Smooth Constrained Mortality Forecasting
with an extension to multi-population forecasts

Giancarlo Camarda

ined

Institut National d'Études Démographiques, France

3rd Human Mortality Database Users Conference

18 Nov 2020

Outline

Mortality forecasting methods: Quick overview

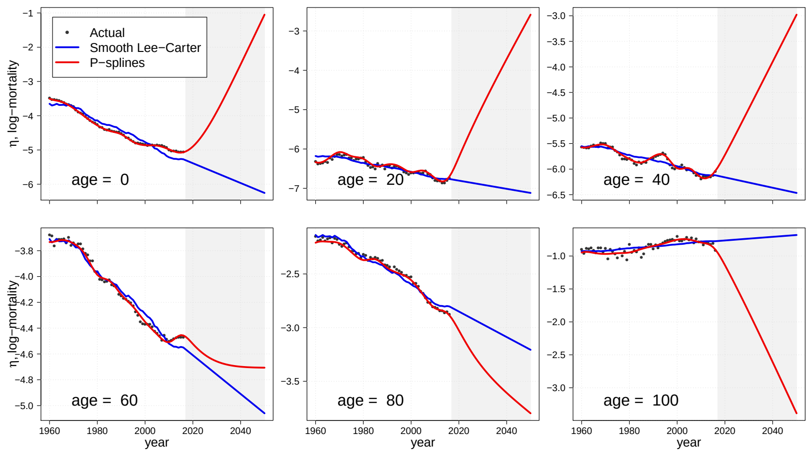
Parametric models
- parsimonious (for adult ages)
- large number of parameters is often necessary
- hard to disentangle physical meaning of each parameter
- simultaneous forecast of all parameters is a challenge
- rigid structure

Lee-Carter variants
- univariate time index condenses mortality development
- fix age effect over time~~fix age effect over time~~
- lack of smoothness in forecast~~lack of smoothness in forecast~~
- extremely large number of parameters~~extremely large number of parameters~~
- rigid structure

Nonparametric approaches, e.g. *P*-splines
- extremely good fit
- flexible structure
- *wisely* parsimonious
- purely data-driven: no demography involved !!

Outcomes from plain approaches

- USA, males, ages 0-105, years 1960-2016, forecast years 2017-2050



- Source: Human Mortality Database
- Delwarde et al. (2007) for the smooth Lee-Carter
- Currie et al. (2004) for the two-dimensional *P*-splines,

The idea

P-splines:

- flexibility (good fit)
- blind adherence extrapolation

Lee-Carter:

- rigidity (bad fit)
- past development drives future mortality

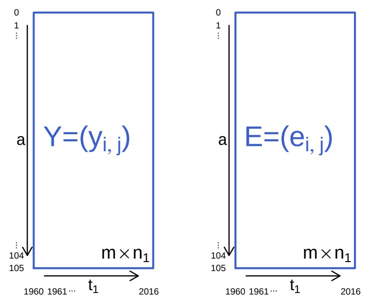
Constrained P-splines, CP-splines:

Incorporate into P-splines demographic information

- flexibility (good fit)
- future development based on prior demographic knowledge

Data structure & basic model

Observed data



Assumption:

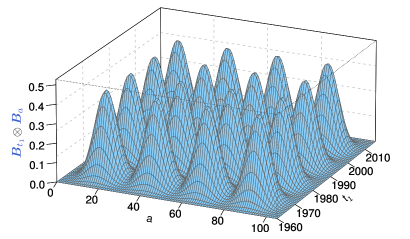
$$y_{ij} \sim \mathcal{P}(e_{ij}\, \mu_{ij})$$

Model:

$$\boldsymbol{y} = \mathrm{vec}(\boldsymbol{Y})$$
$$\boldsymbol{e} = \mathrm{vec}(\boldsymbol{E})$$
$$\ln(\boldsymbol{y}) = \ln(\boldsymbol{e}) + \ln(\boldsymbol{\mu}) = \ln(\boldsymbol{e}) + \boldsymbol{\eta} = \ln(\boldsymbol{e}) + \boldsymbol{B}\,\boldsymbol{\alpha}$$

$\boldsymbol{B} = \boldsymbol{B}_{t_1} \otimes \boldsymbol{B}_a$

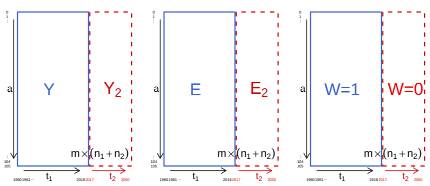
$\boldsymbol{\alpha}$: penalized coefficients

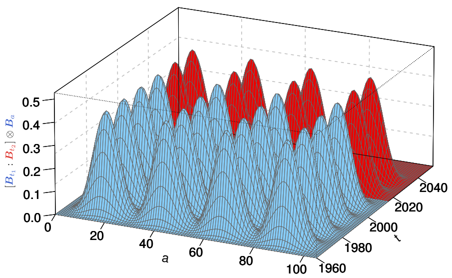


- estimated by a penalized iterative weighted least-squares algorithm

Forecasting with P-splines

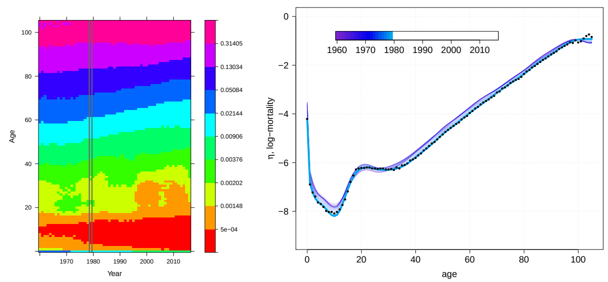
- Forecasting is treated as a missing value problem
- Data are augmented with arbitrary future values and we define a weight matrix
- B-spline bases are augmented too: $\boldsymbol{B} = [\boldsymbol{B}_{t_1} : \boldsymbol{B}_{t_2}] \otimes \boldsymbol{B}_a$
- The original algorithm can be adapted
- No information about mortality structure is used



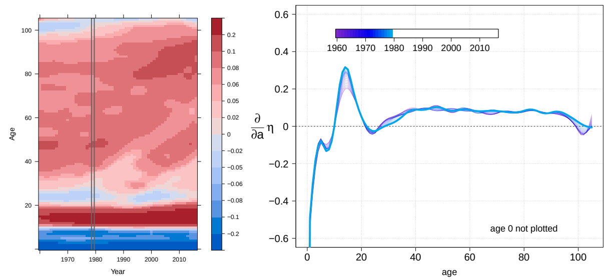


Observing mortality patterns (over ages)

Actual and Smooth log-mortality

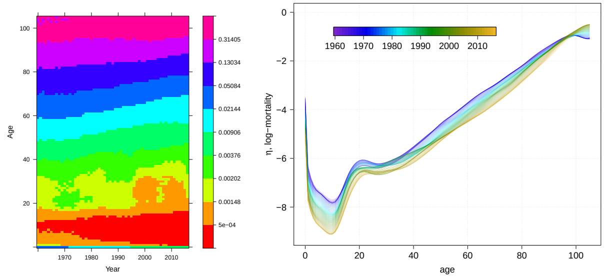


Rate-of-aging from smooth mortality

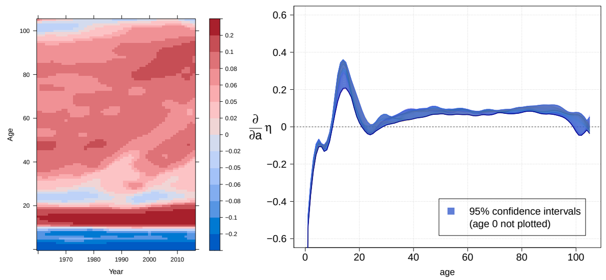


Observing mortality patterns (over ages)

Actual and Smooth log-mortality

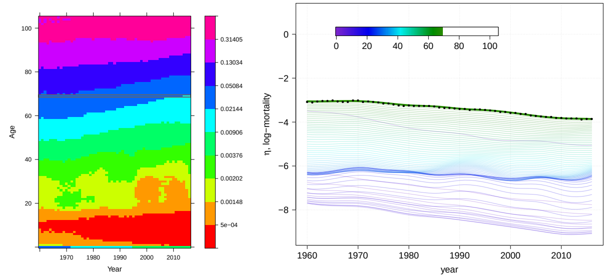


Rate-of-aging from smooth mortality

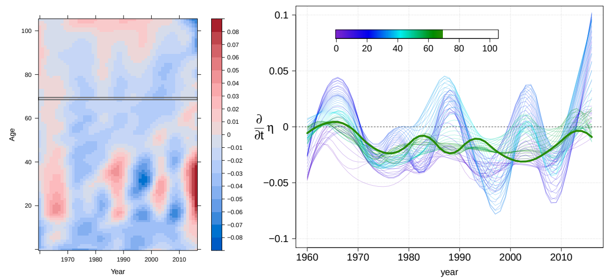


Observing mortality patterns (over years)

Actual and Smooth log-mortality

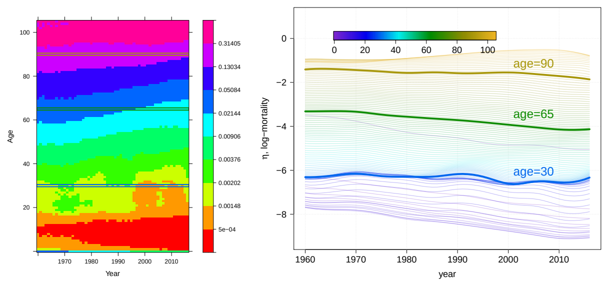


Rate-of-change from smooth mortality

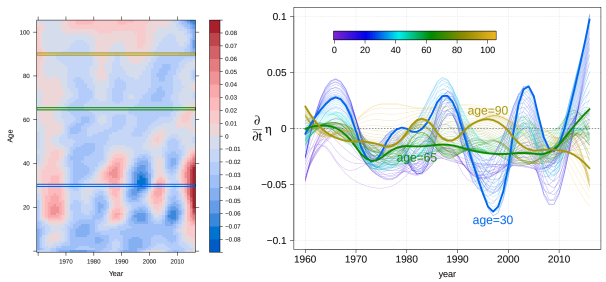


Observing mortality patterns (over years)

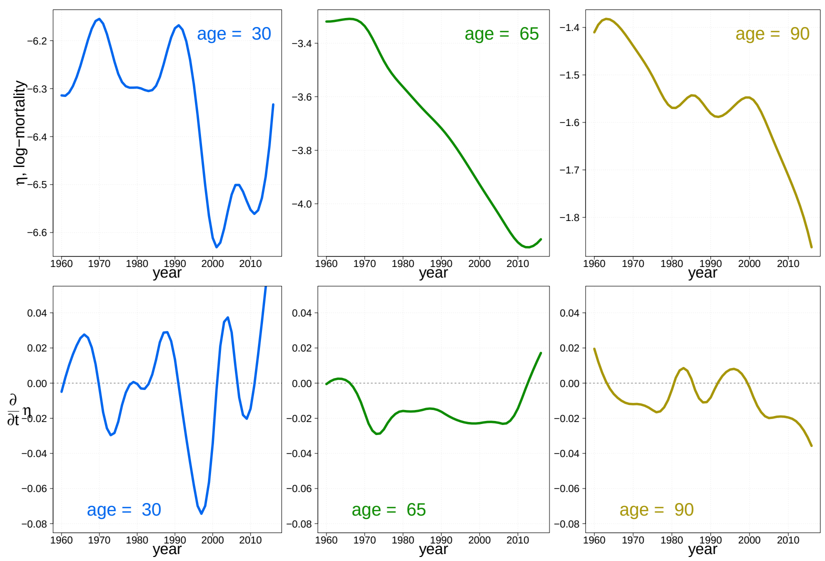
Actual and Smooth log-mortality



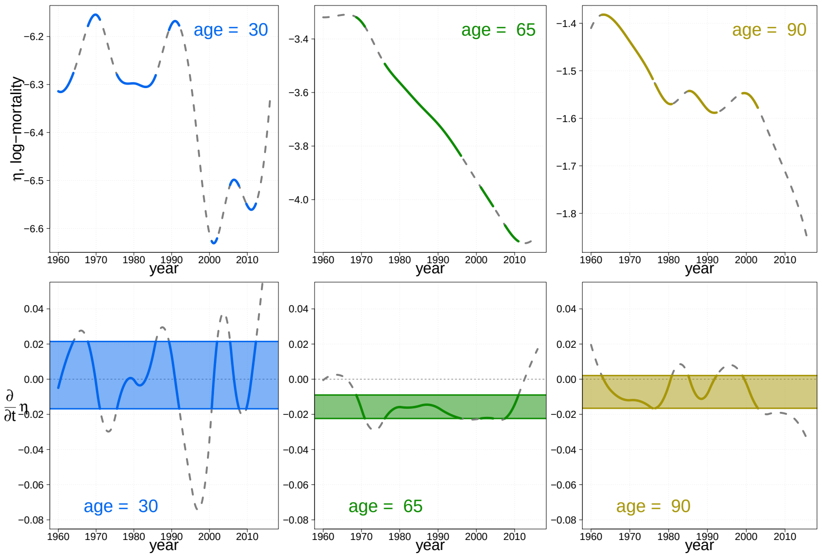
Rate-of-change from smooth mortality



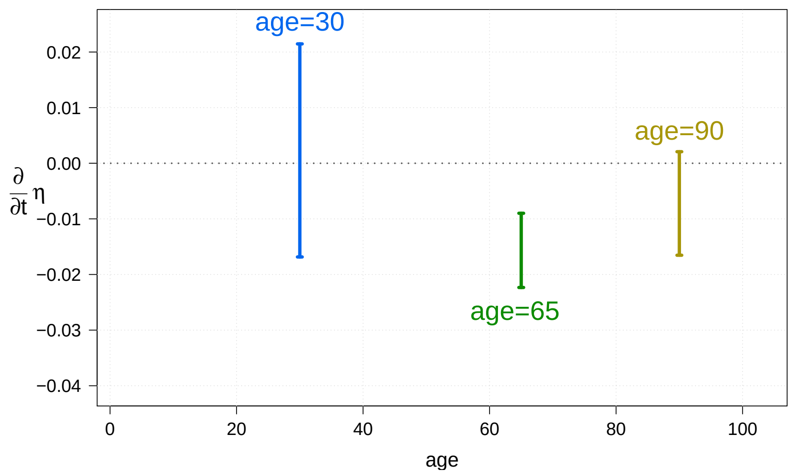
Observing mortality patterns (over years)



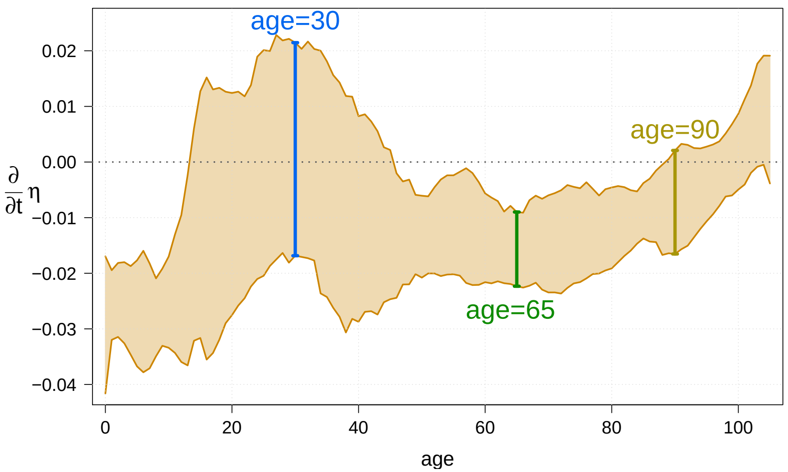
Observing mortality patterns (over years)



Observing mortality patterns (over years)



Observing mortality patterns (over years)

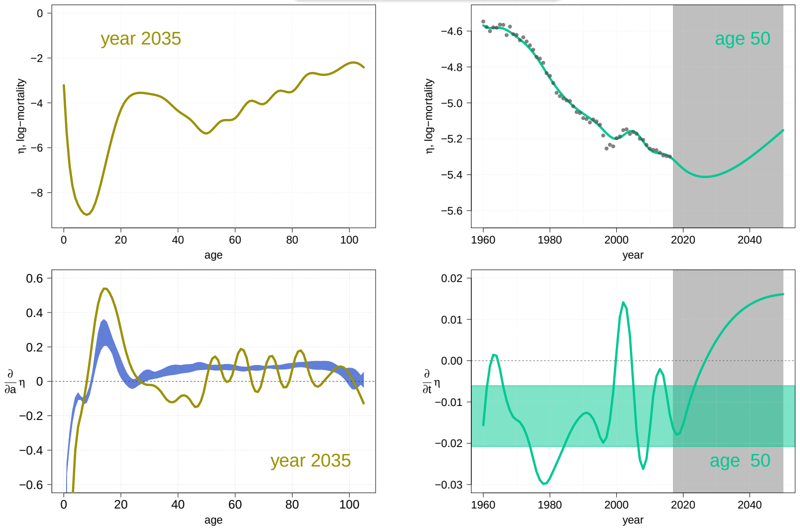


Working on the shape

- Disregarding information about shapes seems unreasonable
- Future mortality *must* follow data-driven age-profiles and rate-of-change
- We constrain derivatives of future mortality to lay within certain confidence intervals of observed derivatives
- Asymmetric penalties are employed for this purpose:
  - within each iteration, whenever current estimations present derivatives (in future years) out of the desired intervals, a penalty intervenes
- Let's see how it works in action for a specific year (2035) and a specific age (50)

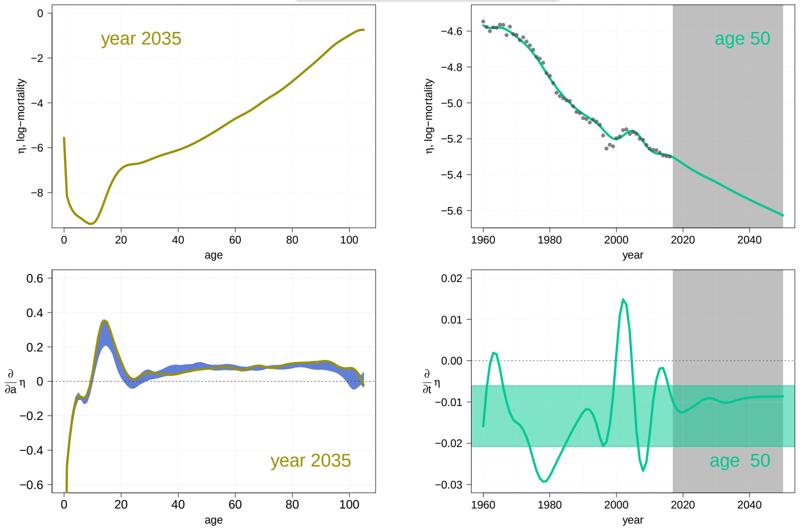
Constraining on derivatives in action

Without asymmetric penalty



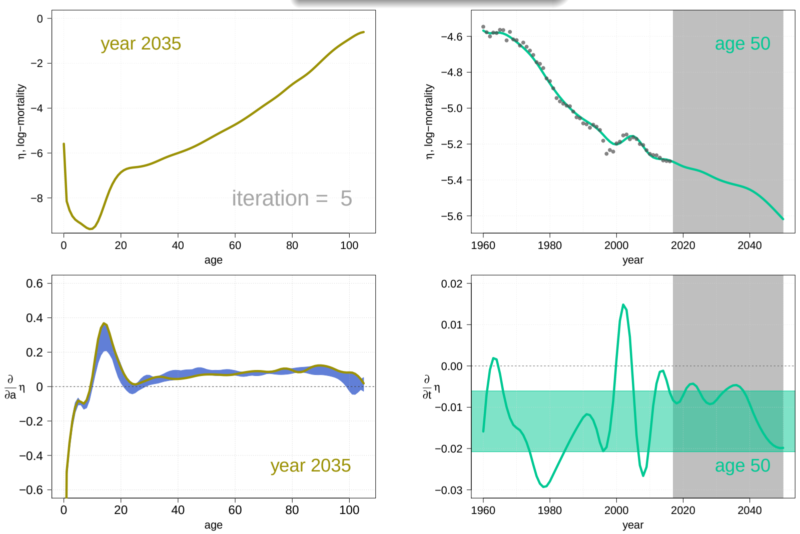
Constraining on derivatives in action

With asymmetric penalty

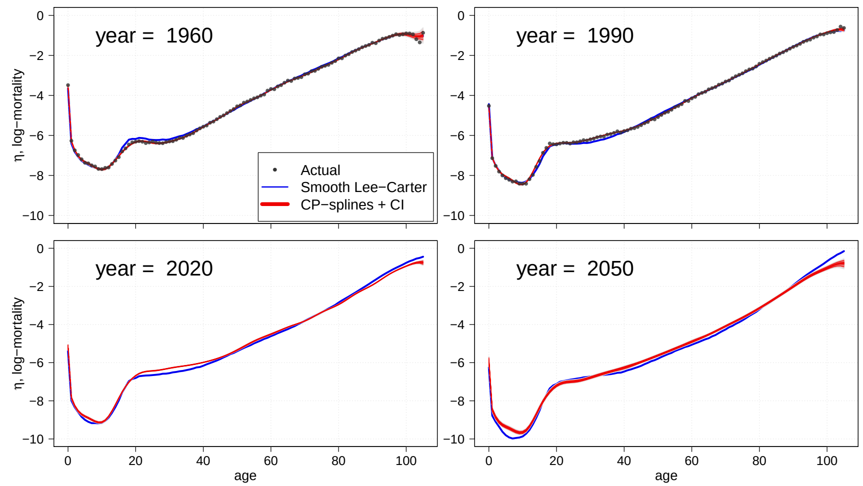


Constraining on derivatives in action

With asymmetric penalty

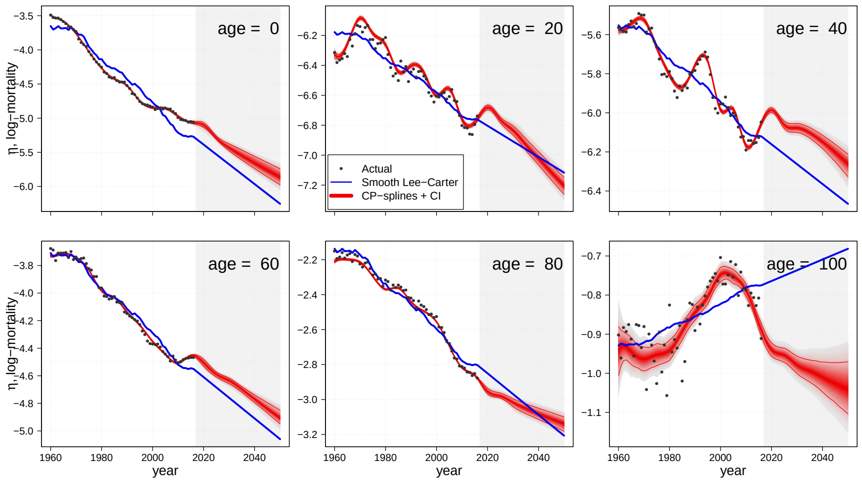


Outcomes for US, males



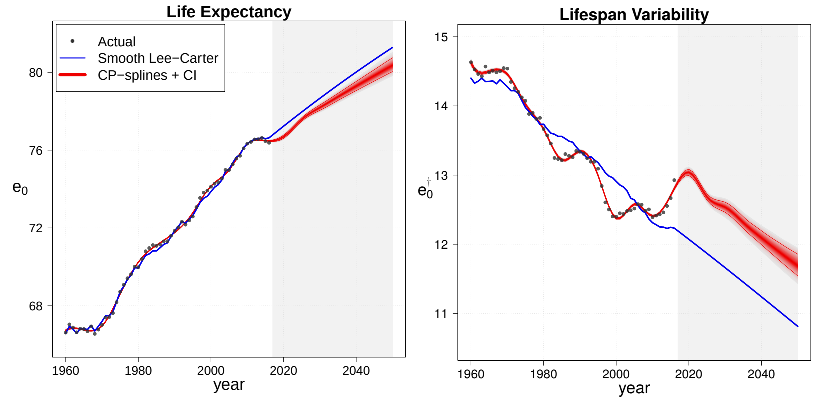
Log-mortality. Ages 0-105, observed years 1960-2016, forecast up to 2050

Outcomes for US, males



Log-mortality. Ages 0-105, observed years 1960-2016, forecast up to 2050

Outcomes for US, males



Life Expectancy and Lifespan variability measure ($e_0^\dagger$).

Ages 0-105, observed years 1960-2016, forecast up to 2050

What I haven't shown here, . . .

- . . . but you can find in the paper below:
  - All associated equations
  - Applications to more populations
  - How infant mortality is addressed in a smoothing setting
  - Bootstrap procedure to obtain confidence intervals
  - Out-of-sample performance
  - Comparison with other alternative methods
  - Effect of changing time-window on the outcomes
  - Sensitivity analysis on confidence level in rate-of-change over t
  - Reproducible R-code

Camarda, C. G. (2019).
Smooth Constrained Mortality Forecasting.
Demographic Research. **41** (38), 1091-1130

Generalizing *CP*-splines

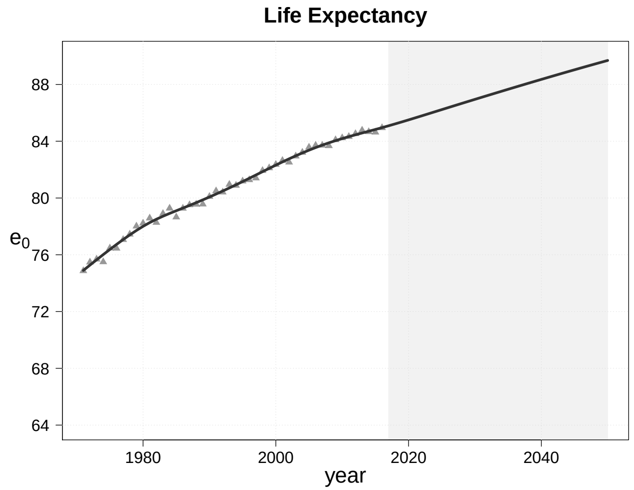
- Recently, there has been a growing strand of research on so-called *coherent* forecasting
- What does it mean? Mortality forecast of
  - males and females from the same population
  - a group of countries
  - more causes of death
  - sub-populations belonging to the same country

The idea:

given forecast values for the whole country, constrain mortality differences between each sub-population and overall country to lay within (a range of) observed past differences

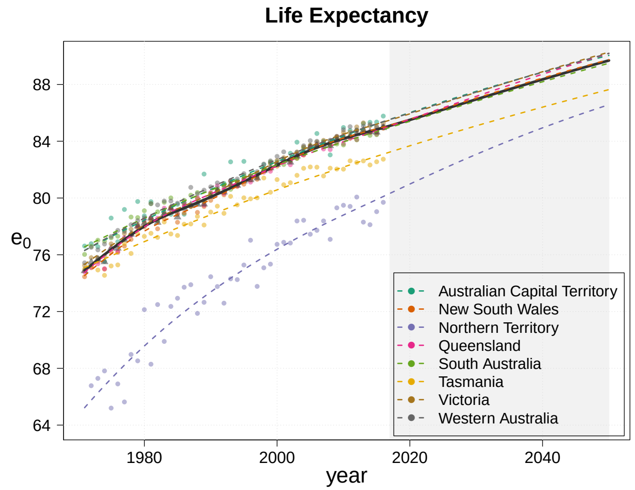
- Let's take Australia and its 8 territories/states
- Data: Females, Ages 0-100, Observed years 1971-2016, forecast up to 2050

CP-splines on Australia and its territories



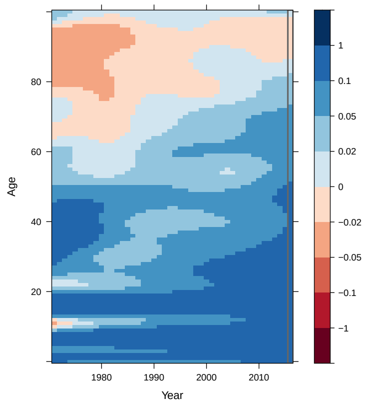
Actual, estimated and forecast life expectancy by *CP*-splines.
Australia

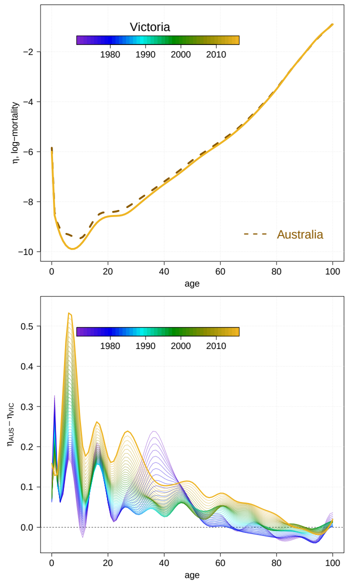
CP-splines on Australia and its territories



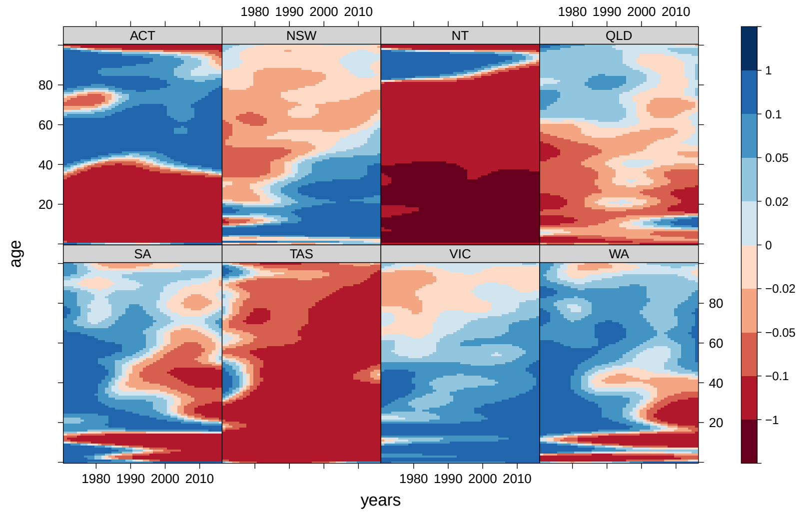
Actual, estimated and forecast life expectancy by *CP*-splines.
Australia and its territories

Differences in log-mortality: only Victoria



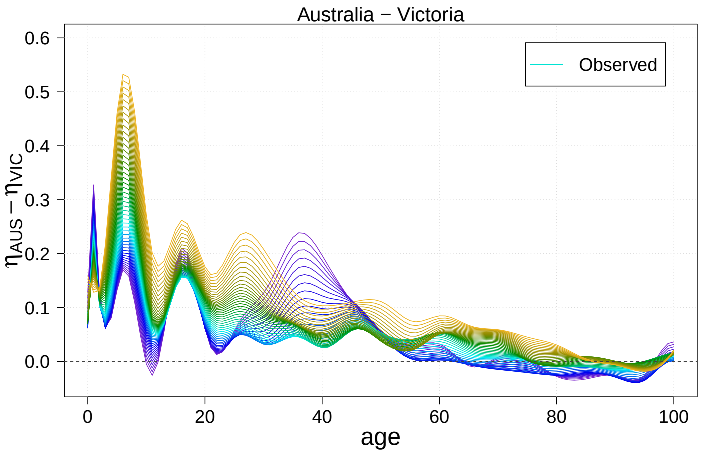


Differences in log-mortality: all territories



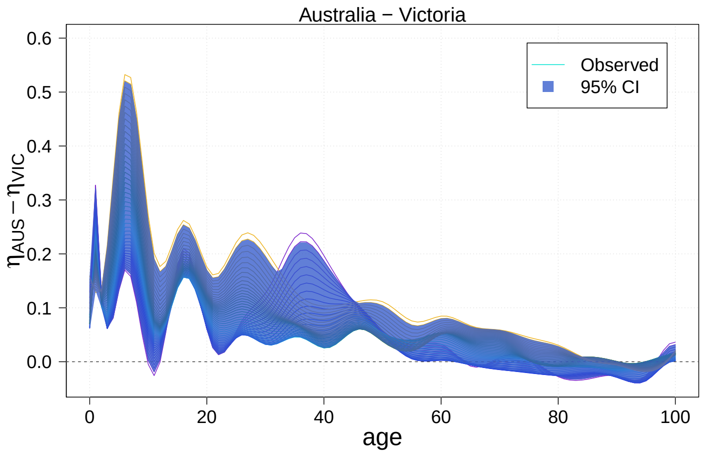
Observed differences in smooth log-mortality: Australia - each territory

Differences in future log-mortality with *CP*-splines



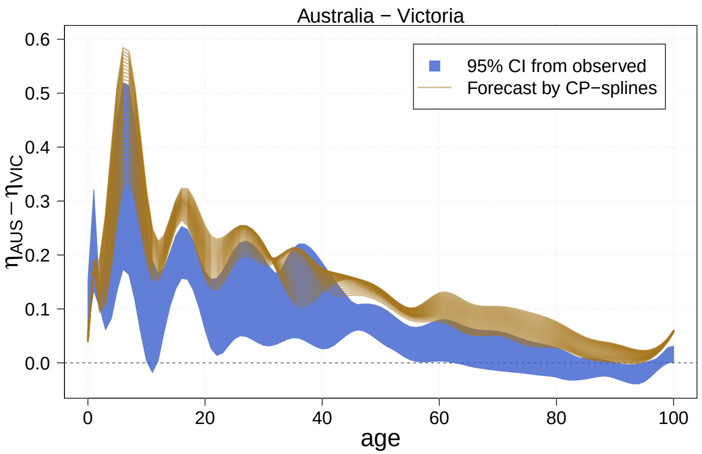
Differences in estimated smooth log-mortality

Differences in future log-mortality with *CP*-splines



Differences with the 95% CI in estimated log-mortality

Differences in future log-mortality with *CP*-splines

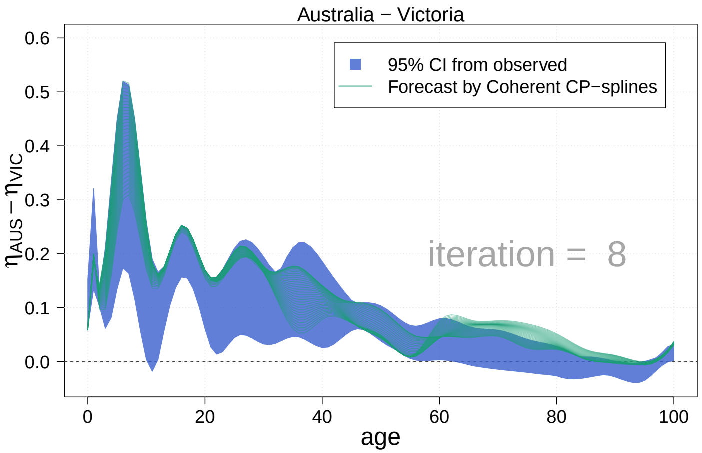


95% CI of estimated & forecast (*CP*-splines) differences in log-mortality

Constraining differences: Coherent *CP*-splines

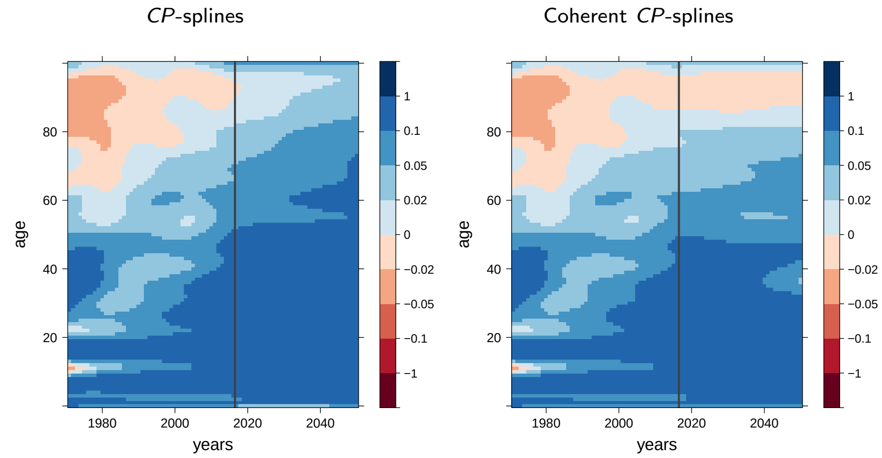
- Incorporate constraints on future differences as *CP*-splines do with relative derivatives
- Asymmetric penalty can be adapted
- Coherence with respect to known overall mortality preserved
- Each sub-population can be treated independently
- We achieve age-specific coherence in future years simultaneously
- Limits:
  - Knowledge about the overall future mortality is necessary
  - Range from past differences are kept in the future

Asymmetric penalty in action on differences



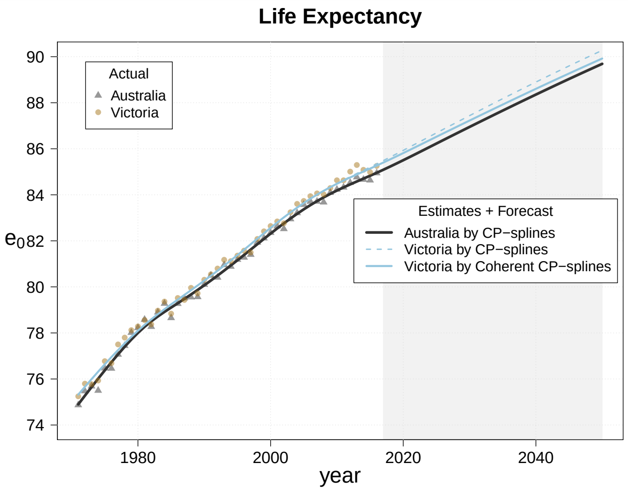
95% CI of estimated & forecast (Coherent *CP*-splines) differences in log-mortality

Forecast differences in log-mortality: Australia - Victoria



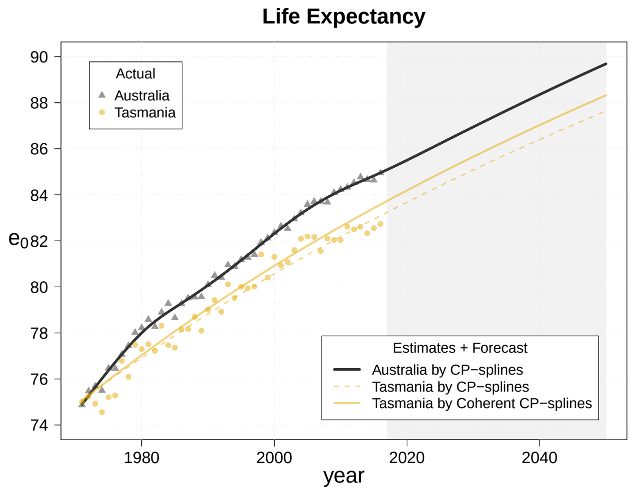
Estimated and forecast differences in log-mortality

Coherent forecast e_0 for Victoria



Actual, estimated and forecast life expectancy by (coherent) *CP*-splines. Australia and Victoria

Coherent forecast e_0 for Tasmania



Actual, estimated and forecast life expectancy by (coherent) *CP*-splines. Australia and Tasmania

Coherent forecast e_0 for all territories in 2050. Females

	CP-splines	Coherent *CP*-splines	Δ
Australia	89.69	-	-
Australian Capital Territory	90.05	89.91	-0.14
New South Wales	89.75	89.64	-0.11
Northern Territory	86.57	86.11	-0.46
Queensland	90.18	89.90	-0.27
South Australia	89.49	89.73	0.24
Tasmania	87.64	88.32	0.68
Victoria	90.28	89.92	-0.36
Western Australia	90.20	90.08	-0.12

Forecast female life expectancy in 2050 by (coherent) *CP*-splines. Australia and its territories

Coherent forecast e_0 for all territories in 2050. Males

	CP-splines	Coherent *CP*-splines	Δ
Australia	87.04	-	-
Australian Capital Territory	88.20	87.92	-0.29
New South Wales	87.47	87.08	-0.40
Northern Territory	81.78	83.13	1.35
Queensland	86.75	86.79	0.04
South Australia	86.88	86.93	0.05
Tasmania	84.94	85.78	0.85
Victoria	87.97	87.47	-0.51
Western Australia	87.49	87.33	-0.16

Forecast male life expectancy in 2050 by (coherent) *CP*-splines. Australia and its territories

Concluding remarks

- We combine a powerful statistical methodology with prior demographic information
- From *P*-splines we gain good fit, flexibility and smooth outcomes
- With additional constraints, we incorporate knowledge about mortality shapes to guide future developments
- We enforce shape constraints by asymmetric penalties on the observed mortality derivatives

CP-splines

- *CP*-splines can be extended to achieve coherent mortality forecast for multiple sub-populations

To-do list

- From a methodological perspective:
  - Addressing cohort effects by constraining derivatives over specific diagonals of the mortality surface
  - Using *CP*-splines to guide expert-based forecast approaches
  - Forecasting of cause of death data
  - Simultaneous *CP*-splines estimation for both sexes
- Further applications:
  - Coherent *CP*-splines on more heterogeneous sub-populations
  - *CP*-spline for creating future mortality scenarios
  - Calibrating the choice of possible future age pattern and time trends based on broader past experiences

3^{rd} HMD Users Conference

Smooth Constrained Mortality Forecasting with an extension to multi-population forecasts

Thanks for your attention.
Comments and questions?

More info and R routines available on:
sites.google.com/site/carlogiovannicamarda

Additional slides

The Penalized IWLS

- Given data:

  $\boldsymbol{d} = \texttt{vec}(\boldsymbol{D})$
  $\boldsymbol{e} = \texttt{vec}(\boldsymbol{E})$
- And model

$$\ln(\boldsymbol{d}) = \ln(\boldsymbol{e}) + \ln(\boldsymbol{\mu}) = \ln(\boldsymbol{e}) + \boldsymbol{\eta} = \ln(\boldsymbol{e}) + \boldsymbol{B}\,\boldsymbol{\alpha}$$

  $\boldsymbol{B} = \boldsymbol{B}_{y_1} \otimes \boldsymbol{B}_x$
  $\boldsymbol{\alpha}$: penalized coefficients
- Estimate $\boldsymbol{\alpha}$ by penalized IWLS:

$$(\boldsymbol{B}'\tilde{\boldsymbol{W}}\boldsymbol{B} + \boldsymbol{P})\tilde{\alpha} = \boldsymbol{B}'\tilde{\boldsymbol{W}}\tilde{\boldsymbol{z}}$$

  where
  - $\tilde{\boldsymbol{z}} = (\boldsymbol{d} - \boldsymbol{e} * \tilde{\mu})/\boldsymbol{e} * \tilde{\mu} + \tilde{\eta}$
  - $\tilde{\boldsymbol{W}} = \mathrm{diag}(\boldsymbol{e} * \tilde{\mu})$

Asymmetric penalty in formulas over ages/1

- $$\frac{\frac{\partial}{\partial \boldsymbol{a}}\hat{\boldsymbol{\mu}}}{\hat{\boldsymbol{\mu}}} = \frac{\partial}{\partial \boldsymbol{a}}\ln(\hat{\mu}) = \frac{\partial}{\partial \boldsymbol{a}}\hat{\boldsymbol{\eta}} = \boldsymbol{D}_a^{t_1}\,\hat{\boldsymbol{\alpha}}\,,$$

  where

$$\boldsymbol{D}_a^{t_1} = \boldsymbol{B}_{t_1} \otimes \boldsymbol{C}_a \quad \text{and} \quad \boldsymbol{C}_a = \frac{1}{h}\left[{}^{q-1}\boldsymbol{B}_a^{k} - {}^{q-1}\boldsymbol{B}_a^{k-1}\right]$$

  with h, q and k being knot-distance, degree and positions of the original B-spline basis, $\boldsymbol{B}_a$.
- δ_L^a and δ_U^a: lower and upper bounds of CI of the derivatives
- Keep same constraints for all years $\Rightarrow$ augment $\boldsymbol{\delta}$ over both dimensions:

$$\boldsymbol{g}_L^a = \boldsymbol{1}_{n_1+n_2} \otimes \boldsymbol{\delta}_L^a$$
$$\boldsymbol{g}_U^a = \boldsymbol{1}_{n_1+n_2} \otimes \boldsymbol{\delta}_U^a$$

- Similar computation is performed over years

Asymmetric penalty in formulas over ages/2

- New penalized IWLS:

$$(\breve{B}' V \tilde{W} \breve{B} + P + P^a + P^t)\tilde{\alpha} = \breve{B}' V \tilde{W} \tilde{z} + p^a + p^t \,.$$

where

$$\begin{aligned} P^a &= P^a_L + P^a_U \\ P^t &= P^t_L + P^t_U \end{aligned} \quad \text{and} \quad \begin{aligned} p^a &= p^a_L + p^a_U \\ p^t &= p^t_L + p^t_U \end{aligned} \,.$$

- As example, lower bounds over ages:

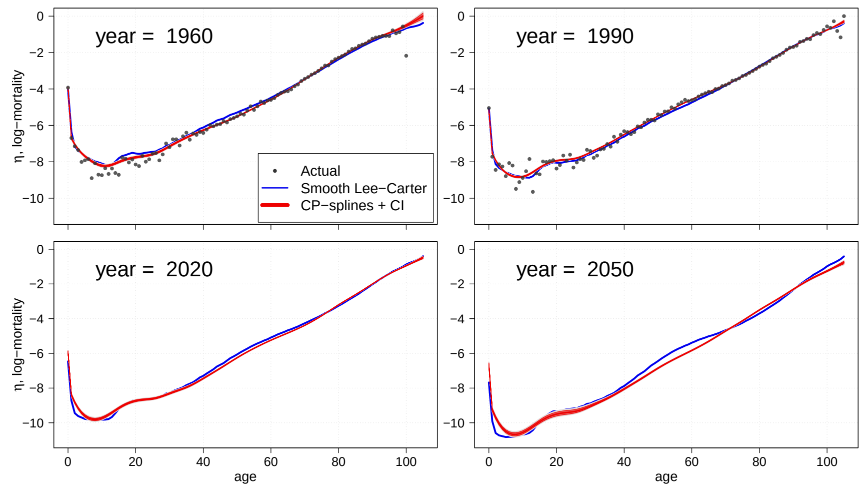
$$\begin{aligned} P^a_L &= \kappa\, D_a^{t_1+t_2\,\prime}\, \mathrm{diag}(s\, v^a_L)\, D_a^{t_1+t_2} \\ p^a_L &= \kappa\, D_a^{t_1+t_2\,\prime}\, \mathrm{diag}(s\, v^a_L)\, g^a_L \end{aligned} \quad \text{with} \quad v^a_L = \begin{cases} 0 & \text{if} \quad D_a^{t_1+t_2}\tilde{\alpha} \geqslant g^a_L \\ 1 & \text{if} \quad D_a^{t_1+t_2}\tilde{\alpha} < g^a_L \end{cases}$$

where

$$v^a_L = \begin{cases} 0 & \text{if} \quad D_a^{t_1+t_2}\tilde{\alpha} \geqslant g^a_L \\ 1 & \text{if} \quad D_a^{t_1+t_2}\tilde{\alpha} < g^a_L \end{cases} ,$$

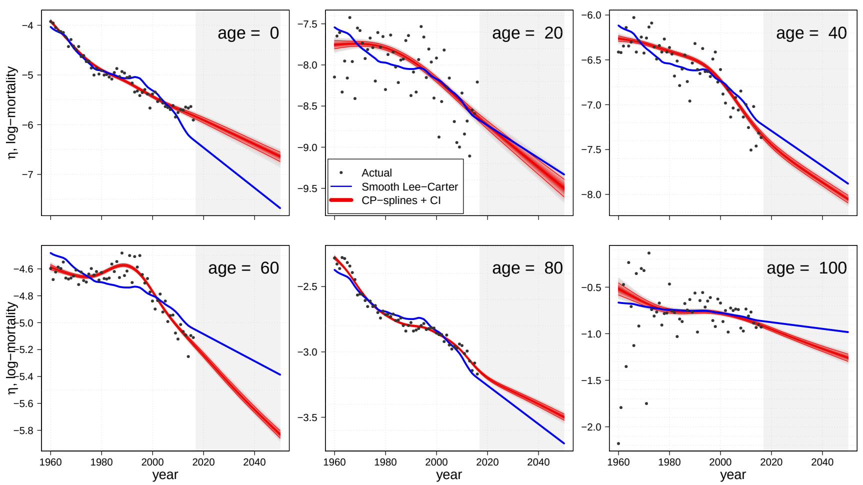
and s is a 0/1 vector equal to 1 when the constraint is to be applied (future years).

Outcomes for Denmark, females



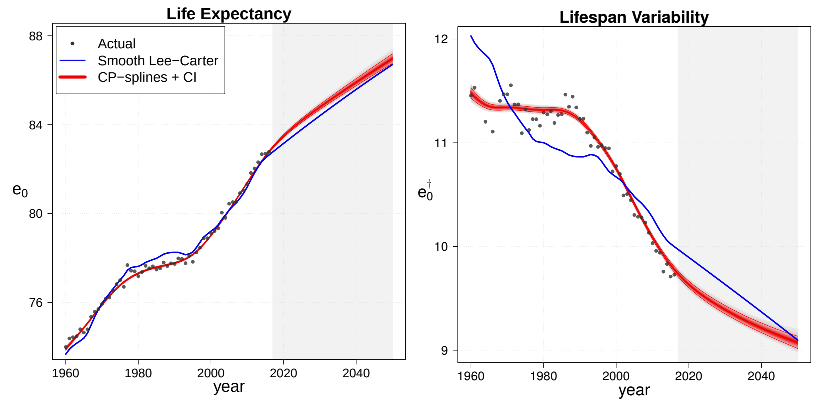
Log-mortality. Ages 0-105, observed years 1960-2016, forecast up to 2050

Outcomes for Denmark, females



Log-mortality. Ages 0-105, observed years 1960-2016, forecast up to 2050

Outcomes for Denmark, females



Life Expectancy and Lifespan variability measure ($e_0^\dagger$).
Ages 0-105, observed years 1960-2016, forecast up to 2050

Confidence level in rate-of-change over time