Actuarial Bayesian Nonparametric Regression modelling for survival data

Dr. Francesco Ungolo

School of Risk & Actuarial Studies, UNSW Business School, UNSW Sydney
ARC Centre of Excellence in Population Ageing Research (CEPAR)

33rd Colloquium on Pensions and Retirement Research
24 November 2025

Introduction

- The analysis of mortality rates is of interest due to their financial impact on portfolio of annuities;

Introduction

- The analysis of mortality rates is of interest due to their financial impact on portfolio of annuities;
- Availability of granular information about pension scheme members provides an opportunity for a more precise estimate of the individual mortality profile (beyond age and sex). For example:

Introduction

- The analysis of mortality rates is of interest due to their financial impact on portfolio of annuities;
- Availability of granular information about pension scheme members provides an opportunity for a more precise estimate of the individual mortality profile (beyond age and sex). For example:
  - Residential area and benefit amount;

Introduction

- The analysis of mortality rates is of interest due to their financial impact on portfolio of annuities;
- Availability of granular information about pension scheme members provides an opportunity for a more precise estimate of the individual mortality profile (beyond age and sex). For example:
  - Residential area and benefit amount;
  - Retirement grounds (ill-health or reaching the retirement age);

Introduction

- The analysis of mortality rates is of interest due to their financial impact on portfolio of annuities;
- Availability of granular information about pension scheme members provides an opportunity for a more precise estimate of the individual mortality profile (beyond age and sex). For example:
  - Residential area and benefit amount;
  - Retirement grounds (ill-health or reaching the retirement age);
  - Relationship status (single, married, children)

Objectives

Aim: Propose a **flexible parametric hazard model** — the Actuarial Dependent Dirichlet Process Mixture (ADDPM) model.

Objectives

Aim: Propose a **flexible parametric hazard model** — the Actuarial Dependent Dirichlet Process Mixture (ADDPM) model.

- Accounts for **linear and non-linear covariate effects**:

Objectives

Aim: Propose a **flexible parametric hazard model** — the Actuarial Dependent Dirichlet Process Mixture (ADDPM) model.

- Accounts for **linear and non-linear covariate effects**:
  - Survival functions may **cross** or **converge** depending on data.

Objectives

Aim: Propose a **flexible parametric hazard model** — the Actuarial Dependent Dirichlet Process Mixture (ADDPM) model.

- Accounts for **linear and non-linear covariate effects**:
  - Survival functions may **cross** or **converge** depending on data.
  - Captures both observed covariates and unobserved heterogeneity.

Objectives

Aim: Propose a **flexible parametric hazard model** — the Actuarial Dependent Dirichlet Process Mixture (ADDPM) model.

- Accounts for **linear and non-linear covariate effects**:
  - Survival functions may **cross** or **converge** depending on data.
  - Captures both observed covariates and unobserved heterogeneity.
- Effectively handles **right censoring** and **left truncation** typical of actuarial datasets.

Objectives

Aim: Propose a **flexible parametric hazard model** — the Actuarial Dependent Dirichlet Process Mixture (ADDPM) model.

- Accounts for **linear and non-linear covariate effects**:
  - Survival functions may **cross** or **converge** depending on data.
  - Captures both observed covariates and unobserved heterogeneity.
- Effectively handles **right censoring** and **left truncation** typical of actuarial datasets.
- Gain insights into how covariates shape **lifetime dependence** and mortality determinants.

The Actuarial Dependent Dirichlet Process Mixture (ADDPM) model (1)

- T_i denotes the random future lifetime;

The Actuarial Dependent Dirichlet Process Mixture (ADDPM) model (1)

- T_i denotes the random future lifetime;
- x_i and z_i are the two vectors of observed individual covariates (e.g. age, medical status);

The Actuarial Dependent Dirichlet Process Mixture (ADDPM) model (1)

- T_i denotes the random future lifetime;
- x_i and z_i are the two vectors of observed individual covariates (e.g. age, medical status);

The distribution of T_i is defined through the hazard function:

$$\mu\left(t \mid h\left(x_i, z_i; \beta, \delta_i\right)\right),$$

The Actuarial Dependent Dirichlet Process Mixture (ADDPM) model (1)

- T_i denotes the random future lifetime;
- x_i and z_i are the two vectors of observed individual covariates (e.g. age, medical status);

The distribution of T_i is defined through the hazard function:

$$\mu\left(t \mid h\left(x_i, z_i; \beta, \delta_i\right)\right), \qquad \text{where}$$

$$h\left(x_i, z_i; \beta, \delta_i\right) = \beta x_i + \delta_i z_i$$

The Actuarial Dependent Dirichlet Process Mixture (ADDPM) model (1)

- T_i denotes the random future lifetime;
- x_i and z_i are the two vectors of observed individual covariates (e.g. age, medical status);

The distribution of T_i is defined through the hazard function:

$$\mu\left(t \mid h\left(x_i, z_i; \beta, \delta_i\right)\right), \qquad \text{where}$$

$$h\left(x_i, z_i; \beta, \delta_i\right) = \beta x_i + \delta_i z_i$$

- The individual-specific parameter δ_i allows to capture the heterogeneity induced by the interaction of z_i with unobserved variables;

The Actuarial Dependent Dirichlet Process Mixture (ADDPM) model (1)

- T_i denotes the random future lifetime;
- x_i and z_i are the two vectors of observed individual covariates (e.g. age, medical status);

The distribution of T_i is defined through the hazard function:

$$\mu\left(t \mid h\left(x_i, z_i; \beta, \delta_i\right)\right), \qquad \text{where}$$

$$h\left(x_i, z_i; \beta, \delta_i\right) = \beta x_i + \delta_i z_i$$

- The individual-specific parameter δ_i allows to capture the heterogeneity induced by the interaction of z_i with unobserved variables;
- Its combination with simpler parametric forms for μ, e.g. Gompertz, allows to capture additional stylized facts of human mortality, such as:

The Actuarial Dependent Dirichlet Process Mixture (ADDPM) model (1)

- T_i denotes the random future lifetime;
- x_i and z_i are the two vectors of observed individual covariates (e.g. age, medical status);

The distribution of T_i is defined through the hazard function:

$$\mu\left(t \mid h\left(x_i, z_i; \beta, \delta_i\right)\right), \qquad \text{where}$$

$$h\left(x_i, z_i; \beta, \delta_i\right) = \beta x_i + \delta_i z_i$$

- The individual-specific parameter δ_i allows to capture the heterogeneity induced by the interaction of z_i with unobserved variables;
- Its combination with simpler parametric forms for μ, e.g. Gompertz, allows to capture additional stylized facts of human mortality, such as:
  - The late-life mortality deceleration (Gavrilov and Gavrilova (2001));

The Actuarial Dependent Dirichlet Process Mixture (ADDPM) model (1)

- T_i denotes the random future lifetime;
- x_i and z_i are the two vectors of observed individual covariates (e.g. age, medical status);

The distribution of T_i is defined through the hazard function:

$$\mu\left(t \mid h\left(x_i, z_i; \beta, \delta_i\right)\right), \qquad \text{where}$$

$$h\left(x_i, z_i; \beta, \delta_i\right) = \beta x_i + \delta_i z_i$$

- The individual-specific parameter δ_i allows to capture the heterogeneity induced by the interaction of z_i with unobserved variables;
- Its combination with simpler parametric forms for μ, e.g. Gompertz, allows to capture additional stylized facts of human mortality, such as:
  - The late-life mortality deceleration (Gavrilov and Gavrilova (2001));
  - Reduced effect of the covariates at older ages.

The Actuarial Dependent Dirichlet Process Mixture (ADDPM) model (2)

The random parameter δ_i is assumed to follow a discrete distribution G following a Dirichlet Process:

$$\delta_i \sim G$$
$$G \sim \text{DP}(\phi, G_0)$$

where

The Actuarial Dependent Dirichlet Process Mixture (ADDPM) model (2)

The random parameter δ_i is assumed to follow a discrete distribution G following a Dirichlet Process:

$$\delta_i \sim G$$
$$G \sim \text{DP}\left(\phi, G_0\right)$$

where

- G_0 is the *base probability measure*, e.g. $G_0 := \underbrace{g\left(\delta \mid 0, \sigma_\delta^2\right)}_{\text{Normal}\left(0,\ \sigma_\delta^2\right)}$
- ϕ is the concentration parameter

The Actuarial Dependent Dirichlet Process Mixture (ADDPM) model (2)

The random parameter δ_i is assumed to follow a discrete distribution G following a Dirichlet Process:

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

where

- G_0 is the *base probability measure*, e.g. $G_0 := \underbrace{g\left(\delta \mid 0, \sigma_\delta^2\right)}_{\mathrm{Normal}\left(0,\ \sigma_\delta^2\right)}$
- ϕ is the concentration parameter

This allows to:

- Model features like multimodality, skewness and heavy tails;

The Actuarial Dependent Dirichlet Process Mixture (ADDPM) model (2)

The random parameter δ_i is assumed to follow a discrete distribution G following a Dirichlet Process:

$$\delta_i \sim G$$
$$G \sim \text{DP}(\phi, G_0)$$

where

- G_0 is the *base probability measure*, e.g. $G_0 := \underbrace{g\left(\delta \mid 0, \sigma_\delta^2\right)}_{\text{Normal}\left(0,\ \sigma_\delta^2\right)}$
- ϕ is the concentration parameter

This allows to:

- Model features like multimodality, skewness and heavy tails;
- Model-base clustering of the units sharing the same δ_i;

The Actuarial Dependent Dirichlet Process Mixture (ADDPM) model (2)

The random parameter δ_i is assumed to follow a discrete distribution G following a Dirichlet Process:

$$\delta_i \sim G$$
$$G \sim \text{DP}(\phi, G_0)$$

where

- G_0 is the *base probability measure*, e.g. $G_0 := \underbrace{g\left(\delta \mid 0, \sigma_\delta^2\right)}_{\text{Normal}\left(0,\ \sigma_\delta^2\right)}$
- ϕ is the concentration parameter

This allows to:

- Model features like multimodality, skewness and heavy tails;
- Model-base clustering of the units sharing the same δ_i;
- Characterize a posteriori groups of units sharing the same mortality profile.

Illustration – UK mid-sized pension scheme dataset

- 18,734 male members;

Illustration – UK mid-sized pension scheme dataset

- 18,734 male members;
- Observation period: 10th November 1992 - 31st December 2009;

Illustration – UK mid-sized pension scheme dataset

- 18,734 male members;
- Observation period: 10th November 1992 - 31st December 2009;
- 172,525.9 exposed-to-risk years and 4,950 observed deaths

Illustration – UK mid-sized pension scheme dataset

- 18,734 male members;
- Observation period: 10th November 1992 - 31st December 2009;
- 172,525.9 exposed-to-risk years and 4,950 observed deaths
- For each individual we observe:

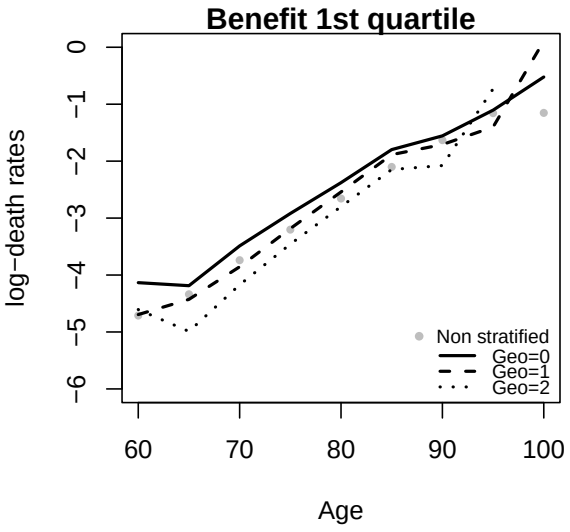
Illustration – UK mid-sized pension scheme dataset

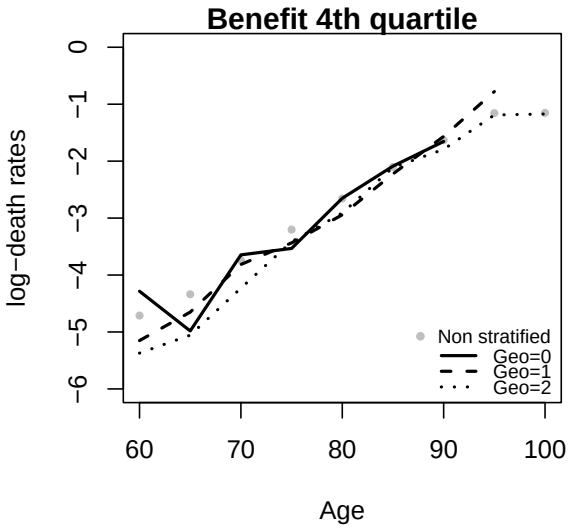
- 18,734 male members;
- Observation period: 10th November 1992 - 31st December 2009;
- 172,525.9 exposed-to-risk years and 4,950 observed deaths
- For each individual we observe:
  - The annual benefit amount;

Illustration – UK mid-sized pension scheme dataset

- 18,734 male members;
- Observation period: 10th November 1992 - 31st December 2009;
- 172,525.9 exposed-to-risk years and 4,950 observed deaths
- For each individual we observe:
  - The annual benefit amount;
  - Geo-demographic feature on the basis of MOSAIC profiler, widely used in the UK for pricing longevity swaps and bulk buy-outs (Richards 2008).

Preliminary data analysis





Model

Based on the preliminary data analysis, we specify:

$$h(x_i, z_i; \beta, \delta_i) = \beta_0 + \beta_1 (\text{age}_i - \overline{\text{age}}) + \beta_2 \text{geo}_{1,i} + \beta_3 \text{geo}_{2,i} + \delta_i \text{ben}_i$$

Model

Based on the preliminary data analysis, we specify:

$$h(x_i, z_i; \beta, \delta_i) = \beta_0 + \beta_1 (\text{age}_i - \overline{\text{age}}) + \beta_2 \text{geo}_{1,i} + \beta_3 \text{geo}_{2,i} + \delta_i \text{ben}_i$$

$$\mu(t \mid h(x_i, z_i; \beta, \delta_i)) = \exp(h(x_i, z_i; \beta, \delta_i))$$

where:

Model

Based on the preliminary data analysis, we specify:

$$h(x_i, z_i; \beta, \delta_i) = \beta_0 + \beta_1 (\text{age}_i - \overline{\text{age}}) + \beta_2 \text{geo}_{1,i} + \beta_3 \text{geo}_{2,i} + \delta_i \text{ben}_i$$

$$\mu(t \mid h(x_i, z_i; \beta, \delta_i)) = \exp(h(x_i, z_i; \beta, \delta_i))$$

where:

- $\overline{\text{age}}$ is the average age as observed in the sample;

Model

Based on the preliminary data analysis, we specify:

$$h(x_i, z_i; \beta, \delta_i) = \beta_0 + \beta_1 (\text{age}_i - \overline{\text{age}}) + \beta_2 \text{geo}_{1,i} + \beta_3 \text{geo}_{2,i} + \delta_i \text{ben}_i$$

$$\mu(t \mid h(x_i, z_i; \beta, \delta_i)) = \exp(h(x_i, z_i; \beta, \delta_i))$$

where:

- $\overline{\text{age}}$ is the average age as observed in the sample;
- $\text{geo}_{k,i}$ (for $k = 1, 2$) is an indicator variable based on the level of the geo-demographic profile;

Model

Based on the preliminary data analysis, we specify:

$$
\begin{aligned}
h\left(x_i, z_i; \beta, \delta_i\right) &= \beta_0 + \beta_1 \left(\text{age}_i - \overline{\text{age}}\right) + \beta_2 \text{geo}_{1,i} + \beta_3 \text{geo}_{2,i} + \delta_i \text{ben}_i \\
\mu\left(t \mid h\left(x_i, z_i; \beta, \delta_i\right)\right) &= \exp\left(h\left(x_i, z_i; \beta, \delta_i\right)\right)
\end{aligned}
$$

where:

- $\overline{\text{age}}$ is the average age as observed in the sample;
- $\text{geo}_{k,i}$ (for $k = 1, 2$) is an indicator variable based on the level of the geo-demographic profile;
- ben_i is the standardized value of the logarithm of the benefit;

Model

Based on the preliminary data analysis, we specify:

$$
\begin{aligned}
h\left(x_i, z_i; \beta, \delta_i\right) &= \beta_0 + \beta_1 \left(\text{age}_i - \overline{\text{age}}\right) + \beta_2 \text{geo}_{1,i} + \beta_3 \text{geo}_{2,i} + \delta_i \text{ben}_i \\
\mu\left(t \mid h\left(x_i, z_i; \beta, \delta_i\right)\right) &= \exp\left(h\left(x_i, z_i; \beta, \delta_i\right)\right)
\end{aligned}
$$

where:

- $\overline{\text{age}}$ is the average age as observed in the sample;
- $\text{geo}_{k,i}$ (for $k = 1, 2$) is an indicator variable based on the level of the geo-demographic profile;
- ben_i is the standardized value of the logarithm of the benefit;
- $\delta_i \sim G$ and $G \sim \text{DP}\left(\phi, N\left(0, \sigma_\delta^2\right)\right)$

Model

Based on the preliminary data analysis, we specify:

$$h(x_i, z_i; \beta, \delta_i) = \beta_0 + \beta_1 (\text{age}_i - \overline{\text{age}}) + \beta_2 \text{geo}_{1,i} + \beta_3 \text{geo}_{2,i} + \delta_i \text{ben}_i$$

$$\mu(t \mid h(x_i, z_i; \beta, \delta_i)) = \exp(h(x_i, z_i; \beta, \delta_i))$$

where:

- $\overline{\text{age}}$ is the average age as observed in the sample;
- $\text{geo}_{k,i}$ (for $k = 1, 2$) is an indicator variable based on the level of the geo-demographic profile;
- ben_i is the standardized value of the logarithm of the benefit;
- $\delta_i \sim G$ and $G \sim \text{DP}\left(\phi, N\left(0, \sigma_\delta^2\right)\right)$

The model is analysed through a fully Bayesian analysis, and a tailored MCMC has been designed at the purpose.

Results - Fitted model

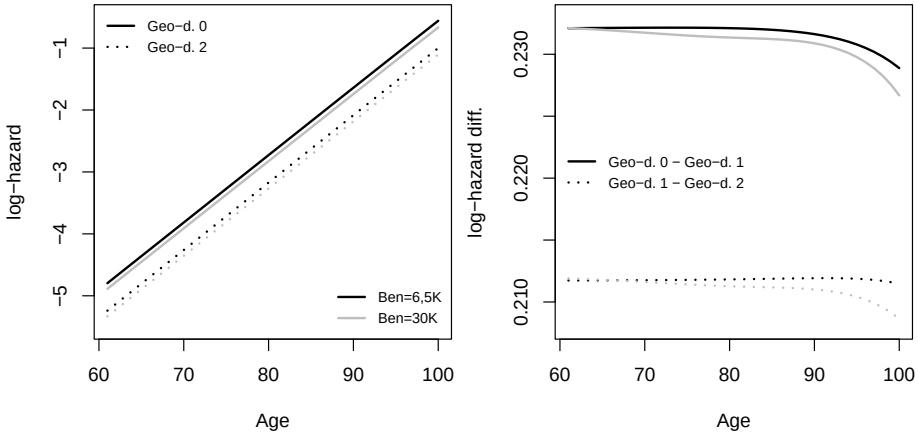


Figure 1: Log-hazard function by geo-demographic level and benefit amount (left) and difference between log-hazard functions by level of geo-demographic profile (0 vs 1) and (1 vs 2) for given benefit amounts (right).

Results - In-sample and out-of-sample Model comparison (1)

The dataset is randomly split into a training set (75% of obs.) and a test set. This random split is repeated 100 times. Model comparison is based on information criteria.

Results - In-sample and out-of-sample Model comparison (1)

The dataset is randomly split into a training set (75% of obs.) and a test set. This random split is repeated 100 times. Model comparison is based on information criteria.

The competing models are:

- Model G0: age-only Gompertz hazard function:

$$\text{G0:} \quad \mu\left(s \mid \text{age}_i; \beta\right) = \exp\left(\beta_0 + \beta_1 \text{age}_i\right)$$

Results - In-sample and out-of-sample Model comparison (1)

The dataset is randomly split into a training set (75% of obs.) and a test set. This random split is repeated 100 times. Model comparison is based on information criteria.

The competing models are:

- Model G0: age-only Gompertz hazard function:

$$\text{G0:} \quad \mu\left(s \mid \text{age}_i; \beta\right) = \exp\left(\beta_0 + \beta_1 \text{age}_i\right)$$

- Model G1: Gompertz hazard function which assumes a log-linear effect of all available covariates:

$$\text{G1:} \quad \mu\left(s \mid \text{age}_i, \text{geo}_{1,i}, \text{geo}_{2,i}, \text{ben}_i; \beta\right) = \exp\left(\beta_0 + \beta_1 \text{age}_i + \beta_2 \text{geo}_{1,i} + \beta_3 \text{geo}_{2,i} + \beta_4 \text{ben}_i\right)$$

Results - In-sample and out-of-sample Model comparison (1)

The dataset is randomly split into a training set (75% of obs.) and a test set. This random split is repeated 100 times. Model comparison is based on information criteria.

The competing models are:

- Model G0: age-only Gompertz hazard function:

$$\text{G0:} \quad \mu\left(s \mid \text{age}_i; \beta\right) = \exp\left(\beta_0 + \beta_1 \text{age}_i\right)$$

- Model G1: Gompertz hazard function which assumes a log-linear effect of all available covariates:

$$\text{G1:} \quad \mu\left(s \mid \text{age}_i, \text{geo}_{1,i}, \text{geo}_{2,i}, \text{ben}_i; \beta\right) = \exp\left(\beta_0 + \beta_1 \text{age}_i + \beta_2 \text{geo}_{1,i} + \beta_3 \text{geo}_{2,i} + \beta_4 \text{ben}_i\right)$$

- Model G2: same as G1, where the benefit covariate is instead modeled as a two-level categorical variable, equal to 1 if the benefit is larger than £8,500 and 0 otherwise, as in Ungolo et. al. (2019).

Results - In-sample and out-of-sample Model comparison (2)

Table 1: AIC and WAIC of the models (the lowest value is shown in bold).

(W)AIC	G0	G1	G2	ADDPM
# Parameters	2	5	5	IS: 4.64; OS: 1.53
In sample	30,875.16	30,771.07	**30,761.73**	30,772.22
# Outperform	0	0	100	0
Out of sample	10,275.02	10,245.07	10,241.73	**10,238.81**
# Outperform	0	0	21	79

- In-sample, model G2 consistently outperforms all models;

Results - In-sample and out-of-sample Model comparison (2)

Table 1: AIC and WAIC of the models (the lowest value is shown in bold).

(W)AIC	G0	G1	G2	ADDPM
# Parameters	2	5	5	IS: 4.64; OS: 1.53
In sample	30,875.16	30,771.07	**30,761.73**	30,772.22
# Outperform	0	0	100	0
Out of sample	10,275.02	10,245.07	10,241.73	**10,238.81**
# Outperform	0	0	21	79

- In-sample, model G2 consistently outperforms all models;
- Excluding G2, the ADDPM consistently outperforms all remaining models;

Results - In-sample and out-of-sample Model comparison (2)

Table 1: AIC and WAIC of the models (the lowest value is shown in bold).

(W)AIC	G0	G1	G2	ADDPM
# Parameters	2	5	5	IS: 4.64; OS: 1.53
In sample	30,875.16	30,771.07	**30,761.73**	30,772.22
# Outperform	0	0	100	0
Out of sample	10,275.02	10,245.07	10,241.73	**10,238.81**
# Outperform	0	0	21	79

- In-sample, model G2 consistently outperforms all models;
- Excluding G2, the ADDPM consistently outperforms all remaining models;
- Out-of-sample, the ADDPM demonstrates superior predictive performance compared to all competing models in 79 out of 100 random splits of the dataset (100 if we exclude G2);

Results - In-sample and out-of-sample Model comparison (2)

Table 1: AIC and WAIC of the models (the lowest value is shown in bold).

(W)AIC	G0	G1	G2	ADDPM
# Parameters	2	5	5	IS: 4.64; OS: 1.53
In sample	30,875.16	30,771.07	**30,761.73**	30,772.22
# Outperform	0	0	100	0
Out of sample	10,275.02	10,245.07	10,241.73	**10,238.81**
# Outperform	0	0	21	79

- In-sample, model G2 consistently outperforms all models;
- Excluding G2, the ADDPM consistently outperforms all remaining models;
- Out-of-sample, the ADDPM demonstrates superior predictive performance compared to all competing models in 79 out of 100 random splits of the dataset (100 if we exclude G2);
- The ADDPM captures the decreasing effect of the benefit level on mortality rates at older ages without sacrificing model parsimony.

Conclusion

- ADDPM extends survival models to allow for a more flexible, interpretable and actuarially meaningful effect of the covariates;

Conclusion

- ADDPM extends survival models to allow for a more flexible, interpretable and actuarially meaningful effect of the covariates;
- The underlying Bayesian workflow allows for the inclusion of prior information by the analyst;

Conclusion

- ADDPM extends survival models to allow for a more flexible, interpretable and actuarially meaningful effect of the covariates;
- The underlying Bayesian workflow allows for the inclusion of prior information by the analyst;
- It captures additional stylized facts of human mortality without sacrificing model parsimony;

Conclusion

- ADDPM extends survival models to allow for a more flexible, interpretable and actuarially meaningful effect of the covariates;
- The underlying Bayesian workflow allows for the inclusion of prior information by the analyst;
- It captures additional stylized facts of human mortality without sacrificing model parsimony;
- The in-sample performance is comparable to that of simpler models, while substantial gains are observed when the ADDPM model is deployed for improving the predictive accuracy;

Conclusion

- ADDPM extends survival models to allow for a more flexible, interpretable and actuarially meaningful effect of the covariates;
- The underlying Bayesian workflow allows for the inclusion of prior information by the analyst;
- It captures additional stylized facts of human mortality without sacrificing model parsimony;
- The in-sample performance is comparable to that of simpler models, while substantial gains are observed when the ADDPM model is deployed for improving the predictive accuracy;
- Future work:
  - Explore mixture structures beyond the Dirichlet Process.

Conclusion

- ADDPM extends survival models to allow for a more flexible, interpretable and actuarially meaningful effect of the covariates;
- The underlying Bayesian workflow allows for the inclusion of prior information by the analyst;
- It captures additional stylized facts of human mortality without sacrificing model parsimony;
- The in-sample performance is comparable to that of simpler models, while substantial gains are observed when the ADDPM model is deployed for improving the predictive accuracy;
- Future work:
  - Explore mixture structures beyond the Dirichlet Process.
  - Extend ADDPM to include **variable selection** for improved predictive performance.

Conclusion

- ADDPM extends survival models to allow for a more flexible, interpretable and actuarially meaningful effect of the covariates;
- The underlying Bayesian workflow allows for the inclusion of prior information by the analyst;
- It captures additional stylized facts of human mortality without sacrificing model parsimony;
- The in-sample performance is comparable to that of simpler models, while substantial gains are observed when the ADDPM model is deployed for improving the predictive accuracy;
- Future work:
  - Explore mixture structures beyond the Dirichlet Process.
  - Extend ADDPM to include **variable selection** for improved predictive performance.
  - Develop faster approximate inference methods (e.g., **Laplace-based techniques**).

References I

- F. Ungolo, T. Kleinow and A.S. Macdonald (2025), *Actuarial Bayesian Nonparametric Regression modelling for survival data*, submitted