Astin Bulletin (2025), **55**, pp. 50–75
doi:10.1017/asb.2024.34

International Actuarial Association
Association Actuarielle Internationale

RESEARCH ARTICLE

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

Francesco Ungolo[1,2] and Patrick J. Laub[1]

[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; Email: f.ungolo@unsw.edu.au

Received: 23 May 2024; **Revised:** 5 September 2024; **Accepted:** 24 October 2024; **First published online:** 2 December 2024

Keywords Dependent lifetimes; survival analysis; Dirichlet process; Bayesian analysis; life insurance; MCMC; mixture models.

Abstract

The analysis of insurance and annuity products issued on multiple lives requires the use of statistical models which account for lifetime dependence. This paper presents a Dirichlet process mixture-based approach that allows to model dependent lifetimes within a group, such as married couples, accounting for individual as well as group-specific covariates. The model is analyzed 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.

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. For example, 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% 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% 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 fields, with copula models being among the most employed ones. The aforementioned paper of Frees *et al*. (1996) analyzes 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 presence of left-truncation and dependent censoring. Bivariate survival data are also modeled using Archimedean

© The Author(s), 2024. Published by Cambridge University Press on behalf of The International Actuarial Association.

https://doi.org/10.1017/asb.2024.34 Published online by Cambridge University Press

copulas for the spouses' stochastic intensities, see Luciano *et al*. (2008), and the non-parametric estimators of Lopez (2012) and Gribkova and Lopez (2015). Youn and Shemyakin (1999) were the first to account for covariates when modeling 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 likely 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. A first alternative to copula models to account for lifetime dependence is the use of Markovian models, drawing on the earlier idea of Norberg (1989), such as the papers of Denuit and Cornet (1999), Spreeuw and Owadally (2013) and Ji *et al*. (2011). In particular, Ji *et al*. (2011) and Spreeuw and Owadally (2013) also consider the temporary impact of bereavement on the mortality of the surviving spouse through a semi-Markov model. Ji *et al*. (2012) discuss the application of this semi-Markov model when analyzing reverse mortgage terminations. We refer to Ji *et al*. (2011) for an account of the key differences between Markov models and copulas to model dependent lifetimes.

Another possibility to model-dependent lifetimes 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, the paper of Gourieroux and Lu (2015) extends a Freund model accounting for the bereavement effect through a common gamma-distributed frailty for the analysis of mortality dependence between couple members, and Yashin and Iachine (1995) develop a correlated gamma frailty model for the analysis of the joint lifetime of Danish twins. The lifetime dependence of the latter was analyzed also in the paper of van den Berg and Drepper (2022) using a mixed proportional hazard model *à la* Abbring and van den Berg (2003) with a frailty which follows a Cherian bivariate Gamma distribution. They find that the dependence can be chiefly attributed either to genetic features or to early childhood upbringing, rather than the bereavement effect. Similarly, the paper of Lu (2017) analyzes the mortality profile of couples in a portfolio of joint life annuities of a French insurer using a flexible bivariate semi-parametric model based on the Gamma distribution, which is motivated by its mathematical tractability.

In the field of biostatistics, a closely related problem is given by modeling 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 current approaches based on copula, Markov multi-state models and random effects are based on specific parametric or semi-parametric assumptions about lifetime dependence. Their possible misspecification may have the impact of yielding biased estimates for the joint (conditional) distributions of interest. The papers of Ungolo and van den Heuvel (2022, 2024) overcome this potential misspecification issue by using a multivariate random effect with a discrete distribution and unknown location and number of levels, which are to be estimated.

On the other hand, existing approaches using random effects do not account for covariates in their distribution in order 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 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.

https://doi.org/10.1017/asb.2024.34 Published online by Cambridge University Press

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 $\left(W_1, \ldots, W_q\right)$ of Γ, then

$$\left(G\left(W_1\right), \ldots, G\left(W_q\right)\right) \sim \text{Dirichlet}\left(\phi G_0\left(W_1\right), \ldots, \phi G_0\left(W_q\right)\right)$$

which we write $G \sim \text{DP}\left(G_0, \phi\right)$. 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* ϕ (see Gelman *et al*., 2013, Chapter 23 and De Iorio *et al*., 2004). The ϕ parameter captures the degree of shrinkage of G toward 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:

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

where $\gamma_k \overset{\text{i.i.d.}}{\sim} G_0$ for $k = 1, 2, \ldots$, and δ_{γ_k} denotes the Dirac measure that 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 $\psi_k \overset{\text{i.i.d.}}{\sim} \text{Beta}\left(1, \phi\right)$ as follows:

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

where $\psi_{1:k} = \left(\psi_1, \ldots, \psi_k\right)$. 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 ith unit, whose density $f\left(\cdot; \beta, \gamma_i\right)$ is indexed by a *global* parameter vector β, common to all units, and by a unit-specific parameter vector γ_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 γ_i are drawn from a discrete distribution G, which follows a Dirichlet process. Therefore, convolving the density of T_i with $G \sim \text{DP}\left(G_0, \phi\right)$ yields a Dirichlet process mixture (DPM) model (Lo, 1984):

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

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

https://doi.org/10.1017/asb.2024.34 Published online by Cambridge University Press

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 γ_i, such as the normal distribution.

This DP construction is flexible since despite its apparent nonparametric nature, it regularizes the distribution of G toward 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 denote the common values of γ_i across the n units in the sample. 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 observe the units in a sequence, these are more likely to be in a certain class with a probability that depends on the number of units already therein, or to belong to a newly created class with a probability which depends on ϕ (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 $t_1, \ldots, t_n$:

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

3. The Augmented Variable DPM model

This section introduces the augmented variable Dirichlet process mixture (AVDPM) model for the analysis of nonexchangeable 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 $T_i = \left(T_{i,1},\ T_{i,2}\right)$, where $T_{i,1}$ and $T_{i,2}$ denote the random future lifetime of husband and wife, respectively, for the ith 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 that 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

https://doi.org/10.1017/asb.2024.34 Published online by Cambridge University Press

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 ith 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 $(T_{i,1}, T_{i,2}, Z_i)$:

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

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_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.

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 $(\gamma_{i,1}, \gamma_{i,2}, \zeta_i)$. The individual-specific fixed covariate vector $x_{i,j}$ directly enters the regression function of $T_{i,j}$. On the other hand, the ith couple-specific covariate vector Z_i is assumed to affect the individual lifetime distribution only indirectly, through $(\gamma_{i,1}, \gamma_{i,2}, \zeta_i)$. Indeed, it is reasonable to assume that individual variables like medical status and age affect the individual mortality *directly*, while couple-specific variables like household income can have an *indirect* effect on mortality through the affordability of good quality healthcare services. This model differs from treating Z_i as an exogenous variable, for example by including this observable variable among the set of fixed covariates within a regression function, such as in the paper of Deresa *et al*., (2022). Both approaches allows to estimate the effect of Z_i on the distribution of $T_{i,j}$, as shown in Equation (3.3) below. In addition, the density of Equation (3.1) allows to have an understanding of the differing strength of the statistical association between $T_{i,1}$ and $T_{i,2}$ as we show in the empirical analysis of Section 6.

We complete the model specification by assuming that the multivariate discrete distribution G is a random draw from a Dirichlet process, and write $G \sim \mathrm{DP}(G_0, \phi)$. For simplicity, we assume $G_0 = G_{0,\gamma} \times G_{0,\zeta}$, that is γ and ζ are independently distributed in the base distribution. This specification still induces a dependence between γ and ζ through the concentration parameter ϕ. 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 ϕ. 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 γ_1 and γ_2, as more reasonable for the analysis of male–female couples.

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 $T_{i,2}$, and that it allows for both positive and negative dependencies between lifetimes, similarly to the model of Lu (2017).

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 (AVDPM)model:

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

https://doi.org/10.1017/asb.2024.34 Published online by Cambridge University Press

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 $\gamma_{i,j}$ and ζ_i.

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

$$f\left(t_{i,j} \mid x_{i,j}, z_i; \beta_j, \gamma^*_{\cdot,j}, \zeta^*_{\cdot}\right) = \frac{\sum_{k=1}^{\infty} \pi_k f\left(t_{i,j} \mid x_{i,j}; \beta_j, \gamma^*_{k,j}\right) f\left(z_i \mid \zeta^*_k\right)}{\sum_{k=1}^{\infty} \pi_k f\left(z_i \mid \zeta^*_k\right)}. \tag{3.3}$$

where $\gamma^*_{\cdot,j} = (\gamma^*_{1,j}, \gamma^*_{2,j}, \ldots)$ and $\zeta^*_{\cdot} = (\zeta^*_1, \zeta^*_2, \ldots)$.

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:

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

where

$$\mathbb{P}\left(T_1 < t, \quad T_2 < t \mid x_1, x_2, z\right) = \frac{\sum_{k=1}^{\infty} \pi_k f\left(z \mid \zeta^*_k\right) \left[\prod_{j=1}^{2} \int_0^t f\left(u \mid x_j; \beta_j, \gamma^*_{k,j}\right) \mathrm{d}u\right]}{\sum_{k=1}^{\infty} \pi_k f\left(z \mid \zeta^*_k\right)}.$$

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

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:

$$\mathbb{P}(W = k \mid Z = z) = \frac{\pi_k f\left(z \mid \zeta^*_k\right)}{\sum_{h=1}^{\infty} \pi_h f\left(z \mid \zeta^*_h\right)}. \tag{3.5}$$

In this way, we can understand how the various couples belonging to the kth 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 analyzed 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 in 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

https://doi.org/10.1017/asb.2024.34 Published online by Cambridge University Press

severity of claims in nonlife 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 December 29, 1988 and December 31, 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 have shown 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 jth member of the ith 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 $Z_i^A = \ln\left(|X_{i,1} - X_{i,2}|\right)$ 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, $Z_i = \left(Z_i^A, Z_i^M\right)$.

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 ith policy:

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

where $\bar{x} = 70$ is set in order to decrease the posterior correlation between α_j and β_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 analyzed a pension scheme dataset with older scheme members.

We consider the male–female lifetime dependence by assuming a couple-specific frailty term $\exp\left(\gamma_{i,j}\right)$ and a joint model for $\left(\gamma_{i,1}, \gamma_{i,2}\right)$. The parameter α_j denotes the log-baseline hazard function and β_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:

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

https://doi.org/10.1017/asb.2024.34 Published online by Cambridge University Press

For the couple-specific covariates, we assume that $Z_i^A \sim \mathrm{N}\left(\zeta_i^A, \sigma_A^2\right)$ and $Z_i^M \sim \text{Bernoulli}\left(\zeta_i^M\right)$. 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 $\left(T_{i,1}, T_{i,2}, Z_i\right)$ has an additional layer of flexibility since the effect of Z_i on the hazard function (through dependent $\gamma_{i,j}$, ζ_i^A and ζ_i^M) is not necessarily proportional nor monotone.

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

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

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 ζ_i^M follow 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 ζ_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 that 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). Indeed, as motivated in Section 5.1, this is a conservative choice since only few components are sufficient to model the additional heterogeneity in the data, while all other components turn out to have a mixture weight nearly equal to zero. 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

https://doi.org/10.1017/asb.2024.34 Published online by Cambridge University Press

heterogeneity in the data. Given the sample size of this dataset and the dimensionality of G, we specify a conjugate Gamma(6, 12) prior for ϕ (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 Σ_γ 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 toward 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 Σ_γ turns out to be sensitive with respect to λ, 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 ζ_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 lesser 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 toward larger values, and the posterior density of of s_A^2 is very flat and centered toward 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 $\exp(\alpha_j) \sim$ Gamma (1, 1) (which ensures conditional prior-posterior conjugacy) and β_j $(j = 1, 2)$ follows *a priori* a truncated normal distribution with mean 0.1 and 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 β around 0.10 (see for example Ungolo *et al*., 2020 and Richards, 2008). Finally, we assume that $\sigma_A^2 \sim$ Inv-Gamma (2, 1).

Our sensitivity analysis shows how the posterior inference for β and σ_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 α 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 that 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 $\mathbf{t} = (t_{1,1}, t_{1,2}, \ldots, t_{n,1}, t_{n,2})$, $\mathbf{x} = (x_{i,1}, x_{i,2}, \ldots, x_{n,1}, x_{n,2})$, $\mathbf{d} = (d_{1,1}, d_{1,2}, \ldots, d_{n,1}, d_{n,2})$, $\mathbf{z^A} = (z_1^A, \ldots, z_n^A)$, $\mathbf{z^M} = (z_1^M, \ldots, z_n^M)$, $\mathbf{a} = (a_1, \ldots, a_n)$, $\alpha = (\alpha_1, \alpha_2)$, $\beta = (\beta_1, \beta_2)$, $\psi = (\psi_1, \ldots, \psi_{K-1})$,

https://doi.org/10.1017/asb.2024.34 Published online by Cambridge University Press

$\gamma^* = \left(\gamma^*_{1,1}, \gamma^*_{1,2}, \ldots, \gamma^*_{K,2}\right)$, $\zeta^{A*} = \left(\zeta_1^{A*}, \ldots, \zeta_K^{A*}\right)$ and $\zeta^{M*} = \left(\zeta_1^{M*}, \ldots, \zeta_K^{M*}\right)$. 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 auxiliary variable 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 $\mathbf{s} = \left(s_{1,1}, \ldots, s_{n,K}\right)$ is given by

$$
\begin{aligned}
&L\left(\alpha, \beta, \gamma^*, \zeta^{A*}, \zeta^{M*}, \sigma_A^2, \psi \mid \mathbf{t}, \mathbf{x}, \mathbf{d}, \mathbf{a}, \mathbf{z}^{\mathbf{A}}, \mathbf{z}^{\mathbf{M}}, \mathbf{s}\right) \\
&\propto \prod_{i=1}^{n} \prod_{k=1}^{K} \pi_k^{s_{i,k}} \left\{ \left[\prod_{j=1}^{2} f\left(t_{i,j} \mid x_{i,j}, d_{i,j}, a_i; \alpha_j, \beta_j, \gamma^*_{k,j}\right) \right] f\left(z_i^A \mid \zeta_k^{A*}, \sigma_A^2\right) \left(\zeta_k^{M*}\right)^{z_i^M} \left(1 - \zeta_k^{M*}\right)^{1 - z_i^M} \right\}^{s_{i,k}}
\end{aligned} \tag{5.1}
$$

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

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

where the numerator is the likelihood contribution for a noninformative 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

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

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:

$$
\begin{aligned}
&p\left(\alpha, \beta, \gamma^*, \zeta^{A*}, \zeta^{M*}, \sigma_A^2, \Sigma_\gamma, m_A, s_A^2, \phi, \psi \mid \mathbf{t}, \mathbf{x}, \mathbf{d}, \mathbf{a}, \mathbf{z}^{\mathbf{A}}, \mathbf{z}^{\mathbf{M}}, \mathbf{s}\right) \\
&\propto L\left(\alpha, \beta, \gamma^*, \zeta^{A*}, \zeta^{M*}, \sigma_A^2, \psi \mid \mathbf{t}, \mathbf{x}, \mathbf{d}, \mathbf{a}, \mathbf{z}^{\mathbf{A}}, \mathbf{z}^{\mathbf{M}}, \mathbf{s}\right) \left[\prod_{j=1}^{2} \underbrace{p\left(\alpha_j \mid 1, 1\right)}_{\text{Gamma}(1,1)} \right] \left[\prod_{j=1}^{2} \underbrace{p\left(\beta_j \mid 0.1, 0.25\right)}_{\text{N}_+(0.1,\ 0.25)} \right]
\end{aligned} \tag{5.4}
$$

https://doi.org/10.1017/asb.2024.34 Published online by Cambridge University Press

$$\times \left[\prod_{k=1}^{K} \underbrace{p\left(\gamma_{k,\cdot}^{*} \mid \mathbf{0}, \Sigma_{\gamma}\right)}_{\text{MVN}\left(\mathbf{0},\Sigma_{\gamma}\right)} \underbrace{p\left(\zeta_{k}^{A*} \mid m_{A}, s_{A}^{2}\right)}_{\text{N}\left(m_{A},s_{A}^{2}\right)} \underbrace{p\left(\zeta_{k}^{M*} \mid 13.31, 4.44\right)}_{\text{Beta}(13.31,4.44)} \right] \underbrace{p\left(m_{A} \mid 3, 0.5\right)}_{\text{N}(3,0.5)} \underbrace{p\left(\Sigma_{\gamma} \middle| 5, \begin{bmatrix} 2 & 1.2 \\ 1.2 & 2 \end{bmatrix}\right)}_{\text{Inv-Wishart}\left(5, \begin{bmatrix} 2 & 1.2 \\ 1.2 & 2 \end{bmatrix}\right)}$$

$$\times \underbrace{p\left(\sigma_{A}^{2} \mid 2, 1\right)}_{\text{Inv-Gamma}(2,1)} \underbrace{p\left(s_{A}^{2} \mid 3,\ 0.5\right)}_{\text{Inv-Gamma}(3,\ 0.5)} \left[\prod_{k=1}^{K-1} \underbrace{p\left(\psi_{k} \mid 1, \phi\right)}_{\text{Beta}(1,\phi)} \right] \underbrace{p\left(\phi \mid 6, 12\right)}_{\text{Gamma}(6,12)} .$$

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 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 1000 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 toward a stationary distribution.

The trace plots of the parameters show that the chains converge toward a stationary distribution for all parameters, except for $\gamma_{k,1}^{*}$, $\gamma_{k,2}^{*}$, ζ_{k}^{A*}, ζ_{k}^{M*}, and π_{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 α_{j} converge toward 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 α_{j} and $\gamma_{k,j}$.

We do not analyze these two aspects for $\left(\gamma_{k,1}^{*}, \gamma_{k,2}^{*}, \zeta_{k}^{A*}, \zeta_{k}^{M*}\right)$ 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 β_{j}.

The overall convergence of the MCMC sampler is further analyzed through the trace plot of the clustering entropy shown in Figure 1 for two chains, as used for the analysis of mixture models (Beraha *et al*. 2023). The entropy is calculated at the ℓth iteration as follows:

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

We note that all chains converge toward 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

https://doi.org/10.1017/asb.2024.34 Published online by Cambridge University Press

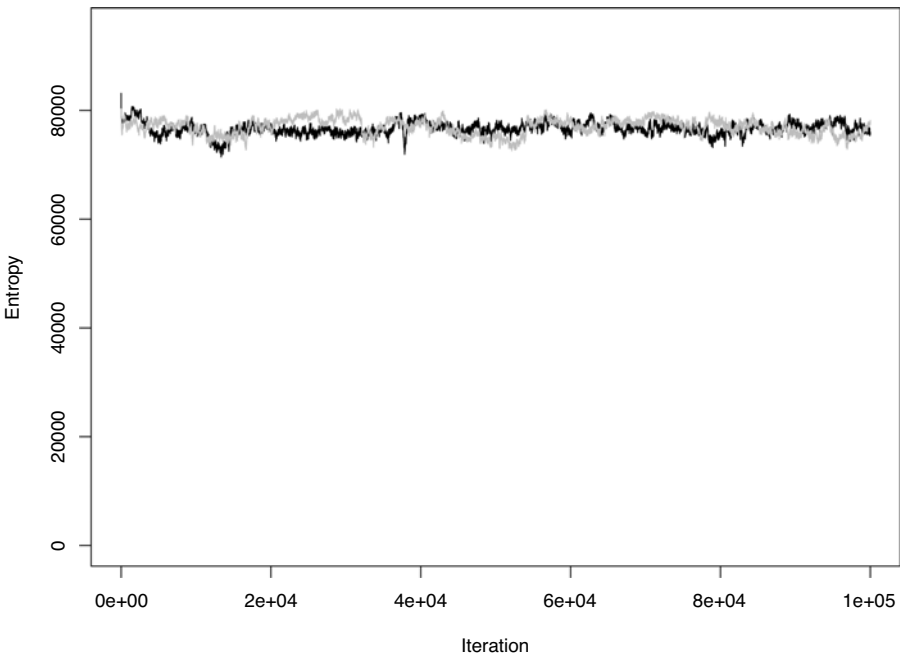


Figure 1. *Trace plot of the entropy across all iterations of the MCMC sampler for two different runs.*

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 shows 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 $\left(T_{i,1}, T_{i,2}, Z_i^A, Z_i^M\right)$.

Table 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.

Six mixture components are associated with mixture weights π_k whose posterior mean totals around 93.5% ($k = 1, 2, 3, 4, 6, 9$), twelve have mixture weights below 0.1%, and the remaining ones are associated with mixture weights below 2%. Each component is characterized by different values of the class-specific parameters $\left(\gamma_{k,1}^*, \gamma_{k,2}^*, \zeta_k^{A*}, \zeta_k^{M*}\right)$.

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 2. For example,

https://doi.org/10.1017/asb.2024.34 Published online by Cambridge University Press

Table 1. *Posterior summaries of α, β, γ_k^*, ζ_k^{A*}, ζ_k^{M*}, π_k for $k = 1, 2, 3, 4, 6, 9$ and ϕ.*

Parameter	Mean	95% Cred. Int.	Parameter	Mean	95% Cred. Int.
α_1	−3.45	(−3.71, −3.27)	α_2	−4.53	(−4.90, −4.12)
β_1	0.1064	(0.0963; 0.1160)	β_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)
ζ_1^{A*}	1.13	(0.99; 1.28)	ζ_1^{M*}	0.66	(0.52; 0.75)
ζ_2^{A*}	1.71	(1.63; 1.81)	ζ_2^{M*}	0.95	(0.93; 0.98)
ζ_3^{A*}	0.50	(−4.15 2.46)	ζ_3^{M*}	0.73	(0.53; 0.90)
ζ_4^{A*}	−0.36	(−1.49; 0.68)	ζ_4^{M*}	0.58	(0.47; 0.77)
ζ_6^{A*}	0.24	(−1.70; 1.90)	ζ_6^{M*}	0.68	(0.53; 0.88)
ζ_9^{A*}	−1.73	(−2.90; −1.02)	ζ_9^{M*}	0.54	(0.42; 0.68)
π_1	0.2291	(0.1321; 0.3167)	π_2	0.4078	(0.3240; 0.4834)
π_3	0.0328	(0.0005; 0.1334)	π_4	0.1205	(0.0272; 0.2311)
π_6	0.1064	(0.0020; 0.2329)	π_9	0.0387	(0.0112; 0.0639)
ϕ	0.97	(0.55; 1.54)			

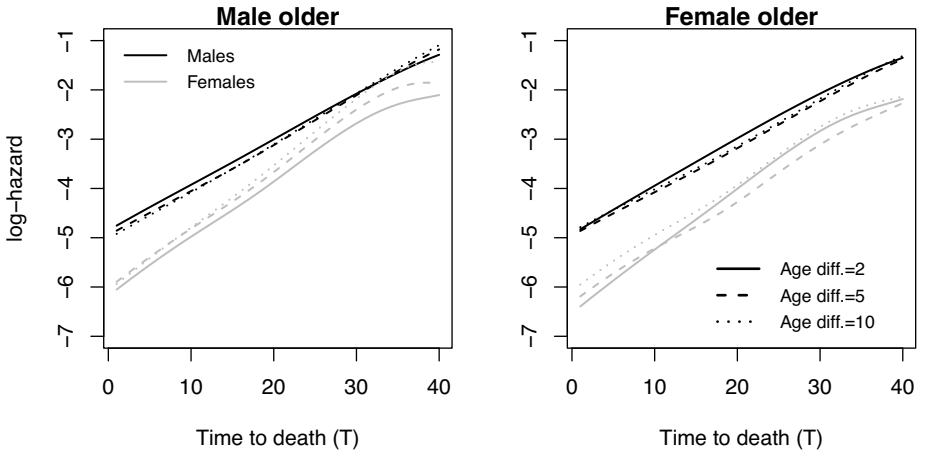


Figure 2. *Hazard function for males (black) and females (gray) aged $x_{\cdot,j} = 60$ within a couple for different values of the covariates male older and age difference (*$\exp(Z^A)$*).*

at younger ages, when the female is five 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.

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 2). This hazard function is calculated as follows:

https://doi.org/10.1017/asb.2024.34 Published online by Cambridge University Press

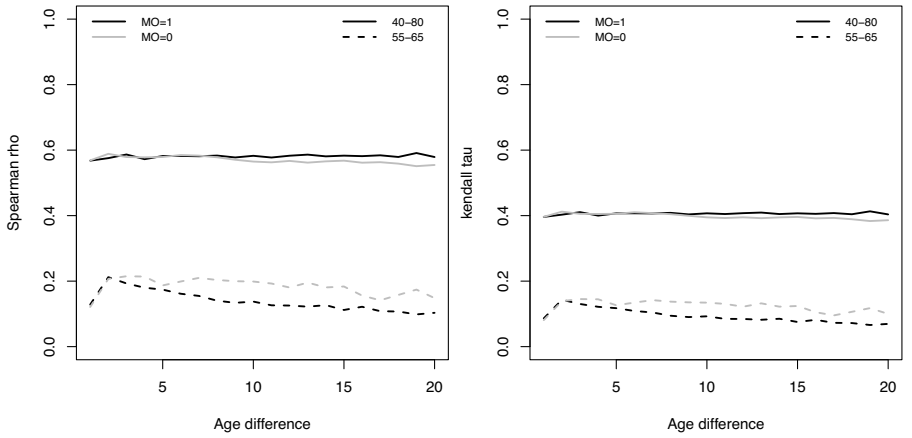


Figure 3. *Value of the Spearman ρ (left panel) and of the Kendall τ (right panel) for different values of AD and MO.*

$$\mu\left(t \mid x_{\cdot j}, z^A, z^M; \alpha_j, \beta_j, \gamma^*_{\cdot j}, \pi, \zeta^{A*}, \zeta^{M*}\right) = \frac{\sum_{k=1}^{K} \pi_k f\left(t \mid x_{\cdot j}; \alpha_j, \beta_j, \gamma^*_{kj}\right) f\left(z^A \mid \zeta_k^{A*}, \sigma_A^2\right) \left(\zeta_k^{M*}\right)^{z^M} \left(1-\zeta_k^{M*}\right)^{1-z^M}}{\sum_{k=1}^{K} \pi_k \left(1 - F\left(t \mid x_{\cdot j}; \alpha_j, \beta_j, \gamma^*_{kj}\right)\right) f\left(z^A \mid \zeta_k^{A*}, \sigma_A^2\right) \left(\zeta_k^{M*}\right)^{z^M} \left(1-\zeta_k^{M*}\right)^{1-z^M}} \tag{6.1}$$

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 analyzed to compare the models.

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 upward 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 years 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 modeling 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 ρ and the Kendall τ 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 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 ρ and τ are much smaller than the previous case, while T_1 and T_2 are still positively associated. For an age difference of one year, ρ and τ are relatively larger and then have a constant value for every value of Z^A, especially for the case where

https://doi.org/10.1017/asb.2024.34 Published online by Cambridge University Press

Table 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	–
$\zeta^*_{\mathrm{AD},\cdot}$	1.13	1.71	−0.36	−1.72	–
$\zeta^*_{\mathrm{MO},\cdot}$	0.66	0.95	0.58	0.54	–

the female is the oldest couple member. In the opposite case where $Z^M = 1$, we observe a decreasing value of ρ and τ 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 ρ and τ 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.

Class analysis

The mixture modeling 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 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 as follows:

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

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

The ith couple is then hard-assigned to the kth 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 2 illustrates the key features of groups 1, 2, 4, and 9, alongside the posterior mean of their class-specific parameters $\left(\gamma^*_{\cdot,1}, \gamma^*_{\cdot,2}, \zeta^{A*}_{\cdot}, \zeta^{M*}_{\cdot}\right)$. The difference between % Composition in Table 2 and π is due to the specific classification rule we use.

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 that 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.

https://doi.org/10.1017/asb.2024.34 Published online by Cambridge University Press

Table 3. *AIC and WAIC of the models (the lowest value is shown in bold).*

(W)AIC	BG	PH	AVDPM
In sample	13,603.41	13,603.11	**13,527.50**
Out of sample	4537.42	4541.64	**4515.14**

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 γ^* posterior mean coefficient. Groups 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)[1] which assume that:

$$\mu^{BG}\left(t \mid x_{i,j}; \alpha_j, \beta_j\right) = \exp\left(\alpha_j + \beta_j\left(x_{i,j} - \bar{x} + t\right)\right)$$

$$\mu^{PH}\left(t \mid x_{i,j}, z_i^A, z_i^M; \alpha_j, \beta_j, \delta_{j,1}, \delta_{j,2}\right) = \exp\left(\alpha_j + \beta_j\left(x_{i,j} - \bar{x} + t\right) + \delta_{j,1} z_i^A + \delta_{j,2} z_i^M\right)$$

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 3), and we chose the model which minimizes the value of these criteria.

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 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.).

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.

[1]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.

https://doi.org/10.1017/asb.2024.34 Published online by Cambridge University Press

Table 4. *AIC and WAIC of the models by gender (the lowest value is shown in bold).*

	Males			Females		
(W)AIC	BG	PH	AVDPM	BG	PH	AVDPM
In sample	9576.52	9579.78	**9498.77**	4026.89	4023.33	**3948.61**
Out of sample	3185.87	3188.66	**3166.74**	1351.55	1352.98	**1330.25**

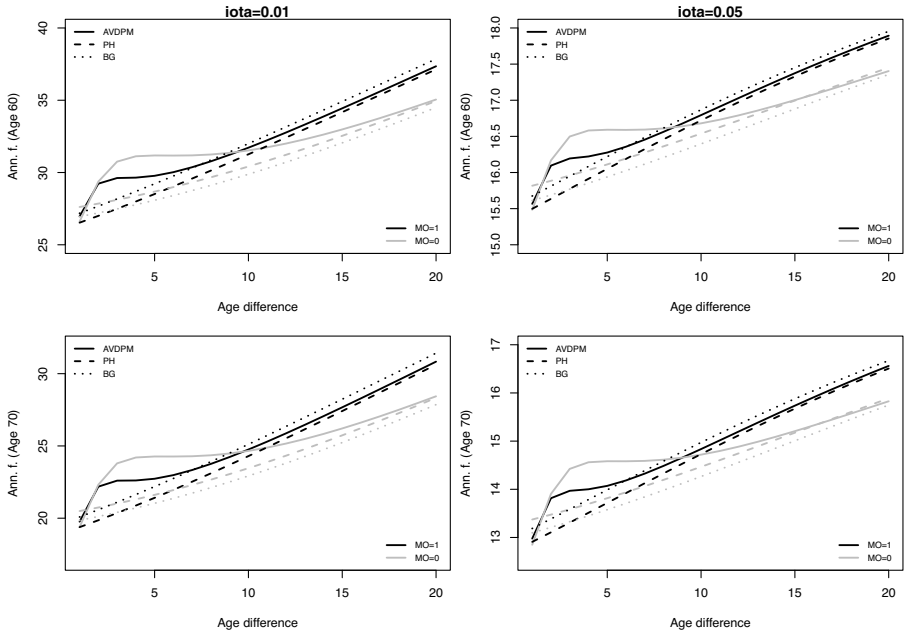


Figure 4. *Last survivor annuity factor calculated using the AVDPM model (solid line), the PH model (dashed line), and the BG model (dotted line) for different values of age difference (*$\exp\left(Z^A\right)$*, x-axis) and* Z^M *when the oldest member is aged 60 (top panel) and 70 (bottom panel), for* $\iota = 1\%$ *(left panel) and* $\iota = 5\%$ *(right panel).*

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):

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

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

where the last survivor function $S_{\overline{x_1,x_2}}(t \mid z)$ was defined in Section 3, and $S_{x_1,x_2}(t \mid z)$ 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 4 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 $Z = \left(Z^A, Z^M\right)$, and two different forces of interest

https://doi.org/10.1017/asb.2024.34 Published online by Cambridge University Press

$\iota = (0.01, 0.05)$. 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 afterward. 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 that 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 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 years 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 and the effect of the price difference due to the covariates. On the other hand, an increase in the age of the two heads (Figure 4 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 analyzed. Similarly, the copula-based dependence model analyzed 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 years, respectively.

8. Extension of AVDPM to the analysis of the joint lifetimes of nonexchangeable units

So far, we have illustrated the method for the case of male–female 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 ith group includes J_i members, with common set of variables z_i, and individual (within group) characteristics $x_{i,j}$ $(j = 1, \ldots, J_i)$. A possible solution is to model the group-specific joint distribution of the lifetimes and Z as:

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

where Q denotes a suitable distribution function for $\gamma_{i,j}$, 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

https://doi.org/10.1017/asb.2024.34 Published online by Cambridge University Press

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 the group-specific covariates. The use of DPM models enhances the flexibility of the random effects distribution and of the standard parametric assumptions for the covariates. The resulting 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 a fully 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.

One limitation of the AVDPM model is that it does not explicitly disentangle the lifetime dependence due to causal effects, as measured through group-specific covariates, from the dependence due to other sources of unobserved heterogeneity. This aspect can be potentially relevant from a pricing perspective (Gourieroux and Lu, 2015). One possibility, beyond the AVDPM of this work, is to consider a joint model *à la* Lu (2017) where the causal effect is part of the regression function, with couple-specific regression coefficients:

$$\mu\left(t \mid x_{i,j}, z_i; \alpha_j, \beta_j, \delta_{i,j}\right) = \exp\left(\alpha_j + \beta_j\left(x_{i,j} - \bar{x} + t\right) + \delta_{i,j} z_i + \gamma_{i,j}\right) \tag{9.1}$$

with $\left(\delta_{i,1}, \delta_{i,2}\right)$ eventually drawn from either a parametric distribution, or a discrete distribution following a Dirichlet process. In this way, we would be able to account for both the effect of z_i on the individual lifetime, as well as the lifetime dependence given z_i. We leave the further analysis of this model for future research.

Richer parametric structures for the Dirichlet processes are available in the literature, such as the generalized Dirichlet process prior (GDP, see Hjort, 2000), which generalizes the basic DP through an additional parameter, the dependent Dirichlet process (DDP, see MacEachern, 2000) which allows for covariate dependent mixture weights π_k and location of the random effects, and the dependent GDP of Barcella *et al*. (2017) which combines DDP and GDPs. Such extensions can be useful to obtain a better characterization of the groups associated with the mixture components. On the other hand, these approaches would require more computationally intensive MCMC algorithms to estimate the posterior distribution of the parameters.

Another 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. We 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.

Supplementary material. The supplementary material for this article can be found at https://doi.org/0.1017/asb.2024.34

https://doi.org/10.1017/asb.2024.34 Published online by Cambridge University Press

References

Abbring, J.H. and van den Berg, G.J. (2003) The identifiability of the mixed proportional hazards competing risks model. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **65**(3), 701–710.

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

Barcella, W., De Iorio, M., Favaro, S. and Rosner, G. L. (2017) Dependent generalized Dirichlet process priors for the analysis of acute lymphoblastic leukemia. *Biostatistics*, **19**(3), 342–358. https://doi.org/10.1093/biostatistics/kxx042

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*, 1471082X231185892. https://doi.org/10.1177/1471082X231185892

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

Blackwell, D. and MacQueen, J.B. (1973) Ferguson distributions via Polya urn schemes. *Annals of Statistics*, **1**(2), 353–355. 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**, 859–877.

Carriere, J.F. (2000) Bivariate survival models for coupled lives. *Scandinavian Actuarial Journal*, **2000**(1), 17–32. 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. 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 Tilleghem, C. and Teghem, S. (2001) Measuring the impact of a dependence among insured life lengths. *Belgian Actuarial Bulletin*, **1**(1), 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. 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: 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*, pp. 223–273. Cambridge: Cambridge University Press.

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

Ferguson, T.S. (1973) A Bayesian analysis of some nonparametric problems. *Annals of Statistics*, **1**(2), 209–230. 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. 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*, 3rd edn. Boca Raton, FL: 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. https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1541-0420.2010.01470.x

Gourieroux, C. and Lu, Y. (2015) Love and death: A Freund model with frailty. *Insurance: Mathematics and Economics*, **63**(C), 191–203. https://ideas.repec.org/a/eee/insuma/v63y2015icp191-203.html

Gribkova, S. and Lopez, O. (2015) Non-parametric copula estimation under bivariate censoring. *Scandinavian Journal of Statistics*, **42**(4), 925–946. https://onlinelibrary.wiley.com/doi/abs/10.1111/sjos.12144

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

Hjort, N.L. (2000) Bayesian analysis for a generalised Dirichlet process prior. Technical Report, Matematisk Institutt, Universitetet i Oslo.

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. https://doi.org/10.1198/016214501750332758

Ji, M., Hardy, M. and Li, J.S.-H. (2011) Markovian approaches to joint-life mortality. *North American Actuarial Journal*, **15**(3), 357–376. https://doi.org/10.1080/10920277.2011.10597625

Ji, M., Hardy, M. and Li, J.S.-H. (2012) A semi-Markov multiple state model for reverse mortgage terminations. *Annals of Actuarial Science*, **6**(2), 235–257.

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. https://doi.org/10.1214/aos/1176346412

https://doi.org/10.1017/asb.2024.34 Published online by Cambridge University Press

Lopez, O. (2012) A generalization of the Kaplan–Meier estimator for analyzing bivariate mortality under right-censoring and left-truncation with applications in model-checking for survival copula models. *Insurance: Mathematics and Economics*, **51**(3), 505–516. https://www.sciencedirect.com/science/article/pii/S0167668712000881

Lu, Y. (2017) Broken-heart, common life, heterogeneity: Analyzing the spousal mortality dependence. *ASTIN Bulletin*, **47**(3), 837–874.

Luciano, E., Spreeuw, J. and Vigna, E. (2008) Modelling stochastic mortality for dependent lives. *Insurance: Mathematics and Economics*, **43**(2), 234–244. https://www.sciencedirect.com/science/article/pii/S0167668708000929

MacEachern, S. (2000) Dependent Dirichlet process. Technical Report, Department of Statistics, The Ohio State University.

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

Norberg, R. (1989) Actuarial analysis of dependent lives. *Bulletin de l'Association Suisse des Actuaires*, 243–255.

Papaspiliopoulos, O. and Roberts, G.O. (2008) Retrospective Markov Chain Monte Carlo methods for Dirichlet process hierarchical models. *Biometrika*, **95**(1), 169–186. 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. 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. 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. *Annals of Applied Probability*, **7**(1), 110–120.

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

Spreeuw, J. and Owadally, I. (2013) Investigating the broken-heart effect: A model for short-term dependence between the remaining lifetimes of joint lives. *Annals of Actuarial Science*, **7**(2), 236–257.

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. 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. 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*, 09622802211057290. PMID: 35077263. 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. https://www.sciencedirect.com/science/article/pii/S0167668724000295

van den Berg, G.J. and Drepper, B. (2022) A unique bond: Twin bereavement and lifespan associations of identical and fraternal twins. *Journal of the Royal Statistical Society Series A: Statistics in Society*, **185**(2), 677–698. https://doi.org/10.1111/rssa.12789

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. 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. https://doi.org/10.1080/03610910601096262

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

Yashin, A.I. and Iachine, I.A. (1995) Genetic analysis of durations: Correlated frailty model applied to survival of Danish twins. *Genetic Epidemiology*, **12**, 529–538.

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*, pp. 34–38.

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

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:

https://doi.org/10.1017/asb.2024.34 Published online by Cambridge University Press

$$\mathbb{P}\left(W_i^{(\ell)}=k \mid t_{i,1}, t_{i,2}, x_{i,1}, x_{i,2}, a_i, d_{i,1}, d_{i,2}, z_i^A, z_i^M\right) \tag{A.1}$$

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

where

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

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 $\psi_K=1)$:

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

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

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

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

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 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

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

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

https://doi.org/10.1017/asb.2024.34 Published online by Cambridge University Press

Step 4.3: Set:

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

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

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

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 γ^* using the acceptance–rejection sampling method:

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

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

Step 5.1.2: Set $\gamma_k{}' = \gamma_k^{*(\ell-1)} + \mathbf{L}_\mathbf{k}^{(\boldsymbol{\ell-1})} \mathbf{h}^{(\boldsymbol{\ell})}$,
where $\mathbf{L}_\mathbf{k}^{(\boldsymbol{\ell-1})}$ is the lower triangular matrix denoting the Cholesky decomposition of $\mathbf{\Sigma}_{\mathbf{p,k}} = \mathbf{L_k} \mathbf{L_k}^\top$. 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:

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

Step 5.3: Set:

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

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

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

$$\mathbf{\Sigma}_{\mathbf{p,k}}^{(\boldsymbol{\ell})} = \mathbf{L}_\mathbf{k}^{(\boldsymbol{\ell-1})} \left(\mathbf{I_2} + \ell^{-0.6} \left(\min(r'', 1) - 0.234\right) \frac{\mathbf{h}^{(\boldsymbol{\ell})} \mathbf{h}^{(\boldsymbol{\ell})\top}}{||\mathbf{h}^{(\boldsymbol{\ell})}||^2}\right) \mathbf{L}_\mathbf{k}^{(\boldsymbol{\ell-1})\top} \tag{A.11}$$

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

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

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, calculated as:

$$\Lambda_1 = 5 + \sum_{k=1}^{K} 1_{\left[n_k^{(\ell)} > 0\right]} \tag{A.12}$$

https://doi.org/10.1017/asb.2024.34 Published online by Cambridge University Press

$$\Lambda_2 = \begin{bmatrix} 2 & 1.2 \\ 1.2 & 2 \end{bmatrix} + \sum_{k=1}^{K} 1_{\left[n_k^{(\ell)}>0\right]} \gamma_k^{*(\ell)} \gamma_k^{*(\ell)^\top}$$

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

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

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

$$\Lambda_{4,k} = \left(\frac{n_k^{(\ell)}}{\sigma_A^{2(\ell-1)}} + \frac{1}{s_A^{2(\ell-1)}} \right)^{-1}$$

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

$$\Lambda_{5,k} = 13.31 + \sum_{\{i:w_i^{(\ell)}=k\}} z_i^M$$

$$\Lambda_{6,k} = 4.44 + \sum_{\{i:w_i^{(\ell)}=k\}} \left(1 - z_i^M\right)$$

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

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

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

Step 10: Sample $\sigma_A^{2(\ell)}$ from a conjugate Inv-Gamma$\left(\Lambda_9^{(\ell)}, \Lambda_{10}^{(\ell)}\right)$ 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$\left(\Lambda_{11}^{(\ell)}, \Lambda_{12}^{(\ell)}\right)$ distribution where:

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

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

https://doi.org/10.1017/asb.2024.34 Published online by Cambridge University Press

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

Step 12.1: Sample $\epsilon \sim \text{Beta}\left(\phi^{(\ell-1)}+1, n\right)$;
Step 12.2: Sample $B \sim \text{Bernoulli}\left(\pi_\epsilon\right)$, where

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

Step 12.3: Sample ϕ:

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

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 ℓth draw from the posterior distribution of the parameters of the AVDPM approach. AIC and WAIC are calculated as follows:

$$\text{AIC} = -2\ln L\left(\hat{\theta} \mid \mathbf{t}, \mathbf{x}, \mathbf{a}, \mathbf{d}, \mathbf{z^A}, \mathbf{z^M}\right) + 2r$$

$$\text{WAIC} = -2\sum_{i=1}^{n} \ln\left(\frac{1}{H}\sum_{\ell=1}^{H} f\left(t_{i,1}, t_{i,2} \mid x_{i,1}, x_{i,2}, z_i^A, z_i^M, a_i, d_{i,1}, d_{i,2}; \theta^{(\ell)}\right)\right) + 2p_{\text{WAIC}}$$

$$\begin{aligned}p_{\text{WAIC}} = 2\sum_{i=1}^{n}\Bigg[& \ln\left(\frac{1}{H}\sum_{\ell=1}^{H} f\left(t_{i,1}, t_{i,2} \mid x_{i,1}, x_{i,2}, z_i^A, z_i^M, a_i, d_{i,1}, d_{i,2}; \theta^{(\ell)}\right)\right) \\ & - \frac{1}{H}\sum_{\ell=1}^{H} \ln f\left(t_{i,1}, t_{i,2} \mid x_{i,1}, x_{i,2}, z_i^A, z_i^M, a_i, d_{i,1}, d_{i,2}; \theta^{(\ell)}\right)\Bigg]\end{aligned}$$

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

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

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

https://doi.org/10.1017/asb.2024.34 Published online by Cambridge University Press

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

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

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

https://doi.org/10.1017/asb.2024.34 Published online by Cambridge University Press