European Doctoral School of Demography 2019-20

Population Projections

Lecture 2

Ugofilippo Basellini

basellini@demogr.mpg.de

MAX-PLANCK-INSTITUT FÜR DEMOGRAFISCHE FORSCHUNG
MAX PLANCK INSTITUTE FOR DEMOGRAPHIC RESEARCH

ined
INSTITUT NATIONAL D'ÉTUDES DÉMOGRAPHIQUES

Online course
2nd June 2020

Brief course summary

Lecture 1: introduction & first simple model of population projections

Lecture 2: cohort component method

Lecture 3: matrix projections

Lecture 4: more matrix projections

Small recap

In Lecture 1, we have seen a first model of population projections based on crude demographic rates

- however, crude rates are very sensitive to age-compositional effects
- the model is reasonable only when:
  - demographic components do not vary much over projection period, *and*
  - the age distribution remains constant
- this is however unlikely, as the age distribution is shaped by fertility, mortality and migration - processes that vary by age
- users are typically interested in age and sex breakdown of projections
- today, we will look at a second model of population projections that explicitly accounts for age distributions

$\Rightarrow$ cohort component method

Introduction I

The cohort component method:

- as reported by Smith and Keyfitz (1977), it can be traced back to Cannan (1895), although it was independently developed by Whelpton (1928, 1936)
- nearly the only method used for population projections, almost universal consensus among social scientists
- employed by the United Nations for their WPP population estimates and projections, as well as by several statistical agencies (i.e. Federal Statistical Office of Germany, EUROSTAT, ...)
- model's intuition: segment the population into different groups exposed to specific fertility, mortality and migration "risks", and compute the changes over time in each group
- typical groups: age and sex; can be extended to include race, nationality, location (regions, rural/urban), educational attainment, ...

Introduction II

The cohort component method:

- discrete-time model (as opposed to continuous-time constant exponential model seen in Lecture 1)
- projection period divided into time intervals of the same length as the age intervals employed
- for each subgroup, three main steps for each projection interval:
  - estimate the number of people still alive at the start of next interval (*mortality*)
  - compute number of births, aggregate them across groups and compute how many survive to the next interval (*fertility* and *infant mortality*)
  - add immigrants and subtract emigrants, compute births from immigrants, project number of migrants and their newborns in the next interval (*migration*, *fertility* and *mortality*)

Cohort component - step 1

Estimate the number still alive at the start of next interval (*mortality*)

- if groups are only age and sex, only need single decrement life tables by sex
- use survivorship ratios to compute survivors in the next period, assigned to the next age group (because age and time intervals congruent)
- if more subgroups, need to take into account transitions to different groups - multistate projections (see, e.g., Rogers 1995)

Closed female population

Simplified example to illustrate the methodology: a closed female population, broken down by 5 year age-groups, closed to migration. What we need:

- ${}_5N_x^F(t)$, the number of females in age group x to $x+5$ at time t
- ${}_5L_x^F$, the number of person-years lived by females between ages x and $x+5$

To ease notation, we drop the left subscript as we consider 5y age groups throughout.

We can then compute survivorship ratios s_x and the projected population in each age group (except the youngest and oldest):

$$\begin{aligned} s_x^F &= \frac{L_{x+5}^F}{L_x^F} \\ N_x^F(t+5) &= N_{x-5}^F(t)\, s_{x-5}^F \end{aligned} \tag{1}$$

Closed female population - exercise

Exercise

Using the provided `dta.swe.1993.Rdata` dataset, compute the projected population by age groups. Do not worry about the first and last age groups, we will adjust them later.

Reminder:

$$
\begin{aligned}
N_x^F(t+5) &= N_{x-5}^F(t)\, s_{x-5}^F \\
&= N_{x-5}^F(t) \frac{L_x^F}{L_{x-5}^F}
\end{aligned}
$$

Closed female population - one possible solution

Example

```
rm(list = ls())
library(tidyverse)
load("dta.swe.1993.Rdata")
## project age groups forward
dta.swe <- as_tibble(dta.swe) %>%
    mutate(sFx=lead(LFx)/LFx,
           NFx5=lag(NFx*sFx))
head(dta.swe[,c(1:4,8,9)])
```

```
# A tibble: 6 x 6
Age  AgeGroup     NFx    LFx   sFx     NFx5
<dbl> <fct>      <dbl>  <dbl> <dbl>    <dbl>
0     00-04     293395 497487 0.999       NA
5     05-09     248369 497138 1.00   293189.
10    10-14     240012 496901 0.999 248251.
15    15-19     261346 496531 0.999 239833.
20    20-24     285209 495902 0.999 261015.
25    25-29     314388 495168 0.998 284787.
```

Closed female population - last age group

For the open-ended age group, we should also consider the number of survivors that were already present in the group at the beginning of the interval. Hence, combining two age groups:

$$ {}_{\infty}N_x^F(t+5) = \left(N_{x-5}^F(t) \frac{L_x^F}{L_{x-5}^F} \right) + \left({}_{\infty}N_x^F(t) \, \frac{T_{x+5}^F}{T_x^F} \right) \tag{2} $$

This requires the open-ended age group in the life table to start at an age (at least) five years older than the population. If we do not have this information, we should use:

$$ {}_{\infty}N_x^F(t+5) = \left(N_{x-5}^F(t) + {}_{\infty}N_x^F(t) \right) \frac{T_x^F}{T_{x-5}^F} \tag{3} $$

We should thus adjust $s_{80}^F = \frac{T_{85}^F}{T_{80}^F} = \frac{L_{85}^F}{L_{80}^F + L_{85}^F}$

Last age group - exercise

Exercise

Adjust the last age group of your projection.

Reminder:

$$s_{80}^F = \frac{L_{85}^F}{L_{80}^F + L_{85}^F}$$

$$_{\infty}N_{85}^F(t+5) = \left(N_{80}^F(t) + {}_{\infty}N_{85}^F(t)\right) s_{80}^F$$

Last age group - one possible solution

Example

```
## adjusting the last age group
dta.swe <- dta.swe %>%
    mutate(sFx=ifelse(test = Age==80,
                      yes  = lead(LFx)/(LFx + lead(LFx)),
                      no   = sFx),
           NFx5=ifelse(test = Age==85,
                       yes  = (NFx+lag(NFx))*lag(sFx),
                       no   = NFx5))
tail(dta.swe[,c(1:4,8,9)],n = 3)
```

```
# A tibble: 3 x 6
Age AgeGroup    NFx    LFx    sFx    NFx5
<dbl> <fct>    <dbl>  <dbl>  <dbl>   <dbl>
75    75-79    183654 350358  0.775 194419.
80    80-84    141990 271512  0.518 142324.
85     85+     112424 291707   NA   131768.
```

Cohort component - step 2

Compute number of births, aggregate them across groups and compute how many survive to the next interval (*fertility* and *infant mortality*)

- ideally: model explicitly couple and union creation and dissolution, births are a by-product of these
- in practice: "female-dominant" approach, births produced by women only
- apply fertility rates to women, disaggregate by sex using sex-ratio at birth
- if more subgroups, additional layer of difficulty - typically assume birth belonging to same group as mother

Adjusting the first age group I

Let F_x denote the fertility rate at age x. During the projection interval, the number of births to women aged x to $x + 5$ is then:

$$B_x[t, t+5] = F_x \underbrace{5 \left[\frac{N_x^F(t) + N_x^F(t+5)}{2} \right]}_{\text{approximation of person-years lived at ages } x \text{ to } x+5}$$

Using Eq. (1), the total number of births during the period, $B[t, t+5]$ can be written as:

$$B[t, t+5] = \sum_x \frac{5}{2} F_x \left(N_x^F(t) + N_{x-5}^F(t) \frac{L_x^F}{L_{x-5}^F} \right)$$

For now, we are interested in female births $\Rightarrow$ use sex-ratio at birth (SRB):

$$B^F[t, t+5] = \frac{1}{1 + SRB} B[t, t+5]$$

Adjusting the first age group II

- number of females in the first age group is obtained by surviving female births through time $t+5$
- assuming that births are distributed evenly during projection period:

$$N_0^F(t+5) = B^F[t, t+5]\frac{L_0^F}{5\,\ell_0} \tag{4}$$

where ℓ_0 is the life-table radix.

- equivalently, we can rewrite Eq. (4) as:

$$N_0^F(t+5) = \sum_x N_x^F(t) b_x^F \tag{5}$$

where $b_x^F = \frac{1}{1+SRB}\frac{L_0^F}{2\ell_0}\left(F_x + s_x^F\, F_{x+5}\right)$.

First age group - exercise

Exercise

Adjust the first age group of your projection, assuming a sex-ratio at birth of 1.05.

Two possible solutions: we can compute

$$b_x^F = \frac{1}{1+SRB}\frac{L_0^F}{2\ell_0}\left(F_x + s_x^F\, F_{x+5}\right),$$

$$N_0^F(t+5) = \sum_x N_x^F(t) b_x^F$$

Alternatively, we can use

$$B_x[t,t+5] = F_x\, 5\left[\frac{N_x^F(t)+N_x^F(t+5)}{2}\right]$$

$$N_0^F(t+5) = \frac{1}{1+SRB}\,\frac{L_0^F}{5\,\ell_0}\sum_x B_x[t,t+5]$$

First age group - one possible solution

Example

```
srb <- 1.05
fact.srb <- 1/(1+srb)
l0 <- 1e5
LF0 <- dta.swe$LFx[dta.swe$Age==0]
dta.swe <- dta.swe %>%
    mutate(bFx=fact.srb * LF0 / (2*l0) * (Fx + sFx*lead(Fx)),
           Bx=Fx*5*(NFx+NFx5)/2,
           NFx5=ifelse(test = Age==0,
                       yes  = fact.srb * LF0 / (5*l0) * sum(Bx,na.rm = T),
                       no   = NFx5)
)
dta.swe$NFx5[1]
## compare with second formula
sum(dta.swe$NFx * dta.swe$bFx,na.rm = T)
```

```
[1] 293573.8
[1] 293573.8
```

Plotting your results - exercise

Exercise

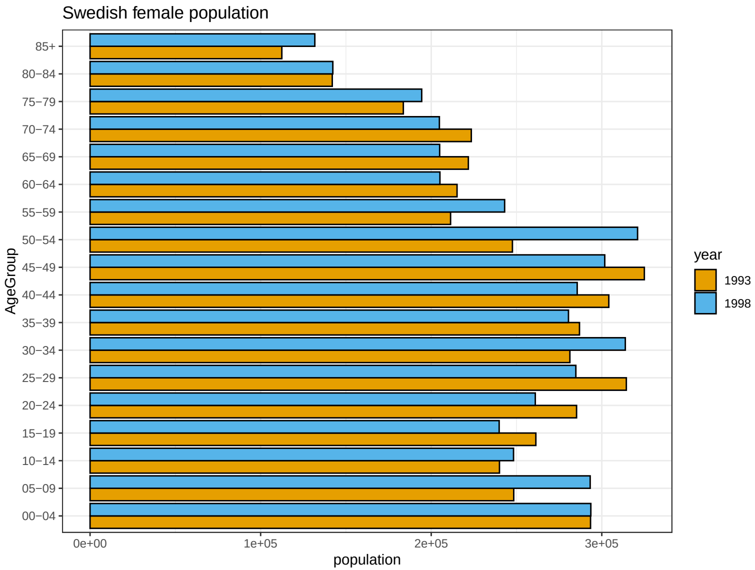
Now, let's plot our results. It would be nice to use a pyramid, and to compare the starting population with the projected one.

Example

```
## long data
dta.swe.l <- dta.swe %>%
    select(AgeGroup,NFx,NFx5) %>%
    rename('1993'=NFx,'1998'=NFx5) %>%
    pivot_longer(-AgeGroup,names_to = "year",values_to = "population")

## plotting
ggplot(dta.swe.l,aes(x=AgeGroup,y=population,fill=year)) +
    geom_bar(stat = "identity",position = "dodge",color = "black") +
    coord_flip() +
    theme_bw() +
    ggtitle("Swedish female population") +
    scale_fill_manual(values=c("#E69F00", "#56B4E9"))
```

Plotting your results - one possible solution



Two-sex closed population

The male population can be projected with the same formulas and using male survivorship ratios

$$s_x^M = \frac{L_{x+5}^M}{L_x^M} \quad ; \quad s_{80}^M = \frac{T_{85}^M}{T_{80}^M}$$

$$N_x^M(t+5) = N_{x-5}^M(t)\, s_{x-5}^M$$

$$_\infty N_{85}^M(t+5) = \Big(N_{80}^M(t) + {}_\infty N_{85}^M(t)\Big)\, s_{80}^M$$

$$b_x^M = \frac{SRB}{1+SRB}\,\frac{L_0^M}{2\ell_0}\left(F_x + s_x^F\, F_{x+5}\right),$$

$$N_0^M(t+5) = \sum_x N_x^F(t) b_x^M$$

Two-sex closed population - exercise

Exercise

Using the same `dta.swe.1993.Rdata` dataset, compute the male projected population by age groups, adjusting the first and last age groups. Then save your data, as this will be useful for tomorrow's lecture.

Reminder:

$$s_x^M = \frac{L_{x+5}^M}{L_x^M} \quad ; \quad s_{80}^M = \frac{T_{85}^M}{T_{80}^M}$$

$$N_x^M(t+5) = N_{x-5}^M(t)\, s_{x-5}^M$$

$$_\infty N_{85}^M(t+5) = \left(N_{80}^M(t) + {}_\infty N_{85}^M(t)\right) s_{80}^M$$

$$b_x^M = \frac{SRB}{1+SRB} \frac{L_0^M}{2\ell_0} \left(F_x + s_x^F\, F_{x+5}\right),$$

$$N_0^M(t+5) = \sum_x N_x^F(t) b_x^M$$

Two-sex closed population - one possible solution

Example

```
fact.srb.M <- srb/(1+srb)
LM0 <- dta.swe$LMx[dta.swe$Age==0]
dta.swe <- dta.swe %>%
    mutate(sMx=lead(LMx)/LMx,NMx5=lag(NMx*sMx),
           sMx=ifelse(test = Age==80,
                      yes  = lead(LMx)/(LMx + lead(LMx)),
                      no   = sMx),
           NMx5=ifelse(test = Age==85,
                       yes  = (NMx+lag(NMx))*lag(sMx),
                       no   = NMx5),
           bMx=fact.srb.M * LM0 / (2*10) * (Fx + sFx*lead(Fx)),
           NMx5=ifelse(test = Age==0,
                       yes  = sum(bMx*NFx,na.rm = T),
                       no   = NMx5))
head(dta.swe[,c(1:3,6,9,13)],n=3)
```

Age	AgeGroup	NFx	NMx	NFx5	NMx5
<dbl>	<fct>	<dbl>	<dbl>	<dbl>	<dbl>
0	00-04	293395	310189	293574.	307798.
5	05-09	248369	261963	293189.	309904.
10	10-14	240012	252046	248251.	261800.

Some final remarks

- cohort component method as the most employed model for population projections
- takes into account age composition of populations
- projection of age groups rather straightforward, except for the youngest and the eldest
- projected population from one interval becomes the baseline for next projection
- several projections can become cumbersome if done one at a time

$\Rightarrow$ matrix algebra can help us to speed this up (tomorrow's lecture)

Assignment

Exercise #2

Take the population for any country (for example, use data from the HMD). Divide the population into 5 years age groups and project the population for 5 years ahead for each sex separately. Plot and compare your results with the baseline population.

Exercise #3

Show how we can derive Eq. (5) starting from Eq. (4)

References

- Cannan, E. (1895). The probability of a cessation of the growth of population in England and Wales during the next century. *The Economic Journal*, **5**(20), 505–515.
- Preston, S. H., Heuveline, P., and Guillot, M. (2001). *Demography. Measuring and Modeling Population Processes*. Blackwell.
- Rogers, A. (1995). *Multiregional Demography*. John Wiley & Sons.
- Smith, D.P. and Keyfitz, N. (*eds.*) (1977). *Mathematical Demography: Selected Papers*. Berlin: Springer Verlag
- Whelpton, P.K. (1928). Population of the United States, 1925 to 1975. *American Journal of Sociology*, **31**, 253–270.
- Whelpton, P.K. (1936). An empirical method of calculating future population. *Journal of the American Statistical Association*, 31(195), 457–473.