ARC Centre of Excellence in Population Ageing Research

Working Paper 2023/08

An Augmented Variable Dirichlet Process Mixture model for the analysis of dependent lifetimes

Francesco Ungolo

University of New South Wales

An Augmented Variable Dirichlet Process Mixture model for the analysis of dependent lifetimes

Francesco Ungolo[1,2,*]

[1]*School of Risk and Actuarial Studies, University of New South Wales, Kensington NSW 2052, Australia*
[2]*ARC Centre of Excellence in Population Ageing Research, University of New South Wales, Kensington NSW 2052, Australia*
[*]*Corresponding author: Francesco Ungolo, f.ungolo@unsw.edu.au*

April 25, 2023

Insurance and annuity products issued on multiple lives require the use of statistical models which account for lifetime dependence. This work presents a Dirichlet Process Mixture-based approach which allows to model dependent lifetimes within a group, such as married couples, accounting for individual as well as group-specific covariates. The approach allows to account for right censoring and left truncation as typical of survival analysis. The model is analysed in a fully Bayesian setting and illustrated to jointly model the lifetime of male-female couples in a portfolio of joint and last survivor annuities of a Canadian life insurer. The model shows an improved in-sample and out-of-sample performance compared to traditional approaches assuming independent lifetimes, and offers additional insights on determinants of the dependence between lifetimes and on their impact on joint and last survivor annuity prices.

Keywords: Dependent lifetimes, Survival Analysis, Dirichlet Processes, Bayesian analysis, Life insurance, MCMC, Mixture models

1. Introduction

The pricing of insurance products issued on multiple lives, such as couple members, requires the use of statistical models which can best predict their future lifetimes. The independence assumption can sensibly reduce the model complexity and ease the implementation of computational routines for pricing. However, this assumption is not tenable in practice. For example, partners are likely to share the same socio-economic characteristics, such that they share the same living standards, and to be exposed to similar risks (Denuit & Cornet (1999), Denuit et al. (2001)).

Furthermore, the use of the simplistic independence assumption can have a material impact on actuarial valuations. Denuit & Cornet (1999) use a Markov model with force of mortality dependent on marital status, and show how the premium of a widow pension annuity is 10 per cent lower compared to the independence case. Using a copula model, Frees et al. (1996) first demonstrate the presence of a positive dependence between husband and wife lifetimes, and show that the annuity value is 5 per cent lower compared to the case of independent lifetimes.

A wealth of approaches have been proposed for the analysis of dependent lifetimes, especially in the biostatistical and in the actuarial field. Copula models are among the most employed approaches for this analysis: Frees et al. (1996) focus on a one-parameter Frank copula with Gompertz marginals for the analysis of the male and female lifetimes within a couple, Carriere (2000) extends this analysis by considering other marginals as well as other types of copulas, while Deresa et al. (2022) focus on the statistical properties of copula models in presence of left-truncation and dependent censoring. Youn & Shemyakin (1999) were the first to account for covariates when modelling dependence. Using a Gumbel copula with Weibull marginals , they account for the age-difference between the spouses, which is found influence the lifetime dependence. In addition, using Gompertz marginals for the time to event for males and females couple members, Dufresne et al. (2018) observe how the gender of the eldest partner has also an influence on the lifetime dependence.

An alternative approach is given by models using random effects (or frailty components, see Vaupel et al. (1979)) to capture the dependence beteween lifetimes. This means that conditional on a latent variable, then lifetimes are independently distributed. For example, Yashin & Iachine (1995) develop a correlated gamma frailty model for the analysis of the joint lifetime of Danish twins. In the field of biostatistics, a closely related problem is given by modelling dependent time to event and time to censoring. Huang & Wolfe (2002) address this problem by assuming that the two random variables have a distribution characterized by the Cox proportional hazard model, whose linear term includes a normally distributed log-frailty component. Gorfine & Hsu (2011) consider other parametric functions for the distribution of the individual frailty.

The common limitation of the aforementioned copula and random effect models is the need of assuming a specific parametric form for the copula, or of the distribution of the random effects. With reference to the latter approach, Ungolo & van den Heuvel (2022) and Ungolo & van den Heuvel (2023) overcome the potential misspecification issue by using a multivariate random effect with a discrete distribution and unknown number of

levels.None of these approaches account for covariates in the distribution of the random effects used to explain the dependence among time to events.

This paper contributes to the literature by proposing the Augmented Variable Dirichlet Process Mixture (AVDPM) model, which briefly consists of a joint probability distribution of the time to events and of the group-specific covariates, where all these variables are independently distributed, conditional on a multivariate random effect, whose distribution is drawn from a Dirichlet Process. In this way, we can flexibly account for the lifetime dependence among units within a group, and at the same time we can account for those common covariates which capture the dependence between lifetimes. In addition, we show how this approach can easily account for right censoring and left truncation. We analyse the resulting model by means of a fully Bayesian analysis, which may include the information available to the researcher.

This paper is organized as follows: Section 2 briefly introduces the Dirichlet Process and the Dirichlet Process Mixture model, and Section 3 presents the Augmented Variable Dirichlet Process Mixture (AVDPM) model for the analysis of dependent lifetimes. Section 4 describes the empirical dataset used for illustrating the model and the additional parametric features for the joint lifetime of male-female couples and for the couple-specific covariates. Section 6 presents the results of the empirical analysis, and Section 7 shows how the model can be used when pricing joint life and last survivor annuities. Section 8 extends the AVDPM to the analysis of more general groups of dependent lifetimes, and Section 9 concludes.

2. Dirichlet Processes and Dirichlet Process Mixtures

The Dirichlet Process (DP) was first introduced by Ferguson (1973) to specify a prior distribution over probability distributions. A random draw $G \sim \mathrm{DP}(G_0, \phi)$ yields a discrete probability distribution over a countably infinite number of points drawn independently from a *base distribution* denoted as G_0, which indexes the DP together with the *concentration parameter* ϕ. This latter captures the degree of shrinkage of G towards G_0, or in other words, the strength of the prior assumption G_0 over G, analogous to the prior assumption about the parameters of a probability distribution in Bayesian statistics.

Suppose we need G to draw samples for a random variable $\gamma \sim G(\cdot)$. Sethuraman (1994) outlines the following construction by means of a mixture distribution with an infinite number of components, indexed by k:

$$G(\gamma) = \sum_{k=1}^{\infty} \pi_k \delta_{\gamma_k}(\gamma), \tag{2.1}$$

where $\gamma_k \overset{iid}{\sim} G_0$ for $k = 1, 2, \ldots$, and $\delta_{\gamma_k}(\cdot)$ is the Dirac measure which assigns unitary mass if $\gamma = \gamma_k$ and zero otherwise. The mixture weights π_k are randomly generated using the so called *stick breaking procedure* (SBP), which rescales a set of i.i.d. random

variables $\psi_k \sim \text{Beta}(1, \phi)$ as follows:

$$\pi_k := \pi_k(\psi_{1:k-1}) = \psi_k \prod_{j=1}^{k-1} (1 - \psi_j). \tag{2.2}$$

where $\psi_{1:k} = (\psi_1, \ldots, \psi_k)$. The SBP definition follows from the decreasing size of the mixture weights as the index k increases. From this characterization we can observe how a random draw from a Dirichlet Process yields a discrete distribution over a countably infinite number of atoms from G_0.

In this paper, we are interested in the probability distribution of a random variable T_i, eventually multivariate, for the ith unit, whose density $f(\cdot; \beta, \gamma_i)$ is indexed by a *global* parameter vector β, common to all units, and by a unit-specific parameter vector γ_i. Unit-specific parameters, or random effects in biostatistics, are introduced in a probability model in order to characterize the heterogeneity among the units, hence to capture the dependence relationship among these.

By convoluting the probability distribution of T_i with the probability distribution of the latent parameter γ_i, we obtain a Dirichlet Process Mixture (DPM) model (Lo 1984):

$$f(t; \beta, G) = \int_{\Omega\gamma} f(t; \beta, \gamma) \, \mathrm{d}G(\gamma) = \sum_{k=1}^{\infty} \pi_k f(t; \beta, \gamma_k). \tag{2.3}$$

where $\Omega\gamma$ denotes the sample space of γ.

We model G as a random draw from a DP to allow for a more flexible distribution of T_i, which can capture complex features in the data, such as fat tails and multimodality, as opposed to any parametric assumptions, such as the Normal distribution. On the other hand, despite the infinitely many parameters of this construction enhancing the flexibility of a parametric model, a DP allows for a degree of regularization of G towards a simple parametric form G_0 through the concentration parameter ϕ.

From another perspective, a discrete distribution for $\gamma_i \sim G$ is akin to the creation of ties among the units, which can configure clusters of observations with the same values of γ. Let $s_i = k$ to indicate that the ith unit belongs to the kth cluster characterized by the parameter γ_k^*. We use the superscript * to distinguish the values of γ represented in a sample of n units from the infinitely many samples from G_0. The clustering procedure, tuned by the parameter ϕ, allows to sequentially group the observations through a sampling process known in the literature as the *Chinese restaurant process* (see Heinz (2014) for an illustration). In words, as we keep observing the units in sequence, these are more likely to be in a certain class with probability which depend on the number of units already therein, or to belong to a newly created class with probability which depends on ϕ (see Blackwell & MacQueen (1973) for a characterization in terms of the Polya urn distribution).

The use of Dirichlet Processes can also be seen as a way to infer the number of components of a mixture distribution, as opposed to strategies based on appropriate model selection criteria (see Ungolo & van den Heuvel (2022) for a discussion). The advantage of the DP specification is to avoid the specification and the inference of several models, which may be time consuming.

Summing up, the DPM model assumes the following data generating process for a sample $t_1, \ldots, t_n$:

$$\begin{aligned} &\pi_k \mid \phi \sim \text{SBP}(\phi), \quad k = 1, 2, \ldots; \\ &\gamma_k^* \mid G_0 \sim G_0, \quad k = 1, 2, \ldots; \\ &s_i \mid \pi_1, \pi_2, \ldots \sim \text{Discrete}(\pi_1, \pi_2, \ldots), \quad i = 1, \ldots, n; \\ &t_i \mid \beta, \gamma_{s_i}^* \sim f\left(t_i; \beta, \gamma_{s_i}^*\right) \quad i = 1, \ldots, n. \end{aligned} \tag{2.4}$$

3. The Augmented Variable DPM model

This section introduces the Augmented Variable Dirichlet Process Mixture (AVDPM) model for the analysis of non-exchangeable joint lifetimes of dependent individuals[1] within a group, where individual as well as group-specific covariates are available. For example, husband and wife lifetimes are likely to be positively associated (Denuit et al. 2001), since they share the same living conditions (e.g. diet, socio-economic factors), are exposed to similar risks (e.g. during a catastrophic event they are likely to be in the same place), or eventually subject to the broken-heart syndrome (Parkes et al. 1969). Other examples include the joint lifetime of the primary and secondary head of an insurance policy, families with husband, wife and one child, and so on. That is, groups of the same size. Without loss of generality, we describe the framework in the context of a model for the joint lifetime of male-female couples. In Section 8 we discuss how to extend the framework to more general cases of different size groups of exchangeable lifetimes.

Let T_1 and T_2 denote the random future lifetime of husband and wife respectively, with individual-specific vector of characteristics X_1 and X_2 (e.g. age, medical status), and couple-specific covariates vector Z, which may include for example household income and geo-demographic profile (an indicator of the socio-economic status, see Ungolo et al. (2019)). Therefore, Z may include any characteristics which may explain the dependence in the couples' lifetime. For example, the household income can explain the heterogeneity in the longevity profile of each couple, since a wealthy couple can access better healthcare services all else being equal, compared to a deprived one. Similarly the geo-demographic profile, capturing the effect of the area where the couple lives (e.g. urban or rural), can also be a proxy for the socio-economic characteristics of a couple.

For the ith couple we assume that conditional on a couple-specific bivariate random effect $(\gamma_{1,i}, \gamma_{2,i})$, then $T_{1,i}$ and $T_{2,i}$ are independently distributed. This is for example the approach followed in Ungolo & van den Heuvel (2022) when analysing a joint model for informative censoring and by Ungolo & van den Heuvel (2023) where the authors develop a joint model for the time to competing risk events.

The AVDPM model "augments" the joint probability distribution of (T_1, T_2) by spec-

[1] A multivariate J-dimensional random variable $\mathbf{X} = (X_1, X_2, \ldots, X_J)$ is exchangeable if its joint distribution does not change when altering the subscript of each element of $\mathbf{X}$, or in other words, we swap their order.

ifying a joint probability model for (T_1, T_2, Z):

$$
\begin{aligned}
&f\left(t_{1,i}, t_{2,i}, z_i \mid x_{1,i}, x_{2,i}; \beta_1, \beta_2, G\right) \\
&= \int_{\Omega_{\gamma,\zeta}} \left[\prod_{j=1}^{J} f\left(t_{j,i} \mid x_{j,i}; \beta_j, \gamma_{j,i}\right)\right] f\left(z_i; \zeta_i\right) \mathrm{d}G\left(\gamma_{1,i}, \gamma_{2,i}, \zeta_i\right)
\end{aligned} \tag{3.1}
$$

where β_j is the parameter specific to the future lifetime density of the husband ($j = 1$) or of the wife ($j = 2$), ζ_i is the couple-specific parameter vector indexing the distribution of Z (which can also include global parameters, i.e. not couple-specific) and $\Omega_{\gamma,\zeta}$ denotes the sample space of $(\gamma_1, \gamma_2, \zeta)$. We generically denote by f the probability density function of continuous variables, and the mass function of the discrete ones. Hence, we model the multivariate distribution G as a random draw from a Dirichlet Process, which we write $G \sim \mathrm{DP}\left(G; G_0, \phi\right)$. For simplicity, we assume $G_0 = G_{0,\gamma} \times G_{0,\zeta}$. A further simplification can be to specify $G_{0,\gamma} = G_{0,\gamma_1} \times G_{0,\gamma_2}$, leaving the distribution of $(\gamma_1, \gamma_2, \zeta)$ fully tuned by the concentration parameter ϕ. However, the former approach allows for a prior specification of a dependence relationship among γ_1 and γ_2 as more reasonable in actuarial practice.

The joint density of equation (3.1) allows to capture the dependence between T_1 and T_2 and between (T_1, T_2) and Z through the joint distribution of $(\gamma_1, \gamma_2, \zeta)$. In this way, we distinguish between the variables which directly affect the individual lifetimes X from those which are couple-specific. For example, in the analysis of couples' lifetimes, Deresa et al. (2022) include the couple specific covariates, such as the age difference, within the set of individual covariates.

As in equation (2.3), we can rewrite the density in (3.1) as a mixture distribution with an infinite number of components, obtaining a DPM:

$$
\begin{aligned}
&f\left(t_{1,i}, t_{2,i}, z_i \mid x_{1,i}, x_{2,i}; \beta_1, \beta_2, G\right) \\
&= \sum_{k=1}^{\infty} \pi_k \left[\prod_{j=1}^{2} f\left(t_j \mid x_{j,i}; \beta_j, \gamma_{j,k}^{*}\right)\right] f\left(z_i; \zeta_k^{*}\right)
\end{aligned} \tag{3.2}
$$

where π_k $(k = 1, 2, \ldots)$ is the mixture weight characterized by the stick-breaking procedure described in Section 2, and the superscript * denotes the unique values of γ_j and ζ.

The data generating process of the data can be summarized as follows:

$$
\begin{aligned}
&G \mid \phi, G_{0,\gamma}, G_{0,\zeta} \sim \mathrm{DP}\left(G; \phi, G_0\right) && && \\
&\left(\gamma_{1,i}, \gamma_{2,i}, \zeta_i\right) \mid G \sim G && && i = 1, \ldots, n \\
&z_i \mid \zeta_i \sim f\left(z_i; \zeta_i\right) && && i = 1, \ldots, n \\
&t_{j,i} \mid x_{j,i}, \beta_j, \gamma_{j,i} \sim f\left(t_{j,i} \mid x_{j,i}; \beta_j, \gamma_{j,i}\right); && j = 1, 2; && i = 1, \ldots, n
\end{aligned} \tag{3.3}
$$

The flexibility of this factorization allows to understand the impact of Z on the dis-

tribution of T_j (or of (T_1, T_2)) by using standard probability calculus:

$$f\left(t_{j,i} \mid x_{j,i}, z_i; \beta_j, \gamma_{j,.}, \zeta_.\right) = \frac{\sum_{k=1}^{\infty} \pi_k f\left(t_{j,i} \mid x_{j,i}; \beta_j, \gamma^*_{j,k}\right) f\left(z_i; \zeta^*_k\right)}{\sum_{k=1}^{\infty} \pi_k f\left(z_i; \zeta^*_k\right)} \tag{3.4}$$

In a similar fashion, we can derive the probability distribution of the time to death of the last survivor $T_{\overline{1,2}} = \max(T_1, T_2)$. The corresponding survivor function, simplistically denoted as $S_{\overline{x_1,x_2}}(t \mid z)$ is useful especially for actuarial calculations, as we illustrate in Section 7:

$$S_{\overline{x_1,x_2}}(t \mid z) := \mathbb{P}\left(T_{\overline{1,2}} > t \mid x_1, x_2, z\right) = 1 - \mathbb{P}\left(T_1 < t \cup T_2 < t \mid x_1, x_2, z\right), \tag{3.5}$$

where

$$\mathbb{P}\left(T_1 < t \cup T_2 < t \mid x_1, x_2, z\right) = \frac{\sum_{k=1}^{\infty} \pi_k f\left(z; \zeta^*_k\right) \int_0^t f\left(u \mid x_1; \beta_1, \gamma^*_{1,k}\right) f\left(u \mid x_2; \beta_2, \gamma^*_{2,k}\right) \mathrm{d}u}{\sum_{k=1}^{\infty} \pi_k f\left(z; \zeta^*_k\right)}$$

Analogous formula can be used for the joint life survival probability, denoted as $S_{x_1,x_2}(t \mid z)$, characterizing the random variable $\min(T_1, T_2)$.

We can also calculate the probability of belonging to a certain class W conditional on the value of Z:

$$\mathbb{P}(W = k \mid Z = z) = \frac{\pi_k f\left(z; \zeta^*_k\right)}{\sum_{h=1}^{\infty} \pi_h f\left(z; \zeta^*_h\right)} \tag{3.6}$$

In this way, we can understand how couples with similar values of W can share the same mortality profile and husband-wife mortality dependence relationships as characterized by the kth class.

Compared to earlier literature in the topic, the AVDPM model extends the mixture model analysed by Ungolo & van den Heuvel (2022), since the authors consider a discrete random component (independent of Z) with unknown number of levels, chosen by means of a model selection procedure. The limitation of this approach is the need of estimating several models, which can be particularly time consuming for larger datasets as mentioned in Section 2. On the other hand, while Ungolo & van den Heuvel (2023) overcome this issue by assuming that the random component is drawn from a (truncated) Dirichlet Process, their model does not account for the statistical association among competing risks due to common factors.

As mentioned in Section 1, to the best of our knowledge, Youn & Shemyakin (1999) and Dufresne et al. (2018) are the sole references accounting for the common variable Z (the age difference between the spouses) within the copula dependence parameter.

Other than the analysis of dependent lifetimes, the AVDPM model can easily find application for the analysis of the time to competing causes of death, where Z may represent for example those genetic factors which impact the dependence between the causes. Alternatively, the AVDPM factorization in equation (3.1) can be used to jointly model dependent frequency and/or severity of claims in non-life and health insurance by type of event, without necessarily specify a strong parametric assumption on the dependence, as can be the case of copula models.

4. Data and parametric model

We showcase the approach outlined in Section 3 to the analysis of the Canadian life insurance dataset initially studied by Frees et al. (1996), and then also by Deresa et al. (2022) and Dufresne et al. (2018) among others. After some data processing operations, briefly described in the Supplementary Material, we have information about 12,139 joint and last-survivor annuity contracts in force between 29/12/1988 and 31/12/1993 (the observation period).

Specifically, we focus on a joint model for the lifetime distribution of individual members of male-female couples, for which we observe the starting date of the contract, the date of birth (thus their age at the start of the contract) and the date of death if any couple member dies within the observation period. To make the inferential process more challenging, hence to prove its robustness, we assume that the death of the first spouse causes a loss of information about the living one, whose lifetime is thus censored. Hence, this dataset contains a large number of censored units: indeed, we observe 1,164 deaths among males, 357 deaths among females, while the remaining units are all right censored. In addition, most units are also subject to left truncation. This means that we are able to observe the annuity contract only if both couple members are alive at the start of the observational period. Right censoring and left truncation must then be taken into account when deriving the likelihood function of the observations, as we show in Section 5.

In this dataset, the age at the start of the contract is the only individual-specific covariate (average of 65.6 for males and 62.6 for females). As in Dufresne et al. (2018) and Youn & Shemyakin (1999), we consider also the age-difference (AD, in absolute value, mean of 4.1 years) and the indicator variable MO, equal to 1 if the male is older than the woman and 0 otherwise (the male is the oldest policyholder in 77.3% of the couples in the dataset). According to their analysis, a model including these two variables captures some additional features of the association between the lifetime of the husband and of the wife. More precisely, they claim that the larger the age difference, the lower the lifetime dependence. In addition, Dufresne et al. (2018) observe how the MO covariate has an influence in the relationship between husband and wife, consequently affecting their lifetime dependence. Therefore, we will consider AD and MO as the two elements of the couple-specific covariates, $Z =$(AD, MO).

We randomly split the dataset into a training set, corresponding to the 75% of the policyholders within the dataset (9,104 units), and use the remaining 25% to test the

predictive ability of the model (3,035 units).

Following Frees et al. (1996) and Carriere (2000) we model the joint future lifetime distribution of the male ($j = 1$) and the female ($j = 2$) within the ith policy in terms of the following hazard function specification:

$$\mu\left(t \mid x_{j,i}; \alpha_j, \beta_j, \gamma_{j,i}\right) = \exp\left(\alpha_j + \beta_j\left(x_{j,i} + t\right) + \gamma_{j,i}\right), \tag{4.1}$$

where α_j denotes the intercept or log-baseline, β_j is the regression coefficient for the age effect, and $\gamma_{j,i}$ is the individual random effect capturing the heterogeneity and the dependence in the couple lifetimes. Given this specification, then the probability density function of $T_{j,i}$ can be written as:

$$f\left(t \mid x_{j,i}; \alpha_j, \beta_j, \gamma_{j,i}\right) = \exp\left[-\int_0^t \mu\left(s \mid x_{j,i}; \alpha_j, \beta_j, \gamma_{j,i}\right) \mathrm{d}s\right] \mu\left(t \mid x_{j,i}; \alpha_j, \beta_j, \gamma_{j,i}\right) \tag{4.2}$$

For the couple-specific covariates, we assume that $\ln\left(\mathrm{AD}_i\right) \sim \mathrm{N}\left(\cdot; \zeta_{\mathrm{AD},i}, \sigma^2_{\mathrm{AD}}\right)$ and $\mathrm{MO}_i \sim \text{Bernoulli}\left(\cdot; \zeta_{\mathrm{MO},i}\right)$. The density of $T_{j,i}$ as function of $x_{j,i}$ and z_i follows from equation (3.4). In this way, we can easily observe how the resulting joint model for (T_1, T_2, Z) has an additional layer of flexibility, since the effect of Z on the hazard function (through $\gamma_{j,i}$) is non-necessarily proportional, nor monotone.

The ith couple-specific parameters are drawn from a random distribution G, $(\gamma_{1,i}, \gamma_{2,i}, \zeta_{\mathrm{AD},i}, \zeta_{\mathrm{MO},i}) \sim G$ as from the generating process described in equation (3.3). We assume that G is a draw from a Dirichlet Process with concentration parameter ϕ and base measure G_0:

$$G_0 = \mathrm{MVN}\left(\begin{bmatrix}\gamma_{1,i} \\ \gamma_{2,i}\end{bmatrix}; \mathbf{0}, \Sigma_\gamma\right) \times \mathrm{N}\left(\zeta_{\mathrm{AD},i}; m_{\zeta_{\mathrm{AD}}}, s^2_{\zeta_{\mathrm{AD}}}\right) \times \mathrm{Beta}\left(\zeta_{\mathrm{MO},i}; 3, 1\right)$$

The model for the base distribution is motivated by the need to carry out a computationally efficient Bayesian inference by exploiting the conditional conjugacy of the parameters whereas possible, whilst keeping the model flexible enough to capture the complex features of the data. The parameters of the Beta distribution for $\zeta_{\mathrm{MO},i}$ are chosen in a way to provide some weak prior information: indeed, this assumption corresponds to a mean of 0.75 and a variance of 0.0375.

5. Inference

First of all, we approximate the DPM model of equation (3.2) by setting an upper bound $K = 25$ to the number of mixture components as in Ungolo & van den Heuvel (2023), which results in the truncated SBP described in Ishwaran & James (2001). This simplifies the implementation of the MCMC sampler compared to the use of the bound-free slice samplers of Walker (2007) and Kalli et al. (2011), or the retrospective sampler of Papaspiliopoulos & Roberts (2008), which avoid.

5.1. Prior distributions

We specify weakly informative and pairwise independent prior distributions for the model parameters as listed below, in order to facilitate the computation of the posterior distribution:

- $\exp(\alpha_j) \sim \text{Gamma}(\alpha_j; 1, 1); \quad j = 1, 2;$
- $\beta_j \sim \text{Uniform}(\beta_j; 0, 5); \quad j = 1, 2;$
- $\Sigma_\gamma \sim \text{Inv-Wishart}\left(\Sigma_\gamma;\ 7,\ 0.001 \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}\right);$
- $m_{\zeta_{\text{AD}}} \sim \text{N}(m_{\zeta_{AD}}; 0, 1);$
- $s^2_{\zeta_{\text{AD}}} \sim \text{Inv-Gamma}\left(s^2_{\zeta_{\text{AD}}}; 2, 2\right);$
- $\sigma^2_{\text{AD}} \sim \text{Inv-Gamma}\left(\sigma^2_{\text{AD}}; 1, 1\right);$
- $\phi \sim \text{Gamma}(\phi; 1, 1)$

The weak prior distribution allows to incorporate some information available to the researcher, such as the positive statistical association between the couple member lifetimes, as we do for the prior distribution of Σ_γ. The uniform distribution for β_j is motivated on the grounds of previous analyses, where the lower bound of zero is set in order to obtain an increasing hazard function at older ages, which ensures the biological reasonableness of the model, while the upper bound is set to a value large enough to ensure the prior is almost non-informative. Previous mortality analyses of a similar model show a value of β around 0.10 (see for example Ungolo et al. (2020) and Richards (2008)).

With the exception of β_j and $\gamma^* = \left(\left(\gamma^*_{1,1}, \gamma^*_{2,1}\right), \ldots, \left(\gamma^*_{1,K}, \gamma^*_{2,K}\right)\right)$, we can obtain closed form updates for all parameters, as shown in Appendix A. The conditional conjugacy of ϕ is discussed in Escobar & West (1995).

5.2. Likelihood

Let $d_{j,i}$ denote an indicator variable which is equal to 1 if the male ($j = 1$) or the female couple member ($j = 2$) is observed to die throughout the observational study and 0 otherwise (hence, the lifetime variable is censored). We assume that the censoring mechanism is ignorable, since the censoring event for a couple member is either caused by the death of the other (whose conditionally independent lifetime is included in the joint model) or from the end of the observation period (Type I censoring). In our dataset we observe each couple until the first member dies. This means that we can only observe $t_i = \min(t_{1,i}, t_{2,i})$. Furthermore, we assume that the distribution of the times to event for each couple are independently distributed, conditional on the covariates and the multivariate random effect.

As discussed in Section 4, data are subject to left truncation, since each contract is observable upon survival of both members at the start of the study. The left-truncation level for each couple is denoted by the variable a_i, denoting the time (in years) from the start of the contract to the start of the observational study. This means that we need to work with the lifetime density function conditional on both couple members being alive at the beginning of the observation period. a_i is equal to zero if the contract starts during the observation period.

Let $\mathbf{t} = (t_1, \ldots, t_n)$, $\mathbf{x} = (x_{1,i}, x_{2,i}, \ldots, x_{1,1}, x_{2,1})$, $\mathbf{d} = (d_{1,1}, d_{2,i}, \ldots, d_{1,n}, d_{2,n})$, $\mathbf{ad}$= $(\mathrm{ad}_1, \ldots, \mathrm{ad}_n)$, $\mathbf{mo}$=$(\mathrm{mo}_1, \ldots, \mathrm{mo}_n)$, $\mathbf{a} = (a_1, \ldots, a_n)$, $\alpha = (\alpha_1, \alpha_2)$, $\beta = (\beta_1, \beta_2)$, $\psi = (\psi_1, \ldots, \psi_{K-1})$, $\gamma^* = \left(\gamma^*_{1,1}, \ldots, \gamma^*_{J,K}\right)$, $\zeta^*_{\mathrm{AD}} = \left(\zeta^*_{\mathrm{AD},1}, \ldots, \zeta^*_{\mathrm{AD},K}\right)$ and $\zeta^*_{\mathrm{MO}} = \left(\zeta^*_{\mathrm{MO},1}, \ldots, \zeta^*_{\mathrm{MO},K}\right)$. We introduce the latent indicator variable $s_{i,k}$, which is equal to 1 if the ith couple belongs to the kth class, and zero otherwise. This facilitates an efficient computation of the posterior distribution (Müller et al. 1996).

The likelihood function of the parameters conditional on $\mathbf{t}$, $\mathbf{x}$, $\mathbf{a}$, $\mathbf{d}$, $\mathbf{ad}$, $\mathbf{mo}$ and $\mathbf{s} = (s_{1,1}, \ldots, s_{n,K})$ is given by:

$$L\left(\alpha, \beta, \gamma^*, \zeta^*_{AD}, \zeta^*_{MO}, \sigma^2_{AD}, \psi \mid \mathbf{t}, \mathbf{x}, \mathbf{d}, \mathbf{a}, \mathbf{mo}, \mathbf{ad}, \mathbf{s}\right) \tag{5.1}$$
$$\propto \prod_{i=1}^{n} \pi_k^{s_{i,k}} \left\{ \left[\prod_{j=1}^{2} \dot{f}\left(t_i \mid x_{j,i}, d_{j,i}, a_i; \alpha_j, \beta_j, \gamma^*_{j,k}\right) \right] \mathrm{N}\left(\ln\left(\mathrm{ad}_i\right); \zeta^*_{AD,k}, \sigma^2_{AD}\right) \mathrm{Bernoulli}\left(\mathrm{mo}_i; \zeta^*_{MO,k}\right) \right\}^{s_{i,k}}$$

where $\mathbf{s} = (s_{1,1}, \ldots, s_{n,K})$.

In order to account for right censoring and left truncation, then by simple algebra we have:

$$\begin{aligned} &\dot{f}\left(t_i \mid x_{j,i}, d_{j,i}, a_i; \alpha_j, \beta_j, \gamma^*_{j,k}\right) \\ &= \frac{\exp\left[-\int_0^{t_i+a_i} \mu\left(s \mid x_{j,i}; \alpha_j, \beta_j, \gamma^*_{j,k}\right) \mathrm{d}s\right] \mu\left(t_i + a_i \mid x_{j,i}; \alpha_j, \beta_j, \gamma^*_{j,k}\right)^{d_{j,i}}}{\exp\left[-\int_0^{a_i} \mu\left(q \mid x_{j,i}; \alpha_j, \beta_j, \gamma^*_{j,k}\right) \mathrm{d}q\right]} \end{aligned} \tag{5.2}$$

whose logarithm is given by:

$$\begin{aligned} &\ln \dot{f}\left(t_i \mid x_{j,i}, d_{j,i}, a_i; \alpha_j, \beta_j, \gamma^*_{j,k}\right) \\ &= -\frac{\exp\left[\beta_j\left(t_i + a_i\right)\right] - \exp\left(\beta_j a_i\right)}{\beta_j} \exp\left(\alpha_j + \beta_j x_{j,i} + \gamma^*_{j,k}\right) + d_{j,i}\left(\alpha_j + \beta_j\left(x_{j,i} + t_i + a_i\right) + \gamma^*_{j,k}\right) \end{aligned} \tag{5.3}$$

5.3. Posterior distribution

The posterior distribution follows as the product of the likelihood and the prior distribution:

$$
\begin{aligned}
& p\left(\alpha, \beta, \gamma^*, \zeta^*_{\mathrm{AD}}, \zeta^*_{\mathrm{MO}}, \sigma^2_{\mathrm{AD}}, \Sigma_\gamma, m_{\zeta_{\mathrm{AD}}}, s^2_{\zeta_{\mathrm{AD}}}, \phi, \psi \middle| \mathbf{t}, \mathbf{x}, \mathbf{d}, \mathbf{mo}, \mathbf{ad}, \mathbf{s}\right) \qquad (5.4) \\
& \propto L\left(\alpha, \beta, \gamma^*, \zeta^*_{\mathrm{AD}}, \zeta^*_{\mathrm{MO}}, \sigma^2_{\mathrm{AD}}, \psi \mid \mathbf{t}, \mathbf{x}, \mathbf{d}, \mathbf{mo}, \mathbf{ad}, \mathbf{s}\right) \left[\prod_{j=1}^{2} \mathrm{Gamma}\left(\exp\left(\alpha_j\right); 1, 1\right) \mathrm{Uniform}\left(\beta_j; 0, 5\right)\right] \\
& \times \left[\prod_{k=1}^{K} \mathrm{MVN}\left(\gamma^*_{\cdot,k}; 0, \Sigma_\gamma\right) \mathrm{N}\left(\zeta^*_{\mathrm{AD},k}; m_{\zeta_{\mathrm{AD}}}, s^2_{\zeta_{\mathrm{AD}}}\right) \mathrm{Beta}\left(\zeta^*_{\mathrm{MO},k}; 3, 1\right)\right] \mathrm{N}\left(m_{\zeta_{\mathrm{AD}}}; 0, 1\right) \\
& \times \text{Inv-Wishart}\left(\Sigma_\gamma;\ 7,\ 0.001 \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}\right) \mathrm{Inv\text{-}Gamma}\left(\sigma^2_{\mathrm{AD}}; 1, 1\right) \mathrm{Inv\text{-}Gamma}\left(s^2_{\zeta_{\mathrm{AD}}}; 2, 2\right) \\
& \times \left[\prod_{k=1}^{K-1} \mathrm{Beta}\left(\psi_k; 1, \phi\right)\right] \mathrm{Gamma}\left(\phi; 1, 1\right)
\end{aligned}
$$

In order to efficiently learn the posterior distribution of equation (5.4), we propose to first Data-Augment the dataset of the missing value of the latent class $s_{i,k}$, and then use a blocked Gibbs sampler scheme (Ishwaran & James (2001)), consisting of a sequential draws of the parameters (exploiting their conditional conjugacy). The steps of this Data Augmentation-Blocked Markov Chain Monte Carlo (MCMC) sampler are detailed in Appendix A. We implement this sampler in R (R Core Team 2013) in order to have a full control over the MCMC sampling process. The code implementing the sampler is available at the GitHub repository `https://github.com/ungolof/AVDPM`.

6. Results

6.1. Convergence

The steps of the MCMC sampler devised for the analysis of the posterior distribution outlined in Section 5.3 are iterated 50,000 times. We discard the first 40,000 iterations (burn-in) and we thin the chain every 20 draws to reduce the degree of autocorrelation between iterations, resulting in a final posterior sample of 500 draws. We run the sampler four times, based on sparse starting values, in order to assess its convergence towards a stationary distribution.

From the analysis of the trace plots we observe how the chains converge towards a stationary distribution for all parameters, and these mix very well, except for γ^*, ζ^*_{MO}, ζ^*_{AD} and $\pi\left(\psi\right)$[2]. We expect this result, which is due to the label switching problem typical of Bayesian mixtures, especially when no label constraint has been placed (see Betancourt (2017), Marin et al. (2005) and Ungolo et al. (2020) for a detailed account about this problem). This issue only affects the interpretation of the groups from the

[2]Plots available upon request.

results of one chain compared to another. Indeed, when looking at the occupancy of the classes across iterations, we note a tendency of the sampler to have a similar number of units.

Nevertheless, this does not represent an issue when making predictions, or when the purpose is to learn the global parameters, such as α and β.

Figure 6.1 shows the posterior mean of the number of units included in each cluster. We observe how the first three classes turn out to be the most populated throughout the iterations, and four classes cover more than 2.5% of the observations. This is also a further evidence of the appropriateness of the upper bound choice $K = 25$.

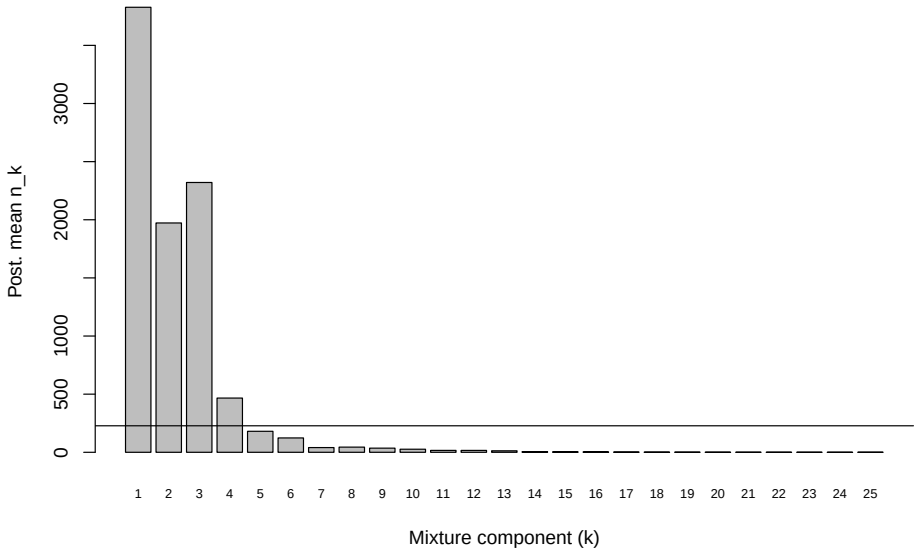

Figure 6.1: Barplot of the posterior average of the mixture component occupancy ($n_k^{(\ell)}$, $k = 1, \ldots, K$) based on the retained posterior sample. The horizontal black line indicates the level of 228 units.

6.2. Model analysis

Model results

Table 6.1 shows the summary statistics of the posterior distribution of the most relevant parameters for the AVDPM approach of this paper. The log-baseline mortality (α) for females is lower than for males, as we can see also from the value of the 95% credible interval extremes which do not overlap. On the other hand, the female members of the couple are characterized by a larger sensitivity of the hazard function with respect to the age as measured by the parameter β compared to males. This is shown in Figure 6.4, where we plot the empirical death rates and compare the resulting hazard function of the AVDPM against those of two other competing models, as we discuss later.

From the analysis of the posterior distribution of the parameters, we observe that four classes cover 94% of the observations, and are characterized by different values of

the class-specific parameters $\left(\gamma^*_{1,k}, \gamma^*_{2,k}, \zeta^*_{\text{AD},k}, \zeta^*_{\text{MO},k}\right)$ for $k = 1, 2, 3, 4$. We note how the additional heterogeneity layer as captured by the random component γ implies a decrease in the hazard function for the males and an increase for the corresponding hazard function for the females, due to the sign of their coefficient, and to the value of the 95% credible interval bands. We add more on this point when analysing the classes.

Table 6.1: Posterior summaries of α, β, γ^*_k, $\zeta^*_{\text{AD},k}$, $\zeta^*_{\text{MO},k}$, π_k for $k = 1, 2, 3, 4$

Parameter	**Mean**	**95% Cred. Int.**	**Parameter**	**Mean**	**95% Cred. Int.**
α_1	-9.3795	$(-10.4781; -8.3154)$	α_2	-15.6532	$(-17.3233; -14.4152)$
β_1	0.0901	(0.0758; 0.1036)	β_2	0.1296	(0.1123; 0.1497)
$\gamma^*_{1,1}$	-0.9139	$(-1.3553; -0.5222)$	$\gamma^*_{2,1}$	1.8088	(1.3064; 2.3155)
$\gamma^*_{1,2}$	-0.7064	$(-1.0829; -0.3504)$	$\gamma^*_{2,2}$	1.6669	(1.1783; 2.1585)
$\gamma^*_{1,3}$	-0.9194	$(-1.2799; -0.3824)$	$\gamma^*_{2,3}$	1.2060	(0.78495; 1.65906)
$\gamma^*_{1,4}$	-3.5996	$(-3.9752; -2.8663)$	$\gamma^*_{2,4}$	1.7884	(1.5285; 2.0137)
$\zeta^*_{\text{AD},1}$	1.6784	(1.59056; 1.78427)	$\zeta^*_{\text{MO},1}$	0.9755	(0.93908; 0.9982)
$\zeta^*_{\text{AD},2}$	0.1326	$(-0.0626; 0.32496)$	$\zeta^*_{\text{MO},2}$	0.6260	(0.57256; 0.68335)
$\zeta^*_{\text{AD},3}$	1.1933	(0.99488; 1.32565)	$\zeta^*_{\text{MO},3}$	0.6406	(0.49215; 0.74976)
$\zeta^*_{\text{AD},4}$	-1.2193	$(-1.5354; -0.9107)$	$\zeta^*_{\text{MO},4}$	0.5054	(0.39599; 0.59236)
π_1	0.4204	(0.3120; 0.5154)	π_2	0.2167	(0.1594; 0.2599)
π_3	0.2549	(0.1630; 0.3570)	π_4	0.0512	(0.03688; 0.0665)

Analysis of the effect of couple-specific covariates

We plot the value of the log-hazard function of the males and females for different values of AD and of the MO covariate (Figure 6.2). Using the posterior mean of the parameters, this conditional hazard function is calculated as:

$$\begin{aligned}
&\mu\left(t \mid x_{j,\cdot}, \text{ad}, \text{mo}; \alpha_j, \beta_j, \gamma^*_{j,\cdot}, \pi, \zeta^*_{\text{AD}}, \zeta^*_{\text{MO}}\right) \\
&= \frac{\displaystyle\sum_{k=1}^{K} \pi_k f\left(t \mid x_{j,\cdot}; \alpha_j, \beta_j, \gamma^*_{j,k}\right) \text{N}\left(\ln\left(\text{ad}\right); \zeta^*_{\text{AD},k}, \sigma^2_{\text{AD}}\right) \text{Bernoulli}\left(\text{mo}; \zeta^*_{\text{MO},k}\right)}{\displaystyle\sum_{k=1}^{K} \pi_k \left(1 - F\left(t \mid x_{j,\cdot}; \alpha_j, \beta_j, \gamma^*_{j,k}\right)\right) \text{N}\left(\ln\left(\text{ad}\right); \zeta^*_{\text{AD},k}, \sigma^2_{\text{AD}}\right) \text{Bernoulli}\left(\text{mo}; \zeta^*_{\text{MO},k}\right)}
\end{aligned} \tag{6.1}$$

where F denotes the cumulative distribution function of the random time to event T, hence $1 - F$ denotes the survival function.

As outlined in Section 4, this model allows for a non monotone effect of the covariates on the hazard function compared to a typical proportional hazard model approach as used in Deresa et al. (2022), and later analysed to compare the models.

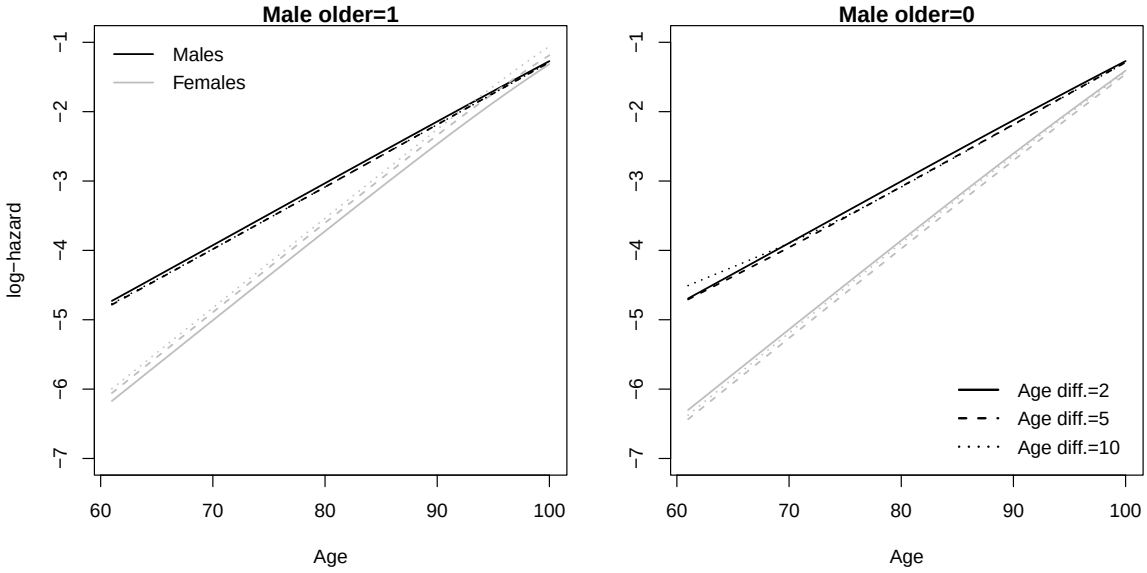

Figure 6.2: Hazard function for males (black) and females (gray) aged 60-100 for different values of the covariates male older and age difference (not in log scale).

First of all, we observe how in case the male is the older couple member (MO=1), there is a slight shift upwards of the hazard function for both males and females, with a more visible effect on the latter.

Furthermore, when MO=1, an age difference of 5 or 10 years has a similar hazard function for the males, while the hazard function increases with the age difference for the females. Conversely, when the female is older, we see that the hazard function values have a non-monotonic relationship with respect to the age difference between the couple members. This can be observed especially for the females hazard function, where the hazard function corresponding to an age difference of 5 years lies below the hazard functions obtainable when the age difference is equal to 2 and 10. In this way, the model captures the interaction between age difference and the MO covariate, since the effect of the age difference is not the same based on different values of the MO covariate.

Dependent lifetime events

The AVDPM of this work is proposed to analyse dependent lifetimes. We generate a sample of 10,000 joint lifetimes for males and females with different ages (sampled from a Uniform(60, 100) distribution, respecting AD and MO[3]), and analyse their statistical association based on different values of the of the couple-specific covariates. Again, we sample the time to event using the posterior mean of the parameters. The statistical association of the time to event is assessed by using the Spearman ρ and the Kendall τ

[3]For example, if we sample for a male an age equal to 65, with AD=2 and MO=1, then the female has an age of 63.

rank correlation coefficient, which are calculated for each value of AD between 0.5 and 20 and MO and plotted in Figure 6.3.

First of all, we note how both statistics agree on the pattern of positive dependence for each value of AD $(0.5, 1, 1.5, \ldots, 20)$ and MO. The spikes in the plots are the consequence of the random lifetime sampling process at each value of (AD, MO). In more detail, the dependence is relatively smaller for an age difference smaller than one year, being slightly larger when the male is the oldest couple member. For an age difference between 1 and 7 years the two statistics have a stable value, regardless on whether the male is the oldest couple member. For an age difference larger than 7 years, the lifetime dependence shows a decreasing trend in case the female is older than males, while in the opposite case, the value of both coefficients remains stable.

These results show how the AVDPM approach of this work can account for the Z-lifetime dependence as by-product.

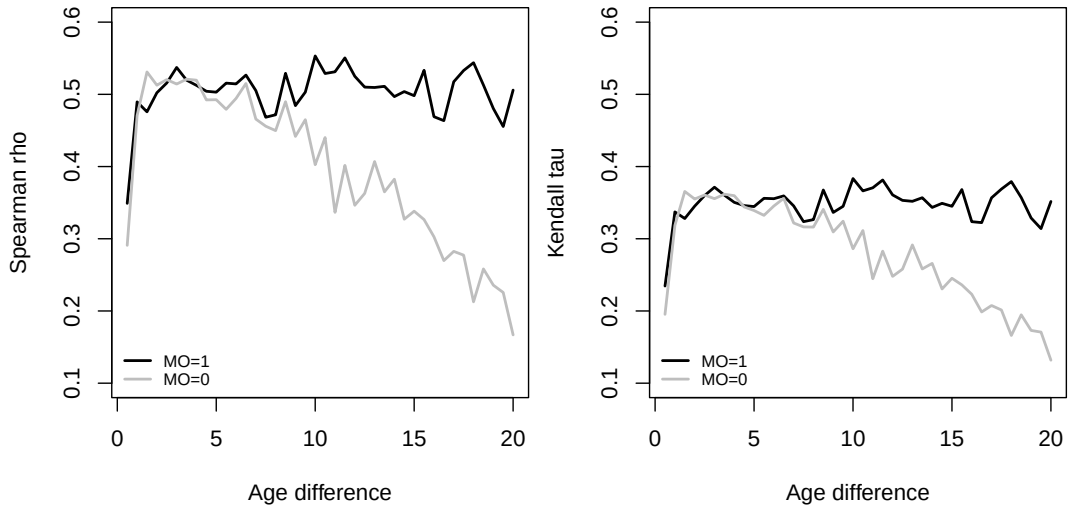

Figure 6.3: Value of the Spearman ρ (left panel) and of the Kendall τ (right panel) for different values of AD and MO.

Class analysis

The mixture modelling nature of AVDPM allows to classify the observations *a posteriori*, which can be helpful to learn further information about the resulting groups. A similar analysis was carried out in Ungolo & van den Heuvel (2023).

At this purpose we use the Bayes' rule: for the ith couple, we calculate the probability to be in the kth class, denoted as $q_{i,k}$, conditional on the observable data and the model

parameters (the posterior mean in this case) as follows:

$$q_{i,k} = \mathbb{P}\left(W_i = k \mid t_i, x_{1,i}, x_{2,i}, a_i, d_i, \ln\left(\mathrm{ad}_i\right), \mathrm{mo}_i\right) \tag{6.2}$$
$$= \frac{\pi_k \left[\prod_{j=1}^{2} f\left(t_i \mid x_{j,i}, d_{j,i}, a_i; \alpha_j, \beta_j, \gamma^*_{j,k}\right)\right] \mathrm{N}\left(\ln\left(\mathrm{ad}_i\right); \zeta^*_{\mathrm{AD},k}, \sigma^2_{\mathrm{AD}}\right) \mathrm{Bernoulli}\left(\mathrm{mo}_i; \zeta^*_{\mathrm{MO},k}\right)}{\sum_{h=1}^{K} \pi_q \left[\prod_{j=1}^{2} f\left(t_i \mid x_{j,i}, d_{j,i}, a_i; \alpha_j, \beta_j, \gamma^*_{j,h}\right)\right] \mathrm{N}\left(\ln\left(\mathrm{ad}_i\right); \zeta^*_{\mathrm{AD},k}, \sigma^2_{\mathrm{AD}}\right) \mathrm{Bernoulli}\left(\mathrm{mo}_i; \zeta^*_{\mathrm{MO},k}\right)}$$

The ith couple is then hard-assigned to the kth class, $w_i = k$ if $q_{i,k} > q_{i,h}$ for $h \neq k$. The first four classes total 97% of the observations. Table 6.2 illustrates the key features of Group 1-4, alongside the posterior mean of their class-specific parameters $\left(\gamma_{1,\cdot}, \gamma_{2,\cdot}, \zeta^*_{\mathrm{AD},\cdot}, \zeta^*_{\mathrm{MO},\cdot}\right)$. The difference between % Composition in Table 6.2 and π is due to the specific classification rule we use.

Table 6.2: Features of the four largest classes resulting from the application of the Bayes' rule.

	Group 1	Group 2	Group 3	Group 4	Train. sample
% Composition	53.7	24.8	13.7	4.9	–
Age male (mean)	66.28	65.0	64.35	64.67	65.57
Age female (mean)	60.27	64.64	67.12	64.66	62.64
Age Diff. (mean)	6.01	0.36	−2.77	0.01	2.93
Male older (in %)	99.8	64.9	24.6	52.0	77.25
$\gamma^*_{1,\cdot}$	−0.91	−0.71	0.92	−3.60	–
$\gamma^*_{2,\cdot}$	1.81	1.67	1.21	1.79	–
$\zeta^*_{\mathrm{AD},\cdot}$	1.68	0.13	1.19	−1.21	–
$\zeta^*_{\mathrm{MO},\cdot}$	0.98	0.63	0.64	0.51	–

A first striking evidence is that these four classes have different features compared to the whole training sample. This means that we can be able to identify groups of couples which have distinctive features.

More than 50% of the couples compose Group 1, which is characterized by the largest age difference among the four groups and the male is the oldest member of the couple for almost all observations.

Group 2 and Group 4 appear to be more similar in terms of characteristics, while characterized by a different value of their corresponding parameters, especially $\gamma^*_{1,\cdot}$ and $\zeta^*_{\mathrm{AD},\cdot}$. Indeed, for Group 4, the men have a lower hazard function, since $\gamma^*_{1,4}$ is considerably lower than $\gamma^*_{1,2}$. This may be indicative of the presence of further sources of heterogeneity which increase males' life expectancy, and which we were not able to observe in this dataset.

Group 3 instead, can be characterized as a subgroup of the training sample where the females are predominantly older than men, and the latter are characterized by a higher hazard function (as we can note from $\gamma^*_{1,3}$) compared to the other three groups.

Comparison with other models

The results of the AVDPM approach are compared with those obtainable by assuming a basic Gompertz (BG) hazard function, and with a proportional hazard model which includes all covariates (PH), similarly to Deresa et al. (2022)[4] which assume that:

$$\mu^{BG}\left(t \mid x_{j,i}; \alpha_j, \beta_j\right) = \exp\left(\alpha_j + \beta_j\left(x_{j,i} + t\right)\right)$$
$$\mu^{PH}\left(t \mid x_{j,i}, \mathrm{ad}_i, \mathrm{mo}_i; \alpha_j, \beta_j, \delta_{j,1}, \delta_{j,2}\right) = \exp\left(\alpha_j + \beta_j\left(x_{j,i} + t\right) + \delta_{j,1}\ln\left(\mathrm{ad}_i\right) + \delta_{j,2}\mathrm{mo}_i\right)$$

with parameters estimated by using maximum likelihood. Therefore, under BG and PH we assume that conditional on the individual age, and the two common covariates, then the males and females lifetimes are independently distributed.

Figure 6.4 shows the log-hazard rates for individuals aged 60 onwards under the three analysed models for both males and females. The log-hazard function for the PH model is plotted at the value of AD=1 and MO=0. For the AVDPM the hazard function is calculated at the value of the posterior mean of the parameters.

For the males, until age 90 all models show a similar fit in terms of slope and baseline. This is because the effect of common covariates on the males is not statistically significant, as shown from the parameter estimates and the standard errors of $\delta_{1,1}$ and $\delta_{1,2}$ (see Table B.1 in Appendix B). The change in the slope of the log-hazard function of the AVDPM approach after age 90 is a consequence of the use of random effects, widely studied in the literature (see Lancaster (1979), Vaupel & Yashin (1985) and Wienke (2014)). For the females, where we have a considerably lower number of deaths (thus a higher degree of censoring), we note as expected a lower empirical log-death rates, which are also more erratic. Despite the similar slope, there is some difference in the fitted log-hazard rates owing the effect of the heterogeneity due to the common covariates.

We quantitatively compare the three models by using the Akaike Information Criteria (Akaike (1974)) for the BG and PH model, and its generalization to a Bayesian framework with latent variables, known as Widely Applicable Information Criteria (WAIC, Watanabe (2009)). Their computation is detailed in Appendix C. AIC and WAIC are computed for both the training sample and the held out part of the dataset (Table 6.3). We chose the model which minimizes the value of these criteria.

[4] Compared to this paper, we consider the fact age increases over time, instead of being a fixed covariate, and take the logarithm of the absolute value of the age difference. In our analysis this shows an improvement in the value of the Akaike Information Criteria. The results are available upon request to the author.

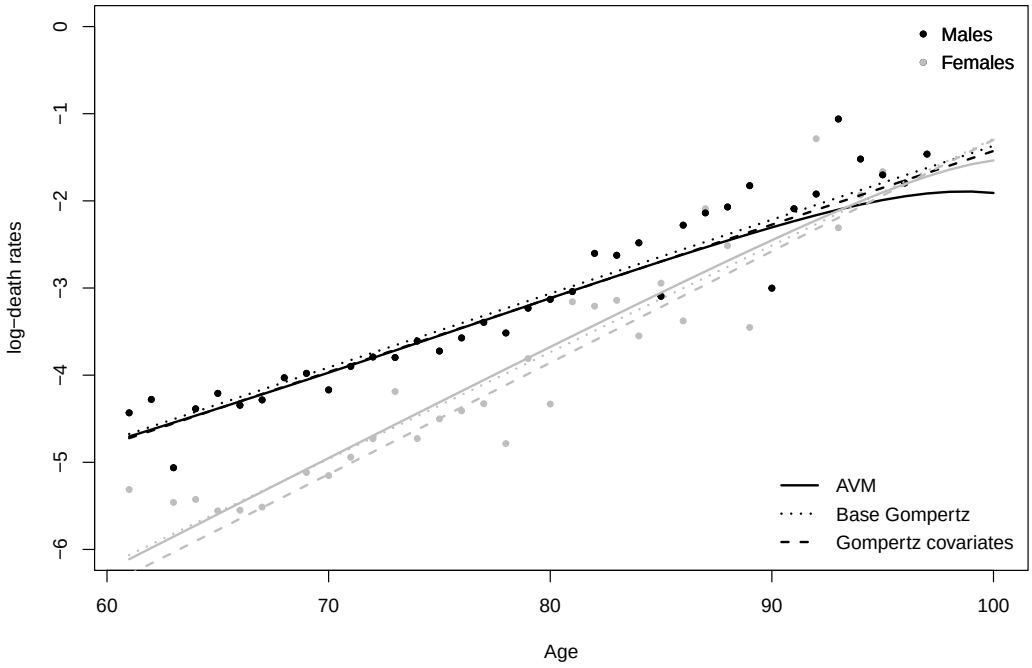

Figure 6.4: Log-hazard functions at age 60-100 for males and females under the BG, PH and AVDPM models.

Table 6.3: AIC and WAIC of the models (the lowest value is shown in bold).

(W)AIC	**BG**	**PH**	**AVDPM**
In sample	10,872.10	10,866.14	**10,857.68**
Out of sample	3,788.19	3,788.11	**3,762.97**

The inclusion of the covariates within a proportional hazard model has the effect to slightly improve the performance of the model compared to the base Gompertz hazard function, as earlier discussed. The benefit of including covariates is very negligible when looking at the out-of-sample performance of the PH model. Conversely, the AVDPM approach shows a smaller value of the WAIC for both the training and test dataset. Therefore, we conclude that the enhanced flexibility of AVDPM yields a better in sample and out of sample fit for these data.

7. Actuarial illustration of AVDPM

Let $Y_{x_1,x_2}(z)$ denote the present value of a cash flow of 1\$ paid continuously to a couple with characteristics z, where the male is aged x_1 and the female x_2, as long as both are alive (joint status). Conversely, let $Y_{\overline{x_1,x_2}}(z)$ denote the present value of a 1\$ cash flow

paid continuously until the death of the last survivor of a similar couple.

Assuming a force of interest ι, we can obtain the annuity factor of these cash flows as the expected value of $Y_{x_1,x_2}(z)$ and $Y_{\overline{x_1,x_2}}(z)$ (Dickson et al. 2013):

$$\overline{a}_{x_1,x_2}(z) = \mathbb{E}\left[Y_{x_1,x_2}(z)\right] = \int_0^\infty \exp(-\iota t)\, S_{x_1,x_2}(t \mid z)\, \mathrm{d}t$$

$$\overline{a}_{\overline{x_1,x_2}}(z) = \mathbb{E}\left[Y_{\overline{x_1,x_2}}(z)\right] = \int_0^\infty \exp(-\iota t)\, S_{\overline{x_1,x_2}}(t \mid z)\, \mathrm{d}t$$

Figure 7.1 shows the percentage difference between the value of the last survivor (LS) annuity factor obtainable under the AVDPM model which accounts for dependence and the PH model described in Section 6. The annuity factors are evaluated at different male ages (60 and 70), different values of Z =(AD, MO) and two different forces of interest $\iota = (0.01, 0.05)$. The same plots for the joint life annuity are shown in Appendix **??** (Figure **??**).

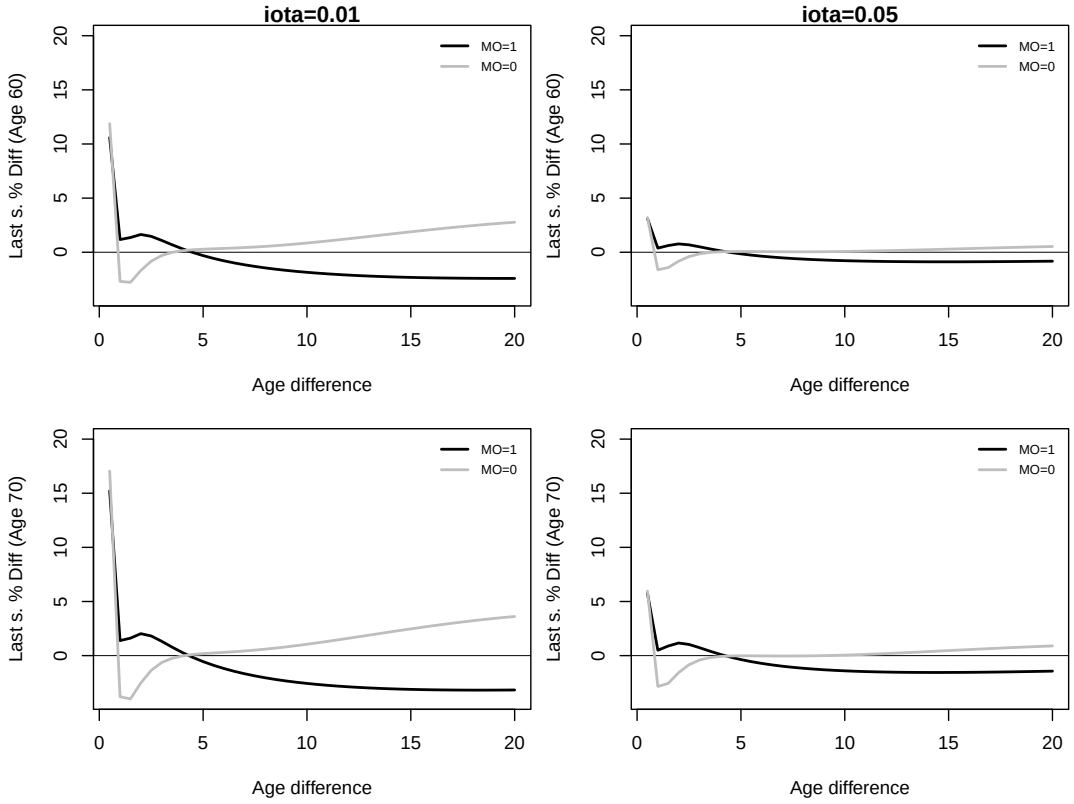

Figure 7.1: Percentage difference between the LS annuity factor under AVDPM and PH for different values of age difference (x-axis) and MO when the oldest member is aged 60 (top panel) and 70 (bottom panel), for $\iota = 1\%$ (left panel) and $\iota = 5\%$ (right panel).

For an age difference of 0.5 we note how the last survivor annuity factor under the

independence PH model is lower compared to the value obtainable under AVDPM. This difference is largest for $\iota = 1\%$. This evidence is in contrast with earlier findings in the literature which use copula models to account for lifetime dependence. Using a Frank copula, Frees et al. (1996, Figure 3) shows how the price of a LS annuity is higher under a model which assumes independent lifetimes, especially for couples where males and females have a similar age. Similarly, Dufresne et al. (2018, Figure 5.2) observe that the difference between independent and dependent lifetime (using a Clayton and a Joe copula) in the last-survivor life expectancy is highest for an age difference around 0, regardless the value of MO.

Then, we observe how the difference between the annuity factor under the PH and AVDPM increases with the age difference. In particular, in case the female is older, the independence assumption underlying the PH model tends to underprice the annuity, while the opposite holds in case MO=1. These evidences are consistent with the findings in Frees et al. (1996, Figure 4), and Deresa et al. (2022), where the latter analyse only the case of a couple where a male and a female are aged 65 and 63 respectively.

The plots show that the lower the value of ι, the larger the percentage difference between the annuity factors obtainable under the PH and the AVDPM models. A similar evidence is observed for annuities starting at older ages.

For the joint life annuity we observe that when the male is aged 60 at contract inception, and is the oldest member of the couple, then the annuity factor under the PH model is lower compared to its AVDPM counterpart. This evidence is also observed by Deresa et al. (2022), although they focus on the sole case of a 65 years old male with MO=1 and AD=2.

Conversely, when the female is the oldest member, the JL annuity factor under the AVDPM model is lower compared to the case we use the PH model. This is observed under the two forces of interest hereby analysed and ages at contract start.

The effect of the force of interest on the percentage difference between annuity factors is the same as for the last survivor annuity, while the older age at contract inception increases the relative value of the annuity factor under independence compared to the case where we use the AVDPM.

8. Extension of AVDPM to the analysis of the joint lifetimes of non-exchangeable units

So far we have illustrated the method for the case of males-females couples. If we want to allow for groups with a different number of exchangeable members, as can be the case of collective insurance policies, we can extend the framework through a hierarchical model with additional layers.

Suppose the ith group includes J_i members, with common set of variables z_i, and individual (within group) characteristics $\mathbf{x}_{j,i}$. A possible solution is to model the group-

specific joint distribution of the lifetimes and $\mathbf{Z}$ as:

$$f\left(t_{1,i},\ldots,t_{J_i,i},\mathbf{z}_i;\beta,G\right)=\int_{\Omega_{\gamma,\eta}}\prod_{j-1}^{J_i}f\left(t_{j,i}\mid x_{j,i};\beta,\gamma_{j,i}\right)f\left(z_i;\zeta_i\right)\mathrm{d}Q\left(\gamma_{j,i};\eta_i\right)\mathrm{d}G\left(\eta_i,\zeta_i\right) \tag{8.1}$$

where Q denotes a suitable distribution function for $\gamma_{j,i}$, indexed by the group specific parameter η_i. Again, Q can also be a random draw from a Dirichlet Process, although we would opt for a simpler known parametric form for computational reasons and also because we are indexing Q with a group-specific parameter η_i whose distribution is drawn from a DP.

9. Conclusion

This paper contributes to the analysis of grouped dependent lifetime events by proposing a joint model for the lifetimes which is augmented of the distribution of the group-specific covariates. The inclusion of multivariate random effects captures the dependence among the lifetimes, and between the lifetimes and group-specific covariates. The use of Dirichlet Process Mixture models enhance the flexibility of the random effects and of the standard parametric assumptions for the covariates. The resulting Augmented Variable DPM (AVDPM) model has been implemented for the empirical analysis of the mortality rates of the male and females members of a couple, which resulted in an enhanced in-sample and out-of-sample fitting performance. We showed how the model output can be used to infer additional information on the nature of the male-female mortality dependence, and how this can affect the price of joint life and last survivor annuities.

The only drawback of the AVDPM presented in this paper is the need of a full Bayesian analysis, which may be computationally expensive. A simpler approach would be to assume a fixed, known number of mixture components, and fit the model parameters using maximum likelihood. The results of similar analysis in the field may be used at this purpose.

Acknowledgments

I wish to thank the Society of Actuaries, through the courtesy of Edward (Jed) Frees and Emiliano Valdez, for allowing the use of the data in this paper.

A. Data Augmentation MCMC sampler

Below, we outline the steps of the Data Augmentation MCMC scheme to sample the parameters from the posterior distribution:

Step 0: Set an initial value for the parameters $\left(\alpha^{(0)},\beta^{(0)},\gamma^{*(0)},\zeta_{AD}^{*(0)},\zeta_{MO}^{*(0)},\sigma_{AD}^{2(0)},\Sigma_{\gamma}^{(0)},m_{\zeta_{\mathrm{AD}}}^{(0)},s_{\zeta_{\mathrm{AD}}}^{2(0)},\phi^{(0)},\psi^{(0)}\right)$;

For $\ell = 1, \ldots, M$:

Step 1: For each unit sample the mixture component $w_i^{(\ell)}$ ($W_i^{(\ell)} \in \{1, \ldots, K\}$) from a discrete distribution with probability:

$$\begin{aligned} &\mathbb{P}\left(W_i^{(\ell)} = k \mid t_i, x_{1,i}, x_{2,i}, a_i, d_i, \ln\left(\mathrm{ad}_i\right), \mathrm{mo}_i\right) \\ &= \frac{\pi_k^{(\ell-1)}\left[\prod_{j=1}^{2} \dot{f}\left(t_i \mid x_{j,i}, d_{j,i}, a_i; \alpha_j^{(\ell-1)}, \beta_j^{(\ell-1)}, \gamma_{j,k}^{*(\ell-1)}\right)\right] f\left(z_i; \zeta_k^{*(\ell-1)}\right)}{\sum_{q=1}^{K} \pi_q^{(\ell-1)}\left[\prod_{j=1}^{2} \dot{f}\left(t_i \mid x_{j,i}, d_{j,i}, a_i; \alpha_j^{(\ell-1)}, \beta_j^{(\ell-1)}, \gamma_{j,q}^{*(\ell-1)}\right)\right] f\left(z_i; \zeta_k^{*(\ell-1)}\right)}, \end{aligned} \tag{A.1}$$

where

$$f\left(z_i; \zeta_k^{*(\ell)}\right) = \mathrm{N}\left(\ln\left(\mathrm{ad}_i\right); \zeta_{\mathrm{AD},k}^{*(\ell)}, \sigma_{\mathrm{AD}}^{2(\ell)}\right) \mathrm{Bernoulli}\left(\mathrm{mo}_i; \zeta_{\mathrm{MO},k}^{*(\ell)}\right). \tag{A.2}$$

Hence, $s_{i,k}^{(\ell)} = 1$ if $w_i^{(\ell)} = k$ and 0 otherwise;

Step 2: Sample the stick-breaking weights ψ and update π:

Step 2.1: Sample $\psi_k^{(\ell)}$ ($k = 1, \ldots, K-1$, with $\psi_K = 1$):

$$\psi_k^{(\ell)} \sim \mathrm{Beta}\left(\psi_k^{(\ell)};\ 1 + \sum_{i=1}^{n} 1_{\left[W_i^{(\ell)} = k\right]},\ \phi^{(\ell-1)} + \sum_{i=1}^{n} 1_{\left[W_i^{(\ell)} > k\right]}\right) \tag{A.3}$$

Step 2.2: Update π_k:

$$\pi_k^{(\ell)} = \psi_k^{(\ell)} \prod_{j<k} \left(1 - \psi_j^{(\ell)}\right) \tag{A.4}$$

Step 3: Sample $\exp\left(\alpha_j^{(\ell)}\right)$ ($j = 1, 2$) from a conjugate Gamma posterior distribution with shape

$$1 + \sum_{i=1}^{n} d_{j,i}$$

and rate

$$1 + \sum_{i=1}^{n} \exp\left(\beta_j^{(\ell-1)} x_{j,i} + \gamma_{j,w_i^{(\ell)}}^{*(\ell-1)}\right) \frac{\exp\left(\beta_j^{(\ell-1)}\left(t_i + a_i\right)\right) - \exp\left(\beta_j^{(\ell-1)} a_i\right)}{\beta_j^{(\ell-1)}}$$

Step 4: Sample $\beta_j^{(\ell)}$ ($j = 1, 2$) using the acceptance-rejection sampling method:

Step 4.1: Sample β_j^* from a Truncated Normal proposal distribution q with mean $\beta_j^{(\ell-1)}$ and variance $\sigma_{p,j}^{2(\ell-1)}$. The proposal distribution is truncated below at the value of 10^{-5}, since a negative value of β_j would make the model biologically implausible, and above at the value of 5, as its uniform prior. The variance of the proposal distribution is iteratively updated using the Robust Adaptive Metropolis (RAM) algorithm of Vihola (2012), described in Step 4.4;

Step 4.2: Compute the ratio[5]

$$r' = \frac{\prod_{i=1}^{n} \dot{f}\left(t_i \mid x_{j,i}, d_{j,i}, a_i; \alpha_j^{(\ell)}, \beta_j^*, \gamma_{j,w_i^{(\ell)}}^{*(\ell-1)}\right)}{\prod_{i=1}^{n} \dot{f}\left(t_i \mid x_{j,i}, d_{j,i}, a_i; \alpha_j^{(\ell)}, \beta_j^{(\ell-1)}, \gamma_{j,w_i^{(\ell)}}^{*(\ell-1)}\right)} \cdot \frac{q_{10^{-5},5}\left(\beta_j^{(\ell-1)};\ \beta_j^*, \sigma_{p,j}^{2(\ell-1)}\right)}{q_{10^{-5},5}\left(\beta_j^*;\ \beta_j^{(\ell-1)}, \sigma_{p,j}^{2(\ell-1)}\right)} \tag{A.5}$$

where $q_{10^{-5},5}(a; b, c)$ denotes the density of the proposal distribution at the value of a, with mean b and variance equal to c;

Step 4.3: Set:

$$\beta_j^{(\ell)} = \begin{cases} \beta_j^* & \text{w.p. } \min(r', 1) \\ \beta_j^{(\ell-1)} & \text{w.p. } 1 - \min(r', 1) \end{cases} \tag{A.6}$$

Step 4.4: Update the standard deviation of the proposal distribution $\sigma_{p,j}$:

$$\sigma_{p,j}^{(\ell)} = \sigma_{p,j}^{(\ell-1)} \sqrt{1 + \ell^{-0.6}\left(\min(r', 1) - 0.234\right)} \tag{A.7}$$

The parameter 0.6 is chosen in accordance to the recommendation in Vihola (2012), who suggests a value between 0.5 and 1, while 0.234 is the desired acceptance probability, chosen following Roberts et al. (1997);

Step 5: Sample γ^* using the acceptance-rejection sampling method[6]:

Step 5.1: For $k = 1, \ldots, K$ sample $\gamma_k' = \left(\gamma_{1,k}', \gamma_{2,k}'\right)^T$ from a bivariate Normal proposal distribution q with mean $\gamma_k^{*(\ell-1)} = \left(\gamma_{1,k}^{*(\ell-1)}, \gamma_{2,k}^{*(\ell-1)}\right)^T$ and variance-covariance matrix $\mathbf{\Sigma}_{\mathbf{p},\mathbf{k}}^{(\ell-\mathbf{1})}$ (the superscript T denotes the transpose):

Step 5.1.1 Sample $\mathbf{h}^{(\ell)} \sim \text{MVN}\left(\mathbf{h}^{(\ell)}; \mathbf{0}_2, \mathbf{I}_2\right)$, where $\mathbf{0}_2$ is a column vector of zeros, and $\mathbf{I}_2$ denotes the 2×2 identity matrix;

[5] To ease computations we first use logarithms and then we take the exponential of the result. We do the same in Step 5.2.

[6] Individually, the elements of γ^* can be sampled in closed form using a conjugate Gamma distribution for $\exp\left(\gamma_{j,k}^*\right)$. However, in order to induce a dependence between $\gamma_{1,k}^*$ and $\gamma_{2,k}^*$, we opted for a bivariate Normal prior, as from the specification in Section 5.1.

Step 5.1.2 Set $\gamma_k' = \gamma_k^{*(\ell-1)} + \mathbf{L_k^{(\ell-1)}}\mathbf{h}^{(\ell)}$,

where $\mathbf{L}^{(\ell-\mathbf{1})}{}_k$ is the lower triangular matrix denoting the Cholesky decomposition of $\mathbf{\Sigma_{p,k}} = \mathbf{L_k L_k^T}$. The variance-covariance matrix of the proposal distribution is again updated using the RAM algorithm described in Step 5.4 for its multivariate version;

Step 5.2: Compute the ratio:

$$r'' = \frac{\prod\limits_{\{i:w_i^{(\ell)}=k\}} \prod\limits_{j=1}^{2} \dot{f}\left(t_i \mid x_{j,i}, d_{j,i}, a_i; \alpha_j^{(\ell)}, \beta_j^{(\ell)}, \gamma_{j,k}'\right)}{\prod\limits_{\{i:w_i^{(\ell)}=k\}} \prod\limits_{j=1}^{2} \dot{f}\left(t_i \mid x_{j,i}, d_{j,i}, a_i; \alpha_j^{(\ell)}, \beta_j^{(\ell)}, \gamma_{j,k}^{*(\ell-1)}\right)} \cdot \frac{q\left(\gamma_k';\ \gamma_k^{*(\ell-1)}, \mathbf{\Sigma_{p,k}^{(\ell-1)}}\right)}{q\left(\gamma_k^{*(\ell-1)};\ \gamma_k', \mathbf{\Sigma_{p,k}^{(\ell-1)}}\right)} \tag{A.8}$$

Step 5.3: Set:

$$\left(\gamma_{1,k}^{*(\ell)}, \gamma_{2,k}^{*(\ell)}\right) = \begin{cases} \left(\gamma_{1,k}', \gamma_{2,k}'\right) & \text{w.p. } \min\left(r'', 1\right) \\ \left(\gamma_{1,k}^{*(\ell-1)}, \gamma_{2,k}^{*(\ell-1)}\right) & \text{w.p. } 1 - \min\left(r'', 1\right) \end{cases} \tag{A.9}$$

Step 5.4: Update the lower triangular Cholesky factor of $\mathbf{\Sigma_{p,k}}$:

Step 5.4.1: Compute $\mathbf{\Sigma_{p,k}^{(\ell)}}$:

$$\mathbf{\Sigma_{p,k}^{(\ell)}} = \mathbf{L_k^{(\ell-1)}} \left(\mathbf{I_2} + \ell^{-0.6}\left(\min\left(r'', 1\right) - 0.234\right) \frac{\mathbf{h}^{(\ell)}\mathbf{h}^{(\ell)^T}}{||\mathbf{h}^{(\ell)}||^2} \right) \mathbf{L_k^{(\ell-1)T}} \tag{A.10}$$

where $||\mathbf{h}||$ denotes the Euclidean norm of $\mathbf{h}$;

Step 5.4.2: Compute $\mathbf{L_k^{(\ell)}}$ as the Cholesky factor of $\mathbf{\Sigma_{p,k}^{(\ell)}}$;

Step 6: Sample Σ_γ from the conjugate posterior which is the Inv-Wishart distribution with degrees of freedom Λ_1 and scale matrix Λ_2:

$$\Sigma_\gamma^{(\ell)} \sim \text{Inv-Wishart}\left(\Sigma_\gamma^{(\ell)}; \Lambda_1^{(\ell)}, \Lambda_2^{(\ell)}\right) \tag{A.11}$$

where

$$\begin{aligned} \Lambda_1^{(\ell)} &= 7 + \sum_{k=1}^{K} 1_{\left[n_k^{(\ell)}>0\right]} \\ \Lambda_2^{(\ell)} &= 7 \times 0.001 \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix} + \sum_{k=1}^{K} 1_{\left[n_k^{(\ell)}>0\right]} \gamma_k^{*(\ell)} \gamma_k^{*(\ell)T} \end{aligned} \tag{A.12}$$

where $n_k^{(\ell)} = \sum\limits_{i=1}^{n} s_{i,k}^{(\ell)} = \sum\limits_{i=1}^{n} 1_{\left[w_i^{(\ell)}=k\right]}$;

Step 7: Sample $\zeta_{\text{AD},k}^{*(\ell)}$ $(k = 1, \ldots, K)$ from a conjugate $\text{N}\left(\zeta_{\text{AD},k}^{*(\ell)}; \Lambda_{3,k}^{(\ell)}, \Lambda_{4,k}^{(\ell)}\right)$ distribution, where:

$$\Lambda_{3,k}^{(\ell)} = \Lambda_4^{(\ell)} \left(\frac{\sum_{\{i:w_i^{(\ell)}=k\}} \ln(\text{ad}_i)}{\sigma_{\text{AD}}^{2(\ell-1)}} + \frac{m_{\zeta_{\text{AD}}}^{(\ell-1)}}{s_{\zeta_{\text{AD}}}^{2(\ell-1)}} \right)$$

$$\Lambda_{4,k}^{(\ell)} = \left(\frac{n_k^{(\ell)}}{\sigma_{\text{AD}}^{2(\ell-1)}} + \frac{1}{s_{\zeta_{\text{AD}}}^{2(\ell-1)}} \right)^{-1}$$

Step 8: Sample $\zeta_{\text{MO},k}^{*(\ell)}$ $(k = 1, \ldots, K)$ from a conjugate $\text{Beta}\left(\zeta_{\text{MO},k}^{*(\ell)}; \Lambda_{5,k}^{(\ell)}, \Lambda_{6,k}^{(\ell)}\right)$ distribution, where:

$$\Lambda_{5,k}^{(\ell)} = 3 + \sum_{\{i:w_i^{(\ell)}=k\}} \text{mo}_i$$

$$\Lambda_{6,k}^{(\ell)} = 1 + \sum_{\{i:w_i^{(\ell)}=k\}} (1 - \text{mo}_i)$$

Step 9: Sample $m_{\zeta_{\text{AD}}}^{(\ell)}$ from a conjugate $\text{N}\left(m_{\zeta_{\text{AD}}}^{*(\ell)}; \Lambda_7^{(\ell)}, \Lambda_8^{(\ell)}\right)$ distribution, where:

$$\Lambda_7^{(\ell)} = \Lambda_8^{(\ell)} \frac{\sum_{\{k:n_k^{(\ell)}>0\}} \zeta_{\text{AD},k}^{*(\ell)}}{s_{\zeta_{\text{AD}}}^{2(\ell-1)}}$$

$$\Lambda_8^{(\ell)} = \left(1 + \frac{\sum_{k=1}^{K} 1_{\left[n_k^{(\ell)}>0\right]}}{s_{\zeta_{\text{AD}}}^{2(\ell-1)}} \right)^{-1}$$

Step 10: Sample $\sigma_{\text{AD}}^{2(\ell)}$ from a conjugate $\text{Inv-Gamma}\left(\sigma_{\text{AD}}^{2(\ell)}; \Lambda_9^{(\ell)}, \Lambda_{10}^{(\ell)}\right)$ distribution where

$$\Lambda_9^{(\ell)} = 1 + 0.5n$$

$$\Lambda_{10}^{(\ell)} = 1 + \sum_{i=1}^{n} \left(\ln(\text{ad}_i) - \zeta_{\text{AD},w_i^{(\ell)}}^{*(\ell)} \right)^2$$

Step 11: Sample $s^{2(\ell)}_{\zeta\text{AD}}$ from a conjugate Inv-Gamma$\left(s^{2(\ell)}_{\zeta\text{AD}}; \Lambda^{(\ell)}_{11}, \Lambda^{(\ell)}_{12}\right)$ distribution where

$$\Lambda^{(\ell)}_{11} = 2 + 0.5 \sum_{k=1}^{K} 1_{\left[n^{(\ell)}_k > 0\right]}$$
$$\Lambda^{(\ell)}_{12} = 2 + \sum_{\{k: n^{(\ell)}_k > 0\}} \left(\zeta^{*(\ell)}_{\text{AD},k} - m^{(\ell)}_{\zeta\text{AD}}\right)^2$$

Step 12: Sample ϕ by following the steps outlined in Escobar & West (1995):

Step 12.1: Sample $\epsilon \sim \text{Beta}\left(\epsilon; \phi^{(\ell-1)} + 1, n\right)$;

Step 12.2: Sample $B \sim \text{Bernoulli}\left(B; \pi_\epsilon\right)$, where

$$\pi_\epsilon = \frac{\sum_{k=1}^{K} 1_{\left[n^{(\ell)}_k > 0\right]}}{n\left(1 - \ln \epsilon\right) + \sum_{k=1}^{K} 1_{\left[n^{(\ell)}_k > 0\right]}} \tag{A.13}$$

Step 12.3: Sample ϕ:

$$\begin{aligned}\phi^{(\ell)} \sim & 1_{[B=1]}\text{Gamma}\left(\phi^{(\ell)}; 1 + \sum_{k=1}^{K} 1_{\left[n^{(\ell)}_k > 0\right]}, 1 - \ln \epsilon\right) \\ & + 1_{[B=0]}\text{Gamma}\left(\phi^{(\ell)}; \sum_{k=1}^{K} 1_{\left[n^{(\ell)}_k > 0\right]}, 1 - \ln \epsilon\right)\end{aligned} \tag{A.14}$$

B. Results of the competing models

Table B.1: Parameter estimates and corresponding standard errors of the Base Gompertz and Proportional hazard models.

	Base Gompertz		**Proportional hazard**	
Parameter	**Estimate**	**St. err.**	**Estimate**	**St. err.**
α_1	-9.85	0.40	-9.88	0.40
β_1	0.0848	0.0054	0.0845	0.0054
$\delta_{1,1}$	–	–	-0.0101	0.0303
$\delta_{1,2}$	–	–	0.0771	0.0912
α_2	-13.53	0.64	-14.09	0.40
β_2	0.1226	0.0086	0.1279	0.0054
$\delta_{2,1}$	–	–	-0.0570	0.0501
$\delta_{2,2}$	–	–	0.2915	0.1411

C. Computation of AIC and WAIC

Let $\hat{\theta}$ denote the set of the parameter estimates using maximum likelihood for BG and PH, and $\theta^{(\ell)}$ the retained ℓth draw from the posterior distribution of the parameters of the AVDPM approach[7]. AIC and WAIC are calculated as follows:

$$\begin{aligned}
\text{AIC} &= -2 \ln L\left(\hat{\theta} \mid \mathbf{t}, \mathbf{x}, \mathbf{a}, \mathbf{d}, \mathbf{ad}, \mathbf{mo}\right) + 2r \\
\text{WAIC} &= -2 \sum_{i=1}^{n} \ln\left(\frac{1}{H} \sum_{\ell=1}^{H} f\left(t_i \mid x_{1,i}, x_{2,i}, \text{ad}_i, \text{mo}_i, a_i, d_i; \theta^{(\ell)}\right)\right) + 2p_{\text{WAIC}} \\
p_{\text{WAIC}} &= 2 \sum_{i=1}^{n} \left[\ln\left(\frac{1}{H} \sum_{\ell=1}^{H} f\left(t_i \mid x_{1,i}, x_{2,i}, \text{ad}_i, \text{mo}_i, a_i, d_i; \theta^{(\ell)}\right)\right)\right. \\
&\quad \left. - \frac{1}{H} \sum_{\ell=1}^{H} \ln f\left(t_i \mid x_{1,i}, x_{2,i}, \text{ad}_i, \text{mo}_i, a_i, d_i; \theta^{(\ell)}\right)\right]
\end{aligned}$$

where $L\left(\hat{\theta} \mid \mathbf{t}, \mathbf{x}, \mathbf{a}, \mathbf{d}, \mathbf{ad}, \mathbf{mo}\right)$ is the likelihood function of the parameters given the data, r denotes the number of parameters of the model under analysis, H the number of draws from the posterior distribution, and

$$\begin{aligned}
& f\left(t_i \mid x_{1,i}, x_{2,i}, \text{ad}_i, \text{mo}_i, a_i, d_i; \theta^{(\ell)}\right) \qquad \text{(C.1)} \\
&= \frac{\sum_{k=1}^{K} \pi_k^{(\ell)} \left[\prod_{j=1}^{2} \dot{f}\left(t_i \mid x_{j,i}, d_{j,i}, a_i; \alpha_j^{(\ell)}, \beta_j^{(\ell)}, \gamma_{j,k}^{*(\ell)}\right)\right] f\left(z_i; \zeta_k^{*(\ell)}\right)}{\sum_{k=1}^{K} \pi_k^{(\ell)} f\left(z_i; \zeta_k^{*(\ell)}\right)}
\end{aligned}$$

In this equation we marginalized the joint distribution of the time to event with respect to the common covariates AD and MO, in order to make the three models comparable among them.

References

Akaike, H. (1974), 'A new look at the statistical model identification', *IEEE Transactions on Automatic Control* **19**(6), 716–723.

Betancourt, M. J. (2017), 'Identifying bayesian mixture models'.
URL: *https://betanalpha.github.io/assets/case_studies/identifying_mixture_models.html*

Blackwell, D. & MacQueen, J. B. (1973), 'Ferguson distributions via polya urn schemes', *Ann. Statist.* **1**(2), 353–355.
URL: *https://doi.org/10.1214/aos/1176342372*

[7] The sample from the posterior distribution which is left after burn-in and thinning of the chain resulting from the MCMC sampler.

Carriere, J. F. (2000), 'Bivariate survival models for coupled lives', *Scandinavian Actuarial Journal* **2000**(1), 17–32.
URL: *https://doi.org/10.1080/034612300750066700*

Denuit, M. & Cornet, A. (1999), Multilife premium calculation with dependent future lifetimes.

Denuit, M., Dhaene, J., Le Bailly De Tilleghem, C. & Teghem, S. (2001), 'Measuring the impact of a dependence among insured life lengths', *Belgian Actuarial Bulletin* pp. 18–39.

Deresa, N., Van Keilegom, I. & Antonio, K. (2022), 'Copula-based inference for bivariate survival data with left truncation and dependent censoring', *Insurance: Mathematics and Economics* **107**, 1–21.
URL: *https://www.sciencedirect.com/science/article/pii/S0167668722000841*

Dickson, D. C. M., Hardy, M. R. & Waters, H. R. (2013), *Actuarial Mathematics for Life Contingent Risks*, International Series on Actuarial Science, Cambridge University Press.

Dufresne, F., Hashorva, E., Ratovomirija, G. & Toukourou, Y. (2018), 'On age difference in joint lifetime modelling with life insurance annuity applications', *Annals of Actuarial Science* **12**(2), 350–371.

Escobar, M. D. & West, M. (1995), 'Bayesian density estimation and inference using mixtures', *Journal of the American Statistical Association* **90**(430), 577–588.
URL: *https://www.tandfonline.com/doi/abs/10.1080/01621459.1995.10476550*

Ferguson, T. S. (1973), 'A bayesian analysis of some nonparametric problems', *Ann. Statist.* **1**(2), 209–230.
URL: *https://doi.org/10.1214/aos/1176342360*

Frees, E. W., Carriere, J. & Valdez, E. (1996), 'Annuity valuation with dependent mortality', *The Journal of Risk and Insurance* **63**(2), 229–261.
URL: *http://www.jstor.org/stable/253744*

Gorfine, M. & Hsu, L. (2011), 'Frailty-based competing risks model for multivariate survival data', *Biometrics* **67**(2), 415–426.
URL: *https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1541-0420.2010.01470.x*

Heinz, D. (2014), Nonparametric mixed membership models, *in* 'Handbook of Mixed Membership Models and Their Applications', Handbook of Modern Statistical Methods, Chapman & Hall, pp. 89–116.

Huang, X. & Wolfe, R. A. (2002), 'A frailty model for informative censoring', *Biometrics* **58**(3), 510–520.

Ishwaran, H. & James, L. F. (2001), 'Gibbs sampling methods for stick-breaking priors', *Journal of the American Statistical Association* **96**(453), 161–173.
URL: *https://doi.org/10.1198/016214501750332758*

Kalli, M., Griffin, J. & Walker, S. (2011), 'Slice sampling mixture models', *Statistics and Computing* **21**, 93–105.

Lancaster, T. (1979), 'Econometric methods for the duration of unemployment', *Econometrica* **47**(4), 939–956.
URL: *http://www.jstor.org/stable/1914140*

Lo, A. Y. (1984), 'On a Class of Bayesian Nonparametric Estimates: I. Density Estimates', *The Annals of Statistics* **12**(1), 351 – 357.
URL: *https://doi.org/10.1214/aos/1176346412*

Marin, J.-M., Mengersen, K. L. & Robert, C. (2005), Bayesian modelling and inference on mixtures of distributions, *in* D. Dey & C. Rao, eds, 'Handbook of Statistics: Volume 25', Elsevier.

Müller, P., Erkanli, A. & West, M. (1996), 'Bayesian curve fitting using multivariate normal mixtures', *Biometrika* **83**(1), 67–79.
URL: *https://doi.org/10.1093/biomet/83.1.67*

Papaspiliopoulos, O. & Roberts, G. O. (2008), 'Retrospective markov chain monte carlo methods for dirichlet process hierarchical models', *Biometrika* **95**(1), 169–186.
URL: *http://www.jstor.org/stable/20441450*

Parkes, C. M., Benjamin, B. & Fitzgerald, R. G. (1969), 'Broken heart: A statistical study of increased mortality among widowers', *BMJ* **1**(5646), 740–743.
URL: *https://www.bmj.com/content/1/5646/740*

R Core Team (2013), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.
URL: *http://www.R-project.org/*

Richards, S. J. (2008), 'Applying survival models to pensioner mortality data', *British Actuarial Journal* **14**(2), 257–303.

Roberts, G. O., Gelman, A. & Gilks, W. R. (1997), 'Weak convergence and optimal scaling of random walk metropolis algorithms', *Ann. Appl. Probab.* **7**(1), 110–120.

Sethuraman, J. (1994), 'A constructive definition of dirichlet priors', *Statistica Sinica* **4**(2), 639–650.
URL: *http://www.jstor.org/stable/24305538*

Ungolo, F., Christiansen, M. C., Kleinow, T. & MacDonald, A. S. (2019), 'Survival analysis of pension scheme mortality when data are missing', *Scandinavian Actuarial Journal* **2019**(6), 523–547.
URL: *https://doi.org/10.1080/03461238.2019.1580610*

Ungolo, F., Kleinow, T. & Macdonald, A. S. (2020), 'A hierarchical model for the joint mortality analysis of pension scheme data with missing covariates', *Insurance: Mathematics and Economics* **91**, 68 – 84.
URL: *http://www.sciencedirect.com/science/article/pii/S0167668720300032*

Ungolo, F. & van den Heuvel, E. R. (2022), 'Inference on latent factor models for informative censoring', *Statistical Methods in Medical Research* **0**(0), 09622802211057290. PMID: 35077263.
URL: *https://doi.org/10.1177/09622802211057290*

Ungolo, F. & van den Heuvel, E. R. (2023), 'A dirichlet process mixture regression model for the analysis of competing risk events', *Available at SSRN: https://ssrn.com/abstract=4355485* .

Vaupel, J. W., Manton, K. G. & Stallard, E. (1979), 'The impact of heterogeneity in individual frailty on the dynamics of mortality', *Demography* **16**(3), 439–454.
URL: *http://www.jstor.org/stable/2061224*

Vaupel, J. W. & Yashin, A. I. (1985), 'The deviant dynamics of death in heterogeneous populations', *Sociological Methodology* **15**, 179–211.
URL: *http://www.jstor.org/stable/270850*

Vihola, M. (2012), 'Robust adaptive Metropolis algorithm with coerced acceptance rate', *Statistics and Computing* **22**, 997–1008.

Walker, S. G. (2007), 'Sampling the dirichlet mixture model with slices', *Communications in Statistics - Simulation and Computation* **36**(1), 45–54.
URL: *https://doi.org/10.1080/03610910601096262*

Watanabe, S. (2009), *Algebraic Geometry and Statistical Learning Theory*, Cambridge University Press, New York, NY, USA.

Wienke, A. (2014), *Frailty Models*, John Wiley Sons, Ltd.
URL: *https://onlinelibrary.wiley.com/doi/abs/10.1002/9781118445112.stat06941*

Yashin, A. I. & Iachine, I. A. (1995), 'Genetic analysis of durations: Correlated frailty model applied to survival of danish twins', *Genetic Epidemiology* **12**.

Youn, H. & Shemyakin, A. (1999), 'Statistical aspects of joint life insurance pricing', *Statistical aspects of joint life insurance pricing* .

University of New South Wales

An Augmented Variable Dirichlet Process Mixture model for the analysis of dependent lifetimes - Supplementary material

Francesco Ungolo[1,2,*]

[1]*School of Risk and Actuarial Studies, University of New South Wales, Kensington NSW 2052, Australia*
[2]*ARC Centre of Excellence in Population Ageing Research, University of New South Wales, Kensington NSW 2052, Australia*
[*]*Corresponding author: Francesco Ungolo, f.ungolo@unsw.edu.au*

April 25, 2023

1 Data cleaning operations

The data preparation steps are listed below:

1. Eliminate same-gender couples, as in Frees et al. (1996) and Deresa et al. (2022);
2. Eliminate records of people unlikely to be couples, such as those including a member with an age lower than 10. This may be an error in the data;
3. Check of records of people alive at the beginning of the observation period;
4. Elimination of duplicated records.

2 Joint life annuity factor

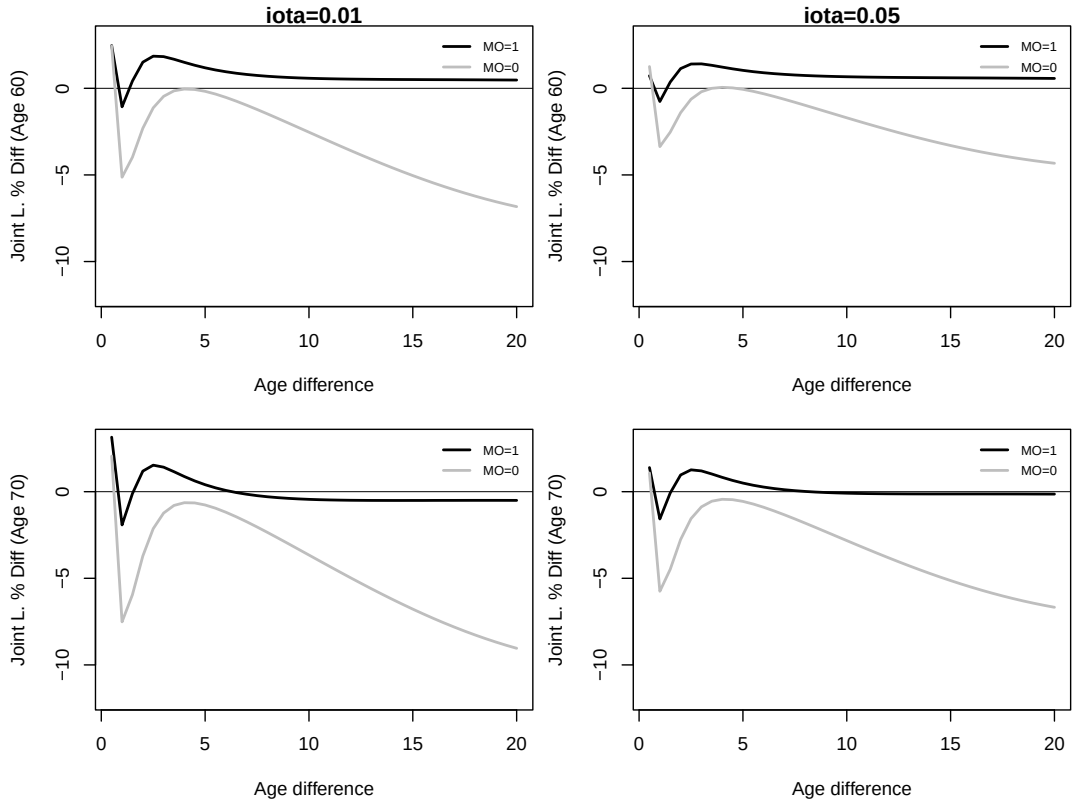

Figure 2.1: Percentage difference between the JL annuity factor under AVDPM and PH for different values of age difference (x-axis) and MO when the oldest member is aged 60 (top panel) and 70 (bottom panel), for $\iota = 1\%$ (left panel) and $\iota = 5\%$ (right panel).

References

Deresa, N., Van Keilegom, I. & Antonio, K. (2022), 'Copula-based inference for bivariate survival data with left truncation and dependent censoring', *Insurance: Mathematics and Economics* **107**, 1–21.
URL: *https://www.sciencedirect.com/science/article/pii/S0167668722000841*

Frees, E. W., Carriere, J. & Valdez, E. (1996), 'Annuity valuation with dependent mortality', *The Journal of Risk and Insurance* **63**(2), 229–261.
URL: *http://www.jstor.org/stable/253744*