

Bayesian Rogers-Castro models for migration

BMSS, October 16 2024

Monica Alexander

Table of contents

1	Overview	1
2	What is a Rogers-Castro model?	1
3	Read in data	2
4	Fit Rogers-Castro model	3
4.1	Another example	5
5	Extensions	6

1 Overview

This Quarto document illustrates how to fit Rogers-Castro models to age-specific migration rates using the `rcbayes` R package. [Here's a paper](#) that gives some more information the package.

Load in the required packages:

```
library(tidyverse)
library(rcbayes)
```

2 What is a Rogers-Castro model?

The Rogers-Castro model (1981) for migration age schedules is multi-exponential parametric model that aims to capture the non-linear characteristic non-linear shape of migration. It has

the form:

$$m(x) = a_1 \exp \{-\alpha_1 x\} + \\ a_2 \exp \{-\alpha_2 (x - \mu_2) - \exp [-\lambda_2 (x - \mu_2)]\} + \\ a_3 \exp \{-\alpha_3 (x - \mu_3) - \exp [-\lambda_3 (x - \mu_3)]\} + \\ a_4 \exp \{\alpha_4 x\} + \\ c$$

where $m(x)$ is the migration rate (in- or out-) for age x . The successive additive components of the model represent pre-working, working, retirement, post-retirement, and overall migration, respectively. Various versions of the model can be estimated; for example, a simpler version would remove the components related to retirement and post-retirement peaks. The `rcbayes` package has a built-in interaction Shiny application that you can use to explore how different parameter values affect the shape of the migration age curve:

```
rcbayes::interact_rc()
```

3 Read in data

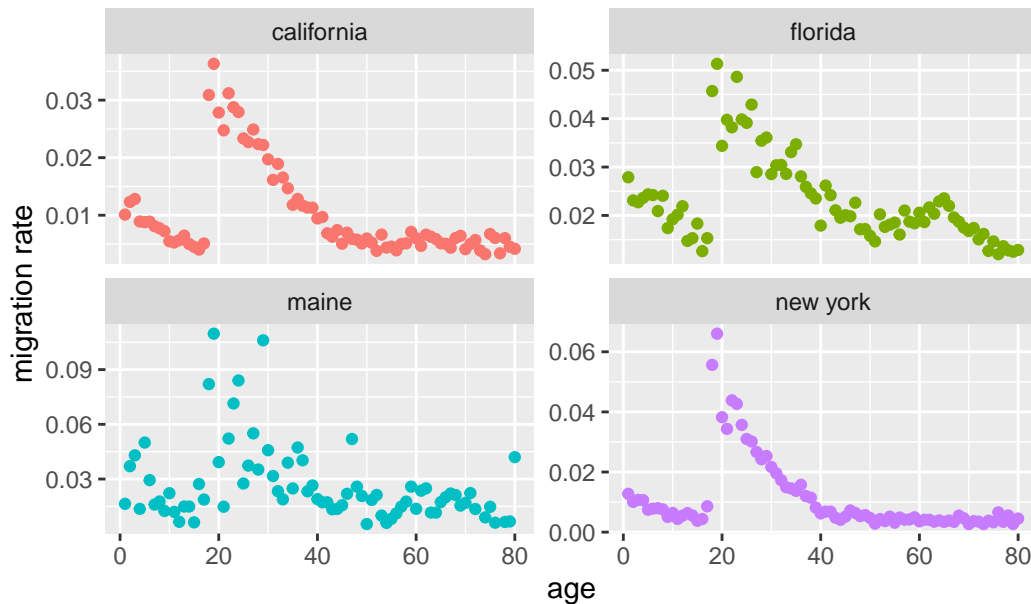
For this example, we'll be using data from the 2022 round of the American Community Survey (ACS). The data below show the population by age and number of in-migrants in the past year by age for each US state.

```
d <- read_csv("../data/us_state_mig.csv")
```

We can plot the in-migration rates observed in the ACS for a few different states. Florida has a clear retirement peak, which is not obvious in the other examples. Maine, which is a smaller state, has much noisier data.

```
d |>
  filter(state %in% c("florida", "new york", "california", "maine")) |>
  ggplot(aes(age, rate, color = state)) +
  geom_point()+
  facet_wrap(~state, scales = "free_y")+
  theme(legend.position = "none")+
  labs(title = "In-migration rates, 2022", y = "migration rate")
```

In-migration rates, 2022



4 Fit Rogers-Castro model

Now we're going to fit a Rogers-Castro migration curve to some ACS data, starting with Florida. To do this, we're going to use the `mig_estimate_rc` function in the `rcbayes` package. This is a wrapper function for `rstan`, and allows you to specify which components of the curve to fit, as well as the usual Stan control settings. For Florida, let's fit the full 13 parameter model.

```
d_florida <- d |> filter(state=="florida")
ages <- d_florida$age
migrants <- d_florida$n_mig
pop <- d_florida$pop

res <- rcbayes::mig_estimate_rc(ages = ages,
                                migrants = migrants,
                                pop = pop,
                                pre_working_age = TRUE,
                                working_age = TRUE,
                                retirement = TRUE,
                                post_retirement = TRUE,
                                control = list(adapt_delta = 0.96),
```

```
seed = 1899,  
refresh = 0)
```

The `res` object has several elements, including a summary of the model parameters, convergence checks, and the estimated fit and uncertainty.

```
names(res)
```

```
[1] "pars_df"      "fit_df"      "check_converge"
```

```
res$pars_df
```

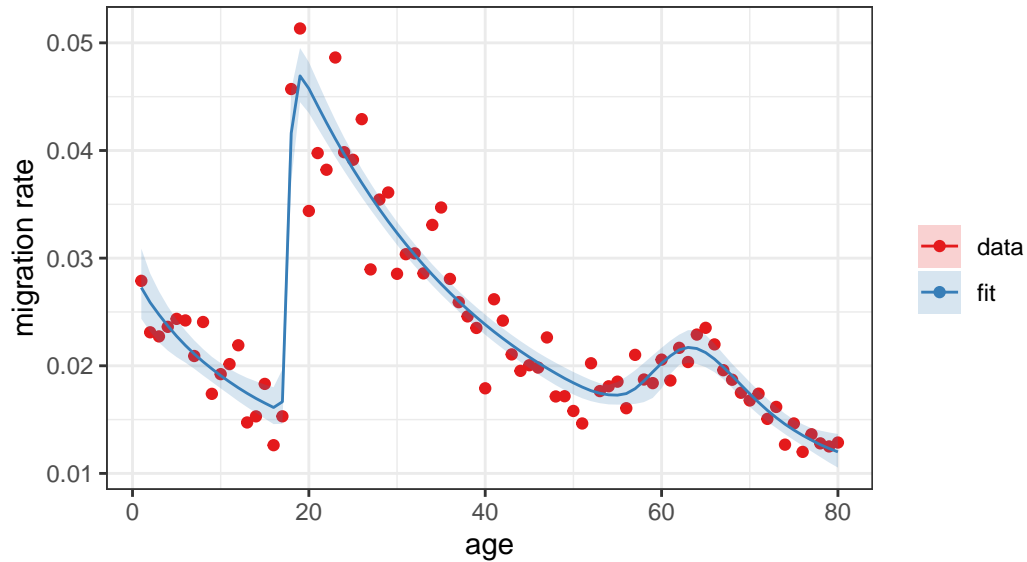
```
# A tibble: 13 x 4  
  variable median lower upper  
  <chr>      <dbl>   <dbl> <dbl>  
1 a1         0.0171  0.00533 0.0270  
2 a2         0.0347  0.0311  0.0382  
3 a3         0.0208  0.0150  0.0268  
4 a4         0.00693 0.000317 0.0222  
5 alpha1     0.0797  0.0407  1.57  
6 alpha2     0.0443  0.0322  0.0587  
7 alpha3     0.202   0.118   0.387  
8 c          0.00432 0.000196 0.0104  
9 lambda2    2.80    1.99    3.94  
10 lambda3   0.172   0.119   0.295  
11 lambda4  -0.00780 -0.0291  0.00937  
12 mu2       17.5    17.3    17.7  
13 mu3       64.9    63.0    66.8
```

Let's pull out the fit dataframe (which also have the data), and plot to see what it looks like.

```
res$fit_df |>  
  ggplot(aes(age, data)) +  
  geom_point(aes(color = "data", fill = "data")) +  
  geom_line(aes(age, median, color = "fit")) +  
  geom_ribbon(aes(age, ymin = lower, ymax = upper, fill = "fit"), alpha = 0.2) +  
  scale_color_brewer(name = "", palette = "Set1") +  
  scale_fill_brewer(name = "", palette = "Set1") +  
  theme_bw() +  
  labs(title = "In-migration rates, Florida, 2022",  
        subtitle = "Data from ACS, with Rogers-Castro fit",  
        y = "migration rate")
```

In-migration rates, Florida, 2022

Data from ACS, with Rogers–Castro fit



4.1 Another example

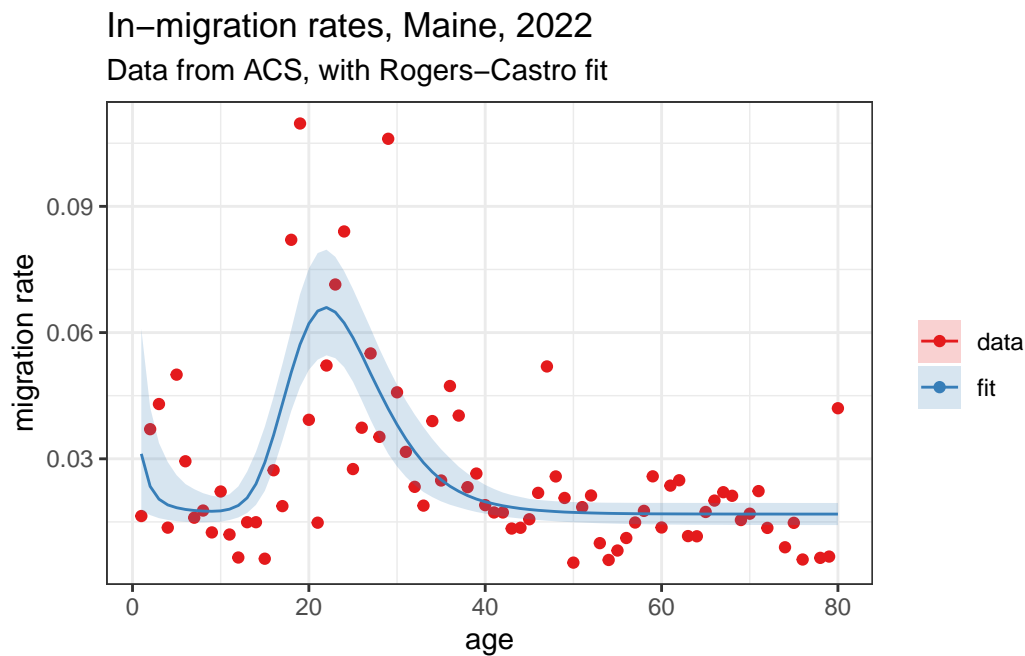
Let's now fit a reduced form of the Rogers Castro model to Maine. Rather than fitting the full model, we'll just fit the simplest model of pre-working age and working age migration.

```
d_maine <- d |> filter(state=="maine")
ages <- d_maine$age
migrants <- d_maine$n_mig
pop <- d_maine$pop

res_maine <- rcbayes::mig_estimate_rc(ages = ages,
                                     migrants = migrants,
                                     pop = pop,
                                     pre_working_age = TRUE,
                                     working_age = TRUE,
                                     retirement = FALSE,
                                     post_retirement = FALSE,
                                     seed = 1865,
                                     refresh = 0)
```

Plotting the results:

```
res_maine$fit_df |>
  ggplot(aes(age, data)) +
  geom_point(aes(color = "data", fill = "data"))+
  geom_line(aes(age, median, color = "fit"))+
  geom_ribbon(aes(age, ymin = lower, ymax = upper, fill = "fit"), alpha = 0.2)+
  scale_color_brewer(name = "", palette = "Set1")+
  scale_fill_brewer(name = "", palette = "Set1")+
  theme_bw()+
  labs(title = "In-migration rates, Maine, 2022",
       subtitle = "Data from ACS, with Rogers-Castro fit",
       y = "migration rate")
```



5 Extensions

The `rcbayes` package was designed for users to be able to fit Rogers-Castro models to a single population. The Stan code for the models in the packages is [here](#). These could be extended to fit, for example, hierarchical models for multiple areas at a time.