Demographic Forecasting

Assignment

Ugofilippo Basellini*

7 May 2025

Exercise 1

Load the mortality data `MORTSWE.Rdata`, and focus on male mortality from 1950 onwards for newborns (age 0). Fit and forecast mortality up to 2050 using two separate models:

- a linear model for death rates
- a generalized linear model for deaths, using exposures as an offset

Plot the fitted and forecast rates on a normal and log scale. What difference do you see between the two approaches?

Exercise 2

Load the mortality data `MORTSWE.Rdata`, and focus on male mortality from 1950 onward for those aged 0-100 (i.e. exclude the 101-110+ age groups). Fit and forecast mortality up to 2050 using a generalized linear model for deaths, using exposures as an offset, for all age groups available (a `for` loop is a convenient way to do so). Compare the observed age-pattern of mortality against the fitted and forecast one. What differences do you observe?

Exercise 3

Load the mortality data `MORTSWE.Rdata`, and focus on mortality from 1950 to 2000 for males aged 20 years. Fit and forecast mortality up to 2022 using two separate models:

- a random walk with drift model for log-rates
- the best possible ARIMA model for log-rates

Plot the forecast rates against the subsequently observed data in 2001-2022. Which model seems to forecast mortality better?

Exercise 4

Load the mortality data `MORTSWE.Rdata`, and focus on male mortality from 1950 onwards for those aged 0-100 (i.e. exclude the 101-110+ age groups). Fit and forecast mortality up to 2050 using the best possible ARIMA model for log-rates for all age groups available (a `for` loop is a convenient way to do so). Compare the observed age-pattern of mortality against the forecast one. Does it seem like a reasonable pattern of mortality?

*Max Planck Institute for Demographic Research, basellini@demogr.mpg.de

[Hint: you will need to replace minus infinite values of $\ln(m_{x,t})$ with `NA` for some age groups]

Exercise 5

Following up on Exercise 4, compute simulations for the future paths of all $\ln(m_{x,t})$, and combine them together to compute a forecast of life expectancy. Use the provided function `LifetableMX.R` to construct a life table from your (log) rates.

[Hint: you will need to create an array, rather than a matrix, where to store your simulations of $\ln(m_{x,t})$]

Exercise 6

Load the mortality data `MORTSWE.Rdata`, and focus on male mortality from 1950 onwards for ages $30 \leq x \leq 100$. Fit and forecast adult mortality up to 2050 using the parametric Gompertz model, fitting a model for each year independently [*hint*: this is a generalized linear model with deaths as response variable, exposures as an offset, and an intercept and age as covariates]. To forecast, fit the most appropriate ARIMA(p,d,q) models to the time-series of the two estimated parameters. Compute the 95% prediction intervals for life expectancy using simulations from the two ARIMA models. [*hint*: you can use the functions inside the `LifetableMX.R` code for constructing life tables and deriving estimates of life expectancy]

Exercise 7

Load the mortality data `MORTSWE.Rdata`, and focus on male mortality from 1950 onwards for ages $x \geq 30$. Fit and forecast adult mortality up to 2050 using two different approaches:

- a Gompertz model with log-linear time trend, i.e. $\ln(m_{x,t}) = \beta_0 + \beta_1 x + \beta_2 t$
- a Gompertz model with flexible time index, i.e. $\ln(m_{x,t}) = \beta_0 + \beta_1 x + \kappa_t$

Plot the life expectancy forecasts of the two models (no need to derive PIs). [*hint*: you can use the functions inside the `LifetableMX.R` code for constructing life tables and deriving estimates of life expectancy]

Exercise 8

Load the mortality data `MORTSWE.Rdata`, and focus on male mortality from 1950 to 2000 for ages $x \leq 100$. Fit and forecast mortality with the LC method, and produce forecasts up to 2022. Evaluate the forecast accuracy of the method by computing a point forecast accuracy measure of your choice on a single measure of interest (e.g. log-rates or e_o).

Exercise 9

Consider again the setting of Exercise 8 (i.e. male mortality from 1950 to 2000 for ages $x \leq 100$). Compare the forecast accuracy of two different methods that we have seen during this course. Which model is more accurate?