Thirty Years of Five-Years-Ago Data:

County Migration from the 1980, 1990, and 2000 Censuses

Dan Mercer, MS
Johns Hopkins University
Bloomberg School of Public Health
Baltimore, MD 21205
(dmercer@jhsph.edu)

draft; do not quote

I. Introduction

This paper uses data from the residence five-years-ago questions from the 1980, 1990, and 2000 censuses (United States Census Bureau 1984, 1995, and 2005) to track county in-, out-, and net migration by age, with a particular focus on the migration of older (retirement-age) persons. A data set is assembled and the paper begins to assess the small area movement of the population over time. While the data are self-reported, and do not measure multiple moves or return moves to the same address in the five-year period, they are geographically exhaustive, covering the entire United States. Time series data and geographic detail are both necessary for proper statistical analysis.

Originally, analogous data from the 1970 census were to be included in the analysis, but time constraints prohibited their use. The migration data for the 1965 to 1970 period were available in paper format only. There was not sufficient time to put them into electronic format and process them.

According to decennial census data, the *gross* number of moves between counties in the five-year period preceding the census is about twice the *gross* number of moves between states. Gross migration is the sum of in-migration and out-migration. Those who move between counties can also move between states, of course. In the 1995 to 2000 period, gross migration between states was about 44.2 million, while gross migration between counties was about 94.8 million (http://factfinder.census.gov/servlet/QTTable?_bm=y&-geo_id=D&-qr_name=DEC_2000_SF4_U_QTP22&-ds_name=D&-_lang=en&-redoLog=false). In the 1985 to 1990 period, gross migration between states was about 43.2 million, while gross migration between counties was about 87.7 million

(http://www.census.gov/population/socdemo/migration/90mig.txt and http://www.census.gov/population/socdemo/migration/net-mig.txt).

Thus, limiting analysis to interstate migration, as many past studies have, eliminates about one-half of all longer distance migration (in contrast to residential mobility). Generally, we do not freely dispose of one-half of our data. The problem is not just one of sample size: inter-county move data have the potential to tell us more about the characteristics of places favored or disfavored by migrants than state data tell us.

Analysis at the state level likely misses a great deal of relevant heterogeneity, both among migrants and among the places they are moving from and moving to. People do not migrate to states: Migration decisions are not typically based only on state characteristics, although such characteristics may certainly play a role. Instead, migration decisions are based on the features and attributes of a specific place. In the absence of accurate measures of migration, we cannot know the effects of place or policy on that migration. Similarly, we cannot begin to accurately assess the impact of migration on place.

<insert figure: age-specific migration rates, interstate and inter-county movers>

II. Migration Data

II.A. Data Sources

There are at least eight sources of United States internal migration data. These cover different years and different geographic levels. Some of these allow both in- and

out-migration to be identified and gross migration to be computed, while others indicate only net migration.

Several well-known panel or longitudinal data sets contain migration information. The Health and Retirement Survey (HRS) has questions about migration. The Panel Study of Income Dynamics (PSID) includes migration data. Migration data can be extracted from the National Longitudinal Surveys (NLS). The Survey of Income and Program Participation (SIPP) covers migration, too.

The Public Use Microdata Samples (PUMS) contain detailed information, including migration, from decennial censuses. The Current Population Survey (CPS) inquires about changed residence. When the United States Census Bureau does its population estimates, migration is a component of the population change. Finally, there are the data from the five-years-ago questions on each of the last five decennial censuses in the U.S.

Many of these data sources represent the population only at the national level. This is true for the HRS, the PSID, the NLS, and the SIPP. The fact that they are nationally representative does not preclude the use local area information or attributes when they are used to estimate relationships, and there is nothing wrong with national data *per se*. Local data can be appended to the nationally representative samples to measure the local effects, but appending local data to nationally representative data likely limits the heterogeneity of the local data used. This will not be the case only if the local characteristics that affect migration vary perfectly with the characteristics on which the national sample is selected.

Nationally representative data cannot, however, present a comprehensive picture of the ability of migration, both in- and out-, to change the age structure of local areas. Aging in place alters age structure, but, in many areas, differential migration by age is the greater contributor to evolving age structure. It is changes in the local age structure that interest me, because I want to understand how changes in local age structure over time affect policy choices.

Some of the data sets mentioned above have large enough sample sizes and are constructed to represent regions, states, or lower levels of geography. The PUMS is an excellent source, but its minimum geographic area covers "super-PUMAs", or micro-data areas of 100,000 persons or more.[1] The CPS is representative at the national and state levels. Population estimates cover counties, but result from model-based estimates in which migration is a residual. Thus, it is left with residual problems: Errors in the other components result in mis-estimation of migration. In some of the population estimates series only net migration is available.

II.B. Five-Years-Ago Data

The "five-years-ago data" that I refer to were asked on the census long form. In 2000, question number 15 asked about residence five years earlier. Since the long form is distributed to approximately one in six households, the data are subject to sampling variability. All data referred to in this paper are point estimates. Confidence intervals have not yet been computed for the data.

---

[1] According to 2000 census data, only 524 of 3,141 counties had 100,000 or more persons. While smaller areas are aggregated for the PUMS—a minimum of 100,000 is not required—this is an interesting comparison of area sizes, and suggests the geographic consolidation that takes place in the PUMS.

The advantage of the five-years-ago data from the decennial censuses lies in their size and their geographic detail. These features permit analysis at decentralized geographic levels where policies, community characteristics, and amenities vary. The effect of policies, characteristics, and amenities on migration and, consequently, on age structure can, therefore, be studied, measured, and assessed. Without adequate sample size at these geographic levels, analysis that fully utilizes local heterogeneity is impossible.

Migrants choose a location based on the specific characteristics of a place, which means they pay attention to local attributes. True, state policies and programs have an impact on the character of localities, but, for that matter, so do federal policies. However, the local area is where people live, and its characteristics are what appeal to (or repel) them. To fully understand what area characteristics cause older persons to change locations, gross- and net-migration by age and the area characteristics must both be known.

A local area population can age dramatically, even over a short period of time, if young persons move out and, simultaneously, old persons move in. The character of a community can change accordingly. In order to understand how changes in age structure can affect the character, policy, and services offered in a local area, more needs to be understood about the age structure dynamics of local areas. The research proposed in this paper initiates that type of research.

Rapid changes are less likely to happen on the state level. Indeed, it is possible that state-level age structure can remain essentially unchanged even when there has been considerable change in age structure at the local level. The higher geographic level at

which most analysis has been done might mask dramatic changes in age structure at lower levels of geography. Only about 20 percent of all moves in the 1995 to 2000 period were between states (He and Schachter, 2003). County-level analysis will capture one-half of all movers in the 1995 to 2000 period, as about half of all moves in that period were intra-county moves.

"Churn" is very important. Net migration measures are insufficient. A place where in-migration and out-migration, measured by levels or rates, are both high differs in important ways from a place where in-migration and out-migration are both low. Both have low net migration, but for the former there are strong pull and push forces while for the latter there is little, apparently, to attract and repel migrants. Variations of this statement can be made for any two areas with similar rates of net migration, whether those rates are positive or negative. Since there is simply more than one way to get to a certain rate of net migration, the in-migration and out-migration detail is crucial to determining the pertinent characteristics of that place.

<insert figure: age-specific inter-county migration rates, 1980, 1990, 2000>

The county migration time series are desirable because such data allow fixed effects models to be estimated. Fixed effects models are intended to eliminate the fixed but unmeasured attributes of places that, uncorrected, lead to statistical endogeneity and biased parameter estimates. Unmeasured attributes are consigned to residual, or error, in statistical regression estimates. Since these unmeasured attributes are often correlated with the measured attributes, the result is that there is correlation between the error term

and the measured (explanatory) variables. This violation of ordinary least squares regression requires statistical correction, or incorrect conclusions will be drawn.

<insert figure: migration rates of persons 65 and over, 1980, 1990, 2000>

As attractive as these data are, they have many deficiencies. First, they are not micro data. There is no way to associate important individual characteristics, such as income, with individual movers (or stayers). While it may be possible to associate area characteristics, such as median household income, with county migration measures, this is much less desirable than having the individual-level data, and the variation in it.

Second, the five-years-ago data likely understates the number of migrants. While I have encountered no exploration of this issue, it seems much more plausible than the opposite. This is mostly due to the fact that multiple moves in the five-year period are not reported. If measurement error in the number of moves were normally distributed with mean zero and constant variance, it would be absorbed in the error term of the regression (Greene, 1997). Such might be the case for net migration. As in- and out-migration both have lower bounds of zero, the measurement error cannot be mean zero. For in- and out-migration, the measurement error is not normally distributed; it is skewed right. Understatement of the number of gross migrants leads, in regression estimation, to mis-measured standard errors and test statistics that are biased. They will be biased toward the alternative, leading to Type I errors. Policy variables, such as local tax features, and amenities may be judged to have an effect on in- or out-migration or both when in fact they have no such effect.

Respondents may be subject to incorrect recollections, thinking their move was more than five years in the past when it was not. Alternately, incorrect recall could lead to report of moves that occurred more than five years before the census. Do these two exactly offset each other? Not likely, especially for each county.

Death rate differences are of concern, too, in the correct reporting of these data, though it is not clear how such differences would affect the estimates. If in-migrants in one area are more likely to die before the census, then in-migration to that area will have been understated. But, the population would also be lower. As a result, since in-migration to an area is, by definition, no larger—and is almost surely smaller—than the area's population, the rate would be understated, too, since in-migration falls proportionally more than population. Similarly, out-migration from another area is understated. This requires more thought and exploration, both in this paper and in other research.

III. Literature Review

Some of the most recent and best literature on the topic is deficient with respect to the principles of area detail and fixed effects bias outlined above.

There has been a great deal of research about migration between states (Cebula, 2002; Conway and Houtenville, 2001; Conway and Houtenville, 1998; Fournier *et. al.*, 1988; Frey, 1996; Serow, 1987). Thus, the characteristics and features associated with states and strong or weak levels or rates of migration have been identified. However, as noted earlier, states are diverse places, with characteristics that vary widely across the

state. It is necessary to have a more location-specific idea of migrant origin and destination.

Theoretical and empirical research indicates that sub-state policies and programs do matter (Clark and Hunter, 1992; Graves and Knapp, 1988; Graves and Waldman, 1991; Walters, 2002). Tax and expenditure policy are set partly at the local level. Thus, to the extent that these factors are important, local characteristics will give a more complete picture of migration stimuli. Knowledge of such characteristics would allow public officials to make policy more attuned to the desires of the local community and to anticipate what types of effects specific policies might have on migration. This is harder to do on a state level because of the diversity of policies, features, and needs at the local level.

Meyer (1987) uses county data from various sources, including decennial census, Public Health Service, and County Business Patterns data. Since she does cover the period from 1940 to 1980, there is the potential to eliminate the fixed effects. However, there is no explicit mention of this in the discussion of methods. Unfortunately, she analyzes net migration, and only in the six New England states, thereby limiting the generalizability of the analysis.

If county-level data are needed to capture and measure relevant heterogeneity, time series data are needed to account for unmeasured, but fixed, effects. Mueser (1989) shows how migration analysis using cross-sectional data results in flawed estimates. Using first differences, Mueser shows that controlling for unmeasured fixed effects substantially changes the estimates. Failure to control for these effects can result in

biased coefficient estimates. This is the principle shortcoming of Clark and Hunter (1992).

Walters (2002) uses 1990 PUMS data to estimate separate relationships for in-migration and out-migration. The problem is that he does PUMA-level analysis (micro data "areas" above county and below state)—there are 732 areas. These are not political jurisdictions, and it is difficult to measure the relevant policy variables. He uses only one PUMS data series, with migration from 1985-1990, and this is subject to potential fixed effects issues. It is not clear why Walters estimates equations for three migrant groups separately rather than using a categorical variable classification to include all of them in one equation. Walters combines data, such as crime and occupation proportions, in ways that make unpalatable assumptions about the geographic distribution of the values of such variables: that they are distributed in proportion to the overall population. Moreover, other data issues, such as use of end-of-period values that are affected by migration during the period, arise; full explication of these issues is beyond the scope of this paper.

IV. A Model

The standard expression of an area's population dynamics, called the demographic balancing equation, is given by the following relationship:

$$(1) \qquad N(t) = N(0) + B(0, t) - D(0, t) + I(0, t) - O(0, t)$$

where N(t) is population size at time t, N(0) is initial population, B is the number of births in the time interval from 0 to t, D is number of deaths in the interval, I is in-migration in the interval, and O is out-migration in the interval. The demographic balancing equation is perfectly flexible in the sense that it applies to any area or age

group (When it applies to a certain age cohort, such as persons 65 to 69, the interpretation of the components changes somewhat; D then refers to those leaving the age group, whether through actual death or aging, and B to those entering.). This equation allows the components of population change to be evaluated and quantified.

The prevalence of older persons is

$$(2) \qquad \text{Old Age Prevalence}_{a,0} = \frac{N_{\geq 65}}{N}$$

The Old Age Prevalence (OAP) for area *a* at time 0 is the ratio of the number of the area's persons 65 and over at time 0 to the total number of the area's persons. Combining the stock nature of the OAP with the flow nature of the demographic balancing equation yields:

$$(3) \qquad \Delta \text{OAP}[0,t]_a = \frac{(I_{\geq 65}[0,t] - O_{\geq 65}[0,t]) + (N_{65-t} - D_{\geq 65}[0,t])}{(I[0,t] - O[0,t]) + (B[0,t] - D[0,t])}$$

I, O, N, D, and B are as defined above and refer to area *a*. The numerator of the change in this ratio is, first, the number of net migrants age 65 and over between time 0 and t to area *a*, plus, second, the net of persons "aging into" the 65 and over age group and the number of deaths in the time period 0 to t to persons who are 65 and over. The denominator is, first, the net migration between 0 and t of persons of all ages, plus the net of births over deaths to persons of all ages between 0 and t.

Borrowing from the theoretical perspective of Graves and Knapp (1988), independent variables are divide into groups based on whether they likely to affect local property values or local wages. I plan to use literature from labor economics (estimates of regional wage determinants) and real estate economics (estimates of regional property values) to help select and categorize variables. The migration decision for any person or

household is described by comparing V_1, the indirect utility derived from living in location "1", with V_2, the indirect utility derived from living in location 2. If $V_1 > V_2$ the move is made. The monthly indirectly utility measure for location "i" for person "j" can be expressed as

(4) V_{ij} = Net income = $Y_j + w_i(T_j) - r_i(A_j)$

where Y_j is non-wage income, w is wage rate and r is the rental cost of land, both of which are functions of the location's amenities. For each individual, T_j is the annual labor supply and A_j is the desired land holding. Y is not place dependent, and T_j is zero for retired persons. Thus net senior migration = $\left(I_{\geq 65}[0,t] - O_{\geq 65}[0,t]\right) = f(r_{it})$ while net total migration $\left(I[0,t] - O[0,t]\right) = f(r_{it}, w_{it})$. Ultimately, OAP will depend on the recent history of rents and wages in an area, e.g. $OAP_{it} = f(r_{it}, r_{it-1}, \ldots r_{it-\tau}, w_{it}, w_{it-1}, \ldots w_{it-\tau}, OAP_{it-1}, \ldots OAP_{it-\tau})$. And since rents and wages depend on local amenities as well as the age structure, a reduced form equation can be derived:

(5) $OAP_{it} = C + \beta_1$'Amenities Affecting $Wages_{it-\tau} + \beta_2$'Amenities Affecting Property $Values_{it-\tau} + OAP_{it-\tau} + \mu_i + \varepsilon_{it}$

V. Analytic Aims

A primary goal of this research is to obtain more geographic detail about migration streams. With the data set assembled for the paper, it will be easy to identify counties that have shown growth over time, those that have slowed in their growth, and those that have maintained consistent growth. From that point, it is possible to enumerate the characteristics associated with such places. I am ultimately concerned with "amenity migration" (Longino, 1990). Sorting this type of migration out from other types, such as

kinship migration, is difficult unless data that includes the proximity of relatives is available. A secondary goal of the research is to begin to identify possible flaws in the data. There are, perhaps, impossibilities suggested by the data, and this analysis will help find them.

As the United States population ages, cohort size will exert different pressure on migration than in past. Different cohort sizes, along with different migration rates by age, combine to alter in-, out-, and net migration levels over time. Thus, it is important to look at rates rather than levels, despite imperfections in computing rates, such as in-migration rates (Preston, 2001). Areas that are unattractive to older persons—metro areas, for example—will experience relatively large outflows when the population ages. But this is a cohort effect, not a place effect. Cohort size effects cannot be allowed to confound the estimates.

Literature

Bean, Frank D., George C. Myers, Jacqueline L. Angel, and Omer R. Galle. "Geographic Concentration, Migration, and Population Redistribution Among the Elderly". In *Demography of Aging*, Linda G. Martin and Samuel H. Preston, editors (pp. 319-355). Washington, D.C.: National Academy Press. 1994.

Cebula, Richard J. "Migration and the Tiebout-Tullock Hypothesis Revisited". *The Review of Regional Studies*, Vol. 32, No. 1 (2002), pp. 87-96.

Clark, David E., T.A. Knapp, N.E. White. "Personal and Location-Specific Characteristics and Elderly Interstate Migration". *Growth and Change*, Vol. 27 (1996), pp. 327-351.

Clark, David E. and William J. Hunter. "The Impact of Economic Opportunity, Amenities and Fiscal Factors on Age-Specific Migration Rates". *Journal of Regional Science*, Vol. 32 (1992), pp. 349-365.

Conway, Karen S., and Andrew J. Houtenville. "Elderly Migration and State Fiscal Policy: Evidence from the 1990 Census". *National Tax Journal*, Vol. 54, No. 1 (2001), pp. 103-123.

Conway, Karen S., and Andrew J. Houtenville. "Do the Elderly 'Vote with Their Feet'?" *Public Choice*, Vol. 97, (1998), pp. 663-685.

Fournier, Gary M., D.W. Rasmussen, and William J. Serow. "Elderly Migration as a Response to Economic Incentives". *Social Science Quarterly*, Vol. 69 (1988), pp. 245-260.

Frey, William H., K.-L. Liaw, and G. Lin. "State Magnets for Different Elderly Migrant Types in the United States". *International Journal of Population Geography*, Vol. 6 (2000), pp. 21-44.

Gale, Lewis R. and Will C. Heath. "Elderly Internal Migration in the United States Revisited". *Public Finance Review*, Vol. 28, No. 2 (2000), pp. 153-170.

Graves, Philip E. and Donald M. Waldman. "Multimarket Amenity Compensation and the Behavior of the Elderly". *American Economic Review*, Vol. 81 (1991), pp. 1382-1390.

Graves, Philip E. and Thomas A. Knapp. "Mobility Behavior of the Elderly". *Journal of Urban Economics*, Vol. 24 (1988), pp. 1-8.

Graves, Philip E. and Peter D. Linneman. "Household Migration: Theoretical and Empirical Results". *Journal of Urban Economics*, Vol. 6 (1979), pp. 383-404.

Greene, William H. *Econometric Analysis*, Third Edition. New Jersey: Prentice Hall. 1997.

Greenwood, Michael J. "Internal Migration in Developed Countries". In *Handbook of Population and Family Economics*, Mark R. Rosenzweig and Oded Stark, eds. Amsterdam: Elsevier (1997).

He, Wan and Jason P. Schachter. "Internal Migration of the Older Population: 1995 to 2000." (*Census 2000 Special Report* CENSR-10). August 2003: U.S. Census Bureau.

Litwak, Eugene and Charles F. Longino, Jr. "Migration Patterns Among the Elderly: A Development Perspective". *The Gerontologist*, Vol. 27, No. 3, (1987), pp. 266-272.

Longino, Charles F. "Geographical Distribution and Migration". In *Handbook of Aging and the Social Sciences*. Third Edition. New York: The Academic Press. 1990.

Meyer, Judith W. "Country Characteristics and Elderly Net Migration Rates". *Research on Aging*, Vol. 9 (1987), pp. 441-452.

Meyer, Judith W. and A. Speare, Jr. "Distinctly Elderly Mobility: Types and Determinants". *Economic Geography*, Vol. 61, No. 1 (1985), pp. 79-88.

Mueser, Peter R. "Measuring the Impact of Locational Characteristics on Migration: Interpreting Cross-sectional Analyses". *Demography*, Vol. 26 (1989), pp. 499-513.

Newbold, K.B. "Determinants of Elderly Interstate Migration in the United Sttes, 1985-1990". *Research on Aging*, Vol. 18 (1996), pp. 451-476.

Plane, D.A. & P.A. Rogerson. "Tracking the Baby Boom, the Baby Bust, and the Echo Generations: How Age Composition Regulates U.S. Migration". *Professional Geographer*, Vol. 43 (1991), pp. 416-430.

Preston, Samuel H., Patrick Heuveline, and Michele Guillot. *Demography: Measuring and Modeling Population Processes*. Oxford, UK: Blackwell. 2001.

Rogers, Andrei. "The Elderly Mobility Transition: Growth, Concentration, and Tempo". Research on Aging, Vol. 11 (1989), pp. 3-32.

Rosenbloom, Joshua L., and William A. Sundstrom. "The Decline and Rise of Interstate Migration in the United States: Evidence from the IPUMS, 1850-1990. NBER Working Paper Series, No. 9847. Cambridge, MA: National Bureau of Economic Research. 2003.

Serow, William J. "Determinants of Interstate Migration: Differences Between Elderly and Nonelderly Movers". *Journal of Gerontology*, Vol. 42, No. 1 (1987), pp. 95-100.

Thomas, D.S. "Age and Economic Differentials in Interstate Migration". *Population Index*, Vol. 24 (1958), pp. 313-325.

United States Census Bureau. 2000 migration data: http://www.census.gov/population/www/cen2000/ctytoctyflow.html. Downloaded May 26, 2004.

United States Census Bureau. *1990 Census of Population and Housing: County-to-County Migration Flow Files (Special Project 312)*. Washington, D.C.: Government Printing Office. 1995.

United States Census Bureau. *Supplementary Report: Gross Migration for Counties: 1975 to 1980*, 1980 Census of Population. Washington, D.C.: Government Printing Office. March 1984.

United States Census Bureau. "Gross Migration by County: 1965 to 1970". *Current Population Reports*, Series P-25, No. 701. Washington, D.C.: Government Printing Office, 1977.

Walters, William H. "Place Characteristics and Later-Life Migration". *Research on Aging*, Vol. 24, No. 2 (2002), pp. 243-277.

Wilson, Franklin D. "Cohort Size Effects and Migration". *International Migration Review*, Vol. 17 (1983), pp. 485-504.

Wiseman, R.F. and C.C. Roseman. "A Typology of Elderly Migration Based on The Decision-making Process". *Economic Geography*, Vol. 55 (1979), pp. 324-337.

Wolf, Douglas A. and Charles F. Longino, Jr. "Our "Increasingly Mobile Society"? The Curious Persistence of a False Belief". *The Gerontologist*, Vol. 45, No. 1 (2005), pp. 5-11.