. Mean across simulations (mean squared error) of the posterior mean for σ, regression coefficient and Kendall's τ under the σ-stable version of the proposed conditional partially PH model and under the σ-stable shared frailty PH model. The results for Kendall's τ under the proposed model are shown for the three different combinations of covariates, where $\tau_{1,1}$, $\tau_{0,0}$ and $\tau_{1,0}$ corresponds to Kendall's τ for pairs of individuals with covariates $(1,1)$, $(0,0)$ and $(1,0)$ (or equivalently $(0,1)$), respectively.

True value				Conditional partially PH model					shared frailty PH model		
r	σ	β	τ	σ	β	$\tau_{1,1}$	$\tau_{0,0}$	$\tau_{1,0}$	σ	β	τ
100	0.25	4	0.75	0.2785	3.6805	0.7205	0.7209	0.7206	0.2485	4.1688	0.7515
				(0.0012)	(0.1389)	(0.0013)	(0.0013)	(0.0013)	(0.0004)	(0.1353)	(0.0004)
	0.50	4	0.50	0.5371	3.6705	0.4644	0.4656	0.4585	0.4874	4.0397	0.5026
				(0.0027)	(0.1658)	(0.0026)	(0.0025)	(0.0031)	(0.0016)	(0.0990)	(0.0016)
	0.75	4	0.25	0.7747	3.7951	0.2588	0.2586	0.2498	0.7338	4.1310	0.2662
				(0.0031)	(0.1094)	(0.0017)	(0.0016)	(0.0013)	(0.0030)	(0.1362)	(0.0030)
200	0.25	4	0.75	0.2751	3.6382	0.7246	0.7254	0.7246	0.2490	4.0582	0.7510
				(0.0008)	(0.1494)	(0.0009)	(0.0009)	(0.0009)	(0.0003)	(0.0709)	(0.0003)
	0.50	4	0.50	0.5343	3.6738	0.4660	0.4668	0.4635	0.4958	4.0354	0.5042
				(0.0017)	(0.1341)	(0.0017)	(0.0017)	(0.0009)	(0.0007)	(0.0456)	(0.0007)
	0.75	4	0.25	0.7782	3.7795	0.2550	0.2562	0.2488	0.7432	4.0621	0.2568
				(0.0022)	(0.1062)	(0.0010)	(0.0011)	(0.0007)	(0.0015)	(0.0588)	(0.0015)
500	0.25	4	0.75	0.2741	3.6300	0.7253	0.7255	0.7263	0.2510	3.9948	0.7490
				(0.0006)	(0.1456)	(0.0007)	(0.0007)	(0.0006)	(0.0001)	(0.0300)	(0.0001)
	0.50	4	0.50	0.5401	3.6525	0.4637	0.4640	0.4684	0.4641	3.9848	0.4960
				(0.00016)	(0.1309)	(0.0016)	(0.0016)	(0.0017)	(0.0004)	(0.0271)	(0.0004)
	0.75	4	0.25	0.7757	3.7925	0.2532	0.2520	0.2443	0.7473	4.0173	0.2527
				(0.0019)	(0.0908)	(0.0008)	(0.0008)	(0.0004)	(0.0008)	(0.0336)	(0.0008)

Table 2: Simulated Data: Data generated from a σ-stable conditional partially PH model. Mean across simulations (mean squared error) of the posterior mean for the parameter σ, the regression coefficient and Kendall's τ under the σ-stable version of the proposed conditional partially PH model and under the σ-stable shared frailty PH model. The results are shown for the three different combination of covariates, where $\tau_{1,1}$, $\tau_{0,0}$ and $\tau_{1,0}$ corresponds to Kendall's τ for pairs of individuals with covariates $(1,1)$, $(0,0)$ and $(1,0)$ (or equivalently $(0,1)$), respectively. The results for the mean squared error under the shared frailty model are shown for $\tau_{1,1}$, $\tau_{0,0}$ and $\tau_{1,0}$, respectively.

	True values					Conditional partially PH model					shared frailty PH model		
r	σ	β	$\tau_{1,1}$	$\tau_{0,0}$	$\tau_{1,0}$	σ	β	$\tau_{1,1}$	$\tau_{0,0}$	$\tau_{1,0}$	σ	β	τ
100	0.25	4	0.6988	0.7166	0.4978	0.2536	3.9626	0.7033	0.7214	0.4980	0.3065	2.0242	0.6932
						(0.0004)	(0.1348)	(0.0005)	(0.0004)	(0.0011)	(0.0054)	(3.9773)	(0.0024; 0.0029; 0.0406)
	0.50	4	0.4512	0.4704	0.2432	0.5072	4.0060	0.4537	0.4723	0.2418	0.5566	2.8378	0.4434
						(0.0027)	(0.1662)	(0.0021)	(0.0024)	(0.0011)	(0.0131)	(1.4087)	(0.0100; 0.0106; 0.0500)
	0.75	4	0.2498	0.2503	0.1155	0.7264	3.9614	0.2904	0.2503	0.1371	0.7526	3.3278	0.2474
						(0.0048)	(0.2142)	(0.0036)	(0.0036)	(0.0009)	(0.0067)	(0.5508)	(0.0067; 0.0067; 0.0241)
200	0.25	4	0.6988	0.7166	0.4978	0.2545	3.9626	0.6644	0.6822	0.4958	0.3186	2.2807	0.6814
						(0.0005)	(0.1362)	(0.0016)	(0.0015)	(0.0007)	(0.0061)	(2.9963)	(0.0017; 0.0027; 0.0351)
	0.50	4	0.4512	0.4704	0.2432	0.5117	3.9192	0.4394	0.4523	0.2605	0.5791	3.0721	0.4209
						(0.0024)	(0.1709)	(0.0013)	(0.0014)	(0.0008)	(0.0090)	(0.9248)	(0.0037; 0.0052; 0.0344)
	0.75	4	0.2498	0.2503	0.1155	0.7274	4.0300	0.2790	0.2707	0.1469	0.7769	3.4714	0.2231
						(0.0043)	(0.2225)	(0.0019)	(0.0017)	(0.0012)	(0.0037)	(0.3303)	(0.0036; 0.0037; 0.0145)
500	0.25	4	0.6988	0.7166	0.4978	0.2515	3.9838	0.7054	0.7238	0.4973	0.2860	1.9973	0.7140
						(0.0001)	(0.0327)	(0.0003)	(0.0002)	(0.0004)	(0.0018)	(4.0308)	(0.0007; 0.0005; 0.0473)
	0.50	4	0.4512	0.4704	0.2432	0.4984	3.9788	0.4607	0.4807	0.2491	0.5240	2.8738	0.4760
						(0.0005)	(0.0374)	(0.0006)	(0.0006)	(0.0003)	(0.0029)	(1.2880)	(0.0029; 0.0023; 0.0564)
	0.75	4	0.2498	0.2503	0.1155	0.7279	3.9737	0.2894	0.2857	0.1359	0.7458	3.3011	0.2542
						(0.0014)	(0.0308)	(0.0021)	(0.0018)	(0.0005)	(0.0020)	(0.5127)	(0.0020; 0.0019; 0.0212)

Table 3: Simulated Data: Mean of LPML for the conditional partially PH model (shared frailty PH model) and percentage in which LPML selects the conditional partially PH model (shared frailty PH model), across simulation. The results are shown for the different simulation settings and true underlying time-to-event regression model assumption.

Simulation Setting		True Model			
		Conditional partially PH model		shared frailty PH model	
r	σ	Mean	%	Mean	%
100	0.25	-152 (-375)	100 (0)	302(303)	0 (1)
	0.50	-122 (-205)	100 (0)	175 (175)	1 (1)
	0.75	-110 (-151)	100 (0)	162 (163)	0 (0)
200	0.25	-396 (-618)	100 (0)	605 (606)	0 (4)
	0.50	-237 (-316)	100 (0)	347 (348)	0 (3)
	0.75	-146 (-177)	100 (0)	324 (324)	1 (0)
500	0.25	-728 (-1855)	100 (0)	1532 (1534)	0 (8)
	0.50	-624 (-1053)	100 (0)	870 (872)	0 (2)
	0.75	-542 (-756)	100 (0)	787 (788)	0 (1)

Table 4: Insurance data: Posterior mean (95 % credible interval) for Kendall's τ under the proposed σ-stable conditional partially PH model. The results are presented for different combinations of gender and age for the members of the cluster.

Subject 1		Subject 2		
Gender	Age (Years)	Gender	Age (Years)	τ
Male	40	Male	40	0.7039 (0.6615, 0.7428)
Male	40	Female	40	0.7029 (0.6681, 0.7465)
Female	40	Female	40	0.7049 (0.6748, 0.7418)
Male	70	Male	70	0.7036 (0.6715, 0.7423)
Male	70	Female	70	0.7030 (0.6694, 0.7380)
Female	70	Female	70	0.7038 (0.6735, 0.7379)
Male	40	Male	70	0.5935 (0.4810, 0.6769)
Male	40	Female	70	0.5822 (0.4582, 0.6708)
Female	40	Male	70	0.6059 (0.5051, 0.6835)
Female	40	Female	70	0.5947 (0.4778, 0.6790)