The American Community Survey's Interstate Migration Data: Strategies for Smoothing Irregular Age Patterns

James Raymer, Andrei Rogers

Abstract

Age- and origin-destination-specific flows obtained from population samples often contain irregularities. The reason for this has mostly to do with the fact that migrations are relatively rare events. Biases in the analysis of migration flows can arise if these irregularities are not corrected for. Furthermore, accurate migration data are needed to understand population change and migration behavior. In this paper, we illustrate some typical examples of age-specific migration flows with irregular patterns, using the 2000-2005 American Community Survey (ACS) data. We then demonstrate how model migration schedules, log-linear models or a combination of both can be used to smooth the irregularities.

Southampton Statistical Sciences Research Institute
Methodology Working Paper M07/13

University of Southampton

The American Community Survey's Interstate Migration Data: Strategies for Smoothing Irregular Age Patterns

James Raymer
Division of Social Statistics
School of Social Sciences
University of Southampton (UK)

Andrei Rogers
Population Program
Institute of Behavioral Science
University of Colorado at Boulder (USA)

ABSTRACT
Age- and origin-destination-specific flows obtained from population samples often contain irregularities. The reason for this has mostly to do with the fact that migrations are relatively rare events. Biases in the analysis of migration flows can arise if these irregularities are not corrected for. Furthermore, accurate migration data are needed to understand population change and migration behavior. In this paper, we illustrate some typical examples of age-specific migration flows with irregular patterns, using the 2000-2005 American Community Survey (ACS) data. We then demonstrate how model migration schedules, log-linear models or a combination of both can be used to smooth the irregularities. The age-specific interstate migration flows observed in the U.S. West Region during 1995-2000, obtained from the 5% Public Use Microdata Sample (PUMS) of the 2000 Census long-form questionnaire, are used to demonstrate the effectiveness of these models. Because we have the corresponding full sample census data, the accuracy of the various smoothed estimates can be assessed. The models are then applied to smooth 2004 ACS migration flow data, which represents a "worse case" type scenario. The results clearly show that more accurate and believable migration data can be provided by applying models to smooth the irregularities in the age patterns caused by relatively small samples.

Key words: American Community Survey, internal migration, age patterns, model migration schedules, log-linear models

INTRODUCTION

Age- and origin-destination-specific migration flows obtained from population samples often contain irregularities due to the fact that migrations are relatively rare events, i.e., most people remain in their region (or state) of residence for a one-year or five-year period. If not examined or corrected, such irregularities can lead to misleading analyses of the data. This issue is particularly relevant in the United States today because of the recent replacement there of the main source of internal migration data. Historically, the decennial censuses have provided researchers with detailed internal migration flow data. Now, it will be the job of the American Community Survey (ACS).

General descriptions of the ACS and external evaluations and comparisons of the 1999-2001 estimates with the 2000 Census data for a number of countries were carried out under contracts with the Census Bureau. A recent special issue of this journal describes the results (i.e., Gage 2006; Gaines 2006; Griffin and Waite 2006; Hough and Swanson 2006; Salvo and Lobo 2006; Scardamalia 2006). Unfortunately, however, no evaluations of migration data were included. The general absence of evaluations of the ACS migration data (Franklin and Plane 2006) led Koerber (2007) to present his exploratory assessments of the migration flows reported in the 2005 ACS data at the 2007 Annual Meeting of the Population Association of America in New York City. But his assessment did not deal with age-specific migration data or their age profiles.

In this paper, we present some examples of ACS migration flows with irregular age patterns and compare them with the corresponding data obtained from the 2000 Census. The multiexponential model migration schedule (Rogers and Castro 1981; Rogers et al. 1978) and the categorical log-linear model (Raymer and Rogers 2007; Rogers et al. 2003b) are then presented for the purpose of smoothing such irregular age-specific migration patterns. The model migration schedule approach can be considered a “bottoms-up” approach that smoothes the age profile of each flow in a migration flow table. The log-linear model, on the other hand, can be considered a “top-down” approach in which higher-order marginal totals of an origin-by-destination-by-age table of migration flows are assumed to be more reliable (and regular) than lower-order marginal totals or cell values. Here, the data may be smoothed by removing two-way and three-way interaction effects. Finally, we show how model migration schedules can be incorporated into log-linear models to form hybrid models for improvements in both fit and parsimony. These approaches are first applied to smooth age-specific migration flows between states in the U.S. West region during the 1995-2000 period, obtained from the 5% Public Use Microdata Sample (PUMS) of the 2000 census long-form questionnaire. Since we have the full sample census data, we can assess the relative reliability and accuracy of such an exercise. We then go on to smooth the irregularities found in the 2004 age-specific interstate ACS public use migration data, which represents a "worse case" scenario. The paper ends with a discussion.

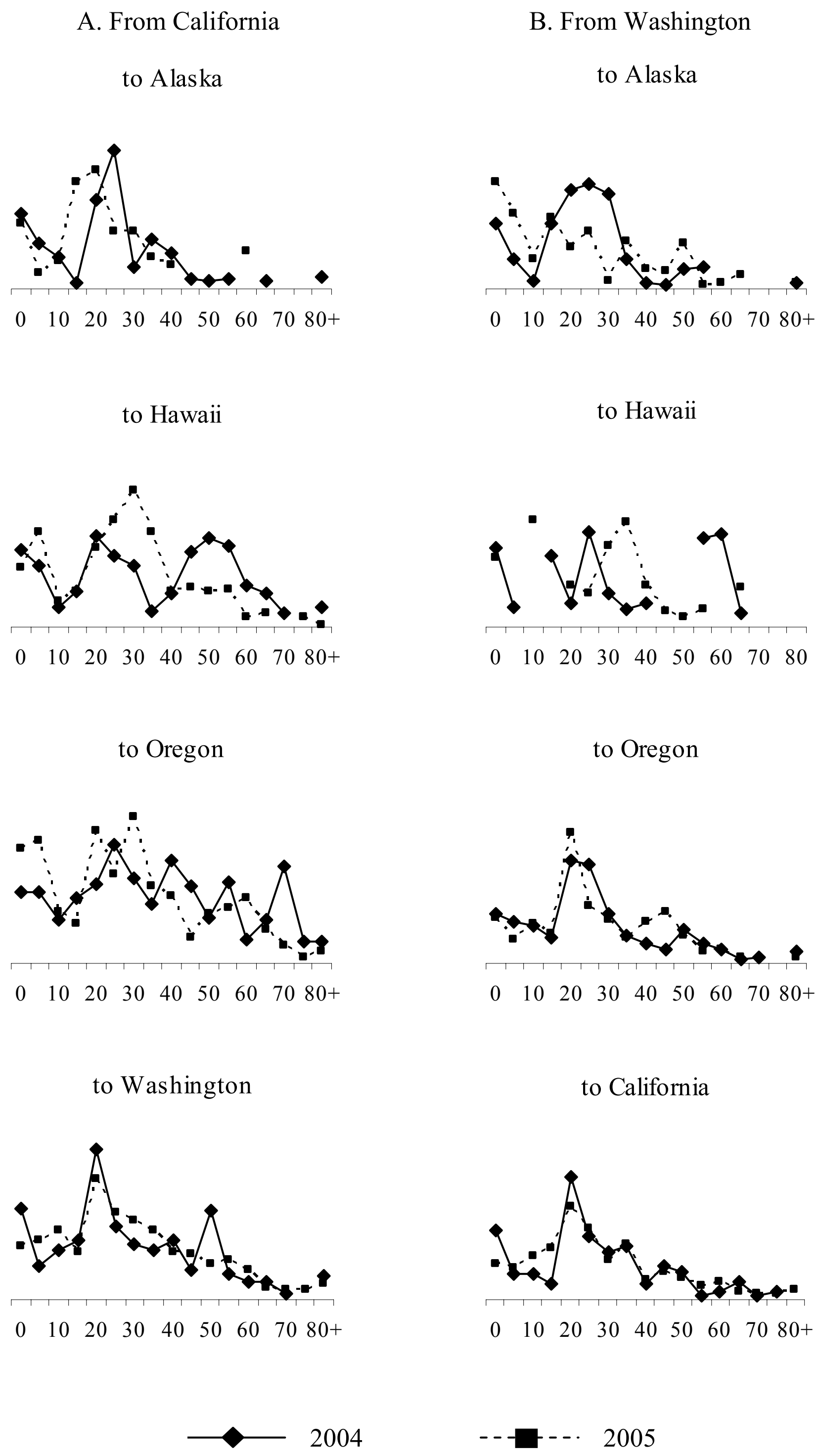
AGE-SPECIFIC MIGRATION DATA FROM THE ACS PUBLIC USE FILES, 2000-2005

The age-specific patterns of migration from the *annual* ACS data are examined in this section from 2000 to 2005 for the purpose of identifying the more reliable structures contained in these data. The data were obtained from the IPUMS (Integrated Public Use Microdata Series) website (http://usa.ipums.org/usa/), which represent 5% samples of the

ACS. The motivation for this research comes from finding many interstate migration flows with irregular age-specific shapes, such as those set out in Figure 1 for migration from California and Washington to other states in the Pacific region (i.e., Alaska, Hawaii, and Oregon) during 2004 and 2005. According to Mather et al. (2005), the ACS data should be averaged over five years to obtain estimates of quality that are of similar quality to past census data (see also Griffin and Waite 2006 for an overview). We focus on the annual data because we want to highlight the irregularities found in each of the samples publicly available. For simplicity, we only examine the migration patterns between these five states and four other regions: Mountain, Northeast, Midwest and South (i.e., a 9-region system with, 72 interregional flows). Keep in mind that California, Oregon and Washington have relatively large populations, whereas Alaska and Hawaii have small populations. Clearly, the annual public use ACS migration data breaks down at the interstate level if disaggregated into age groups, even from large states, such as California and Washington, but the situation is much worse for flows from small population states. Interestingly, Franklin and Plane (2006) were more concerned about the ACS not being able to provide detailed migration data at the county-to-county level. Given the current sampling frame, it appears that most of the inter-county data would be unreliable at any level of detail, unless pooled over a very long period.

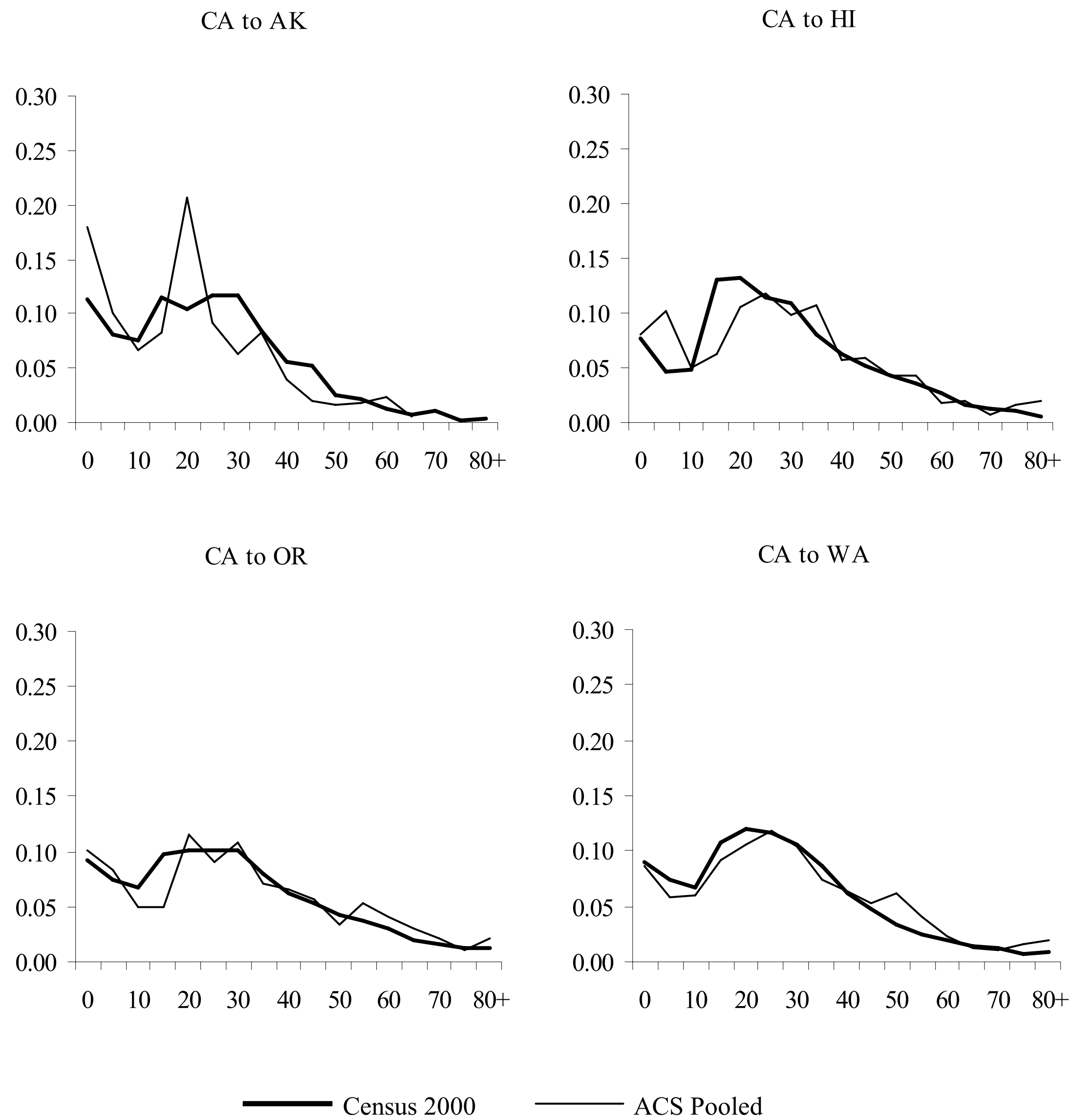
The problem with analyzing migration flow data without correcting for irregularities, such as those presented in Figure 1, is that one might misinterpret the data. As a general rule, we expect age patterns of migration to change gradually over time. For example, Raymer et al. (2006) examined annual interregional migration flows in Italy from 1970 to 2000 obtained from population registers and found strong stability and regularity in the age patterns over time (see also Raymer and Rogers 2007 and Rogers et al. 2002 for stability in U.S. and Mexico age-specific *census* migration data over time). Such stabilities do not appear in the ACS data. For example, the labor force peak of the migration flow from California to Alaska came at an older age than the corresponding peak in 2005. Was this caused by a shift in behavior or by irregularities due to sample size? In 2004, the migration flow from California to Hawaii contains a large retirement peak, whereas in 2005, it does not. Why? Unlike the above two examples, the age profiles of migration for the California-Washington, Washington-Oregon and Washington-California flows are more consistent over time with age patterns we generally expect to find (albeit with some small irregularities).

Next, consider a comparison of age profiles of migration from California to Alaska, Hawaii, Oregon and Washington presented in Figure 2, which were obtained from the 2000 Census (full sample) and from the ACS, pooled from 2000 to 2005. Even when pooled over six years, the ACS data still exhibit some irregularities in comparison with the corresponding census data. Moreover, the shapes of the ACS data are substantially different from those of the Census data for three out of the four flows. Are these differences in shapes due to problems with the sample, accumulated over time, or because of differences in the question asked? Note, the migration question of the ACS survey asks persons 1 year and over where they lived 1 year ago, whereas the migration question of the Census covers persons 5 years and over and where they lived five years ago. Thus, the two sets of measures are incompatible and give rise to a "1-year/5-year" problem that complicates a strict comparison between the two (Rogers et al. 2003a).



Note: y-axis = level (proportion) and x-axis = age

Figure 1. Age compositions of interstate migration from California and Washington to the other states in the Pacific Division, ACS data, 2004 and 2005



Notes: (1) AK = Alaska, CA = California, HI = Hawaii, OR = Oregon, and WA = Washington; (2) y-axis = proportions; (3) x-axis = age.

Figure 2. A comparison of interstate migration age compositions from California to the other states in the Pacific region: ACS 2000-2005 pooled (one-year interval) and Census 2000 full sample (five-year interval)

So, what aspects of the annual ACS migration data may be considered reliable? One way to answer this question is to examine the various age and spatial structures over time contained in the migration flows between the five Pacific states and the Mountain, Northeast, Midwest and South regions. We do this using a multiplicative component model (Raymer et al. 2006; Raymer and Rogers 2007). Such a categorical data approach to analysis allows us to identify the more stable aspects of the migration flow data over time, which can then be used to guide the (log-linear) smoothing of the age-specific irregularities in the data.

The multiplicative component model for an origin (*O*) by destination (*D*) by age (*A*) table of migration flows is specified as

$$n_{ijx} = (T)(O_i)(D_j)(A_x)(OD_{ij})(OA_{ix})(DA_{jx})(ODA_{ijx}) \qquad i \neq j \qquad (1)$$

where n_{ijx} is an observed flow of migration from origin *i* to destination *j* for age group x (i.e., 0-4, 5-9, ..., 80+ years, measured at the beginning of the one-year or five-year time interval). There are eight multiplicative components in total: an overall level, three main effects, three two-way interaction components and a single three-way interaction component. Note, for analysis and estimation purposes, the three-way interaction component ODA_{ijx} is generally ignored because (1) the other seven components capture nearly all of the patterns and (2) because it has a relatively complex interpretation (Raymer et al. 2006).

The components are calculated with reference to the total level in the migration flow tables. The *T* component represents the total number of all migrants in the system,

$$T = \sum_{ijx} n_{ijx} = n_{+++} \,. \qquad (2)$$

The main effect components, O_i, D_j and A_x, represent proportions all migration from each origin, to each destination, and in each age group, respectively, i.e.,

$$O_i = \frac{\sum_{jx} n_{ijx}}{\sum_{ijx} n_{ijx}} = \frac{n_{i++}}{T} \qquad (3)$$

$$D_j = \frac{\sum_{ix} n_{ijx}}{\sum_{ijx} n_{ijx}} = \frac{n_{+j+}}{T} \qquad (4)$$

$$A_x = \frac{\sum_{ij} n_{ijx}}{\sum_{ijx} n_{ijx}} = \frac{n_{++x}}{T} \qquad (5)$$

The two-way interaction components represent the ratios of observed migration to expected migration (for the case of no interaction) and are calculated as

$$OD_{ij} = \frac{n_{ij+}}{(T)(O_i)(D_j)} \qquad (6)$$

$$OA_{ix} = \frac{n_{i+x}}{(T)(O_i)(A_x)} = \frac{n_{i+x}/n_{i++}}{n_{++x}/n_{+++}} = \frac{p_{i+x}}{A_x} \qquad (7)$$

$$DA_{jx} = \frac{n_{+jx}}{(T)(D_j)(A_x)} = \frac{n_{+jx}/n_{+j+}}{n_{++x}/n_{+++}} = \frac{p_{+jx}}{A_x} \qquad (8)$$

where p_{ijx} denotes the age composition of migration, expressed in proportions (i.e., $p_{ijx} = n_{ijx} / n_{ij+}$). The OD_{ij} component captures the association or "connectedness" between origins and destinations. The OA_{ix} and DA_{jx} components represent the age-specific deviations in the age compositions of in-migration and out-migration (i.e., p_{i+x}

and p_{+jx}, respectively) from the overall age composition of migration (i.e., p_{++x} or A_x). Finally, although not analyzed or estimated in this chapter, the ODA_{ijx} component is calculated as:

$$ODA_{ijx} = \frac{n_{ijx}}{(T)(O_i)(D_j)(A_x)(OD_{ij})(OA_{ix})(DA_{jx})}. \quad (9)$$

The overall levels (T) are set out in Figure 3 for the ACS data from 2000 to 2005. As we expect stability over time, it appears that the overall levels are being captured adequately, with the levels ranging from 4.2 million in 2000, down to 3.9 million in 2003 and up to 4.5 million in 2005. As illustrated in Figures 4 and 5, the O_i, D_j and A_x components also exhibited relative stability in their patterns over time with the expected patterns appearing, for example, with California sending and receiving the largest share of migrants in the five states of the Pacific region, and Alaska and Hawaii the least. In other words, it appears that the origin, destination and age main effect components of the ACS data are generally reasonable and reliable over time.

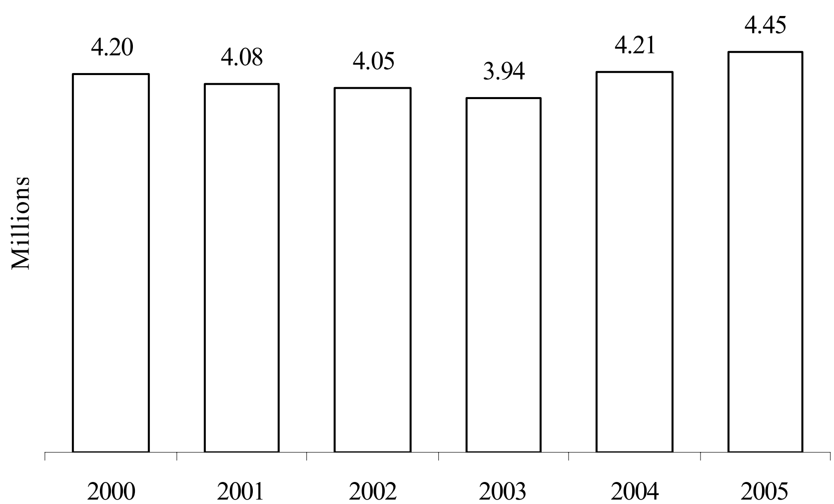


Figure 3. Overall levels of annual migration between states in the Pacific and the Mountain, Northeast, Midwest, and South regions, ACS data, 2000-2005

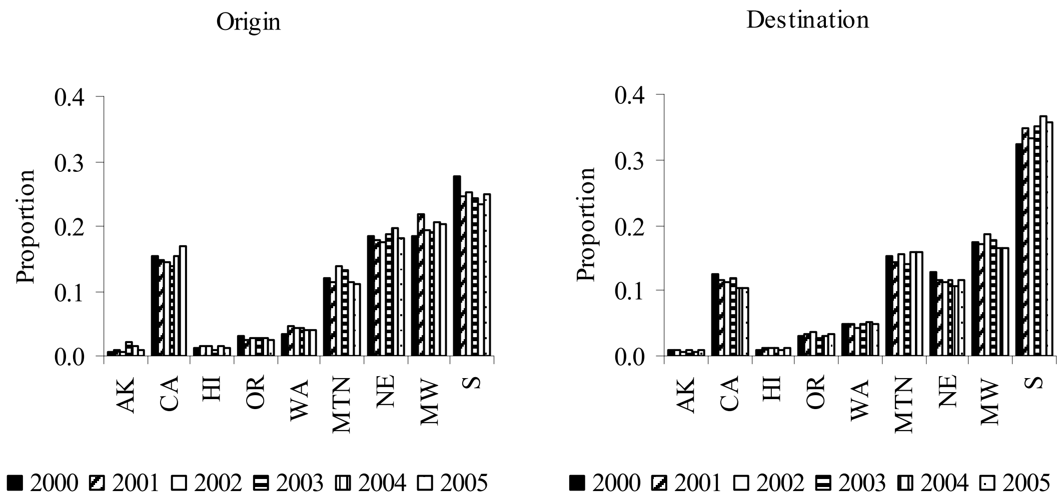


Figure 4. Proportions of all migrants from and to states in the Pacific and the Mountain, Northeast, Midwest, and South regions, ACS data, 2000-2005

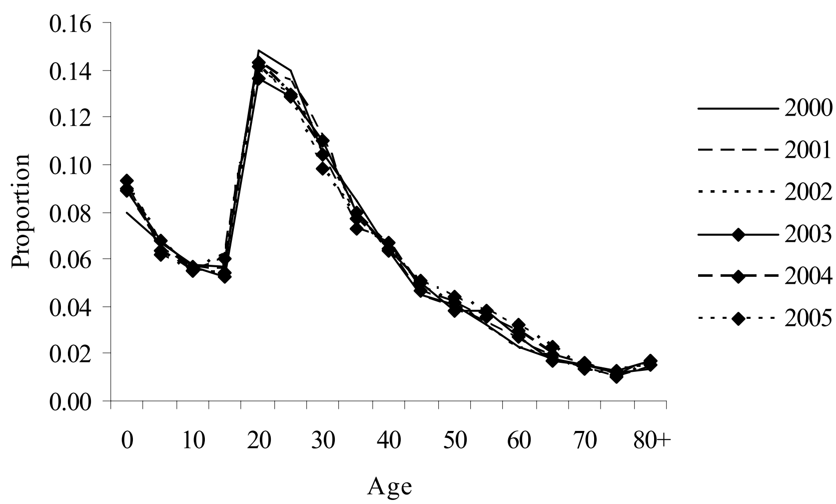
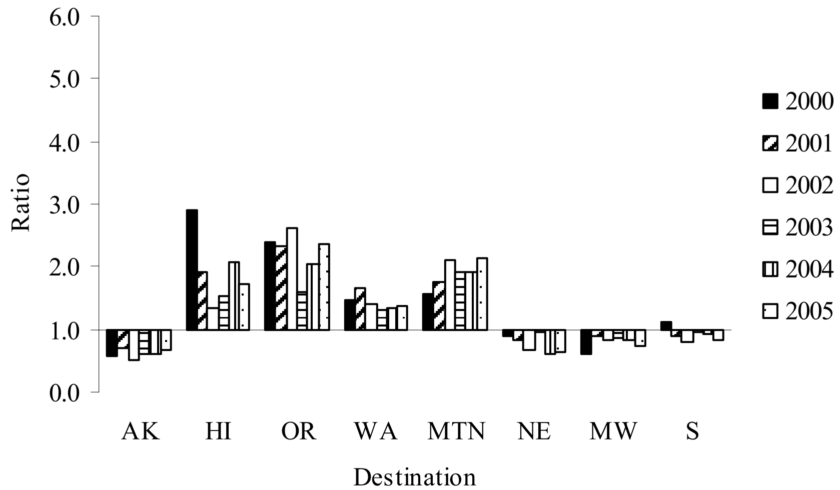


Figure 5. Proportions of migration by age between states in the Pacific and the Mountain, Northeast, Midwest, and South regions, ACS data, 2000-2005

The OD_{ij} components are compared over time in Figure 6 for migration flows from California and Washington. Here, there also appears to be some stability over time in the patterns with the expected trends (i.e., we expect to find strong associations in the migration patterns between neighboring states or regions and weaker associations between non-neighboring states or regions). For the purposes of this paper, we assume that these associations are reliable, even though the patterns are not as regular as those found in the main effect components. This means that, based on the analysis of the T, O_i,

D_j, A_x and OD_{ij} components, we conclude that the aggregate interstate migration levels of the ACS can be trusted to some extent, which leaves us to focus on the age profiles of migration which clearly are not reliable. Future research should explore the reliability of the aggregate origin-destination-specific flows (e.g., the very high association between Washington and Alaska observed in 2005). Note, we have not explored the marginal structures of the age patterns (captured by the OA_{ix} and DA_{jx} components) here. Later, we show that they are more reliable than the age- and origin-destination-specific patterns (see Figure 1) but that they still contain irregularities that have to be smoothed.

A. From California



B. From Washington

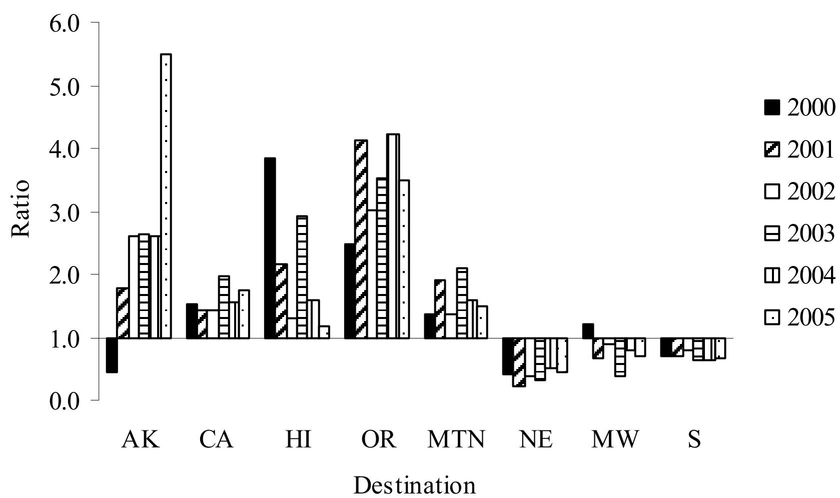


Figure 6. Origin-destination associations of migration from California and Washington to states in the Pacific and the Mountain, Northeast, Midwest, and South regions, ACS data, 2000-2005

MODELS FOR SMOOTHING AGE-SPECIFIC MIGRATION FLOWS

Three methods for smoothing irregular age-specific migration data are presented in this section. The *multiexponential model migration schedule* is a “bottoms-up” approach that smoothes each age-specific flow by fitting a non-linear curve to the available data. The *log-linear model*, on the other hand, is a “top-down” approach that smoothes each age-specific flow by removing particular interactions contained in a three-way contingency table (e.g., an origin-by-destination-by-age table). Finally, the hybrid *log-linear with offset model* may be used to combine the two approaches, for example, by including *smoothed* marginal age structures in the offset.

Model Migration Schedules

Model migration schedules can be used to smooth irregularities found in age-specific migration flow data. In this paper, we focus on the age compositions of migration, p_{ijx}, assuming that the *aggregate* origin-destination-specific flows, n_{ij+}, are reliable. The smoothed age-specific migration flows are obtained by $\hat{n}_{ijx} = \hat{p}_{ijx} n_{ij+}$, where $\hat{n}_{ijx}$ and $\hat{p}_{ijx}$ denote predicted age-specific flows and proportions, respectively, of interregional migration.

In examining the 2000 Census migration data, there were no retirement peaks found in the flows between the five Pacific region states and the Mountain, Northeast, Midwest and South regions. Therefore, we just apply the seven-parameter model schedule to smooth the irregular age compositions of migration. [1] This model is specified as

$$\hat{p}_{ijx} = a_0 + a_1 \exp(-\alpha_1 x) + a_2 \exp\{-\alpha_2 (x - \mu_2) - \exp[\lambda_2 (x - \mu_2)]\} \qquad (10)$$

and consists of three components: a constant minimum level of migration, a negative exponential curve that represents child migrant flows and a double-exponential curve that represents the young adult migrants around the age of the “labor force peak” (Rogers and Castro 1981). The *a* parameters are level parameters, whereas the α, μ and λ parameters affect only the shapes of the curves.

Log-Linear Models

Log-linear models use maximum likelihood methods for parameter estimation and assume that the counts are Poisson distributed. The *saturated* model for an origin-by-destination-by-age table, analogous to the multiplicative component model (Equation 1), perfectly predicts the flows. This model is specified as

$$\log n_{ijx} = \tau + \tau_i^O + \tau_j^D + \tau_x^A + \tau_{ij}^{OD} + \tau_{ix}^{OA} + \tau_{jx}^{DA} + \tau_{ijx}^{ODA}, \qquad (11)$$

with τ denoting the log-linear model parameters. Migration flow tables can be smoothed by simply dropping various two-way or three-way interaction terms. For example, the *unsaturated* model:

$$\log \hat{n}_{ijx} = \tau + \tau_i^O + \tau_j^D + \tau_x^A + \tau_{ij}^{OD}, \qquad i \neq j \qquad (12)$$

[1] Schedules with retirement peaks can be modeled by adding another (four-parameter) double-exponential curve to the seven-parameter model migration schedule equation.

provides estimates of migration flows that are consistent the observed aggregate levels (i.e., n_{ij+}) but have a single age profile of migration applied to all flows, represented by τ_x^A.

Incorporating Model Migration Schedules in Log-Linear Models

The motivation for this third approach comes from recent work on estimating age-specific migration flows in the context of internal migration in the U.S. (Raymer and Rogers 2007) and international migration in Europe (Raymer forthcoming). Various structures can be included with unsaturated models via offsets. For example, the model,

$$\log \hat{n}_{ijx} = \log n^*_{ijx} + \tau + \tau_i^O + \tau_j^D + \tau_{ij}^{OD} . \tag{13}$$

provides estimates of migration flows that are consistent the observed aggregate levels of flows (i.e., n_{ij+}) but *borrow* age profiles of migration from the offset or auxiliary data set, n^*_{ijx}. With the offset, model migration schedules of the reported data, or of the aggregate flows of in-migration or out-migration, can be incorporated to improve the prediction of all age-specific interregional migration flows while, at the same time, predicting flows that fit the margins of the origin-destination-age migration table that are believed to be reliable.

In the next section, we test the accuracy with which model migration schedules and unsaturated log-linear models fit the U.S. 2000 Census data, where we have both the full sample and the PUMS 5% sample (of the full sample). The smoothed estimates of the PUMS data are compared against the full sample data by using the square of the correlation coefficient (r^2) as a goodness-of-fit test. The r^2 is appropriate as we are mainly interested in the assessing the predicted *shapes* of the age-specific flows.

SMOOTHING THE 2000 PUMS MIGRATION DATA

The accuracies of the model migration schedule and unsaturated log-linear model approaches are assessed in this section by comparing their fits to the 2000 PUMS migration data with the corresponding full sample data. For this application, the unsaturated log-linear model includes all two-way interactions, specified as

$$\log \hat{n}_{ijx} = \tau + \tau_i^O + \tau_j^D + \tau_x^A + \tau_{ij}^{OD} + \tau_{ix}^{OA} + \tau_{jx}^{DA}, \qquad i \neq j . \tag{14}$$

The PUMS migration data have some of the same problems exhibited by the ACS data, although they are not as prevalent or significant. In fact, the PUMS data captures most of the full-sample age-specific interstate migration patterns. To identify irregularities, we expanded our analysis to include migration between twenty states or regions, that is, migration between the thirteen states in the West region and the seven divisions outside the West region (i.e., New England, Middle Atlantic, East North Central, West North Central, South Atlantic, East South Central, and West South Central). Furthermore, because of the generally high correspondence between the full sample and the PUMS data, we focus on migration from Colorado, because of its relatively low levels of migration and more irregular patterns.

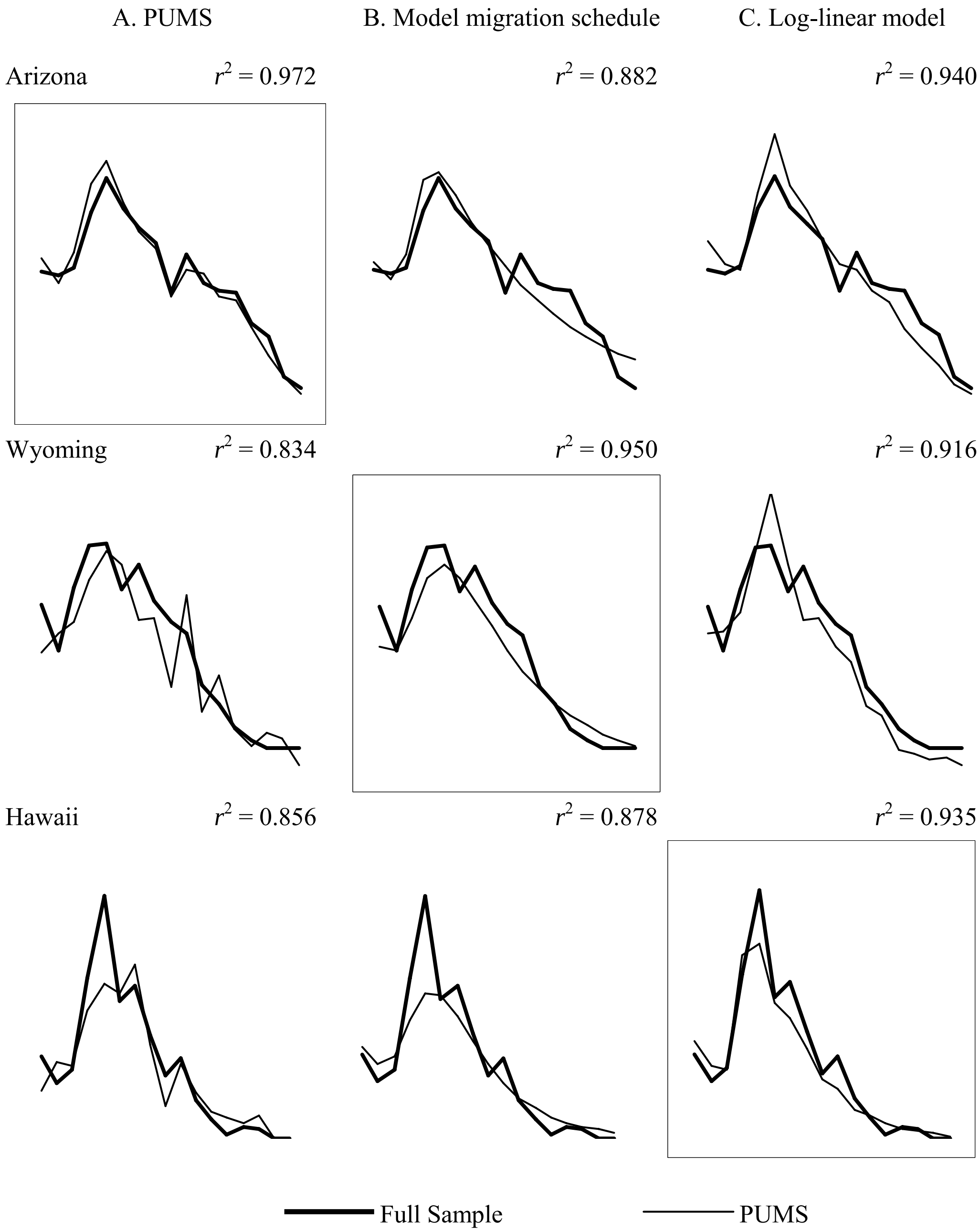
The r^2 goodness-of-fit measures comparing the PUMS, the model migration schedule fits, and the unsaturated log-linear model fits are set out in Table 1. Here, we find that, on average, the unsaturated log-linear model produced the best results, both in terms of individual fits and in terms of variance in the estimated patterns. However, there were many instances where the model migration schedule performed better, and even a couple of instances where the unaltered PUMS data represented the best fit to the full sample data.

Table 1. Goodness-of-fit tests (r^2, square of the correlation coefficient) of age compositions of migration from Colorado, Census 2000 data, 1995-2000

Destination	5% PUMS	Model Migration Schedule	Log-Linear Model
Alaska	0.942	0.885	**0.975**
Arizona	**0.972**	0.882	0.940
California	0.988	0.983	**0.998**
Hawaii	0.856	0.878	**0.935**
Idaho	0.858	0.955	**0.958**
Montana	0.774	0.809	**0.952**
Nevada	0.938	**0.970**	0.905
New Mexico	0.940	0.936	**0.944**
Oregon	0.972	0.968	**0.980**
Utah	0.975	0.951	**0.983**
Washington	**0.993**	0.980	0.977
Wyoming	0.834	**0.950**	0.916
New England	0.980	**0.985**	0.975
Middle Atlantic	0.989	**0.990**	0.976
East North Central	0.977	0.976	**0.996**
West North Central	0.984	**0.992**	0.976
South Atlantic	0.994	**0.997**	0.970
East South Central	0.977	**0.983**	0.974
West South Central	**0.993**	0.979	0.989
Average	0.944	0.950	**0.964**
Minimum	0.774	0.809	0.905
Maximum	0.994	0.997	0.998
Standard Deviation	0.065	0.051	0.026

Notes: (1) Predicted values are compared to full sample 2000 Census data. (2) Best fits are set in boldface.

For a better understanding on how the PUMS, model schedule fits, and log-linear model fits differ, a selection of flows representing age-specific migration from Colorado to Arizona, Wyoming and Hawaii are set out in Figure 7 with each flow representing a situation of best fit. The PUMS data best represented the Colorado to Arizona flow because it both corresponded to the full sample pattern and captured the unusual dip after the labor force peak. The model schedule simply fitted a line through the dip. And the log-linear model produced a sharper labor force peak, which came from the marginal age structures of out-migration from Colorado and in-migration to Arizona. For the Colorado to Wyoming flow, the PUMS data exhibited irregular patterns compared to the full sample data. Here, the model migration schedule was able to produce a curve closer to the full sample flow by fitting a line through the irregular patterns. This was, however, not the case for the Colorado to Hawaii flow, where the model schedule fit was unable to represent the sharp labor force peak because it was not captured in the PUMS data. In this case, the log-linear model, again relying on marginal age structures with a sharp labor force peak, performed better.



Notes: (1) y-axis = level (count) and x-axis = age. (2) See Table 1 for goodness-of-fits.

Figure 7. Observed and predicted age-specific migration flows from Colorado to Arizona, Wyoming and Hawaii, Census 2000 data, 1995-2000 flows: PUMS 5%, model migration schedule fits of PUMS 5% data, and unsaturated log-linear model fits of PUMS 5% data

SMOOTHING THE 2004 ACS MIGRATION DATA

The 2004 ACS data are smoothed in this section, using the same geography set out in the second section of this paper. In addition to the model migration schedule and the unsaturated log-linear model (with all two-way interactions included), we also fitted the log-linear with offset model (or "hybrid log-linear model"), specified as

$$\log \hat{n}_{ijx} = \log n^*_{ijx} + \tau + \tau^O_i + \tau^D_j + \tau^A_x + \tau^{OD}_{ij} , \qquad (15)$$

where the offset contains smoothed marginal age patterns. Specifically, this involved obtaining smooth estimates of the OA_{ix} and DA_{jx} components (see Equations 7 and 8). Here, model schedules were fitted to the aggregate in-migration and out-migration age compositions (i.e., $\hat{p}_{i+x}$ and $\hat{p}_{+jx}$) and then divided by the overall age composition of migration A_x (or p_{++x}; see also Figure 5). The offset n^*_{ijx} was constructed by multiplying the smoothed estimates of OA_{ix} and DA_{jx} by all the other components, except ODA_{ijx} (see Equation 1; also refer to Raymer et al. 2006 for a more detailed discussion of the methodology and parameter constraints). This hybrid log-linear model is considered our best model because it (1) relies on the marginal structures, which are considered more reliable, and (2) produces smoothed age patterns throughout.

Some selected results from the three model fits are illustrated in Figure 8. The Hawaii to Alaska flow represents a situation where only seven data points are available. The Hawaii to California and California to Oregon flows are cases where the patterns are highly irregular. And, the Washington to Oregon flow contains an age profile that is fairly regular with the exception of a small peak at the 50-54 age group. The hybrid log-linear model produces the best results, particularly for this situation. Clearly the unsaturated model is inappropriate because the marginal structures are highly irregular. Model migration schedules have the advantage of making the most use out of the reported data but they involve a large amount of work and can fail when the data are highly irregular. For the 2004 ACS data, we were able to fit model schedules to nearly all of the 72 flows. However, eight flows were deemed very difficult or impossible to fit model schedules to. These included the Alaska-Hawaii, Alaska-Oregon, Alaska-Northeast, Hawaii-Oregon, Oregon-Hawaii, Washington-Hawaii, Northeast-Oregon and Midwest-Oregon flows. Also, as can be seen from the model schedule fits, the resulting schedules are heavily dependent on the available data (e.g., the California to Oregon flow).

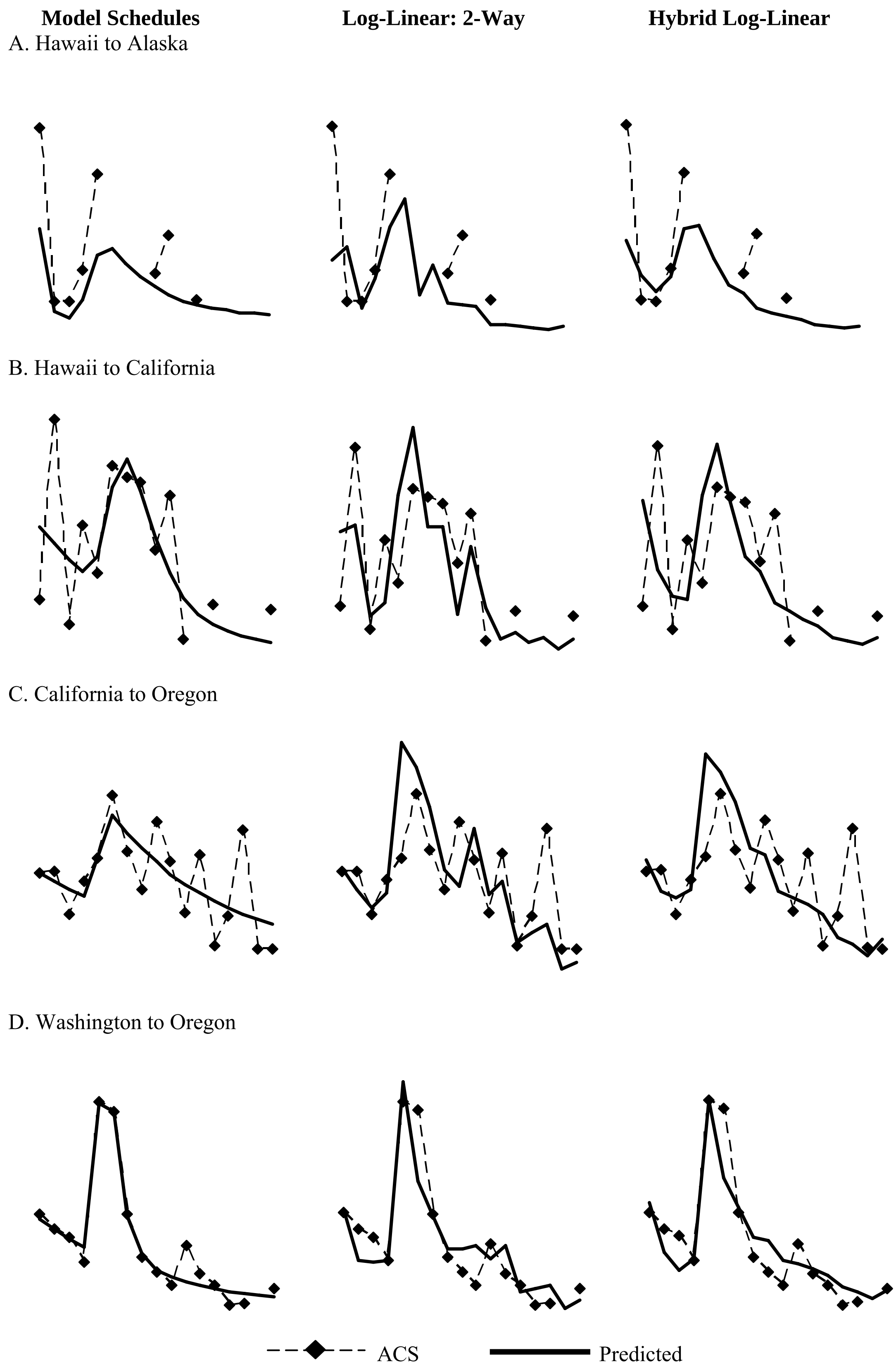


Figure 8. Selected ACS and predicted age-specific migration flows, 2004: Model migration schedules, unsaturated log-linear model, and hybrid log-linear model

SUMMARY AND DISCUSSION

The annual age-specific interstate migration data from the ACS public use files exhibit many irregularities, even when pooled over time. So what does the migration researcher interested in age-specific patterns do in such a situation? They should use the full sample data (if possible), pool the samples over several years, or somehow correct for the irregularities in the patterns before using them. We have shown how model migration schedules, unsaturated log-linear models, or a combination of both can be applied to smooth irregular age patterns in migration. We tested the effectiveness of these approaches by applying them to the 2000 PUMS data and comparing them with the corresponding full sample data. Future work should look into the issues related to pooling the migration data over time and how explanatory factors or auxiliary data (e.g., migration data from the Internal Revenue Service register) can be incorporated to further improve the estimation process.

Some 3 million households received the 2005 ACS questionnaire, giving rise to annual data for about 750 countries with more than 80 percent of the U.S. population represented (Mather et al. 2005). Counties with population less than 65,000 will be represented by three- or five-year averages. The ACS is a valuable source for annual socioeconomic data for states and counties. But analysts studying age- and origin-destination-specific migration flows will be confronted by issues revolving around the relatively small sample sizes associated with high levels of disaggregation and temporal measurement (i.e., the change from a five-year to a one-year migration time interval and "averaging" over time). In other words, with the elimination of the Census long-form questionnaire, migration research in the United States is going to be more difficult. Researchers will need to be careful when using the publicly available ACS data to analyze migration patterns, particularly those disaggregated by origin, destination, age, sex or other characteristics.

Acknowledgements

An earlier version of this paper was presented at the 2007 *Annual Meeting of the Population Association of America*, New York, New York. Support for this research came from the National Institute for Child Health and Human Development. The authors would like to thank Wanran Ma for downloading and organizing the data used in this paper.

REFERENCES

Franklin, R. S., & Plane, D. A. (2006). Pandora's box: The potential and peril of migration data from the American Community Survey. *International Regional Science Review*, 29(3), 231-246.

Gage, L. (2006). Comparison of Census 2000 and American Community Survey 1999-2001 estimates: San Francisco and Tulare counties, California. *Population Research and Policy Review*, 25(3), 243-256.

Gaines, L. M. (2006). Introduction to the special ACS issue of population research and policy review. *Population Research and Policy Review*, 25(3), 197-199.

Griffin, D. H., & Waite, P. J. (2006). American Community Survey overview and the role of external evaluations. *Population Research and Policy Review*, 25(3), 201-223.

Hough Jr., G. C., & Swanson, D. A. (2006). An evaluation of the American Community Survey: Results from the Oregon test site. *Population Research and Policy Review*, 25(3), 257-273.

Koerber, W. K. (2007). *Domestic Migration Flows for States from the 2005 ACS*. Paper presented at the 72nd Annual Meeting of the Population Association of America, New York.

Mather, M., Rivers, K. L., & Jacobsen, L. A. (2005). The American Community Survey. *Population Bulletin*, 60(3).

Raymer, J. (forthcoming). Obtaining an overall picture of population movement in the European Union. In J. Raymer & F. Willekens (Eds.), *International migration in Europe: Data, models and estimates*. Chichester: Wiley.

Raymer, J., Bonaguidi, A., & Valentini, A. (2006). Describing and projecting the age and spatial structures of interregional migration in Italy. *Population, Space and Place*, 12(5), 371-388.

Raymer, J., & Rogers, A. (2007). Using age and spatial flow structures in the indirect estimation of migration streams. *Demography*, 44(2), 199-223.

Rogers, A., & Castro, L. J. (1981). Model migration schedules. RR-81-30, International Institute for Applied Systems Analysis, Laxenburg, Austria.

Rogers, A., Raquillet, R., & Castro, L. J. (1978). Model migration schedules and their applications. *Environment and Planning A*, 10, 475-502.

Rogers, A., Raymer, J., & Newbold, K. B. (2003a). Reconciling and translating migration data collected over time intervals of differing widths. *The Annals of Regional Science*, 37(4), 581-601.

Rogers, A., Willekens, F. J., & Raymer, J. (2002). Capturing the age and spatial structures of migration. *Environment and Planning A*, 34, 341-359.

Rogers, A., Willekens, F. J., & Raymer, J. (2003b). Imposing age and spatial structures on inadequate migration-flow datasets. *The Professional Geographer*, 55(1), 56-69.

Salvo, J. J., & Lobo, A. P. (2006). Moving from a decennial census to a continuous measurement survey: Factors affecting nonresponse at the neighborhood level. *Population Research and Policy Review*, 25(3), 225-241.

Scardamalia, R. (2006). The American Community Survey: General commentary on the findings from external evaluations. *Population Research and Policy Review*, 25(3), 293-303.