



**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 and Patrick J. Laub

---

This paper can be downloaded without charge from the ARC Centre of Excellence in Population Ageing Research Working Paper Series available at [www.cepar.edu.au](http://www.cepar.edu.au)

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

---

Francesco Ungolo<sup>1,2,\*</sup> and Patrick J. Laub<sup>1</sup>

<sup>1</sup>*School of Risk and Actuarial Studies, University of New South Wales, Kensington NSW 2052, Australia*

<sup>2</sup>*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*

May 23, 2024

The analysis of insurance and annuity products issued on multiple lives requires 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 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 inferential approach allows to account for right censoring and left truncation, which are common features of data in survival analysis. The model shows improved in-sample and out-of-sample performance compared to traditional approaches assuming independent lifetimes, and offers additional insights into the determinants of the dependence between lifetimes and their impact on joint and last survivor annuity prices.

**Keywords:** Dependent lifetimes, Survival Analysis, Dirichlet Process, 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 assumption of independent lifetimes 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 socioeconomic characteristics, which affect their living standards and their exposure to similar risks (Denuit and Cornet, 1999; Denuit et al., 2001).

Furthermore, the use of the simplistic independence assumption can have a material impact on actuarial valuations. Denuit and Cornet (1999) use a Markov model where the force of mortality depends on marital status, and show how the premium of a widow pension annuity is 10 per cent lower compared to the case where lifetime independence between husband and wife is assumed. Frees et al. (1996) employ a one-parameter copula model, demonstrating 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, with copula models being among the most employed ones. The aforementioned paper of Frees et al. (1996) analyses a one-parameter Frank copula with Gompertz marginals for the lifetimes of the male and the female within a couple. Carriere (2000) extends this analysis by considering other marginal distributions for the future lifetime, as well as other types of copulas, and Deresa et al. (2022) focus on the statistical properties of copula models in the presence of left-truncation and dependent censoring. Youn and Shemyakin (1999) were the first to account for covariates when modelling dependence. More precisely, they employ a Gumbel copula with Weibull marginals, with dependence parameter which accounts for the age-difference between the spouses. They find that this feature is found to influence the statistical association between husband and wife lifetimes. Dufresne et al. (2018) use Gompertz marginals for the time to death of the male and of the female within a couple, and observe how the gender of the eldest partner has also an influence on the lifetime dependence.

An alternative to copulas is given by models using random effects (or frailty components, see Vaupel et al., 1979) to capture the dependence between lifetimes. This means that conditional on a latent variable, then lifetimes are independently distributed. For example, Yashin and 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 and Wolfe (2002) address this problem by assuming a Cox proportional hazard model for the hazard function of two random variables, whose linear term includes a normally distributed log-frailty component. Gorfine and 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 and van den Heuvel \(2022, 2024\)](#) overcome this potential misspecification issue by using a multivariate random effect with a discrete distribution and unknown number of levels. However, 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 latent variable, whose discrete probability 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. The model parameters are estimated by means of a fully Bayesian analysis, which may include the information available to the researcher. In addition, we show how this approach can easily account for right censoring and left truncation.

This paper is organized as follows: Section [2](#) briefly introduces the Dirichlet Process and the Dirichlet Process Mixture model, and Section [3](#) presents the 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 [5](#) describes the Bayesian inferential framework for the analysis of the AVDPM model. 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, and how it compares with approaches assuming independent lifetimes. 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 probability model for the random masses of a discrete distribution  $G$ . We thus say that  $G$  follows a Dirichlet Process on the measurable space  $(\Gamma, \mathcal{G})$  if for any measurable partition  $(W_1, \dots, W_q)$  of  $\Gamma$ , then

$$(G(W_1), \dots, G(W_q)) \sim \text{Dirichlet}(\phi G_0(W_1), \dots, \phi G_0(W_q))$$

which we write  $G \sim \text{DP}(G_0, \phi)$ . Hence, a random draw from the DP yields an almost sure discrete probability distribution over a countably infinite number of points drawn independently from a *base distribution* denoted as  $G_0$ , which specifies the DP together with the *concentration parameter*  $\phi$  (see [Gelman et al., 2013](#), Chapter 23 and [De Iorio et al., 2004](#)). The  $\phi$  parameter 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.

Sethuraman (1994) outlines a construction of  $G$  as a mixture distribution with a countably infinite number of components, indexed by  $k$ :

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

where  $\gamma_k \stackrel{\text{i.i.d.}}{\sim} G_0$  for  $k = 1, 2, \dots$ , and  $\delta_{\gamma_k}$  denotes the Dirac measure which assigns unitary mass if  $\gamma = \gamma_k$  and zero otherwise. The mixture weights  $\pi_k$  are randomly generated using the so called *stick breaking procedure* (SBP), which rescales a set of i.i.d. random variables  $\psi_k \stackrel{\text{i.i.d.}}{\sim} \text{Beta}(1, \phi)$  as follows:

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

where  $\psi_{1:k} = (\psi_1, \dots, \psi_k)$ . The ‘‘stick breaking’’ definition refers to 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 a flexible model for the probability distribution of a random variable  $T_i$ , eventually multivariate, for the  $i$ th unit, whose density  $f(\cdot; \beta, \gamma_i)$  is indexed by a *global* parameter vector  $\beta$ , common to all units, and by a unit-specific parameter vector  $\gamma_i$ . Unit-specific parameters, commonly referred as random effects in biostatistics, are introduced in a probability model to characterize the heterogeneity among the units due to unobservable variables.

To achieve this goal, we assume that  $\gamma_i$  are drawn from a discrete distribution  $G$  which follows a Dirichlet Process. Therefore, convolving the density of  $T_i$  with  $G \sim \text{DP}(G_0, \phi)$ , yields a Dirichlet Process Mixture (DPM) model (Lo, 1984):

$$f(t | \beta, G) = \int_{\Omega_\gamma} f(t | \beta, \gamma) dG(\gamma) = \sum_{k=1}^{\infty} \pi_k f(t | \beta, \gamma_k) \quad (2.3)$$

where  $\Omega_\gamma$  denotes the sample space of  $\gamma$ .

The DPM model hereby defined, allows 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 the case where we specify a parametric model for  $\gamma_i$ , such as the Normal distribution.

This DP construction is flexible since despite its apparent nonparametric nature, it regularizes the distribution of  $G$  towards a simple parametric form  $G_0$  through the concentration parameter  $\phi$ .

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  $\gamma$ . Let  $s_i = k$  to indicate that the  $i$ th unit belongs to the  $k$ th cluster characterized by the parameter  $\gamma_k^*$ . We use the superscript  $*$  to denote the common values of  $\gamma_i$  across the  $n$  units in the sample. The clustering procedure, tuned by the parameter  $\phi$ , 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 observe the units in a sequence, these are more likely to be in a certain class with a probability which depends on the number of units already therein, or to belong to a newly created class with a probability which depends on  $\phi$  (see [Blackwell and MacQueen, 1973](#) for a characterization in terms of the Pólya urn distribution).

The use of Dirichlet Processes can also be seen as a method to infer the number of components of a mixture distribution, as opposed to strategies based on appropriate model selection criteria (see [Ungolo and van den Heuvel, 2022](#) for a discussion). Their advantage is to avoid the specification and the estimation of several models with different numbers of mixture components, which can be time consuming.

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

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

### 3. The Augmented Variable DPM model

This section introduces the Augmented Variable Dirichlet Process Mixture (AVDPM) model for the analysis of non-exchangeable joint dependent lifetimes 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, socioeconomic 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 lifetimes of the primary and secondary head of an insurance policy, families with husband, wife and one child, and so on. 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 groups with different numbers of exchangeable lifetimes.

Let  $T_i = (T_{i,1}, T_{i,2})$ , where  $T_{i,1}$  and  $T_{i,2}$  denote the random future lifetime of husband and wife respectively for the  $i$ th couple, with individual-specific vector of characteristics  $x_{i,1}$  and  $x_{i,2}$  (such as age, medical status, and so on) fixed and observable, and couple-specific covariates vector  $Z_i$ , which can include for example household income and geo-demographic profile (an indicator of the socioeconomic status, see [Ungolo et al., 2019](#)), which we treat as a random variable. Therefore,  $Z_i$  includes any features which can explain the statistical association between the lifetimes of the husband and of the wife. For example, the household income can explain the heterogeneity in the longevity profile of each couple member, 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 socioeconomic characteristics of a couple.

For the  $i$ th couple we assume that conditional on a couple-specific bivariate random effect  $\gamma_i = (\gamma_{i,1}, \gamma_{i,2})$ , then  $T_{i,1}$  and  $T_{i,2}$  are independently distributed. This is for example the approach followed in Ungolo and van den Heuvel (2022) when analyzing a joint model for the time to event and the time to censoring, and by Ungolo and van den Heuvel (2024) where they develop a joint model for the time to competing risk events.

The AVDPM model “augments” the joint probability distribution of  $(T_{i,1}, T_{i,2})$  by specifying a joint probability model for  $(T_{i,1}, T_{i,2}, Z_i)$ :

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

where  $\beta_j$  is the parameter specific to the future lifetime density of the husband ( $j = 1$ ) or of the wife ( $j = 2$ ),  $\zeta_i$  is the couple-specific parameter vector indexing the distribution of  $Z_i$  (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, and write  $G \sim \text{DP}(G_0, \phi)$ .

For simplicity, we assume  $G_0 = G_{0,\gamma} \times G_{0,\zeta}$ , that is  $\gamma$  and  $\zeta$  are independently distributed in the base distribution. This specification still induces a dependence between  $\gamma$  and  $\zeta$  through the concentration parameter  $\phi$ . A further simplification can be to specify  $G_{0,\gamma} = G_{0,\gamma_1} \times G_{0,\gamma_2}$ , leaving the distribution of  $(\gamma_{i,1}, \gamma_{i,2}, \zeta_i)$  fully tuned by  $\phi$ . However, the former approach allows to account for the prior information from the researcher about the joint distribution of the parameters, including the statistical association between  $\gamma_1$  and  $\gamma_2$ , as more reasonable for the analysis of male-female couples.

The joint distribution of equation (3.1) allows to capture the dependence between  $T_{i,1}$  and  $T_{i,2}$  and between  $(T_{i,1}, T_{i,2})$  and  $Z_i$  through the joint distribution of  $(\gamma_{i,1}, \gamma_{i,2}, \zeta_i)$ . In this way, we distinguish between the covariates which directly affect the individual lifetimes  $x$  from those features which are couple-specific. For example, in the analysis of couple member lifetimes, Deresa et al. (2022) include the couple specific covariates, such as the age difference, within the set of individual covariates. In this paper we do not model the randomness of the individual covariates, since we are rather interested in the effect of the common covariates on the lifetime dependence.

An additional feature of this factorization, which enhances the flexibility of this joint model is that we can specify different parametric models for  $T_{i,1}$  and  $T_{i,2}$ .

As in equation (2.3), we can rewrite the density in (3.1) as a mixture distribution with

an infinite number of components, obtaining the Augmented Variable DPM model:

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

where  $\pi_k$  ( $k = 1, 2, \dots$ ) is the mixture weight characterized by the stick-breaking procedure described in Section 2, and the superscript  $*$  denotes the unique values of  $\gamma_{i,j}$  and  $\zeta_i$ .

The flexibility of this factorization allows to understand the impact of  $Z_i$  on the distribution of  $T_{i,j}$  (or of  $(T_{i,1}, T_{i,2})$ ) by using standard probability calculus:

$$f(t_{i,j} \mid x_{i,j}, z_i; \beta_j, \gamma_{j,j}^*, \zeta^*) = \frac{\sum_{k=1}^{\infty} \pi_k f(t_{i,j} \mid x_{i,j}; \beta_j, \gamma_{k,j}^*) f(z_i \mid \zeta_k^*)}{\sum_{k=1}^{\infty} \pi_k f(z_i \mid \zeta_k^*)}. \quad (3.3)$$

where  $\gamma_{j,j}^* = (\gamma_{1,j}^*, \gamma_{2,j}^*, \dots)$  and  $\zeta^* = (\zeta_1^*, \zeta_2^*, \dots)$ .

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

$$S_{\overline{x_1, x_2}}(t \mid z) := \mathbb{P}(T_{\overline{1:2}} > t \mid x_1, x_2, z) = 1 - \mathbb{P}(T_1 < t, T_2 < t \mid x_1, x_2, z), \quad (3.4)$$

where

$$\mathbb{P}(T_1 < t, T_2 < t \mid x_1, x_2, z) = \frac{\sum_{k=1}^{\infty} \pi_k f(z \mid \zeta_k^*) \left[ \prod_{j=1}^2 \int_0^t f(u \mid x_j; \beta_j, \gamma_{k,j}^*) du \right]}{\sum_{k=1}^{\infty} \pi_k f(z \mid \zeta_k^*)}$$

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  $T_{1,2} = \min(T_1, T_2)$ .

The DPM naturally clusters each couple into different classes. If we say that  $W$  is the random class allocation for a couple, we can calculate the probability of belonging to a certain class  $k$  conditional on the value of  $Z$ :

$$\mathbb{P}(W = k \mid Z = z) = \frac{\pi_k f(z; \zeta_k^*)}{\sum_{h=1}^{\infty} \pi_h f(z; \zeta_h^*)}. \quad (3.5)$$



In this way, we can understand how the various couples belonging to the  $k$ th class can share the same mortality profile and husband-wife mortality dependence relationships.

Compared to earlier literature in the field, the AVDPM model extends the mixture model analysed by Ungolo and 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. Ungolo and van den Heuvel (2024) overcome this issue by assuming that the random component is drawn from a discrete distribution following a Dirichlet Process as in this paper. However, their joint lifetime model does not account for the statistical association among competing risks due to common factors.

To the best of our knowledge, the aforementioned papers (Section 1) of Youn and Shemyakin (1999) and Dufresne et al. (2018) are the sole contributions accounting for the common variable  $Z$  (the absolute value of the age difference between the spouses and gender of the oldest partner) within the copula dependence parameter.

The factorization implied by the AVDPM model can also find application for the analysis of the time to competing causes of death, where  $Z$  accounts for those genetic factors affecting the dependence between the causes. Alternatively, the AVDPM model can be used to jointly model dependent frequency and/or severity of claims in non-life and health insurance by type of event or line of business, without necessarily specifying a strong parametric assumption on the dependence, as can be the case with 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 by Deresa et al. (2022) and Dufresne et al. (2018) among others. After some data processing operations, briefly described in Section 1 of the Supplementary Material, we have information about 13,482 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. This dataset contains a large number of censored units: indeed, we only observe 1,424 deaths among males and 500 deaths among females, while the remaining units are all right censored. In addition, most couples' lifetime data are 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 65.54 for males and 62.64 for females), which we denote as  $x_{i,j}$  for the  $j$ th member of the  $i$ th couple. As in Dufresne et al. (2018) and Youn and Shemyakin

(1999), we consider also the age-difference (mean of 2.91 years) through the random variable  $Z_i^A = \ln(|X_{i,1} - X_{i,2}|)$  and the indicator variable  $Z_i^M$ , which is equal to 1 if the male is older than the female and 0 otherwise (the male is the oldest policyholder in 76.9% 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  $Z_i^M$  covariate has an influence in the relationship between husband and wife, consequently affecting their lifetime dependence. Therefore, we will consider  $Z_i^A$  and  $Z_i^M$  as the two elements of the couple-specific covariates,  $Z_i = (Z_i^A, Z_i^M)$ .

We randomly split the dataset into a training set, corresponding to 75% of the policyholders within the dataset (10,112 units), and use the remaining 25% to assess the out of sample performance of the model (3,370 units).

Following Frees et al. (1996) and Carriere (2000) we specify a Gompertz-type hazard function to characterize the probability distribution of the male ( $j = 1$ ) and the female ( $j = 2$ ) lifetime within the  $i$ th policy:

$$\mu(t \mid x_{i,j}; \alpha_j, \beta_j, \gamma_{i,j}) = \exp(\alpha_j + \beta_j(x_{i,j} - \bar{x} + t) + \gamma_{i,j}), \quad (4.1)$$

where  $\bar{x} = 70$  is set in order to decrease the posterior correlation between  $\alpha_j$  and  $\beta_j$ , which improves the convergence of the MCMC sampler. The choice of  $\bar{x}$  is based on an indicative average value of  $x$  as calculated across males and females. For example, Ungolo et al. (2020) set  $\bar{x} = 77.5$  since they analysed a pension scheme dataset with older scheme members.

We consider the male-female lifetime dependence by assuming a couple-specific frailty term  $\exp(\gamma_{i,j})$ , and a joint model for  $(\gamma_{i,1}, \gamma_{i,2})$ . The parameter  $\alpha_j$  denotes the log-baseline hazard function and  $\beta_j$  measures the log-linear increase in the hazard function due to the individual age. The probability density function of  $T_{i,j}$  can be written as:

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

For the couple-specific covariates, we assume that  $Z_i^A \sim N(\zeta_i^A, \sigma_A^2)$  and  $Z_i^M \sim \text{Bernoulli}(\zeta_i^M)$ . The density of  $T_{i,j}$  as a function of  $x_{i,j}$  and  $z_i$  follows from equation (3.3). In this way, we can easily observe how the resulting joint model for  $(T_{i,1}, T_{i,2}, Z_i)$  has an additional layer of flexibility, since the effect of  $Z_i$  on the hazard function (through dependent  $\gamma_{i,j}$ ,  $\zeta_i^A$  and  $\zeta_i^M$ ) is not necessarily proportional nor monotone.

The  $i$ th couple-specific parameters are drawn from a multivariate random discrete distribution  $G$ ,  $(\gamma_{i,1}, \gamma_{i,2}, \zeta_i^A, \zeta_i^M) \sim G$ , where  $G$  is a draw from a Dirichlet Process with concentration parameter  $\phi$  and base measure  $G_0$ :

$$G_0 = \underbrace{g(\gamma_{i,1}, \gamma_{i,2} \mid \mathbf{0}, \Sigma_\gamma)}_{\text{MVN}(\mathbf{0}, \Sigma_\gamma)} \underbrace{g(\zeta_i^A \mid m_A, s_A^2)}_{\text{N}(m_A, s_A^2)} \underbrace{g(\zeta_i^M \mid 13.31, 4.44)}_{\text{Beta}(13.31, 4.44)}. \quad (4.3)$$

This 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 where possible, whilst keeping the model simple and flexible enough to capture the complex features of the data.

The parameters of the Beta distribution for  $\zeta_i^M$  follows from an elicitation process where we specify a weakly informative distribution. Indeed, we expect that for most contracts the male couple member is the oldest in the couple, therefore we set  $\zeta_i^M$  to have mean 0.75 and standard deviation equal to 0.1 in the base distribution. In our analysis we have also tried a prior distribution with a standard deviation of 0.05, which did not significantly impact the final results in terms of the convergence of the chain towards a stationary distribution, posterior distribution of the other parameters, and cluster allocation of the couples.

For all the other parameters of the base distribution, we assume these are random to enhance the robustness of the inference, as from the prior elicitation process and the sensitivity analysis described in Section 5.1.

## 5. Inference

First of all, we approximate the mixture distribution of equation (3.2) by setting an upper bound  $K = 25$  to the number of mixture components as in Ungolo and van den Heuvel (2024), which results in the truncated SBP of Ishwaran and James (2001). This simplifies the implementation of the Markov Chain Monte Carlo (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 and Roberts (2008).

### 5.1. Prior elicitation, specification and sensitivity

In this paper we carefully consider the specification of the prior distribution of the parameters. This is motivated by the number of layers of the hierarchical model hereby proposed, and the large number of coarsened observations in this dataset, which can hinder the convergence of any MCMC scheme, and cause poor mixing whenever different starting values are used.

As already observed by Dunson (2010) and Beraha et al. (2023), specifying base distributions with large variances, implies that all couples will tend to be allocated in the same mixture component, which does not capture the heterogeneity in the data.

Therefore, we specify at least weakly informative prior distributions, especially for the lowest levels of the model. Where possible, we assume that the parameters (or groups of parameters) are pairwise independently distributed, and use conditionally conjugate prior distributions in order to facilitate the computation of the posterior distribution.

Previous analyses of mixtures involving proportional hazard models (see for example Ungolo et al. 2020 and Ungolo and van den Heuvel 2022) show how three components can be useful to explain the data. Given the sample size of this dataset and the dimensionality of  $G$ , we specify a conjugate Gamma(6, 12) prior for  $\phi$  (Escobar and West, 1995), which yields a right-skewed prior probability distribution for the number of occupied components (the number of classes with at least one observation) with mean 5.5 and mode equal to 4. Conversely, a lesser informative Gamma(1, 1) prior increases the

number of occupied components a priori and a posteriori, without any impact on the posterior inference for the other parameters.

The parameter  $\Sigma_\gamma$  follows a conjugate Inverse-Wishart distribution with positive-definite scale matrix  $\lambda \times \begin{bmatrix} 1 & 0.6 \\ 0.6 & 1 \end{bmatrix}$  and 5 degrees of freedom. We set the scalar  $\lambda = 2$  in order to strengthen the prior assumption towards a covariance matrix which induces a positive correlation between  $\gamma_{i,1}$  and  $\gamma_{i,2}$ . In our prior-posterior sensitivity analysis, we notice how the posterior distribution of  $\Sigma_\gamma$  turns out to be sensitive with respect to  $\lambda$ , hence the specification of an informative prior distribution plays an important role in the inference for this parameter. The posterior distribution for all other parameters is robust with respect to this assumption.

The location parameter of  $\zeta_i^A$ ,  $m_A$  is drawn from a conjugate Normal distribution with mean 3 and variance 0.5, while the scale  $s_A^2$  is assumed to follow an Inverse-Gamma(3, 0.5) distribution, ensuring that mean and variance of  $s_A^2$  are defined. Our sensitivity analysis, which used a less informative prior on these two parameters, showed how the posterior distribution of  $m_A$  is robust with respect to the specification of a weaker prior distribution with a larger variance, while the opposite holds for  $s_A^2$ . In the latter case, we observe a shift in the posterior distribution for  $m_A$  towards larger values, and the posterior density of  $s_A^2$  is very flat and centered towards very large values. This evidence suggests the necessity to carefully specify the prior distribution for  $s_A^2$ .

We complete the full specification of the prior distribution by assuming that  $\exp(\alpha_j) \sim \text{Gamma}(1, 1)$  (which ensures conditional prior-posterior conjugacy) and  $\beta_j$  ( $j = 1, 2$ ) follows *a priori* a truncated normal distribution with mean 0.1, variance 0.25, bounded below at zero to ensure biologically reasonable mortality rates, which should increase with age. The location parameter was chosen on the basis of previous analyses of similar models, which estimated a value of  $\beta$  around 0.10 (see for example Ungolo et al., 2020 and Richards, 2008). Finally, we assume that  $\sigma_A^2 \sim \text{Inv-Gamma}(2, 1)$ .

Our sensitivity analysis shows how the posterior inference for  $\beta$  and  $\sigma_A^2$  is robust with respect to the specification of weaker prior distributions, and that the posterior distribution of the other model parameters are unaffected by this. On the other hand, the posterior distribution for  $\alpha$  can be affected by its posterior correlation with the parameters  $\gamma_{k,1}^*$  and  $\gamma_{k,2}^*$ , as we discuss in Section 6.

## 5.2. Likelihood

Let  $d_{i,j}$  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 right censored). We assume that the censoring mechanism can be ignored, since the censoring event for a couple member is caused by the end of the observation period (Type I censoring). Furthermore, we assume that the individual times to event for each couple member 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 level of left

truncation 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. Note that  $a_i$  is equal to zero if the contract starts during the observation period.

Let  $\mathbf{t} = (t_{1,1}, t_{1,2}, \dots, t_{n,1}, t_{n,2})$ ,  $\mathbf{x} = (x_{i,1}, x_{i,2}, \dots, x_{n,1}, x_{n,2})$ ,  $\mathbf{d} = (d_{1,1}, d_{1,2}, \dots, d_{n,1}, d_{n,2})$ ,  $\mathbf{z}^{\mathbf{A}} = (z_1^{\mathbf{A}}, \dots, z_n^{\mathbf{A}})$ ,  $\mathbf{z}^{\mathbf{M}} = (z_1^{\mathbf{M}}, \dots, z_n^{\mathbf{M}})$ ,  $\mathbf{a} = (a_1, \dots, a_n)$ ,  $\alpha = (\alpha_1, \alpha_2)$ ,  $\beta = (\beta_1, \beta_2)$ ,  $\psi = (\psi_1, \dots, \psi_{K-1})$ ,  $\gamma^* = (\gamma_{1,1}^*, \gamma_{1,2}^*, \dots, \gamma_{K,2}^*)$ ,  $\zeta^{A*} = (\zeta_1^{A*}, \dots, \zeta_K^{A*})$  and  $\zeta^{M*} = (\zeta_1^{M*}, \dots, \zeta_K^{M*})$ . We introduce the latent indicator variable  $s_{i,k}$ , which is equal to 1 if the  $i$ th couple belongs to the  $k$ th 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  $\mathbf{t}$ ,  $\mathbf{x}$ ,  $\mathbf{a}$ ,  $\mathbf{d}$ ,  $\mathbf{z}^{\mathbf{A}}$ ,  $\mathbf{z}^{\mathbf{M}}$  and the latent allocation variable  $\mathbf{s} = (s_{1,1}, \dots, s_{n,K})$  is given by:

$$L(\alpha, \beta, \gamma^*, \zeta^{A*}, \zeta^{M*}, \sigma_A^2, \psi \mid \mathbf{t}, \mathbf{x}, \mathbf{d}, \mathbf{a}, \mathbf{z}^{\mathbf{A}}, \mathbf{z}^{\mathbf{M}}, \mathbf{s}) \quad (5.1)$$

$$\propto \prod_{i=1}^n \prod_{k=1}^K \pi_k^{s_{i,k}} \left\{ \left[ \prod_{j=1}^2 f(t_{i,j} \mid x_{i,j}, d_{i,j}, a_i; \alpha_j, \beta_j, \gamma_{k,j}^*) \right] f(z_i^{\mathbf{A}} \mid \zeta_k^{A*}, \sigma_A^2) (\zeta_k^{M*})^{z_i^{\mathbf{M}}} (1 - \zeta_k^{M*})^{1-z_i^{\mathbf{M}}} \right\}^{s_{i,k}}$$

where  $f(\cdot \mid \zeta_k^{A*}, \sigma_A^2)$  denotes the probability density function of the normal distribution with mean  $\zeta_k^{A*}$  and variance  $\sigma_A^2$ , and  $(\zeta_k^{M*})^{z_i^{\mathbf{M}}} (1 - \zeta_k^{M*})^{1-z_i^{\mathbf{M}}}$  is the probability mass function of the Bernoulli-distributed random variable  $Z_i^{\mathbf{M}}$  with parameter  $\zeta_k^{M*}$ . The  $j$ th couple member likelihood contribution  $f(t_{i,j} \mid x_{i,j}, d_{i,j}, a_i; \alpha_j, \beta_j, \gamma_{k,j}^*)$  is derived to account for right censoring and left truncation by simple algebra as follows:

$$f(t_{i,j} \mid x_{i,j}, d_{i,j}, a_i; \alpha_j, \beta_j, \gamma_{k,j}^*)$$

$$= \frac{\exp \left[ - \int_0^{t_{i,j}+a_i} \mu(s \mid x_{i,j}; \alpha_j, \beta_j, \gamma_{k,j}^*) ds \right] \mu(t_{i,j} + a_i \mid x_{i,j}; \alpha_j, \beta_j, \gamma_{k,j}^*)^{d_{i,j}}}{\exp \left[ - \int_0^{a_i} \mu(q \mid x_{i,j}; \alpha_j, \beta_j, \gamma_{k,j}^*) dq \right]} \quad (5.2)$$

where the numerator is the likelihood contribution for a non-informative right-censored observation, which is divided by the probability of being alive between policy inception (time 0) until the start of the observation period. Given the specification of the hazard function in equation (4.1), the logarithm of the likelihood contribution is given by:

$$\ln f(t_{i,j} \mid x_{i,j}, d_{i,j}, a_i; \alpha_j, \beta_j, \gamma_{k,j}^*)$$

$$= - \frac{\exp[\beta_j(t_{i,j} + a_i)] - \exp(\beta_j a_i)}{\beta_j} \exp(\alpha_j + \beta_j(x_{i,j} - \bar{x}) + \gamma_{k,j}^*)$$

$$+ d_{i,j}(\alpha_j + \beta_j(x_{i,j} - \bar{x} + t_{i,j} + a_i) + \gamma_{k,j}^*) . \quad (5.3)$$

### 5.3. Posterior distribution

The (unnormalized) posterior distribution follows as the product of the likelihood and the prior distribution. In this latter, we assume that all parameters are pairwise independently distributed, with densities generically denoted by  $p$ :

$$\begin{aligned}
& p(\alpha, \beta, \gamma^*, \zeta^{A*}, \zeta^{M*}, \sigma_A^2, \Sigma_\gamma, m_A, s_A^2, \phi, \psi \mid \mathbf{t}, \mathbf{x}, \mathbf{d}, \mathbf{a}, \mathbf{z}^A, \mathbf{z}^M, \mathbf{s}) \tag{5.4} \\
& \propto L(\alpha, \beta, \gamma^*, \zeta^{A*}, \zeta^{M*}, \sigma_A^2, \psi \mid \mathbf{t}, \mathbf{x}, \mathbf{d}, \mathbf{a}, \mathbf{z}^A, \mathbf{z}^M, \mathbf{s}) \left[ \prod_{j=1}^2 \underbrace{p(\alpha_j \mid 1, 1)}_{\text{Gamma}(1,1)} \right] \left[ \prod_{j=1}^2 \underbrace{p(\beta_j \mid 0.1, 0.25)}_{\text{N}_+(0.1, 0.25)} \right] \\
& \times \left[ \prod_{k=1}^K \underbrace{p(\gamma_{k,\cdot}^* \mid \mathbf{0}, \Sigma_\gamma)}_{\text{MVN}(\mathbf{0}, \Sigma_\gamma)} \underbrace{p(\zeta_k^{A*} \mid m_A, s_A^2)}_{\text{N}(m_A, s_A^2)} \underbrace{p(\zeta_k^{M*} \mid 13.31, 4.44)}_{\text{Beta}(13.31, 4.44)} \right] \underbrace{p(m_A \mid 3, 0.5)}_{\text{N}(3, 0.5)} \underbrace{p\left(\Sigma_\gamma \mid 5, \begin{bmatrix} 2 & 1.2 \\ 1.2 & 2 \end{bmatrix}\right)}_{\text{Inv-Wishart}\left(5, \begin{bmatrix} 2 & 1.2 \\ 1.2 & 2 \end{bmatrix}\right)} \\
& \times \underbrace{p(\sigma_A^2 \mid 2, 1)}_{\text{Inv-Gamma}(2,1)} \underbrace{p(s_A^2 \mid 3, 0.5)}_{\text{Inv-Gamma}(3, 0.5)} \left[ \prod_{k=1}^{K-1} \underbrace{p(\psi_k \mid 1, \phi)}_{\text{Beta}(1, \phi)} \right] \underbrace{p(\phi \mid 6, 12)}_{\text{Gamma}(6, 12)}.
\end{aligned}$$

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 and James, 2001), consisting of a sequential draws of the parameters (exploiting their conditional conjugacy where possible). 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 100,000 times. We discard the first 80,000 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 1,000 draws. We run the sampler four times, based on sparse starting values, in order to assess the mixing of the chains and the convergence of the MCMC sampler towards a stationary distribution.

The trace plots of the parameters show that the chains converge towards a stationary distribution for all parameters, except for  $\gamma_{k,1}^*$ ,  $\gamma_{k,2}^*$ ,  $\zeta_k^{A*}$ ,  $\zeta_k^{M*}$  and  $\pi_k$  for those mixture components where few units are allocated at each iteration of the MCMC sampler.

Furthermore, the overlap of the four trace plots for each parameter show a good mixing for all of those, and the marginal densities are all unimodal with similar location. The chains for the parameter  $\alpha_j$  converge towards a stationary distribution, although they do not mix with one another. Nevertheless, the marginal posterior densities for this parameter from all the chains do not differ significantly. This result is likely to be connected with the correlation in the posterior distribution between  $\alpha_j$  and  $\gamma_{k,j}$ . We do not analyze these two aspects for  $(\gamma_{k,1}^*, \gamma_{k,2}^*, \zeta_k^{A*}, \zeta_k^{M*})$  due to the label switching problem which characterizes mixture distributions as discussed in [Betancourt \(2017\)](#), [Ungolo et al. \(2020\)](#) and [Ungolo and van den Heuvel \(2022\)](#). This issue affects solely the interpretation of the groups from the 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  $\beta_j$ .

The overall convergence of the MCMC sampler is further analysed through the trace plot of the clustering entropy shown in [Figure 6.1](#) for two chains, as used for the analysis of mixture models ([Beraha et al., 2023](#)). The entropy is calculated at the  $\ell$ th iteration as follows:

$$\text{Entropy}(\ell) = \sum_{k=1}^K n_k^{(\ell)} \cdot \ln n_k^{(\ell)}$$

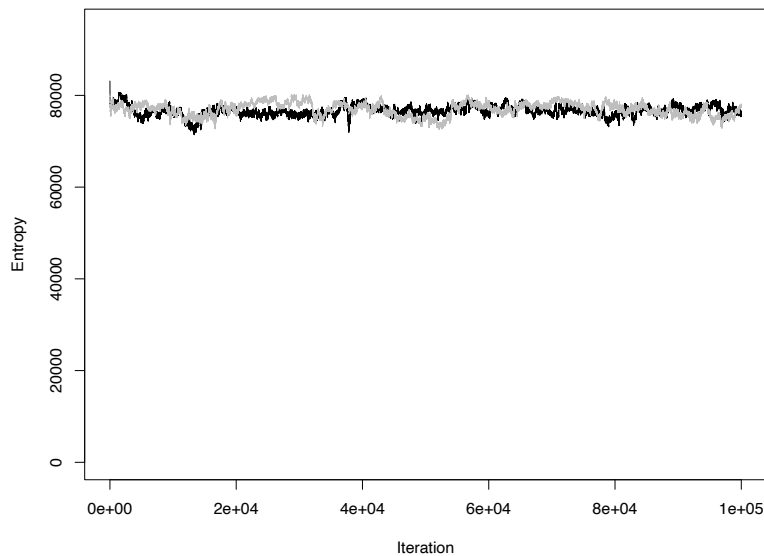


Figure 6.1: Trace plot of the Entropy across all iterations of the MCMC sampler for two different runs.

We note that all chains converge towards the same value. This is also observed when these chains are compared to those obtainable under the different prior distributions

discussed in Section 5.1. This is an indication of the stability of the allocation of each couple to the groups characterized by the mixture components.

## 6.2. Model analysis

### Model results

The analysis of the histogram of the number of occupied classes across all iterations shows that a minimum of 8 and a maximum of 24 mixture components have at least one observation (Figure 2.1 in the Supplementary material). The posterior mean of the number of couples allocated in each mixture component show that only six of these contain at least 2.5% (253 couples) of the total number of observations, and the sum of these six components totals 9,458 observations (93.5% of the total). These two results show that a truncation level of  $K = 25$  on the number of mixtures is sufficient to approximate the joint distribution of  $(T_{i,1}, T_{i,2}, Z_i^A, Z_i^M)$ .

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 ( $\alpha$ ) 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  $\beta$  compared to males.

The six classes which cover more than 95% of the total observations are characterized by different values of the class-specific parameters  $(\gamma_{k,1}^*, \gamma_{k,2}^*, \zeta_k^{A*}, \zeta_k^{M*})$  for  $k = 1, 2, 3, 4, 6, 9$ . These parameters, which capture the level of heterogeneity among the observations, allow to obtain hazard functions which can take a flexible shape, and where the effect of the other covariates is not necessarily proportional, nor monotone as we can see for the age difference in Figure 6.2. For example, at younger ages, when the female is years older than the male couple member, the hazard function for the females is higher compared to the case of an age difference of two years, while this relationship reverts around age 75.



Table 6.1: Posterior summaries of  $\alpha$ ,  $\beta$ ,  $\gamma_k^*$ ,  $\zeta_k^{A*}$ ,  $\zeta_k^{M*}$ ,  $\pi_k$  for  $k = 1, 2, 3, 4, 6, 9$  and  $\phi$ .

Parameter	Mean	95% Cred. Int.	Parameter	Mean	95% Cred. Int.
$\alpha_1$	-3.45	(-3.71, -3.27)	$\alpha_2$	-4.53	(-4.90, -4.12)
$\beta_1$	0.1064	(0.0963; 0.1160)	$\beta_2$	0.1440	(0.1274; 0.1614)
$\gamma_{1,1}^*$	-1.00	(-1.75; -0.53)	$\gamma_{2,1}^*$	-2.14	(-3.74; -0.93)
$\gamma_{1,2}^*$	-0.72	(-0.94; -0.49)	$\gamma_{2,2}^*$	-0.39	(-0.86; 0.16)
$\gamma_{1,3}^*$	0.39	(0.15; 0.68)	$\gamma_{2,3}^*$	-1.21	(-1.67; -0.81)
$\gamma_{1,4}^*$	-0.51	(-0.89; -0.26)	$\gamma_{2,4}^*$	-0.39	(-0.87; 0.02)
$\gamma_{1,6}^*$	-0.19	(-0.44; 0.17)	$\gamma_{2,6}^*$	-0.32	(-0.97; 0.14)
$\gamma_{1,9}^*$	-1.19	(-1.48; -0.84)	$\gamma_{2,9}^*$	-1.19	(-1.32; -0.97)
$\zeta_1^{A*}$	1.13	(0.99; 1.28)	$\zeta_1^{M*}$	0.66	(0.52; 0.75)
$\zeta_2^{A*}$	1.71	(1.63; 1.81)	$\zeta_2^{M*}$	0.95	(0.93; 0.98)
$\zeta_3^{A*}$	0.50	(-4.15; 2.46)	$\zeta_3^{M*}$	0.73	(0.53; 0.90)
$\zeta_4^{A*}$	-0.36	(-1.49; 0.68)	$\zeta_4^{M*}$	0.58	(0.47; 0.77)
$\zeta_6^{A*}$	0.24	(-1.70; 1.90)	$\zeta_6^{M*}$	0.68	(0.53; 0.88)
$\zeta_9^{A*}$	-1.73	(-2.90; -1.02)	$\zeta_9^{M*}$	0.54	(0.42; 0.68)
$\pi_1$	0.2291	(0.1321; 0.3167)	$\pi_2$	0.4078	(0.3240; 0.4834)
$\pi_3$	0.0328	(0.0005; 0.1334)	$\pi_4$	0.1205	(0.0272; 0.2311)
$\pi_6$	0.1064	(0.0020; 0.2329)	$\pi_9$	0.0387	(0.0112; 0.0639)
$\phi$	0.97	(0.55; 1.54)			

### Analysis of the effect of couple-specific covariates

We plot the value of the log-hazard function of males and females for different values of  $Z^A$  and  $Z^M$  (Figure 6.2). This hazard function is calculated as follows:

$$\begin{aligned}
 & \mu(t | x_{\cdot,j}, z^A, z^M; \alpha_j, \beta_j, \gamma_{\cdot,j}^*, \pi, \zeta^{A*}, \zeta^{M*}) \\
 &= \frac{\sum_{k=1}^K \pi_k f(t | x_{\cdot,j}; \alpha_j, \beta_j, \gamma_{k,j}^*) f(z^A | \zeta_k^{A*}, \sigma_A^2) (\zeta_k^{M*})^{z^M} (1 - \zeta_k^{M*})^{1-z^M}}{\sum_{k=1}^K \pi_k (1 - F(t | x_{\cdot,j}; \alpha_j, \beta_j, \gamma_{k,j}^*)) f(z^A | \zeta_k^{A*}, \sigma_A^2) (\zeta_k^{M*})^{z^M} (1 - \zeta_k^{M*})^{1-z^M}} \quad (6.1)
 \end{aligned}$$

where  $F$  denotes the cumulative distribution function of the random time to event  $T$ , hence  $1 - F$  denotes the survival function. Both numerator and denominator are averaged over the parameter draws.

As observed, this model allows for a nonmonotone 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.

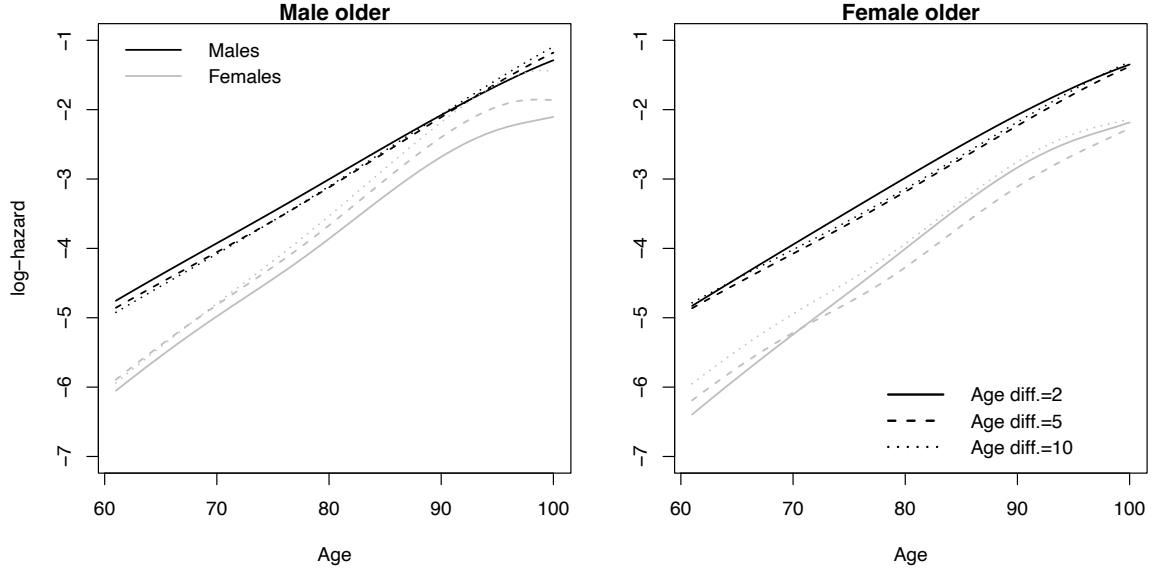


Figure 6.2: Hazard function for males (black) and females (gray) aged 60-100 for different values of the covariates male older and age difference ( $\exp(Z^A)$ ).

First of all, we notice that in the case where the male is the older couple member ( $Z^M=1$ ), there is a slight shift upwards of the hazard function at each age and for all all values of  $Z^A$ . The only exception occurs when a male is aged between 60 and 68 and  $Z^A = 10$ . A similar finding can be observed when analyzing the fitted hazard function for the females.

These results demonstrate how the model is capable to flexibly capture the effect of interactions between  $X$ ,  $Z^A$  and  $Z^M$  without an explicit modelling assumption about these. We remark how these three variables do not have any linear relationship among them following the definition of  $Z^A$  in Section 4.

### Dependent lifetime events

We analyze the statistical association between  $T_1$  and  $T_2$  by calculating the Spearman  $\rho$  and the Kendall  $\tau$  rank correlation coefficients over 20,000 random samples of  $(T_1, T_2)$  generated for each value of the age difference between 1 and 20 and of  $Z^M$  (Figure 6.3). For each sample, we first draw the male age from a Uniform(40, 80), with the female age following from the value of  $Z^A$  and  $Z^M$ .

We note how the model captures the presence of a positive statistical association between males and females lifetime. However, we note how this statistical association is unchanged with the values of  $Z^A$  and  $Z^M$ . As additional experiment, we then narrow the endpoints of the Uniform distribution of the male individual age and repeat the same experiment. First, we note how both  $\rho$  and  $\tau$  are much smaller than the previous case, while  $T_1$  and  $T_2$  are still positively associated. For an age difference of one year,  $\rho$  and  $\tau$  are relatively larger, and then have a constant value for every value of  $Z^A$ , especially

for the case where the female is the oldest couple member. In the opposite case where  $Z^M = 1$ , we observe a decreasing value of  $\rho$  and  $\tau$  as  $Z^A$  increases. A different midpoint of the Uniform distribution for the males age does not bring any change in the results. The modest impact of the age difference on the values of  $\rho$  and  $\tau$  might be due to the direct relationship between the age and the computation of the age difference. We remark how we perform this analysis on the basis of the solely available joint covariates. In any case, we could show how the AVDPM approach of this work can account for the  $Z$ -lifetime dependence as by-product.

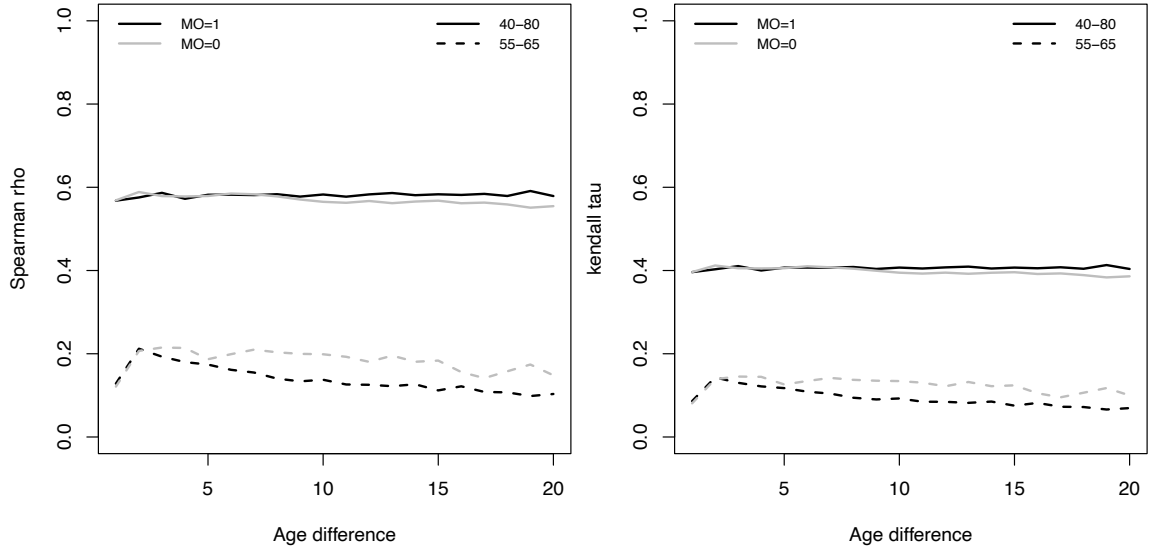


Figure 6.3: Value of the Spearman  $\rho$  (left panel) and of the Kendall  $\tau$  (right panel) for different values of AD and MO.

## 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 and van den Heuvel \(2024\)](#).

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

$$\begin{aligned}
 q_{i,k} &= \mathbb{P}(W_i = k \mid t_{i,1}, t_{i,2}, x_{i,1}, x_{i,2}, a_i, d_i, z_i^A, z_i^M) \\
 &= \frac{\pi_k \left[ \prod_{j=1}^2 f(t_{i,j} \mid x_{i,j}, d_{i,j}, a_i; \alpha_j, \beta_j, \gamma_{k,j}^*) \right] f(z_i^A \mid \zeta_k^{A*}, \sigma_A^2) (\zeta_k^{M*})^{z_i^M} (1 - \zeta_k^{M*})^{1-z_i^M}}{\sum_{h=1}^K \pi_h \left[ \prod_{j=1}^2 f(t_{i,j} \mid x_{i,j}, d_{i,j}, a_i; \alpha_j, \beta_j, \gamma_{h,j}^*) \right] f(z_i^A \mid \zeta_h^{A*}, \sigma_A^2) (\zeta_h^{M*})^{z_i^M} (1 - \zeta_h^{M*})^{1-z_i^M}}
 \end{aligned} \tag{6.2}$$

where we average both the numerator and the denominator (separately) over the posterior draws of the parameters.

The  $i$ th couple is then hard-assigned to the  $k$ th class, by setting  $w_i = k$  if  $q_{i,k} > q_{i,h}$  for  $h \neq k$ . The four largest classes total 98.4% of the observations. Table 6.2 illustrates the key features of groups 1, 2, 4 and 9, alongside the posterior mean of their class-specific parameters ( $\gamma_{\cdot,1}^*$ ,  $\gamma_{\cdot,2}^*$ ,  $\zeta^{A*}$ ,  $\zeta^{M*}$ ). The difference between % Composition in Table 6.2 and  $\pi$  is due to the specific classification rule we use.

Table 6.2: Features of the four largest classes as determined by Bayes' rule.

	Group 1	Group 2	Group 4	Group 9	Train. sample
% Composition	29.25	50.73	12.99	5.39	—
Age male (mean)	64.68	66.29	64.96	64.56	65.55
Age female (mean)	65.22	60.28	64.89	64.55	62.63
log( Age Diff.  ) (mean)	0.74	1.71	−0.47	−2.07	0.91
Male older (in %)	53.75	98.71	56.39	55.05	77.26
$\gamma_{\cdot,1}^*$	−1.00	−0.71	−0.51	−1.19	—
$\gamma_{\cdot,2}^*$	−2.14	−0.39	−0.39	−1.19	—
$\zeta_{AD,\cdot}^*$	1.13	1.71	−0.36	−1.72	—
$\zeta_{MO,\cdot}^*$	0.66	0.95	0.58	0.54	—

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.

Around 50% of the couples compose Group 2, which is characterized by the largest log-absolute age difference among the four groups and the males are the oldest member of the couple for almost all observations.

Group 1 is characterized by the lowest percentage of couples where the males are the oldest couple members, and where females have an average age higher than males. Furthermore, this group is characterized by the lowest hazard function for males and females, given their  $\gamma^*$  posterior mean coefficient. Group 4 and 9 are almost similar in terms of average age of the males and females therein, as well as the percentage of older males, while these differ in terms of the log-absolute age difference in couples.

## 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. \(2022\)](#)<sup>1</sup> which assume that:

$$\begin{aligned}\mu^{BG}(t | x_{i,j}; \alpha_j, \beta_j) &= \exp(\alpha_j + \beta_j(x_{i,j} - \bar{x} + t)) \\ \mu^{PH}(t | x_{i,j}, z_i^A, z_i^M; \alpha_j, \beta_j, \delta_{j,1}, \delta_{j,2}) &= \exp(\alpha_j + \beta_j(x_{i,j} - \bar{x} + t) + \delta_{j,1}z_i^A + \delta_{j,2}z_i^M)\end{aligned}$$

with parameters estimated by using maximum likelihood (results in Section 3 of the Supplementary Material). Therefore, under BG and PH models we assume that conditional on the individual age, and the two common covariates, then the males and females lifetimes are independently distributed.

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

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

(W)AIC	BG	PH	AVDPM
<b>In sample</b>	13,603.41	13,603.11	<b>13,527.50</b>
<b>Out of sample</b>	4,537.42	4,541.64	<b>4,515.14</b>

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.

These results are confirmed also when calculating AIC and WAIC by gender (Table [6.4](#)). The use of the AVDPM yields a better in sample and out of sample performance for both males and females (the sum of the WAICs by gender is not equal to the overall WAIC as consequence of the Jensen’s inequality, as we can observe from its computation in Appendix [B](#)).

---

<sup>1</sup>Compared to this paper, we consider the fact that 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 authors.

Table 6.4: AIC and WAIC of the models by gender (the lowest value is shown in bold).

(W)AIC	Males			Females		
	BG	PH	AVDPM	BG	PH	AVDPM
<b>In sample</b>	9,576.52	9,579.78	<b>9,498.77</b>	4,026.89	4,023.33	<b>3,948.61</b>
<b>Out of sample</b>	3,185.87	3,188.66	<b>3,166.74</b>	1,351.55	1,352.98	<b>1,330.25</b>

## 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  $\iota$ , 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):

$$\begin{aligned}\bar{a}_{x_1, x_2}(z) &= \mathbb{E}[Y_{x_1, x_2}(z)] = \int_0^\infty \exp(-\iota t) S_{x_1, x_2}(t | z) dt; \\ \bar{a}_{\overline{x_1, x_2}}(z) &= \mathbb{E}[Y_{\overline{x_1, x_2}}(z)] = \int_0^\infty \exp(-\iota t) S_{\overline{x_1, x_2}}(t | z) dt,\end{aligned}$$

where the last survivor function  $S_{\overline{x_1, x_2}}(t | z)$  was defined in Section 3, and  $S_{x_1, x_2}(t | z)$  denotes the corresponding joint survivor function.

We analyze the effect of the covariates by comparing the annuity value under the AVDPM with the annuity value obtainable using the BG model in order to assess the effect of the covariates (as well as the dependence), and with the value obtained using the PH model to assess the effect of the dependence.

Figure 7.1 shows the value of the last survivor annuity factor obtainable under the AVDPM model and the other two competing models PH and BG described in Section 6. The annuity factors are evaluated at different male ages (60 and 70), different values of  $Z = (Z^A, Z^M)$  and two different forces of interest  $\iota = (0.01, 0.05)$ . The same plot for the joint life annuity is shown in Section 4 of the Supplementary Material (Figure 4.1).

The plot shows that the AVDPM annuity value is always higher than the value obtained using the PH model. When the male is the oldest couple member, the difference between these two values is highest when the age difference is within five years, and narrows afterwards. Such annuity price difference is also negligible for an age difference of one year. This means that neglecting the effect of the dependence has the effect of underpricing the last survivor annuity, especially when the age difference is between two and five years. A similar evidence is obtained when the female is the oldest couple member, whenever the male is lesser than ten years younger. 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) show that the price of a last survivor annuity is higher under a model which assumes independent lifetimes,

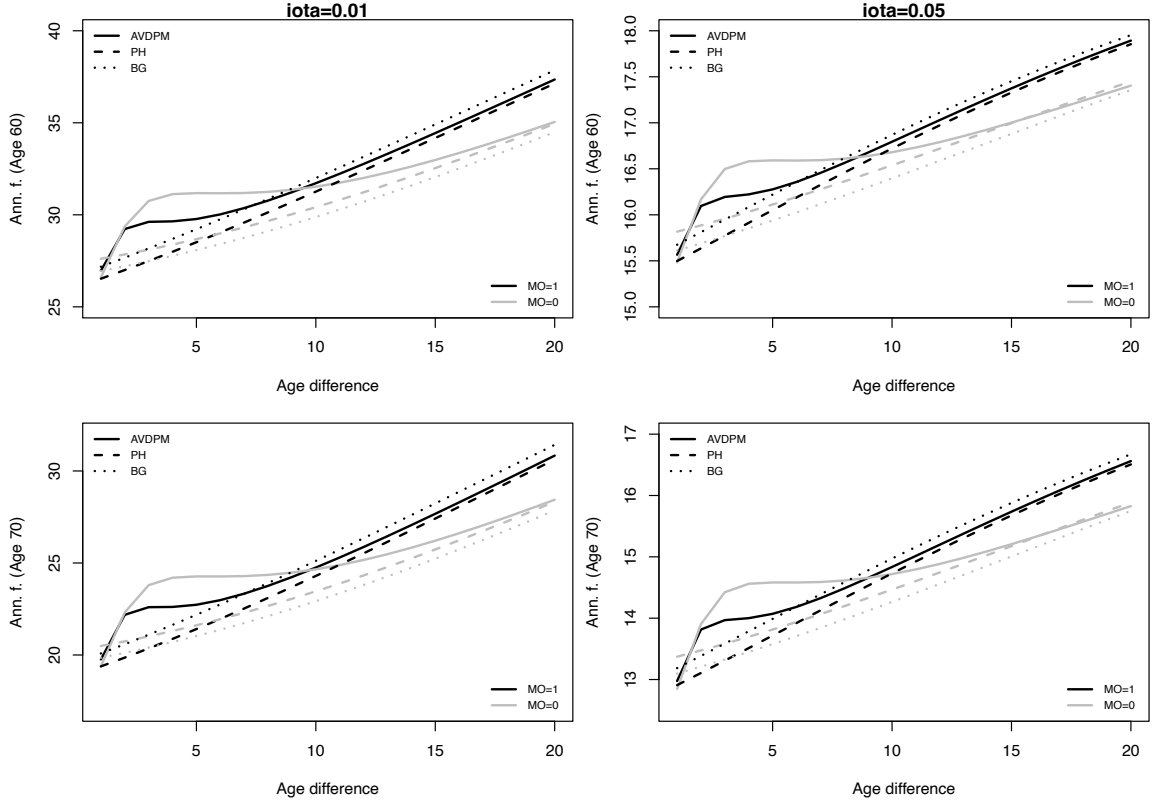


Figure 7.1: Last survivor annuity factor calculated using the AVDPM model (solid line), the PH model (dashed line) and the BG model (dotted line) for different values of age difference ( $\exp(Z^A)$ , x-axis) and  $Z^M$  when the oldest member is aged 60 (top panel) and 70 (bottom panel), for  $\iota = 1\%$  (left panel) and  $\iota = 5\%$  (right panel).

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  $Z^M$ .

When comparing the annuity factor using the AVDPM with the value obtained using a Base Gompertz model which assumes independent lifetimes and does not account for the effect of the covariates, we note that using latter in annuity pricing yields a lower value compared to the former when the female is the older couple member. Conversely, when  $Z^M = 1$ , the BG model underprices the last survivor annuity for an age difference lower than eight years. These evidences are consistent with the findings in [Frees et al. \(1996\)](#), Figure 4), and [Deresa et al. \(2022\)](#), where the latter analyze only the case of a couple where the male is aged 65 and the female is two years younger.

A higher interest rate, which decreases the value of the annuity all else being equal, yields a reduction of both the effect of the dependence, as well as of the effect of the

price difference due to the covariates. On the other hand, an increase in the age of the two heads (Figure 7.1 bottom panel) yields a higher percentage difference between the annuity value priced using the AVDPM model and the two models which assume independence. Therefore, in this latter case, the dependence turns out to have a higher impact in the pricing of the last survivor annuity.

For the joint life annuity, we note that when the male is the oldest couple member, the joint life annuity factor under the AVDPM is always higher compared to the value obtainable under the two models assuming independent lifetimes for an age difference higher than a year, and the same is obtained in case  $Z^M = 0$  and the age difference is between 1 and 15 years. This is obtained for both male ages and for the two interest rates hereby analysed. Similarly, the copula-based dependence model analysed by [Deresa et al. \(2022\)](#) yields a higher joint life annuity value compared to the independence case, although they focus on the sole case where males and females are again aged 65 and 63 respectively.

## 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 differing 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  $i$ th group includes  $J_i$  members, with common set of variables  $z_i$ , and individual (within group) characteristics  $x_{i,j}$  ( $j = 1, \dots, J_i$ ). A possible solution is to model the group-specific joint distribution of the lifetimes and  $Z$  as:

$$f(t_{i,1}, \dots, t_{i,J_i}, z_i | x_i; \beta, G) = \int_{\Omega_{\gamma, \eta}} \left[ \prod_{j=1}^{J_i} f(t_{i,j} | x_{i,j}; \beta, \gamma_{i,j}) \right] f(z_i | \zeta_i) dQ(\gamma_{i,j} | \eta_i) dG(\eta_i, \zeta_i) \quad (8.1)$$

where  $Q$  denotes a suitable distribution function for  $\gamma_{i,j}$ , indexed by the group specific parameter  $\eta_i$ . 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  $\eta_i$  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 the group-specific covariates. The



use of Dirichlet Process Mixture models enhances the flexibility of the random effects distribution and of the standard parametric assumptions for the covariates. The resulting Augmented Variable DPM (AVDPM) model has been illustrated through the empirical analysis of the mortality rates of the male and female 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 full Bayesian analysis of the AVDPM model of this work allows for the incorporation of the prior information about possible parameter values from the user. An alternative to the full Bayesian analysis can be the assumption of a fixed, known number of mixture components, and fit the model parameters using maximum likelihood. The results of similar analysis in the field can be used at this purpose. In addition, an approach based on the use of random effects allows for the specification of different parametric forms for the lifetime of each couple or group member.

An important future research avenue is the development of fast approximation techniques for the computation of posterior distribution of the parameters (for example, based on Variational Inference techniques (Blei et al., 2016)), which would allow to quickly explore different parametric models for the data on hand.

## 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

$$\left( \alpha^{(0)}, \beta^{(0)}, \gamma^{*(0)}, \zeta^{A*(0)}, \zeta^{M*(0)}, \sigma_A^{2(0)}, \Sigma_{\gamma}^{(0)}, m_A^{(0)}, s_A^{2(0)}, \phi^{(0)}, \psi^{(0)} \right);$$

For each iteration  $\ell = 1, \dots, M$ :

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

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

where

$$f\left(z_i \mid \zeta_k^{*(\ell)}\right) = f\left(z_i^A \mid \zeta_k^{A*(\ell)}, \sigma_A^{2(\ell)}\right) \left(\zeta_k^{M*(\ell)}\right)^{z_i^M} \left(1 - \zeta_k^{M*(\ell)}\right)^{1-z_i^M}. \quad (\text{A.2})$$

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

**Step 2:** Sample the stick-breaking weights  $\psi$  and update  $\pi$ :

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

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

**Step 2.2:** Update  $\pi_k$ :

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

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

$$\begin{aligned} \Gamma_{j,1} &= 1 + \sum_{i=1}^n d_{i,j} \\ \Gamma_{j,2} &= 1 + \sum_{i=1}^n \frac{\exp\left[\beta_j^{(\ell-1)}(t_{i,j} + a_i)\right] - \exp\left(\beta_j^{(\ell-1)} a_i\right)}{\beta_j^{(\ell-1)}} \exp\left(\beta_j^{(\ell-1)}(x_{i,j} - \bar{x}) + \gamma_{w_i^{(\ell)},j}^{*(\ell-1)}\right) \end{aligned} \quad (\text{A.5})$$

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

**Step 4.1:** Sample  $\beta_j^*$  from a Truncated Normal proposal distribution  $q$  with mean  $\beta_j^{(\ell-1)}$  and variance  $\sigma_{p,j}^{2(\ell-1)}$ . The proposal distribution has truncation bounds given by the parameters of the uniform distribution specified as 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

$$r' = \frac{\left[ \prod_{i=1}^n f\left(t_{i,j} \mid x_{i,j}, d_{i,j}, a_i; \alpha_j^{(\ell)}, \beta_j^*, \gamma_{j,w_i^{(\ell)}}^{*(\ell-1)}\right) \right] p\left(\beta_j^* \mid 0.1, 0.25\right) q_{10^{-5},5}\left(\beta_j^{(\ell-1)} \mid \beta_j^*, \sigma_{p,j}^{2(\ell-1)}\right)}{\left[ \prod_{i=1}^n f\left(t_{i,j} \mid x_{i,j}, d_{i,j}, a_i; \alpha_j^{(\ell)}, \beta_j^{(\ell-1)}, \gamma_{j,w_i^{(\ell)}}^{*(\ell-1)}\right) \right] p\left(\beta_j^{(\ell-1)} \mid 0.1, 0.25\right) q_{10^{-5},5}\left(\beta_j^* \mid \beta_j^{(\ell-1)}, \sigma_{p,j}^{2(\ell-1)}\right)} \quad (\text{A.6})$$

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

**Step 4.3:** Set:

$$\beta_j^{(\ell)} = \begin{cases} \beta_j^* & \text{w.p. } \min(r', 1) \\ \beta_j^{(\ell-1)} & \text{w.p. } 1 - \min(r', 1) \end{cases} \quad (\text{A.7})$$

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

$$\sigma_{p,j}^{(\ell)} = \sigma_{p,j}^{(\ell-1)} \sqrt{1 + \ell^{-0.6} (\min(r', 1) - 0.234)} \quad (\text{A.8})$$

The parameter 0.6 is chosen following [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  $\gamma^*$  using the acceptance-rejection sampling method:

**Step 5.1:** For  $k = 1, \dots, K$  sample  $\gamma'_k = (\gamma'_{k,1}, \gamma'_{k,2})^\top$  from a bivariate Normal proposal distribution  $q$  with mean  $\gamma_k^{*(\ell-1)} = (\gamma_{k,1}^{*(\ell-1)}, \gamma_{k,2}^{*(\ell-1)})^\top$  and variance-covariance matrix  $\Sigma_{\mathbf{p},\mathbf{k}}^{(\ell-1)}$ :

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

**Step 5.1.2** Set  $\gamma'_k = \gamma_k^{*(\ell-1)} + \mathbf{L}_{\mathbf{k}}^{(\ell-1)} \mathbf{h}^{(\ell)}$ ,

where  $\mathbf{L}^{(\ell-1)}_{\mathbf{k}}$  is the lower triangular matrix denoting the Cholesky decomposition of  $\Sigma_{\mathbf{p},\mathbf{k}} = \mathbf{L}_{\mathbf{k}} \mathbf{L}_{\mathbf{k}}^\top$ . 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:

$$r'' = \frac{p(\gamma'_k | \mathbf{0}, \Sigma_{\gamma}^{(\ell-1)}) \left[ \prod_{\{i:w_i^{(\ell)}=k\}} \prod_{j=1}^2 f_j(t_{i,j} | x_{i,j}, d_{i,j}, a_i; \alpha_j^{(\ell)}, \beta_j^{(\ell)}, \gamma'_{k,j}) \right] q(\gamma'_k | \gamma_k^{*(\ell-1)}, \Sigma_{\mathbf{p},\mathbf{k}}^{(\ell-1)})}{p(\gamma_k^{*(\ell-1)} | \mathbf{0}, \Sigma_{\gamma}^{(\ell-1)}) \left[ \prod_{\{i:w_i^{(\ell)}=k\}} \prod_{j=1}^2 f_j(t_{i,j} | x_{i,j}, d_{i,j}, a_i; \alpha_j^{(\ell)}, \beta_j^{(\ell)}, \gamma_{k,j}^{*(\ell-1)}) \right] q(\gamma_k^{*(\ell-1)} | \gamma'_k, \Sigma_{\mathbf{p},\mathbf{k}}^{(\ell-1)})} \quad (\text{A.9})$$

**Step 5.3:** Set:

$$\left( \gamma_{k,1}^{*(\ell)}, \gamma_{k,2}^{*(\ell)} \right) = \begin{cases} \left( \gamma'_{k,1}, \gamma'_{k,2} \right) & \text{w.p. } \min(r'', 1) \\ \left( \gamma_{k,1}^{*(\ell-1)}, \gamma_{k,2}^{*(\ell-1)} \right) & \text{w.p. } 1 - \min(r'', 1) \end{cases} \quad (\text{A.10})$$

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

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

$$\Sigma_{\mathbf{p},\mathbf{k}}^{(\ell)} = \mathbf{L}_{\mathbf{k}}^{(\ell-1)} \left( \mathbf{I}_2 + \ell^{-0.6} (\min(r'', 1) - 0.234) \frac{\mathbf{h}^{(\ell)} \mathbf{h}^{(\ell)\top}}{\|\mathbf{h}^{(\ell)}\|^2} \right) \mathbf{L}_{\mathbf{k}}^{(\ell-1)\top} \quad (\text{A.11})$$

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

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

**Step 6:** Sample  $\Sigma_\gamma$  from the conjugate posterior which is the Inv-Wishart distribution with degrees of freedom  $\Lambda_1$  and scale matrix  $\Lambda_2$ , calculated as:

$$\begin{aligned} \Lambda_1 &= 5 + \sum_{k=1}^K 1_{[n_k^{(\ell)} > 0]} \\ \Lambda_2 &= \begin{bmatrix} 2 & 1.2 \\ 1.2 & 2 \end{bmatrix} + \sum_{k=1}^K 1_{[n_k^{(\ell)} > 0]} \gamma_k^{*(\ell)} \gamma_k^{*(\ell)\top} \end{aligned} \quad (\text{A.12})$$

$$\text{where } n_k^{(\ell)} = \sum_{i=1}^n s_{i,k}^{(\ell)} = \sum_{i=1}^n 1_{[w_i^{(\ell)} = k]};$$

**Step 7:** Sample  $\zeta_k^{A*(\ell)}$  ( $k = 1, \dots, K$ ) from a conjugate  $N(\Lambda_{3,k}, \Lambda_{4,k})$  distribution, where:

$$\begin{aligned} \Lambda_{3,k} &= \Lambda_{4,k} \left( \frac{\sum_{\{i:w_i^{(\ell)}=k\}} z_i^A}{\sigma_A^{2(\ell-1)}} + \frac{m_A^{(\ell-1)}}{s_A^{2(\ell-1)}} \right) \\ \Lambda_{4,k} &= \left( \frac{n_k^{(\ell)}}{\sigma_A^{2(\ell-1)}} + \frac{1}{s_A^{2(\ell-1)}} \right)^{-1} \end{aligned}$$

**Step 8:** Sample  $\zeta_k^{M*(\ell)}$  ( $k = 1, \dots, K$ ) from a conjugate Beta( $\Lambda_{5,k}, \Lambda_{6,k}$ ) distribution, with posterior parameters calculated as follows:

$$\begin{aligned} \Lambda_{5,k} &= 13.31 + \sum_{\{i:w_i^{(\ell)}=k\}} z_i^M \\ \Lambda_{6,k} &= 4.44 + \sum_{\{i:w_i^{(\ell)}=k\}} (1 - z_i^M) \end{aligned}$$

**Step 9:** Sample  $m_A^{(\ell)}$  from a conjugate  $N(\Lambda_7, \Lambda_8)$  distribution, where:

$$\Lambda_7 = \Lambda_8 \frac{\sum_{\{k: n_k^{(\ell)} > 0\}} \zeta_k^{A*^{(\ell)}}}{s_A^{2(\ell-1)}}$$

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

**Step 10:** Sample  $\sigma_A^{2(\ell)}$  from a conjugate Inv-Gamma $(\Lambda_9^{(\ell)}, \Lambda_{10}^{(\ell)})$  distribution where:

$$\Lambda_9^{(\ell)} = 2 + 0.5n$$

$$\Lambda_{10}^{(\ell)} = 1 + \sum_{i=1}^n \left( z_i^A - \zeta_{w_i^{(\ell)}}^{A*^{(\ell)}} \right)^2$$

**Step 11:** Sample  $s_A^{2(\ell)}$  from a conjugate Inv-Gamma $(\Lambda_{11}^{(\ell)}, \Lambda_{12}^{(\ell)})$  distribution where:

$$\Lambda_{11}^{(\ell)} = 3 + 0.5 \sum_{k=1}^K 1_{[n_k^{(\ell)} > 0]}$$

$$\Lambda_{12}^{(\ell)} = 0.5 + \sum_{\{k: n_k^{(\ell)} > 0\}} \left( \zeta_k^{A*^{(\ell)}} - m_A^{(\ell)} \right)^2$$

**Step 12:** Sample  $\phi$  by following the steps outlined in [Escobar and West \(1995\)](#):

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

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

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

**Step 12.3:** Sample  $\phi$ :

$$\phi^{(\ell)} \sim 1_{[B=1]} \text{Gamma} \left( 1 + \sum_{k=1}^K 1_{[n_k^{(\ell)} > 0]}, 1 - \ln \epsilon \right) \quad (\text{A.14})$$

$$+ 1_{[B=0]} \text{Gamma} \left( \sum_{k=1}^K 1_{[n_k^{(\ell)} > 0]}, 1 - \ln \epsilon \right)$$

## B. 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  $\ell$ th draw from the posterior distribution of the parameters of the AVDPM approach. AIC and WAIC are calculated as follows:

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

where  $L \left( \hat{\theta} \mid \mathbf{t}, \mathbf{x}, \mathbf{a}, \mathbf{d}, \mathbf{z}^{\mathbf{A}}, \mathbf{z}^{\mathbf{M}} \right)$  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 retained draws from the posterior distribution, and

$$\begin{aligned} &f \left( t_{i,1}, t_{i,2} \mid x_{i,1}, x_{i,2}, z_i^{\mathbf{A}}, z_i^{\mathbf{M}}, a_i, d_{i,1}, d_{i,2}; \theta^{(\ell)} \right) \tag{B.1} \\ &= \frac{\sum_{k=1}^K \pi_k^{(\ell)} \left[ \prod_{j=1}^2 f \left( t_{i,j} \mid x_{i,j}, d_{i,j}, a_i; \alpha_j^{(\ell)}, \beta_j^{(\ell)}, \gamma_{k,j}^{*(\ell)} \right) \right] f \left( z_i; \zeta_k^{*(\ell)} \right)}{\sum_{k=1}^K \pi_k^{(\ell)} f \left( z_i; \zeta_k^{*(\ell)} \right)}. \end{aligned}$$

In this equation we take the joint distribution of  $T_{i,1}$  and  $T_{i,2}$  conditional to the values of  $Z_i^{\mathbf{A}}$  and  $Z_i^{\mathbf{M}}$ , in order to make the three models comparable among them.

When computing the WAIC by gender, for the males we compute:

$$\begin{aligned} &f \left( t_{i,1} \mid t_{i,2}, x_{i,1}, x_{i,2}, z_i^{\mathbf{A}}, z_i^{\mathbf{M}}, a_i, d_{i,1}, d_{i,2}; \theta^{(\ell)} \right) \tag{B.2} \\ &= \frac{\sum_{k=1}^K \pi_k^{(\ell)} \left[ \prod_{j=1}^2 f \left( t_{i,j} \mid x_{i,j}, d_{i,j}, a_i; \alpha_j^{(\ell)}, \beta_j^{(\ell)}, \gamma_{k,j}^{*(\ell)} \right) \right] f \left( z_i \mid \zeta_k^{*(\ell)} \right)}{\sum_{k=1}^K \pi_k^{(\ell)} f \left( t_{i,2} \mid x_{i,2}, d_{i,2}, a_i; \alpha_2^{(\ell)}, \beta_2^{(\ell)}, \gamma_{k,2}^{*(\ell)} \right) f \left( z_i \mid \zeta_k^{*(\ell)} \right)}. \end{aligned}$$

Similar calculations are performed for the WAIC of the AVDPM for the females.

## References

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

- Beraha, M., Guglielmi, A., Quintana, F. A., Iorio, M. D., Eriksson, J. G. and Yap, F. (2023), ‘Childhood obesity in Singapore: A Bayesian nonparametric approach’, *Statistical Modelling* **0**(0), 1471082X231185892.  
**URL:** <https://doi.org/10.1177/1471082X231185892>
- Betancourt, M. J. (2017), ‘Identifying Bayesian mixture models’.  
**URL:** [https://betanalpha.github.io/assets/case\\_studies/identifying\\_mixture\\_models.html](https://betanalpha.github.io/assets/case_studies/identifying_mixture_models.html)
- Blackwell, D. and MacQueen, J. B. (1973), ‘Ferguson distributions via Polya urn schemes’, *Ann. Statist.* **1**(2), 353–355.  
**URL:** <https://doi.org/10.1214/aos/1176342372>
- Blei, D., Kucukelbir, A. and McAuliffe, J. (2016), ‘Variational inference: A review for statisticians’, *Journal of the American Statistical Association* **112**.
- Carriere, J. F. (2000), ‘Bivariate survival models for coupled lives’, *Scandinavian Actuarial Journal* **2000**(1), 17–32.  
**URL:** <https://doi.org/10.1080/034612300750066700>
- De Iorio, M., Müller, P., Rosner, G. and MacEachern, S. (2004), ‘An ANOVA model for dependent random measures’, *Journal of the American Statistical Association* **99**(465), 205–215.  
**URL:** <https://doi.org/10.1198/016214504000000205>
- Denuit, M. and Cornet, A. (1999), Multilife premium calculation with dependent future lifetimes.
- Denuit, M., Dhaene, J., Le Bailly De Tillegem, C. and Teghem, S. (2001), ‘Measuring the impact of a dependence among insured life lengths’, *Belgian Actuarial Bulletin* pp. 18–39.
- Deresa, N., Van Keilegom, I. and 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. and Waters, H. R. (2013), *Actuarial Mathematics for Life Contingent Risks*, International Series on Actuarial Science, Cambridge University Press.
- Dufresne, F., Hashorva, E., Ratovomirija, G. and Toukourou, Y. (2018), ‘On age difference in joint lifetime modelling with life insurance annuity applications’, *Annals of Actuarial Science* **12**(2), 350–371.
- Dunson, D. B. (2010), *Nonparametric Bayes applications to biostatistics*, Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press, p. 223–273.

- Escobar, M. D. and 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. and 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>
- Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A. and Rubin, D. B. (2013), *Bayesian Data Analysis, Third Edition*, Chapman & Hall/CRC Texts in Statistical Science, Taylor & Francis.
- Gorfine, M. and 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. and Wolfe, R. A. (2002), ‘A frailty model for informative censoring’, *Biometrics* **58**(3), 510–520.
- Ishwaran, H. and 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. and Walker, S. (2011), ‘Slice sampling mixture models’, *Statistics and Computing* **21**, 93–105.
- 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>
- Müller, P., Erkanli, A. and 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. and 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. and 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. and 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. and 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. and 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. and 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. and van den Heuvel, E. R. (2024), ‘A Dirichlet Process Mixture Regression model for the analysis of competing risk events’, *Insurance: Mathematics and Economics* **116**, 95–113.  
**URL:** <https://www.sciencedirect.com/science/article/pii/S0167668724000295>
- Vaupel, J. W., Manton, K. G. and 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>
- 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.

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

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

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

---

Francesco Ungolo<sup>1,2,\*</sup> and Patrick J. Laub<sup>1</sup>

<sup>1</sup>*School of Risk and Actuarial Studies, University of New South  
Wales, Kensington NSW 2052, Australia*

<sup>2</sup>*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*

May 23, 2024

## 1 Data cleaning operations

The data preparation steps are listed below:

1. Discard observations with same-gender couples, as in [Frees et al. \(1996\)](#) and [Deresa et al. \(2022\)](#);
2. Eliminate records of couples where males and females aged lesser than 40 at the start of the observation period. In this way, the Gompertz model is a reasonable fit for these data;
3. Check of records of couples where both members are alive at the start of the observation period;
4. Elimination of duplicated records.

## 2 Occupied mixture components throughout the MCMC iterations

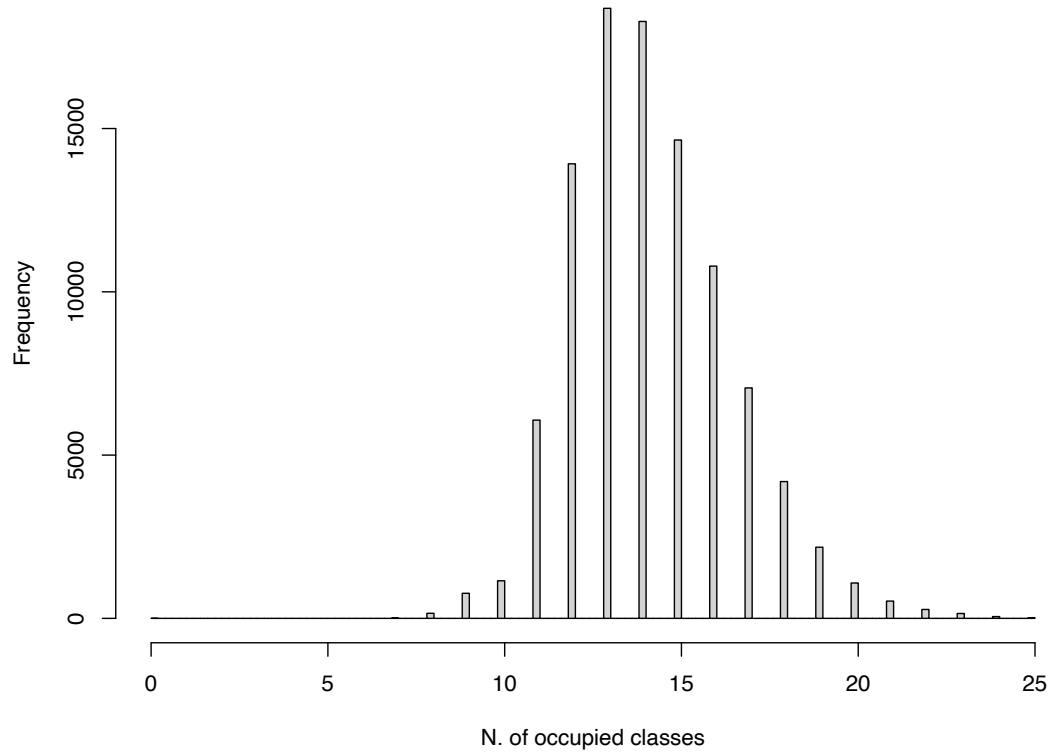


Figure 2.1: Histogram of the number of occupied classes throughout the MCMC iterations.

### 3 Results of the competing models

Table 3.1: Parameter estimates and corresponding standard errors of the Base Gompertz and Proportional hazard models.

Parameter	Base Gompertz		Proportional hazard	
	Estimate	St. err.	Estimate	St. err.
$\alpha_1$	-3.92	0.0367	-3.97	0.0715
$\beta_1$	0.0991	0.0046	0.0987	0.0046
$\delta_{1,1}$	—	—	-0.0035	0.0279
$\delta_{1,2}$	—	—	0.0697	0.0827
$\alpha_2$	-4.85	0.0581	-5.10	0.1143
$\beta_2$	0.1243	0.0074	0.1309	0.0077
$\delta_{2,1}$	—	—	0.0053	0.0448
$\delta_{2,2}$	—	—	0.3234	0.1208

## 4 Joint life annuity factor

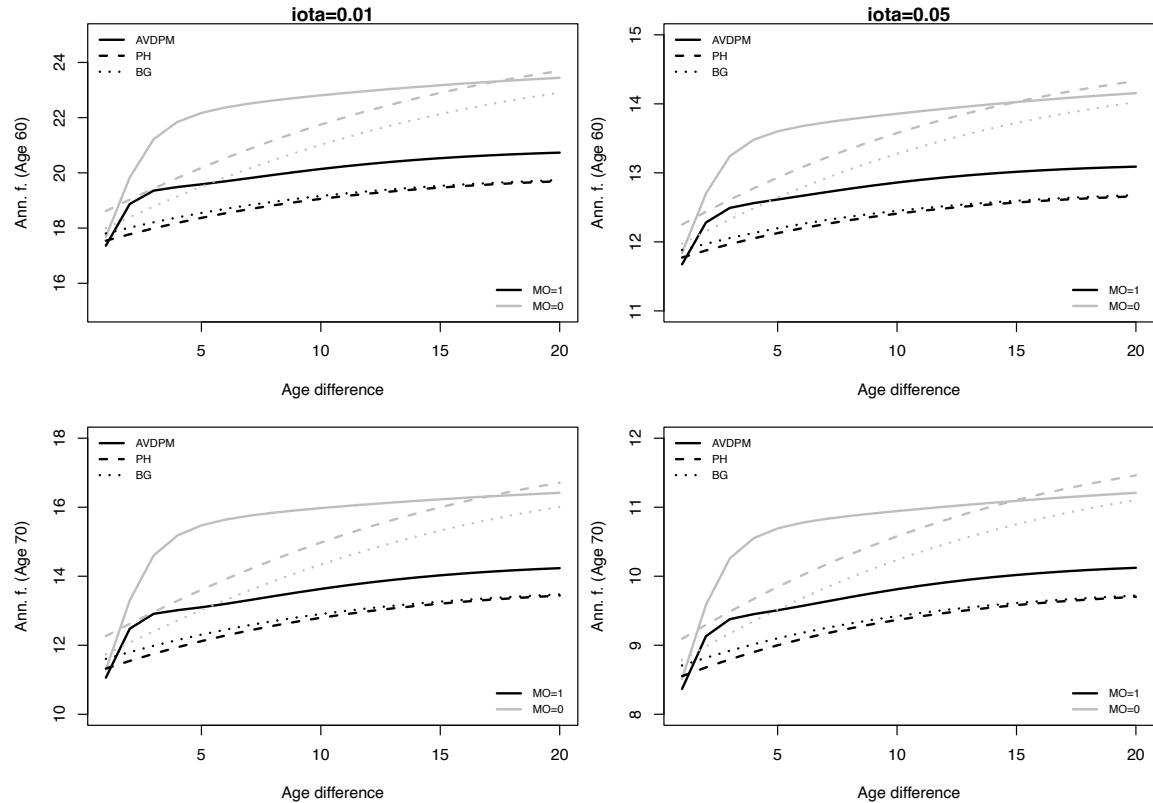


Figure 4.1: Joint life annuity factor calculated using the AVDPM model (solid line), the PH model (dashed line) and the BG model (dotted line) for different values of age difference ( $\exp(Z^A)$ , x-axis) and  $Z^M$  when the oldest member is aged 60 (top panel) and 70 (bottom panel), for  $\iota = 1\%$  (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>