Mathematical Population Studies, 9: 231–263, 2001

0889-8480/01 $12.00 + .00

MODELING INTERREGIONAL MIGRATION FLOWS: CONTINUITY AND CHANGE

ANDREI ROGERS[a], FRANS WILLEKENS[b], and JAMES RAYMER[a]

[a]*Population Program, University of Colorado, Boulder, Colorado, USA*
[b]*Population Research Center, University of Groningen, The Netherlands*

This paper addresses the question of how to formally represent the spatial structure of an observed origin-destination-specific pattern of interregional migration flows. Such a representation allows an analyst to compare the spatial structures of different migration regimes and contrast their changes over time. It also facilitates the indirect estimation of migration flows, in the absence of such data, by allowing the analyst to impose a particular age or spatial structure when observed flow data are inadequate, partial, or completely nonexistent. In this paper, we focus on the level and allocation aspects (or the *generation* and *distribution* components) of age-specific interregional migration flows. We find that over time these flows exhibit strong regularities that can be captured by generalized linear models, which can then be used in situations where data are inadequate or missing to indirectly estimate interregional migration patterns.

KEYWORDS: Migration spatial structure; Estimation; Log-linear models; Logit models

INTRODUCTION

The notion of spatial interaction patterns and the spatial structures that arise from them have long fascinated population geographers.

Acknowledgements: This research is being supported by a grant from the National Science Foundation (BCS-9986203).

Address correspondence to Andrei Rogers, Population Program, Institute of Behavioral Science, University of Colorado Boulder, CO 80309-0484, USA. E-mail: andrei.rogers@colorado.edu

Early papers that developed a "social physics" perspective of interaction at a distance led to a host of gravity model formulations (Olsson, 1965). Further work led to reformulations that put forward the ideas of "entropy maximization" (Wilson, 1970) and "information minimization" (Snickars and Weibull, 1977; Plane, 1984). Other generalizations produced general models of movement (Alonso, 1978; Mueser, 1989) and related statistical formulations (Willekens, 1980; Alonso, 1986; Lin, 1999).

Underlying the various spatial interaction models of migration patterns has been a recognition that the decision to move both shapes is shaped by the population geography within which the movement takes place. The spatial distribution of potential destination locations, their attributes, and the interlinkages that connect them shape migration flows in ways which accord them spatial structure that exhibits both continuities and changes over time.

The aim of this paper is (1) to model time series of age-specific and origin-destination specific migration flows, and (2) to identify a few factors that characterize structural stability (continuity) and structural change in both *age profiles* and *spatial interaction patterns*. In general, the problem can be viewed as one of modeling a multidimensional contingency table that includes (but is not restricted to) the dimensions of space, time and age. Age patterns and spatial patterns change gradually, not abruptly or at random. The goal is to identify the stable patterns, and to capture the structural or systematic changes by constructing parsimonious models of spatial interaction.

Our identification of stable age and spatial patterns and the few factors that capture structural changes in them should benefit three fields of research: (1) the study of changes in migration flows, (2) the indirect estimation of migration flows (since the identified stable patterns may facilitate the indirect estimation of migration when data are incomplete) and (3) the forecasting of migration.

This article builds on work by Willekens and Baydar (1986) on migration forecasting using generalized linear models, by Rogers and Wilson (1996) and Lin (1999) on modelling of structural change in migration patterns, by Rogers and Castro (1981) on parameterized models of age-specific migration, and by Willekens *et al.* (1981) and Willekens (1982, 1983) on including a priori information with offsets in a log-linear modeling framework. The article continues research on identifying migration spatial structure (Rogers *et al.*, 2002) and on the indirect estimation of migration (Rogers *et al.*, 2001). What separates this paper from the work above is the linkage of known structures, or "standard migration structures," with generalized linear models in order to infer migration data sets using indirect methods of estimation. The emphasis on the formal modeling of the generation and distribution components of migration spatial structures also leads to new directions and insights that contribute to the development of age- and origin-destination-specific migration data.

MODELS

In studies of migration, two types of data are often distinguished, movement data and transition data. Movement data represent the number of migrations (events) during an interval of a given length (e.g., one or five years). Transition data represent the number of persons by place of residence at two points in time (e.g., at time of the census and five years prior to the census). To highlight the difference, a distinction is often made between *migrations* (events) and *migrants* (persons). The data modeled in this paper are transition data. They are the interregional (4-region) migration streams reported in the 1960, 1970, 1980 and 1990 U.S. censuses, disaggregated by age (5-year age groups).

In order to specify the probability model that underlies an observed migration pattern or structure, it is assumed that the probability distribution underlying the migration transition data is known. Specifically, that the number of transitions recorded during a given period follows a Poisson distribution or a multinomial (or binomial) distribution, depending on whether the total number of people recorded is assumed to be free or fixed. The multinomial model may be derived from the Poisson model by conditioning on the sample size (Azzalini, 1996, p. 251). The parameters of the two models are related: the parameter of the Poisson model is the expected number of transitions during a unit interval (of, say, one or five years); the parameter of the multinomial (or binomial) model is the proportion experiencing a transition. That observation is used in this paper to (1) demonstrate the relation between different models of migration and (2) to accommodate various data types, drawn from different sources, into a single modeling framework.

The regression-type model that is associated with the Poisson model is the *log-linear model* or *Poisson regression model*; the model associated with the multinomial (or binomial) model is the *logit model* or the related *logistic regression model*. The relation between the logit model and the log-linear model has been known for some time

(Haberman, 1978, 1979). The relation between the logistic regression model and the log-linear model was established later.

To illustrate our approach, let $N_{ij}(x)$ denote the random variable measuring the number of persons aged x to $x + h$ (h is the length of age interval) who reside in region i at the first measurement (beginning of interval) and in j at the second measurement (end of interval), and let $n_{ij}(x)$ be a realization of $N_{ij}(x)$, i.e., the number of migrants observed. In this article, a migrant is defined as a person who lives in one of the four regions five years prior to the census and in another region at the time of the census. Persons who reside in the same region at the two measurements are referred to as stayers. They include persons who did not change their residence, persons who changed residence within the region and persons who moved out of the region but returned before the end of the interval. Migrants are persons with different residences at the two measurements.

The observed number of migrants, $n_{ij}(x)$, may be viewed as determined by three components: (1) the number of persons of age x residing in region i at the beginning of the interval [$n_{i+}(x)$, where + denotes summation over the subscript] who survive to the end of the interval (and therefore are included in the second measurement), (2) the proportion of that number who are migrants, i.e., persons aged x in i at the beginning of the interval who reside outside of i at the end of the interval: $\bar{S}_i(x) = \sum_j n_{ij}(x)/n_{i+}(x)$ where $j \neq i$, and (3) the destination proportion of migrants, i.e., the conditional probability that a migrant leaving i at start of interval, resides in j at the end of that interval: $\bar{S}_{j|i}(x) = n_{ij}(x)/\sum_j n_{ij}(x)$ with $\bar{S}_{j|i}(x) \geq 0$, $\sum_j \bar{S}_{j|i}(x) = 1$, and $j \neq i$. The second component in the above triple is often referred to as the *generation component*. The third component is known as the *distribution component*. The expected number of persons of age x residing in i at the beginning of the interval and in j at the end of the interval may be expressed as the product of the expected number of persons aged x residing in i at the beginning of the interval, the generation component, and the distribution component, i.e., $E[N_{ij}(x)] = E[N_{i+}(x)]\bar{S}_i(x)\bar{S}_{j|i}(x)$. In the demographic literature, the product $\bar{S}_i(x)\bar{S}_{j|i}(x)$ is known as the *multiregional conditional survivorship proportion*, $\bar{S}_{ij}(x)$ (Rogers, 1995).

In this article, log-linear and logit models are applied to study the level of out-migration. That level is expressed in terms of number of migrants or proportion of migrants. The log-linear model predicts the number of out-migrants and the number of stayers, i.e., the number of persons by migrant status. The logit model predicts the proportion of out-migrants.

Models for Describing Migration Patterns

Because saturated log-linear and logit models perfectly account for (i.e., "predict") the observed data, they are useful for describing the different "effects" shaping migration flow tables. For example, consider a two-way, origin-by-destination, migration flow table. The *multiplicative log-linear model* to describe that table is specified as:

$$\hat{n}_{ij} = \tau\tau_i^O\tau_j^D\tau_{ij}^{OD}, \tag{1}$$

where $\hat{n}_{ij}$ denotes the predicted number of migrants or stayers and the τs denote the parameters in the model, consisting of an overall effect (τ), an origin main effect (τ_i^O), a destination main effect (τ_j^D) and an origin-destination interaction effect (τ_{ij}^{OD}). Note that the models in this article are expressed in their *multiplicative* forms, found by exponentiating the linear additive model form.

To describe the multiregional conditional survivorship proportions ($\bar{S}_{ij}$) of the above migration flow table, a saturated multinomial logit model can be applied. In the multinomial logit model, a dependent variable is specified and the predicted values are equal to odds, which then can be converted into probabilities. The *multinomial logit model* is specified as:

$$\hat{\theta}_{ij} = \frac{\hat{S}_{ij}}{\hat{S}_{ik}} = \nu_j\nu_{ij}^O, \tag{2}$$

where $\hat{\theta}_{ij}$ is the predicted odds of migrating from origin i to destination j relative to migrating from origin i to destination k, where k equals the reference category. The parameters denoted by ν represent the multiplicative logit parameters. Note that in the multinomial logit model, for each destination j there exist separate parameters that control for the other destinations in the model. In the case of a binary variable, the proportion in the first category is $\theta/(1+\theta)$, where θ is the odds. If the variable is polytomous with N categories, the proportion in category i is equal to $\theta_i/(\theta_1+\theta_2+\ldots+\theta_N)$, where θ_i is the odds of being in category i rather than in the reference category (proportion in i over proportion in the reference category).

Instead of focusing on the complete migration flow table discussed above, the models in this article represent separately the *generation* and *distribution* components of that table. The *generation component* represents the proportion of persons leaving a region for another region. To distinguish between movers and stayers, one may introduce an additional subscript to denote migrant status. Let m denote migrant status, with $m = 1$ denoting migrants and $m = 2$ denoting stayers. Hence n_{mi} denotes the number of persons living in region i at the beginning of the interval by migrant status m. If the level of migration is expressed as the proportion of migrants, the logit model is the appropriate model. It predicts the odds of being a migrant rather than a stayer. The *odds* that a resident of region i is a migrant rather than a nonmigrant is the ratio of the number of migrants to the number of nonmigrants: n_{1i}/n_{2i}. The corresponding saturated logit model then is:

$$\hat{\theta}_i = \frac{\hat{n}_{1i}}{\hat{n}_{2i}} = \nu\nu_i^O, \tag{3}$$

where ν is the appropriate multiplicative logit parameter.

Finally, the multinomial model for the *distribution component* is specified as:

$$\hat{\theta}_{j|i} = \frac{\hat{S}_{j|i}}{\hat{S}_{k|i}} = \nu_{j|i}, \tag{4}$$

where the parameter $\nu_{j|i}$ denotes the odds of choosing destination region j in reference to destination region k, *given the origin region i.* The response variable for this model is the destination.

Relational Models of Migration

Relational models typically make use of a "standard" and a mathematical rule for relating that standard to a particular data set in order to predict the values of a particular variable. Much of the complexity in the variable is captured through the standard, and the model parameters represent deviations from that standard. For example, the Brass (1974) model for mortality and the Coale-Trussell (1974) model for fertility are considered to be relational models. The log-linear model *with offset* also can be viewed as a *relational* model, with the offset being any set of specified values that need to be imposed on a particular data setting. The method of offsets rescales information from one setting to add up to the marginal totals of another.

The log-linear model with offset is a flexible platform for modeling migration flows. Generally, the offset is applied to a *main effects* log-linear model, specified as:

$$\hat{n}_{ij} = \tau\tau_i^O\tau_j^D. \tag{5}$$

The main effects log-linear model predicts migration flows based on the marginal totals of the migration flow table. Because of the generally strong diagonal effects in migration tables (which in saturated models are captured by the origin-destination interaction effects), the main effects model generally does not predict the observed migration flows accurately. The sum of the predicted values, however, does equal the observed marginal totals.

When an offset is included in a main effects model, a spatial structure is imposed on the data. The offset can consist of any specified set of values that need to be imposed on a particular table. Such a model, specified as

$$\hat{n}_{ij} = n^*_{ij}\xi\xi^O_i\xi^D_j, \tag{6}$$

borrows the interaction term from the offset (n^*_{ij}). Note that the * denotes the offset and the ξs denote the parameters of the log-linear model with offset, which predicts changes expressed in the form of odds between the predicted values (above) and the values of the offset:

$$\hat{\theta}_{ij} = \frac{\hat{n}_{ij}}{n^*_{ij}} = \frac{\tau\tau^O_i\tau^D_j\tau^{OD*}_{ij}}{\tau^*\tau^{O*}_i\tau^{D*}_j\tau^{OD*}_{ij}} = \xi\xi^O_i\xi^D_j. \tag{7}$$

The ξ-parameters represent the ratios of the predicted saturated log-linear model parameters to the corresponding observed saturated log-linear model parameters. For example, the ξ-parameter for the origin effect (ξ^O_i) is equal to the corresponding τ-parameter for the predicted migration flows (τ^O_i) divided by the τ-parameter associated with the historical migration flow table (τ^{O*}_i). Notice that in this model, the interaction parameters for the predicted and historical migration flow tables are the same in the numerator and denominator, thus they drop out. The main advantage of the offset approach to the indirect estimation of migration is that the parameters of the model indicate how they collectively determine each predicted flow value. Relational models are used later in this article to specify age-specific generation and distribution components.

DESCRIBING THE STRUCTURE OF MIGRATION

We identify and model the structure of regional out-migration in this section and the next section by means of two classes of generalized linear models: log-linear and logit models. For our examples, we consider four categorical variables: region of origin, region of destination, time period, and age group. We begin by first modeling the origin and time effects of the generation and distribution components of migration, without considering age. After the underlying structures in the origin-time models are identified, we then incorporate age effects in the overall models.

The Generation Component

The data in Table 1 describe the number of region-specific (Northeast, Midwest, South and West) out-migrants and stayers during the 1955–1960, 1965–1970, 1975–1980 and 1985–1990 periods. The out-migration data represent five-year transition data that were developed from information on current and previous (five years prior to the census) places of residence obtained at the time of the census.

To examine the structure of the origin and time dependent data in Table 1, we use saturated log-linear and logit models. The saturated time-dependent log-linear model may be specified as:

$$\hat{n}_{mit} = \tau\tau_m^M \tau_i^O \tau_t^T \tau_{mi}^{MO} \tau_{mt}^{MT} \tau_{it}^{OT} \tau_{mit}^{MOT}, \tag{8}$$

TABLE 1
Observed number of U.S. persons by migrant status: 1955–1960 to 1985–1990

Period	Region of Origin	Migrant	Stayer	Total	Proportion Migrants
1955–1960	Northeast	1,683,195	37,731,535	39,414,730	0.0427
	Midwest	2,544,231	42,858,625	45,402,856	0.0560
	South	2,434,713	44,861,691	47,296,404	0.0515
	West	1,062,128	21,351,360	22,413,488	0.0474
	Total	7,724,267	146,803,211	154,527,478	0.0500
1965–1970	Northeast	2,142,805	42,593,965	44,736,770	0.0479
	Midwest	2,780,878	49,124,857	51,905,735	0.0536
	South	2,703,705	52,729,059	55,432,764	0.0488
	West	1,781,277	28,350,333	30,131,610	0.0591
	Total	9,408,665	172,798,214	182,206,879	0.0516
1975–1980	Northeast	3,014,910	43,123,461	46,138,371	0.0653
	Midwest	3,464,705	51,136,449	54,601,154	0.0635
	South	2,917,481	67,094,529	70,012,010	0.0417
	West	2,084,055	37,901,623	39,985,678	0.0521
	Total	11,481,152	199,256,062	210,737,214	0.0545
1985–1990	Northeast	2,720,077	44,379,438	47,099,515	0.0578
	Midwest	3,168,926	52,301,415	55,470,341	0.0571
	South	3,353,748	72,887,476	76,241,224	0.0440
	West	2,272,415	43,732,986	46,005,401	0.0494
	Total	11,515,166	213,301,315	224,816,481	0.0512

where $\hat{n}_{mit}$ represents the predicted *number* of persons who exhibit migrant status m, leaving region of origin i during time period t. The associated logit model, which predicts the odds of being a migrant versus being a stayer, is specified as:

$$\hat{\theta}_{it} = \frac{\hat{S}_{it}}{1 - \hat{S}_{it}} = \nu\nu_i^O \nu_t^T \nu_{it}^{OT}. \tag{9}$$

The parameter values for the two models above (Equations 8 and 9) are set out in Table 2. Both models utilize the last category of each variable as the reference category; in other words, they use the "dummy" or "cornered effect" coding system (Wrigley, 1985, pp. 132–136). Throughout this article, when we discuss log-linear or logit models, the parameter values will be defined with respect to the last category of the variables in the model. Other common coding systems utilize the geometric mean (effect coding) or the first category (dummy coding) as the reference category. Since there are three variables in the log-linear model specified in Equation (7), there exist the three reference categories associated with the variables of migrant status, region of origin and time period, i.e., "Stayer," "West," and "1985–1990," respectively. For the logit model specified in Equation (9), there exist two reference categories corresponding to the independent variables of region of origin (West) and time period (1985–1990), plus the reference category of stayer implicit in the dependent variable, i.e., migrant status.

The parameters of the log-linear and logit models that contain the same qualitative variables are in this illustration fundamentally equivalent. The close relationships between them are illustrated in Table 2, in which the overall effect (ν), main effects (ν_i^O and ν_t^T) and interaction effects (ν_{it}^{OT}) parameters of the logit model correspond to the migrant status main effect (τ_m^M), the two-way interaction effects with migrant status (τ_{mi}^{MO} and τ_{mt}^{MT}) and the three-way interaction effect parameters (τ_{mit}^{MOT}), respectively, of the log-linear model. There are fewer parameters in the logit model because migrant status is specified as the dependent variable (with migrant = 1 and stayer = 2).

The parameters of the log-linear and logit models in their multiplicative forms may be interpreted as odds (main effects) and odds ratios (interaction effects) with regard to the reference categories (see Alba (1988) for a complete discussion). Such interpretations of the parameters are fairly straightforward, as long as one keeps in mind the reference categories. For example, consider the saturated log-linear and logistic model parameters in Table 2. When the models are

TABLE 2
Modeling migrant status, origin, and time: comparison of saturated log-linear and logit model parameters

Log-Linear		Logit	
τ	43,732,986	ν	0.0520
τ_1^M	0.0520	ν_1^O	1.1796
τ_1^O	1.0148	ν_2^O	1.1661
τ_2^O	1.1959	ν_3^O	0.8855
τ_3^O	1.6666	ν_1^T	0.9574
τ_1^T	0.4882	ν_2^T	1.2092
τ_2^T	0.6483	ν_3^T	1.0582
τ_3^T	0.8667	ν_{11}^{OT}	0.7603
τ_{11}^{MO}	1.1796	ν_{21}^{OT}	1.0234
τ_{12}^{MO}	1.1661	ν_{31}^{OT}	1.2320
τ_{13}^{MO}	0.8855	ν_{12}^{OT}	0.6788
τ_{11}^{MT}	0.9574	ν_{22}^{OT}	0.7727
τ_{12}^{MT}	1.2092	ν_{32}^{OT}	0.9216
τ_{13}^{MT}	1.0582	ν_{13}^{OT}	1.0779
τ_{11}^{OT}	1.7414	ν_{23}^{OT}	1.0567
τ_{21}^{OT}	1.6784	ν_{33}^{OT}	0.8930
τ_{31}^{OT}	1.2607		
τ_{12}^{OT}	1.4805		
τ_{22}^{OT}	1.4489		
τ_{32}^{OT}	1.1160		
τ_{13}^{OT}	1.1212		
τ_{23}^{OT}	1.1282		
τ_{33}^{OT}	1.0621		
τ_{111}^{MOT}	0.7603		
τ_{121}^{MOT}	1.0234		
τ_{131}^{MOT}	1.2320		
τ_{112}^{MOT}	0.6788		
τ_{122}^{MOT}	0.7727		
τ_{132}^{MOT}	0.9216		
τ_{113}^{MOT}	1.0779		
τ_{123}^{MOT}	1.0567		
τ_{133}^{MOT}	0.8930		

Note: (A) M = Migrant Status (Migrant = 1 and Stayer = 2), O = Region of Origin (Northeast = 1, Midwest = 2, South = 3, and West = 4), and T = Time Period (1955–1960 = 1, 1965–1970 = 2, 1975–1980 = 3, and 1985–1990 = 4). (B) Reference category equals last category of each variable.

saturated, one can relate the parameters directly to the observed data in Table 1. The overall effect parameter for the saturated log-linear model is 43,732,986 and corresponds to the stayers in the West during the 1985–1990 period (n_{244}). The overall effect parameter for the saturated logit model is 0.0520 and corresponds to the odds of being a migrant versus being a stayer from the West during the 1985–1990 period (note that $\nu = \tau_1^M = 2{,}272{,}415/43{,}732{,}986 = 0.0520$).

The main effect parameters for *migrant status* (τ_m^M) are interpreted as the *odds* of being a migrant (or stayer) to being a stayer in the West during the 1985–1990 time period. The parameter value is $\tau_1^M = 0.0520$, which means that the odds of being a migrant to being a stayer in the West during the 1985–1990 period are roughly 52 to 1000. The main effect parameters for *region of origin* (τ_i^O) are interpreted as the odds of being a stayer in the Northeast, Midwest, or South during the 1985–1990 period to being a stayer in the West during the 1985–1990 period. The parameter values are $\tau_1^O = 1.0148$, $\tau_2^O = 1.1959$, and $\tau_3^O = 1.6666$. The fact that all three $\tau_i^O > 1$, simply means that there were more stayers in the Northeast, Midwest and South than in the West during the 1985–1990 period. The main effect parameters for *time period* (τ_t^T) are interpreted as the odds of being a stayer in the West during the 1955–1960, 1965–1970, and 1975–1980 and periods versus being a stayer in the West during the 1985–1990 period. The parameter values are $\tau_1^T = 0.4882$, $\tau_2^T = 0.6483$, and $\tau_3^T = 0.8667$, which means there were roughly half as many stayers in the West during the 1955–1960 period, two-thirds as many in the 1965–1970 period, and nine-tenths as many in the 1975–1980 period compared with the stayers in the West during the 1985–1990 period.

The two-way interaction parameters for, say, migrant status and region of origin (τ_{mi}^{MO}) are interpreted as *odds ratios*. The parameters equal the ratio of (1) the odds of being a migrant from region i to the odds of being a migrant from the West during the 1985–1990 period to (2) the odds of being a stayer in region i to being a stayer in the West during the 1985–1990 period. For example, the migrant status-region effect (τ_{mi}^{MO}) for the Northeast ($\tau_{11}^{MO} = 1.1796$) implies that, during the 1985–1990 period, a person was 1.18 times as likely to be an out-migrant from the Northeast than to be an out-migrant from the West.

The three-way interaction parameters in this model (τ_{mit}^{MOT}) are *ratios of two odds ratios*. These parameters are interpreted as the odds of (1) being a migrant from region i during time t relative to being a migrant from the West during time t to (2) being a migrant from region i during 1985–1990 relative to being a migrant from the West during 1985–1990. For example, the odds of being a migrant from the Northeast relative to the West during the 1955–1960 period (i.e., $(1{,}683{,}195/37{,}731{,}535)/(1{,}062{,}128/21{,}351{,}360) = 0.0446/0.0497 = 0.8968$) were less than the odds of being a migrant from the Northeast relative to the West during the 1985–1990 period (i.e., 0.0613/0.0520 = 1.1796). Therefore the parameter $\tau_{111}^{MOT} = 0.8968/1.1796 = 0.7603$.

The Distribution Component

The destination of out-migrants may depend on three variables: region of origin, time period and age group. Table 3A shows the number of migrants by origin, destination and time period. Intra-regional migration is omitted. The row totals are equal to the number of out-migrants in Table 1 of the generation component. The corresponding totals for the distribution component (proportions) are shown in Table 3B. The data indicate, for example, that over the four census intervals examined 59% of the persons who left the Northeast region went to the South, 24% to the West, and 17% to the Midwest. The South has been the most attractive region, and not only for persons from the Northeast. Half of the migrants from the Midwest went to the South, 37% to the West, and 12% to the Northeast. The South also has been the most attractive region for residents of the West. More than half of the migrants from that region moved to the South (51%), 33% to the Midwest and 16% to the Northeast.

To examine the spatial structure of migration destinations, a saturated multinomial logit model may be specified to reproduce all of the period-specific distribution proportions in Table 3B:

$$\hat{\theta}_{jt|i} = \frac{\hat{S}_{jt|i}}{\hat{S}_{kt|i}} = \nu_{j|i}\nu_{jt|i}^{T}, \tag{10}$$

where $\nu_{j|i}$ is the intercept for destination j and $\nu_{jt|i}^{T}$ is the period effect for destination j. Both the intercept and the period effect are for a given origin i. The reason each region of origin is modeled separately is because a simultaneous estimation for all regions of origin is not possible due to the missing diagonal elements (representing intra-regional migration) in the origin-destination matrix of destination proportions.

The regression coefficients for this model are shown in Table 4 for the multiplicative model. Since destination and time both are categorical variables, reference categories must be defined for each. The reference categories in this illustration are the West region, when the regions of origin are the Northeast, Midwest and South, and the South when the region of origin is the West. The 1985–1990 time period is the reference category for all regions of origin.

TABLE 3

Observed numbers and proportions of U.S. interregional migrants by origin and destination: 1955–1960 to 1985–1990

Origin Region	Period	A. Numbers					B. Proportions				
		Destination Region					Destination Region				
		Northeast	Midwest	South	West	Total	Northeast	Midwest	South	West	Total
Northeast	1955–1960	—	353,818	882,783	446,594	1,683,195	—	0.2102	0.5245	0.2653	1.0000
	1965–1970	—	481,391	1,138,625	522,789	2,142,805	—	0.2247	0.5314	0.2440	1.0000
	1975–1980	—	461,932	1,799,666	753,312	3,014,910	—	0.1532	0.5969	0.2499	1.0000
	1985–1990	—	357,318	1,821,896	540,863	2,720,077	—	0.1314	0.6698	0.1988	1.0000
	Total	—	1,654,459	5,642,971	2,263,559	9,560,989	—	0.1730	0.5902	0.2367	1.0000
Midwest	1955–1960	314,030	—	1,088,395	1,141,806	2,544,231	0.1234	—	0.4278	0.4488	1.0000
	1965–1970	413,172	—	1,328,880	1,038,826	2,780,878	0.1486	—	0.4779	0.3736	1.0000
	1975–1980	350,373	—	1,845,327	1,269,005	3,464,705	0.1011	—	0.5326	0.3663	1.0000
	1985–1990	378,395	—	1,765,905	1,024,626	3,168,926	0.1194	—	0.5573	0.3233	1.0000
	Total	1,455,968	—	6,028,507	4,474,264	11,958,739	0.1217	—	0.5041	0.3741	1.0000
South	1955–1960	568,877	966,659	—	899,177	2,434,713	0.2337	0.3970	—	0.3693	1.0000
	1965–1970	671,361	1,093,461	—	938,883	2,703,705	0.2483	0.4044	—	0.3473	1.0000
	1975–1980	694,648	1,081,859	—	1,140,974	2,917,481	0.2381	0.3708	—	0.3911	1.0000
	1985–1990	848,986	1,241,907	—	1,262,855	3,353,748	0.2531	0.3703	—	0.3766	1.0000
	Total	2,783,872	4,383,886	—	4,241,889	11,409,647	0.2440	0.3842	—	0.3718	1.0000
West	1955–1960	161,197	381,808	519,123	—	1,062,128	0.1518	0.3595	0.4888	—	1.0000
	1965–1970	292,979	616,363	871,935	—	1,781,277	0.1645	0.3460	0.4895	—	1.0000
	1975–1980	287,366	676,870	1,119,819	—	2,084,055	0.1379	0.3248	0.5373	—	1.0000
	1985–1990	389,396	704,566	1,178,453	—	2,272,415	0.1714	0.3101	0.5186	—	1.0000
	Total	1,130,937	2,379,606	3,689,329	—	7,199,872	0.1571	0.3305	0.5124	—	1.0000

TABLE 4
Modeling destination by region of origin and time period: parameters of the multinomial logit model

Region of Origin	Parameter	
Northeast*	$\nu_{2\|1}$	0.6606
	$\nu^T_{21\|1}$	1.1992
	$\nu^T_{22\|1}$	1.3938
	$\nu^T_{23\|1}$	0.9282
	$\nu^T_{3\|1}$	3.3685
	$\nu^T_{31\|1}$	0.5868
	$\nu^T_{32\|1}$	0.6466
	$\nu^T_{33\|1}$	0.7092
Midwest*	$\nu_{1\|2}$	0.3693
	$\nu^T_{11\|2}$	0.7447
	$\nu^T_{12\|2}$	1.0770
	$\nu^T_{13\|2}$	0.7476
	$\nu_{3\|2}$	1.7235
	$\nu^T_{31\|2}$	0.5531
	$\nu^T_{32\|2}$	0.7422
	$\nu^T_{33\|2}$	0.8437
South*	$\nu_{1\|3}$	0.6723
	$\nu^T_{11\|3}$	0.9411
	$\nu^T_{12\|3}$	1.0636
	$\nu^T_{13\|3}$	0.9056
	$\nu_{2\|3}$	0.9834
	$\nu^T_{21\|3}$	1.0932
	$\nu^T_{22\|3}$	1.1843
	$\nu^T_{23\|3}$	0.9642
West**	$\nu_{1\|4}$	0.3304
	$\nu^T_{11\|4}$	0.9397
	$\nu^T_{12\|4}$	1.0169
	$\nu^T_{13\|4}$	0.7766
	$\nu_{2\|4}$	0.5979
	$\nu^T_{21\|4}$	1.2302
	$\nu^T_{22\|4}$	1.1823
	$\nu^T_{23\|4}$	1.0110

Note: (A) The 1st subscript in $\nu^T_{jt|i}$ refers to region of destination (Northeast = 1, Midwest = 2, South = 3, and West = 4). (B) The 2nd subscript in $\nu^T_{jt|i}$ refers to time period (1955–1960 = 1, 1965–1970 = 2, 1975–1980 = 3, and 1985–1990 = 4). (C) The 3rd subscript in $\nu^T_{jt|i}$ refers to region of origin (Northeast = 1, Midwest = 2, South = 3, and West = 4). (D) Reference category for the variable of time T = 1985–1990. (E) * Reference category = West. (F) ** Reference category = South.

The parameters are interpreted as odds and odds ratios. For example, consider the migration from the Northeast to the Midwest. The intercept is 0.6606, which is the odds that a migrant who leaves the Northeast during the reference period (1985–1990) selects the Midwest as the destination rather than the West (which is the reference category). Since the model is a saturated model, the same number can be obtained, using Table 3, by dividing the *observed* number (or proportion) of out-migrants from the Northeast selecting the Midwest as a destination (i.e., n_{124}) over the number (or proportion) selecting the West as the destination (i.e., n_{144}): $357{,}318/540{,}863 = 0.6606$ (or $0.1314/0.1988 = 0.6606$). The intercept is the odds of destination choice in the reference period. It is a measure of the *preference* for the Midwest rather than the West in the reference period. High odds indicate a high preference.

Consider, for example, the out-migration from the Northeast to the Midwest in 1965–1970. The parameter of the logit model is 1.3938. It is greater than one, hence the odds of migrants going to the Midwest relative to the West is larger in 1965–1970 ($481{,}391/522{,}789 = 0.9208$) than in the reference period 1985–1990 ($357{,}318/540{,}863 = 0.6606$). The odds ratio is a measure of *the change in preference* for the Midwest as a destination compared with the West. Odds ratios significantly different from one indicate significant change. Odds ratios equal to one or close to one indicate relative stability.

This discussion illustrates the significance of the selection of reference categories and, consequently, of the coding applied in statistical modeling involving categorical variables. The selection of reference categories does not receive much attention in the literature, and most packages are not flexible when it comes to their selection. For instance, SPSS selects, by default, the last category of a categorical variable as the reference category (in this case, the region West and the period 1985–1990). As a consequence, the coefficients of the regression models express characteristics of particular migration flows relative to the characteristics of the migration to the West in 1985–1990. In many studies, the interpretation of the regression coefficients may benefit from a more appropriate choice of reference categories. In general, the selection of reference category should be determined by the research question rather than by the software. For instance, the investigation of changes in the distribution component over time may be enhanced if the characteristics of the distribution component are expressed relative to the characteristics in the base period (1955–1960).

DESCRIBING THE STRUCTURE OF AGE-SPECIFIC MIGRATION

The Generation Component

Regional *age patterns* of migration have varied substantially in the United States since 1960. The age structure of a resident or migrant population is normally expressed by a set of age-specific proportions that specify how that population is distributed across a full range of ages (or age groups). The sum of all such proportions is unity. Age profiles of residents or of migrants also can be expressed in the form of proportions across a full range of states (e.g., out-migrants or stayers), in which each age-specific proportion is the ratio of the number of out-migrants who *survive* to the end of the unit time interval to the number who were residents of region i at the start of that same time interval (i.e., by $\bar{S}_i(x)$). This is formally referred to in each case as the *age-specific generation component.* Such age-specific generation components are shown in Figure 1. These proportions or propensities tend to exhibit considerable stability over time.

The age patterns of migration proportions and their variations across different origin and destination pairs can be modeled by generalized linear models, provided that at least one parameter is associated with each age group. This necessarily makes the model more cumbersome, particularly when interactions are considered, such as origin-age or period-age effects.

One of the advantages of applying generalized linear models is the decompositional information one receives when the effects of the underlying hierarchical structures are examined. We begin with an analysis of the age-specific saturated logit model parameters. The saturated logit model used for the analysis of this sub-section is specified as:

$$\hat{\theta}_{it}(x) = \nu\nu_i^O \nu_t^T \nu^A(x)\nu_{it}^{OT}\nu_i^{OA}(x)\nu_t^{TA}(x)\nu_{it}^{OTA}(x). \tag{11}$$

This model is the same as the model specified in Equation (9), but with the effect of age added. The saturated model distinguishes four hierarchical effects with regard to age: (1) the main effects, (2) the region effects, (3) the period effects, and (4) the region-period effects. The age profiles associated with the age-specific main, region, and period effects are set out in Figure 2.

The age main effects profile (Figure 2A) corresponds to the age-specific conditional survivorship proportions of out-migration from

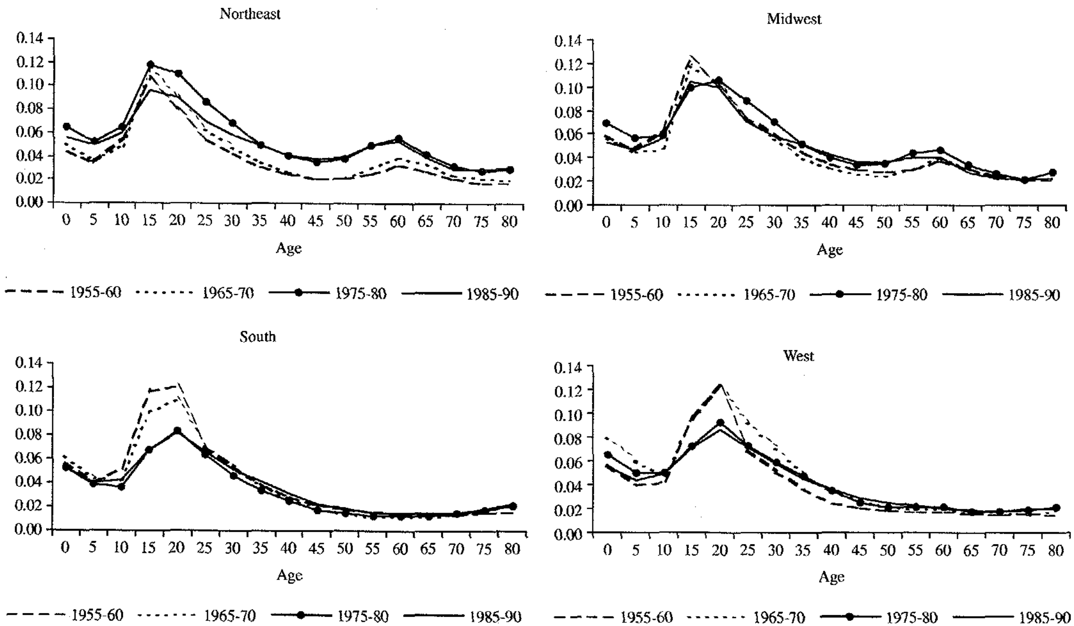

FIGURE 1 Observed U.S. regional conditional survivorship proportions of out-migration flows by age: 1955–1960 to 1985–1990.

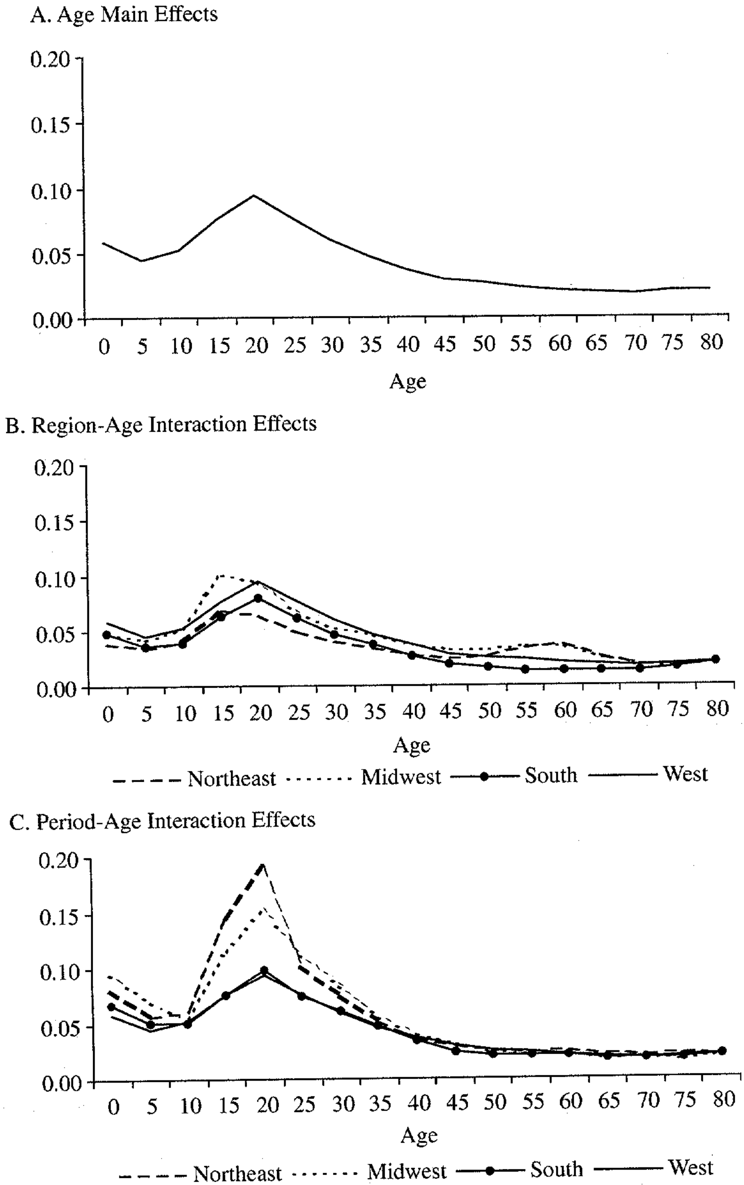

FIGURE 2 Age-specific conditional survivorship proportions of out-migration: The main and first-order interaction effects of a saturated logit model.

the West during the 1985–1990 period, and also represents the base for all the age profiles shown in Figures 2B and 2C. Subsequently, the age profiles corresponding to the West region (Figure 2B) or the 1985–1990 period (Figure 2C) are exactly the same as the age main effects profile set out in Figure 2A. The other age profiles set out in Figures 2B and 2C representing region-age and period-age effects, respectively, deviate from that single profile. For example, to calculate the *combined region-age* effect for, say, the Northeast region, one needs to multiply together the intercept parameter (ν), the 17 age main effect parameters ($\nu^A(x)$), the single region main effect parameter for the Northeast (ν_1^O), and the 17 Northeast-age interaction effect parameters ($\nu_1^{OA}(x)$). Likewise to calculate the *combined period-age* effect for, say, the 1955–1960 period (Figure 2C), one can multiply together the intercept parameter (ν), the 17 age parameters ($\nu^A(x)$), the single period main effect parameter for the 1955–1960 period (ν_1^T), and the 17 period-age interaction effect parameters associated with the 1955–1960 period ($\nu_1^{TA}(x)$).

When the period effects are interacted with age (Figure 2C), one notices that the principal differences in the age profiles that result are that the proportions of out-migrants in the young adult ages (ages 15–29) have declined over time (along with the corresponding early ages of childhood [ages 0–9]). The retirement peak only becomes noticeable when the region and age effects are interacted (Figure 2B). So, from this information, we can conclude that the period-age effects are more sensitive to changes in levels of out-migration, whereas the region-age effects are more sensitive to changes in the shapes of the out-migration schedules.

The graphs in Figure 2 represent age-specific effects within a saturated logit model. Next, we examine the predicted values from the complete list of *unsaturated* models corresponding to the saturated model discussed above. This will serve two purposes: (1) it will help us to find the simplest model that adequately predicts out-migration proportions, and (2) it will allow us to obtain a better indication of how each separate main and interaction effect contributes to the overall fit of the model.

The logit model that perfectly predicts age-specific out-migration proportions by region (Northeast, Midwest, South, and West) and period (1955–1960 to 1985–1990) can be compared to its many *unsaturated* logit model specifications of lower order. In total, there exist fifteen possible permutations of our hierarchical logit framework that includes region, period, and age (set out in Table 5) with the number of parameters ranging from a maximum of 272 (4 regions $\times$ 4 periods $\times$

TABLE 5
Comparisons of various unsaturated logit models: generation component

Fitted Marginals				R^2	X^2	G^2	# of Indep. Parameters	Re-scaled X^2	Re-scaled G^2
	Interactions		Model	(1)	(2)	(3)	(4)	(2)*(4)	(3)*(4)
Main	1st Order	2nd Order							
1. [O]			Region	0.028	3.982	9.365	4	15.928	37.460
2. [T]			Period	0.003	4.045	9.220	4	16.181	36.880
3. [A]			Age	0.870	0.594	0.418	17	10.100	7.108
4. [O][T]			Region and period main effects	0.032	3.990	9.212	7	27.932	64.485
5. [O][A]			Region and age main effects	0.889	0.459	0.440	20	9.187	8.799
6. [T][A]			Period and age main effects	0.872	0.588	0.416	20	11.765	8.310
7. [O][T][A]			Main effects	0.891	0.452	0.411	23	10.388	9.447
8. [O][T][A]	[OT]		Main effects with region-period interaction	0.921	0.343	0.409	32	10.973	13.097
9. [O][T][A]	[OA]		Main effects with region-age interaction	0.934	0.227	0.126	71	16.121	8.979
10. [O][T][A]	[TA]		Main effects with period-age interaction	0.916	0.347	0.349	71	24.621	24.798
11. [O][T][A]	[OT][OA]		Main effects with region-period and region-age interactions	0.965	0.120	0.090	80	9.621	7.211
12. [O][T][A]	[OT][TA]		Main effects with region-period and period-age interactions	0.947	0.246	0.341	80	19.714	27.285
13. [O][T][A]	[OA][TA]		Main effects with region-age and period-age interactions	0.961	0.127	0.048	119	15.099	5.678
14. [O][T][A]	[OT][OA][TA]		First-order interaction	0.992	0.029	0.029	128	3.696	3.649
15. [O][T][A]	[OT][OA][TA]	[OTA]	Saturated	1.000	0.000	0.000	272	0.000	0.000

Note: (A) O = Region of Origin, T = Period, and A = Age. (B) R^2 (coefficient of determination), X^2 (Pearson statistic), and G^2 (likelihood ratio statistic) values are based on comparisons between predicted and observed *proportions*.

17 age groups) for the saturated model, to a minimum of 4 for the region-only or time-only models.

Although models that contain only 4 parameters do not accurately predict the observed data, there are several "relatively simple" unsaturated models that do the job quite well. The most successful of these is the one that includes only the single variable of age. This model requires only 17 parameters (one for each age group) yet it has X^2 and G^2 values that are close to those of models of higher order (i.e., the main effects [O][T][A] model). And, when this "age only" model is rescaled to account for the number of parameters, it has a G^2 value lower than any other unsaturated model, except for the main effects model with region-age and period-age interactions ([OA][TA]; 119 parameters) and the first-order interaction model ([OT][OA][TA]; 128 parameters). A comparison of the 15 models (14 unsaturated models plus the saturated model) is set out in Table 5.

The models in Table 5 can be separated into three major groups: main effects models ([O], [T] and [A]); interaction effects models ([OT], [OA] and [TA]); and the saturated model ([OTA]). We use Knoke and Burke's (1980) notation for abbreviating models by putting capital letters to represent the model effects of region [O], period [T], and age [A] in brackets. When there are only single letters in each bracket, a main effects model is represented. When more than one letter appears in a bracket, an interaction effects model is assumed. The letters in the brackets also define the hierarchical structure of the generalized linear model, e.g., a region-period interaction model may be specified as:

$$[OT] = [O][T][A][OT]. \tag{12}$$

To demonstrate the differences in the predicted values, consider the main effects [O][T][A], the first order interaction [OT][OA][TA], and the saturated [OTA] models. The predicted out-migration proportions from the first-order interaction model and the saturated model are nearly identical. The main effects model, on the other hand, assumes a constant age profile for the entire data set. That particular age profile corresponds to the average age profile for all time periods and all regions ($\hat{S}_{++}(x)$). The "level" of out-migration of this age profile is then adjusted according to the effects of period and region. Subsequently, the age profile of the [O][T][A] model includes an "average" retirement peak, one that under-predicts the Northeast's and Midwest's retirement peaks and over-predicts the South's and West's (nonexistent) retirement peaks.

The Distribution Component

The distribution component differs according to the age of the migrant. To include age, two approaches may be followed. First, age may be viewed as a stratification variable and be represented by a separate model for each age group. Second, age may be viewed as a predictor variable and be included in the set of independent variables.

Consider the age-specific destination proportions (distribution component) from the Northeast to the Midwest, South and West for the four time periods in this study (Figure 3). Those migrating to the Midwest and West regions were mostly persons of working age and their children, whereas those migrating to the South were mostly persons of older ages (i.e., elderly, aged 60+).

The attractiveness of the South for migrants of all age groups from the Northeast has increased over time. During the 1955–1960 period, about 45% of the migrants aged 20–29 went to the South from the Northeast. In the late 1960s, the percentage was about the same. But in the late 1970s it had increased to over 50%, and in the late 1980s it was close to 60%. For all periods, persons of retirement age, 60 to 64 year-olds who, say, left the Northeast, went disproportionately to the South. In the late 1950s, three out of four of these out-migrants from the Northeast went to the South. Ten years later, it was four out of five. It then increased to more than six out of seven in the most recent period (the late 1980s). The South always has been a major destination for migrants from the Northeast, and its importance has increased over the years. The overall age composition of the migrant destination proportions to the South did not change much. The share of persons of working age increased, the share of children decreased and the share of the oldest persons (75 +) increased substantially.

The observed age-specific migration flows from the Northeast to the Midwest, South and West during the 1985–1990 period are set out in Table 6, along with the predicted destination proportions of two multinomial logit models: a main effects only model, and a saturated model. Both models predicted destination proportions for all four time periods, but because of space constraints, only the predicted values for the last period (1985–1990) are shown. The main difference between the two models is that the main effects model has no interaction effect. The saturated model does, and thus predicts the observed data perfectly. The main effects (unsaturated) multinomial logit

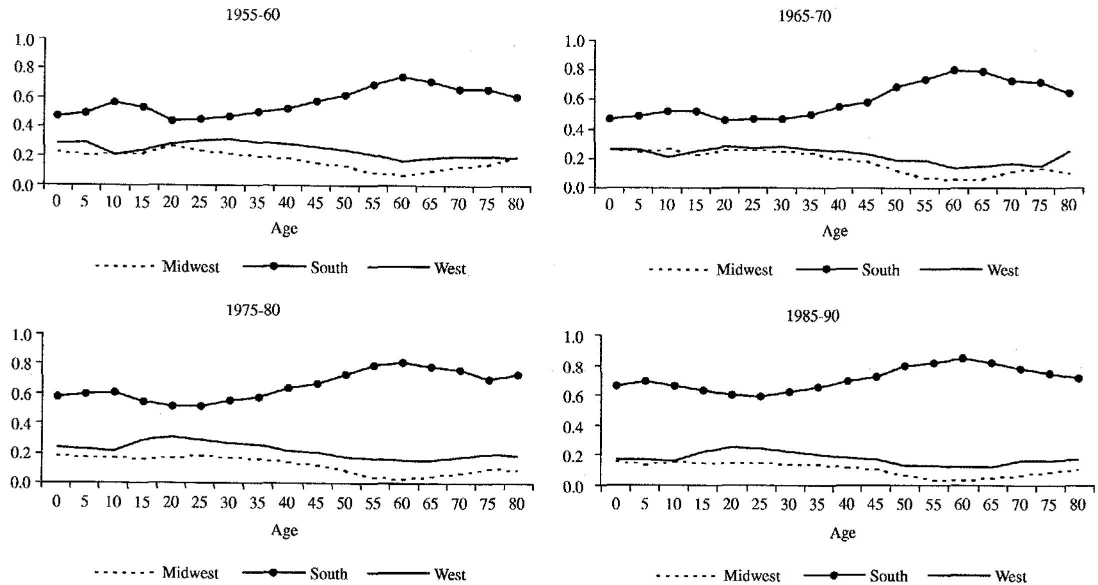

FIGURE 3 Observed age-specific destination proportions of U.S. migrants from the Northeast to the Midwest, South, and West: 1955–1960 to 1985–1990.

TABLE 6

Predicted destination proportions of migration from the Northeast to the Midwest, South, and West during the 1985–1990 period: main effects and saturated models

	Observed Flows				Predicted Destination Proportions							
					Main Effects				Saturated			
Age	Midwest	South	West	Total	Midwest	South	West	Total	Midwest	South	West	Total
0	29,686	121,068	31,798	182,552	0.1620	0.6359	0.2021	1.0000	0.1626	0.6632	0.1742	1.0000
5	21,461	107,224	26,740	155,425	0.1451	0.6572	0.1977	1.0000	0.1381	0.6899	0.1720	1.0000
10	33,553	134,265	33,489	201,307	0.1556	0.6775	0.1669	1.0000	0.1667	0.6670	0.1664	1.0000
15	55,923	232,479	81,184	369,586	0.1429	0.6434	0.2137	1.0000	0.1513	0.6290	0.2197	1.0000
20	56,286	226,348	96,186	378,820	0.1608	0.5940	0.2453	1.0000	0.1486	0.5975	0.2539	1.0000
25	47,577	183,081	76,198	306,856	0.1606	0.5990	0.2404	1.0000	0.1550	0.5966	0.2483	1.0000
30	33,802	145,213	53,260	232,275	0.1497	0.6234	0.2269	1.0000	0.1455	0.6252	0.2293	1.0000
35	25,725	118,284	36,875	180,884	0.1402	0.6489	0.2109	1.0000	0.1422	0.6539	0.2039	1.0000
40	15,519	84,162	22,261	121,942	0.1224	0.6893	0.1883	1.0000	0.1273	0.6902	0.1826	1.0000
45	9,883	66,419	15,667	91,969	0.1056	0.7216	0.1728	1.0000	0.1075	0.7222	0.1704	1.0000
50	6,390	71,452	12,129	89,971	0.0733	0.7841	0.1427	1.0000	0.0710	0.7942	0.1348	1.0000
55	5,295	99,769	16,466	121,530	0.0425	0.8267	0.1308	1.0000	0.0436	0.8209	0.1355	1.0000
60	4,568	105,657	14,831	125,056	0.0332	0.8548	0.1119	1.0000	0.0365	0.8449	0.1186	1.0000
65	4,049	60,628	9,262	73,939	0.0464	0.8353	0.1183	1.0000	0.0548	0.8200	0.1253	1.0000
70	2,717	31,494	6,622	40,833	0.0647	0.7975	0.1379	1.0000	0.0665	0.7713	0.1622	1.0000
75	2,466	19,373	4,191	26,030	0.0854	0.7705	0.1441	1.0000	0.0947	0.7443	0.1610	1.0000
80	2,418	14,980	3,704	21,102	0.0903	0.7480	0.1617	1.0000	0.1146	0.7099	0.1755	1.0000
Total	357,318	1,821,896	540,863	2,720,077								

Note: The predicted values come from models that included all time periods. The 1985–1990 period is only shown because of space constraints.

model for destination by time period and age, for a given region, is specified as:

$$\hat{\theta}_{jt|i}(x) = \frac{\hat{S}_{jt|i}(x)}{\hat{S}_{kt|i}(x)} = \nu_{j|i}\nu^{T}_{jt|i}\nu^{A}_{j|i}(x). \quad (13)$$

The difference between the predicted and observed values is due to the interaction between age and period, i.e., the changes in the age structure of migrants from the Northeast to the Midwest. The model omits the effects of interaction. In 1985–1990 the Midwest was less attractive to 20–24 year-old migrants from the Northeast than expected on the basis of average attractiveness during the past four decades. The South has become more attractive to the elderly, since the observed flow exceeds the flow predicted by the model. Table 6 shows the observed and the predicted destination proportion of migrants from the Northeast to the Midwest in 1985–1990. The main effects model predicts the migration flows adequately. Further improvement can be obtained by including interaction effects.

IMPOSING AGE STRUCTURES OF MIGRATION WITH OFFSETS

The Generation Component

Including age into the generalized linear model necessarily complicates the modeling process because a separate parameter is required for each age group (17 in our example). Consequently, methods that reduce the number of such parameters while maintaining the full detail of the age effects, are much desired. In this section, we offer the notion of model migration schedules as a possible method for reducing the number of parameters.

After modeling the age structures of hundreds of schedules of interregional migration, Rogers and Castro (1981) found sufficient regularities in age patterns to put forward a "standard" schedule and several variants. In particular, they put forward three basic categories of multiexponential model migration schedules: (1) a *standard* 7-parameter model, (2) a 9-parameter *elderly return migration* model, and (3) an 11-parameter *elderly retirement peak* model.

The *standard* age profile of migration can be defined as the sum of three components: (1) a constant curve; (2) a single negative exponential curve of the *prelabor force* ages, with its descent parameter; and,

(3) a left-skewed unimodal curve of the *labor force* ages positioned at a point on the age axis and exhibiting parameters of ascent and descent. The elderly retirement peak age profile requires an 11-parameter model migration schedule. To estimate this function, one simply adds another double-exponential function to the earlier equation.

The three age structures set out in Figure 4 represent the model migration schedules that we applied to our regional and time period data. First, a 7-parameter model migration schedule was applied to the entire observed out-migration data set to obtain a single "average" age profile of migration (Figure 4A). Second, for the purpose of improving the predicted out-migration flows, an 11-parameter model migration schedule was applied separately to the out-migration proportions from the Northeast and Midwest (Figure 4B), and to the out-migration proportions from the South and West (Figure 4C). Because the out-migration proportions from the South and West do not contain evidence of a retirement migration peak, the relevant 4 parameters were set to zero, essentially forcing the model to be a 7-parameter model migration schedule.

To impose an age profile of migration, we used the model migration schedules presented in Figure 4 to define the age structures: one that assumed a single age profile for the entire data set (Figure 4A), and another that differentiated between the age profiles of the Northeast/ Midwest and South/West outflows (Figures 4B and 4C). Such age profiles were imposed using a log-linear model with an *offset*. The model we used to predict conditional out-migration proportions was:

$$\hat{S}_{it}(x) = \bar{S}^{*}_{it}(x)\xi\xi^{O}_{i}\xi^{T}_{t}, \tag{14}$$

where $\bar{S}^{*}_{it}(x)$ is the offset, or in our case, the age profiles presented in Figure 4. *This model requires parameters only for the region and period main effects, because the offset essentially replaces the age main effects parameters*. So, the level of the age-specific offset rises or falls according to the specific effects of the region of origin and the time period. The parameters are set out in Table 7.

The predicted proportions of log-linear offset models (both the 7-parameter and the 11-parameter versions) can be contrasted to the main effects model (Table 5), which required 23 model parameters. As set out in Table 7B, the results from the log-rate model with the 7-parameter offset ($R^2 = 0.877$) were nearly the same as those of the main effects model ($R^2 = 0.891$), except that fewer parameters were required, making the re-scaled X^2 and G^2 goodness-of-fit measures

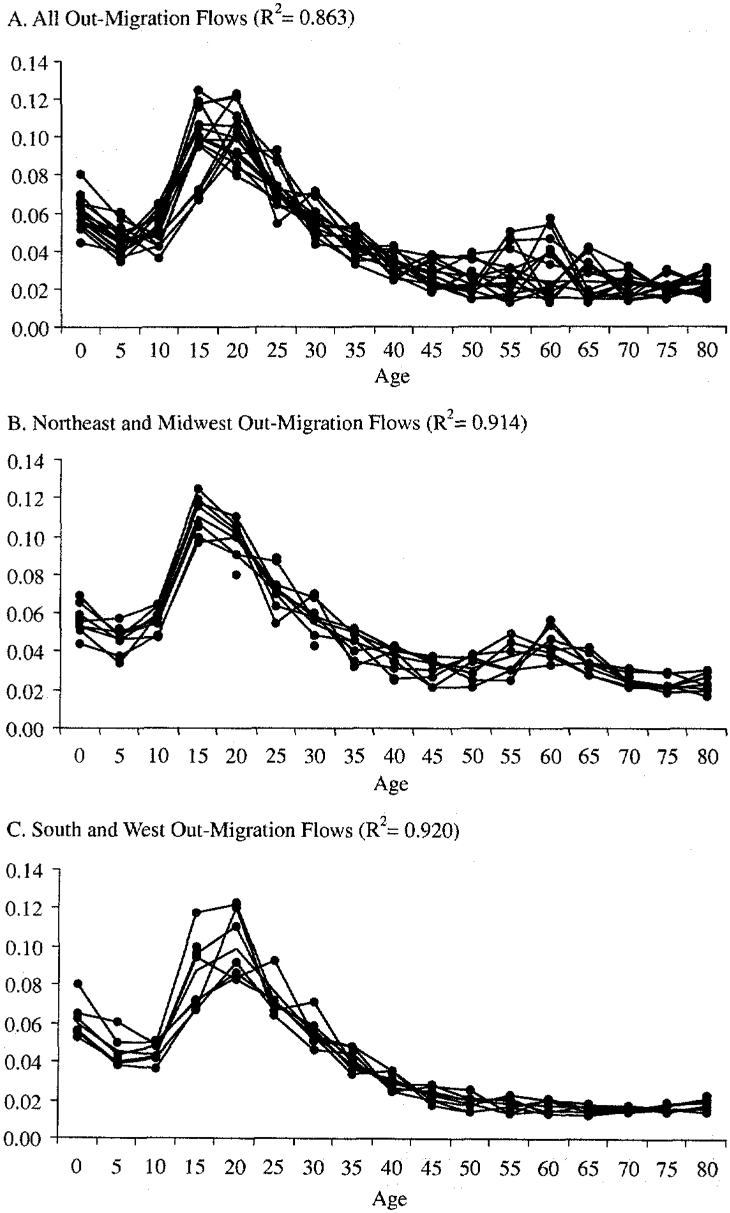

FIGURE 4 Model migration schedules of observed regional out-migration proportions: 1955–1960 to 1985–1990.

TABLE 7
Parameters and goodness-of-fit measures of the log-linear model with offset [R][T] with age structure as the offset

	Model Migration Schedule Offset	
	7-Parameter	11-Parameter
(A) Parameters of the Log-linear Model with offset		
Intercept	0.9549	1.0576
Northeast	1.1247	0.9264
Midwest	1.1748	0.9676
South	0.8927	0.8927
1955–60	0.9563	0.9563
1965–70	0.9940	0.9940
1975–80	1.0463	1.0463
(B) Goodness-of-Fit Measures		
R^2	0.877	0.922
X^2	0.519	0.265
G^2	0.488	0.262
Number of Indep. Parameters	14	18
Re-scaled X^2	7.265	4.770
Re-scaled G^2	6.828	4.724

Note: (A) R = region of origin and T = time period. (B) The number of independent parameters include the model migration schedule parameters and the log-linear model with offset parameters, respectively: $14 = 7 + 7$ and $18 = 11 + 7$.

significantly lower in comparison. The log-linear offset model with the 11-parameter offset ($R^2 = 0.922$) improved the model fit substantially, but at the expense of 4 additional parameters.

The Distribution Component

We have seen that model migration schedules may be used to impose a particular age profile of out-migration rates. The corresponding age-specific destination proportions, however, do not exhibit the typical age pattern exhibited by the generation component. Consequently, the standard model age profiles are not suited for representing the destination component. Instead, one may select age-specific destination proportions representative of one period and *impose* them (i.e., their spatial structure) to define the destination components of other periods.

Although offsets were used in the previous sub-section, they were not applied there to entire matrices of flows. Now they will be. Consider the question: What would have been the destination of migrants from the Northeast in 1965–1970 if the age-specific destination proportions of 1955–1960 had been imposed onto the age structure of the out-migrants from the Northeast during the period 1975–1980, and if the total (all ages combined) destination proportions of the latter period were to be preserved? The model of proportions to be estimated is a logit model (or logistic regression) with an offset. *But generally available software packages (e.g., SPSS) have no procedures to estimate multinomial logit models with offsets. Therefore, log-linear models need to be estimated instead.*

In the case of log-linear models, the offset is the number of migrants. For example, the offset used may represent the number of migrants leaving the Northeast to move to the other regions in 1955–1960. Consider the observed number of migrants from the Northeast shown in Table 6. Three unsaturated models have been estimated, the parameter values of which are shown in Table 8. The first model includes the age effect only. The number of migrants predicted in any period, including the initial period 1955–1960, is the number of migrants in

TABLE 8
Log-linear model used to predict the 1975–1980 migration destinations using 1955–1960 destinations as offset

Parameter		Model 1	Model 2	Model 3
Intercept		3.1826	2.8002	3.1867
Age	0	0.3518	0.3638	0.3638
	5	0.4067	0.4184	0.4184
	10	0.4109	0.4162	0.4162
	15	0.4771	0.4863	0.4863
	20	0.5100	0.5323	0.5323
	25	0.4651	0.4836	0.4836
	30	0.4017	0.4155	0.4155
	35	0.4092	0.4195	0.4195
	40	0.4056	0.4133	0.4133
	45	0.4284	0.4309	0.4309
	50	0.4659	0.4627	0.4627
	55	0.5619	0.5483	0.5483
	60	0.5147	0.4958	0.4958
	65	0.4994	0.4857	0.4857
	70	0.5475	0.5389	0.5389
	75	0.7216	0.7116	0.7116
Destination	Midwest		0.9198	0.9198
	South		1.2478	1.2478
Period	1955–1960			0.6188
	1965–1970			0.7878
	1975–1980			1.1084

1955–1960 multiplied by an age factor. The age factor is $\tau^{A}_{j|i}(x)$. The τ-parameters are set out in Table 8 in the column titled "Model 1". For instance, the number of persons aged 20–24 in 1985 and residing in the Northeast at that time, who then lived in the West at the time of the 1990 census is predicted to be $56,139[(3.1826)(0.5100)] = 91,113$. The observed number of migrants in 1985–1990 was 96,186. The migration to the West therefore is larger than one would expect on the basis of the average age pattern of migration during the past four decades and the age and destination patterns of migration from the Northeast in the late fifties.

Note that the number of migrants predicted by Model 1 is the same, irrespective of the period that is considered. Since it does not include a period effect, it predicts the same number for 1955–1960, 1965–1970, 1975–1980 and 1985–1990. The observed numbers are 56,139, 75,926, 137,185, and 96,186, respectively. (The fact that the model predicts the flow in 1985–1990 relatively accurately is purely accidental.) The predicted number of migrations to the destinations can be obtained by applying the same age effect to the migration during the period 1955–1960. For instance, the number of migrants from the Northeast to the South in 1985–1990, of age 20 in 1985, is predicted to be the number in 1955–1960 multiplied by the age effect: $88,260[(3.1826)(0.5100)] = 143,245$. This figure differs considerably from its observed counterpart figure of 226,348. When we introduce a destination effect to account for the differing attractiveness of the regions, the predictive performance of the model improves (Model 2). Such a model predicts that 83,688 migrants, aged 20–24 in 1985, change residence from the Northeast to the West during the 1985–90 period: 56,319[(2.8002) (0.5323)].

Finally, Model 3 adds a period effect to account for the changing levels of migration. This model is the unsaturated model that includes the main effects of destination, age and period (sometimes referred to as [D] [T] [A]), plus an offset which is the migration pattern in 1955–1960:

$$\hat{n}_{1j4}(x) = n_{1j4}(x)\xi_1\xi^{D}_{j|1}\xi^{T}_{4|1}\xi^{A}_{1}(x).$$

From the late 1950s to the late 1970s, migration increased, and in the late 1980s it declined again. The number of migrants of a given age, moving to a given destination, during a given time period depends on the main effects of destination, period and age.

Consider migrants aged 20–24 in 1985. The number of persons of that age who migrated from the Northeast to the West in the period

1985–1990 depends on the average age patterns of migration, the average attractiveness of the West during the past four decades and the level of migration out of the Northeast in 1985–1990. In addition, it also depends on the number of migrants aged 20–24 years moving from the Northeast to the West in 1955–1960, namely, the offset:

$$\begin{aligned}\hat{n}_{144}(20) &= n_{144}(20)\xi\xi_4^D\xi_4^T\xi^A(20)\\ &= 56,139[(3.1867)(0.5323)] = 95,236.\end{aligned}$$

CONCLUSION

In this article, we have focused on the structure of age-specific out-migration across regions and over time. In doing so, we have found several interesting things worth summarizing. First, age patterns of migration exhibit strong regularities. Origin effects appear to represent differences in the shapes of migration, whereas time effects tend to represent differences in the levels of migration. Once one finds that strong regularities in overall age profiles of migration exist, one can then choose different techniques to represent changes in such profiles across regions and over time.

When all variables are discrete, as is the case in this article, log-linear models do not differ from Poisson regression models, and logit models are identical to logistic regression models. These equivalences are used extensively in this article to estimate the models using standard statistical software (SPSS) and to enhance the interpretation of the regression coefficients or parameters of the models. Another relation between categories of models was used to detect patterns in the time series of migration data, namely, the relation that exists between the log-linear model and the logit model. That relation is particularly useful when prior information, e.g., a historical pattern of destination proportions, is introduced to enhance the study of change.

The models used in this article have two major strengths. First, they decompose observed patterns of migration to identify the separate effects of all variables considered in the definition of the pattern. The variables considered have been region of origin, region of destination, time period and age, together with their interactions. The effect of the region of destination, i.e., the attractiveness of the destination, is expressed in terms of odds. A region is attractive if a migrant is more likely to select that region than another region (i.e., the reference region). The odds are a logical measure of attractiveness in migration analyses. The attractiveness of a destination may depend on the current residence of migrants. The relative attractiveness then is expressed in terms of odds ratios.

Time is a major variable in this article. When the effect of time is weak, either its main effect or its interaction effect with another variable, then the migration pattern is deemed stable. In this case the parameter of the log-linear or the logit model is not much different from unity in the multiplicative specification of the model. Second, the models accommodate partial information on the destination choice and then impose a spatial structure and test its significance as a predictor of migration patterns. The spatial structure often used is a historical pattern, and the question then is whether that historical pattern is still adequate to predict more recent or current patterns of migration. Such predictions are most accurate when the migration spatial structure has been relatively stable.

Generation and distribution have been the two principal components used in this article to represent continuity and change in interregional migration flows. Taken together they offer the analyst a methodology for identifying the spatial structure of a migration pattern, assessing its relative stability over time and, if needed, of imposing that particular spatial structure onto a different migration setting.

REFERENCES

Alba, R. D. (1988). Interpreting the parameters of log-linear models. In J. S. Long, *Common Problems/Proper Solutions: Avoiding Error in Quantitative Research*, Newbury Park: Sage. pp. 258–287.

Alonso, W. (1978). A theory of movement. In N. M. Hansen (ed.), *Human Settlement Systems*. Cambridge: Ballinger.

Alonso, W. (1986). Systemic and log-linear models: from here to there, then to now, and this to that. *Discussion Paper 86–10*, Center for Population Studies, Harvard University, Cambridge, Massachusetts.

Azzalini, A. (1996). *Statistical inference: Based on the likelihood.* London: Chapman and Hall.

Brass, W. (1974). Perspectives in population prediction: illustrated by the statistics of England and Wales. *Journal of the Royal Statistical Society A*, **137**: 532–570.

Coale, A. J. and Trussell, J. (1974). Model fertility schedules: Variations in the age structure of childbearing in human populations. *Population Index*, **40**: 185–258.

Haberman, S. J. (1978). *Analysis of qualitative data: New developments* (Vol. 1). New York: Academic Press.

Haberman, S. J. (1979). *Analysis of qualitative data: New developments* (Vol. 2). New York: Academic Press.

Knoke, D. and Burke P. J. (1980). *Log-linear models*. Newbury Park: Sage.

Lin, G. (1999). Assessing structural change in U. S. migration patterns: A log-rate modeling approach. *Mathematical Population Studies*, **7**: 217–237.

Mueser, P. (1989). The spatial structure of migration: An analysis of flows between states in the USA over three decades. *Regional Studies*, **23**: 185–200.

Olsson, G. (1965). *Distance and human interaction: A review and bibliography*. Philadelphia: Regional Science Research Institute.

Plane, D. A. (1984). A systematic demographic efficiency analysis of U. S. interstate population exchange, 1935–1980. *Economic Geography*, **60**: 294–312.

Rogers, A. (1995). *Multiregional demography: Principles, methods, and extensions*. Chichester: John Wiley.

Rogers, A. and Castro, L. (1981). Model migration schedules. Research Report, International Institute for Applied Systems Analysis, Laxenburg, Austria.

Rogers, A. and Wilson, R. T. (1996). Representing structural change in U.S. migration patterns. Geographical Analysis, 28: 1–17.

Rogers, A., Willekens, F. J., Little, J. and Raymer, J. (2002). Describing migration spatial structure. *Papers in Regional Science*, forthcoming.

Rogers, A., Willekens, F. J. and Raymer, J. (2001). The indirect estimation of migration flows. *Working Paper WP-01-01*, Population Program, University of Colorado, Boulder, Colorado.

Snickars, F. and Weibull, J. W. (1977). A minimum information principle: Theory and practice. *Regional Science and Urban Economics*, **7**: 137–168.

Willekens, F. J. (1980). Entropy, multiproportional adjustment and the analysis of contingency tables. *Systemi Urbani*, **2**: 171–201.

Willekens, F. J. (1982). Multidimensional population analysis with incomplete data. In K. Land and A. Rogers (eds), *Multidimensional Mathematical Demography*. New York: Academic Press. pp. 43–111.

Willekens, F. J. (1983). Log-linear modeling of spatial interaction. *Papers of the Regional Science Association*, **52**: 187–205.

Willekens, F. J. and Baydar, N. (1986). Forecasting place-to-place migration with generalized linear models. In R. I. Woods and P. Rees (eds), *Population structures and models: Developments in spatial demography*. London: Allen and Unwin. pp. 203–244.

Willekens, F. J., Por, A. and Raquillet, R. (1981). Entropy, multiproportional, and quadratic techniques for inferring patterns of migration from aggregate data. *A Journal of International Applied Systems Analysis*, **4**: 83–124.

Wilson, A. G. (1970). *Entropy in urban and regional modeling*. London: Pion.

Wrigley, N. (1985). *Categorical data analysis for geographers and environmental scientists*. London: Longman.