An augmented variable Dirichlet process mixture model for the analysis of dependent lifetimes

Francesco Ungolo

School of Risk & Actuarial Studies, UNSW Business School, Sydney

Longevity 20 Conference

Introduction

- The pricing of insurance products issued on multiple lives, e.g. reversionary annuities, requires the use of statistical models which can best predict their future lifetimes;

Introduction

- The pricing of insurance products issued on multiple lives, e.g. reversionary annuities, requires the use of statistical models which can best predict their future lifetimes;
- Assuming independent lifetimes reduces model complexity, at the expense of a material impact on actuarial valuations if this assumption does not hold:

Introduction

- The pricing of insurance products issued on multiple lives, e.g. reversionary annuities, requires the use of statistical models which can best predict their future lifetimes;
- Assuming independent lifetimes reduces model complexity, at the expense of a material impact on actuarial valuations if this assumption does not hold:
  - For example, partners are likely to share the same socio-economic characteristics, such that they share the same living standards, and are exposed to similar risks;

Introduction

- The pricing of insurance products issued on multiple lives, e.g. reversionary annuities, requires the use of statistical models which can best predict their future lifetimes;
- Assuming independent lifetimes reduces model complexity, at the expense of a material impact on actuarial valuations if this assumption does not hold:
  - For example, partners are likely to share the same socio-economic characteristics, such that they share the same living standards, and are exposed to similar risks;
  - Denuit and Cornet (1999) and Frees et. al. (1996) showed how the last survivor annuity value under dependent lifetime models is lower compared to the case of independent lifetimes;

Introduction

- The pricing of insurance products issued on multiple lives, e.g. reversionary annuities, requires the use of statistical models which can best predict their future lifetimes;
- Assuming independent lifetimes reduces model complexity, at the expense of a material impact on actuarial valuations if this assumption does not hold:
  - For example, partners are likely to share the same socio-economic characteristics, such that they share the same living standards, and are exposed to similar risks;
  - Denuit and Cornet (1999) and Frees et. al. (1996) showed how the last survivor annuity value under dependent lifetime models is lower compared to the case of independent lifetimes;
- Typical approaches to model dependence include the use of copula, or models using random effects, especially in biostatistics.

Objectives

Aim: propose a joint model for the lifetime distribution such that it is possible to:

- Account for individual and group-specific covariates;

Objectives

Aim: propose a joint model for the lifetime distribution such that it is possible to:

- Account for individual and group-specific covariates;
- Being able to handle right censoring and left truncation;

Objectives

Aim: propose a joint model for the lifetime distribution such that it is possible to:

- Account for individual and group-specific covariates;
- Being able to handle right censoring and left truncation;
- Gain an understanding of the impact of the covariates on the lifetime dependence.

Earlier literature

- Focus on parametric models:
  - Copula models (Frees et. al. (1996), Carriere (2000), Deresa et. al. (2022), Youn and Shemaykin (1999) and Dufresne et. al. (2018));

Earlier literature

- Focus on parametric models:
  - Copula models (Frees et. al. (1996), Carriere (2000), Deresa et. al. (2022), Youn and Shemaykin (1999) and Dufresne et. al. (2018));
  - Frailty models (Vaupel et. al. (1979), Yashin and Yachine (1985), Huang and Wolfe (2002));

Earlier literature

- Focus on parametric models:
  - Copula models (Frees et. al. (1996), Carriere (2000), Deresa et. al. (2022), Youn and Shemaykin (1999) and Dufresne et. al. (2018));
  - Frailty models (Vaupel et. al. (1979), Yashin and Yachine (1985), Huang and Wolfe (2002));
- There is an increasing need for of flexible and adaptable statistical models capable to handle complex problems, which can be interpretable at the same time.

The Augmented Variable model for husband and wife (1)

- T_1 and T_2 denote the random (future) lifetime of husband (1) and wife (2);

The Augmented Variable model for husband and wife (1)

- T_1 and T_2 denote the random (future) lifetime of husband (1) and wife (2);
- x_1 and x_2 are the vector of individual specific covariates (e.g. age, medical status);

The Augmented Variable model for husband and wife (1)

- T_1 and T_2 denote the random (future) lifetime of husband (1) and wife (2);
- x_1 and x_2 are the vector of individual specific covariates (e.g. age, medical status);
- z denotes the couple-specific covariates (e.g. household income, geo-demographic profile and so on);

For the *i*th couple, the AV model "augments" the joint probability distribution of (T_1, T_2) through a joint probability model for (T_1, T_2, Z):

The Augmented Variable model for husband and wife (1)

- T_1 and T_2 denote the random (future) lifetime of husband (1) and wife (2);
- x_1 and x_2 are the vector of individual specific covariates (e.g. age, medical status);
- z denotes the couple-specific covariates (e.g. household income, geo-demographic profile and so on);

For the ith couple, the AV model "augments" the joint probability distribution of (T_1, T_2) through a joint probability model for (T_1, T_2, Z):

$$f\left(t_{i,1}, t_{i,2}, z_i \mid x_{i,1}, x_{i,2}\right) = \int_{\Omega_{\gamma,\zeta}} \left[\prod_{j=1}^{J} f\left(t_{i,j} \mid x_{i,j}; \gamma_{i,j}\right)\right] f\left(z_i; \zeta_i\right) \mathrm{d}G\left(\gamma_{i,1}, \gamma_{i,2}, \zeta_i\right)$$

The Augmented Variable model for husband and wife (1)

- T_1 and T_2 denote the random (future) lifetime of husband (1) and wife (2);
- x_1 and x_2 are the vector of individual specific covariates (e.g. age, medical status);
- z denotes the couple-specific covariates (e.g. household income, geo-demographic profile and so on);

For the ith couple, the AV model "augments" the joint probability distribution of (T_1, T_2) through a joint probability model for (T_1, T_2, Z):

$$f\left(t_{i,1}, t_{i,2}, z_i \mid x_{i,1}, x_{i,2}\right) = \int_{\Omega_{\gamma,\zeta}} \left[\prod_{j=1}^{J} f\left(t_{i,j} \mid x_{i,j}; \gamma_{i,j}\right)\right] f\left(z_i; \zeta_i\right) \mathrm{d}G\left(\gamma_{i,1}, \gamma_{i,2}, \zeta_i\right)$$

The latent multivariate random vector $(\gamma_{i,1}, \gamma_{i,2}, \zeta_i)$ induces dependence among $T_{i,1}$, $T_{i,2}$ and Z_i.

The Augmented Variable model for husband and wife (2)

$$f\left(t_{i,1}, t_{i,2}, z_i \mid x_{i,1}, x_{i,2}\right) = \int_{\Omega_{\gamma,\zeta}} \left[\prod_{j=1}^{J} f\left(t_{i,j} \mid x_{i,j}; \gamma_{i,j}\right)\right] f\left(z_i \mid \zeta_i\right) \mathrm{d}G\left(\gamma_{i,1}, \gamma_{i,2}, \zeta_i\right)$$

The Augmented Variable model for husband and wife (2)

$$f\left(t_{i,1}, t_{i,2}, z_i \mid x_{i,1}, x_{i,2}\right) = \int_{\Omega_{\gamma,\zeta}} \left[\prod_{j=1}^{J} f\left(t_{i,j} \mid x_{i,j}; \gamma_{i,j}\right)\right] f\left(z_i \mid \zeta_i\right) \mathrm{d}G\left(\gamma_{i,1}, \gamma_{i,2}, \zeta_i\right)$$

- The joint density 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 G of $(\gamma_{i,1}, \gamma_{i,2}, \zeta_i)$;

The Augmented Variable model for husband and wife (2)

$$f\left(t_{i,1}, t_{i,2}, z_i \mid x_{i,1}, x_{i,2}\right) = \int_{\Omega_{\gamma,\zeta}} \left[\prod_{j=1}^{J} f\left(t_{i,j} \mid x_{i,j}; \gamma_{i,j}\right)\right] f\left(z_i \mid \zeta_i\right) \mathrm{d}G\left(\gamma_{i,1}, \gamma_{i,2}, \zeta_i\right)$$

- The joint density 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 G of $(\gamma_{i,1}, \gamma_{i,2}, \zeta_i)$;
- We distinguish between the variables which directly affect the individual lifetimes $X_{i,j}$ from those which are couple-specific Z_i;

The Augmented Variable model for husband and wife (2)

$$f\left(t_{i,1}, t_{i,2}, z_i \mid x_{i,1}, x_{i,2}\right) = \int_{\Omega_{\gamma,\zeta}} \left[\prod_{j=1}^{J} f\left(t_{i,j} \mid x_{i,j}; \gamma_{i,j}\right)\right] f\left(z_i \mid \zeta_i\right) \mathrm{d}G\left(\gamma_{i,1}, \gamma_{i,2}, \zeta_i\right)$$

- The joint density 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 G of $(\gamma_{i,1}, \gamma_{i,2}, \zeta_i)$;
- We distinguish between the variables which directly affect the individual lifetimes $X_{i,j}$ from those which are couple-specific Z_i;
- We can also understand how Z_i affects the dependence between $T_{i,1}$ and $T_{i,2}$.

Dirichlet Process (1): Features

Let $\theta = (\gamma_1, \gamma_2, \zeta)$:

Dirichlet Process (1): Features

Let $\theta = (\gamma_1, \gamma_2, \zeta)$:

- A flexible specification for $G(\theta)$ is to assume it is a random draw from a Dirichlet Process (DP)

Dirichlet Process (1): Features

Let $\theta = (\gamma_1, \gamma_2, \zeta)$:

- A flexible specification for $G(\theta)$ is to assume it is a random draw from a Dirichlet Process (DP)

$$G(\theta) \sim \mathrm{DP}(\phi, G_0) \tag{1}$$

where

Dirichlet Process (1): Features

Let $\theta = (\gamma_1, \gamma_2, \zeta)$:

- A flexible specification for $G(\theta)$ is to assume it is a random draw from a Dirichlet Process (DP)

$$G(\theta) \sim \mathrm{DP}(\phi, G_0) \tag{1}$$

  where

  - G_0 is the *base probability measure*, e.g.

$$G_0 := \underbrace{g(\gamma_1, \gamma_2 \mid \mathbf{0}, \Sigma_\gamma)}_{\mathrm{MVN}(\mathbf{0},\ \Sigma_\gamma)} \times \underbrace{g(\zeta \mid m, s)}_{\mathrm{Gamma}(m,\ s)}$$

  - ϕ is the concentration parameter

Dirichlet Process (2): A discrete probability measure

- DP yields a discrete probability measure G, which induces a clustering structure among the individual random effects θ_i.
  - Let θ_k^* ($k = 1, \ldots, K$, with $K \leq n$) denote the unique values of θ_i. These are i.i.d. samples from G_0;

Dirichlet Process (2): A discrete probability measure

- DP yields a discrete probability measure G, which induces a clustering structure among the individual random effects θ_i.
  - Let θ_k^* ($k = 1, \ldots, K$, with $K \leq n$) denote the unique values of θ_i. These are i.i.d. samples from G_0;
  - The probability attached to each outcome of θ is

$$\pi_k = \psi_k \prod_{j=1}^{k-1} (1 - \psi_j)\,; \quad \psi_k \sim \text{Beta}\,(1, \phi)$$

Dirichlet Process (2): A discrete probability measure

- DP yields a discrete probability measure G, which induces a clustering structure among the individual random effects θ_i.
  - Let θ_k^* ($k = 1, \dots, K$, with $K \leq n$) denote the unique values of θ_i. These are i.i.d. samples from G_0;
  - The probability attached to each outcome of θ is

$$\pi_k = \psi_k \prod_{j=1}^{k-1} (1 - \psi_j)\,; \quad \psi_k \sim \text{Beta}\,(1, \phi)$$

This is also known as *stick-breaking process* (SBP, Sethuraman (1994))

The Augmented Variable Dirichlet Process Mixture model

- The joint lifetime density for the ith couple can be written as a Dirichlet Process Mixture:

$$
\begin{aligned}
f\left(t_{i,1}, t_{i,2}, z_i \mid x_{i,1}, x_{i,2}\right) &= \int_{\Omega_{\gamma,\zeta}} \left[\prod_{j=1}^{2} f\left(t_{i,j} \mid x_{i,j}; \gamma_{i,j}\right)\right] f\left(z_i \mid \zeta_i\right) \mathrm{d}G\left(\gamma_{i,1}, \gamma_{i,2}, \zeta_i\right) \\
&= \sum_{k=1}^{\infty} \pi_k \left[\prod_{j=1}^{2} f\left(t_{i,j} \mid x_{i,j}; \gamma^*_{j,k}\right)\right] f\left(z_i \mid \zeta^*_k\right)
\end{aligned}
$$

The Augmented Variable Dirichlet Process Mixture model

- The joint lifetime density for the ith couple can be written as a Dirichlet Process Mixture:

$$
\begin{aligned}
f\left(t_{i,1}, t_{i,2}, z_i \mid x_{i,1}, x_{i,2}\right) &= \int_{\Omega_{\gamma,\zeta}} \left[\prod_{j=1}^{2} f\left(t_{i,j} \mid x_{i,j}; \gamma_{i,j}\right)\right] f\left(z_i \mid \zeta_i\right) \mathrm{d}G\left(\gamma_{i,1}, \gamma_{i,2}, \zeta_i\right) \\
&= \sum_{k=1}^{\infty} \pi_k \left[\prod_{j=1}^{2} f\left(t_{i,j} \mid x_{i,j}; \gamma^*_{j,k}\right)\right] f\left(z_i \mid \zeta^*_k\right)
\end{aligned}
$$

In this way, we aim at carrying out statistical inferences which are robust with respect to misspecification of the joint probability distribution of (T_1, T_2, Z).

Data - The Canadian life insurance data set

- 12,139 joint and last-survivor annuity contracts;

Data - The Canadian life insurance data set

- 12,139 joint and last-survivor annuity contracts;
- Observation period: 29th December 1988 - 31th December 1993;

Data - The Canadian life insurance data set

- 12,139 joint and last-survivor annuity contracts;
- Observation period: 29th December 1988 - 31th December 1993;
- For each couple we observe:

Data - The Canadian life insurance data set

- 12,139 joint and last-survivor annuity contracts;
- Observation period: 29th December 1988 - 31th December 1993;
- For each couple we observe:
    - *Starting date* of the insurance contract;

Data - The Canadian life insurance data set

- 12,139 joint and last-survivor annuity contracts;
- Observation period: 29th December 1988 - 31th December 1993;
- For each couple we observe:
  - *Starting date* of the insurance contract;
  - *Date of birth* of each couple member (if within the observational plan);

Data - The Canadian life insurance data set

- 12,139 joint and last-survivor annuity contracts;
- Observation period: 29th December 1988 - 31th December 1993;
- For each couple we observe:
  - *Starting date* of the insurance contract;
  - *Date of birth* of each couple member (if within the observational plan);
  - *Date of death* if within the observational plan (otherwise the individual lifetime is censored);

Data - The Canadian life insurance data set

- 12,139 joint and last-survivor annuity contracts;
- Observation period: 29th December 1988 - 31th December 1993;
- For each couple we observe:
  - *Starting date* of the insurance contract;
  - *Date of birth* of each couple member (if within the observational plan);
  - *Date of death* if within the observational plan (otherwise the individual lifetime is censored);
- Most observations are left-truncated;

Data - The Canadian life insurance data set

- 12,139 joint and last-survivor annuity contracts;
- Observation period: 29th December 1988 - 31th December 1993;
- For each couple we observe:
  - *Starting date* of the insurance contract;
  - *Date of birth* of each couple member (if within the observational plan);
  - *Date of death* if within the observational plan (otherwise the individual lifetime is censored);
- Most observations are left-truncated;
- 1,164 deaths were observed among the males and 357 among females

Data - The Canadian life insurance data set

- 12,139 joint and last-survivor annuity contracts;
- Observation period: 29th December 1988 - 31th December 1993;
- For each couple we observe:
  - *Starting date* of the insurance contract;
  - *Date of birth* of each couple member (if within the observational plan);
  - *Date of death* if within the observational plan (otherwise the individual lifetime is censored);
- Most observations are left-truncated;
- 1,164 deaths were observed among the males and 357 among females

Hence, the age is the only covariate that we can observe.

Data - The Canadian life insurance data set

- Earlier work (Youn and Shemaykin (1999), Dufresne et. al. (2018) and Deresa et. al. (2022) have shown how age difference (AD, in absolute value) and whether the male is oldest (MO) capture additional features of the association between the lifetime of the husband and of the wife;

Data - The Canadian life insurance data set

- Earlier work (Youn and Shemaykin (1999), Dufresne et. al. (2018) and Deresa et. al. (2022) have shown how age difference (AD, in absolute value) and whether the male is oldest (MO) capture additional features of the association between the lifetime of the husband and of the wife;
- The MO covariate has an influence in the relationship between husband and wife, consequently affecting their lifetime dependence;

Data - The Canadian life insurance data set

- Earlier work (Youn and Shemaykin (1999), Dufresne et. al. (2018) and Deresa et. al. (2022) have shown how age difference (AD, in absolute value) and whether the male is oldest (MO) capture additional features of the association between the lifetime of the husband and of the wife;
- The MO covariate has an influence in the relationship between husband and wife, consequently affecting their lifetime dependence;

In our model we consider:

$$X = \text{age}$$

$$Z = (\text{AD}, \text{MO})$$

Model (1)

- Following Frees et. al. (1996), we assume that for the ith couple, males ($j = 1$) and females ($j = 2$) lifetime distribution is characterized by the following hazard function

Model (1)

- Following Frees et. al. (1996), we assume that for the ith couple, males ($j = 1$) and females ($j = 2$) lifetime distribution is characterized by the following hazard function

$$\mu_{i,j}(t) \quad = \quad \exp\left(\alpha_j + \beta_j\left(x_{i,j} + t\right) + \gamma_{i,j}\right), \quad \text{hence with p.d.f.}$$

Model (1)

- Following Frees et. al. (1996), we assume that for the ith couple, males ($j = 1$) and females ($j = 2$) lifetime distribution is characterized by the following hazard function

$$\begin{aligned} \mu_{i,j}(t) &= \exp\left(\alpha_j + \beta_j (x_{i,j} + t) + \gamma_{i,j}\right), \quad \text{hence with p.d.f.} \\ f_{i,j}(t) &= \exp\left(-\int_0^t \mu_{i,j}(s)\,\mathrm{d}s\right)\mu_{i,j}(t) \end{aligned}$$

Model (1)

- Following Frees et. al. (1996), we assume that for the ith couple, males ($j = 1$) and females ($j = 2$) lifetime distribution is characterized by the following hazard function

$$\mu_{i,j}(t) = \exp\left(\alpha_j + \beta_j (x_{i,j} + t) + \gamma_{i,j}\right), \quad \text{hence with p.d.f.}$$

$$f_{i,j}(t) = \exp\left(-\int_0^t \mu_{i,j}(s)\,\mathrm{d}s\right)\mu_{i,j}(t)$$

- For the couple-specific covariates AD and MO we assume:

$$\ln(\mathrm{AD}_i) \sim \mathrm{N}\left(\zeta_{\mathrm{AD},i}, \sigma^2_{\mathrm{AD}}\right)$$
$$\mathrm{MO}_i \sim \mathrm{Bernoulli}\left(\zeta_{\mathrm{MO},i}\right)$$

Model (2)

- For the dependence between T_1, T_2 and Z:

$$(\gamma_{i,1}, \gamma_{i,2}, \zeta_{\mathrm{AD},i}, \zeta_{\mathrm{MO},i}) \sim G$$

$$G \sim \mathrm{DP}\left(\phi, \quad G_0\right)$$

where the *base probability* distribution given by

$$G_0 = \underbrace{g\left(\gamma_{i,1}, \gamma_{i,2} \mid \mathbf{0}, \Sigma_\gamma\right)}_{\mathrm{MVN}(\mathbf{0},\Sigma_\gamma)} \underbrace{g\left(\zeta_i^A \mid m_A, s_A^2\right)}_{\mathrm{N}\left(m_A, \;\; s_A^2\right)} \underbrace{g\left(\zeta_i^M \mid 3, 1\right)}_{\mathrm{Beta}(3,1)}$$

Model (2)

- For the dependence between T_1, T_2 and Z:

$$(\gamma_{i,1}, \gamma_{i,2}, \zeta_{\mathrm{AD},i}, \zeta_{\mathrm{MO},i}) \sim G$$

$$G \sim \mathrm{DP}\left(\phi, \quad G_0\right)$$

where the *base probability* distribution given by

$$G_0 = \underbrace{g\left(\gamma_{i,1}, \gamma_{i,2} \mid \mathbf{0}, \Sigma_\gamma\right)}_{\mathrm{MVN}(\mathbf{0},\Sigma_\gamma)} \underbrace{g\left(\zeta_i^A \mid m_A, s_A^2\right)}_{\mathrm{N}\left(m_A, \ s_A^2\right)} \underbrace{g\left(\zeta_i^M \mid 3, 1\right)}_{\mathrm{Beta}(3,1)}$$

- To complete the parametric specification, we assume conditionally conjugate weakly informative prior for the other parameters whereas possible;

Model (2)

- For the dependence between T_1, T_2 and Z:

$$(\gamma_{i,1}, \gamma_{i,2}, \zeta_{\mathrm{AD},i}, \zeta_{\mathrm{MO},i}) \sim G$$

$$G \sim \mathrm{DP}\left(\phi, \quad G_0\right)$$

where the *base probability* distribution given by

$$G_0 = \underbrace{g\left(\gamma_{i,1}, \gamma_{i,2} \mid \mathbf{0}, \Sigma_\gamma\right)}_{\mathrm{MVN}(\mathbf{0},\Sigma_\gamma)} \underbrace{g\left(\zeta_i^A \mid m_A, s_A^2\right)}_{\mathrm{N}\left(m_A,\ s_A^2\right)} \underbrace{g\left(\zeta_i^M \mid 3, 1\right)}_{\mathrm{Beta}(3,1)}$$

- To complete the parametric specification, we assume conditionally conjugate weakly informative prior for the other parameters whereas possible;
- The posterior distribution of the parameters is estimated using a simple Data Augmentation Blocked MCMC sampler;

Fitted log-hazard functions

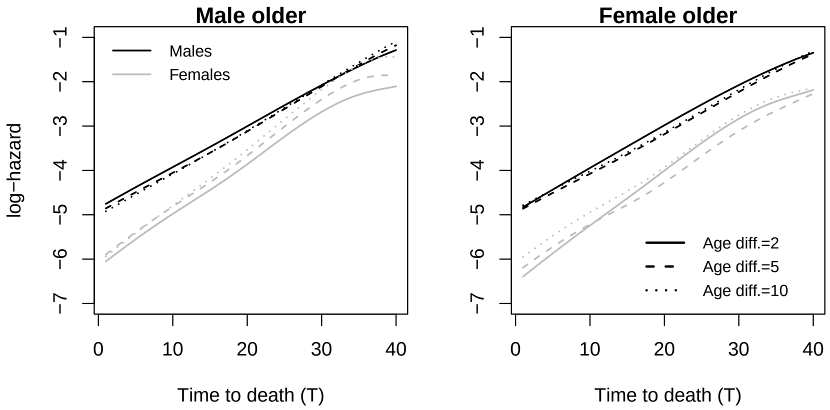

Figure: Hazard function for males (black) and females (gray) aged 60-100 for different values of the covariates male older and age difference (not in log scale).

Model comparison

- We compare the results of our model with:

Base Gompertz: $\mu_{i,j}^{BG}(t) = \exp\left(\alpha_j + \beta_j\left(x_{i,j} + t\right)\right)$

Prop. Hazard: $\mu_{i,j}^{PH}(t) = \exp\left(\alpha_j + \beta_j\left(x_{i,j} + t\right) + \delta_{j,1}\ln\left(\mathrm{ad}_i\right) + \delta_{j,2}\mathrm{mo}_i\right)$

- The data set is split into two parts: 75% couples in the training and the remaining 25% in the testing sample.

Table: 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	4,537.42	4,541.64	**4,515.14**

Model comparison

- We compare the results of our model with:

$$\text{Base Gompertz:} \quad \mu_{i,j}^{BG}(t) = \exp\left(\alpha_j + \beta_j\left(x_{i,j} + t\right)\right)$$

$$\text{Prop. Hazard:} \quad \mu_{i,j}^{PH}(t) = \exp\left(\alpha_j + \beta_j\left(x_{i,j} + t\right) + \delta_{j,1}\ln\left(\mathrm{ad}_i\right) + \delta_{j,2}\mathrm{mo}_i\right)$$

Model comparison

- We compare the results of our model with:

$$\text{Base Gompertz:} \quad \mu_{i,j}^{BG}(t) = \exp\left(\alpha_j + \beta_j\left(x_{i,j} + t\right)\right)$$

$$\text{Prop. Hazard:} \quad \mu_{i,j}^{PH}(t) = \exp\left(\alpha_j + \beta_j\left(x_{i,j} + t\right) + \delta_{j,1}\ln\left(\mathrm{ad}_i\right) + \delta_{j,2}\mathrm{mo}_i\right)$$

- The data set is split into two parts: 75% couples in the training and the remaining 25% in the testing sample.

Model comparison

- We compare the results of our model with:

$$\text{Base Gompertz:} \quad \mu^{BG}_{i,j}(t) = \exp\left(\alpha_j + \beta_j\left(x_{i,j} + t\right)\right)$$

$$\text{Prop. Hazard:} \quad \mu^{PH}_{i,j}(t) = \exp\left(\alpha_j + \beta_j\left(x_{i,j} + t\right) + \delta_{j,1}\ln\left(\mathrm{ad}_i\right) + \delta_{j,2}\mathrm{mo}_i\right)$$

- The data set is split into two parts: 75% couples in the training and the remaining 25% in the testing sample.

Table: 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	4,537.42	4,541.64	**4,515.14**

Extensions of the model

- The AVDPM can also be extended to the analysis of groups with differing number of non-exchangeable individuals (e.g. collective insurance);

Extensions of the model

- The AVDPM can also be extended to the analysis of groups with differing number of non-exchangeable individuals (e.g. collective insurance);
- Suppose the ith group includes J_i members, with common set of variables z_i, and individual (within ith group) characteristics $\mathbf{x}_{i,j}$.

Extensions of the model

- The AVDPM can also be extended to the analysis of groups with differing number of non-exchangeable individuals (e.g. collective insurance);
- Suppose the ith group includes J_i members, with common set of variables z_i, and individual (within ith group) characteristics $\mathbf{x}_{i,j}$. The group-specific joint distribution of the lifetimes and $\mathbf{Z}$ can be written as:

Extensions of the model

- The AVDPM can also be extended to the analysis of groups with differing number of non-exchangeable individuals (e.g. collective insurance);
- Suppose the ith group includes J_i members, with common set of variables z_i, and individual (within ith group) characteristics $\mathbf{x}_{i,j}$. The group-specific joint distribution of the lifetimes and $\mathbf{Z}$ can be written as:

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

Extensions of the model

- The AVDPM can also be extended to the analysis of groups with differing number of non-exchangeable individuals (e.g. collective insurance);
- Suppose the ith group includes J_i members, with common set of variables z_i, and individual (within ith group) characteristics $\mathbf{x}_{i,j}$. The group-specific joint distribution of the lifetimes and $\mathbf{Z}$ can be written as:

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

where Q denotes a suitable distribution function for $\gamma_{i,j}$ indexed by the DP-distributed η_i;

Conclusions

- The AVDPM allows for a simple and flexible specification of the joint distribution of dependent lifetimes, which

Conclusions

- The AVDPM allows for a simple and flexible specification of the joint distribution of dependent lifetimes, which
  - Incorporates individual-specific information;

Conclusions

- The AVDPM allows for a simple and flexible specification of the joint distribution of dependent lifetimes, which
  - Incorporates individual-specific information;
  - Accounts for the group-specific covariates;

Conclusions

- The AVDPM allows for a simple and flexible specification of the joint distribution of dependent lifetimes, which
  - Incorporates individual-specific information;
  - Accounts for the group-specific covariates;
  - Allows for an understanding of the impact of the group-specific covariates over the dependence between individual lifetimes

Conclusions

- The AVDPM allows for a simple and flexible specification of the joint distribution of dependent lifetimes, which
    - Incorporates individual-specific information;
    - Accounts for the group-specific covariates;
    - Allows for an understanding of the impact of the group-specific covariates over the dependence between individual lifetimes
- The model turns out to improve the predictive performance for the held-out dataset and to carry out robust statistical inferences:

Conclusions

- The AVDPM allows for a simple and flexible specification of the joint distribution of dependent lifetimes, which
  - Incorporates individual-specific information;
  - Accounts for the group-specific covariates;
  - Allows for an understanding of the impact of the group-specific covariates over the dependence between individual lifetimes
- The model turns out to improve the predictive performance for the held-out dataset and to carry out robust statistical inferences:
  - For the case of non-exchangeable lifetimes, different parametric laws can be specified (e.g. Gompertz for males and Makeham for females);

Conclusions

- The AVDPM allows for a simple and flexible specification of the joint distribution of dependent lifetimes, which
    - Incorporates individual-specific information;
    - Accounts for the group-specific covariates;
    - Allows for an understanding of the impact of the group-specific covariates over the dependence between individual lifetimes
- The model turns out to improve the predictive performance for the held-out dataset and to carry out robust statistical inferences:
    - For the case of non-exchangeable lifetimes, different parametric laws can be specified (e.g. Gompertz for males and Makeham for females);
- The model can be easily estimated by means of a fully Bayesian analysis which may account for the prior information of the researcher;

Conclusions

- The AVDPM allows for a simple and flexible specification of the joint distribution of dependent lifetimes, which
  - Incorporates individual-specific information;
  - Accounts for the group-specific covariates;
  - Allows for an understanding of the impact of the group-specific covariates over the dependence between individual lifetimes
- The model turns out to improve the predictive performance for the held-out dataset and to carry out robust statistical inferences:
  - For the case of non-exchangeable lifetimes, different parametric laws can be specified (e.g. Gompertz for males and Makeham for females);
- The model can be easily estimated by means of a fully Bayesian analysis which may account for the prior information of the researcher;
- An analysis of grouped units can be obtained as by-product to obtain further insight on the dependence relationships.

References I

- F. Ungolo and P.J. Laub (2025), *An Augmented Variable Dirichlet Process Mixture model for the analysis of dependent lifetimes*, ASTIN Bulletin, vol. 55 (1), pag. 50-75.