Demographic Methods

Monica Alexander

Week 5: Migration

Overview

- Migration intro
  - Data sources
  - Adding into a cohort component projection framework
- Migration models
  - Two population Leslie Matrix
  - Gravity and log-linear models
  - Model age patterns
- Interesting data
- Next week

Migration

Back to the demographic accounting identity

Recall from week 1:

$$P(t+1) = P(t) + B[t, t+1] - D[t, t+1] + I[t, t+1] - O[t, t+1]$$

or

$$P(t+1) = P(t) + B[t, t+1] - D[t, t+1] + N[t, t+1]$$

where $N[t, t+1]$ is net migration.

- migration is the third component of population change
- generally less contribution to population change, but becoming more important
- we've largely ignored it until now, focusing on natural increase/decrease
- adding migration stuffs up (messes up) our idea of generational renewal, etc

Why is migration important?

From a demographic perspective, migration changes the age and sex structure of a population

- can offset natural decrease
- reduce dependency ratios (economic implications)

Old age dependency ratio:

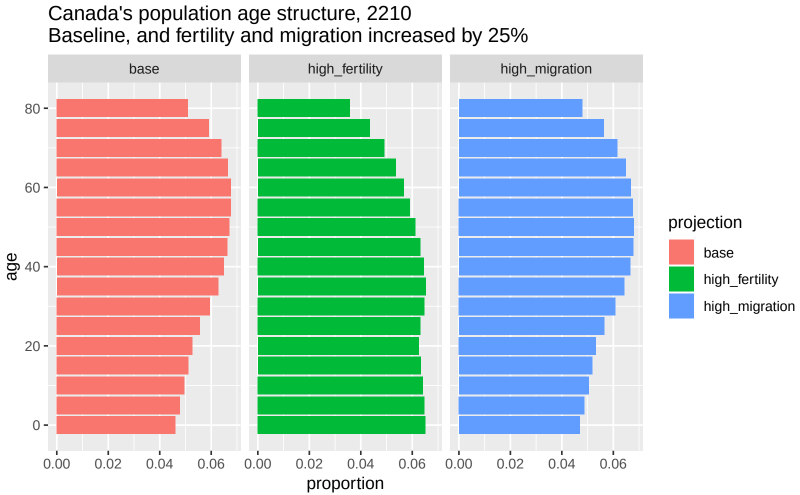
$$\frac{\text{Population aged 65+}}{\text{Population aged 15-64}}$$

Child dependency ratio:

$$\frac{\text{Population aged 0-14}}{\text{Population aged 15-64}}$$

Is increasing migration as effective as increasing fertility?

Migration versus fertility effects on population age structure



Issues

Definition issues:

- not biological
- geographical element
- depends on intent or subsequent behavior
- migration versus mobility (temporal component)
- contextually dependent (international versus internal)

Measurement issues:

- separate from vital registration
- immigration easier than emigration
- illegality issues

Measurement

$$P(t+1) = P(t) + B[t, t+1] - D[t, t+1] + N[t, t+1]$$

If we knew population size (e.g. from census), births and deaths completely accurately, net migration is just the residual

But in reality, we are usually dealing with

$$P(t+1) = P(t) + B[t, t+1] - D[t, t+1] + N[t, t+1] + \text{error}$$

Measures of migration

Stock: Population that are living in geographic area of interest (state, country) who were not born in that area. Gives idea of overall magnitude but no information about recent changes.

Flows: The number of people moving into (or out of) a geographic area over a time period of interest. Changes in stocks gives some information about flows.

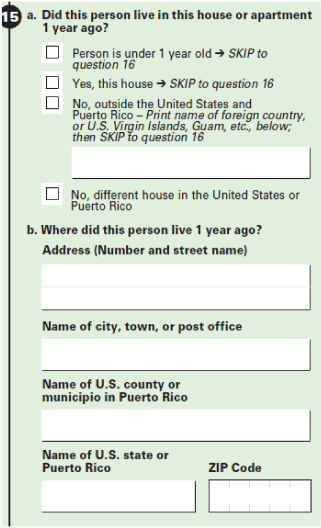
Migration rates: What is the denominator? For emigration it is as before, but for immigration it is not usually the population at risk.

Data sources

Traditional data come from censuses or surveys

- Inferred from population change
- Birthplace
- Place of previous residence
- Duration of current residence

Example: ACS



15 a. Did this person live in this house or apartment 1 year ago?

- ☐ Person is under 1 year old ➔ *SKIP to question 16*
- ☐ Yes, this house ➔ *SKIP to question 16*
- ☐ No, outside the United States and Puerto Rico – *Print name of foreign country, or U.S. Virgin Islands, Guam, etc., below; then SKIP to question 16*
- ☐ No, different house in the United States or Puerto Rico

b. Where did this person live 1 year ago?

Address (Number and street name)

Name of city, town, or post office

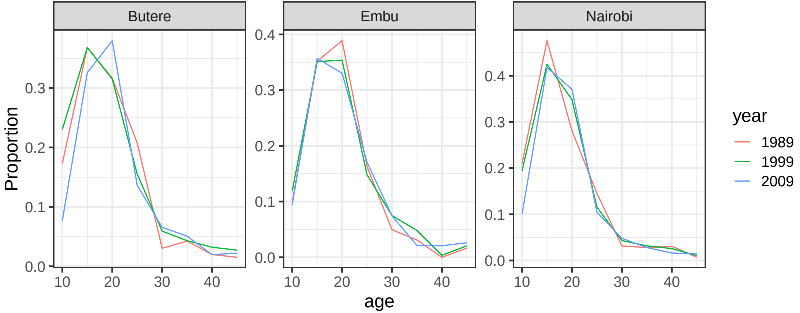
Name of U.S. county or municipio in Puerto Rico

Name of U.S. state or Puerto Rico

ZIP Code

Example: Kenya

From the Census we know district of residence one year ago. Can work out net migration by looking at the difference between in and out migrants. Proportion of migrants by age:



Adding migration into a cohort component projection framework

Adding migration into a CCP

Remember from last week, we have the Leslie Matrix population projection set-up:

$$\begin{bmatrix} \text{kids} & \text{kids} & \text{kids} & \dots & \text{kids} \\ \frac{{}_5L_{10}}{{}_5L_0} & 0 & 0 & \dots & 0 \\ 0 & \frac{{}_5L_{15}}{{}_5L_{10}} & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 0 \end{bmatrix}$$

where the kids entries are:

$${}_nL_0 \cdot \frac{1}{2}({}_nF_x + {}_nF_{x+n} \cdot \frac{{}_nL_{x+n}}{{}_nL_x}) \cdot f_{fab}$$

Adding migration into a CCP

Call this A. Multiply population vector $K(t)$:

$$K(t) = \begin{bmatrix} {}_5K_0 \\ {}_5K_5 \\ {}_5K_{10} \\ \cdots\cdots \\ {}_\infty K_\omega \end{bmatrix}$$

by A to get $K(t+n)$.

$$K(t+n) = AK(t)$$

Adding migration into a CCP

Think about the period t to $t+n$ (e.g. often $n=5$ so this is a five-year period).

- if all migrants arrived at the end of the period, then we would have

$$K(t+5) = \begin{bmatrix} \text{kids} & \text{kids} & \text{kids} & \dots & \text{kids} \\ \frac{{}_5L_{10}}{{}_5L_0} & 0 & 0 & \dots & 0 \\ 0 & \frac{{}_5L_{15}}{{}_5L_{10}} & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 0 \end{bmatrix} \times K(t) + \begin{bmatrix} {}_5K_0(t)_5i_0 \\ {}_5K_5(t)_5i_5 \\ {}_5K_{10}(t)_5i_{10} \\ \dots \\ {}_\infty K_\omega(t)_\infty i_\omega \end{bmatrix}$$

where ${}_ni_x$ is the number of net migrants expressed as a proportion of population size, so that ${}_nK_x \cdot {}_ni_x = {}_nI_x$ is the number of net migrants.

Adding migration CCP

Perhaps a slightly better approximation is assuming that half the migrants arrive at the start of the period and half arrive at the end. Then we have:

$$K(t+5) = \begin{bmatrix} \text{kids} & \text{kids} & \text{kids} & \dots & \text{kids} \\ \frac{{}_5L_{10}}{{}_5L_0} & 0 & 0 & \dots & 0 \\ 0 & \frac{{}_5L_{15}}{{}_5L_{10}} & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 0 \end{bmatrix} \times \begin{bmatrix} {}_5K_0(t) + {}_5K_0(t)_5i_0/2 \\ {}_5K_5(t) + {}_5K_5(t)_5i_5/2 \\ {}_5K_10(t) + {}_5K_10(t)_5i_{10}/2 \\ \dots\dots\dots \\ {}_\infty K_\omega(t) +_\infty K_\omega(t)_5i_0/2 \end{bmatrix}$$

$$+ \begin{bmatrix} {}_5K_0(t)_5i_0/2 \\ {}_5K_5(t)_5i_5/2 \\ {}_5K_{10}(t)_5i_{10}/2 \\ \dots\dots\dots \\ {}_\infty K_\omega(t)_\infty i_\omega/2 \end{bmatrix}$$

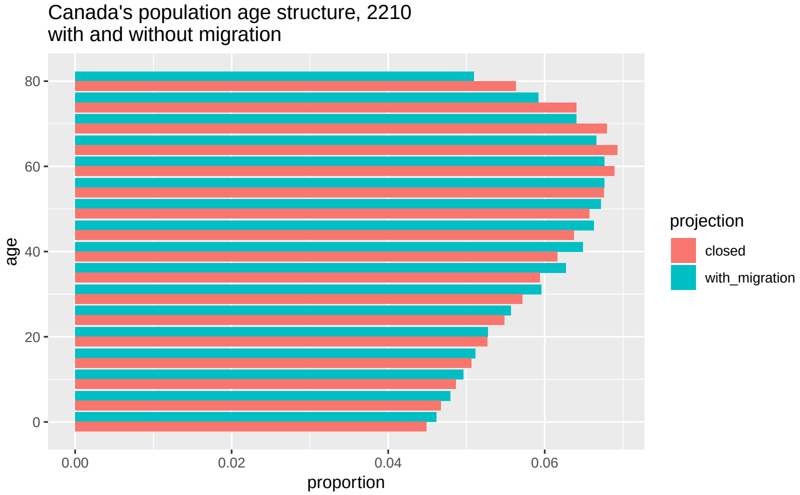
So half the migrants come in at the start and have time to die/give birth, and half come in at the end of the period.

Adding migrants to CCP

We are assuming that as soon as migrants arrive they experience the same mortality and fertility conditions as the native population.

- In practice, depending on the size of the population, this probably doesn't matter much
- How realistic is this assumption in practice?
  - fertility: usually depends on age at migration, education etc (Adsera and Ferrer (2011) illustrate this for Canada)
  - mortality: depends on migrant group, often observed to be higher in high-skilled migrant groups ('healthy migrant effect')

Example population projection with migration



Switching the focus: implied net migrants

- Given:
  - Population vectors $K(t)$ from two time points
  - Survival ratios
  - Fertility rates
- Can calculate the implied $K(t+5)$ based on surviving the initial population forward
- Compare this to the observed end population, and this gives you an estimate for the number of net migrants

Migration models

Overview

- The focus here is on aggregate models, not individual models
  - But there is an increasing use of agent based models in migration estimation and projection (group at University of Southampton (Bijak et al.))
- Demographers are mostly interested in two dimensions:
  1. net flows into areas (including pairwise relationships between areas)
  2. age patterns of migration

Two population Leslie Matrix

- Interested in capturing movements between two areas (e.g. urban-rural movements)
- Split Leslie Matrix into distinct submatrices based on four populations: urban natives, rural natives, urban migrants and rural migrants.

$$\begin{bmatrix} M^{[\text{city native}]}_{[\text{city native}]} & M^{[\text{city migrant}]}_{[\text{city native}]} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & M^{[\text{city migrant}]}_{[\text{city migrant}]} & M^{[\text{village native}]}_{[\text{city migrant}]} & M^{[\text{village migrant}]}_{[\text{city migrant}]} \\ \mathbf{0} & \mathbf{0} & M^{[\text{village native}]}_{[\text{village native}]} & M^{[\text{village migrant}]}_{[\text{village native}]} \\ M^{[\text{city native}]}_{[\text{village migrant}]} & M^{[\text{city migrant}]}_{[\text{village migrant}]} & \mathbf{0} & M^{[\text{village migrant}]}_{[\text{village migrant}]} \end{bmatrix}$$

- migrant to migrant transitions represent return migration (could simplify by assuming no return migration)
- migrant to native transitions represent births of migrants

Two population Leslie Matrix

$$\begin{bmatrix} M^{[\text{city native}]}_{[\text{city native}]} & M^{[\text{city migrant}]}_{[\text{city native}]} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & M^{[\text{city migrant}]}_{[\text{city migrant}]} & M^{[\text{village native}]}_{[\text{city migrant}]} & M^{[\text{village migrant}]}_{[\text{city migrant}]} \\ \mathbf{0} & \mathbf{0} & M^{[\text{village native}]}_{[\text{village native}]} & M^{[\text{village migrant}]}_{[\text{village native}]} \\ M^{[\text{city native}]}_{[\text{village migrant}]} & M^{[\text{city migrant}]}_{[\text{village migrant}]} & \mathbf{0} & M^{[\text{village migrant}]}_{[\text{village migrant}]} \end{bmatrix}$$

The elements of each M are: Fertility rates for the four populations; Survival probabilities for the four populations; Probabilities of migration.

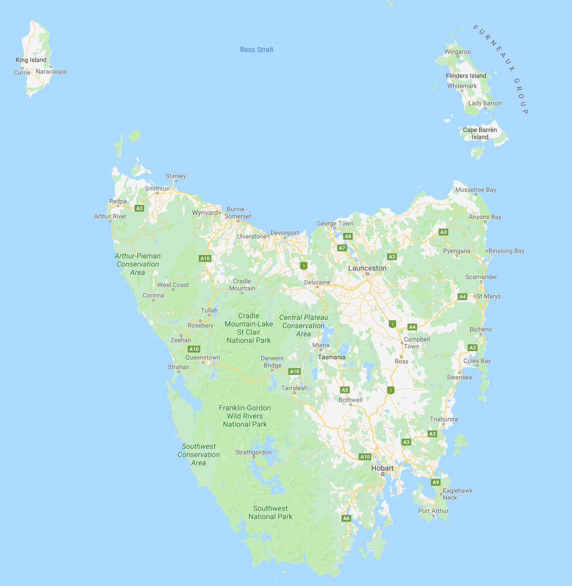
Assuming

- survival probabilities are positive
- migration rates are positive
- fertility rates are positive for any two adjacent age groups

the there exists a stable population $\tilde{K}$ such that $M\tilde{K} = \lambda\tilde{K}$.

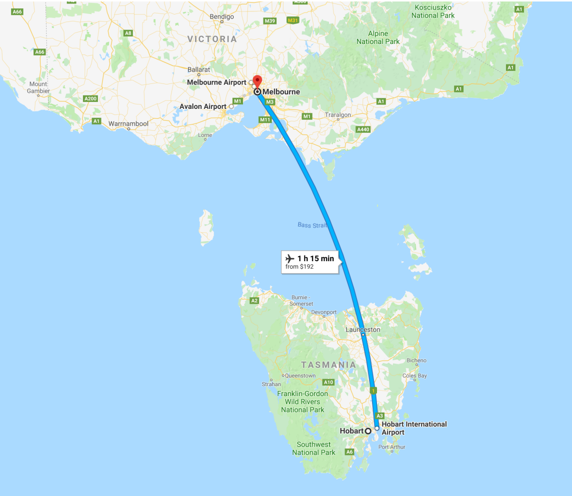
So can do our eigendecomposition on the whole matrix or each submatrix.

Example



King Island
Currie
Naracoopa
Bass Strait
FURNEAUX GROUP
Wingaroo
Flinders Island
Whitemark
Lady Barron
Cape Barren Island
Stanley
Smithton
Redpa
Arthur River
Wynyard
Burnie - Somerset
Ulverstone
Devonport
George Town
Musselroe Bay
Ansons Bay
Pyengana
Binalong Bay
Launceston
Arthur-Pieman Conservation Area
West Coast
Corinna
Cradle Mountain
Deloraine
Scamander
St Marys
Tullah
Rosebery
Cradle Mountain-Lake St Clair National Park
Central Plateau Conservation Area
Bicheno
Campbell Town
Zeehan
Miena
Tasmania
Ross
Coles Bay
Queenstown
Strahan
Derwent Bridge
Swansea
Tarraleah
Bothwell
Franklin-Gordon Wild Rivers National Park
Triabunna
Southwest Conservation Area
Strathgordon
Hobart
Eaglehawk Neck
Port Arthur
Southwest National Park
A2
A10
A8
A7
A3
A4
A5
A9
A6
1

Hobart

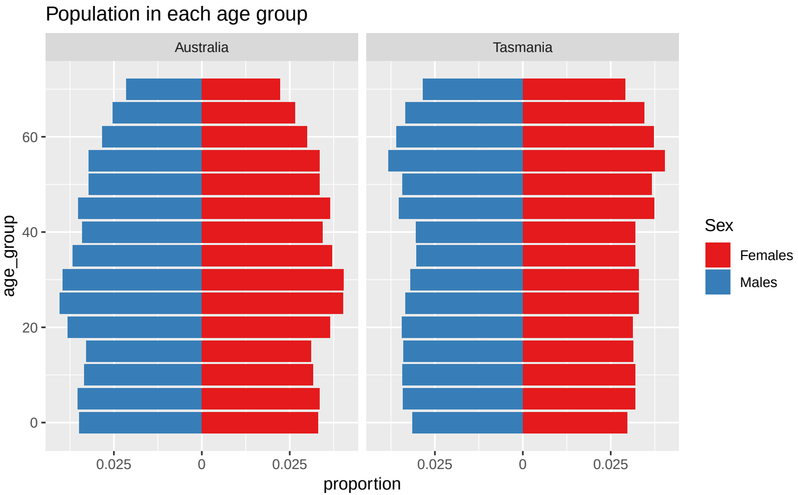


Bendigo
VICTORIA
Kosciuszko
National Park
Alpine
National Park
Ballarat
Mount
Gambier
Melbourne Airport
Melbourne
Avalon Airport
Warrnambool
Lorne
Traralgon
Bass Strait
1 h 15 min
from $192
Burnie -
Somerset
Devonport
Launceston
TASMANIA
Bicheno
Coles Bay
Queenstown
Strahan
Franklin-Gordon
Wild Rivers
National Park
Hobart
Hobart International
Airport
Southwest
National Park
Port Arthur

Melbourne


I am a web developer, mystery blogger and jazz kitten.
I would never be caught dead in neo-hippie "bush-doof" couture, a small, inefficient beanie, or anything less than extraordinary.
My style is bucolic socialist with improvised elements (like jazz).
Osh Kosh B'Gosh overalls I found in a vintage store in Tokyo.
A tote by my favourite feminist provocateurs, Ladies of Leisure, that says "Feeling Myself" because self-love is underrated, and my shoes are Nike.
I admire the style of Trotsky in leather, Albert Einstein, John Coltrane.
MEET MELBOURNE MAN
The world has gone crazy for our hyper-hipster – but is he just having us on?
PAGE 7
Samuel David Hains, who featured in *M Magazine*'s Street Seen column at the weekend, has shot to global prominence with his quirky stylings.
Photo: Tara Kenny

Population age pyramids

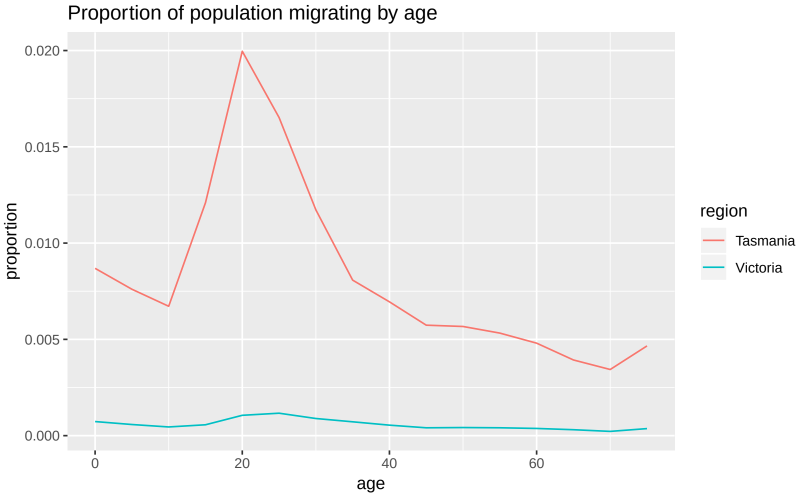


Set up two-population Leslie Matrix

$$
\begin{bmatrix}
M^{[\text{city native}]}_{[\text{city native}]} & M^{[\text{city migrant}]}_{[\text{city native}]} & \mathbf{0} & \mathbf{0} \\
\mathbf{0} & M^{[\text{city migrant}]}_{[\text{city migrant}]} & M^{[\text{village native}]}_{[\text{city migrant}]} & M^{[\text{village migrant}]}_{[\text{city migrant}]} \\
\mathbf{0} & \mathbf{0} & M^{[\text{village native}]}_{[\text{village native}]} & M^{[\text{village migrant}]}_{[\text{village native}]} \\
M^{[\text{city native}]}_{[\text{village migrant}]} & M^{[\text{city migrant}]}_{[\text{village migrant}]} & \mathbf{0} & M^{[\text{village migrant}]}_{[\text{village migrant}]}
\end{bmatrix}
$$

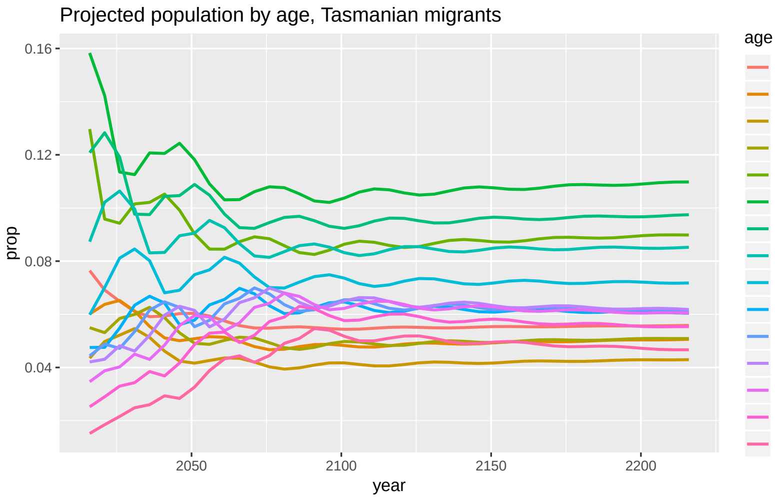
- Tasmania is 'village'
- Victoria is 'city'
- Note: assume that migrants are always migrants, but they give birth to natives

Migration by age

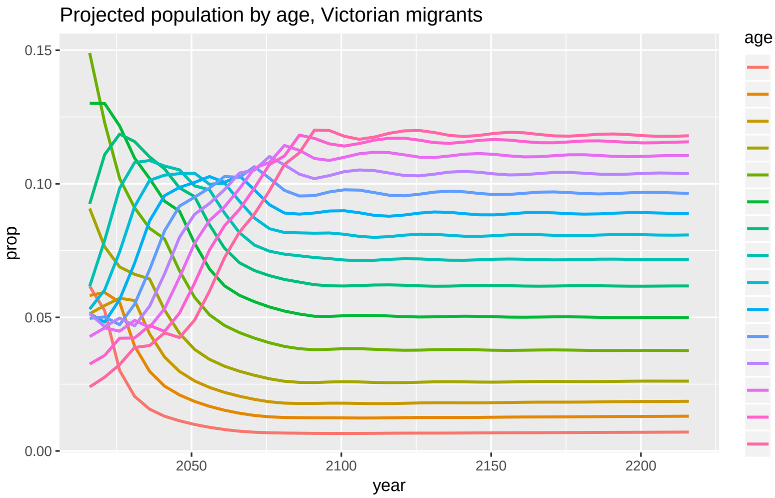


- majority (95%) of Tasmanians do not return after migrating
- majority (80%) of Victorians return after migrating

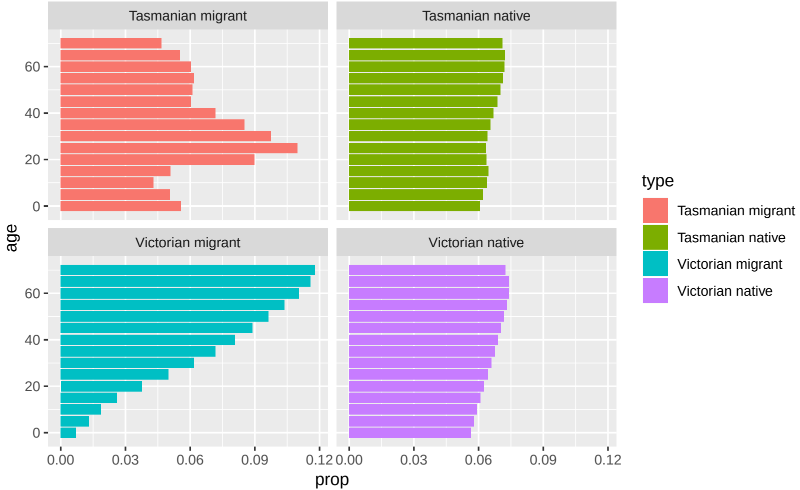
Projections of migrant populations



Projections of migrant populations



Stable populations



Gravity model

Resembles Newton's law of gravity. The idea that spatial interaction is related to stocks and inversely related to distance.

$$M_{ij} = G\frac{K_i^{\alpha}K_j^{\beta}}{D_{ij}^{\gamma}}$$

- G is some proportionality constant
- K_i is population of area i
- K_j is population of area j
- α is strength of 'push' factors
- β is strength of 'pull' factors
- D_{ij} is distance

Gravity model

On the log scale

$$\log M_{ij} = \log G + \alpha \log K_i + \beta \log K_j - \gamma \log D_{ij}$$

- Easy to fit
- But:
  - predicts symmetric flows between i and j
  - doesn't include push/pull factors other than population and distance

A more general representation

$$\log M_{ij} = \mu + \alpha_i + \beta_j + \gamma_{i,j}$$

for $i \neq j$.

- Push/pull factors need not just be population size
- Interaction effect need not be a linear function of distance

Note that this is a saturated model, not really a model, but a way of representing the data.

To make identifiable, set $\alpha_0 = \beta_0 = \gamma_{1j} = \gamma_{i1} = 0$.

Log-linear models

Quasi-independent generally fits reasonably well to large populations:

$$\log M_{ij} = \mu + \alpha_i + \beta_j$$

for $i \neq j$. 'Quasi' because we are not including the diagonal terms in the model.

Could also put some structure on the interaction term, for example

$$\log M_{ij} = \mu + \alpha_i + \beta_j + \gamma d_{ij}$$

$d_{ij} = 1$ when areas are adjacent and zero otherwise.

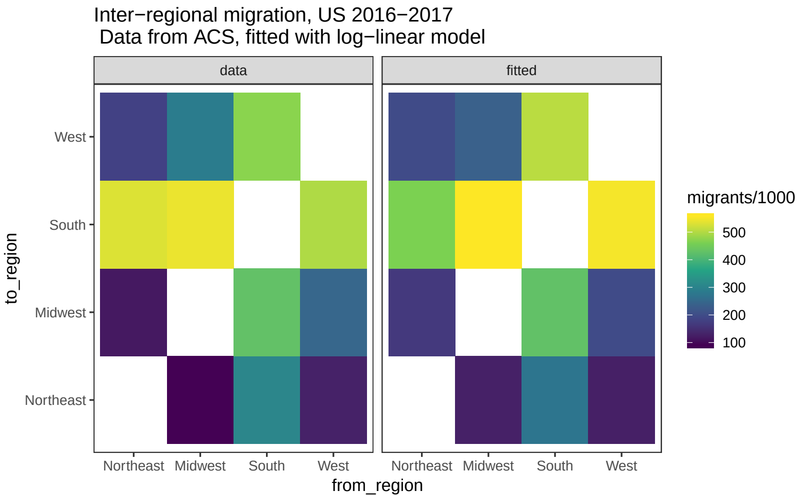
Example

US inter-regional movements, 2016-2017. Looking at migrants only, so diagonals are 0.

	2017				
2016	Northeast	Midwest	South	West	Total
Northeast	0	116096	529678	179832	825606
Midwest	92079	0	540847	284665	917591
South	304617	427617	0	471788	1204022
West	128870	246592	496148	0	871610
Total	525566	790305	1566673	936285	3818829

Fit

using `glm` in R:



Using log-linear models with missing data

In some cases, we may not have information on all cells.

Example: migration between countries (e.g. Europe)

- probably know marginal totals (i.e. number of migrants in each countries) M_{i+} and M_{+j}
- may only have information on source / destination breakdown for some countries

As long as we know marginals and have some information on the cells, we can use a log-linear model to fill in estimates (Raymer 2007).

Using log-linear models with missing data

Model set-up:

y_{ij} are observed flows with some error.

$$\begin{aligned} y_{ij} &\sim N(M_{ij}, \sigma^2) \\ \log M_{ij} &\sim N(\mu + \alpha_i + \beta_j + \gamma d_{ij}, \tau^2) I \left[\sum_j M_{ij} = M_{i+}, \sum_i M_{ij} = M_{+j} \right] \\ \alpha &\sim N(0, \sigma_\alpha^2) \\ \beta &\sim N(0, \sigma_\beta^2) \\ \gamma &\sim N(0, \sigma_\gamma^2) \end{aligned}$$

with hyperpriors on variance parameters.

Model age patterns of migration

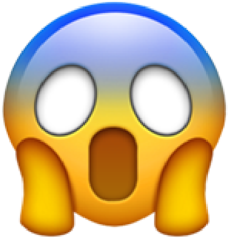
Migration often occurs in conjunction with some transition in the life course:

- Adult migration peaks at young adult ages (work, education)
- Second peak around retirement
- Child peak associated with parents' migration

Model age patterns of migration

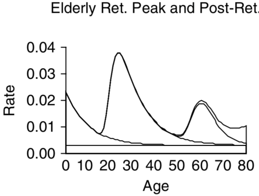
Multi-exponential model of migration (Rogers and Castro, 1981).
Migration rate $m(x)$ is modeled as:

$$
\begin{aligned}
m(x) &= a_1 \exp(-\alpha_1 x) + a_2 \exp\left(-\alpha_2(x-\mu_2) - \exp[\gamma_2(x-\mu_2)]\right) \\
&+ a_3 \exp\left(-\alpha_3(x-\mu_3) - \exp[\gamma_3(x-\mu_3)]\right) \\
&+ a_4 \exp\left(\gamma_4 x\right) + c
\end{aligned}
$$



Model age patterns of migration

$$
\begin{aligned}
m(x) &= a_1 \exp(-\alpha_1 x) + a_2 \exp\left(-\alpha_2(x-\mu_2) - \exp[\gamma_2(x-\mu_2)]\right) \\
&+ a_3 \exp\left(-\alpha_3(x-\mu_3) - \exp[\gamma_3(x-\mu_3)]\right) \\
&+ a_4 \exp\left(\gamma_4 x\right) + c
\end{aligned}
$$



- α's and γ's are rates of descent and ascent
- μ'a are peak ages
- all others are intensity parameters
- Can fit a model with all or some of these components, depending on context

Fitting Rogers-Castro

- Standard approach is Excel's solver or `optim` in R
- Super sensitive to initial conditions
- Unlikely to converge with 11 or 13 parameter versions

But MCMC is a reasonable option!

- Can use sensible priors and constraints on parameter space
- Can extend to include data model (e.g. Poisson)
- Can extend to smooth rates over time

RC in Stan

A lot of my week has been playing around with this

- Initial/simpler code: https://github.com/MJAlexander/demographic-methods/blob/master/code/rc.stan
- More flexible code: https://github.com/MJAlexander/demographic-methods/blob/master/code/rc_flexible.stan

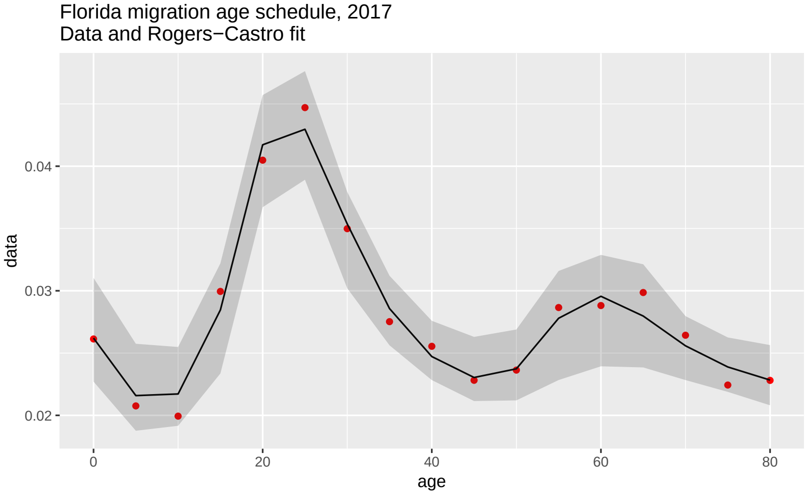
Or, fit using DemoTools.

```r
#devtools::install_github("timriffe/DemoTools")
library(DemoTools)
ages <- 0:80
mx_FL <- c(0.02345, 0.02195,    0.0212, 0.02185,    0.02095,
           0.02075, 0.0198, 0.02065,    0.01815,    0.01865,
           0.01945, 0.02165,    0.0224, 0.02255,    0.02665,
           0.02875, 0.02915,    0.03335,    0.0355, 0.03755,
           0.0404,  0.0411, 0.04205,    0.04355,    0.04275,
           0.0423,  0.04075,    0.0387, 0.0392, 0.0374,
           0.0336,  0.03185,    0.02975,    0.0295, 0.02855,
           0.02685, 0.0251, 0.0246, 0.02425,    0.0241,
           0.02285, 0.02165,    0.0236, 0.02085,    0.02145,
           0.0204,  0.02025,    0.02125,    0.0229, 0.01985,
           0.0234,  0.02175,    0.02185,    0.02385,    0.0248,
           0.0259,  0.02545,    0.0271, 0.02605,    0.0277,
           0.0296,  0.0278, 0.02945,    0.0288, 0.02775,
           0.0296,  0.02615,    0.0234, 0.0259, 0.02635,
           0.0256,  0.0248, 0.0221, 0.0234, 0.0233,
           0.02245, 0.02235,    0.0212, 0.02075,    0.01915,
           0.01995)

rc_res <- mig_estimate_rc(ages, mx_FL_sim,
                          pre_working_age = TRUE,
                          working_age = TRUE,
                          retirement = TRUE,
                          post_retirement = FALSE,
                          # (optional) arguments for Stan
                          chains = 4,
                          iter = 2000,
                          control = list(adapt_delta = 0.8, max_treedepth = 10))
rc_res[["fit_df"]] %>%
  ggplot(aes(ages, data)) +
  geom_point(aes(color = "data")) +
  geom_line(aes(x = age, y = median, color = "fit")) +
  geom_ribbon(aes(ymin = lower, ymax = upper), alpha = 0.2) +
  scale_color_manual(name = "", values = c(data = "red", fit = "black")) +
  ylab("migration rate")
```

Example

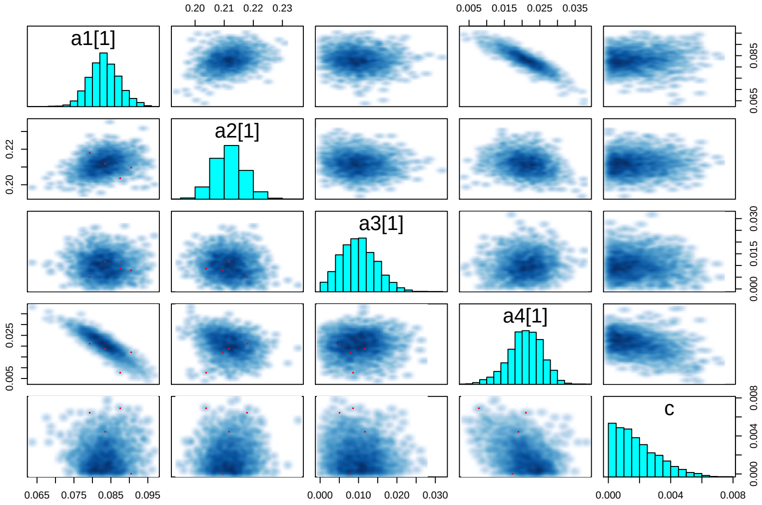
For Florida in 2017: $\mu_1 = 24.5$, $\mu_2 = 60$.



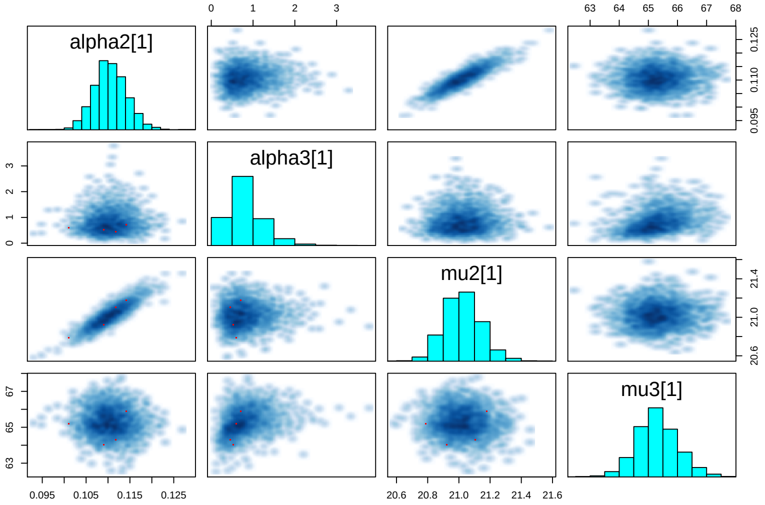
RC in Stan

- Works well in terms of convergence
- But still issues with divergent transitions, particurly when the fourth component is added
- Not surprising given there's strong correlation between a lot of the parameters
- In most cases doesn't seem like a huge issue but still would be nice to improve

Joint posterior samples



Joint posterior samples



Can we make this more stable?

$$\begin{aligned} m(x) &= a_1 \exp(-\alpha_1 x) + a_2 \exp\left(-\alpha_2(x-\mu_2) - \exp[\gamma_2(x-\mu_2)]\right) \\ &+ a_3 \exp\left(-\alpha_3(x-\mu_3) - \exp[\gamma_3(x-\mu_3)]\right) \\ &+ a_4 \exp\left(\gamma_4 x\right) + c \end{aligned}$$

Some ideas that may or may not work:

- model ratio of e.g. a_1 and a_4... have strong prior about what should be bigger, might be less correlation in estimates of ratios
- more constraints on priors
- shrinkage prior on fourth component (but issues with these anyway)

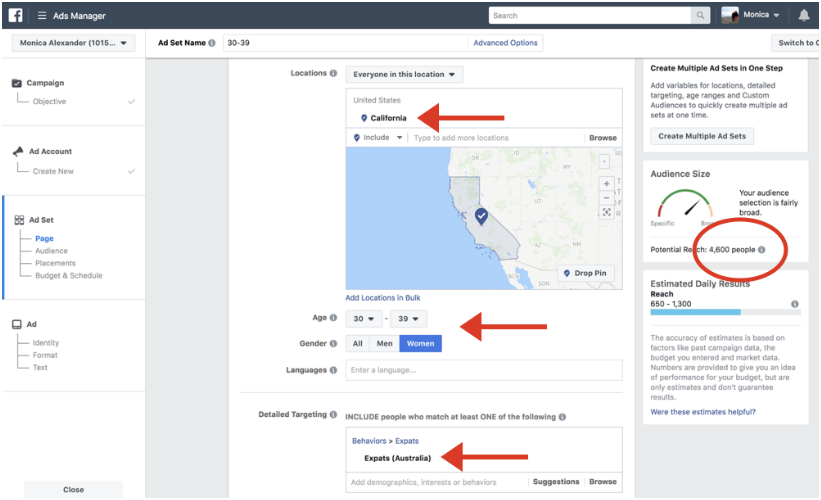
Why am I mentioning this?

Interesting data sets

Migration data

- Often migration information available from traditional/official sources is sparse or delayed in release
- Lots of examples of researchers using interesting sources of information to try and measure migration and mobility
  - cell phone data (mobility, short-term)
  - school enrollments
  - tax filings
  - social media (Twitter, Facebook/Instagram, LinkedIn)

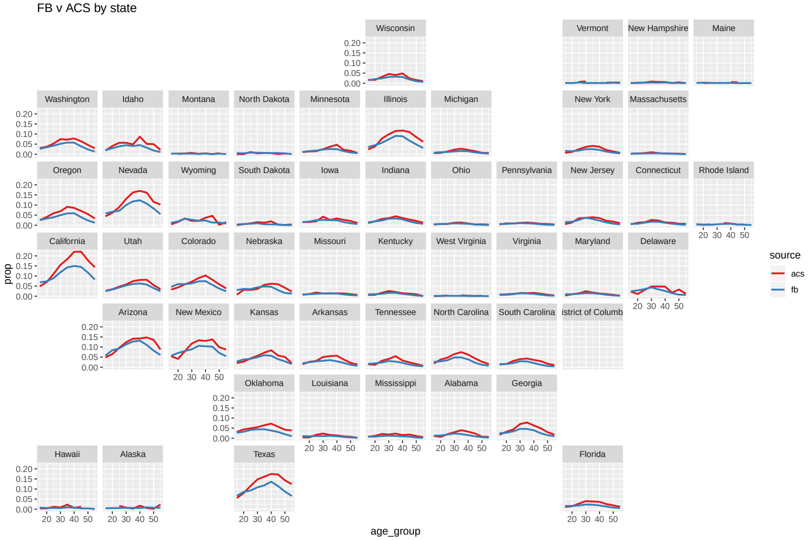
Social media and migration



X

Example: Mexican migrant age distributions in the USA

Facebook versus ACS



Bias is substantial, but can be modeled

- Adjust Facebook data based on gold standard
- Build time series model incorporating historical trends and new Facebook data
- Projections are informed by both