Pension Systems / Demography & Mortality

Lecture notes: Mortality – part II

University of Copenhagen, Autumn 2021

Snorre Jallbjørn & Søren F. Jarner

Recap – Notation and operators

Suppose $v(x,t)$ is a demographic function of interest and $w(x,t)$ is some weighting function, where x is age and t is time.

- Mean value: Let a *bar* over a function denote the (weighted) average over the x-variable

$$\bar{v}(t) \coloneqq \frac{\int_0^\infty v(x,t)w(x,t)\mathrm{d}x}{\int_0^\infty w(x,t)\mathrm{d}x}$$

  - Note: Not clear from notation which weights are used!

- Time derivative: Let a *dot* over a function denote its derivative w.r.t. time

$$\dot{v}(x,t) \coloneqq \frac{\partial}{\partial t}v(x,t)$$

- Relative derivative: Let an *acute accent* over a function denote its relative deriv. w.r.t. time

$$\acute{v}(x,t) = \frac{\dot{v}(x,t)}{v(x,t)} \coloneqq \frac{\partial}{\partial t}\log v(x,t)$$

Recap – Vaupel & Canudas-Romo decomposition

Covariance can be decomposed into $\mathrm{Cov}_w(v,u) = \overline{vu} - \bar{v}\bar{u}$

The mortality improvement rate is $\rho(x,t) \coloneqq -\acute{\mu}(x,t) = -\frac{\partial}{\partial t}\log\mu(x,t)$

- The average improvement rate is $\bar{\rho}(t) = \int_0^\infty \rho(x,t)\mu(x,t)S(x,t)\mathrm{d}x$

The Vaupel & Canudas-Romo decomposition

$$\dot{e}(0,t) = \bar{\rho}(t)e^\dagger(0,t) + \mathrm{Cov}_f(\rho,e),$$

where $e^\dagger(x,t) = \frac{1}{S(x,t)}\int_x^\infty \mu(u,t)S(u,t)e(u,t)\mathrm{d}u$ is the average number of years of life lost due to death.

Interpretation of $\dot{e}(0,t)$:

- Direct effect: $\bar{\rho}$ is the proportion of deaths averted (lifes saved), while $e^\dagger$ is the avg. years gained per saved life.
- Secondary effect: Covariance term captures heterogeniety in ρ

Recap – Life table entropy

The life table entropy conditional on survival to age x is

$$\mathcal{H}(x,t) = \frac{\int_x^\infty \big(I(u,t) - I(x,t)\big) S(u,t)\mathrm{d}u}{\int_x^\infty S(u,t)\mathrm{d}u} = \frac{e^\dagger(x,t)}{e(x,t)}.$$

It is a measures of concavity of the survival function (and the mortality compression/expansion process).

Two extremes:

1. Maximum equality $(\mathcal{H} = 0)$: A population in which everyone attains the maximum life span (age ω)
2. Maximum disparity $(\mathcal{H} = 1)$: A population in which no age is favored at death, i.e. $\mu(x) = \mu, \forall x \geq 0$

Note that $\mathcal{H}$ can exceed 1. In particular, whenever $e(x)$ is monotonically increasing, e.g. $\mu(x) = \frac{1}{(1+x)^2} \Longrightarrow \mathcal{H} = 2$

By the Vaupel & Canudas-Romo decomposition: $\acute{e}(0,t) = \bar{\rho}(t)\mathcal{H}(0,t) + \frac{\mathrm{Cov}_f(\rho,e)}{e(x,t)}$

The life table entropy can thus be viewed as the life expectancy elasticity subject to a change in mortality

Agenda

#1 Diverging projections and coherence

#2 Li-Lee model

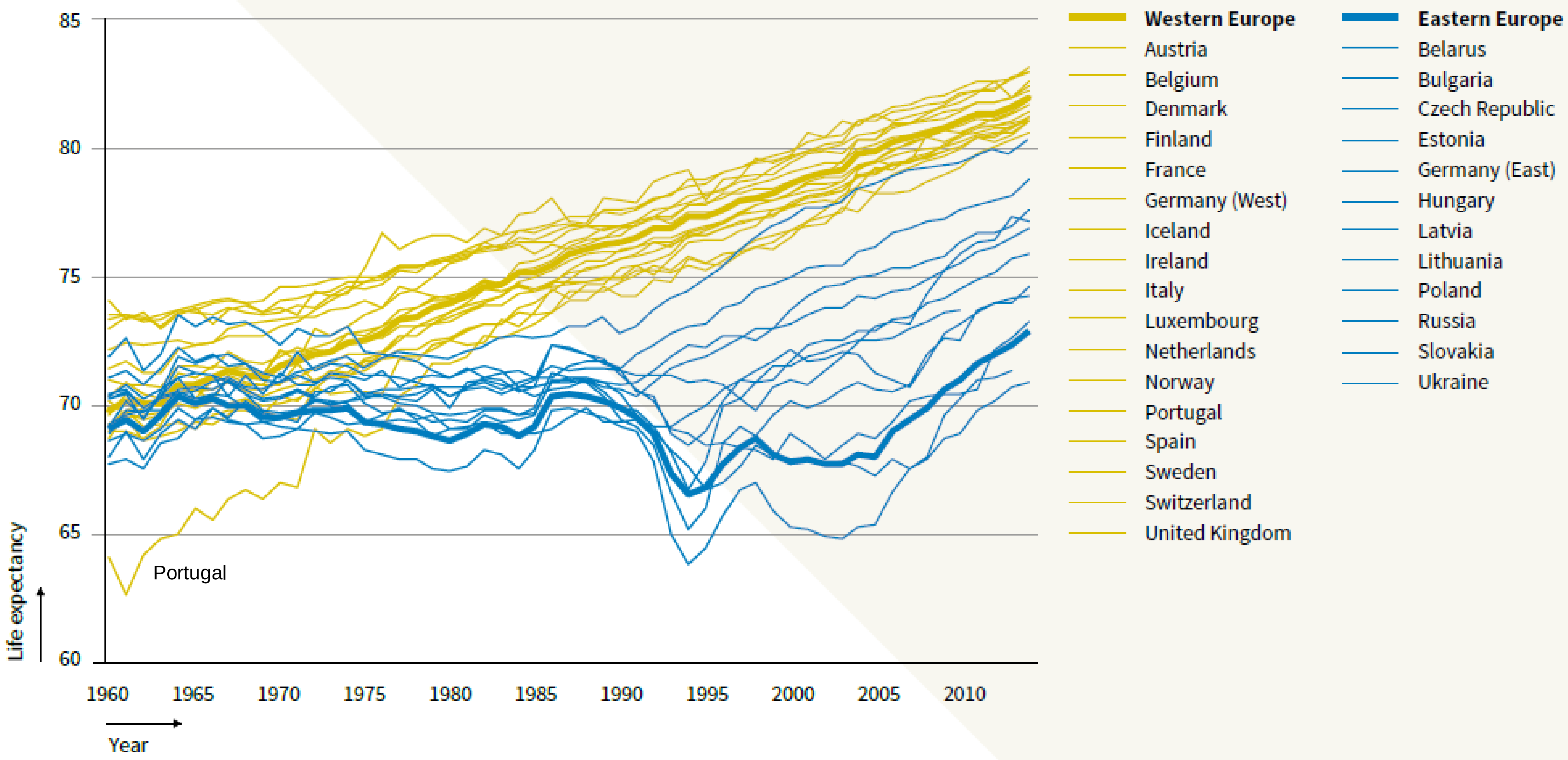
#3 The SAINT projection methodology

#4 Cointegration and modelling the gender gap

#1

Diverging projections and coherence

Common life expectancy evolution in Westen Europe



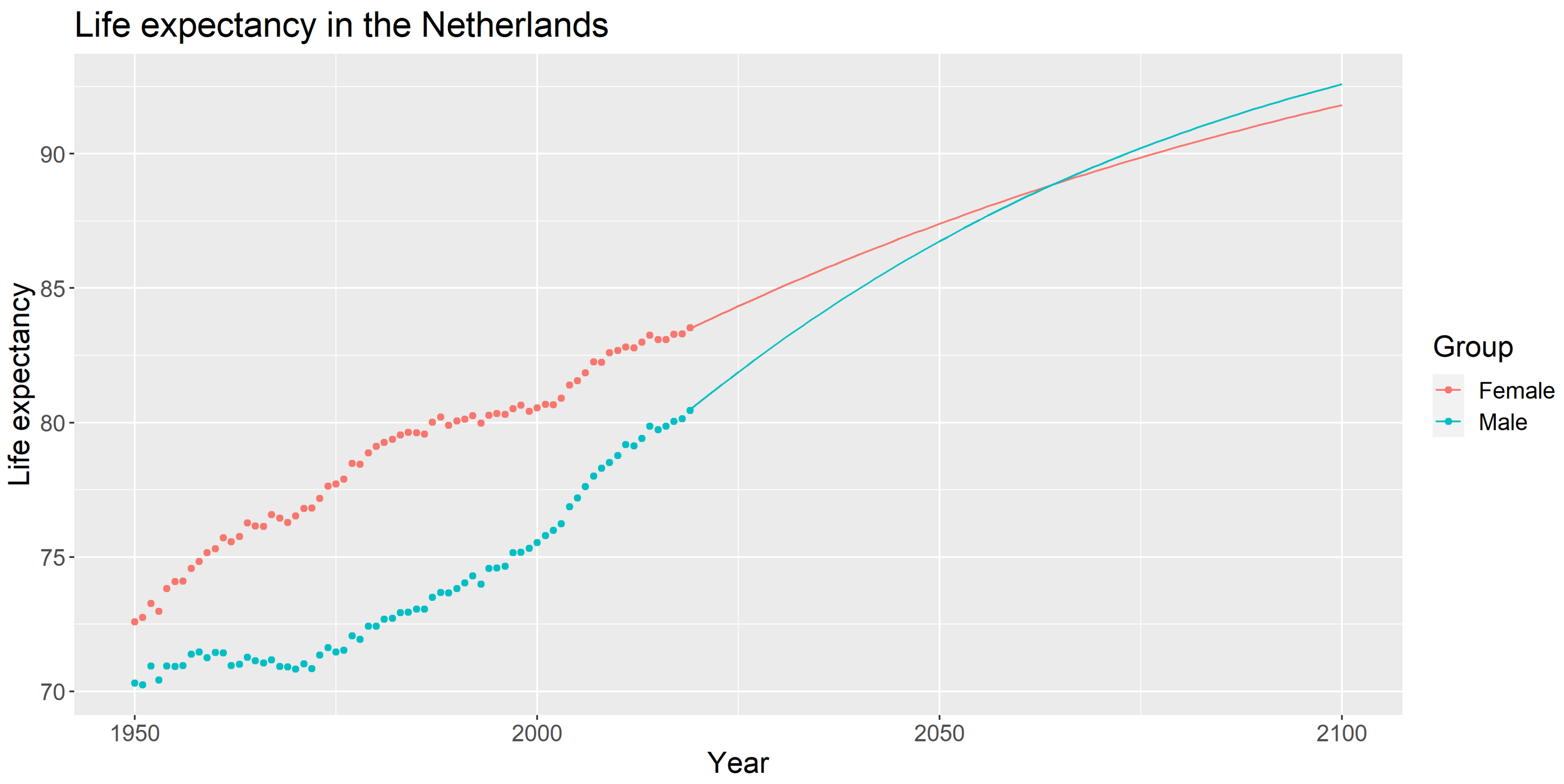
Source data: Human Mortality Database, accessed 26/8/2018

Reproduced from Janssen (2018)

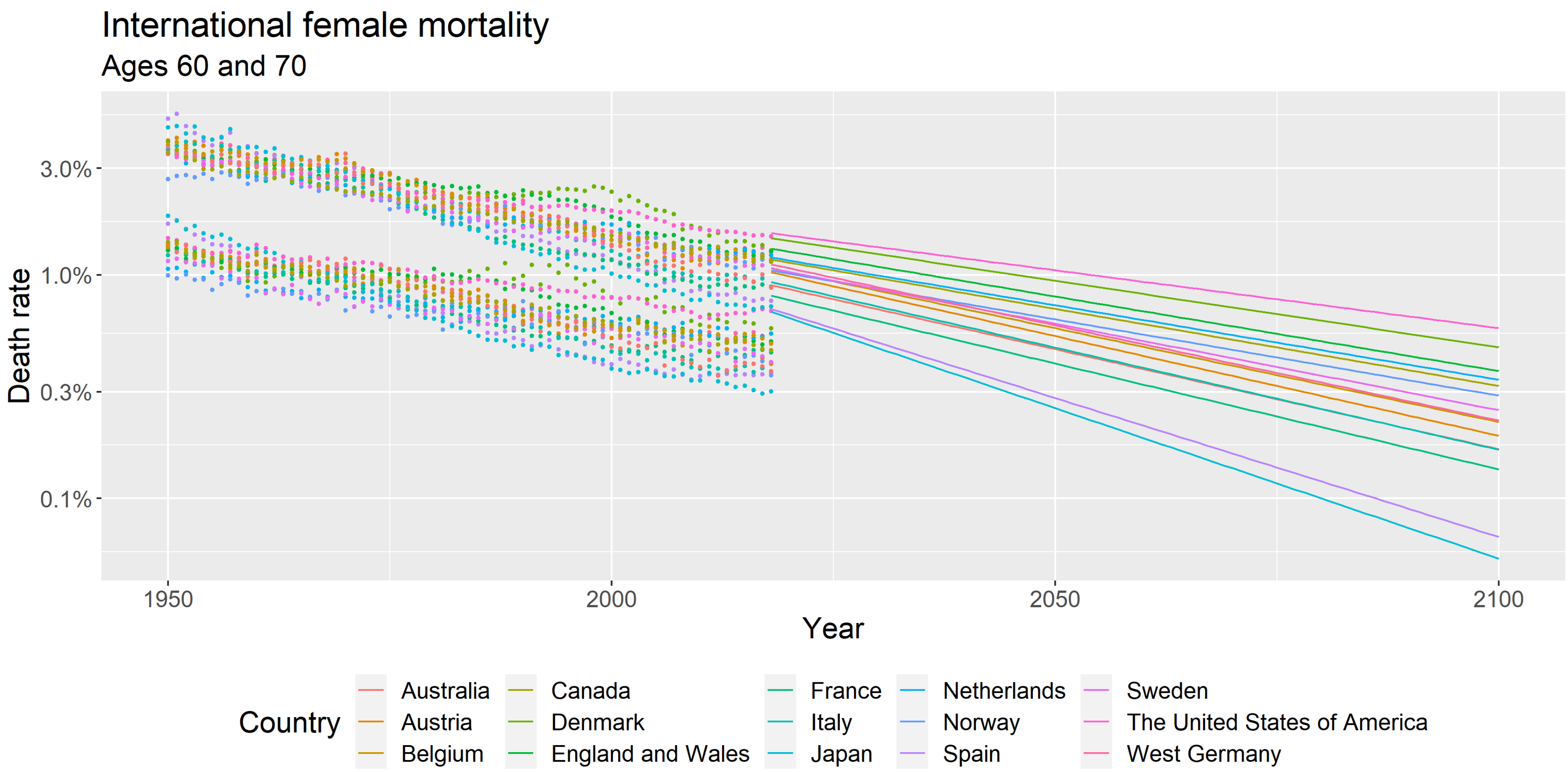
Incoherent mortality projections

- **Convergence of mortality levels**
  - Similar mortality evolution in most developed countries, due to similarities in socio-economic factors, lifestyle, level of treatment etc.
  - Mortality levels are likely to continue to evolve in parallel with temporary country-specific deviations
- **Incoherent mortality projections**
  - Separate analyses exaggerate short-term differences and lead to *diverging* projections
  - Seems highly implausible in the light of historic similarities
  - Same problem for sex-specific forecasts
    - Divergence and mortality crossover

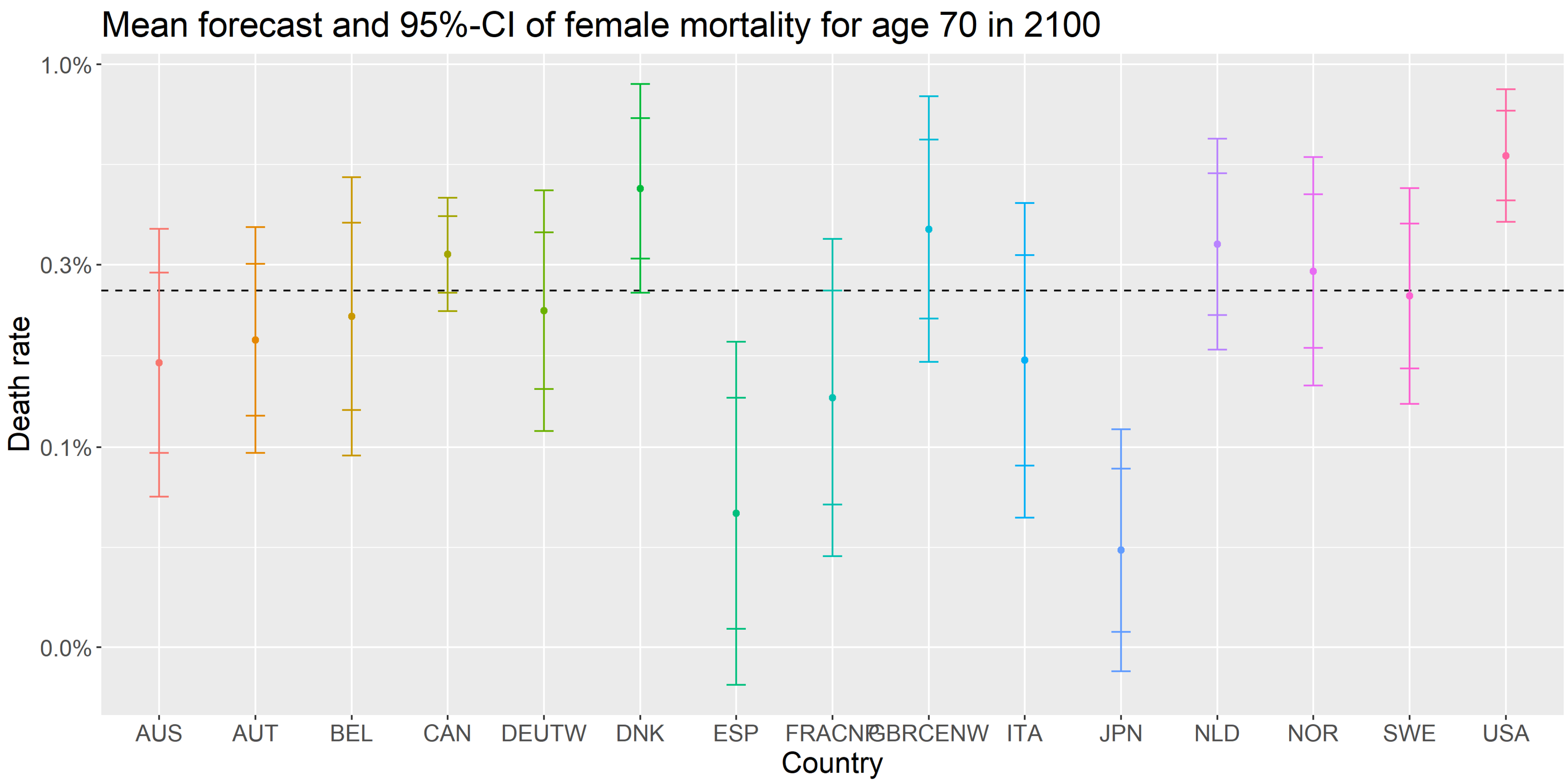
Life expectancy crossover from separate LC-analyses



Divergent Lee-Carter projections from separate analyses



Large variation in projected mortality and levels of uncertainty



Coherence

Let $\mu_i(x,t)$ denote the force of mortality in population i (age x, time t)

A two-population mortality forecast is said to be ***coherent*** if the relative mortality rates converge for each age

$$\frac{\hat{\mu}_1(x,t)}{\hat{\mu}_2(x,t)} \to R(x), \text{for } t \to \infty$$

for positive age-specific constants $R(x)$.

- Formalizes the notion of mortality rates “staying together’’
- … but it is a very strict requirement!

- In the following, we will look at ways of achieving coherence

Coherent mortality modelling: Short-term deviations from a long-term trend

- **Stylized facts**
  - **All countries appear to follow the same long-term trend**
    - But improvements occur at different times in the individual countries
    - Variation in annual improvement rates differ between countries
  - **Separate analyses**
    - Diverging projections with non-overlapping confidence intervals
    - Unreasonable variation in forecasting uncertainty
      - inability of RW to distingush between short- and long-term uncertainty
  - **Coherent mortality projections**
    - Model common long-term trend
    - Allow country-specific short-term deviations from trend

#2

The Li-Lee model

The Li-Lee model

- **The Li-Lee model**
  - Introduced by Li & Lee (2005). ”*Coherent mortality forecasts for a group of populations: An extension of the Lee-Carter method*”
  - The Lee-Carter model:

$$\log \mu(x,t) = \alpha_x + \beta_x \kappa_t \text{ where } \kappa_t = \kappa_{t-1} + \theta + \omega_t \text{ is modelled as a RWD}$$

    - To avoid long-run divergence in mean mortality forecasts for a group using the LC method, it is a sufficient condition that all populations in the group have the same β_x and the same drift term θ.
  - The common factor model:

$$\log \mu_i(x,t) = \alpha_{x,i} + B_x K_t$$

    - B_x and K_t are common factors, determining the long-term trend
  - The augmented common factor model (Li-Lee model):

$$\log \mu_i(x,t) = \alpha_{x,i} + B_x K_t + \beta_{x,i} \kappa_{t,i}$$

    - $\beta_{x,i}$ and $\kappa_{t,i}$ are population-specific factors, controlling short-term deviations from the trend

Time dynamics

- The joint K_t-index is assumed to follow a random walk with drift

$$K_t = K_{t-1} + \theta_K + \sigma_K \omega_t, \text{ where } \omega_t \text{ is iid } N(0,1)$$

- For the model to be coherent, $\kappa_{t,i}$ must be forecasted as a stationary process
  - e.g. a random walk without drift or a first order autoregressive model
  - usually the latter is preferred:

$$\kappa_{t,i} = \phi_i \kappa_{t-1,i} + \theta_i + \sigma_i \omega_{t,i}, \text{ where } \omega_{t,i} \text{ is iid } N(0,1) \text{ and } |\phi_i| < 1$$

  - This accomodates short-term differences from the main trend…
  - … but this difference diminishes over time

Identification issues

Like its ancestor, the Lee-Li model suffers from an identifiability problem

- For any $c_1, c_2 \in \mathbb{R}$ and $d_1, d_2 \in \mathbb{R} \setminus \{0\}$, the model is invariant under the transformation

$$\left\{\alpha_{x,i}, B_x, K_t, \beta_{x,i}, \kappa_{t,i}\right\} \mapsto \left\{\alpha_{x,i} + d_1 B_x + d_2 \beta_{x,i}, \frac{B_x}{d_1}, d_1(K_t - c_1), \frac{\beta_{x,i}}{d_2}, d_2(\kappa_{t,i} - c_2)\right\}$$

  - K_t and $\kappa_{t,i}$ are only defined up to a linear transformation
  - B_x and $\beta_{x,i}$ up to a multiplicative constant
  - $\alpha_{x,i}$ up to an additive constant
- Note that the invariant transformations are the straightforward extensions of the LC ones
- Typical constraints are therefore

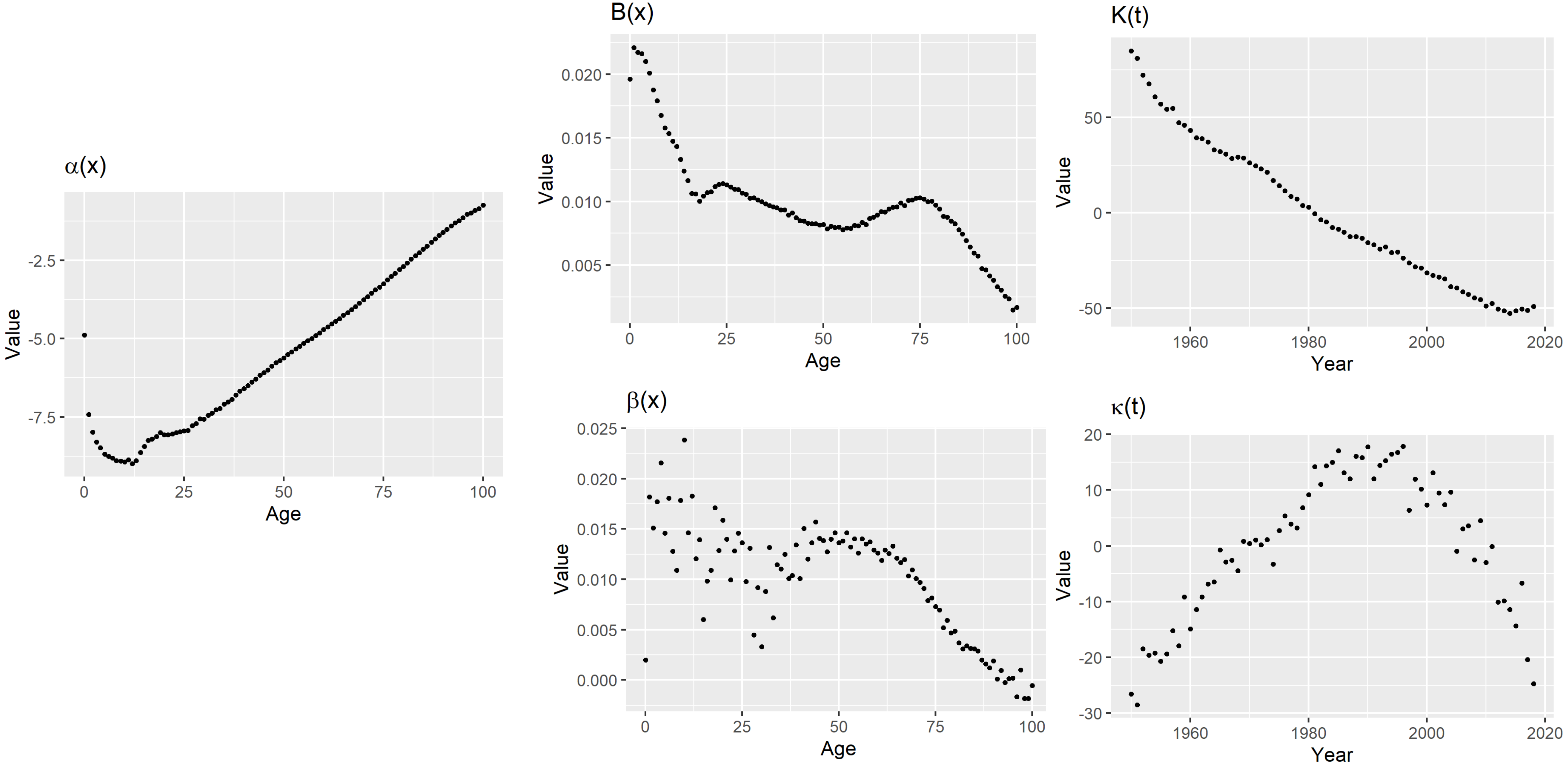
$$\sum_x B_x = 1, \sum_t K_t = 0, \sum_x \beta_{x,i} = 1, \sum_t \kappa_{t,i} = 0$$

Model fitting

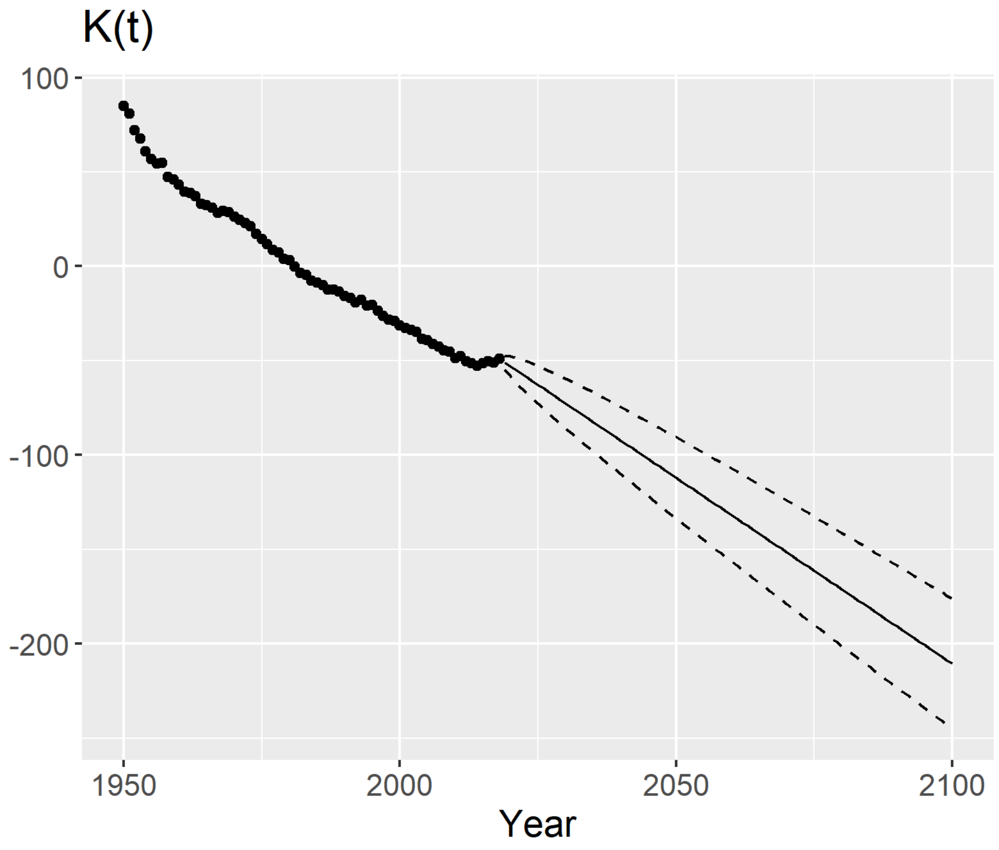
- Li and Lee (2005) propose a two-step SVD procedure to estimate model parameters
  1. Estimate B_x and K_t by applying the standard Lee-Carter SVD procedure to the pooled population dataset. Adjust K_t to fit the groups average life expectancy.
  2. Estimate $\hat{\alpha}_{x,i} = \frac{1}{T-1}\sum_t \log m_i(x,t)$ as the time-average of the log empirical death rates. Then use SVD on the residual matrix $\{\log m_i(x,t) - \hat{\alpha}_{x,i} - \hat{B}_x \hat{K}_t\}$ to determine $\beta_{x,i}$ and $\kappa_{t,i}$.

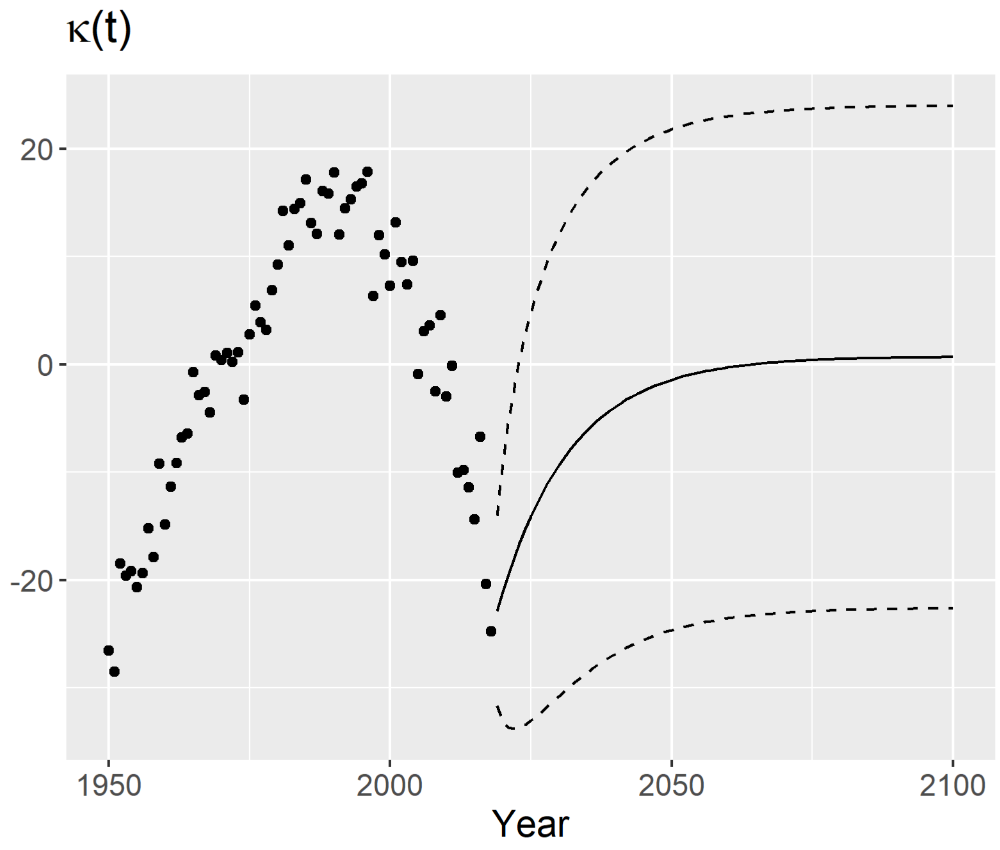
- Alternatively, we could use the Poisson assumption $D_i(x,t)|E_i(x,t) \sim \mathrm{Pois}\big(E_i(x,t)\mu_i(x,t)\big)$
  - Log-likelihood: $l = \sum_{x,t,i} D_i(x,t)\log\big(E_i(x,t)\mu_i(x,t)\big) - E_i(x,t)\mu_i(x,t) - \log(D_i(x,t)!)$
  - Derive iterative update scheme using Newton-Raphson: $\theta' = \theta - \frac{\partial l/\partial\theta}{\partial^2 l/\partial\theta^2}$
  - Adjust parameters following each iteration such that identification constraints are satisfied

Li-Lee model parameters example, Denmark (1/2)

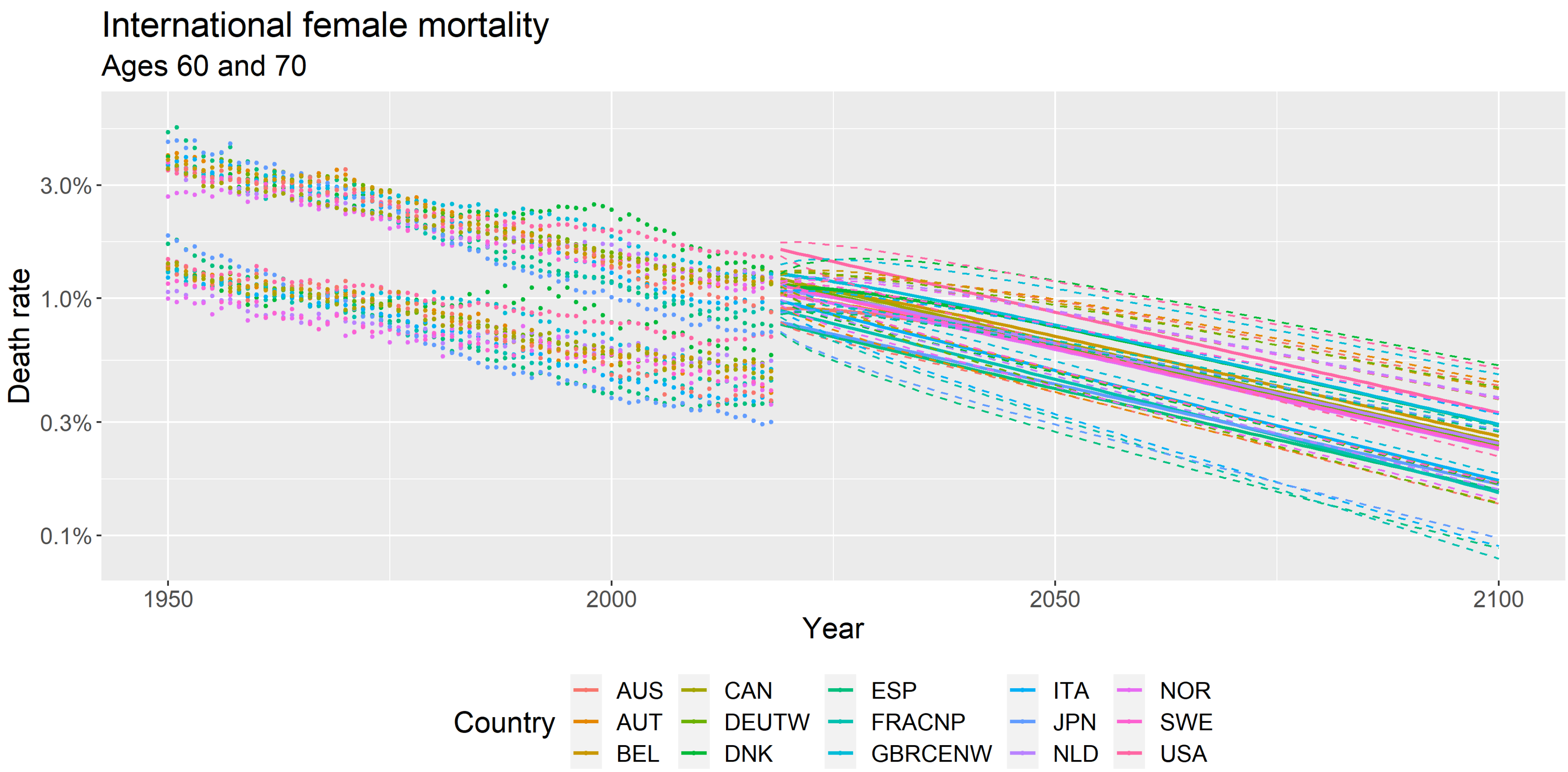


Li-Lee model parameters example, Denmark (2/2)





Li-Lee model coherent forecast example



#3

The SAINT methodology

Spread Adjusted InterNational Trend

- **Two-step projection method**

1. Model and project ***reference*** population mortality
2. Model and project the “distance” (spread) between ***target*** population mortality and the reference level of mortality

- Rationale: Easier to extract a long-term trend from a large dataset, than a small dataset, since idiosyncratic features are typically more pronounced in the latter.

- **Model for population i**

$$\log \mu_i(x,t) = \log \mu_{\mathrm{ref}}(x,t) + \log\left(y_{i,t}^{\top} r_x\right)$$

- The spread, $\log \mu_{\mathrm{i}}(x,t) - \log \mu_{\mathrm{ref}}(x,t)$, is parameterized by r_x
  - e.g. level, slope and curvature
  - for example, $r_x = (1, x - x_0, (x - x_0)^2)^{\top}$
- Multivariate, *stationary* time series model for $y_{i,t}^{\top}$ controlling length and magnitude of deviations (spread)

Example (1/4): Trend and spread model

- Let $\mu_{\mathrm{ref}}(x,t) = \exp(\alpha_x + \beta_x \kappa_t)$, i.e. Lee-Carter

- Use a three-dimensional VAR(1) model for $y_{i,t}^{\top}$

$$y_{i,t} = A_i y_{i,t-1} + c_i + \varepsilon_{i,t}, \qquad \varepsilon_{i,t} \sim N_3(0, \Omega_i)$$

- In this case, the forecast variance is

$$\mathrm{Var}(\log \mu_i(x, T+h)) = \underbrace{h \hat{\beta}_x \hat{\sigma}^2}_{\to \infty} + \underbrace{\mathrm{Var}\left(y_{i,T+h}^{\top} r_x\right)}_{\to k_i < \infty} \quad \text{for } \mathrm{h} \to \infty$$

- First term: Lee-Carter variance of the common long-term trend
- Second term: The (bounded) variance of the country-specific spread

Example (2/4): Spread parametrization

Country specific mortality

$$\mu_{\text{target}}(x,t) = \mu_{\text{ref}}(x,t)\exp(y_{0,t} + r_1(x)y_{1,t} + r_2(x)y_{2,t})$$

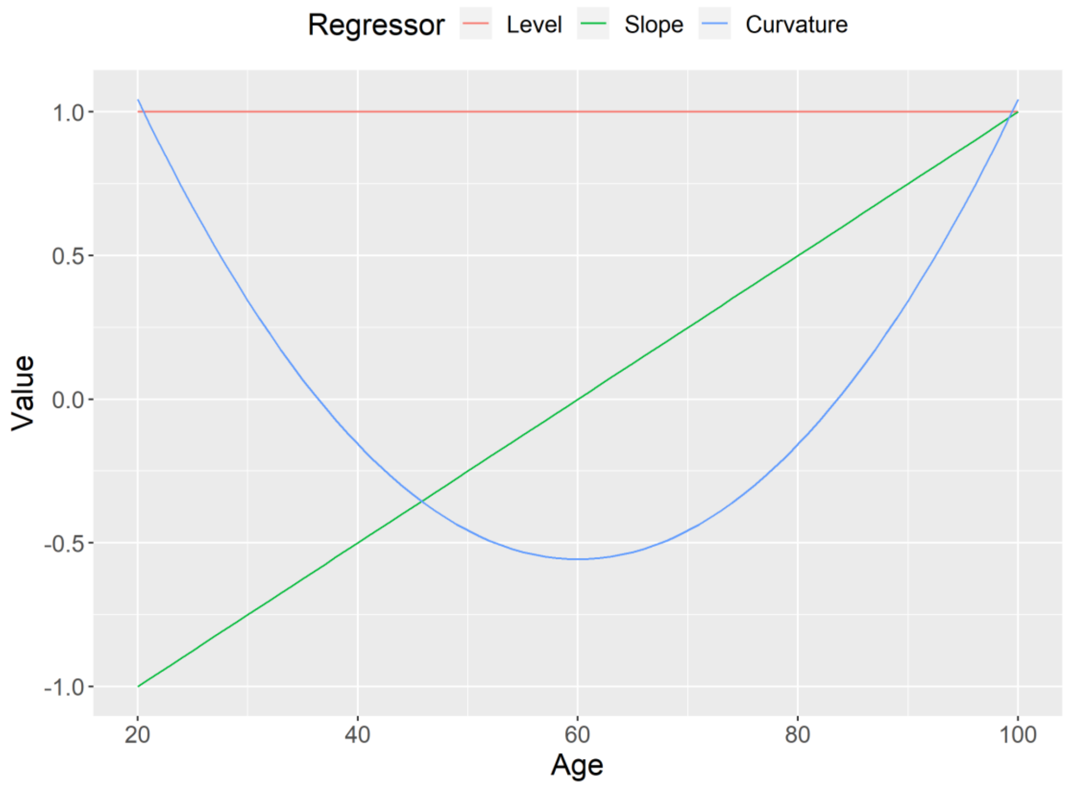
with

$$r_1(x) = (x-60)/40$$
$$r_2(x) = \frac{\left(x^2 - 120x + \frac{9130}{3}\right)}{1000}$$

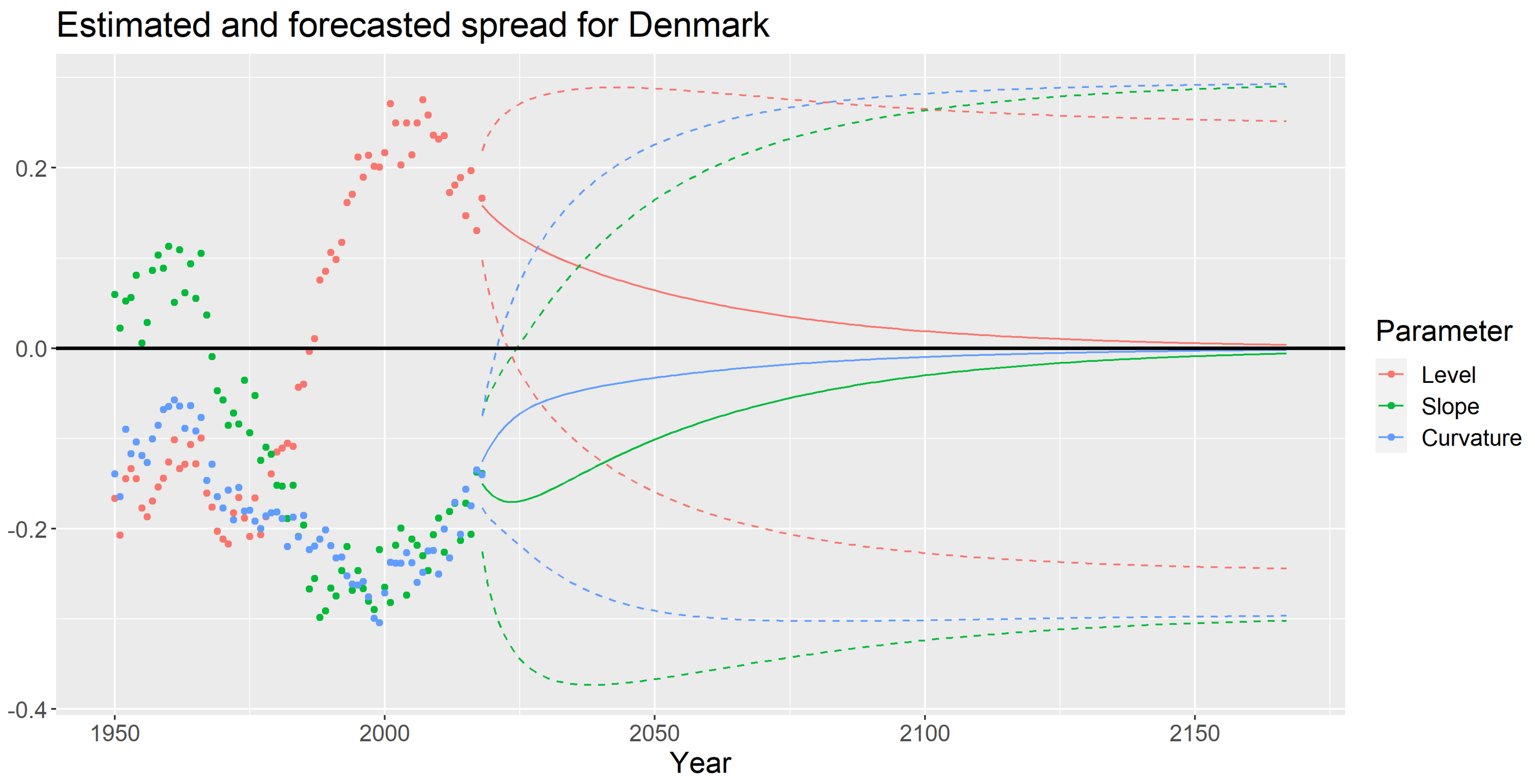
Mean zero, orthogonal regressors normalized to (about) 1 at age 20 and 100

$$\left(y_{0,t}, y_{1,t}, y_{2,t}\right)^{\top} = A\left(y_{0,t-1}, y_{1,t-1}, y_{2,t-1}\right)^{\top} + e_t, e_t \sim N_3(0, \Omega)$$

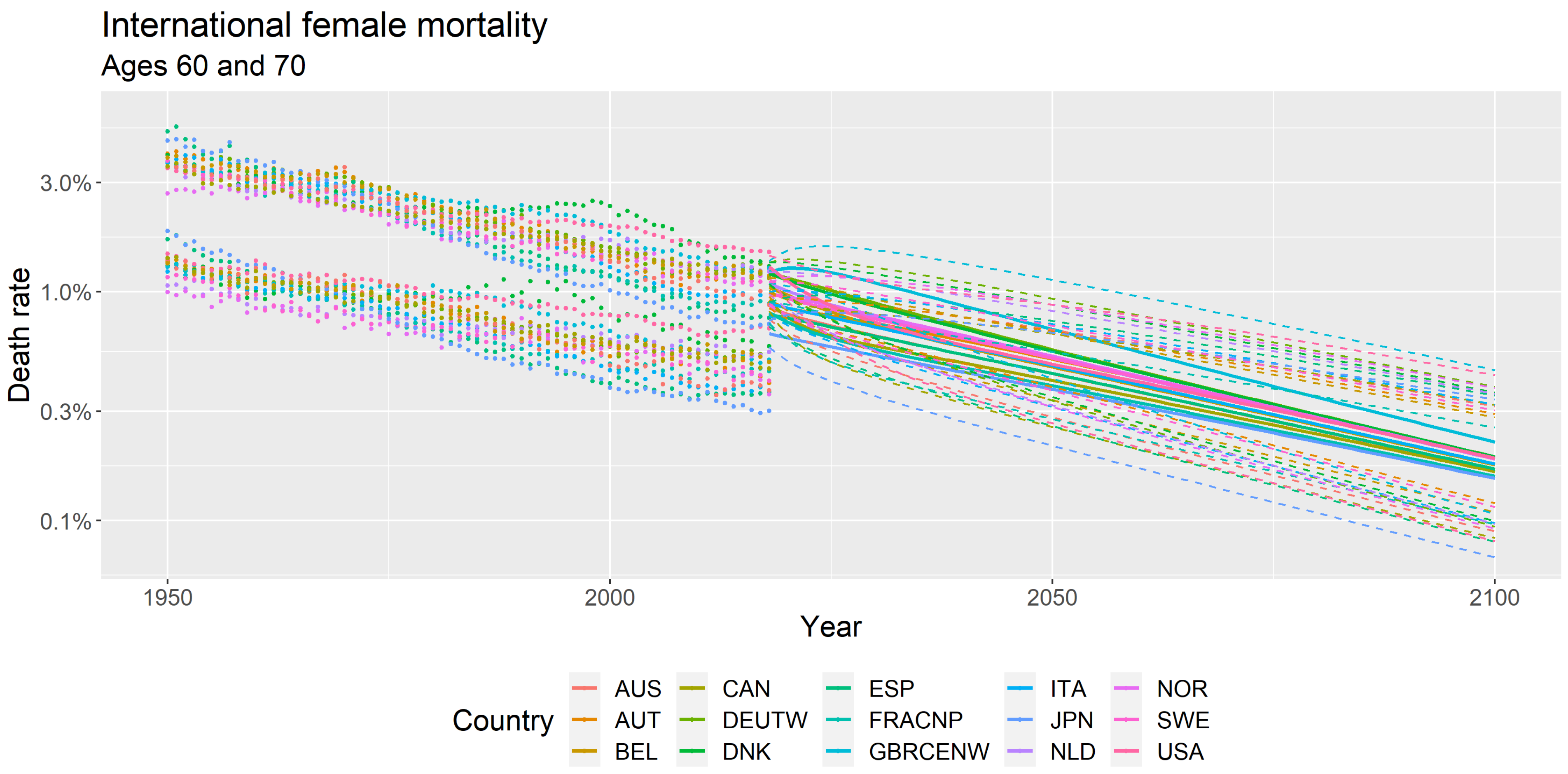
- The spread is *assumed* to fluctuate around zero
  - that is, no mean term included in the model
- The spread controls the length and magnitude of deviations



Example (3/4): Spread parameter forecast



Example (4/4): Coherent projections and levels of uncertainty



A short digression on the mortality benchmark of the Danish FSA (1/2)

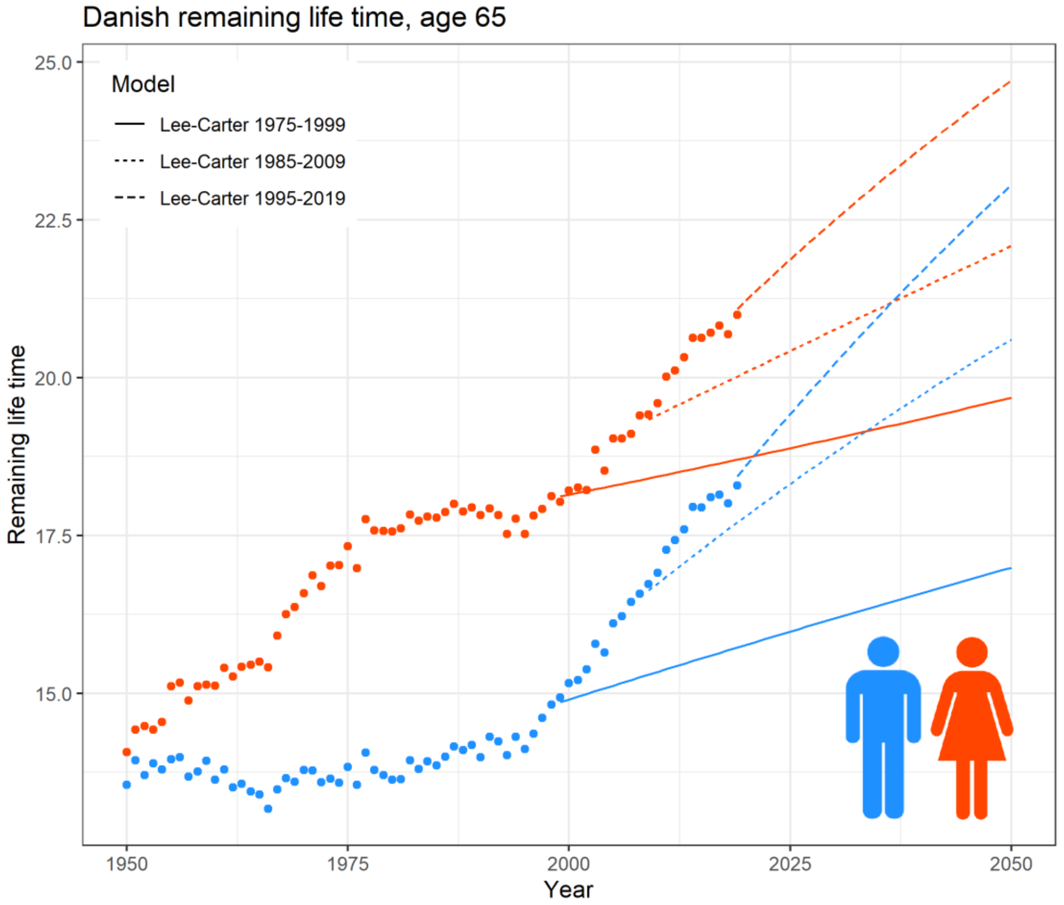
The longevity benchmark used by the Danish FSA is based on the SAINT projection methodology

- **Reference trend**
  - Based on pooled data provided by Danish life insurance companies and multi-employer pension funds
  - Reduction factor mortality model: $\mu_s(x, T+h) = \mu_s(x,T) \cdot \left(1 - R_s(x)\right)^h$ for $s \in \{\text{female}, \text{male}\}$
    - Each year, the FSA provides the current observed mortality level $\mu_s(x,T)$ and the reduction factor $R_s(x)$
    - Essentially, these quantities are determined by a Lee-Carter model
- **Company specific mortality**
  - Given the benchmark, each company estimates its own company-specific mortality relative to the benchmark

    $\mu_s^{\text{company}}(x,t) = \mu_s^{FSA}(x,t) \exp\left(y_i^\top r_x\right)$, where r_x contains 3 regressors specified by the FSA
  - The y-parameters are estimated by the company and subject to a significance test in which non-significant parameters are set to zero.
  - The estimation is performed each year and is based on the mortality experience of the company's portfolio over the last 5 years.

A short digression on the mortality benchmark of the Danish FSA (2/2)

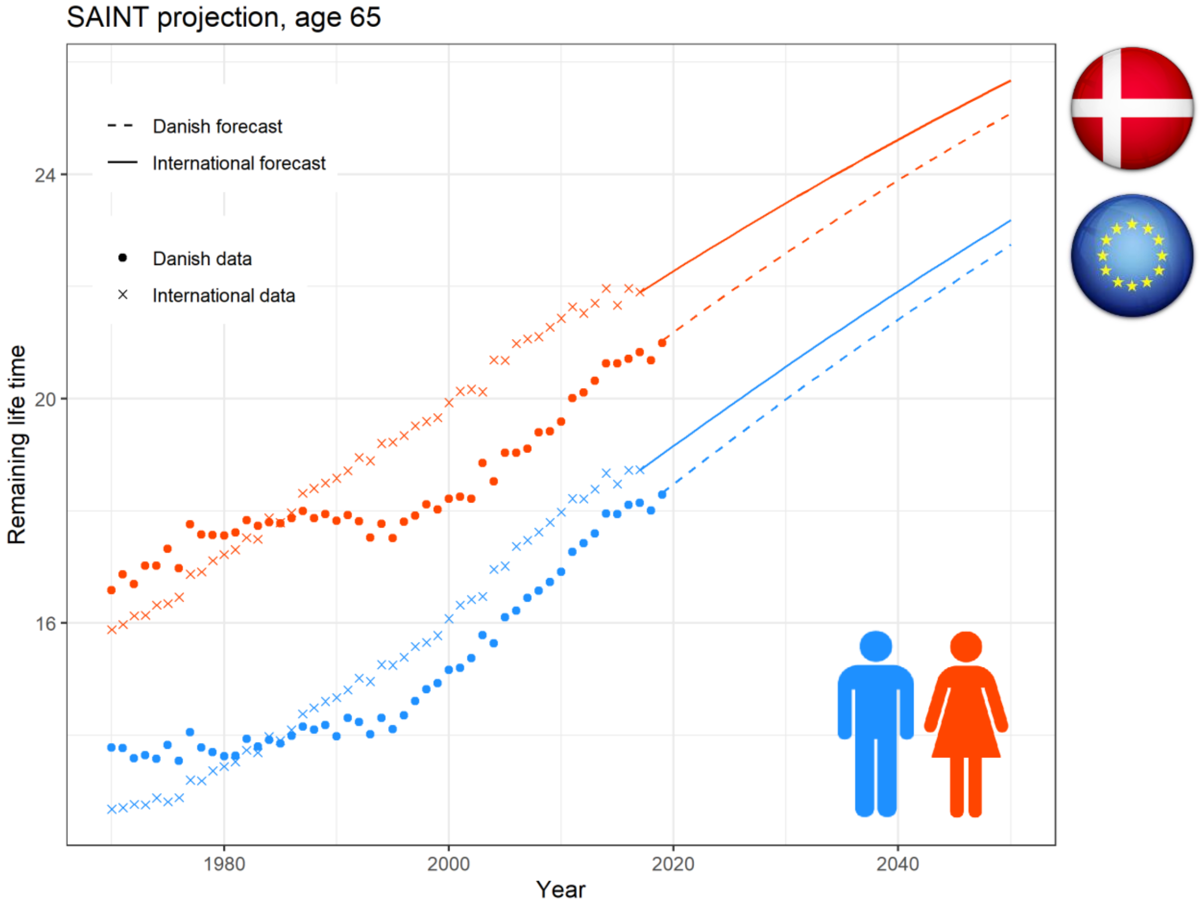
- **Danish FSA life expectancy benchmark**
  - **Based on the Lee & Carter (1992) model**
    - Log-linear projection of age-specific death rates
  - **The model preserves historical trends**
    - In situations with varying rates of improvements (e.g. Denmark), the projection is very sensitive to the window of estimation
  - **The figure shows three different projections**
    - Very different even though periods are overlapping
    - The FSA previously used the last 30 years for estimation, but now uses only the last 20 years of data to avoid the “kink” in 1995.
    - Shortening the estimation window solves the “update” problem, but seems problematic for long-term projections



SAINT model currently used at ATP

- **SAINT = Spread Adjusted InterNational Trend**
  - Forecasting of small populations with the aid of larger reference populations
  - ATP uses SAINT to forecast Danish/ATP mortality using Western European mortality as reference
  - *Based on the empirical "fact" that large populations are often much more stable than small populations*
- **Projection of Danish/ATP mortality in two steps**

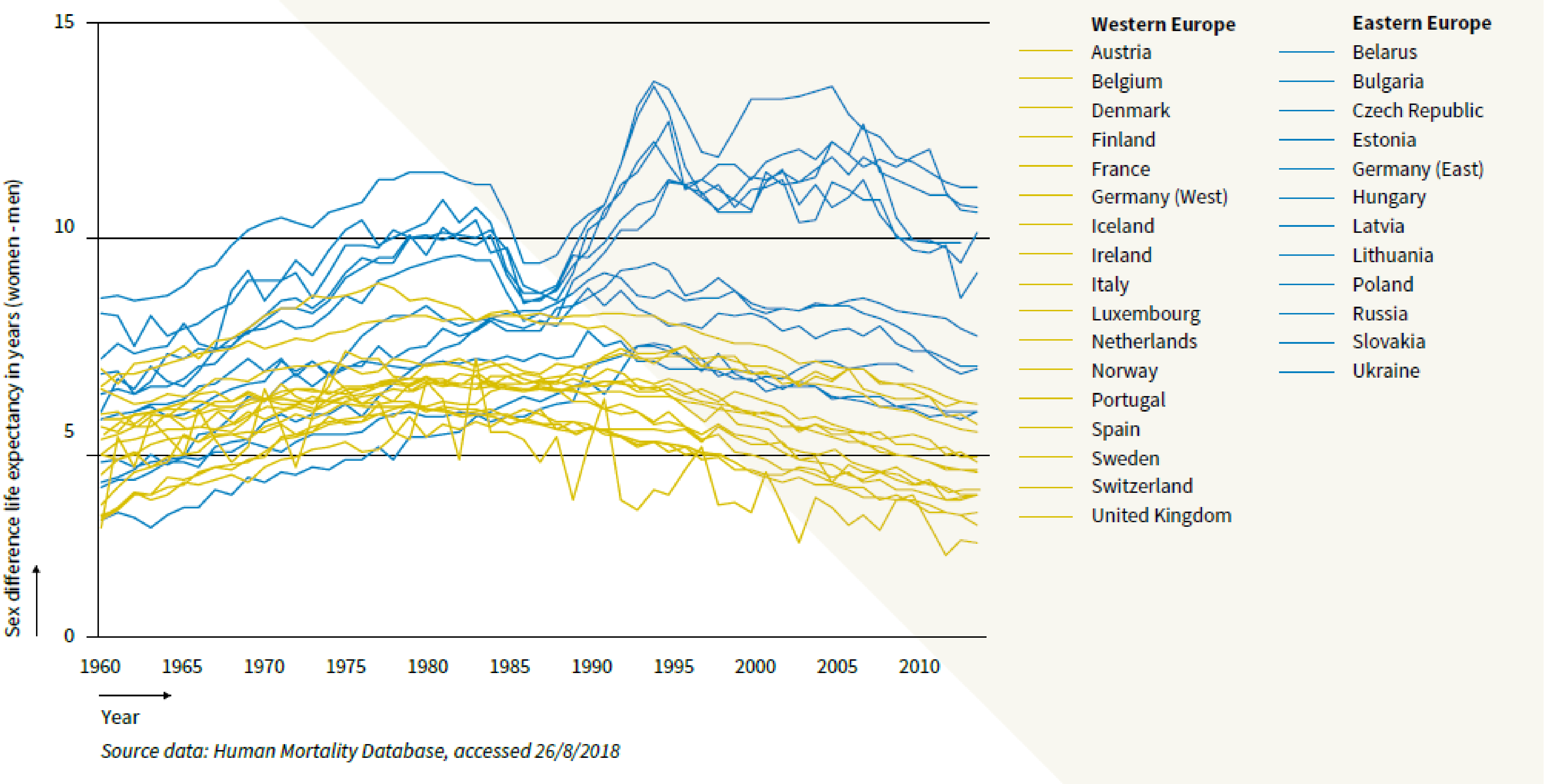
1. **Project international mortality based on a "long" estimation period**
   - Large international dataset consisting of 18 countries, primarily Western Europe
   - *Frailty*-component gives rise to increasing rates of improvements in old-age mortality
   - Gender gap preserved through a simultaneous projection of both sexes using *cointegration* techniques
2. **Project the spread between Denmark/ATP and the international trend**
   - Denmark is assumed to converge on the international trend



#4

Modelling the gender gap

Common gender gap pattern in Western Europe



Reproduced from Janssen (2018)

Preserving gender gap by cointegration

Model parameters by an error-correction model on the form

$$\begin{pmatrix}\Delta X_t^f \\ \Delta X_t^m\end{pmatrix} = \begin{pmatrix}a_1 \\ a_2\end{pmatrix}\left(X_{t-1}^f - X_{t-1}^m\right) + \begin{pmatrix}b_1 \\ b_2\end{pmatrix} + \varepsilon_t$$

$$= \begin{pmatrix}a_1 \\ a_2\end{pmatrix}\left(D_{t-1} - b_D\right) + \begin{pmatrix}b_\Delta \\ b_\Delta\end{pmatrix} + \varepsilon_t$$

where X is the parameter of interest

$D_t = X_t^f - X_t^m$ is the difference at time t

b_D = stationary difference between parameters

b_Δ = common slope of the two parameters

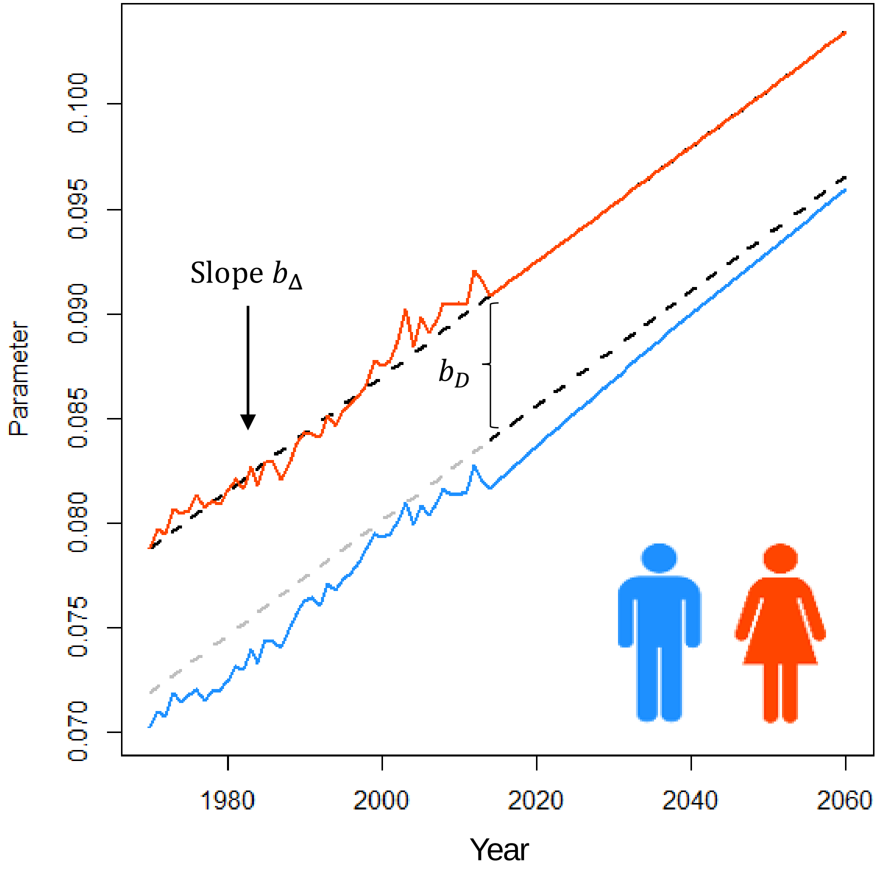


Figure: Example for Gompertz slope

- Historic slope for females continues ($a_1 = 0$)
- Stationary difference reduced to 75% of jump-off value
- Males approaching females with half-life of 25 years

Example: Simulated cohort life expectancy of a 60 year old