Modeling Age-Specific Migration for U.S. states

Crystal Yu
Hana Ševčíková
Adrian E. Raftery
Sara R. Curran

University of Washington

Abstract

People often make moves in preparation for, or in response to, major life course events and transitions. This can include entering higher education, obtaining a job, marriage, family, and retirement. However, data on migration events and flows are often scarce and of varying data quality. Moreover, the dynamic nature of migration makes it an extremely difficult phenomenon to empirically study, much less predict. As migration becomes an increasingly important driver of population change in both national and sub-national settings, it is imperative to have high-quality, granular migration data, as well as methods for informing and producing migration and population estimates and projections. In this analysis, we draw on previous research on model migration schedules, data from the American Community Survey, and use a Bayesian computational framework to produce probabilistic estimates of age-specific migration rates and counts with prediction intervals for U.S. states. This approach allows us to address potential concerns with migration data availability and reliability, and provides measures of uncertainty surrounding the migration estimates to indicate the uncertainty inherent in predicting migration. These estimates will provide a more complete picture of migration patterns at the state level, and inform probabilistic population projections for all U.S. states.

Introduction

In a 1981 paper, Rogers and Castro observed that migration follows a fairly regular pattern over the life course, i.e., by age. They suggested this pattern can be modeled using a parametric, multi-exponential curve. This model curve, or schedule, would capture the rate (intensity) of migration over the life course; specifically, moves that occur during early childhood, young adulthood, and retirement ages. The model migration schedule suggested by Rogers and Castro (1981) is approximated by a curve with a high rate of migration during the young adulthood years, with sharp declines in migration intensities during late childhood and middle adulthood years. Retirement migration may also be prominent in some areas, and can be modeled with an increase in migration rates around peak retirement ages. This archetypal pattern is shown in Figure 1.

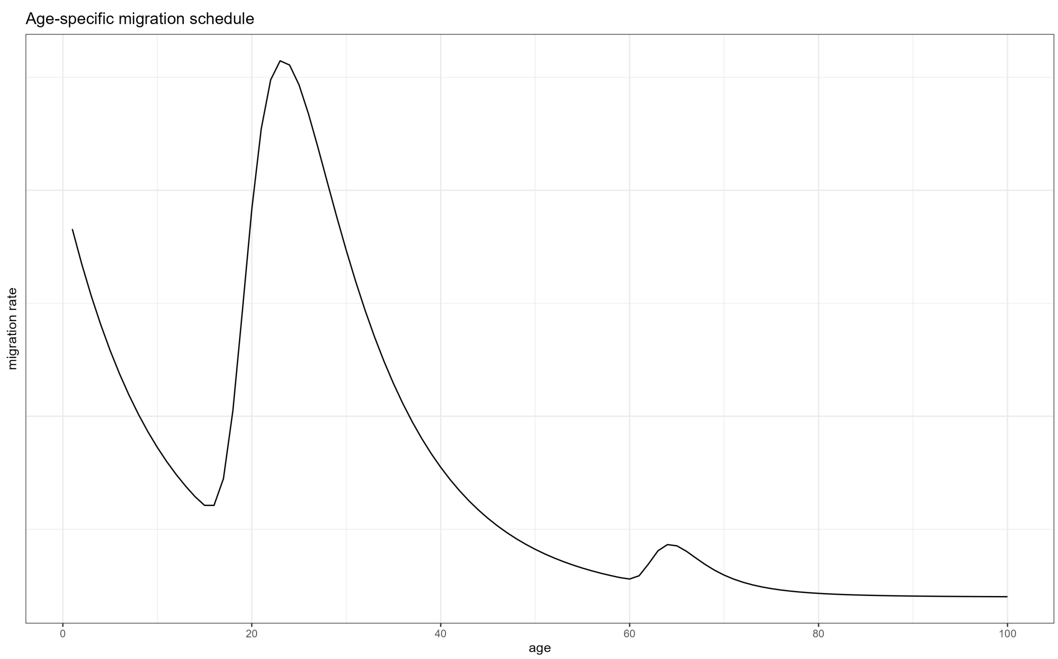

Figure 1: Example of a Rogers-Castro model migration schedule illustrating the patterns in migration rates by age, r_x, over the life course. It shows that migration rates are highest during the young adult years, and low among teenagers and middle aged individuals. There is a similar rate of decline in migration rates for young children and young parents due to co-dependency, i.e., parents relocating to areas more amenable to starting a family and raising children. Some areas may also exhibit a retirement migration bump in the ages leading up to and through retirement; the bump has been exaggerated here to illustrate the phenomenon. This general shape of migration over the life course has been observed across different geographies, although in some places the curve peaks are centered at different ages, and the magnitude of the peaks differ.

The general trend in migration rates by age depicted in Figure 1 has been observed across different geographies (Rogers and Castro 1981; Bernard et al. 2014). However, variations in the peak age of migration, as well as intensity of migration, have been observed (Bernard et al. 2014). Moreover, in some areas, retirement migration is less prominent (Rogers et al. 2010). Although some research has suggested a need to additionally model post-retirement migration, in which older adults move towards areas with better access to health facilities and nursing amenities (Rogers 1988; Rogers and Little 1994; Plane et al. 2005), as well as traditional college-age migration (Wilson 2010), the general pattern portrayed in Figure 1 reflects the observed migration patterns over the life course for many geographies.

Theoretical focus

The full model migration schedule originally proposed by Rogers and Castro (1981) uses 11 parameters to capture the increase and decrease in migration rates over the life course. They also propose a reduced schedule with 7 parameters for areas without retirement migration. Although several researchers have explored the estimation and use of model migration schedules, including Rogers et al. (2005), Raymer and Rogers (2008), Rogers et al. (2010), Bernard et al. (2014), Bernard and Bell (2015), and Wilson (2010, 2020), producing stable parameter estimates for a migration curve has proven to be challenging. These difficulties stem from limited data on migration, as well as computational difficulties in producing stable estimates for a large number of nonlinear parameters. Although Rogers et al. (2005) proposed several linear approaches for estimating the model schedule parameters, the estimated values often vary depending on the starting values as well as the methodological approach used, making them difficult to be used and compared across different areas.

Several researchers have explored a probabilistic approach to estimating migration rates, and with limited data. For instance, Congdon (2008) demonstrated the use of several Bayesian modeling approaches to model migration rates between Scotland and England. He shows that both parametric and nonparametric modeling approaches generate reasonable results, though some model specifications and priors capture specific features of the migration curve better than others. He further suggests that a Bayesian approach and random effect models work well for estimating migration schedules. Alexander et al. (2022), Somerset et al. (2023), and Yeung et al. (2023) have explored using a Bayesian approach to estimate migration rates using U.S. American Community Survey data. They suggest that a Bayesian approach can yield stable Roger-Castro parameter estimates, and that the resulting migration rate estimates are reasonable. They illustrate this using selected 1- and 5-year age-specific migration data from the U.S. Census Bureau American Community Survey, although they focused on states with larger populations and presumably, more stable age-specific migration patterns. It is unclear how well their approach works for geographic areas with smaller populations and potentially more volatile age-specific migration patterns.

In a 1970 publication, the United Nations (UN) provides an overview of several different methods to estimate internal migration. This includes age-specific rates of in-, out-, and net migration. However, they caution that the interpretation and reliability of these rates are contingent on the quality and type of data from the areas being examined. They also note that migration estimates are often calculated using indirect methods, which can obscure notable spatial patterns and phenomena such as temporary migration (United Nations 1970). Moreover, most of these estimates are of net migration, not of in- and out-flows. As Rogers (1990) has argued, examining in- and out-migration flows separately provides a more comprehensive understanding of migration patterns and trends.

High quality, detailed data on migration remains a challenge today (United Nations 2024). In the most recent revision of the World Population Prospects, released in 2024, the UN continues to primarily work with net migration total counts, then uses model patterns to estimate the projected number of net migrants by age (United Nations 2024). These model patterns are based on Rogers-Castro migration schedules, and are used to distribute total net migrants into age-specific estimates of migrants. Several model patterns are currently in use, distinguishing between family, female labor, and male labor migration patterns (Riffe et al. 2019). Each of these model patterns are based on Rogers and Castro (1981)'s reduced migration schedule, using 7 parameters to capture the changes in migration rates during the childhood and young adulthood ages. These patterns are derived from data from countries with more extensive demographic data collection systems (United Nations 2024). The reliance on these potentially limited model patterns to estimate age-specific migration may obscure more nuanced patterns occurring in specific geographic areas. Such nuances might be particularly important to observe in places where populations are rapidly aging or where there has been a recent change in overall population growth rates. A renewed examination of the Rogers-Castro migration schedule, and construction of area-specific migration curves, would be a methodological improvement and contribution.

Research methods

In this analysis, we draw on previous research on migration schedules, Bayesian modeling approaches, and recent data from the U.S. Census Bureau American Community Survey to estimate age-specific migration rates and counts for U.S. states. Specifically, we focus on inter-state migration, or migrants relocating from one U.S. state to another U.S. state. Further, we separately examine in- and out-migration, as the substantive interpretation and the population at risk for each are different. With out-migration, the population at risk is the population residing in the geographic area a given amount of time ago. In contrast, for in-migration, the population at risk consists of everyone in the world residing outside the geographic area of interest a specified time period ago. Using this population at risk does not give useful predictions, however, since the risk of migrating to a given area varies hugely for people who live in different places.

Separate examination of in- and out-migration, and more precise estimates of out-migration rates and counts, can be used to expand our knowledge and understanding of migration patterns and trends (Rogers 1990). It can also inform better population estimates and projections. Further, a probabilistic approach allows us to quantify uncertainty about both current and future migration, which is important given the often large measurement errors in the data and the large variations in migration rates over time.

As introduced in Rogers and Castro (1981), the Rogers-Castro migration curve models migration rates for one geographic area at selected time point or period. In this analysis, we extend this to simultaneously examine multiple geographies and multiple time points. That is, we examine $r_{x,s,t}$, the inter-state in- or out-migration rate at age x, in U.S. state s, at year t. This rate is theoretically different for each age x, state s, and time point t. We model $r_{x,s,t}$ using the 11-parameter model introduced in Rogers and Castro (1981):

$$
\begin{aligned}
r_{x,s,t} := {} & a_{1_{s,t}}\, exp(-\alpha_{1_{s,t}}\, x) + & (1)\\
& a_{2_{s,t}}\, exp\left(-\alpha_{2_{s,t}}[x - \mu_{2_{s,t}}] - exp[-\lambda_{2_{s,t}}(x - \mu_{2_{s,t}})]\right) + & (2)\\
& a_{3_{s,t}}\, exp\left(-\alpha_{3_{s,t}}[x - \mu_{3_{s,t}}] - exp[-\lambda_{3_{s,t}}(x - \mu_{3_{s,t}})]\right) + & (3)\\
& c_{s,t} & (4)
\end{aligned}
$$

Each numbered line of the equation corresponds to one component of the full migration schedule. Line 1 captures childhood migration; line 2 young adult migration; line 3 retirement migration; and line 4 a constant level of migration present throughout the life course.

We use a Bayesian computational approach to estimate the migration curve parameters and to produce probabilistic estimates of age-specific out-migration rates for all U.S. states. This computational approach helps alleviate some of the difficulties of estimation due to limited data, by allowing for borrowing of information across age groups and areas. Moreover, this approach produces prediction intervals, which help capture the uncertainty inherent in predicting the migration process. We then use these estimated rates as input for simulating estimated in- and out-migrant counts under a negative binomial distribution. The negative binomial overdispersion factor will help capture state-specific differences in in- and out-migration propensities and data availability. Although we examine U.S. states in this analysis, the methods used here can also be applied to other national and subnational geographic units.

Data

The U.S. Census Bureau American Community Survey (ACS) contains data that can be used to estimate migration schedules for U.S. states. The ACS is an annual household survey administered by the U.S. Census Bureau to gather information about the demographics, social, economic and housing characteristics of the U.S. population. It collects information on individual Americans' primary demographic and socioeconomic details, including age, sex, current state and area of residence, and whether the respondent moved one year ago. About 3.5 million households in the U.S. are surveyed each year for the ACS (U.S. Census Bureau 2020).

The ACS microdata files contain individual-level information on various demographic characteristics, including age, sex, state and place of residence, migration status in the past year and, if applicable, previous state of residence. These data can be used to calculate the observed age-specific in- and out-migration rates and counts over the past year for each state. This information can then be used to produce state-specific model migration schedules.

Using ACS microdata, we first tabulate observed migration counts and rates by age for each U.S. state. We then use this information to estimate state-specific migration schedules, namely, all 11 parameters of the Rogers-Castro curve. We also introduce extra parameterss to capture state- and time-specific migration propensity and data effects. We then compare the results to observed age-specific migration rates and counts for all U.S. states from 2006 to 2022. We will also assess the model fit and perform validation checks.

Expected findings

Using ACS microdata, we estimate in- and out-migration schedules for all U.S. states from 2006 to 2022. We separately examine in- and out-migration for each state. For each state and year combination, we estimate the Rogers-Castro curve and additional state- and time-specific migration propensity and data effects. We use this information to generate in- and out- age-specific migration rate and count estimates for all U.S. states, as well as prediction intervals to represented the inherent uncertainty surrounding migration estimates. The final data product will be estimated age-specific in- and out-migration counts, with prediction intervals, for all states and years being examined.

References

- Alexander, Monica, Jessie Yeung, and Tim Riffe. 2022. Rogers Castro Migration Models with rcbayes. https://cran.r-project.org/web/packages/rcbayes/vignettes/intro_to_rcbayes.html
- Bernard, Aude, Martin Bell, and Elin Charles-Edwards. 2014. Improved measures for the cross-national comparison of age profiles of internal migration. Population Studies 68(2): 179-195.
- Bernard, Aude and Martin Bell. 2015. Smoothing internal migration age profiles for comparative research. Demographic Research 32: 915-948.
- Congdon, Peter. 2008. Models for migration age schedules: a Bayesian perspective with an application to flows between Scotland and England. Pp. 193-205 in International migration in Europe: data, models and estimates, edited by James Raymer and Frans Willekens. Chichester, England: Wiley.
- Plane, D.A., C.J. Henrie, and M.J. Perry. 2005. Migration up and down the urban hierarchy and across the life course. Proceedings of the National Academies of Science 102(43): 15313-15318.
- Raymer, James and Andrei Rogers. 2008. Applying model migration schedules to represent age-specific migration flows. Pp. 175-192 in International migration in Europe: data, models and estimates, edited by James Raymer and Frans Willekens. Chichester, England: Wiley.
- Riffe, Tim, José M. Aburto, Monica Alexander, Sean Fennell, Ilya Kashnitsky, Marius D. Pascariu, and Patrick Gerland. 2019. DemoTools: An R package of tools for aggregate demographic analysis. https://github.com/timriffe/DemoTools/
- Rogers, Andrei. 1988. Age patterns of elderly migration: An international comparison. Demography 25(3): 355-370.
- Rogers, Andrei. 1990. Requiem for the net migrant. Geographical Analysis 22(4): 283-300.
- Rogers, Andrei and Luis J. Castro. 1981. Model migration schedules. International Institute for Applied Systems Analysis, Working Paper 81-30. https://pure.iiasa.ac.at/id/eprint/1543/1/RR-81-030.pdf
- Rogers, Andrei and Jani S. Little. 1994. Parameterizing age patterns of demographic rates with the multiexponential model schedule. Mathematical Population Studies 4(3): 175-195.
- Rogers, Andrei, Luis J. Castro, and Megan Lea. 2005. Model migration schedules: Three alternative linear parameter estimation methods. Mathematical Population Studies 12(1): 17-38.
- Rogers, Andrei, Jani S. Little, and James Raymer. 2010. The indirect estimation of migration: Methods for dealing with irregular, inadequate, and missing data. New York: Springer Science & Business Media.
- Somerset, Emily, Jessie Yeung, and Monica Alexander. 2023. Modeling age patterns of migration in small areas: A flexible framework to account of unexpected deviations. https://submissions.mirasmart.com/PAA2023/ViewSubmissionFile.aspx?sbmID=1511&mode=html&validate=false&key=8eznj2QI2

- United Nations. 1970. Methods of Measuring Internal Migration. New York: United Nations. https://www.un.org/development/desa/pd/sites/www.un.org.development.desa.pd/files/files/documents/2020/Jan/un_1970_manual_vi_-_methods_of_measuring_internal_migration_0.pdf

- United Nations, Department of Social and Economic Affairs, Population Division. 2024. World Population Prospects 2024: Methodology of the United Nations population estimates and projections. UN DESA/POP/2024/DC/NO.10. https://population.un.org/wpp/Publications/Files/WPP2024_Methodology_Advance_Unedited.pdf

- U.S. Census Bureau. 2020. Understanding and Using American Community Survey Data: What All Data Users Need to Know. Washington DC: U.S. Government Publishing Office. https://www.census.gov/content/dam/Census/library/publications/2020/acs/acs_general_handbook_2020.pdf

- Wilson, Tom. 2010. Model migration schedules incorporating student migration peaks. Demographic Research 23: 191-221.

- Wilson, Tom. 2020. Modeling age patterns of internal migration at the highest ages. Spatial Demography 8(2): 175-192.

- Yeung, Jessie, Monica Alexander, and Tim Riffe. 2023. Bayesian implementation of Rogers-Castro model migration schedules: An alternative technique for parameter estimation. Demographic Research 49: 1201-1228.