PAA 2018 Extended Abstract

Immobility and Repeat Migration in the U.S.: Evidence from Linked IRS and Census Survey Records

Mark J. Ellis[a]
Thomas B. Foster[b]
Lee Fiorio[a]

[a] University of Washington, Department of Geography and
Center for Studies in Demography and Ecology

[b] U.S. Bureau of Census, Center for Administrative Records Research and Applications

Contact Information:
Mark Ellis
ellism@uw.edu

Abstract

This paper utilizes a novel dataset linking individual level IRS and SSA administrative records to decennial census and ACS microdata to examine interstate migration longitudinally. We make use of the these linked administrative and survey data to compare the socio-demographic characteristics of the immobile population (those who never move between 2000 and 2016) to those of the mobile population (those who move at least once), and then compare the characteristics of the mobile population by the frequency (moving once versus twice versus three times, etc.) and type (onward moves to new destinations, or return moves to prior locations - i.e. circulations) of moves made between 2000 and 2016. In addition, this project will also assess the relative contribution of changes in long-term immobility, as well as repeat migration frequency and type, to changes in annual migration rates.

1. Introduction

This paper utilizes linked individual IRS and Census Bureau survey data on migration to investigate differences in the longitudinal migratory behavior of subsets of the U.S. population. Our interest is in the social, demographic, and geographic characteristics associated with interstate migration. The usual research strategy is to compare these characteristics between those who move and those who stay over a single period, usually defined as a year. We aim to make use of the longitudinal structure of these linked administrative and survey data to make a different set of comparisons, first comparing the characteristics of the immobile population (those who never move between 2000 and 2016) to those of the mobile population (those who move at least once), and then differentiating the mobile population by the frequency and types of moves made between 2000 and 2016 (those who move once vs twice vs three times, etc; and those who make onward moves to new destinations, or return moves to prior locations - i.e. circulations).

These comparisons will provide further insight into the drivers of migration in two ways: first, by identifying and defining the characteristics of the immobile population more precisely than in previous research; and second, by providing measurements of the characteristics of repeat migrants by type and frequency of move. We are also interested in the effect of changes in the shares of the population who are immobile or repeat migrants on changes on aggregate U.S. migration rates. For example, we calculate how much of the well-reported decline in U.S interstate migration is due to an increase in the percentage of population that never moves (i.e. who are "stuck") versus a decrease in the rate of repeat migration (Kaplan and Schulhofer-Wohl 2017).

2. Background and Motivation

Broadly speaking, our analysis of U.S. migration proceeds along two lines. The *first* makes use of cross-sectional data on migration based on survey questions inquiring where the respondent lived either one year or five years ago. The Current Population Survey (CPS) and the American Community Survey (ACS) supply these data at one-year intervals, and the long form of the decennial census provided five-year data for most of the twentieth century, ending in 2000. These data sources provide additional information on the characteristics of movers, including their age, gender, education, race, and nativity that allow for more detailed analysis of subpopulation differentials in migration propensity and directionality.

In predicting who moves in a given single-period, such as across state lines in the U.S., human capital-based migration theory predicts and empirical research using the above sources has repeatedly confirmed several trends (e.g. Sjaastad 1962, Schwartz 1976, Wright and Ellis 2018). Individuals who are young are most likely to move. Specifically, annual U.S. interstate migration rates peak for those in their early to mid-twenties and decline by half by the time individuals reach the age of 40. Also, education and specialized training enhances the likelihood of moving. Specifically, those in their twenties with four-year degrees are over twice as likely to migrate across state lines annually as those of equivalent age without a degree (Wright and Ellis 2018). There is variation in these trends by social group (gender, race, nativity) but the pattern of these age-education profiles is consistent across groups.

While the ACS and CPS provide rich data for migration analyses, the relatively small sample sizes of the ACS and CPS make it difficult to observe migration reliably for small subpopulations, such as a particular age cohort of a smaller racial or nativity group across small geographies (e.g. Conway and Rork 2016, Franklin and Plane 2006). The public releases of county-to-county flows by the Internal Revenue Service, based on address comparisons between current and previous year tax returns, do not suffer from small sample problems, but these publicly available administrative migration data contain no information on the characteristics of movers.

The *second* line of inquiry in migration analysis makes use of panel data to observe the migration behavior for individuals longitudinally. Research using these data shows that annual migration decisions are most appropriately considered as part of a sequence of onward or return moves and that human capital theory can be extended to be predictive of these sequences (DaVanzo 1983). Specifically, educated migrants are more likely to make onward moves whereas those with less formal education are more likely to return. Other factors differentiate the frequency and direction of subsequent mobility behavior, including the accumulation of experience and investment in a destination place, the distance of first move and whether adverse conditions in the origin motivated the move (e.g. unemployment). Much of this repeat migration research has been conducted with rich but small longitudinal samples, such as the Panel Study of Income Dynamics (PSID) (e.g. Clark and Withers 2007). The PSID's current sample is 24,000 individuals, which is sufficient for observing longitudinal trends in some reasonably large subpopulations, but not in specific places (e.g. metropolitan areas) or across relatively small geographic units, such as counties (Cushing and Poot 2003).

In this paper, we propose bridging these two lines of inquiry by building a population-sized longitudinal dataset of migration using linked tax-based administrative and Census Bureau survey data. We use these data to investigate migration sequences by population subgroups that current small sample panel data cannot reliably observe. We also examine the relative contribution of changes in long-term immobility, as well as repeat migration frequency and type, on changes in annual interstate migration rates. For example, we assess whether changes in the fraction of the population engaging in repeat migration is the prime driver of changes in annual interstate migration rates or whether such change primarily is a function of change in the share of the population that is long-term immobile.

3. Data and Methods

The project utilizes annual IRS data which, when linked over time, allow for a longitudinal analysis of migration that is novel in terms of detail and scale. Specifically, it uses three sources of Federal Tax Information (FTI) from the Internal Revenue Service (IRS) and one source from the Social Security Administration (SSA).[1] The first source of FTI, IRS 1040s for the tax years 2000 through 2016, is our primary source of longitudinal information. These data are scrubbed of Personally Identifiable Information (PII) like names and Social Security Numbers (SSNs), but retain much of the data individuals report when filing income taxes each year including addresses and incomes. The second source of FTI,

[1] Access to these sensitive administrative records sources was granted by the Internal Revenue Service and U.S. Census Bureau under the authority of Title 13 and Title 26 of U.S. Code and following research approval at the Center for Administrative Records Research and Applications at the Census Bureau. Given confidentiality concerns, data used in this analysis are not available for public use.

IRS 1099 "information returns" for tax years 2003 to 2016 that are created by the IRS to flag sources of income for individuals. They provide insight into a population typically overlooked in studies of migration relying solely on 1040s – namely, those who do not and/or are not required to file income taxes. The final source of FTI is the universe of IRS W2s for tax years 2005 to 2016. As opposed to IRS 1040s, which aggregate individual incomes from various sources for spouses filing jointly, IRS W2s contain detailed earnings information for *individual earners*, regardless of whether earners file separately or jointly. We also rely on administrative records from the Social Security Administration, which provide sex, date of birth, date of death, and place of birth information for the population found in IRS records. These data allow us to identify loss of individuals from our longitudinal data due to death, improving the precision of our longitudinal approach.

To the administrative records sources discussed above, we add two sources of restricted-use Census Bureau microdata – 2010 Census records and 1-year American Community Survey (ACS) responses from 2005 through 2016. The 2010 Census serves as a benchmark against which the population found in IRS records can be compared and contrasted. Fully implemented in 2005, the ACS is an ongoing survey sampling over 2.8 million housing unit addresses per year and providing a wealth of information on the demographic and socioeconomic characteristics of the U.S. population.[2]

All Personally Identifiable Information (PII) such as names, Social Security Numbers (SSNs), and Individual Taxpayer Identification Numbers (ITINs) are removed from administrative and Census Bureau records to protect privacy and reduce the risk of disclosing personal information. They are replaced with unique, anonymous personal identifiers which facilitate individual-level record linkage across data sources. These unique identifiers are assigned on SSNs or ITINs, when available, and probabilistically assigned using other PII, such as date of birth and address, when an SSN or ITIN is not available (Wagner and Layne 2014). Over 99 percent of individuals found in IRS administrative records receive a unique person identifier because SSNs or ITINs are required on tax documents. While the rates are not as high for Census Bureau surveys, they are still encouraging. Over 91 percent of individuals in the 2010 Census were assigned unique person identifiers, while the comparable rate for ACS surveys is approximately 94 percent. We restrict our analyses to the universe of individuals with person identifiers in each of the data sources above, as they are required for record linkage.[3]

Migration in the IRS administrative records, the primary interest of this paper, is inferred by longitudinally linking individuals across tax years, and then checking for changes in addresses found in 1040 and 1099 records. We can then infer migration for individuals in the IRS records by comparing addresses for an individual at two points in time, using this comparison to create retrospective measures of interstate migration between tax years *Y-1* and *Y.* We use linked data from SSA, decennial census, and

[2] The ACS design and methodology report can be found at https://www.census.gov/programs-surveys/acs/methodology/design-and-methodology.html. Additional information on ACS methodology, variable definitions, code lists, comparisons, statistical testing, and the accuracy of estimates can be found at https://www.census.gov/programs-surveys/acs/technical-documentation/code-lists.2010.html.

[3] We are aware that patterns in the failure to assign unique identifiers are non-random and may introduce bias in subsequent analyses. Research on ACS microdata has shown that racial/ethnic minorities and those of lower socioeconomic status are less likely to receive identifiers than non-Hispanic Whites and those of higher socioeconomic status (Bond et al. 2014).

the ACS to provide demographic information for the longitudinal tax migration records to distinguish the migration sequences among population subgroups and to identify censoring due to mortality.

Our previous work has demonstrated the feasibility of these linking and analysis procedures, showing how the addition of 1099s reduces selectivity bias in IRS records by including those who do not file 1040s, and by demonstrating the utility of IRS data for following individuals (i.e. their survivability) and observing their locational and other behavior over time (Foster, Ellis and Fiorio 2018a). We have made use of these data for preliminary analysis of the internal migration of the foreign-born (Foster, Ellis and Fiorio 2018b).

4. Paper Specific Outputs

By using these linked administrative and Census Bureau survey data we can parse the effects of human capital and other drivers of longitudinal interstate migration sequences reported from analyses of small panel-based datasets, such as the PSID, in much finer ways. With these near-population scale longitudinal data we can better differentiate the likelihood of being immobile (i.e. never migrating across state lines), or of repeat migration and its types (onward vs return) by age and education, thus enhancing our understandings of the relationship between the basic categories of human capital theory and longitudinal migration behaviors. Most importantly, we can extend this differentiation by gender, race, and nativity. **Our first set of outputs will therefore describe the characteristics of the population by interstate migration type observed longitudinally over the 2000-2016 period (i.e. no migration, one migration, repeat migration, return or circular migration, onward migration). We will also report estimates of a hazard model of the likelihood of interstate migration based on these data.**

In addition to investigating how the characteristics of populations relate to interstate longitudinal migration sequence frequencies and types, this project will also assess the effect of changes in the population share of these sequences to changes in annual interstate migration rates. One of the explanations for the decline in U.S. migration is the decline in repeat migration, postulated as the result of better access to information about destinations today than in the past (Kaplan and Schulhofer-Wohl 2017). The linked administrative and Census Bureau survey data that we will use will allow us to explore this effect in much greater detail across subgroups of the population than is possible with public-use data. **Accordingly, our second set of outputs will describe yearly changes in the relative share of subgroups of the population who are non-movers, first-time migrants and repeat migrants (distinguishing the latter by frequency and type: onward from return) and the contribution of each to changes in annual interstate migration rates. We will calculate the effect of changes in these relative shares on changes in migration rates by age cohort, gender, levels of educational attainment, and race/ethnicity, assessing how they affect aggregate changes in interstate migration rates for the U.S. as a whole.**

5. References

Bond, Brittany, J. David Brown, Adela Luque, and Amy O'Hara. 2014. "The Nature of the Bias When Studying Only Linkable Person Records: Evidence from the American Community Survey." CARRA Working Paper #2014-08. Available from https://census.gov/library/working-papers/2014/adrm/carra-wp-2014-08.html.

Clark, William A. V. and Suzanne Davies Withers. 2007. "Family Migration and Mobility Sequences in the United States: Spatial Mobility in the Context of the Life Course." *Demographic Research* 17:591–622.

Conway, Karen Smith and Jonathan C. Rork. 2016. "How has Elderly Migration Changed in the 21st Century? What the Data Can and Can't Tell Us." *Demography* 53(4): 1011-1025.

Cushing, Brian and Jacques Poot. 2003. "Crossing Boundaries and Borders: Regional Science Advances in Migration Modelling." *Papers in Regional Science* 83(1):317–38.

DaVanzo, J. 1983. Repeat Migration in the United States: Who Moves Back and Who Moves On? *The Review of Economics and Statistics* 65 (4):552–559.

Foster, Thomas B., Mark Ellis, and Lee Fiorio. 2018a. *The Opportunities and Challenges of Linked IRS Administrative and Census Survey Records in the Study of Migration*. CARRA-WP-2018-06. Washington DC: Center for Administrative Records Research and Applications, U.S. Census Bureau.

Foster, Thomas B., Mark Ellis, and Lee Fiorio. 2018b. *Foreign-Born and Native-Born Migration in the U.S.: Evidence from IRS Administrative and Census Survey Records*. CARRA-WP-2018-07. Washington DC: Center for Administrative Records Research and Applications, U.S. Census Bureau.

Franklin, Rachel S. and David A. Plane. 2006. "Pandora's Box: The Potential and Peril of Migration Data from the American Community Survey." *International Regional Science Review* 29(3): 231-246.Kaplan, G., and S. Schulhofer-Wohl. 2017. Understanding the Long-Run Decline in Interstate Migration. *International Economic Review* 58 (1):57–94.

Schwartz, A. 1976. Migration, Age, and Education. *Journal of Political Economy* 84 (4, Part 1):701–719.

Sjaastad, LA. 1962. "The Costs and Returns of Human Migration." Journal of Political Economy 70(5): 80-93.

Wagner, Deborah and Mary Layne. 2014. "The Person Identification Validation System (PVS): Applying the Center for Administrative Records Research and Applications' (CARRA) record linkage software." CARRA Working Paper Series #2014-01. Washington, DC. Available from https://www.census.gov/library/working-papers/2014/adrm/carra-wp-2014-01.html.

Wright and Ellis 2018. "Where STEM graduates move: human capital, employment patterns, and interstate migration in the U.S.", Forthcoming, Population, Space and Place.