Lecture 2 - Part 2: The Lee-Carter model

Giancarlo Camarda

Institut national d'études démographiques

Ugofilippo Basellini

Max Planck Institute for Demographic Research

IDEM 117

Advances in Mortality Forecasting

International Advanced Studies in Demography

17 – 21 January, 2022

The Lee-Carter model (1992)

- Proposed in 1992 to model and forecast US mortality
- After almost 30 years, Lee-Carter (LC) still widely employed by variety of users: governments, private companies, international organizations, ...
- The landmark model in mortality forecasting
- One of the firstly introduced *stochastic* mortality models
- An **extrapolation** method:
  - model the mortality surface over age and time
  - extrapolate trends in the future, assuming that observed trends will continue
- Simplicity, robustness and objectivity have made the model so successful
- Nonetheless, some limitations of the model have stimulated several extensions over the years

The LC model

- A simple log-bilinear functional form for mortality rates $m_{x,t}$ at age x and time t

$$\ln(m_{x,t}) = \alpha_x + \beta_x \kappa_t + \epsilon_{x,t} \quad (1)$$

where:

  - α_x is the general shape of log-mortality at age x
  - β_x is the rate of mortality improvement at age x
  - κ_t is the general level of mortality at time t
  - $\epsilon_{x,t}$ is the error term with mean 0 and variance σ^2_ϵ, reflecting residual age-specific influences not captured by the model

- The model is undetermined: if $\boldsymbol{\theta}_1 = [\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\kappa}]$ is a solution, then for any scalar c:
  - $\boldsymbol{\theta}_2 = [\boldsymbol{\alpha} - \boldsymbol{\beta} c, \boldsymbol{\beta}, \boldsymbol{\kappa} + c]$ is also a solution
  - $\boldsymbol{\theta}_3 = [\boldsymbol{\alpha}, \boldsymbol{\beta} c, \boldsymbol{\kappa} / c]$ is also a solution

- Two constraints introduced to ensure model identification:

$$\sum_x \beta_x = 1 \quad \text{and} \quad \sum_t \kappa_t = 0 \quad (2)$$

The LC model: a schematic view

$$\ln\left(m_{x,t}\right) \simeq \alpha_x + \beta_x \kappa_t$$

$$\begin{pmatrix} \ln(m_{0,1960}) & \ln(m_{0,1961}) & \dots & \ln(m_{0,2018}) \\ \ln(m_{1,1960}) & \ln(m_{1,1961}) & \dots & \ln(m_{1,2018}) \\ \ln(m_{2,1960}) & \ln(m_{2,1961}) & \dots & \ln(m_{2,2018}) \\ \vdots & \vdots & \ddots & \vdots \\ \ln(m_{105,1960}) & \ln(m_{105,1961}) & \dots & \ln(m_{105,2018}) \end{pmatrix} \simeq$$

$$\begin{pmatrix} \alpha_0 \\ \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_{105} \end{pmatrix} + \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_{105} \end{pmatrix} \begin{pmatrix} \kappa_{1960} & \kappa_{1961} & \dots & \kappa_{2018} \end{pmatrix}$$

$$\underbrace{59}_{\text{years}} \times \underbrace{106}_{\text{ages}} = \underbrace{6254}_{\text{cells}} \simeq \underbrace{106}_{\alpha_i} + \underbrace{106}_{\beta_i} + \underbrace{59}_{\kappa_j} - \underbrace{2}_{\text{constraints}} = \underbrace{269}_{\text{parameters}}$$

Model estimation

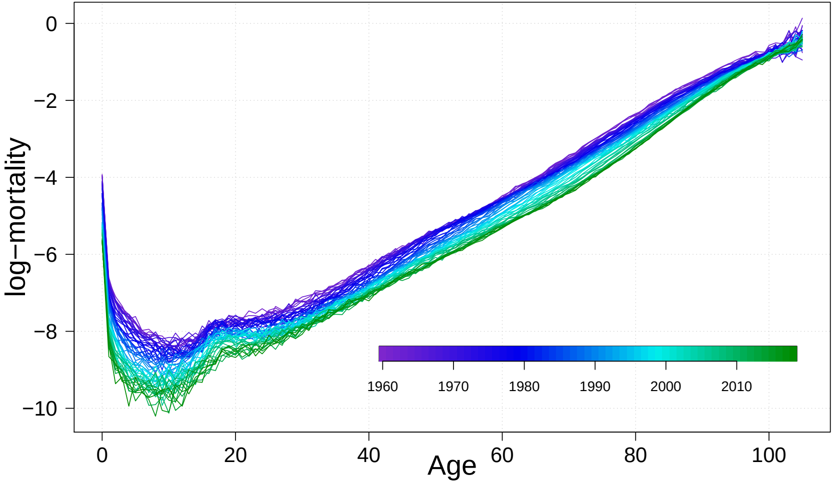
- The model is estimated by minimizing the residual sum of squares:

$$\sum_{x,t}\Big(\ln\left(m_{x,t}\right)-\alpha_x-\beta_x\kappa_t\Big)^2 \tag{3}$$

- A singular value decomposition (SVD) is employed to minimize Eq. (3):
  - $\hat{\alpha}_x$ is the average of the observed $\ln\left(m_{x,t}\right)$

Estimating LC: an example

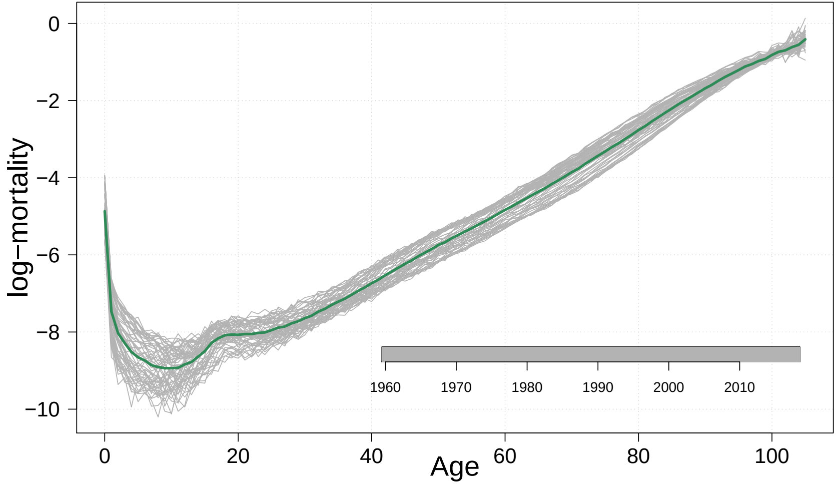
- observed mortality rates $\ln(m_{x,t})$



Females aged 0–105+ in England & Wales, 1960–2018.
Source (all figures): Human Mortality Database (2021)

Estimating LC: an example

- $\hat{\alpha}_x$ = average of observed mortality rates $\ln(m_{x,t})$



Females aged 0–105+ in England & Wales, 1960–2018.
Source (all figures): Human Mortality Database (2021)

Model estimation

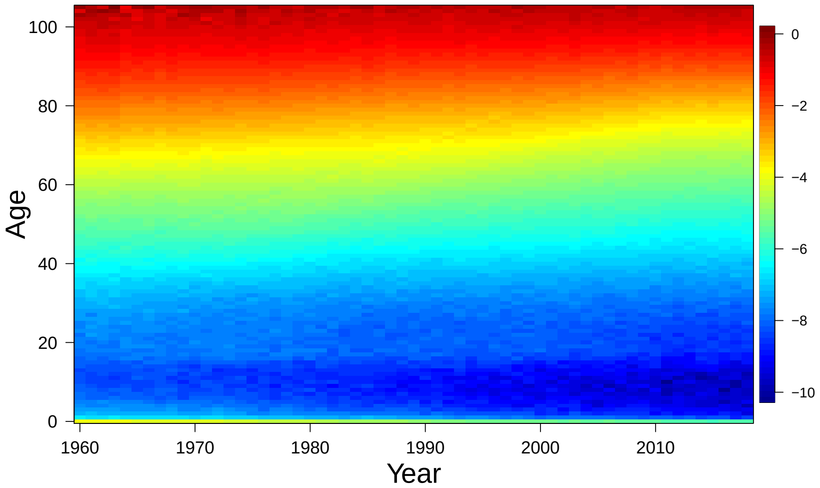
- The model is estimated by minimizing the residual sum of squares:

$$\sum_{x,t}\Big(\ln\left(m_{x,t}\right)-\alpha_x-\beta_x\kappa_t\Big)^2$$

- A singular value decomposition (SVD) is employed to minimize Eq. (3):
  - $\hat{\alpha}_x$ is the average of the observed $\ln\left(m_{x,t}\right)$
  - $\hat{\beta}_x$ and $\hat{\kappa}_t$ are the first left- and right-singular vectors of the SVD of the matrix $\ln\left(m_{x,t}\right)-\hat{\alpha}_x$

Estimating LC: an example

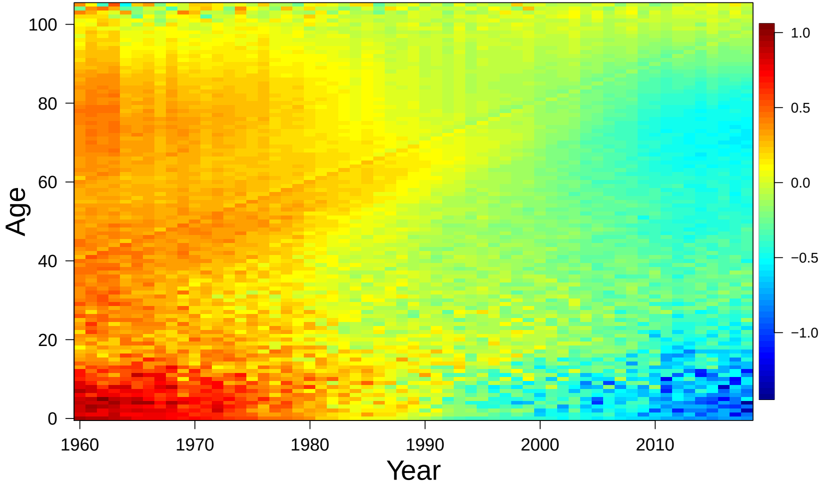
- $\boldsymbol{M} = (\ln(m_{x,t}))$: matrix of observed mortality rates



Females aged 0–105+ in England & Wales, 1960–2018.

Estimating LC: an example

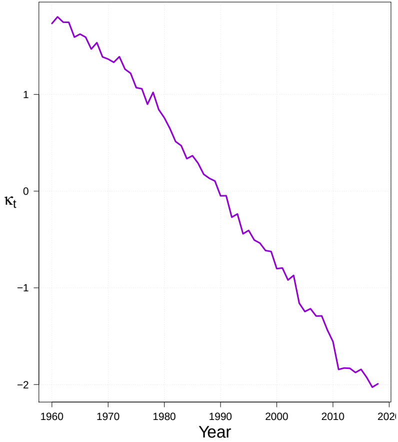
- $\tilde{M} = (\ln(m_{x,t}) - \hat{\alpha}_x)$: matrix of "centered" mortality rates



Females aged 0–105+ in England & Wales, 1960–2018.

Estimating LC: an example

- From SVD of $\tilde{M}$, and using the constraints in Eq.(2), we get $\hat{\beta}_x$ and $\hat{\kappa}_t$



Females aged 0–105+ in England & Wales, 1960–2018.

Model estimation

- The model is estimated by minimizing the residual sum of squares:

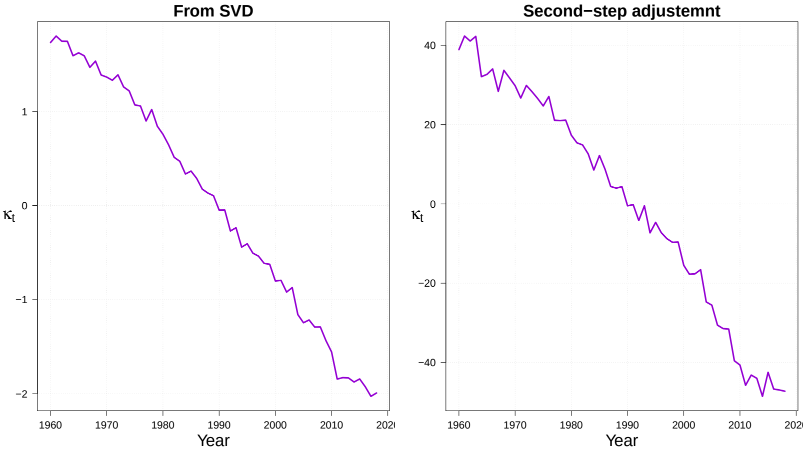
$$\sum_{x,t}\Big(\ln\left(m_{x,t}\right)-\alpha_x-\beta_x\kappa_t\Big)^2$$

- A singular value decomposition (SVD) is employed to minimize Eq. (3):
  - $\hat{\alpha}_x$ is the average of the observed $\ln\left(m_{x,t}\right)$
  - $\hat{\beta}_x$ and $\hat{\kappa}_t$ are the first left- and right-singular vectors of the SVD of the matrix $\ln\left(m_{x,t}\right)-\hat{\alpha}_x$
- In a second-step estimation, the parameter $\hat{\kappa}_t$ is adjusted so that the fitted deaths match the observed deaths in all years, i.e.

$$\sum_x \hat{y}_{x,t}=\sum_x y_{x,t} \quad \text{for all } t$$

Estimating LC: an example

- Adjusting $\hat{\kappa}_t$ to match observed number of deaths at each year t



Females aged 0–105+ in England & Wales, 1960–2018.

Forecasting with LC

- Forecasting "made simple": choose an appropriate time-series model for $\hat{\kappa}_t$ and extrapolate it
- The forecast $\hat{\kappa}_{T+h}$ allows one to derive the entire age-pattern of mortality at time $T+h$, i.e.:

$$\ln\left(\hat{m}_{x,T+h}\right) = \hat{\alpha}_x + \hat{\beta}_x \hat{\kappa}_{T+h}$$

- LC suggest a random walk model (i.e. ARIMA(0,1,0)) with drift:

$$\kappa_t = \kappa_{t-1} + c + e_t$$

where c is a constant (drift) and e_t the error term (purely random process)

- From our time-series lecture: $\hat{\kappa}_{T+h|T} = \hat{\kappa}_T + c\,h$
- Prediction intervals for $\hat{m}_{x,T+h}$ and summary measure $\hat{e}_{0,T+h}$ in period life tables from those of $\hat{\kappa}_{T+h}$ (Lee and Miller 2001)

Forecasting with LC: a schematic view

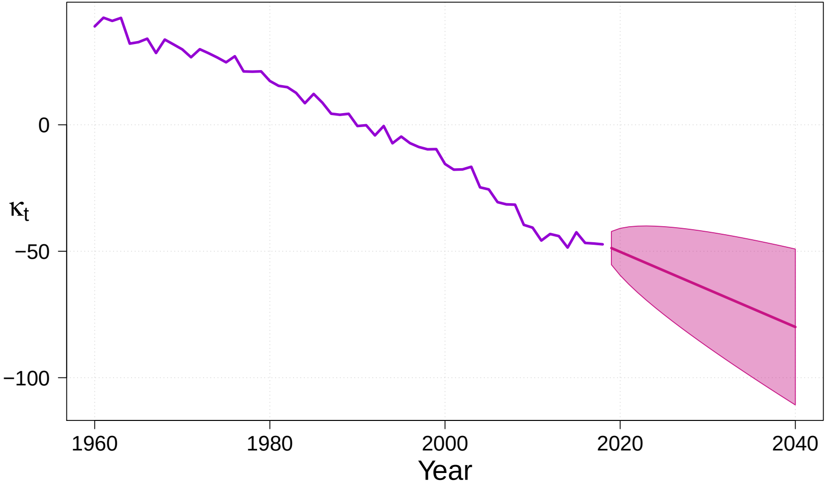
$$\ln(m_{x,t}) \simeq \alpha_x + \beta_x \kappa_t \qquad \ln(m_{x,T+h}) \simeq \alpha_x + \beta_x \kappa_{T+h}$$

$$\begin{pmatrix} \ln(m_{0,1960}) & \dots & \ln(m_{0,2018}) & \ln(m_{0,2019}) & \dots & \ln(m_{0,2040}) \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \ln(m_{105,1960}) & \dots & \ln(m_{105,2018}) & \ln(m_{105,2019}) & \dots & \ln(m_{105,2040}) \end{pmatrix} \simeq$$

$$\begin{pmatrix} \alpha_0 \\ \vdots \\ \alpha_{105} \end{pmatrix} + \begin{pmatrix} \beta_0 \\ \vdots \\ \beta_{105} \end{pmatrix} \begin{pmatrix} \kappa_{1960} & \dots & \kappa_{2018} & \kappa_{2019} & \dots & \kappa_{2040} \end{pmatrix}$$

Forecasting LC: an example

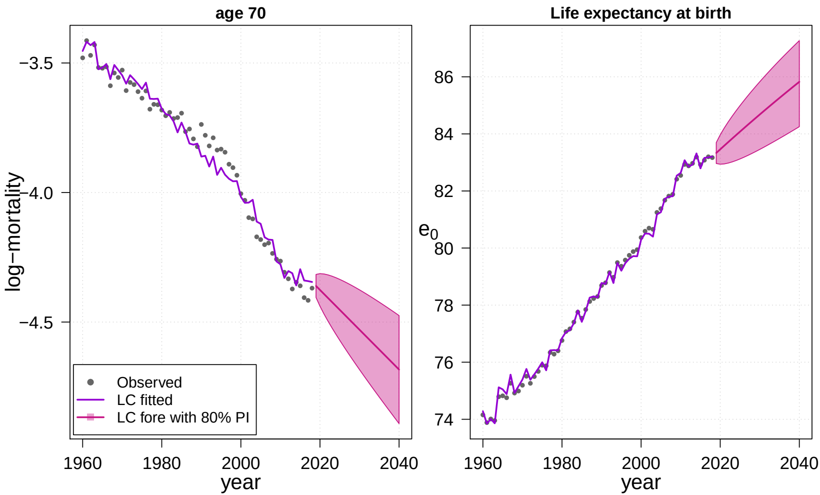
- Forecasting $\hat{\kappa}_t$ using a random walk with drift



Females aged 0–105+ in England & Wales, 1960–2018, forecast 2019–2040.

Forecasting LC: an example

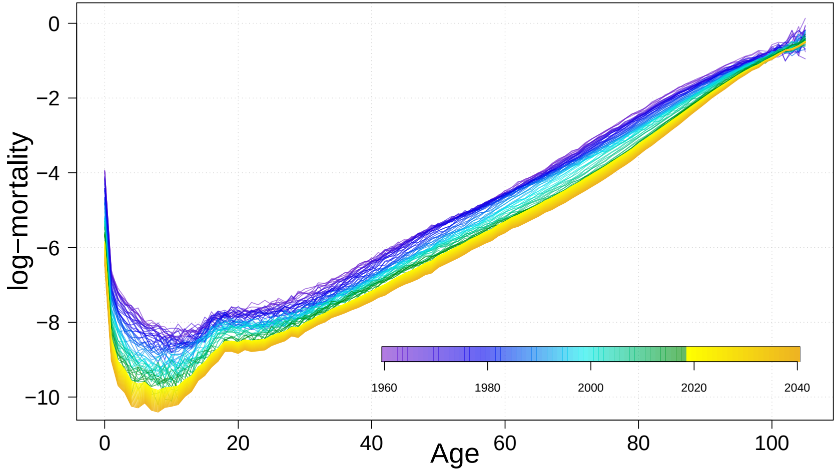
- Prediction intervals derived from those of $\hat{\kappa}_{T+h}$



Females aged 0–105+ in England & Wales, 1960–2018, forecast 2019–2040.

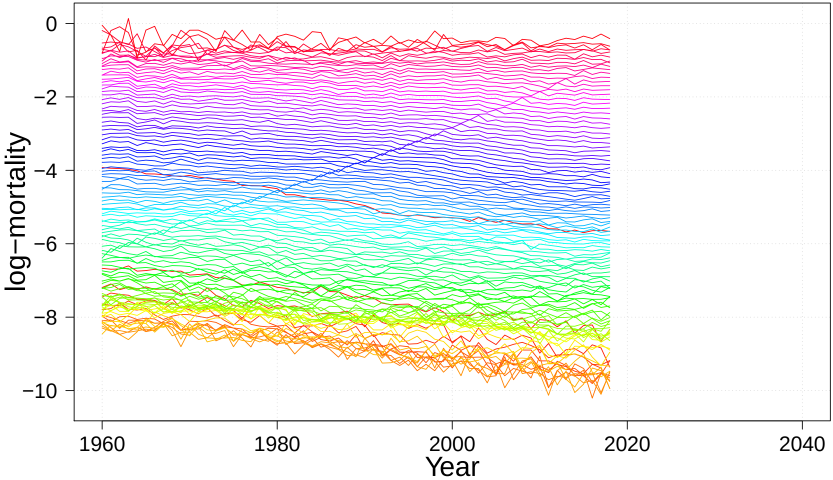
Forecasting LC: an example

- $\hat{\kappa}_{T+h}$ allows one to derive forecast rates $\hat{\boldsymbol{M}} = \left(\ln\left(\hat{m}_{x,T+h}\right)\right)$



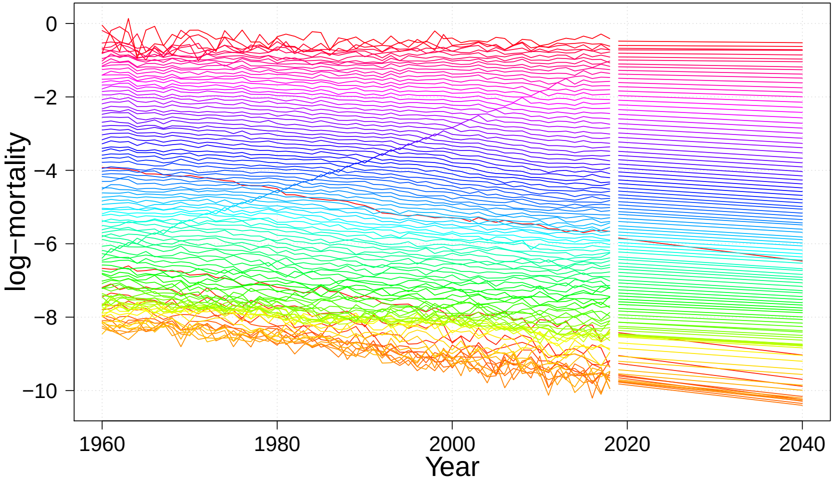
Females aged 0–105+ in England & Wales, 1960–2018, forecast 2019–2040.

Forecasting LC: an example



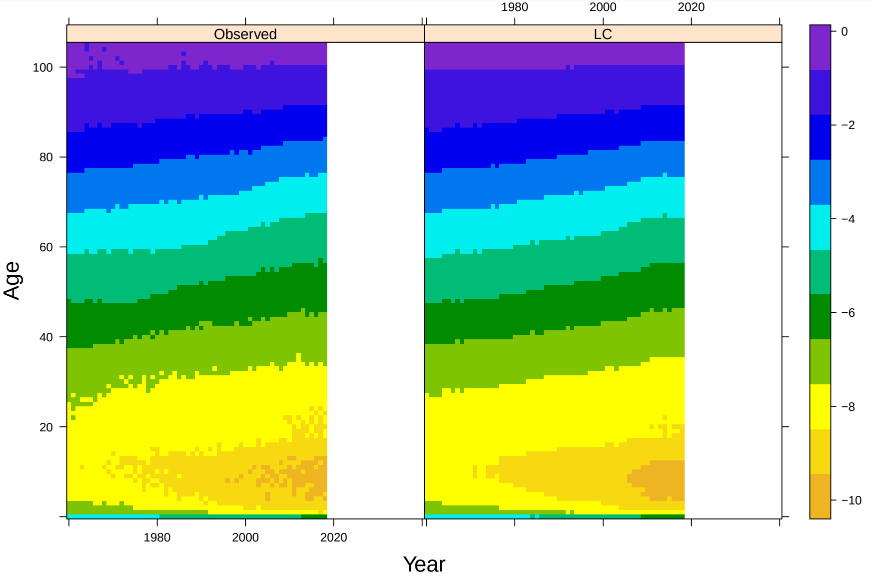
Females aged 0–105+ in England & Wales, 1960–2018, forecast 2019–2040.

Forecasting LC: an example



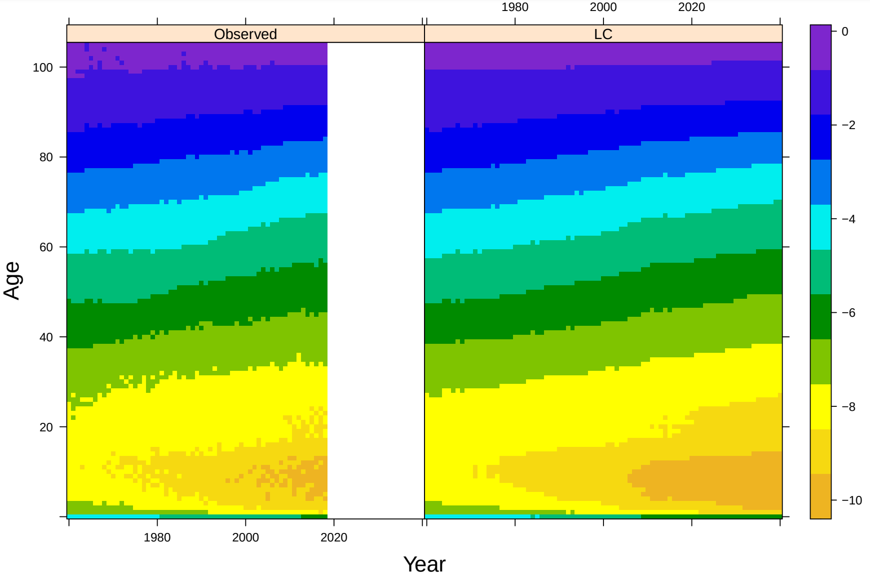
Females aged 0–105+ in England & Wales, 1960–2018, forecast 2019–2040.

Forecasting LC: an example



Females aged 0–105+ in England & Wales, 1960–2018, forecast 2019–2040.

Forecasting LC: an example



Females aged 0–105+ in England & Wales, 1960–2018, forecast 2019–2040.

The LC model: a summary

Advantages:

- simple functional form
- univariate time index condenses mortality development $\Rightarrow$ forecasting made "simple"
- stochastic model ($\Rightarrow$ probabilistic intervals), no expert opinions
- more accurate than previous methodologies

Disadvantages:

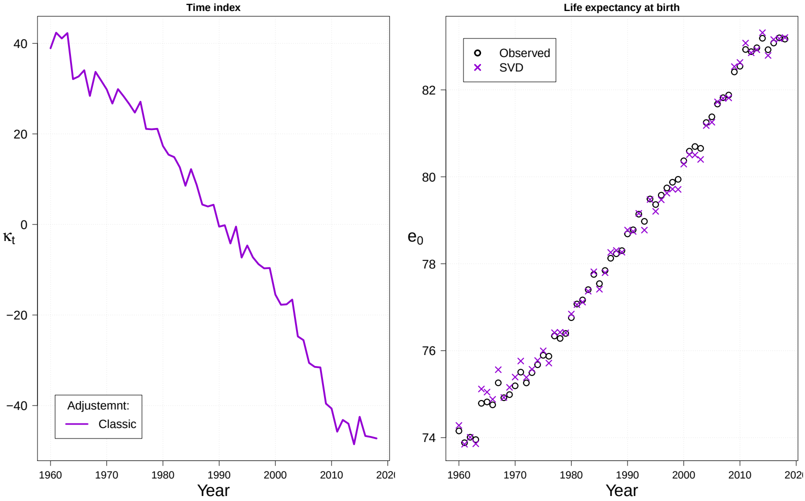
- "jump-off" bias
- Normality assumption (from SVD)
- jagged fitted and forecast age profile, lacking smoothness
- fixed age-pattern of mortality decline
- rigid structure

$\Rightarrow$ several extensions proposed to overcome some of these issues

Adjusting for the jump-off bias

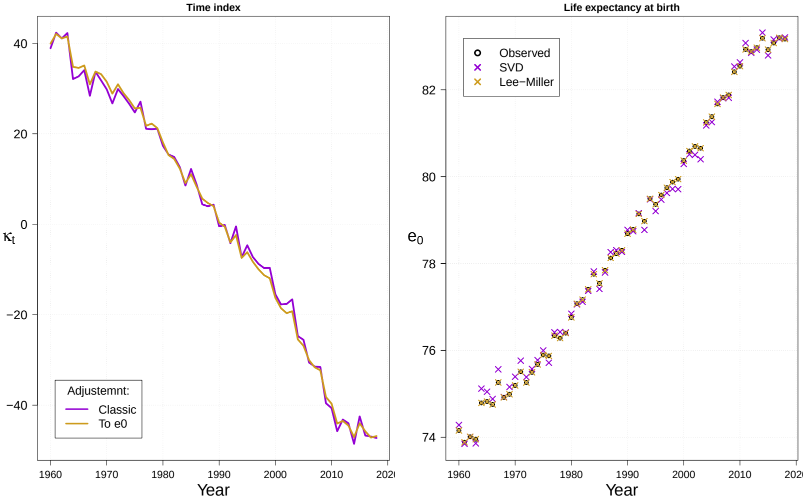
- The LC model does not fit perfectly the data in the jump-off year
- When discrepancy between observed and fitted e_0 is large, forecasts will be biased since the first year
- Two possible solutions identified by Lee and Miller (2001):
  - Set $\alpha_x = \ln(m_{x,T})$, where T is the last observed year
    This will however extrapolate idiosyncratic features of mortality in the jump-off year
  - Perform the second-step adjustment of κ_t to match the observed value of e_0 in each year t

Correcting the e_0 bias: an example



Females aged 0–105+ in England & Wales, years 1960–2018.

Correcting the e_0 bias: an example



Females aged 0–105+ in England & Wales, years 1960–2018.

The LC model: a summary

Advantages:

- simple functional form
- univariate time index condenses mortality development $\Rightarrow$ forecasting made "simple"
- stochastic model ($\Rightarrow$ probabilistic intervals), no expert opinions
- more accurate than previous methodologies

Disadvantages:

- ~~"jump-off" bias~~
- Normality assumption (from SVD)
- jagged fitted and forecast age profile, lacking smoothness
- fixed age-pattern of mortality decline
- rigid structure

The Poisson LC model

- The least-squares estimation via SVD assumes that errors $\epsilon_{x,t}$ are homoskedastic (with equal variance) and normally distributed
- Unreasonable assumption: $\ln(m_{x,t})$ is more variable at older than younger ages (due to a smaller number of deaths)

  $\Rightarrow$ Brouhns et al. (2002) embedded LC in a Poisson framework

- Data: deaths and exposures over ages x and years t, i.e. $y_{x,t}$ and $e_{x,t}$
- Assumption: $Y_{x,t} \sim \mathcal{P}(e_{x,t}\,\mu_{x,t})$, with $\mu_{x,t} = \mu_{x,t}^{LC} = \exp(\alpha_x + \beta_x \kappa_t)$, i.e. the force of mortality is assumed to have the log-bilinear form of the LC model (with parameters subject to same constraints in Eq.(2))
- Estimation: $\boldsymbol{\theta} = [\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\kappa}]$ derived by maximizing the (log-)likelihood:

$$\ln \mathcal{L}(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\kappa} | \boldsymbol{Y}, \boldsymbol{E}) \propto \sum_{x,t} [y_{x,t} \ln(\mu_{x,t}) - e_{x,t}\mu_{x,t}] \qquad (4)$$

Poisson LC: estimation

- The log-likelihood can be maximized by Newton-Raphson method:
  1. begin from some starting values
  2. at each iteration step $\nu + 1$, a single set of parameters is updated keeping the other two fixed with the updating scheme:

$$\hat{\theta}^{(\nu+1)} = \hat{\theta}^{(\nu)} - \frac{\partial \mathcal{L}^{(\nu)} / \partial \theta}{\partial^2 \mathcal{L}^{(\nu)} / \partial \theta^2}$$

  3. stop when the iterations result in a tiny difference in the parameters or log-likelihood (e.g. 10^{-6})

Poisson LC: estimation

- Set some starting values, e.g. $\hat{\alpha}_x^{(0)} = \sum_t \ln(m_{x,t})/T$, $\hat{\beta}_x^{(0)} = 1$ and $\hat{\kappa}_t^{(0)} = 0$
- Update the parameters using the formulas:

$$\hat{\boldsymbol{\alpha}}^{(\nu+1)} = \hat{\boldsymbol{\alpha}}^{(\nu)} - \frac{\sum_t \left(y_{x,t} - \hat{y}_{x,t}^{(\nu)}\right)}{-\sum_t \hat{y}_{x,t}^{(\nu)}}$$

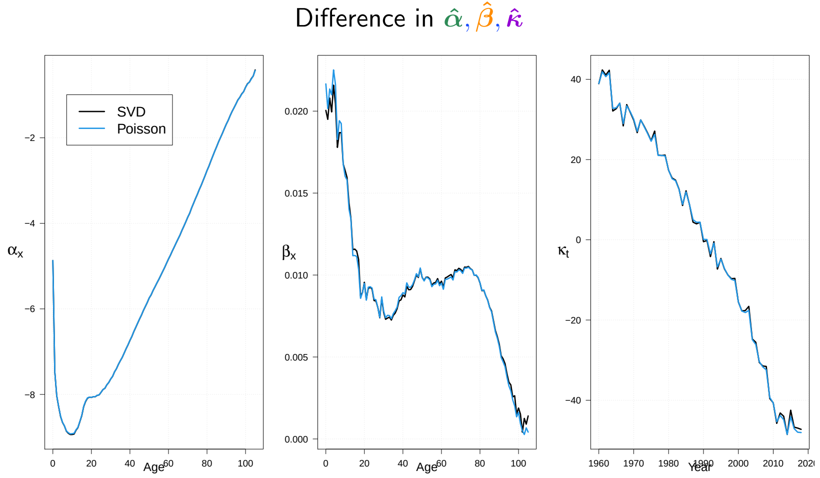
$$\hat{\boldsymbol{\kappa}}^{(\nu+2)} = \hat{\boldsymbol{\kappa}}^{(\nu+1)} - \frac{\sum_x \left(y_{x,t} - \hat{y}_{x,t}^{(\nu+1)}\right) \hat{\boldsymbol{\beta}}^{(\nu+1)}}{-\sum_x \hat{y}_{x,t}^{(\nu+1)} \left(\hat{\boldsymbol{\beta}}^{(\nu+1)}\right)^2}$$

$$\hat{\boldsymbol{\beta}}^{(\nu+3)} = \hat{\boldsymbol{\beta}}^{(\nu+2)} - \frac{\sum_t \left(y_{x,t} - \hat{y}_{x,t}^{(\nu+2)}\right) \hat{\boldsymbol{\kappa}}^{(\nu+2)}}{-\sum_t \hat{y}_{x,t}^{(\nu+2)} \left(\hat{\boldsymbol{\kappa}}^{(\nu+2)}\right)^2}$$

where $\hat{y}_{x,t}^{(\nu)} = e_{x,t} \exp\left(\hat{\alpha}_x^{(\nu)} + \hat{\beta}_x^{(\nu)} \hat{\kappa}_t^{(\nu)}\right)$

- Reach convergence in the iteration process
- Set the constraints $\sum_x \beta_x = 1$, $\sum_t \kappa_t = 0$
- No need of second step adjustment!!

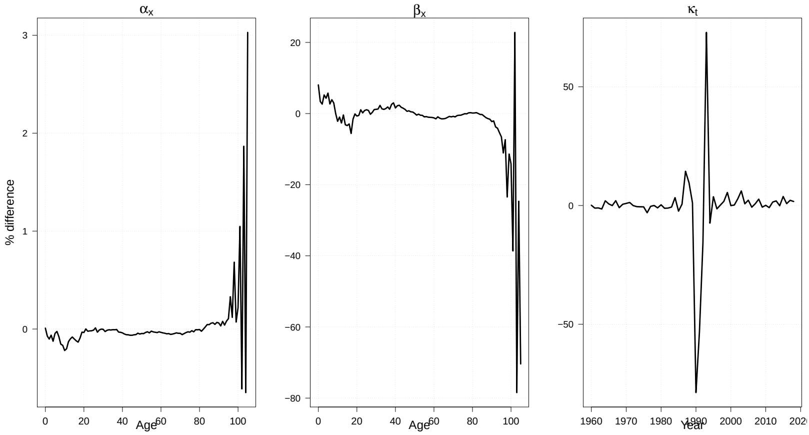
LC SVD vs Poisson: an example



Females aged 0–105+ in England & Wales, years 1960–2018.

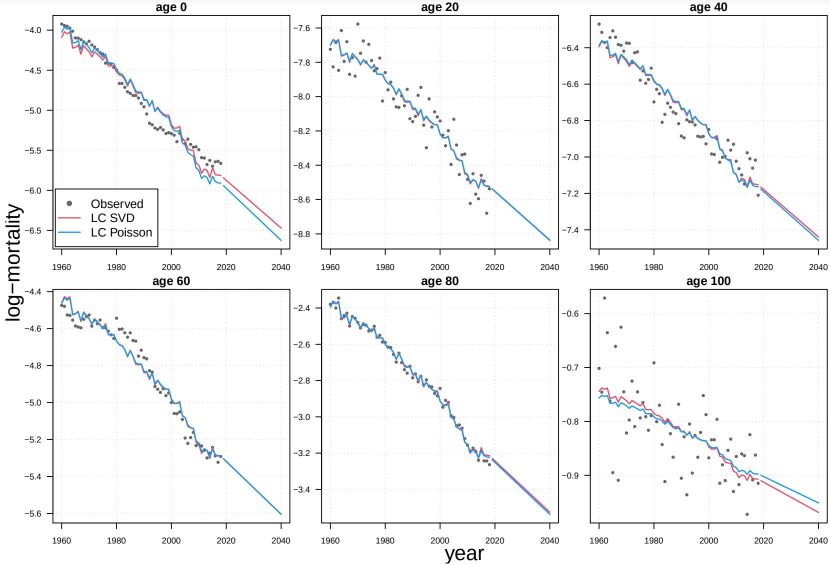
LC SVD vs Poisson: an example

Percentage difference in $\hat{\alpha}$, $\hat{\beta}$, $|\hat{\kappa}|$



Females aged 0–105+ in England & Wales, years 1960–2018.

LC SVD vs Poisson: an example



Females aged 0–105+ in England & Wales, 1960–2018, forecast 2019–2040.

The LC model: a summary

Advantages:

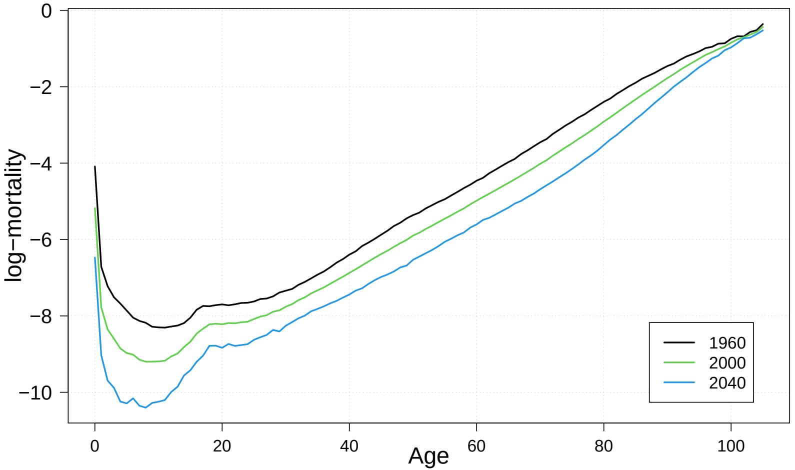
- simple functional form
- univariate time index condenses mortality development $\Rightarrow$ forecasting made "simple"
- stochastic model ($\Rightarrow$ probabilistic intervals), no expert opinions
- more accurate than previous methodologies

Disadvantages:

- ~~"jump-off" bias~~
- ~~Normality assumption (from SVD)~~
- jagged fitted and forecast age profile, lacking smoothness
- fixed age-pattern of mortality decline
- rigid structure

The Smooth LC model

- Fitted and forecast mortality rates over ages are jagged, lacking smoothness in the mortality age profile



The Smooth LC model

- Fitted and forecast mortality rates over ages are jagged, lacking smoothness in the mortality age profile
- Unreasonable outcomes that become more problematic as the forecast horizon grows
- To overcome this, Delwarde et al. (2007) and Currie (2013) penalized the (log-)likelihood to enforce smoothness in α and β:

$$\ln \mathcal{L}^P(\cdot) = \ln \mathcal{L}(\cdot) - \frac{1}{2}\lambda_\alpha \boldsymbol{\alpha}' D' D \,\boldsymbol{\alpha} - \frac{1}{2}\lambda_\beta \boldsymbol{\beta}' D' D \,\boldsymbol{\beta} \qquad (5)$$

  where D is the second order difference matrix, while λ_α and λ_β control the smoothness of α and β

- λ_α and λ_β can be selected by BIC minimization

Smooth LC: estimation

- Again with the Newton-Raphson method!
- Set some starting values, e.g. those from LC SVD: $\hat{\alpha}_x^{(0)} = \alpha_x^{LC}$, $\hat{\beta}_x^{(0)} = \beta_x^{LC}$ and $\hat{\kappa}_t^{(0)} = \kappa_t^{LC}$
- Update the parameters using the formulas:

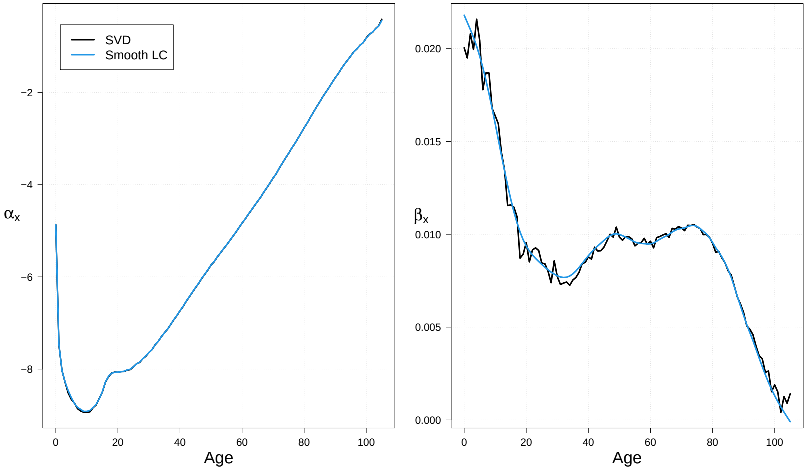
$$\left(C_\alpha^{(\nu)} + P_\alpha\right)\hat{\boldsymbol{\alpha}}^{(\nu+1)} = \left(C_\alpha^{(\nu)} + P_\alpha\right)\hat{\boldsymbol{\alpha}}^{(\nu)} + \sum_t \left(y_{x,t} - \hat{y}_{x,t}^{(\nu)}\right)$$

$$C_\kappa^{(\nu+1)}\hat{\boldsymbol{\kappa}}^{(\nu+2)} = C_\kappa^{(\nu+1)}\hat{\boldsymbol{\kappa}}^{(\nu+1)} + \sum_x \hat{\boldsymbol{\beta}}^{(\nu+1)}\left(y_{x,t} - \hat{y}_{x,t}^{(\nu+1)}\right)$$

$$\left(C_\beta^{(\nu+2)} + P_\beta\right)\hat{\boldsymbol{\beta}}^{(\nu+3)} = \left(C_\beta^{(\nu+2)} + P_\beta\right)\hat{\boldsymbol{\beta}}^{(\nu+2)} + \sum_t \hat{\boldsymbol{\kappa}}^{(\nu+2)}\left(y_{x,t} - \hat{y}_{x,t}^{(\nu+2)}\right)$$

  where $P_\alpha = \lambda_\alpha D'D$, $P_\beta = \lambda_\beta D'D$, and the C are diagonal matrices with elements $C_\alpha^{(\nu)} = \sum_t \hat{y}_{x,t}^{(\nu)}$, $C_\kappa^{(\nu)} = \sum_x (\hat{\beta}_x^{(\nu)})^2 \hat{y}_{x,t}^{(\nu)}$ and $C_\beta^{(\nu)} = \sum_t (\hat{\kappa}_x^{(\nu)})^2 \hat{y}_{x,t}^{(\nu)}$
- Reach convergence and set usual constraints $\sum_x \beta_x = 1$, $\sum_t \kappa_t = 0$
- Also here, no need of second step adjustment!!

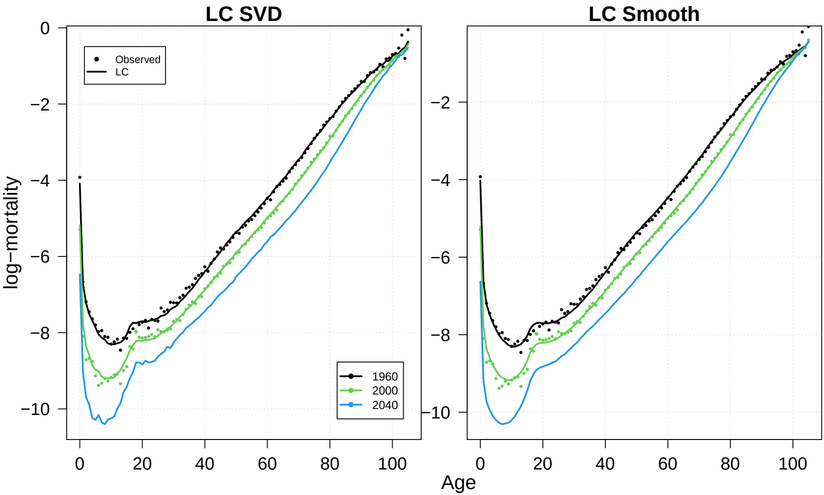
LC SVD vs LC smooth: an example



Females aged 0–105+ in England & Wales, 1960–2018. Smoothing parameters for LC smooth model: λ_α=100 and λ_β=10,000, chosen by BIC minimization

LC smooth: an example

- Fitted and forecast mortality rates over ages now smooth



The LC model: a summary

Advantages:

- simple functional form
- univariate time index condenses mortality development $\Rightarrow$ forecasting made "simple"
- stochastic model ($\Rightarrow$ probabilistic intervals), no expert opinions
- more accurate than previous methodologies

Disadvantages:

- ~~"jump-off" bias~~
- ~~Normality assumption (from SVD)~~
- ~~jagged fitted and forecast age profile, lacking smoothness~~
- fixed age-pattern of mortality decline
- rigid structure

Other LC extensions (single population)

- Booth et al. (2002): adjusting κ_t to match the age-at-death distribution & determining optimal fitting period
- Renshaw and Haberman (2003): adding more than one principal components, i.e. $\ln(m_{x,t}) = \alpha_x + \sum_k \beta_x^k \kappa_t^k$
- Koissi et al. (2006): residual bootstrap to include parameter uncertainty in forecasts
- Renshaw and Haberman (2006): including cohort effects, i.e. $\ln(m_{x,t}) = \alpha_x + \beta_x^{(1)} \kappa_t + \beta_x^{(0)} \gamma_{t-x}$
- Hyndman and Ullah (2007): smooth underlying data (functional data) & additional principal components
- Li et al. (2013): rotation of $\beta_x \Rightarrow$ ~~fixed age-pattern of mortality decline~~
- Camarda and Basellini (2021): smoothing, decomposing and forecasting the three components of mortality (childhood, early-adulthood and senescence), i.e. $m_{x,t} = \sum_k \exp\left(\alpha_x^k + \beta_x^k \kappa_t^k\right)$

References

- Booth, H., Maindonald, J. and Smith, L. (2002). Applying Lee–Carter under conditions of variable mortality decline. *Population Studies*, 56, 325–336.
- Camarda, C.G., and Basellini, U. (2021). Smoothing, decomposing and forecasting mortality rates. *European Journal of Population*, 1–34.
- Hyndman, R.J. and Ullah, M.S. (2007). Robust forecasting of mortality and fertility rates: A functional data approach. *Computational Statistics & Data Analysis*, 51, 4942–4956.
- Koissi, M.-C., Shapiro, A.F., and Högnäs, G. (2006). Evaluating and extending the Lee–Carter model for mortality forecasting: Bootstrap confidence interval. *Insurance: Mathematics and Economics*, 38, 1–20.
- Lee, R.D., and Miller, T. (2001). Evaluating the Performance of the Lee–Carter method for forecasting mortality. *Demography*, 38, 537–549.
- Lee, R.D., and Carter, L.R. (1992). Modeling and forecasting U.S. mortality. *Journal of the American Statistical Association*, 87, 659–671.
- Li, N., Lee, R.D., and Gerland, P. (2013). Extending the Lee–Carter method to model the rotation of age patterns of mortality-decline for long-term projection. *Demography*, 50, 2037–2051.
- Renshaw, A., and Haberman, S. (2003). Lee–Carter Mortality Forecasting with Age-specific Enhancement. *Insurance: Mathematics and Economics*, 33, 255–272.
- Renshaw, A., and Haberman, S. (2006). A cohort-based Extension to the Lee–Carter model for mortality reduction factors. *Insurance: Mathematics and Economics*, 38, 556–570.