Session II:
Overview of projection methods

7 March 2016

Cheryl Sawyer, Lina Bassarsky

Population Estimates and Projections Section

www.unpopulation.org

Outline

I. Choosing a projection method
II. Trend extrapolation methods
III. Cohort-component projections
IV. Country presentations on national experience with population projections

I. Choosing a projection method

Criteria for choosing a projection method

1. Desired outputs
   - Length of time horizon → short-term (e.g. next 5 years), medium-term, long-term, etc.
   - Amount of subject detail → total population, broad age groups, five-year age groups, single years of age, subgroups

2. Input data
   - Data availability
   - Data quality
   - Missing data → plausible assumptions

3. Projection tools

II. Trend extrapolation methods

- Based on the continuation of observable historical trends. Mathematical models are fitted to historical data, then used to project population values
- Historical data may be "borrowed" from population growth observed in other populations
- Usually applied to the figure of total population only

- Most appropriate when the total population growth is fairly regular, or when the past is a good predictor of the future trends
  - ▶ when vital rates and age structure can be assumed to be constant
- …or if the only thing one knows about the population is its total size.

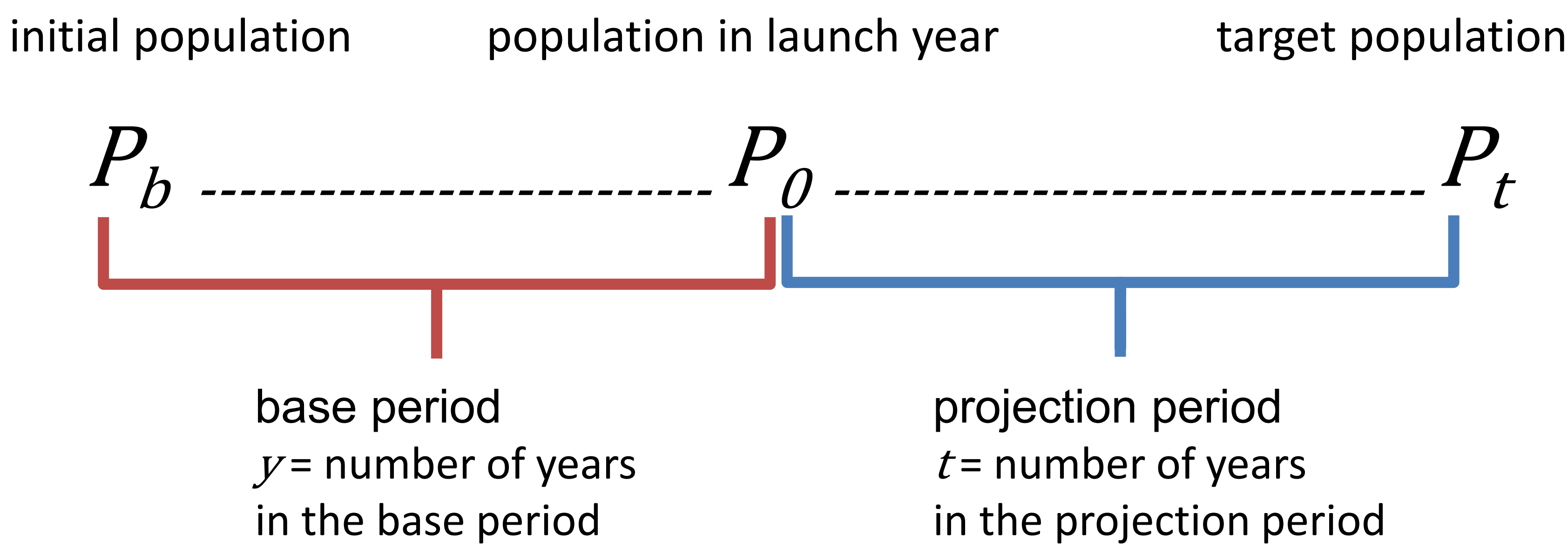
- Not appropriate
  - where the present age structure has marked peculiarities
  - where important or rapid socio/economic changes are foreseen or expected on the grounds of current conditions of experiences of other populations

- Usually not useful for policy-related purposes.
- It can be used for making quick estimates or for illustrative or comparison purposes, but is not recommended for making official population projections.

- Methods that require data for 2 dates → base period (an initial year and the launch year)
  - Linear growth
  - Geometrical growth
  - Exponential growth
- Methods that require data for more dates
  - Logistic curve
  - Other (linear trend, polynomial curve, ARIMA)
- Ratio extrapolation methods → population of a smaller area is expressed as a proportion of its “parent” larger population area
  - Constant share, shift share, share of growth

Source: Siegel, Jacob S. & Swanson, David A. (editors) (2004) *The Methods and Materials of Demography,* Elsevier Academic Press, second edition, Chapter 21

[Terminology for trend extrapolation methods section]



Linear growth

- Assumptions:
  - Population will change by the same (constant) absolute amount of change per unit of time (e.g. a year) ▶ a decreasing relative change.
  - Change is experienced at the end of unit of time
- Steps:
  - Compute the average absolute change Δ during the base period, apply the same amount for each unit of time through the target year.

Linear growth

- Computation:

Δ = annual amount of population change:

$$\Delta = (P_0 - P_b)/y$$

P_t = population in the target year:

$$P_t = P_0 + \Delta t$$

Where:

P_0 = population in the launch year

P_b = initial population (base period)

y = number of years in the base period

t = number of years in the projection (between the launch and the target years)

Geometric growth

- Assumptions:
  - Population will change by the same percentage rate per unit of time as during the base period
  - Change is occurring at discrete intervals

- Steps:
  - Compute the average annual geometric rate of change r, apply the rate to the launch population.

Geometric growth

- Computation:

r = average annual geometric rate of change

$$r = \left[\left(\frac{P_0}{P_b}\right)^{1/y}\right] - 1$$

P_t = population in the target year

$$P_t = P_0 * (1 + r)^t$$

Where:

P_0 = population in the launch year

P_b = initial population (base period)

y = number of years in the base period

t = number of years in the projection (between the launch and the target years)

Exponential growth

- Assumptions:
  - Yields similar results to geometric growth, BUT change occurs continuously.
- Steps:
  - Compute the average annual exponential rate of change r, apply the rate to the launch population.

Exponential growth

- Computation:

r = average annual exponential rate of change

$$r = \left[ln \left(\frac{P_0}{P_b} \right) \right] / \ y$$

P_t = population in the target year

$$P_t = \ P_0 * e^{rt}$$

Where:

P_0 = population in the launch year

P_b = initial population (base period)

y = number of years in the base period

e = base of the system of natural logarithms

t = number of years in the projection (between the launch and the target years)

- It can be observed that:

$$(1 + rt) \ \leq \ (1 + r)^t \leq e^{rt}$$

Exponential growth

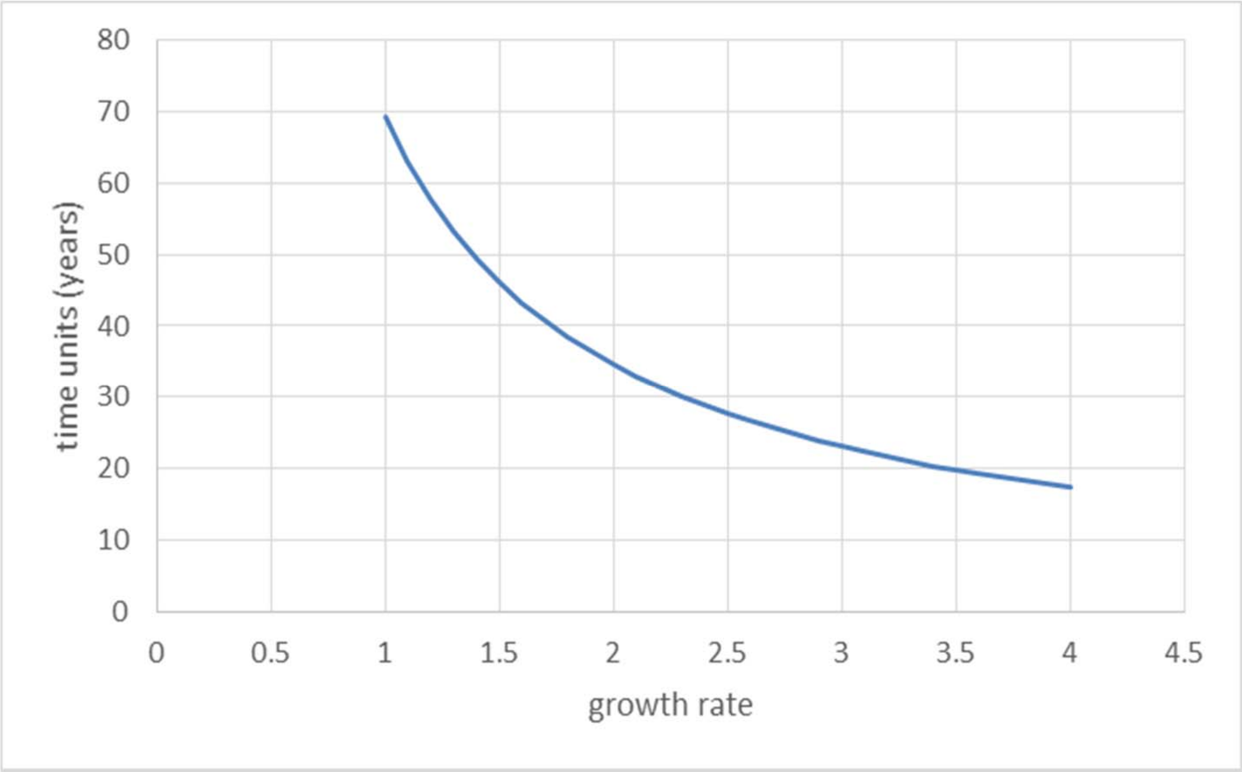
- Time required for one population (P_0) to double ($2P_0$), given a known growth rate

$$2P_0 = P_0 e^{rt}$$

$$2 = e^{rt}$$

$$\ln 2 = rt$$

$$t = \frac{\ln 2}{r}$$



Logistic growth

- The logistic function: exhibits an S-shape and describes a diffusion process where new behavior is gradually being adopted by people
- It can be used as a model representing the behaviour of various populations and demographic phenomena, such as population proportions and projection of small areas
- Population growth: trend of growth with increasing annual increments until a maximum. Beyond that point, growth diminishes until a minimum or becomes negligible.

▶ limits exist to population growth

- Currently mostly used in projection of life expectancy at birth and total fertility (growing from an initial level to an upper or lower asymptote).

Logistic growth

- Computation:

$$P_t = \frac{a}{[1 + b(e^{-ct})]}$$

Where

a, b, c → three parameters

a is the upper asymptote

b & c define the shape of the logistic curve

t is the time

Determine the magnitude of the upper asymptote (a) and the time (t) required to reach it.

- The general form of a logistic function can be expressed as:

①

$$P(t) = \frac{k}{1 + \exp[-\alpha(t - \beta)]}$$

k = asymptote (saturation level)

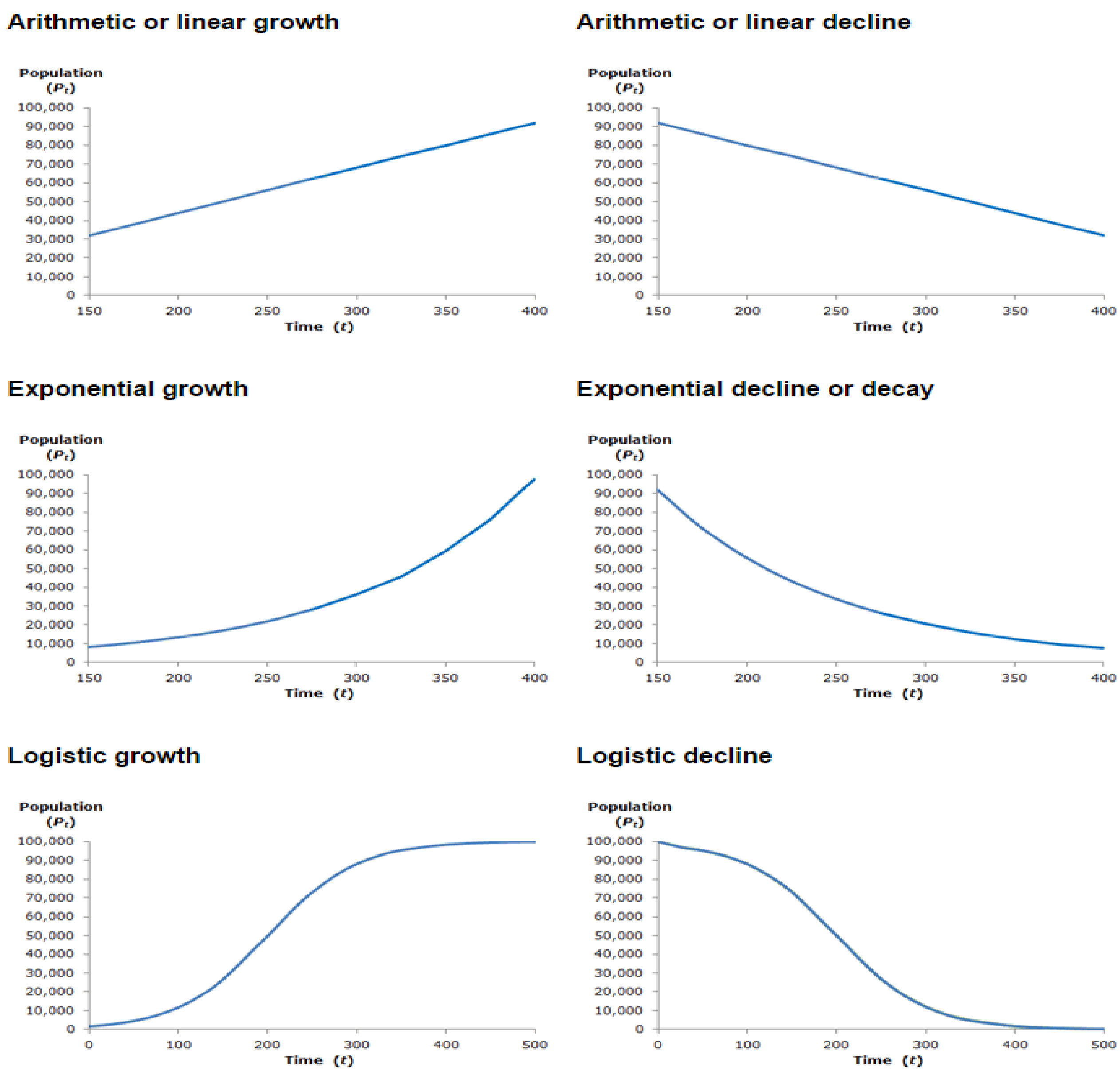
β = length of time to midpoint = $\mathbf{t_m}$

α = growth rate of the curve

②

$$P(t) = \frac{k}{1 + \exp[-\frac{Ln(81)}{\Delta t}(t - t_m)]}$$

where: $\Delta t = \frac{\ln(81)}{\alpha}$ = Duration for the growth process to proceed from 10% to 90% of (k)



Source:
http://papp.iussp.org

III. Cohort-component projections

Cohort-component method (CCM)

- Most commonly used method that accounts for age distribution
- Benefit of taking into account the age distribution

a) Population accounting concept

b) Data required

c) Model implementation, example

Population accounting concept

Projection period

P_t ________ P_{t+n}

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

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

$\mathbf{P_t}$ is the population at time t

$\mathbf{B_{t,t+n}}$ and $\mathbf{D_{t,t+n}}$ are number of births and deaths occurring between t and t + n.

$I_{t,t+n}$ and $E_{t,t+n}$ are number of immigrants and emigrants from the country during the same period

The Cohort component method (CCM)
for population estimation

The population enumerated in the first census is projected to the reference date of the second census, based on intercensal estimates. The expected population from the projection is then compared with the actual population enumerated in the second census (e.g. 10 years after).

Data required:

1. Population enumerated in two censuses by age and sex
2. Age specific fertility rates for women aged 15 to 49 (in 5-year age groups), assumed to represent the level and age structure of fertility during the intercensal period
3. Life table survival ratios for males and females, assumed to be representative of mortality conditions during the intercensal period
4. An estimate of sex ratio at birth
5. Estimates of the level and age pattern of net international migration during the intercensal period if the level of net migration is substantial

The Cohort component method (CCM)

for population ~~estimation~~ projection

at the launch year of the projection

Assumptions about trends during the projection span

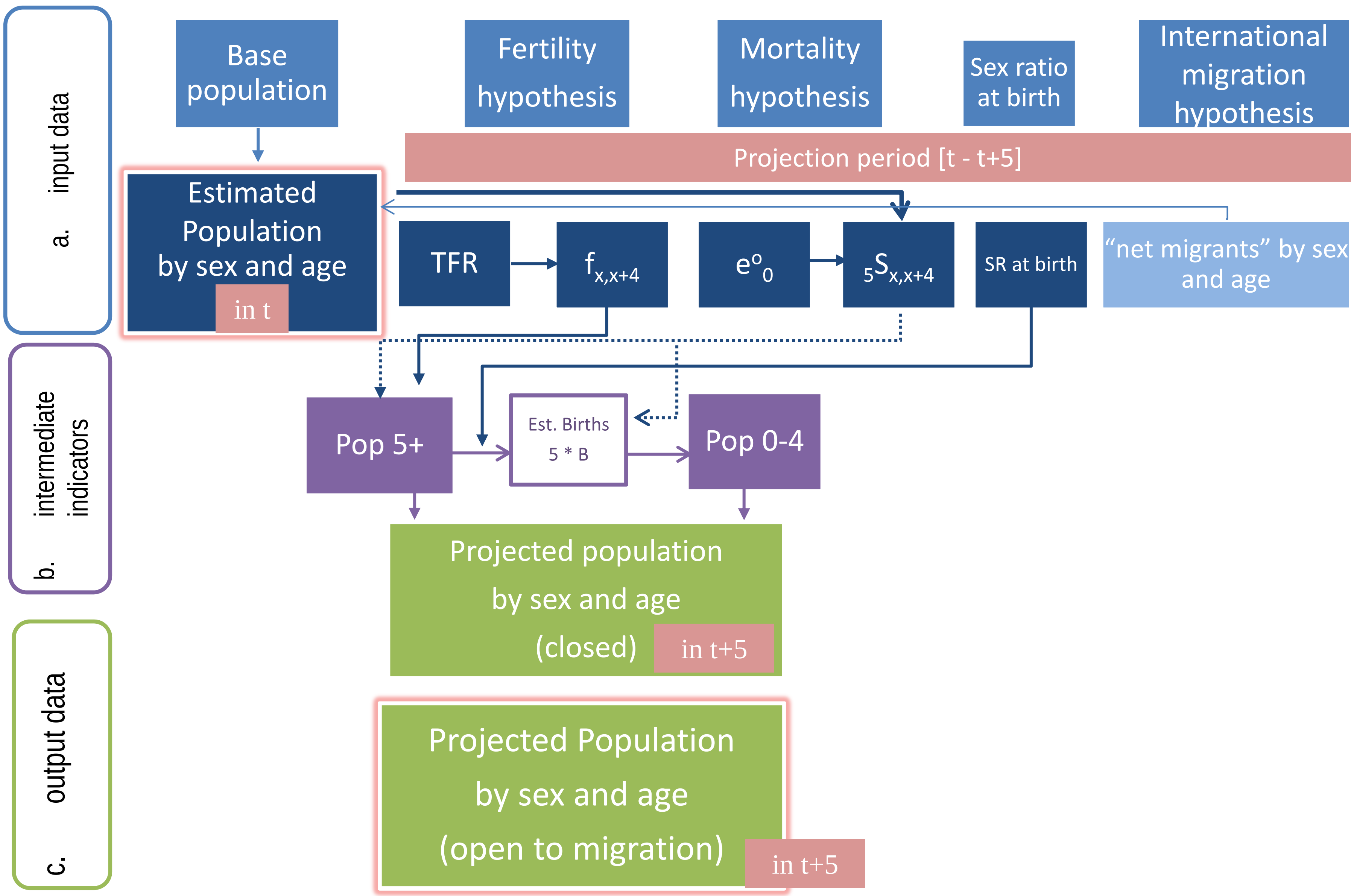
Data required:

1. Population ~~enumerated in two censuses~~ by age and sex
2. Age specific fertility rates for women aged 15 to 49 (in 5-year age groups), ~~assumed to represent the level and age structure of fertility during the intercensal period~~
3. Life table survival ratios for males and females, ~~assumed to be representative of mortality conditions during the intercensal period~~
4. An estimate of sex ratio at birth
5. Estimates of the level and age pattern of net international migration ~~during the intercensal period if the level of net migration is substantial~~

Computational steps

1. Survive the base population in t by age and sex to t+n in projection period
   Multiply each age group population by life table survival ratios (open-ended interval requires special handling)
2. Make any necessary adjustments for migration
3. Calculate the total number of births (total, by sex) during t, t+n
4. Apply survivorship to these births to determine number that survive to t+n
5. Obtain population projection at t+n
6. Repeat steps 1-5 for the next projection period

CCM


a. input data
Base population
Fertility hypothesis
Mortality hypothesis
Sex ratio at birth
International migration hypothesis
Projection period [t - t+5]
Estimated Population by sex and age
in t
TFR
$f_{x,x+4}$
e^o_0
$_5S_{x,x+4}$
SR at birth
"net migrants" by sex and age
b. intermediate indicators
Pop 5+
Est. Births
5 * B
Pop 0-4
c. output data
Projected population by sex and age (closed)
in t+5
Projected Population by sex and age (open to migration)
in t+5

- Survival ratios measure how many people in the younger age group will survive into the older age group based on the age-specific mortality rates used to construct the life table.
- Computed from two adjoining $_5L_x$ values of a life table

5Lx represents the number of person-years lived in each age group relative to the life table.

- In other words, $S_{x,x+n}$ is the survivorship ratio of persons who are aged x to x+n at the start of projection interval t to t+n.

- Used to project forward the population in each subgroup at the **beginning of the time interval** in order to estimate the **number still alive** at the beginning of the next interval accounting for those who survive each specific age interval

$$P_{x+n}(t+n) = P_x(t) * S_{x,x+n}$$

$$S_{x,x+n} = \frac{{}_nL_{x+n}}{{}_nL_x}$$

for all age groups except the ***last*** and the ***first***

$$S_{b,0} = \frac{{}_nL_0}{n \times l_0}$$

for the ***first age group***

$${}_\infty P_{85}(t+n) = [{}_5P_{80}(t) + {}_\infty P_{85}(t)] \times \frac{T_{85}}{T_{80}}$$

for the last ***age group***

International migration hypothesis (for open populations)

- If net international migration is substantial, the "survived" cohort population must be adjusted to reflect the effects of migration
- Assumptions are more often formulated in terms of ***absolute numbers*** and not by rates (related to policies)
- Migrants are also exposed to giving births and dying

  = migrants have the same fertility and mortality level as the non-migrant population
- The introduction of net migrants by age group at the mid-point of the projection period and the survival of net migrants to the end of the period:

  = an equal distribution of net migrants across time interval

CCM Model implementation – Example

A simplified Cohort-component method projection of the female population of "Westlands", from 2010 to 2015

Still need to estimate youngest cohorts based on fertility data

Population values are expressed in thousands

1. Survive base population by age and sex

First age group (0-4 → 5-9)

= **1,084** * (459092 / 469097) = 1,061

$$S_{x,x+n} = \frac{{}_nL_{x+n}}{{}_nL_x}$$

2. Adjust for migration

= 1061-0.9 = **1,060**

3. Survive the open-ended age group:

$${}_{\infty}P_{85}(t+n) = [{}_5P_{80}(t) + {}_{\infty}P_{85}(t)] \times \frac{T_{85}}{T_{80}}$$

= **(18+6)** * 39221 / (39221 + 91395) = **7**

Age group	Population in 2010	Person-years lived - ${}_5L_x$ (e_0=64)	Net migrant population	Population in 2015
0-4	1,084	469097	-2.1	
5-9	922	459092	-0.9	1,060
10-14	793	454840	-0.8	912
15-19	706	450962	-4.0	782
20-24	623	445347	-6.9	690
25-29	528	438622	-6.4	607
30-34	442	431379	-4.7	515
35-39	355	423529	-3.1	431
40-44	277	414509	-2.0	345
45-49	223	403997	-1.3	268
50-54	181	390632	-0.8	215
55-59	145	372490	-0.5	172
60-64	111	346186	-0.3	134
65-69	92	306649	-0.2	98
70-74	67	248840	-0.1	74
75-79	41	172444	-0.1	46
80-84	18	91395	0.0	22
85+	6	39221	0.0	7
Total	6,613		-34.4	

CCM Model implementation – Example

A simplified Cohort-component method projection of the female population of "Westlands", from 2010 to 2015

4. Calculate the total number of births during interval

- By multiplying projected age-specific fertility rates (ASFRs), with the average number of women in each reproductive age group → an estimation of the surviving women who gave birth during interval

$B_{15\text{-}19} = (706+782)/2 * 0.0967 = 360$

- Sum up births by age of mother to obtain total births

$B = 2{,}728$

Age group	Population in 2010	Population in 2015	Age-specific fertility rates	Births by age of mother
0-4	1,084			
5-9	922	1,060		
10-14	793	912		
15-19	706	782	0.0967	360
20-24	623	690	0.2089	686
25-29	528	607	0.2405	682
30-34	442	515	0.2239	536
35-39	355	431	0.1626	320
40-44	277	345	0.0780	121
45-49	223	268	0.0192	24
50-54	181	215		
55-59	145	172		
60-64	111	134		
65-69	92	98		
70-74	67	74		
75-79	41	46		
80-84	18	22		
85+	6	7		
Total	6,613		5.15	2728
Female births				

average=744

CCM Model implementation – Example

A simplified Cohort-component method projection of the female population of "Westlands", from 2010 to 2015

$P^{f}_{2010} = 6{,}613 \rightarrow P^{f}_{2015} = 7{,}632$

group	2010	Person-years lived - $_5L_x$ (e_0=64)	Net migrant population	Population in 2015	Age-specific fertility rates	Births by age of mother
0-4	1,084	469097	-2.1			
5-9	922	459092	-0.9	1,060		
10-14	793	454840	-0.8	912		
			-4.0	782	0.0967	360
			-6.9	690	0.2089	686
			-6.4	607	0.2405	682
			-4.7	515	0.2239	536
			-3.1	431	0.1626	320
			-2.0	345	0.0780	121
			-1.3	268	0.0192	24
			-0.8	215		
			-0.5	172		
			-0.3	134		
			-0.2	98		
70-74	67	248840	-0.1	74		
75-79	41	172444	-0.1	46		
80-84	18	91395	0.0	22		
85+	6	39221	0.0	7		
Total	6,613		-34.4		5.15	2728
Female births						1337

5. Apply the sex ratio at birth to obtain female births (consider sex ratio assumptions)

$B^{f} = 2{,}728/2.04 = 1{,}337$

6. Apply survivorship to these births to determine the surviving $_5P_0^{f}$ in 2015

$$S_{b,0} = \frac{_nL_0}{n \times l_0}$$

$_5P_0^{f} = 1{,}337 * (469097/500{,}000) \approx 1{,}254$

7. Adjust for migration of the group age (-2.1)

$_5P'^{f}_0 = 1{,}254 - 2.1 = 1{,}252$

CCM Model implementation – Example

A simplified Cohort-component method projection of the female population of "Westlands", from 2010 to 2015

"**5.** Apply the sex ratio at birth to obtain female births (consider sex ratio assumptions)"

From a ratio to a proportion:

Sex ratio (male/females) = 1.04 males per 1.00 females

Total males+females = 2.04

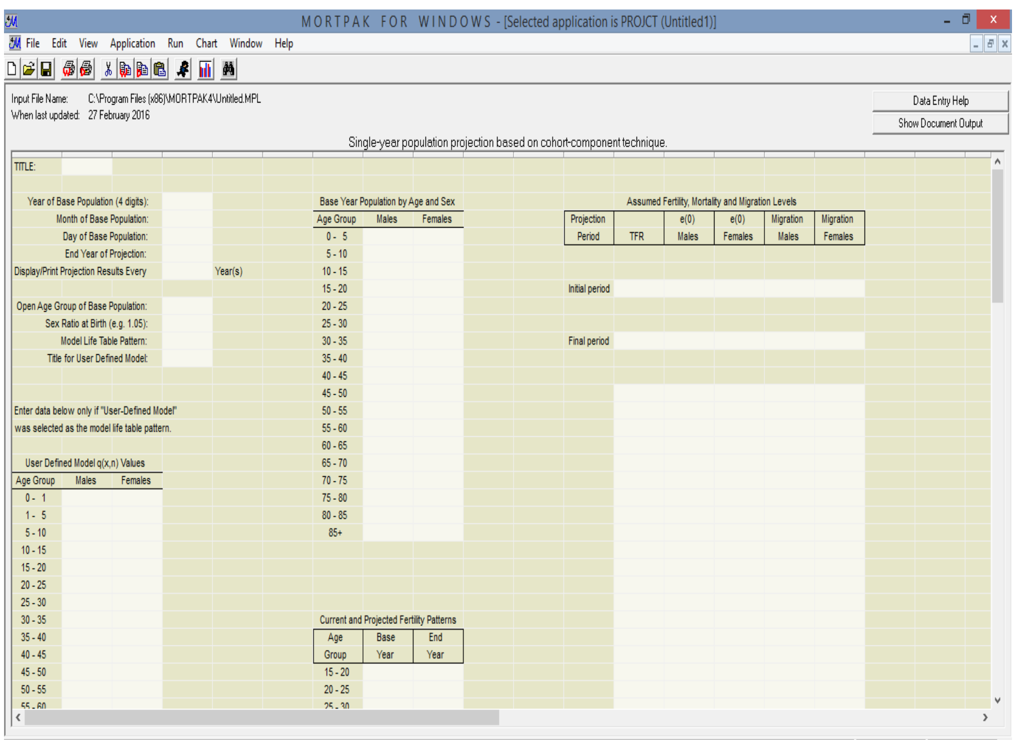
Proportion males = 1.04/2.04 = 0.51

Proportion females = 1.00/2.04 = 0.49

E.g.: $B^f = 2{,}728/2.04 \;\; = 2{,}728 * (1.00/2.04) = \; 2{,}728 * (1\text{-}1.04/2.04\,) = 1{,}337$

Session II: Hands-on exercise

1. Open the file: S II Projection methods exercise.doc
2. Work with Excel file: S II Projection methods exercise.xls
3. Work with MORTPAK PROJCT application



Thank you

Questions?

>> until 11 March:

>> After 11 March: sawyerc@un.org

bassarsky@un.org