Forecasting mortality with the Lee-Carter method

Ugofilippo Baselllini
Max Planck Institute for Demographic Research

Guest lecture for the M.Sc. course "Mortality Analysis" taught by Prof. Roland Rau

University of Rostock

Rostock, 10 January 2024

Preliminaries

- Presentations
- All materials (slides & codes) available at https://github.com/ubasellini/LC-short-course
- Materials are (mostly) derived from a wider IDEM course jointly taught with Giancarlo Camarda (INED): https://github.com/ubasellini/IDEM117-AdvancesMortalityForecasting

Mortality forecasting

- Crucial for sustainability of pensions, insurances, elderly care; predicting population ageing and projecting populations; ...
- Until the 1980s, the methods used to forecast mortality were **deterministic**, based on mathematical formulae or expert judgment
- Revived interest in recent years following the introduction of the Lee-Carter method in 1992
- One of the firstly introduced **stochastic** mortality models $\Rightarrow$ the model "revolutionized probabilistic mortality and population forecasting" (Raftery 2023)

The Lee-Carter method (1992)

- Proposed in 1992 to model and forecast US mortality
- After 30+ years, Lee-Carter (LC) still widely employed by variety of users: governments, private companies, international organizations, ...
- The landmark model in mortality forecasting
- 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 (see Basellini et al. (2023) for a recent comprehensive review)

The LC method

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

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

  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

- Modelling log-rates $\Rightarrow$ fitted and forecast rates constrained to be positive
- Log transformation partially counters heteroscedasticity of observed rates

The LC method

- 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:

$$\sum_x \beta_x = 1 \qquad \text{and} \qquad \sum_t \kappa_t = 0 \tag{2}$$

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

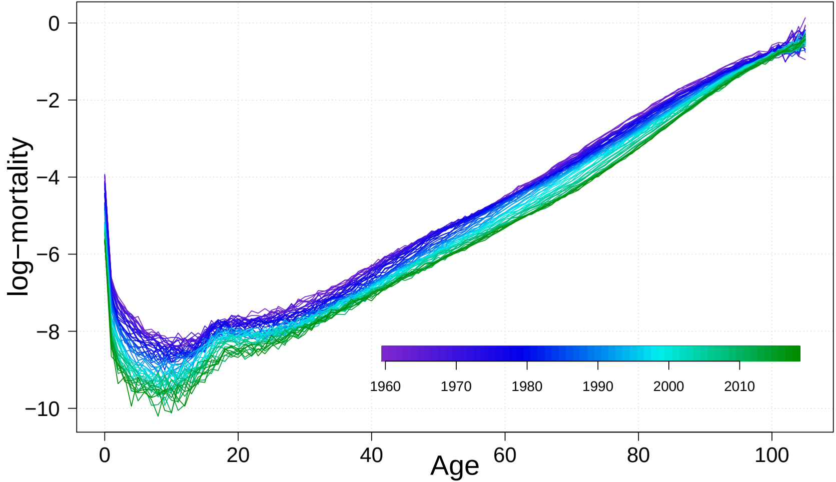
- The model is estimated by minimizing the residual sum of squares:

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

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

Estimating LC: an example

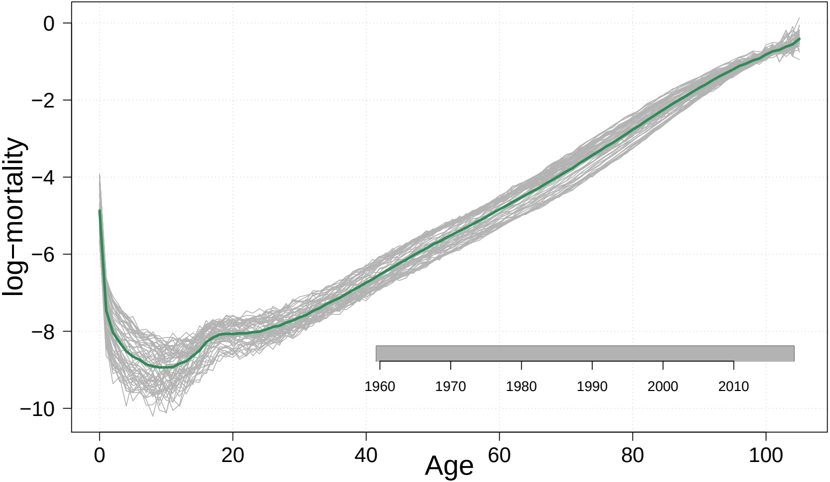
- observed mortality rates $\ln\left(m_{x,t}\right)$



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

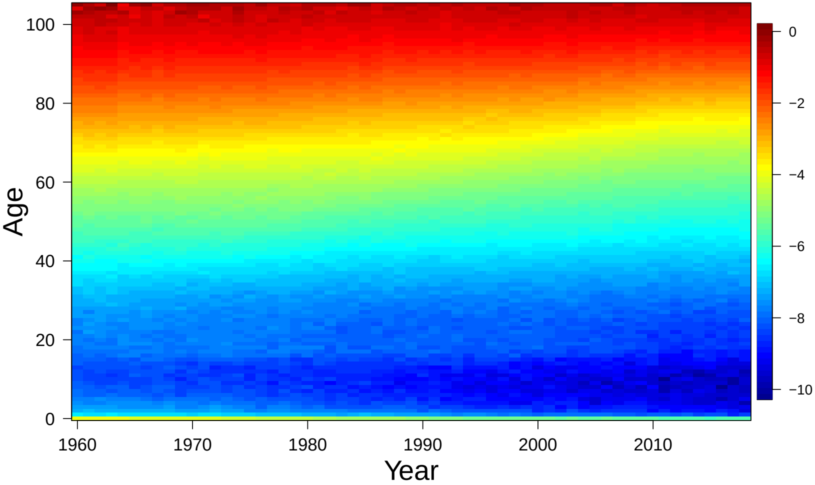
- The model is estimated by minimizing the residual sum of squares:

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

- A singular value decomposition (SVD) is employed to minimize Eq. (3) & derive an ordinary least squares (OLS) solution:
  - $\hat{\alpha}_x$ is the average of the observed $\ln\left(m_{x,t}\right)$
  - $\hat{\beta}_x$ and $\hat{\kappa}_t$ are derived from 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

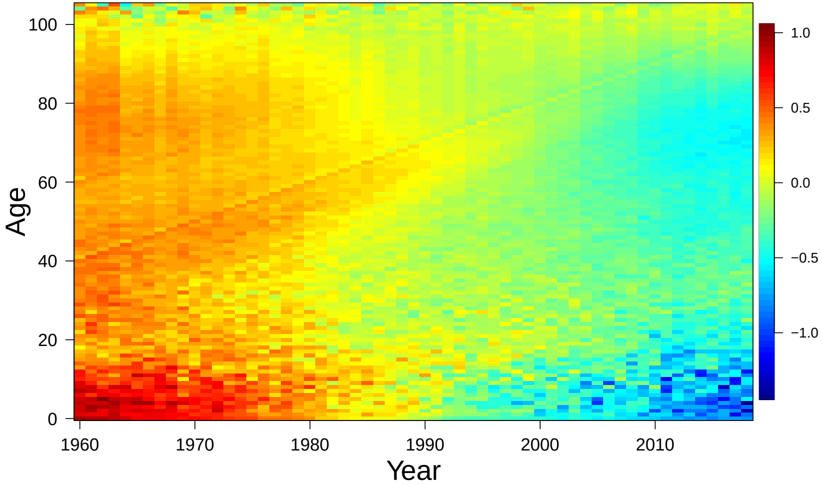
- $\boldsymbol{M} = (\ln(m_{x,t}))$: matrix of observed mortality rates



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

Estimating LC: an example

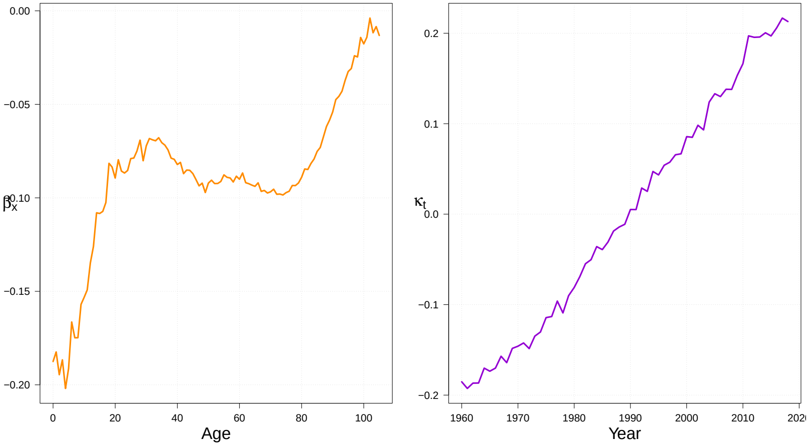
- $\tilde{M} = (\ln(m_{x,t}) - \hat{\alpha}_x)$: matrix of "centered" mortality rates



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

Estimating LC: an example

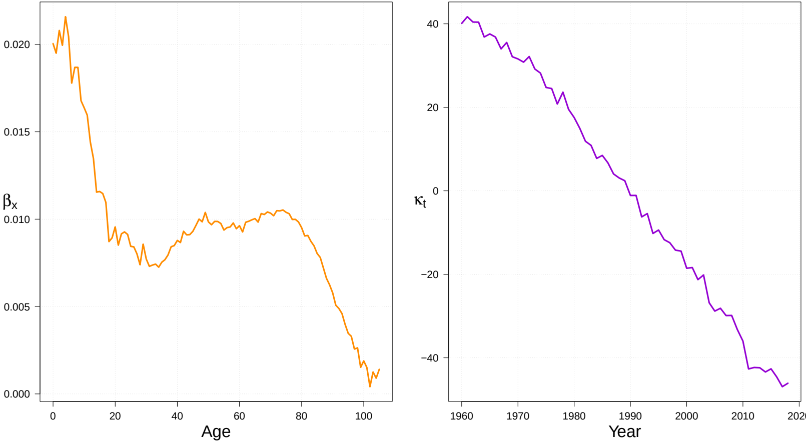
- From SVD of $\tilde{M}$



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

Estimating LC: an example

- Adjusting the left-singular vector to sum to 1, and multiplying the right-singular vector for the leading singular value and the sum of the left-singular vector, 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:

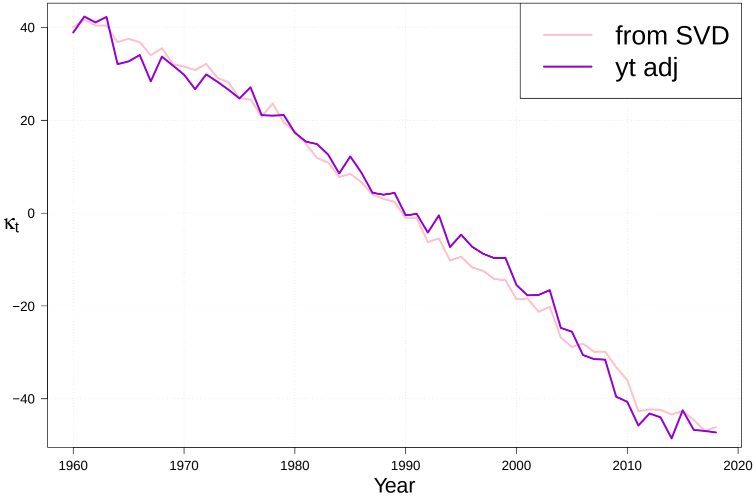
$$\sum_{x,t} \Big(\ln(m_{x,t}) - \alpha_x - \beta_x \kappa_t \Big)^2$$

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

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

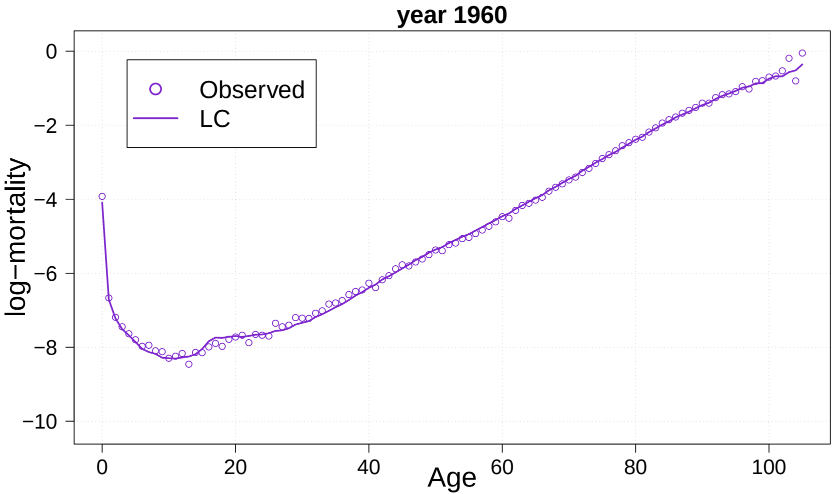
Estimating LC: an example

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



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

Estimating LC: an example



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

Time-series analysis: a short introduction

- A *time-series* is a collection of observations made sequentially through time. Let us focus on discrete time series recorded at equal intervals of time
- Suppose we have an observed time series $y_1, y_2, \ldots, y_T$ and we wish to forecast future values such as y_{T+h}. The integer h is called *forecasting horizon*, and we denote by $\hat{y}_{T+h}$ the forecast made at time T for h steps ahead
- Forecasting methods may be broadly identified into three types:
  - *Judgemental forecasts*: based on subjective judgement and intuition
  - *Univariate methods*: forecasts depend on present and past values of a single series
  - *Multivariate methods*: forecasts depend, at least partly, on values of one or more additional variables (predictors)

Descriptive techniques

- In order to *forecast*, description and modelling of data is a prerequisite
- Always start from plotting your data!!
- This can help to identify two main sources of variation in many time series: i) trend, and ii) seasonal variation
- These variations are typically removed before time-series modelling via differencing $\Rightarrow$ helps to stabilize the **mean**
- The time plot may also help to decide whether a variable needs to be transformed $\Rightarrow$ helps to stabilize the **variance**

Stationary stochastic processes

- A *stochastic* time series is one where future values can only partly be determined by past values
- A process is defined *stationary* if its properties do not change through time
- More formally, let y_t be the realization of the underlying random variable Y_t, and the observed time-series $\boldsymbol{y} = [y_1, y_2, \ldots, y_T]$ be a realization of the stochastic process
- A stochastic process is second-order stationary if its first and second moments are finite and do not change over time, i.e.:

$$\mathbb{E}[Y_t] = \mu$$
$$\mathrm{COV}[Y_t, Y_{t+k}] = \mathbb{E}[(Y_t - \mu)(Y_{t+k} - \mu)] = \gamma_k$$

  for all t and for $k = 0, 1, 2, \ldots$ (note, for $k = 0$, $\gamma_0 = \sigma^2$)
- In simpler words, a stationary series has constant mean, constant variance and no predictable patterns in the long-term

Differencing

- Powerful tool to stabilize the mean and obtain stationary time-series
- First-order differencing: $y'_t = y_t - y_{t-1}$, y'_t is the *change* between observations of y_t (composed by $T-1$ values)
- If y'_t still non-stationary, second-order differencing: $y''_t = y'_t - y'_{t-1} = y_t - 2y_{t-1} + y_t - 2$ (composed by $T-2$ values)
- Almost never necessary to go beyond y''_t
- For seasonal data, seasonal differencing: $y'_t = y_t - y_{t-m}$, where m is the number of seasons, y'_t is the change between one year to the next
- In addition to correlogram, two main tests for determining the required order of differencing:
  - Augmented Dickey Fuller test: H_0 data are non-stationary and non-seasonal
  - Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test : H_0 data are stationary and non-seasonal

Random walk

- Random walk models are widely used for non-stationary data (e.g. economic and financial data):

$$Y_t = Y_{t-1} + Z_t$$

  where Z_t is a purely random process (with zero mean and constant variance σ_z^2)

- Random walk is non-stationary (variance increases through time), but first-order differences (purely random process) is stationary
- Typical features:
  - sudden and unpredictable changes of direction
  - long periods of apparent trends up/down
- Forecasts for the random walk model are simply given by the value of the last observation (i.e. naïve forecast), i.e.:

$$\hat{y}_{T+h|T} = y_T$$

Random walk with drift

- A closely related model that allows first differences to have non-zero mean is the **random walk with drift**:

$$Y_t = c + Y_{t-1} + Z_t$$

where Z_t is a purely random process, and c is a constant
  - if $c > 0$, y_t will tend to drift upwards
  - if $c < 0$, y_t will tend to drift downwards
- The estimate of the drift c is given by the average of the changes between consecutive observations, i.e.

$$c = \frac{1}{T-1} \sum_{t=2}^{T} (y_t - y_{t-1}) = \frac{y_T - y_1}{T-1}$$

- Forecasts for the random walk model with drift are given by:

$$\hat{y}_{T+h|T} = y_T + c\,h$$

$\Rightarrow$ equivalent to drawing a line between the first and last observations, and extrapolating it into the future

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)
- For this time-series model: $\hat{\kappa}_{T+h|T} = \hat{\kappa}_T + c\,h$
- Simulated future trajectories of $\hat{\kappa}_{T+h}$ to construct prediction intervals for $\hat{m}_{x,T+h}$ and other summary measures (e.g., $\hat{e}_{0,T+h}$)
- (Coale and Guo (1989) adjustment for forecast rates at ages 85+)

Forecasting with LC: a schematic view

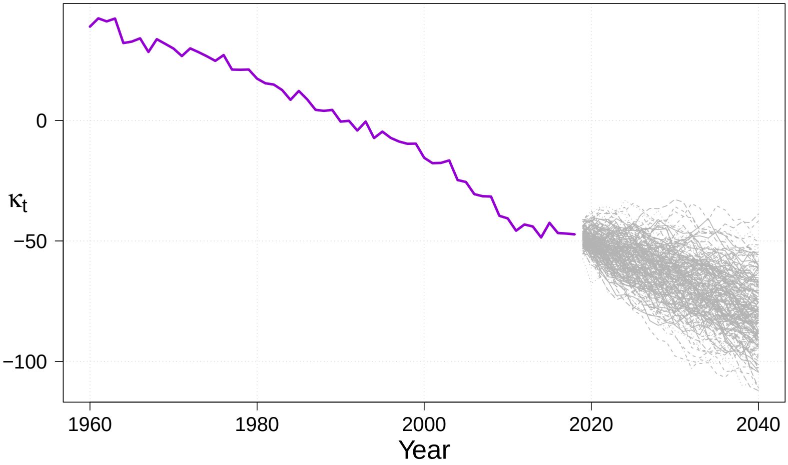
$$\ln\left(m_{x,t}\right) \simeq \alpha_x + \beta_x \kappa_t \qquad \ln\left(m_{x,T+h}\right) \simeq \alpha_x + \beta_x \kappa_{T+h}$$

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

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

Forecasting LC: an example

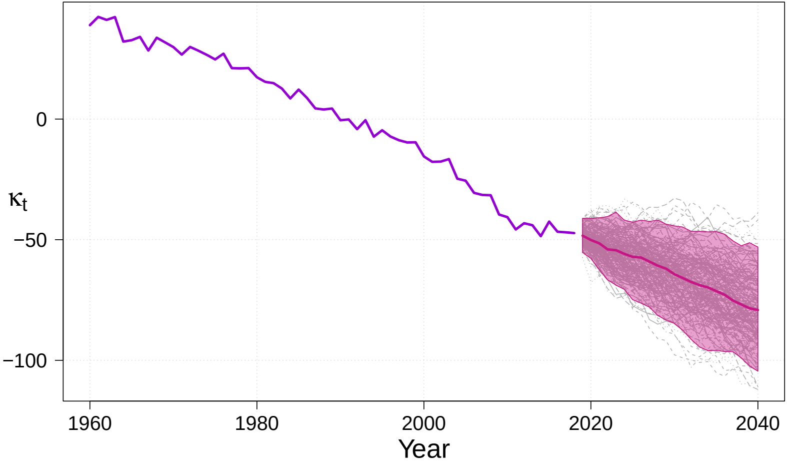
- Simulated future paths of $\hat{\kappa}_t$ from random walk with drift



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

Forecasting LC: an example

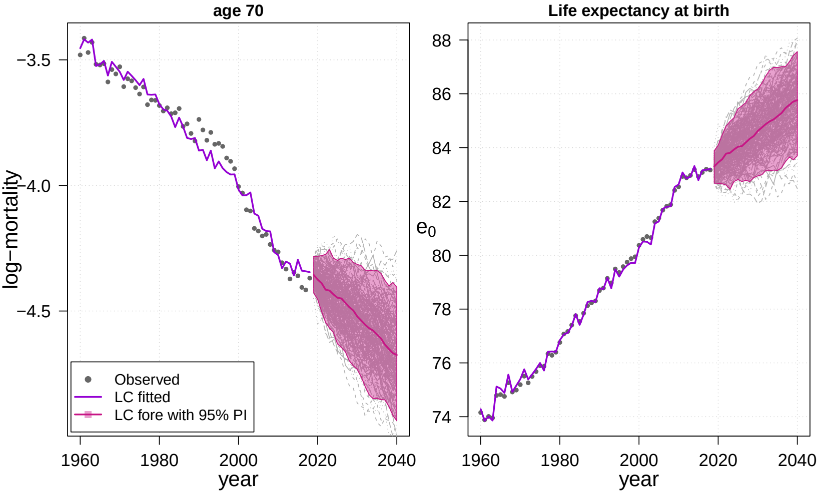
- Simulated future paths of $\hat{\kappa}_t$ from random walk with drift $\Rightarrow$ prediction intervals from empirical percentiles



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

Forecasting LC: an example

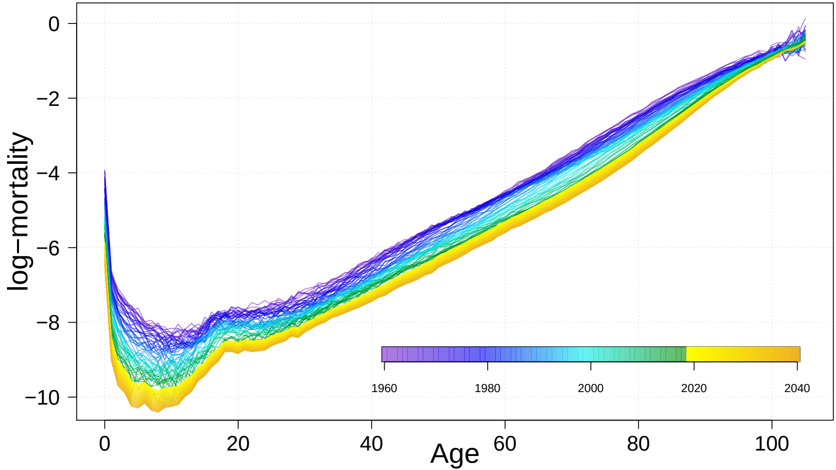
- Prediction intervals from simulated future $\hat{\kappa}_{T+h}$



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

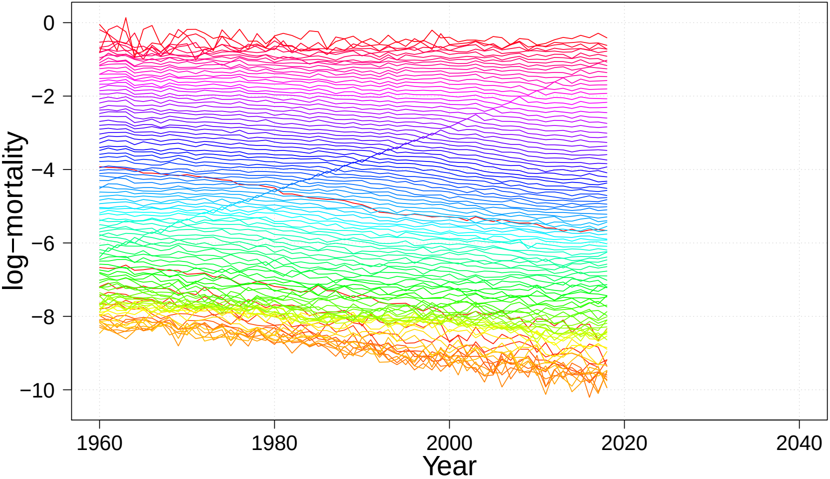
Forecasting LC: an example

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



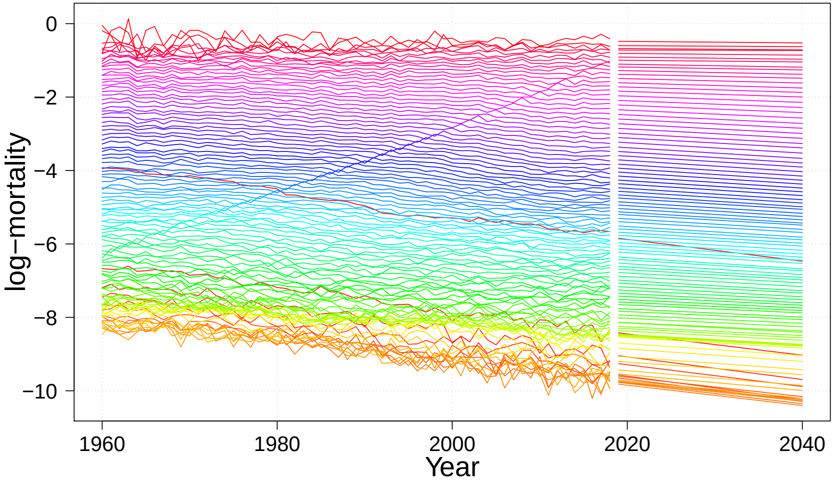
Females aged 0–105+ in England & Wales, 1960–2018, forecast 2019–2040.

Forecasting LC: an example



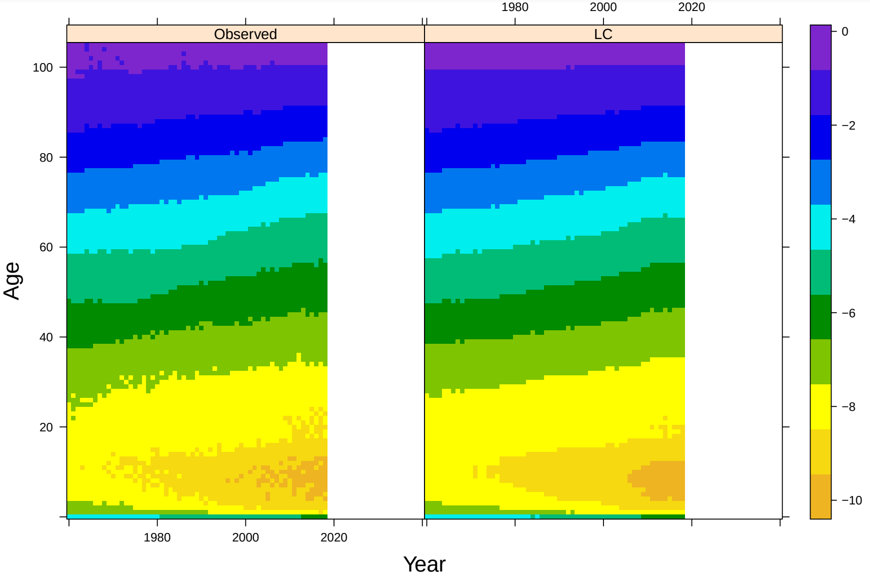
Females aged 0–105+ in England & Wales, 1960–2018, forecast 2019–2040.

Forecasting LC: an example



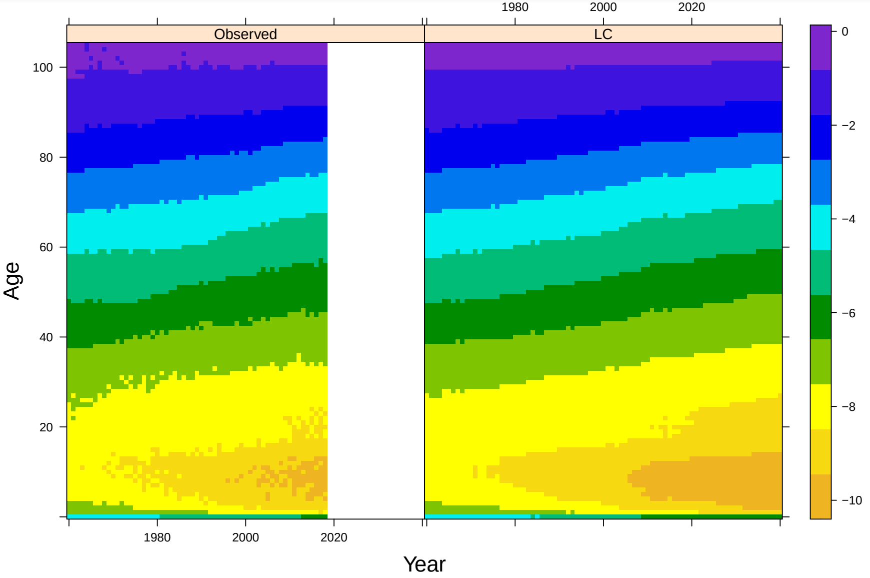
Females aged 0–105+ in England & Wales, 1960–2018, forecast 2019–2040.

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 & powerful method: forecast rates derived by modeling single time index by standard time-series model
- linear time index very often captures well historical decline in mortality
- stochastic model ⇒ probabilistic intervals
- extrapolative approach, 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

⇒ several extensions proposed to overcome some of these issues

Some LC extensions

Disadvantages:

- ~~"jump-off" bias~~ $\Rightarrow$ observed jump-off rates (Lee and Miller 2001)
- ~~Normality assumption (from SVD)~~ $\Rightarrow$ Poisson LC (Brouhns et al. 2002)
- ~~jagged fitted and forecast age profile, lacking smoothness~~ $\Rightarrow$ Smooth LC (Delwarde et al. 2007)
- ~~fixed age-pattern of mortality decline~~ $\Rightarrow$ Li et al. (2013)
- 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
- Camarda and Basellini (2021): smoothing, decomposing and forecasting the three components of mortality (childhood, early-adulthood and senescence), i.e. $m_{x,t} = \sum_k \exp\left(\alpha_x^k + \beta_x^k \kappa_t^k\right)$
- for a comprehensive review, see Basellini et al. (2023)

References

- Basellini, U., Camarda, C.G. and Booth, H. (2023). Thirty years on: A review of the Lee–Carter method for forecasting mortality. *International Journal of Forecasting*, 39(3), 1033–1049.
- Booth, H., Maindonald, J. and Smith, L. (2002). Applying Lee–Carter under conditions of variable mortality decline. *Population Studies*, 56, 325–336.
- Brouhns, N., Denuit, M., and Vermunt, J. K. (2002). A Poisson log-bilinear regression approach to the construction of projected lifetables. *Insurance: Mathematics & Economics*, 31, 373–393.
- Camarda, C.G., and Basellini, U. (2021). Smoothing, decomposing and forecasting mortality rates. *European Journal of Population*, 1–34.
- Coale, A., and Guo, G. (1989). Revised regional model life tables at very low levels of mortality. *Population Index*, 613–643.
- Delwarde, A., Denuit, M., and Eilers, P. H. C. (2007). Smoothing the Lee-Carter and Poisson log-bilinear models for mortality forecasting: A penalized log-likelihood approach. *Statistical Modelling*, 7, 29–48.
- 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.

References

- 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.
- Raftery, A. (2023). The Lee-Carter Method and Probabilistic Population Forecasts. *International Journal of Forecasting*.
- 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.

Forecasting mortality with the Lee-Carter method

Thank you for your attention !!

Ugofilippo Basellini

Laboratory of Digital and Computational Demography

Max Planck Institute for Demographic Research (MPIDR)

basellini@demogr.mpg.de

Website: [MPIDR]; [Personal]