A Bayesian pooling model for mortality graduation with Age-Varying Smoothness

Viviana G. R. Lobo

viviana@dme.ufrj.br
Universidade Federal do Rio de Janeiro
Departamento de Métodos Estatísticos and Laboratório de Matemática Aplicada

Joint with Thais C. O. Fonseca (DME, LabMA/UFRJ),
Mariane B. Alves (DME, LabMA/UFRJ) and Luiz Figueiredo (LabMA, IM/UFRJ)

Longevity 19

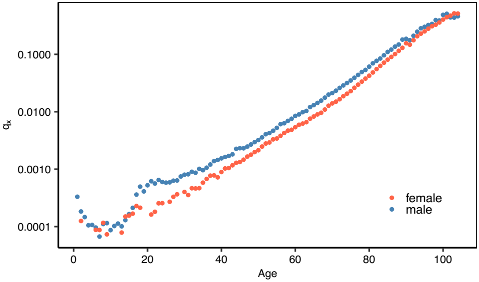
Joint work with:

MARIANE BRANCO
(DME/IM, LABMA)

THAÍS FONSECA
(DME/IM, LABMA)

LUIZ FIGUEIREDO
(IM, LABMA)

Motivation



Modelling multiple populations independently ignores associations between mortality patterns across different groups.

Key benefits of joint models include:

- borrowing strength;
- quantification of joint uncertainty;
- imposing constraints;
- handling missing data.

Literature

- Tang, Dodd and Forster (2022) propose a model to smooth mortality rates based on a P-spline with adaptive penalty to allow for varying smoothness over the age domain.
- Additionally, the joint modelling technique facilitates information sharing across sexes, thereby preventing divergence and intersection between the mortality curves.
- Alexander et al. (2017) consider a Bayesian hierarchical approach to account for correlation in the joint modelling of subnational mortality in the context of missing data.

Multivariate mortality data

- For each population j and mid-year age x, we define the central mortality rate as

$$m_x^{(j)} = \frac{D_x^{(j)}}{E_x^{(j)}}, \ j = 1, \dots, J.$$

- Define the log death rate for the j-th population as

$$Y_x^{(j)} = \log\left(\frac{D_x^{(j)}}{E_x^{(j)}}\right),$$

  which follows a Normal distribution:

$$Y_x^{(j)} \sim \mathcal{N}\left(\mu_x^{(j)}, V^{(j)}\right).$$

- The probability of death $q_x^{(j)}$ at age x for the j-th population:

$$q_x^{(j)} = 1 - \exp(-m_x^{(j)}) = 1 - \exp\left[-\exp\left(y_x^{(j)}\right)\right].$$

The multivariate dynamical linear mortality model

- Dynamic Linear Models (DLM) are traditionally used in time series analysis to handle auto-correlated observations over time (West and Harrison, 1997).

- We adopt a specific DLM specification to create graduated mortality tables.

- This approach recongnises the association between mortality rates of neighbouring ages and ensures smoothness in the estimated mortality curve across ages.

The multivariate dynamical linear mortality model

- We consider a second-order polynomial trend model to generate graduated mortality tables for the $j-th$ population, $j = 1, \ldots, J$, as follows:

$$Y_x^{(j)} = \mu_x^{(j)} + v_x^{(j)}, \quad v_x^{(j)} \sim \mathcal{N}(0, V^{(j)})$$

OBSERVATION EQUATION

$$\begin{aligned} \mu_x^{(j)} &= \mu_{x-1}^{(j)} + \beta_{x-1}^{(j)} + w_{x,1}^{(j)}, \quad w_{x,1}^{(j)} \sim \mathcal{N}(0, W_{x,1}^{(j)}) \\ \beta_x^{(j)} &= \beta_{x-1}^{(j)} + w_{x,2}^{(j)}, \quad w_{x,2}^{(j)} \sim \mathcal{N}(0, W_{x,2}^{(j)}), \end{aligned}$$

EVOLUTION EQUATION

DYNAMICAL LEVEL OF THE LOG-MORTALITY

LOCAL SLOPE OF THE LOG-MORTALITY

The model considers how mortality at age x is influenced by mortality at age $x-1$.

Components of the proposed joint DLM

The general proposed joint DLM is given by:

$$\mathbf{Y}_x = \mathbf{F}\boldsymbol{\theta}_x + \boldsymbol{\nu}_x, \quad \boldsymbol{\nu}_x \sim \mathcal{N}_2(\mathbf{0}, \mathbf{V}_x)$$

$$\boldsymbol{\theta}_x = \mathbf{G}\boldsymbol{\theta}_{x-1} + \boldsymbol{\omega}_x, \quad \boldsymbol{\omega}_x \sim \mathcal{N}_p(\mathbf{0}, \mathbf{W}_x).$$

For $J = 2$ populations, $\mathbf{Y}'_x = \left(Y_x^{(1)}, Y_x^{(2)}\right)$:

$$\mathbf{F} = blockdiag(F_x^{(1)}, F_x^{(2)}) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}, \quad \mathbf{G} = blockdiag(G_x^{(1)}, G_x^{(2)}) = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad \boldsymbol{\theta}_x = \begin{pmatrix} \mu_x^{(1)} \\ \beta_x^{(1)} \\ \mu_x^{(2)} \\ \beta_x^{(2)} \end{pmatrix}, \quad \boldsymbol{V}_x = \begin{pmatrix} V_x^{(1)} & \sigma_{12} \\ \sigma_{21} & V_x^{(2)} \end{pmatrix}.$$

Age-Varying Smoothness

- Discount factors control how quickly past mortality rates lose relevance for estimates at age x.
- Different discount factors may be applied to distinct age intervals.
- Smoother or more volatile fits can be made depending on the expected behaviour of the theoretical mortality law for a population across different age groups.

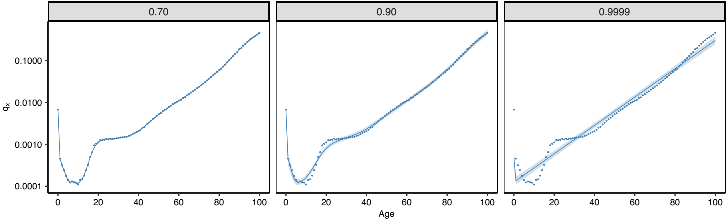


Figure: USA, 2010 male population, from Human Mortality Database via `BayesMortalityPlus` package.

Mortality in England and Wales (2010-2012, ONS)

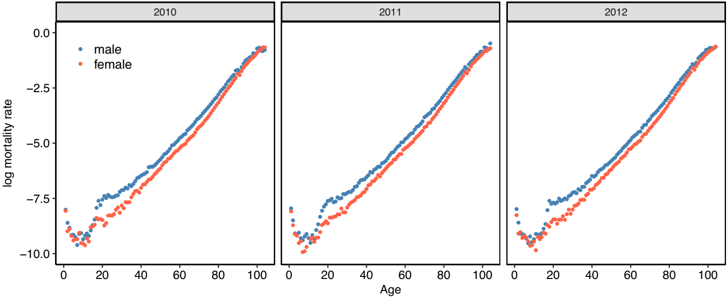


Figure: Raw mortality rates in log-scale for England and Wales from 2010 to 2012 and 1-104 years old. Blue and red dots denote the male ($j = 1$) and female ($j = 2$) populations, respectively.

Three studies are considered to evaluate the performance of the multivariate dynamical linear model.

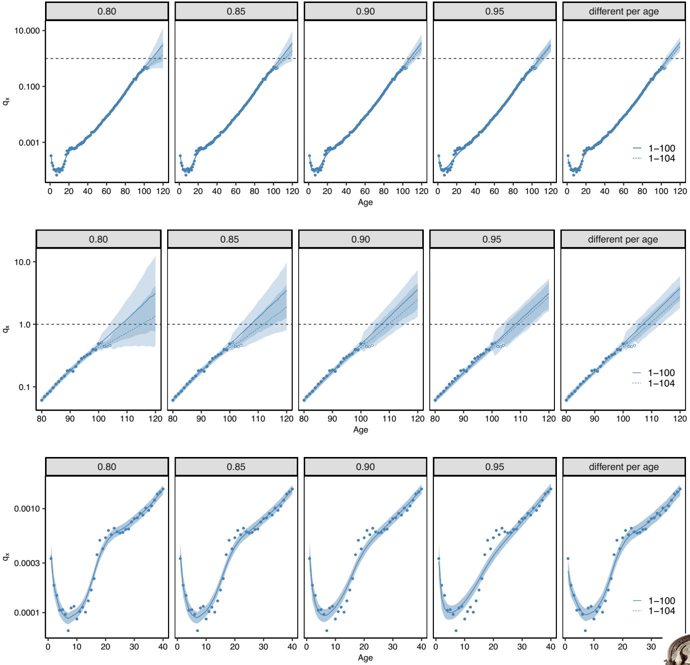
Study 1: Mortality extrapolation at advanced ages with age-varying smoothness

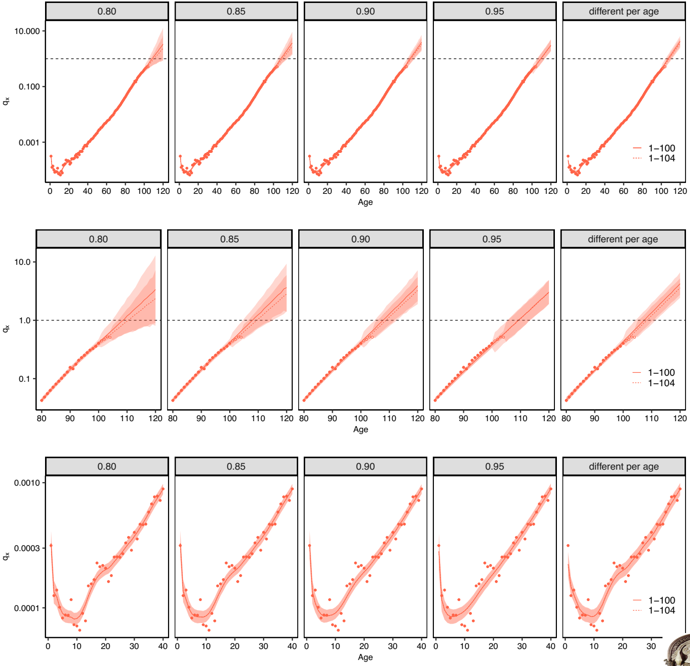
- Two scenarios are considered as presented by Tang, Dodd and Forster (2022):
  (a) Full dataset (ages 1 to 104)
  (b) Dataset with ages 1 to 100 (excluding the last four age groups)

- Univariate models for male and female mortality rates with extrapolation up to age 120 years via `BayesMortalityPlus` package available in `R`.

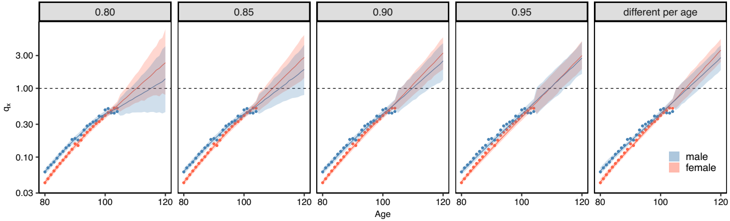
- Our proposed model controls the smoothness of the fit within each age interval:
  (a) fixed discount factors ($\delta_x = 0.80, 0.85, 0.90, 0.95$) applied uniformly across all ages.
  (b) flexible discount factor across different age ranges, denoted by δ_x^* with:

$$\delta_{1,1:5} = 0.99, \delta_{2,6:35} = 0.80, \delta_{3,36:85} = 0.85, \delta_{4,86+} = 0.99.$$


0.80
0.85
0.90
0.95
different per age
10.000
0.100
0.001
q_x
Age
1–100
1–104
10.0
1.0
0.1
0.0010
0.0003
0.0001



0.80
0.85
0.90
0.95
different per age
10.000
0.100
0.001
q_x
0 20 40 60 80 100 120
Age
1–100
1–104
10.0
1.0
0.1
80 90 100 110 120
0.0010
0.0003
0.0001
0 10 20 30 40



- ☹ Issues with univariate models: cross-over extrapolations and divergence.
- ☹ Ensuring convergence: apply constraints and use transition functions to blend and smooth male and female curves from a selected starting age.

An alternative is to consider the joint modelling of subpopulations.

Study 2: Joint modelling preventing cross-over and convergence

- We consider the same dataset analysed in Study 1.
- Here, our goal is to illustrate the benefits of jointly modelling male and female mortality rates.
- We consider the age range from 1 to 104 for the modelling process.
- As presented in Study 1, the utilization of age-varying factors yields satisfactory predictions with robust uncertainty quantification, without constraining the age range.

Study 2: Joint modelling preventing cross-over and convergence

The following joint model is adopted to reflect the fact that male ($j = 1$) mortality is higher than or equal to female ($j = 2$) mortality:

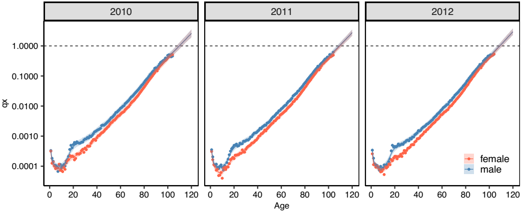
≥ 0

$$Y_x^{(1)} = \mu_x^{(1)} + \boxed{\mu_x^{(2)}} + v_x^{(1)}$$
$$Y_x^{(2)} = \mu_x^{(2)} + v_x^{(2)}$$

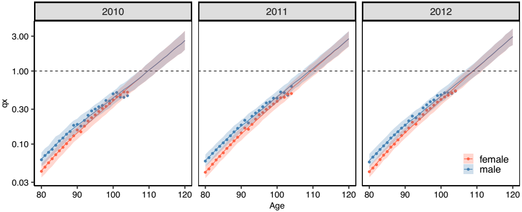
THEN $Y_x^{(1)} \geq Y_x^{(2)}$

$$\mu_x^{(1)} = \mu_{x-1}^{(1)} + \beta_{x-1}^{(1)} + w_{x,1}^{(1)}$$
$$\mu_x^{(2)} = \mu_{x-1}^{(2)} + \beta_{x-1}^{(2)} + w_{x,1}^{(2)}$$
$$\beta_x^{(1)} = \beta_{x-1}^{(1)} + w_{x,2}^{(1)}$$
$$\beta_x^{(2)} = \beta_{x-1}^{(2)} + w_{x,2}^{(2)},$$

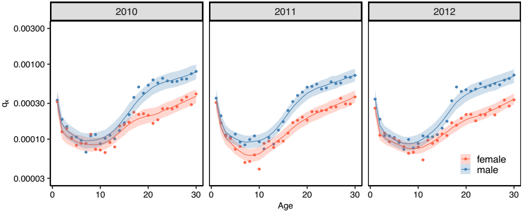
$$\mathbf{F}_x = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}, \quad \mathbf{G}_x = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad \boldsymbol{\theta}_x = \begin{pmatrix} \mu_x^{(1)} \\ \beta_x^{(1)} \\ \mu_x^{(2)} \\ \beta_x^{(2)} \end{pmatrix}, \quad \boldsymbol{V}_x = \begin{pmatrix} V_x^{(1)} & \sigma_{12} \\ \sigma_{21} & V_x^{(2)} \end{pmatrix}.$$



2010
2011
2012
1.0000
0.1000
0.0100
0.0010
0.0001
qx
0
20
40
60
80
100
120
Age
female
male

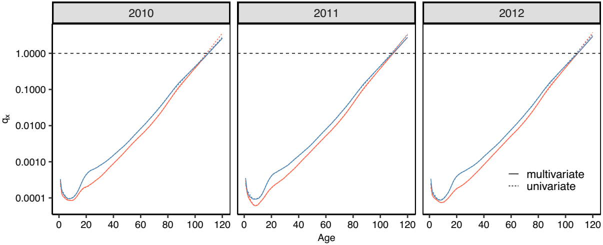


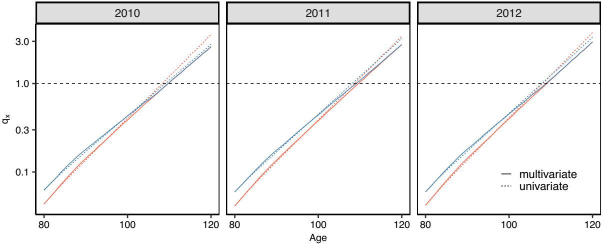
2010
2011
2012
3.00
1.00
0.30
0.10
0.03
qx
80
90
100
110
120
Age
female
male



2010
2011
2012
0.00300
0.00100
0.00030
0.00010
0.00003
q_x
0
10
20
30
Age
female
male

Study 2: Univariate *versus* Multivariate





Study 3: Joint modelling for missing data

- Focusing on the year 2010 (ages 1-104), we construct various scenarios by selectively removing mortality data, particularly from younger and older age groups.

- These scenarios are designed to test the robustness of the proposed model when faced with incomplete data across different age ranges.

scenario	**age groups**	**% of missing data**
a	4-8 (Female)	$\approx 5\%$
b	4-10, 15-17 (Female)	$\approx 10\%$
c	3-16 (Female)	$\approx 15\%$
d	1-25 (Female)	$\approx 25\%$
e	1-16, 23-41 (Female)	$\approx 33\%$
f	1-45 (Female)	$\approx 43\%$
g	1-4 (Male), 5-17 (Female)	$\approx 16\%$
h	80+ (Male)	$\approx 24\%$

Study 3: Joint modelling for missing data

- We introduce a model with a common term as follows:

$$
\begin{aligned}
Y_x^{(1)} &= \alpha_x + \mu_x^{(1)} + v_x^{(1)} \\
Y_x^{(2)} &= \alpha_x + \mu_x^{(2)} + v_x^{(2)}
\end{aligned}
$$

$$
\alpha_x = \alpha_{x-1} + \mu_{x-1}^{(1)} + \mu_{x-1}^{(2)} + w_x^*
$$

$$
\begin{aligned}
\mu_x^{(1)} &= \mu_{x-1}^{(1)} + \beta_{x-1}^{(1)} + w_{x,1}^{(1)} \\
\mu_x^{(2)} &= \mu_{x-1}^{(2)} + \beta_{x-1}^{(2)} + w_{x,1}^{(2)} \\
\beta_x^{(1)} &= \beta_{x-1}^{(1)} + w_{x,2}^{(1)} \\
\beta_x^{(2)} &= \beta_{x-1}^{(2)} + w_{x,2}^{(2)}
\end{aligned}
$$

$$
\mathbf{F}'_x = \begin{pmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1 \\ 0 & 0 \\ 1 & 1 \end{pmatrix}, \quad
\mathbf{G}_x = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 \end{pmatrix}, \quad
\boldsymbol{\theta}_x = \begin{pmatrix} \mu_x^{(1)} \\ \beta_x^{(1)} \\ \mu_x^{(2)} \\ \beta_x^{(2)} \\ \alpha_x \end{pmatrix}, \quad
\boldsymbol{V}_x = \begin{pmatrix} V_x^{(1)} & \sigma_{12}^2 \\ \sigma_{21}^2 & V_x^{(2)} \end{pmatrix}.
$$

- The distribution for missing data is given by:

$$
\mathbf{Y}_x^{(miss)} \mid \mathbf{Y}_x^{(obs)}, \boldsymbol{\theta}_x, V \sim \mathcal{N}_q\left(\mathbf{F}'_x \boldsymbol{\theta}_x^{(miss)} + V_{m,m} V_{0,0}^{-1}\left(\mathbf{y}_x^{(obs)} - \mathbf{F}'_x \boldsymbol{\theta}_x^{(obs)}\right), V_{m,m} - V_{m,o} V_{0,0}^{-1} V_{0,m}\right),
$$

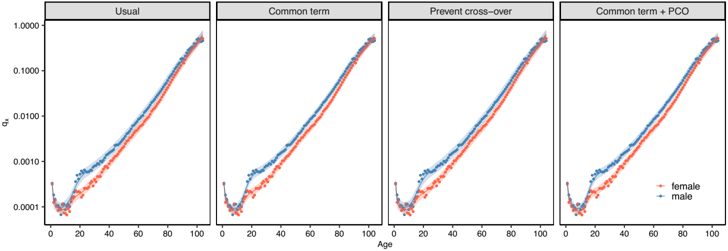
$$
\boldsymbol{\theta}_x = \begin{pmatrix} \boldsymbol{\theta}_x^{(obs)} \\ \boldsymbol{\theta}_x^{(miss)} \end{pmatrix}, \quad V = \begin{pmatrix} V_{o,o} & V_{o,m} \\ V_{m,o} & V_{m,m} \end{pmatrix}
$$

Study 3: Joint modelling for missing data

- Four competing models are considered here:

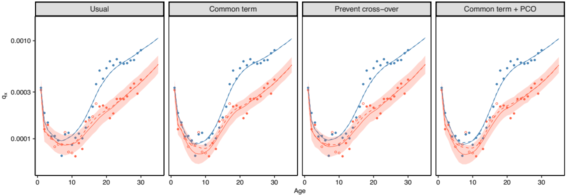
  1. the usual joint model;
  2. the joint model with a common term;
  3. the joint model that prevents cross-over;
  4. combined model w/ a common term and cross-over prevention.

- No missing data results:



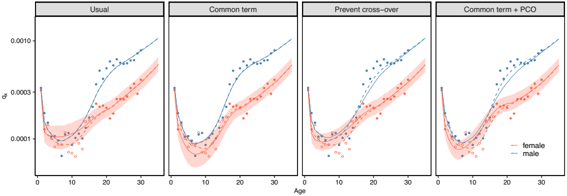
4-10, 15-17

10%



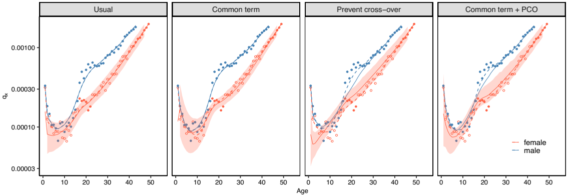
3-16

15%

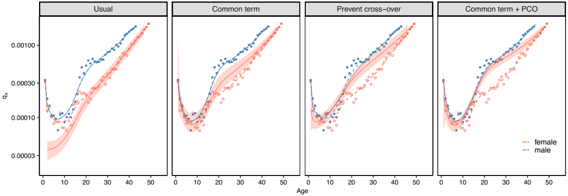


1-16, 23-41

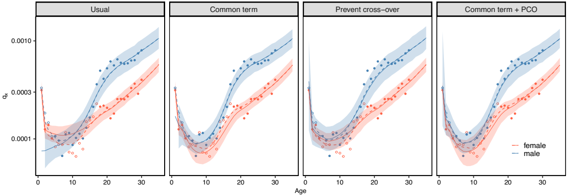
33%



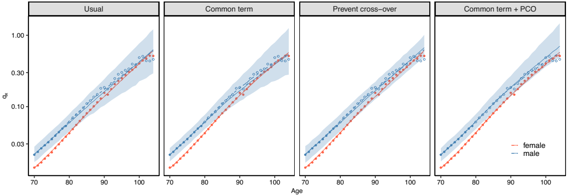
1-45
43%



5-17, 1-4
16%



80+
24%



Study 3: Model comparison

competing models	scenario	% missing	measures		
			mspe	mape	wci
Bivariate usual	b	$\approx 10\%$	0.00097	0.00850	0.02414
Bivariate w/ CT			0.00063	0.00777	0.02021
Bivariate w/ PCO			0.00088	0.00803	0.02345
Bivariate w/ CT+ PCO			**0.00052**	**0.00704**	**0.02020**
Bivariate usual	c	$\approx 15\%$	0.00099	0.00855	0.02017
Bivariate w/ CT			0.00064	0.00779	0.01901
Bivariate w/ PCO			0.00089	0.00813	0.01909
Bivariate w/ CT+ PCO			**0.00053**	**0.00711**	**0.01833**
Bivariate usual	e	$\approx 33\%$	0.00098	0.00851	0.02063
Bivariate w/ CT			0.00066	0.00787	**0.01729**
Bivariate w/ PCO			0.00081	0.00759	0.02174
Bivariate w/ CT+ PCO			**0.00051**	**0.00700**	0.02092
Bivariate usual	f	$\approx 43\%$	0.00098	0.00854	0.01591
Bivariate w/ CT			**0.00075**	**0.00811**	**0.01159**
Bivariate w/ PCO			0.00115	0.00930	0.02081
Bivariate w/ CT+ PCO			0.00081	0.00850	0.01310
Bivariate usual	g	$\approx 16\%$	0.00113	0.00902	0.02899
Bivariate w/ CT			**0.00076**	**0.00832**	**0.02398**
Bivariate w/ PCO			0.00123	0.00940	0.02890
Bivariate w/ CT+ PCO			0.00080	0.00850	0.02437
Bivariate usual	h	$\approx 24\%$	0.00122	0.00900	0.04882
Bivariate w/ CT			**0.00066**	**0.00686**	0.04781
Bivariate w/ PCO			0.00153	0.01005	**0.03509**
Bivariate w/ CT+ PCO			0.00188	0.01219	0.04692

Concluding remarks

- The model uses an age-varying smoothness parameter to enhance accuracy and reliability in mortality estimates based on data credibility.
- Joint modelling of male and female mortality rates prevents unrealistic divergences and maintains coherent mortality trajectories for both sexes.
- The model effectively addresses missing information by using available data from one subpopulation to improve estimates in another, ensuring more reliable extrapolations and mitigating the impact of incomplete data.

Main References

- Tang, K. H., Dodd, E., and Forster, J. J. (2022). *Joint modelling of male and female mortality rates using adaptive p-splines*. Annals of Actuarial Science, 16:119–135.
- `BayesMortalityPlus`: Bayesian mortality models. Laboratório de Matemática Aplicada (LabMA/UFRJ), `R package`, GPL-3 license.
- Moura, L; Figueiredo, L.F.V., Lobo, V.G.R., Fonseca, T. C.O. and Alves, M.B. *BayesMortalityPlus: A package in R for Bayesian mortality modelling*. ArXiv: `https://arxiv.org/pdf/2306.01575`.
- West, M. and Harrison, J. (1997). *Bayesian forecasting and dynamic models*. Springer, 2 edition.
- Office for National Statistics (2019). *Population estimates and deaths by single year of age for England and Wales and the UK, 1961 to 2018*.

Thank you!