Introduction to demographic & mortality forecasting

Christina Bohk-Ewald

As part of IMPRS-PHDS course on Population Health
MPIDR, Germany
November 19, 2020

Brief round of introduction

Who are you? What is your study background?

What is your experience in demographic forecasting?

What are you most interested in with respect to demographic forecasting?

Today's lecture:
Introduction to demographic & mortality forecasting

- **Demographic forecasting**
  - What they are about
  - How they can look like in real life
  - What they are good for
  - How they are generated in general
  - What might be potential sources of error and how to account for them

- **Mortality forecasting**
  - What kind of methods there are
  - The Lee-Carter method
  - New directions

Today's lab:
Hands-on exercise in mortality forecasting with R

Forecast US mortality for women and men 50 years ahead:

1. Load mortality data from the Human Mortality Database.
2. Implement and use the Lee-Carter method to fit and forecast mortality for US women and men 50 years ahead based on base periods of different length.
3. Analyze and compare the LC mortality forecasts for US women and men. Explain the impact of the length of the base period on the US mortality forecast.
4. Thinking about the basic procedure of the LC method and its underlying assumptions, how plausible are the US mortality forecasts and how could they be improved?

Please make sure to have installed R package `fds`. And to have at hand your username and password for the Human Mortality Database.

What is: Demography & Forecasting

- **Demography** is the science of populations (*demos*) and their measurement (*graphy*).

- **Forecasting** is the process of making statements about likely future development of variable(s) of interest.

What is: Demographic forecasting

- **Demographic forecasts** predict how populations will develop over time in the future
- Population balance equation:

$$P_{t+n} = P_t + B_{[t,t+n]} - D_{[t,t+n]} + I_{[t,t+n]} - E_{[t,t+n]}$$

Population forecasts	Fertility forecasts	Mortality forecasts	Migration forecasts

$\rightarrow$ Demographic forecasting comprises all these components

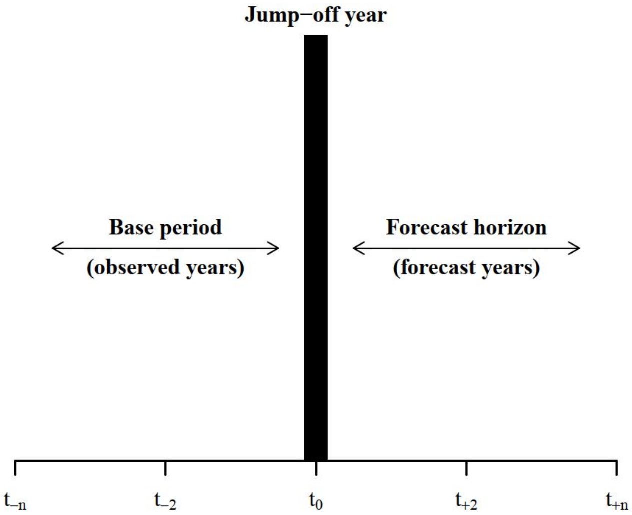
Typical questions

- How long will people live?
- How many of the remaining years of life will people spend in good health, in poor health, or in work and in retirement?
- How many children will people have in 5 and in 50 years from now?
- How many people will live worldwide in upcoming years?
- *What if ...?*

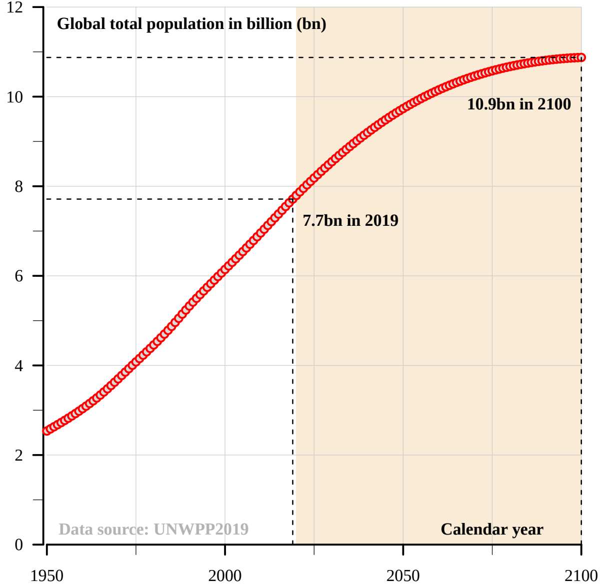
Demographic forecasts in the real world

Example: United Nations World Population Prospects 2019
($\longrightarrow$ published only in June this year)

Some terminology first...



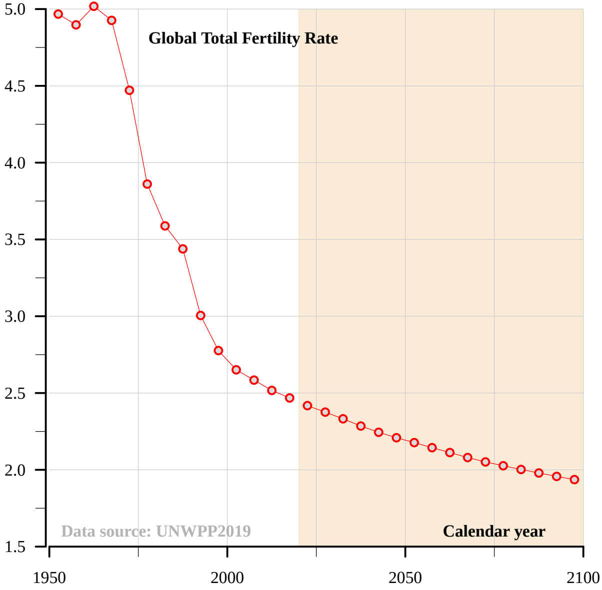
Global population size and growth



[one billion is equal to one thousand million]

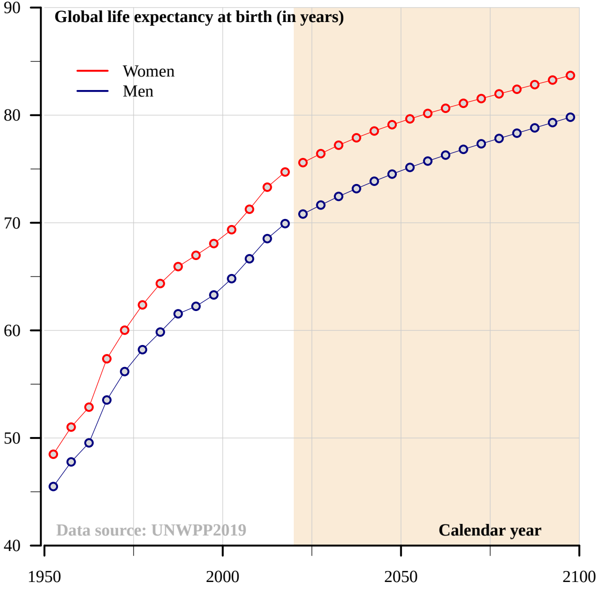
1. Global population size is forecasted to increase
2. ...but less and less with each forecast year

Global Total Fertility Rate



1. Global TFR is forecasted to decline ...but less and less with each forecast year

Global life expectancy at birth



1. Global e_0 is higher for women than for men
2. Global e_0 is forecasted to increase ...but less and less with each forecast year

Societal relevance

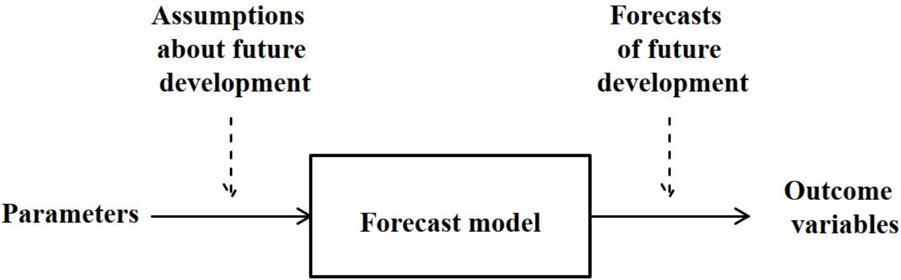
Reliable demographic forecasts lead to informed decisions, policies, and programs affecting e.g. people's social & economic welfare

- They serve as basis for further analysis
  (that predict demand for resources and services in upcoming years)

- Assist in planning, decision making, and creating goals
  in areas such as health care, education, housing,
  energy consumption, transport, retirement planning

→ applies to different geographical levels and types of organization

Basic procedure



Basic procedure

Demographic forecasting is a form of prediction
(that you already know from regression analysis / statistics).

Demographic forecast methods often:

- extrapolate past developments into the future
- take time into account
- assume that future is a continuation of the past
- assume that functional relationships between model parameters (expressed in the forecast model) are valid throughout time

$\longrightarrow$ underlying assumptions might be questionable?

Potential sources of error

"Prediction is very difficult, especially about the future."

attributed to Niels Bohr

Forecast errors are deviations between forecasts and their true realizations.

What are potential sources of error?
What could possibly go wrong?

Potential sources of error

- model misspecification (all models are wrong)
- missing / incorrect data (as input for forecast model)
- errors in model implementation
- unexpected events (leading to gross shifts in demographic behavior)
- ...

→ contribute to forecast uncertainty

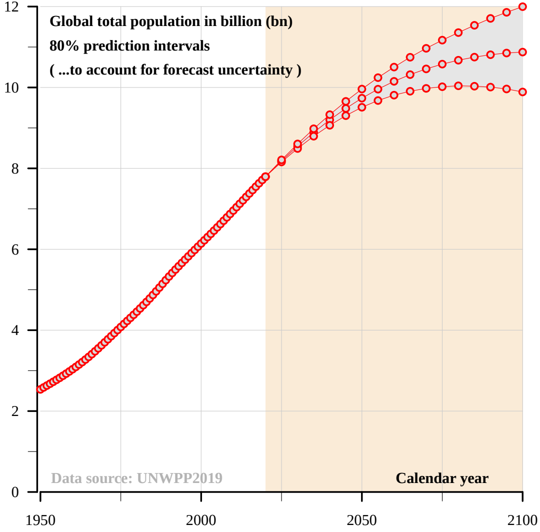
Potential sources of error

Anecdote about the inductivist chicken:

> "The man who has fed the chicken every day throughout its life at last wrings its neck instead, showing that more refined views as to the uniformity of nature would have been useful to the chicken."
>
> **Bertrand Russell (1912)**

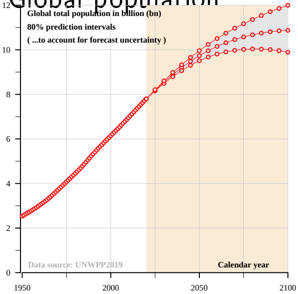
Probabilistic forecasts

...quantify forecast uncertainty

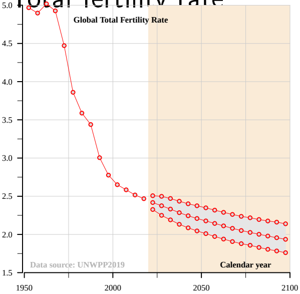


Probabilistic forecasts quantify forecast uncertainty

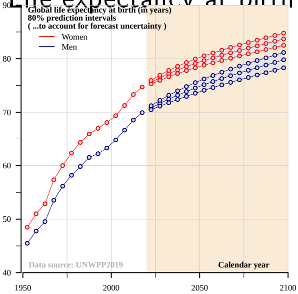
Global population



Total fertility rate



Life expectancy at birth



$\longrightarrow$ deterministic forecasts versus probabilistic forecasts

What you have learned today about demographic forecasting

- Define demographic forecasting
- List typical questions
- Apply basic terminology of demographic forecasting
- Describe UN forecast of the global population
- Describe societal relevance
- Describe basic procedure how to generate demographic forecasts
- Describe potential sources of error
  and know that they can be accounted for in probabilistic forecasts

Today's class in the morning: Introduction to demographic & mortality forecasting

- **Demographic forecasting**
  - What they are about
  - How they can look like in real life
  - What they are good for
  - How they are generated in general
  - What might be potential sources of error and how to account for them
- **Mortality forecasting**
  - What kind of methods there are
  - The Lee-Carter method
  - New directions

Mortality forecasting

Some empirical core questions:

- How long will people live in 10, 50, 100 years from now?
- How large will be mortality of 65-year-olds in 1-100 years from now?
- ...

Some urgent empirical questions of today:

- How many people will die from COVID-19?
- How large will be excess mortality from COVID-19?
- ...

Mortality forecasting

Some methodological core questions:

- How to best capture and forecast mortality dynamics, for example by age and calendar year (and birth cohort) with ongoing time?
- To what extent should we include factors that affect mortality dynamics? What are reliable sources of information?
- What are the characteristics of a *good* statistical forecast model and how to assess the accuracy of demographic forecasts?

Mortality forecasting approaches

1. Extrapolation methods

- Model trends in mortality over age and time (and birth cohort)
- Are objective & data-driven
- Assume that basic trends in mortality were regular and would continue in years ahead

Mortality forecasting approaches

2. Explanation methods

- Take into account mortality that is e.g. attributable to health-related behavior (such as tobacco smoking) and/or causes of death
- Consider explanatory mechanisms / risk factors of mortality
- Are prone to model misspecification (due to high complexity)

Mortality forecasting approaches

3. Expert-based methods

- Use expert opinion to e.g. interpolate mortality between start and target value
- Are subjective & opinion-driven
- Might be biased (as experts tend to be overly confident)

Mortality forecasting approaches

1. Extrapolation methods
   - Model trends in mortality over age and time (and birth cohort)
   - Are objective & data-driven but assume basic trends in mortality to be regular and to continue in years ahead
2. Explanation methods
   - Take into account mortality that is e.g. attributable to health-related behavior (such as tobacco smoking) and/or causes of death
   - Consider explanatory mechanisms / risk factors of mortality but are prone to model misspecification (due to high complexity)
3. Expert-based methods
   - Use expert opinion to e.g. interpolate mortality between start and target value
   - Are subjective & opinion-driven and might be biased (as experts tend to be overly confident)
4. Mixture of methods above

$\longrightarrow$ Overview in e.g. Booth (2006) and Booth and Tickle (2008)

Mortality forecasting approaches

What approach would you prefer and why?

Would you have thought of another way to forecast mortality?

The Lee-Carter method

Modeling and Forecasting U.S. Mortality

RONALD D. LEE and LAWRENCE R. CARTER*

Time series methods are used to make long-run forecasts, with confidence intervals, of age-specific mortality in the United States from 1990 to 2065. First, the logs of the age-specific death rates are modeled as a linear function of an unobserved period-specific intensity index, with parameters depending on age. This model is fit to the matrix of U.S. death rates, 1933 to 1987, using the singular value decomposition (SVD) method; it accounts for almost all the variance over time in age-specific death rates as a group. Whereas e_0 has risen at a decreasing rate over the century and has decreasing variability, $\mathbf{k}(t)$ declines at a roughly constant rate and has roughly constant variability, facilitating forecasting. $\mathbf{k}(t)$, which indexes the intensity of mortality, is next modeled as a time series (specifically, a random walk with drift) and forecast. The method performs very well on within-sample forecasts, and the forecasts are insensitive to reductions in the length of the base period from 90 to 30 years; some instability appears for base periods of 10 or 20 years, however. Forecasts of age-specific rates are derived from the forecasts of $\mathbf{k}$, and other life table variables are derived and presented. These imply an increase of 10.5 years in life expectancy to 86.05 in 2065 (sexes combined), with a confidence band of plus 3.9 or minus 5.6 years, including uncertainty concerning the estimated trend. Whereas 46% now survive to age 80, by 2065 46% will survive to age 90. Of the gains forecast for person-years lived over the life cycle from now until 2065, 74% will occur at age 65 and over. These life expectancy forecasts are substantially lower than direct time series forecasts of e_0, and have far narrower confidence bands; however, they are substantially higher than the forecasts of the Social Security Administration's Office of the Actuary.

KEY WORDS: Demography; Forecast; Life expectancy; Mortality; Population; Projection.

From 1900 to 1988, life expectancy in the United States rose from 47 to 75 years. If it were to continue to rise at this same linear rate, life expectancy would reach 100 years in 2065, about seventy five years from now. The increase would

Next we fit the demographic model to U.S. data and evaluate its historical performance. Using standard time series methods, we then forecast the index of mortality and generate associated life table values at five-year intervals. Because we

Google Scholar | Modeling and forecasting US mortality

Artikel — Ungefähr 101.000 Ergebnisse (**0,10** Sek.)

Beliebige Zeit
Seit 2019
Seit 2018
Seit 2015
Zeitraum wählen...

Modeling and forecasting US mortality
RD Lee, LR Carter - Journal of the American statistical association, 1992 - Taylor & Francis
Time series methods are used to make long-run forecasts, with confidence intervals, of age-specific mortality in the United States from 1990 to 2065. First, the logs of the age-specific death rates are modeled as a linear function of an unobserved period-specific intensity ...
Zitiert von: 2888 Ähnliche Artikel Alle 17 Versionen

1. Golden standard to forecast mortality.
2. Published in 1992 and widely used since then.
3. Extrapolation method. Simple and robust.
4. Many extensions since 1992.

Use Lee-Carter model to fit and forecast US female mortality

We will look at the broad idea first
before we will also briefly look at methodological and technical details.

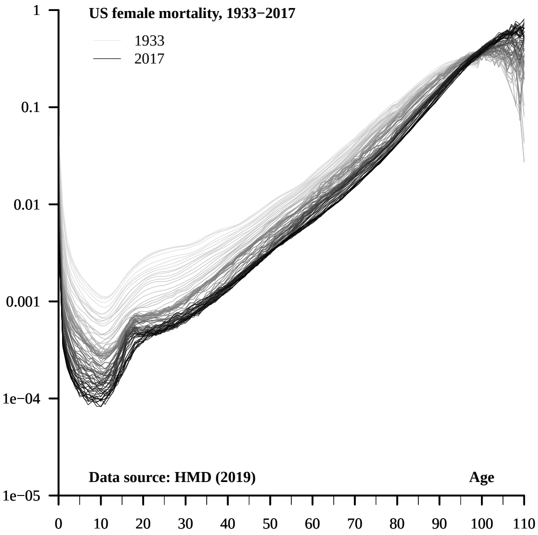
Example of US female mortality.

Two-step procedure:

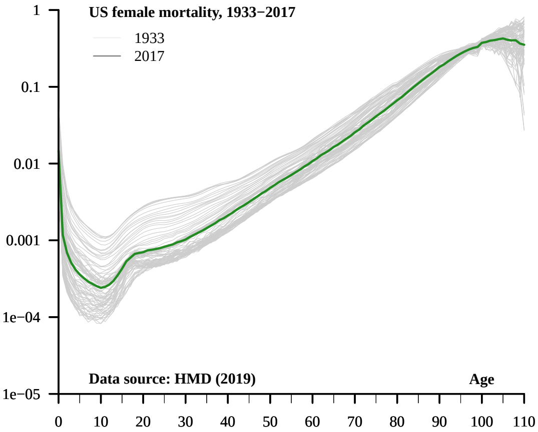
1. **Fit** Lee-Carter **model to** US female **mortality** by age and over time **in base period**.

2. **Forecast** US female **mortality** over time **in upcoming years**.

1. Fit Lee-Carter model to US female mortality, 1933-2017

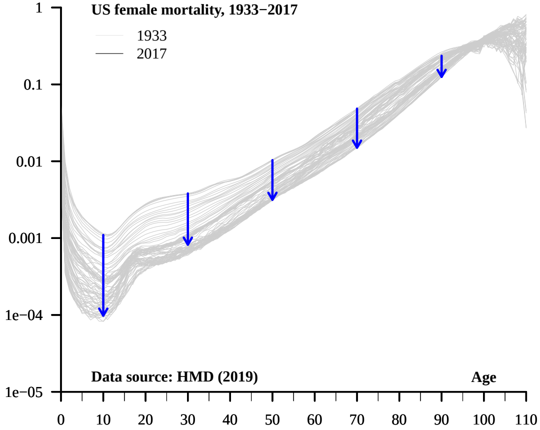


...with only few model parameters



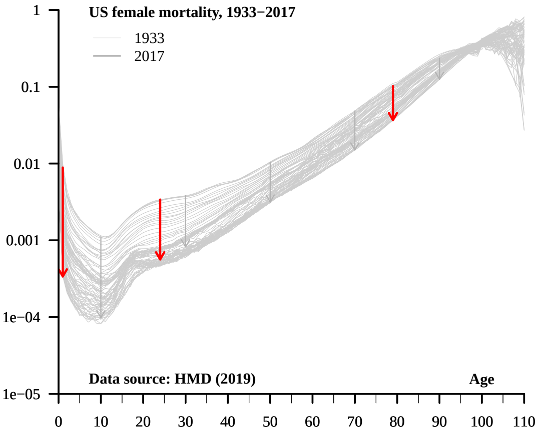
$$\log m_{x,t} = \alpha_x + \beta_x \kappa_t$$

...with only with few model parameters



$$\log m_{x,t} = \alpha_x + \beta_x \kappa_t$$

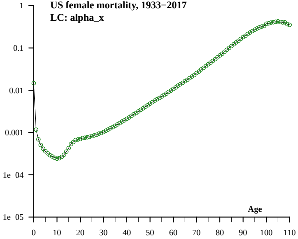
...with only with few model parameters

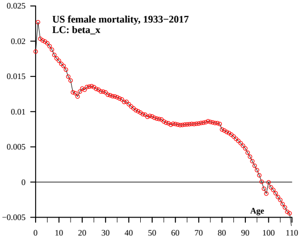


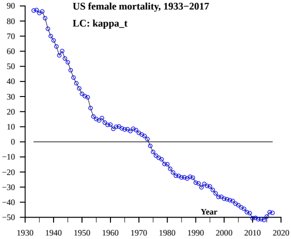
$$\log m_{x,t} = \alpha_x + \beta_x \kappa_t$$

Lee-Carter model fitted to US female mortality in base period 1933-2017

$$\log m_{x,t} = \alpha_x + \beta_x \kappa_t$$





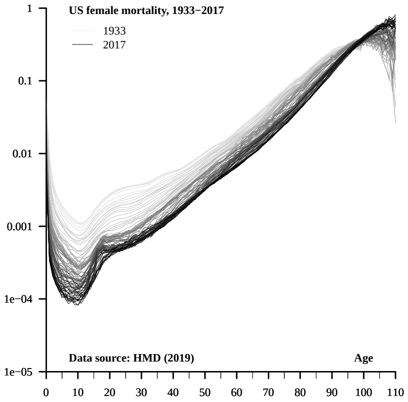


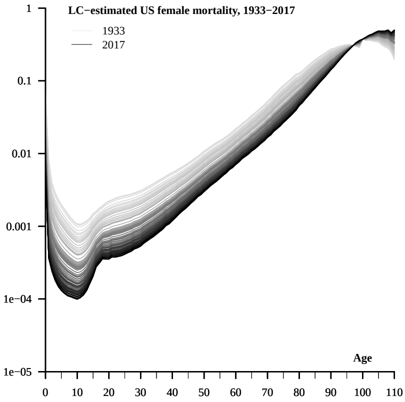
- α_x is the general shape of mortality across age x

- β_x is the change of mortality at age x
- $\beta_x > 0$: mortality decline, $\beta_x < 0$: mortality increase

- κ_t is an index of the level of mortality over time t
- direction and slope indicate strong mortality decline

...and one more thing

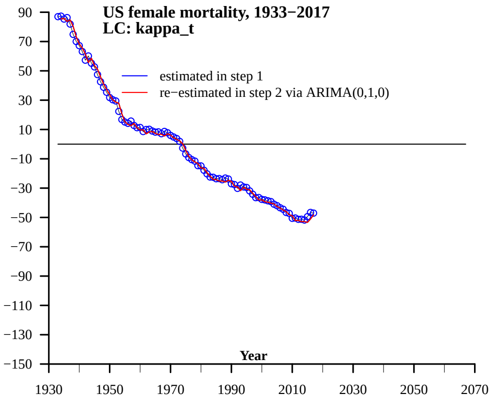




to account for residuals between LC-estimated and observed mortality:

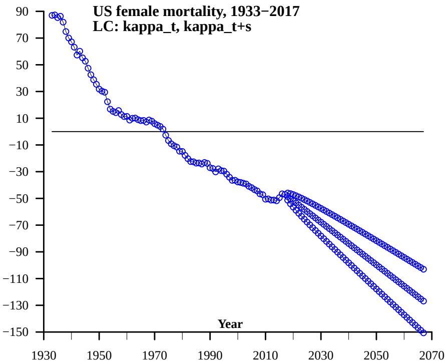
$$\log m_{x,t} = \alpha_x + \beta_x \kappa_t + \epsilon_{x,t}$$

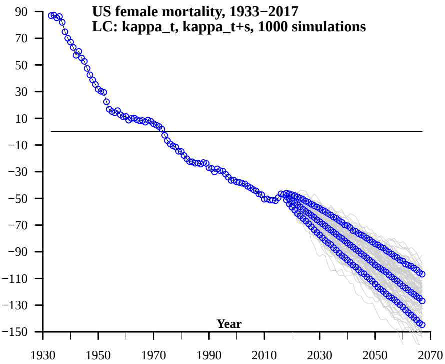
2. Use fitted LC-model to forecast US female mortality s years ahead via time index k_t



- κ_t is an index of the level of mortality over time t
- **Fit** estimated κ_t in base period using a time series model, e.g. ARIMA(0,1,0)
  - Random walk with drift: $\kappa_t = \kappa_{t-1} + \delta + \epsilon_t$
    - with δ being a drift term
    - and ϵ_t being an error term

2. Use fitted LC-model to forecast US female mortality s years ahead via time index k_t

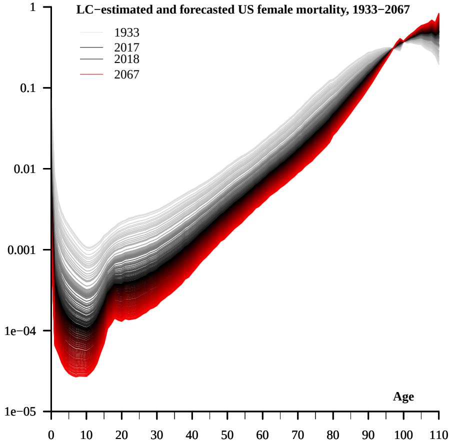


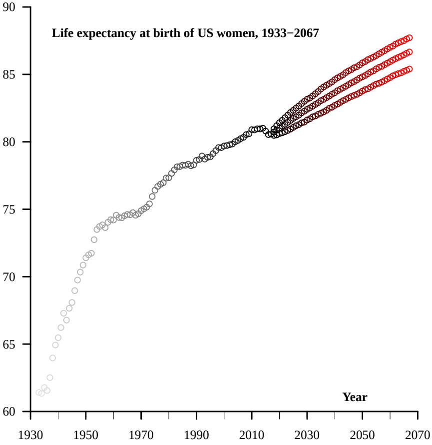


ARIMA(0,1,0), random walk with drift: $\kappa_t = \kappa_{t-1} + \delta + \epsilon_t$

80% **prediction intervals** based on statistical theory (left) & simulation (right)

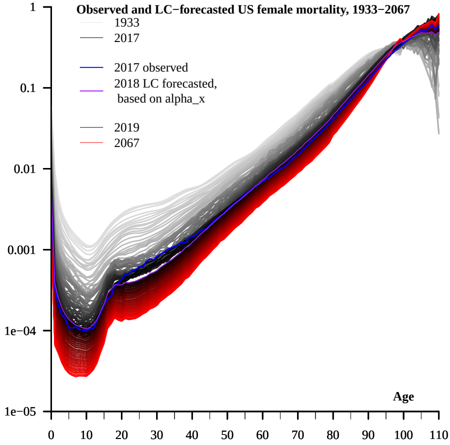
Forecasting US female mortality 50 years ahead using base period 1933-2017

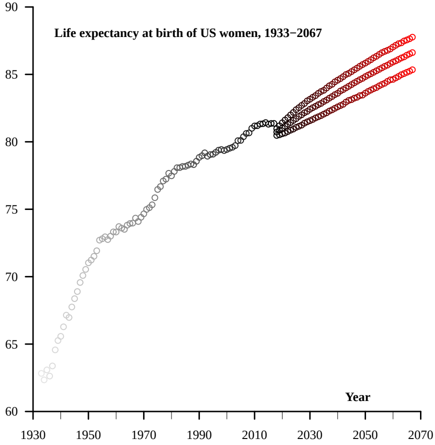




κ_t point estimates are based on median of 1000 simulated trajectories

Forecasting US female mortality 50 years ahead using base period 1933-2017, **jump-off-bias**

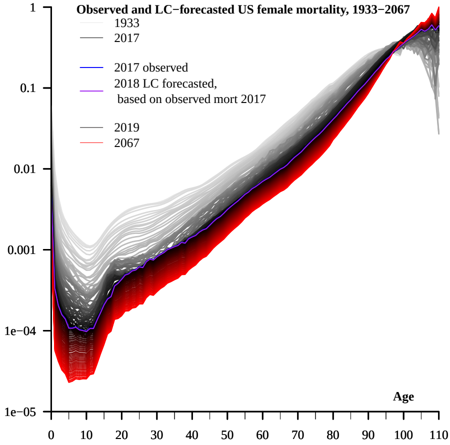


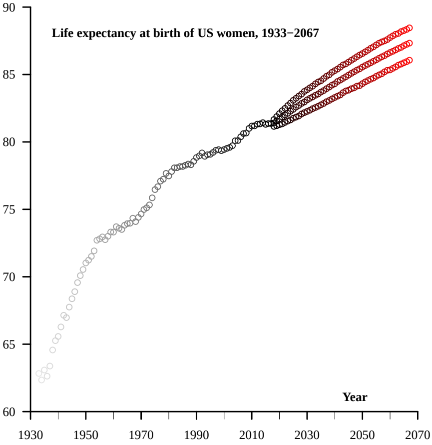


κ_t point estimates are based on median of 1000 simulated trajectories

Forecasting US female mortality 50 years ahead using base period 1933-2017, **corrected for jump-off-bias:**

$$\log m_{x,t} = m_{x,2017} + \beta_x \kappa_t^* + \epsilon_{x,t}$$





κ_t point estimates are based on median of 1000 simulated trajectories

Think about it.

What does the Lee-Carter model do?

What are the main steps?

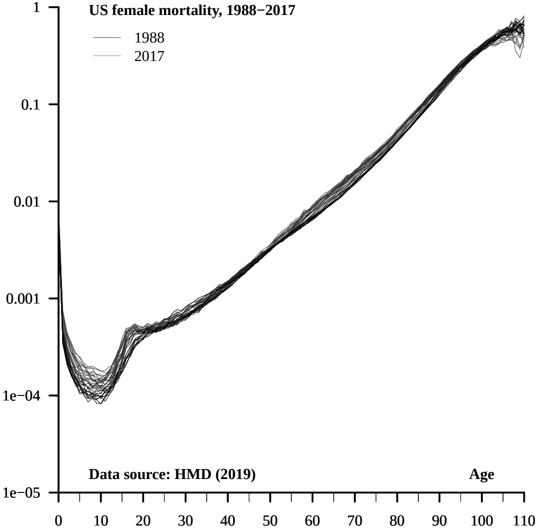
Could the US female mortality forecasts differ
although we always use the Lee-Carter model to generate them?

What are high-impact settings & parameters of the LC model?

What impacts US female mortality forecast with LC model?

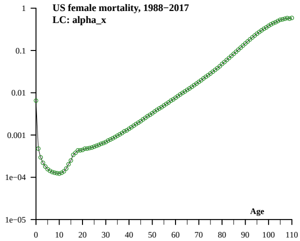
- Observed levels and trends in base period (α_x, β_x, and κ_t)
- Fitting procedure (e.g. singular value decomposition, maximum likelihood)
- Forecast time index κ_t
  - Time series model
  - Prediction intervals (based on e.g. simulation or statistical theory)
- Implementation (e.g. different R-packages)

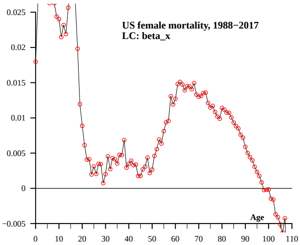
Impact base period: fit LC model to US female mortality, 1988-2017

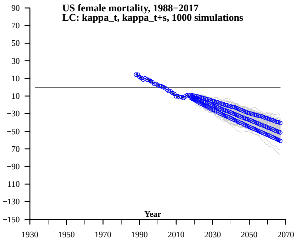


Lee-Carter model fitted to mortality in base period: 1988-2017 (focus on more recent trends)

$$\log m_{x,t} = \alpha_x + \beta_x \kappa_t + \epsilon_{x,t}$$





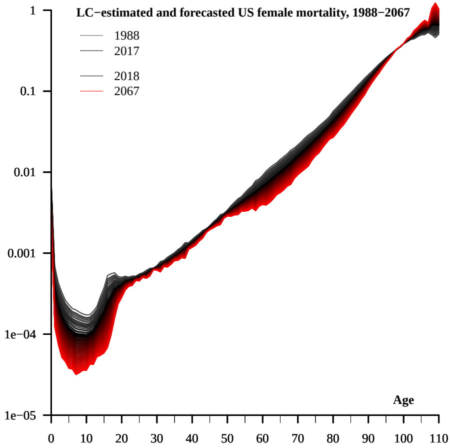


- α_x is the general shape of mortality across age x

- β_x is the change of mortality at age x
- $\beta_x > 0$: mortality decline, $\beta_x < 0$: mortality increase

- κ_t is an index of the level of mortality over time t
- direction and slope indicate moderate mortality decline

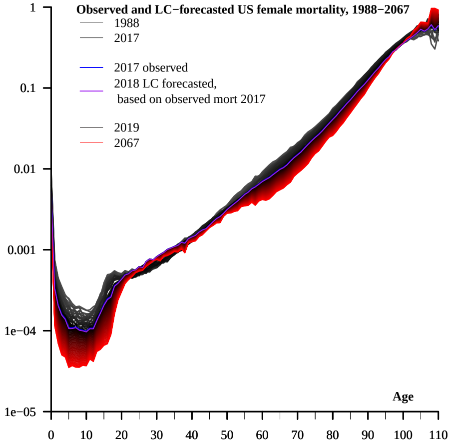
Forecasting US female mortality 50 years ahead using base period 1988-2017

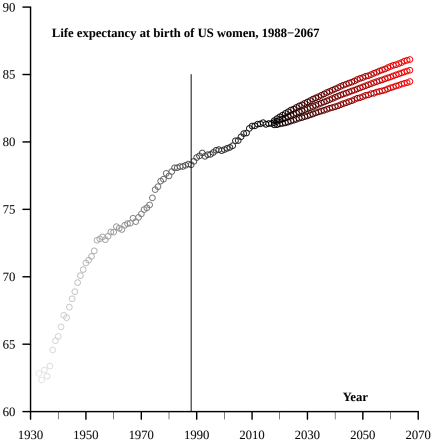


κ_t point estimates are based on median of 1000 simulated trajectories

Forecasting US female mortality 50 years ahead using base period 1988-2017, **corrected for jump-off-bias:**

$$\log m_{x,t} = m_{x,2017} + \beta_x \kappa^*_t + \epsilon_{x,t}$$





Think about it.

What are the benefits of the Lee-Carter model?

What trends does the Lee-Carter model capture?

What are the caveats concerning the Lee-Carter model?

What trends does it not capture?

What you have learned today about mortality forecasting

- Describe different approaches to forecast mortality
- Explain the method of Lee and Carter
- Reflect on pros and cons of the Lee-Carter method

Learning material for today's class

You can find the learning material of today (slides and hands-on exercise) on `https://github.com/christina-bohk-ewald/2020-IMPRS-PHDS-forecasting-mortality`.

Recommended learning material for today's class

- **Lee, R. D., & Carter, L. R. (1992)**
  Modeling and forecasting U.S. mortality. Journal of the American Statistical Association, 87(419), 659-671.
- **Booth, H. (2006)**
  Demographic forecasting: 1980 to 2005 in review. International Journal of Forecasting, 22(3), 547-581.
- **Booth, H., & Tickle, L. (2008)**
  Mortality modelling and forecasting: A review of methods. Annals of Actuarial Science, 3(1-2), 3-43.

Recommended learning material for today's class

- **Bohk, C., & Rau, R. (2016)**
  Changing mortality patterns and their predictability: the case of the United States. In Dynamic Demographic Analysis (pp. 69-89). Springer, Cham.
- **Rau, R., Bohk-Ewald, C., Muszyńska, M., & Vaupel, J. (2017)**
  Visualizing Mortality Dynamics in the Lexis Diagram (Vol. 44). Springer.
- **Bohk-Ewald, C., Ebeling, M., & Rau, R. (2017)**
  Lifespan disparity as an additional indicator for evaluating mortality forecasts. Demography, 54(4), 1559-1577.
- **Bohk-Ewald, C., Li, P., & Myrskylä, M. (2018)**
  Forecast accuracy hardly improves with method complexity when completing cohort fertility. Proceedings of the National Academy of Sciences, 115(37), 9187-9192.

Recommended learning material for today's class

- **UNWPP2019**: `https://population.un.org/wpp/`
  Publications, Graphs, & Data files.

- **Raftery, A. E., Gerland, P., and Ševčíková, H. (2013)**
  Bayesian probabilistic projections of life expectancy for all countries. Demography, 50(3), 777–801.
- **Alho, J. and Spencer, B. (1997)**
  The practical specification of the expected error of population forecasts. Journal of Official Statistics, 13(3), 203–225.
- **Preston, S., Heuveline, P., and Guillot, M. (2000)**
  Demography: measuring and modeling population processes
  Blackwell Publishers Ltd.
- **Alho, J. and Spencer, B. (2006)**
  Statistical demography and forecasting
  Springer Science & Business Media.

So close before lab session

Time to let it sink in and, perhaps, for a cup of tea :-)

...before we continue with the lab session.

See you in 10 minutes.

For the lab: please make sure to have installed R package `fds`. And to have at hand your username and password for the Human Mortality Database.

Today's lab session

Forecast US mortality for women and men 50 years ahead:

1. Load mortality data from the Human Mortality Database.
2. Implement and use the Lee-Carter method to fit and forecast mortality for US women and men 50 years ahead based on base periods of different length (i.e., taking the last 30, 50, 70, and 85 years).
3. Analyze and compare the LC mortality forecasts for US women and men. Explain the impact of the length of the base period on the US mortality forecast.
4. Thinking about the basic procedure of the LC method and its underlying assumptions, how plausible are the US mortality forecasts and how could they be improved?

Please make sure to have installed R package `fds`. And to have at hand your username and password for the Human Mortality Database.

Methodological and technical details on generating mortality forecasts with the LC method

1. Fit the Lee-Carter model to mortality by age and time in base period.
2. Forecast mortality by age s years ahead.

1. Fit LC model to mortality in base period in 8 steps

1. Put mortality rates $m_{x,t}$ in matrix by age (rows) and year (columns)
2. Calculate natural logarithm of mortality rates: $\ln m_{x,t}$
3. Calculate α_x as mean mortality over time for each age x
4. Calculate central (or normalized) log mortality rates $M_{x,t}$ as difference between $\ln m_{x,t}$ and α_x
5. Estimate β_x and κ_t applying singular value decomposition (SVD) to central log mortality rates ($M_{x,t}$)

1. Fit LC model to mortality in base period in 8 steps

5. Estimate β_x and κ_t applying singular value decomposition (SVD) to central log mortality rates ($M_{x,t}$)
   1. $svd(M_{x,t}) = UDV$; with $M[x,t]$, $U[t,t]$, $D[1,t]$, and $V[x,x]$
   2. $\beta_x = \frac{V[,1]}{\sum V[,1]}$
   3. $\kappa_t = D[1,1]U[,1]sumV[,1]$
   4. Check that $\sum \beta_x = 1$ and $\sum \kappa_t = 0$

1. Fit LC model to mortality in base period in 8 steps

6. Plot α_x, β_x, κ_t for plausibility checks

7. If desired, re-fit κ_t to e.g. total death counts, deaths counts by age, life expectancy with iterative process.

8. Fit mortality in base period putting parameter values into LC model function: $\log \hat{m}_{x,t} = \alpha_x + \beta_x \kappa_t$

2. Forecast mortality by age s years ahead in 3 steps

1. **Fit estimated κ_t in base period** using a time series model, e.g. ARIMA(0,1,0)
   - Lee and Carter suggest random walk with drift, ARIMA(0,1,0): $\kappa_{t+s} = \kappa_{t-1} + \delta + \epsilon_t$
     - δ is a drift term
     - ϵ_t is an error term

2. Forecast mortality by age s years ahead in 3 steps

2. **Forecast κ_t s years ahead** with fitted time series model:
   $\kappa_{t+s} = \kappa_{t-1} + \delta + \epsilon_t$
   - Point and interval forecasts of κ_t can be based on simulation:
     - ★ Simulate N trajectories for κ_{t+s}
       using estimate of δ (of fitted ARIMA(0,1,0) model)
     - ★ Draw $\epsilon_t \sim \mathcal{N}(0, \sigma^2_{\epsilon_t})$
       for each trajectory and year t,
       with $\sigma^2_{\epsilon_t}$ being the estimated variance of the residuals
       of the fitted ARIMA(0,1,0) model
     - ★ Determine median and 80% prediction intervals for κ_{t+s}
       using quantiles (0.1, 0.5, and 0.9)
       of the distribution comprising the N trajectories

2. Forecast mortality by age s years ahead in 3 steps

3. **Forecast age-specific mortality s years ahead**
   inserting parameter values into LC model function:
   $\log m_{x,t+s} = \alpha_x + \beta_x \kappa_{t+s} + \epsilon_{x,t}$

Many R implementations of the Lee-Carter model out there

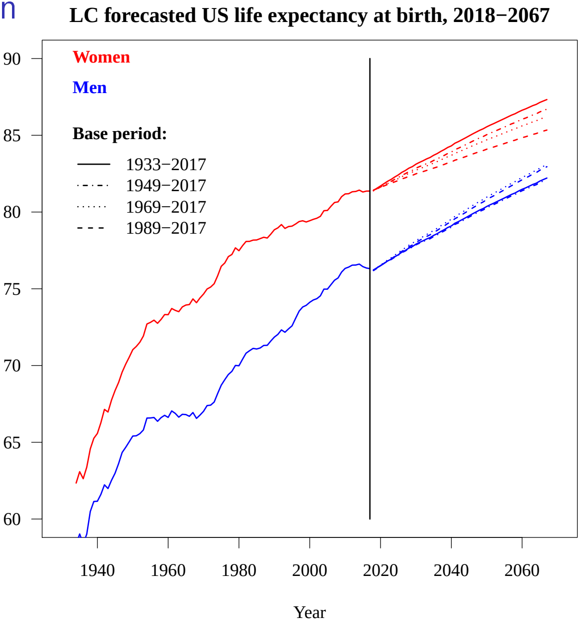
R-packages and related functions (non-exhaustive):

- `demography: lca(), forecast.lca()`
- `StMoMo:`
  Vignette available at `https://cran.r-project.org/web/packages/StMoMo/vignettes/StMoMoVignette.pdf`
- `MortalityForecast: model.LeeCarter(), predict()`
- `ilc`
- `LifeMetrics`
  available at `http://www.macs.hw.ac.uk/~andrewc/lifemetrics/`

→ helpful to apply LC method but please do not use them as black box

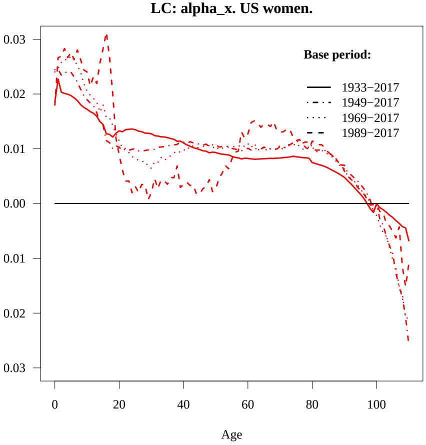
→ To **really understand LC method** it is good to **implement it yourself**!

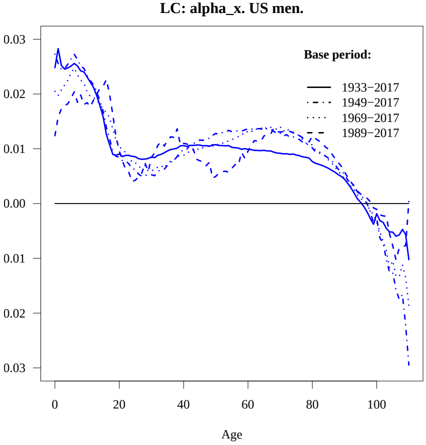
Lab session


LC forecasted US life expectancy at birth, 2018−2067
Women
Men
Base period:
1933−2017
1949−2017
1969−2017
1989−2017
90
85
80
75
70
65
60
1940
1960
1980
2000
2020
2040
2060
Year

Lab session

Note that these plots show β_x, not α_x:





Lab session

What do yo think:

Thinking about the basic procedure of the LC method
and its underlying assumptions,
how plausible are the US mortality forecasts
and how could they be improved?

Extensions of the LC method

- Re-fit κ_t in step 1.7 to match
  - Total death counts (e.g. Lee and Carter (1992))
  - Life expectancy at birth (e.g. Lee and Miller (2001))
  - Death counts by age and time (e.g. Booth et al. (2002))
  - Lifespan inequality (e.g. Rabbi and Mazzuco (2020))
  - ...
- Consider mortality of multiple populations (e.g. Li and Lee (2005))
- Consider cohort effects (e.g. Renshaw and Haberman (2006))
- Consider multiple functional principal components to capture non-random patterns (e.g. Hyndman and Ullah (2007))
- Consider time-variant age pattern of mortality change (e.g. Li et al. (2013))
- ...

$\rightarrow$ Overview of some LC extensions in e.g. Booth (2006)

Other directions in mortality forecasting

Coherent forecasting to consider developments of other populations:

- Multiple countries (e.g. Li and Lee (2005))
- Women and men (e.g. Hyndman et al. (2013))
- Bayesian methodology of recent UNWPP (e.g. Raftery et al. (2013))

Approaches to consider mortality attributable to health-related behavior:

- Smoking (e.g. Janssen et al. (2013))
- ...

Approaches to capture more closely mortality by age over time:

- Rates of mortality improvement (e.g. Bohk-Ewald and Rau (2017))
- Distribution of ages at death (e.g. Basellini and Camarda (2019))
- ...

...

Other directions in mortality forecasting

Other current / future directions:

- Account for climate change (e.g. air quality)
- Account for COVID-19 pandemic
- ...

Other directions in mortality forecasting

What would you do?

How would you forecast mortality?

Take your time to reflect on it...

Thank you for your time and attention!

Bohk-Ewald@demogr.mpg.de