Papers Reg. Sci. 81, 29–48 (2002)

Papers in
Regional
Science

Describing migration spatial structure*

Andrei Rogers[1], Frans Willekens[2], Jani Little[1], James Raymer[1]

[1] Population Program, Institute of Behavioral Science, University of Colorado, Boulder, Colorado 80309-0484, USA (e-mail: andrei.rogers@colorado.edu)
[2] Population Research Centre, Faculty of Spatial Sciences, University of Groningen, PO Box 800, 9700 AV Groningen, The Netherlands

Received: 11 October 2000 / Accepted: 3 April 2001

Abstract. The age structure of a population is a fundamental concept in demography and is generally depicted in the form of an age pyramid. The spatial structure of an interregional system of origin-destination-specific migration streams is, however, a notion lacking a widely accepted definition. We offer a definition in this article, one that draws on the log-linear specification of the geographer's spatial interaction model. We illustrate our definition with observed migration data, we discuss extensions and special cases, and proceed to contrast our definition and associated empirical findings against another measure having an alternative definition.

JEL classification: R23, C25

Key words: Migration, spatial structure, log-linear models

1 Introduction

The notion of age structure is a central concept in demography, however, the notion of spatial structure, and for example, that of migration, is not. Spatial structure has been used to develop functional representations of age patterns of a population or that of a migrant stream, and is furthermore the basis for constructing so-called model schedules: mathematical expressions to describe age patterns of mortality, fertility and migration. Conversely, migration has no such widely accepted mathematical representation.

As a demographic process, migration stands apart from fertility and mortality because of the explicitly spatial nature of migration. Unlike fertility and mortality

* This research is supported by a grant from the National Science Foundation (BCS-9986203). The authors gratefully acknowledge the helpful comments of Kao-Lee Liaw, Richard Rogers and two anonymous reviewers.

processes, which affect the population of only one region, aggregate migration flows interact within a multi-regional system in which departures from each region affect the populations of several other regions. The result is a decrease of people from each region of origin and a differential addition of people to each region of destination. A representation of this complex process and associated data structure must therefore arise from a model that incorporates the influences of population sizes at the origins and destinations, and one that also includes some sort of “separation” or “interaction” factor between each pair of origins and destinations.

We define migration spatial structure as a particular description of a matrix of inter-regional migration flows that provide an analyst with the means to: (1) reconstruct that matrix of flows, (2) identify the implied relative “push” at each origin and “pull” of each destination (here called “emissiveness” and “attractiveness”, respectively), and (3) express the origin-destination-specific levels of spatial interaction implied by that matrix. Spatial interaction is here understood to reflect the degree of deviation exhibited by that matrix when compared to the corresponding matrix generated under the assumption of no spatial interaction, i.e., a situation in which origin-destination-specific migration flows, rates, or probabilities do not depend on regions of origin. The larger the deviation the stronger the degree of spatial interaction.

The linkage between age structure and the analysis of fertility, mortality and migration processes is central to demographic study. Standardised mathematical representations of age patterns, called model schedules, allow demographers to accurately define age-specific patterns with continuous functions described with few parameters. The corresponding mathematical representation of spatial patterns calls for a somewhat more complex statistical structure. A powerful, yet conceptually simple instrument for the study of aggregate migration spatial structure is offered by the family of generalised linear models. In particular is the log-linear model, which provides a more readily interpretable mathematical representation of migration flow structure than the flow matrix itself. Just as model schedules are used to compare across time and place, the log-linear specification can be employed as an especially valuable statistical model for comparing inter-regional migration structures across time. The parameters of the log-linear model can not only gauge the relative emissiveness and attractiveness (Liaw and Rogers 1999) of specific regions, but also for the level of interaction between pairs of regions. Because the parameters of the model are interpretable and can be used to characterise migration spatial structure, the log-linear model can potentially standardise and enhance a demographic analysis. Furthermore, the model can be used in the indirect estimation of migration flows, particularly when such flows cannot be derived from available data.

The main contributions of the log-linear approach are that the parameters of the model capture different features of the spatial structure of migration, with one set of parameters representing the effect of the sizes of origin populations, another set representing the corresponding effects of the sizes of destination populations, and still another set representing the strengths of the linkages between these two populations. This parameterisation facilitates comparisons of spatial structures. The method can be applied to a multi-regional system comprised of many regions and to theoretical as well as observed spatial structures. A log-linear approach also decomposes the spatial structure into contributing structural factors. For example, the number of migrants from a region, i say, to another region, j, depends on the size and composition of the population of region i and on the size and composition of the population of region j.

But size and composition alone are insufficient to characterise the flows of migrants. A history of migration from i to j may be a greater determinant of current migration structure than the particular characteristics of the two regions. We present here a method able to capture the effect of historical migration patterns on contemporary migration. The method of offsets (Knudsen 1992) facilitates the quantitative assessment of historical changes among observed structures and their influences on contemporary spatial structures. A noteworthy observation is that this method belongs to the family of log-linear models. Only the log-linear model is able to isolate the various factors of the spatial structure most effectively (even more than the logit model and the log-rate model).

We next discuss a brief overview of previous efforts to describe the spatial structure of migration. We then focus on the merits of the general spatial interaction model and emphasise the functional equivalence between this model and the log-linear model. We also offer an exposition of the log-linear model, showing how it can be used to represent the components of migration spatial structure, illustrating its use with particular numerical examples, and considering its extensions and wider implications. The results of the numerical examples of the log-linear approach are compared with results obtained with the neutral migration process, which also endeavours to describe various attributes of migration spatial structure through the definition of indices of origin emissiveness and destination attractiveness (Liaw and Rogers 1999). A final discussion and a number of conclusions complete the article.

2 Representing spatial structures of migration: The log-linear model

2.1 Overview

The literature on migration is curiously ambiguous on the definition of migration spatial structure and how it should be measured. An early effort to describe the structure of migration is Shryock (1964, p. 267), who advanced a set of preference indices focused on the ratio of actual to expected number of migrants in a stream; the latter defined as being proportionate to both the population at origin and the population at destination. Conversely, Clayton (1977, p. 109) defined migration spatial structure as the way in which origins and destinations are linked in terms of their exchanges of migrants. He implemented his definition by identifying those regions (states in his application) that represent major origins and destinations in the interstate migration system. He used nodal and principal component analyses to identify such places and delineated a number of migration fields.

Plane (1984) and Manson and Groop (1996), among others, relied on the widely used notion of migration efficiency (or efficacy) and applied it to interstate migration matrices to identify changes in migration system structure. In a recent co-authored article with Mulligan, Plane adopted the well-known Gini index of concentration to identify the spatial focus exhibited by a set of origin-destination-specific migration flows. The aim was to measure the strength of that concentration by the departure from equality in the distribution of migration streams exhibited by an observed origin-destination-specific matrix of flows (Plane and Mulligan 1997).

Finally, Mueser (1989) fitted a generalised spatial interaction model to data on migration flows between U.S. states over a 30 year period. Mueser's work is important in that it demonstrates the ability of the spatial interaction model to clearly represent the structural components of migration. Due to his reliance on the spatial interaction model, he was able to decompose migration structure into the sending effects of each region, the region's ability to draw migrants, and the inter-regional interaction or separation effects. His findings on migration flow stability conflict with Plane (1984) and Manson and Groop (1996), probably due to differences in methodological approaches. Instead of instability, he finds great stability in the separation effects, i.e., the relative attachments between regions over time. Mueser concludes that there are changes in the relative volumes of migration streams, but these are because of the relative desirability or attraction of different locations rather than to the interactions between them.

2.2 The spatial interaction model and the log-linear model

The spatial interaction model, once so popular in human geography (Haynes and Fotheringham 1984), has proven to be the most useful method for representing the spatial structure of migration (Willekens 1983; Alonso 1986; Mueser 1989). The generality of the method incorporates most models used to examine migration streams, including the gravity model, entropy maximisation, information minimisation, bi-proportional adjustment, the systemic model of movement, random utility models based on choice theory, and the log-linear model. A formal equivalence exists between the log-linear model and the gravity model and entropy maximisation model (e.g., see Willekens 1980, 1982a,b, 1983; Scholten and Van Wissen 1985; Bennett and Haining 1985; Aufhauser and Fischer 1985; and Alonso 1986).

The log-linear (generalised linear) model is a powerful instrument for the analysis of complex data structures; the fact that it can express traditional models of spatial interaction enhances the opportunities for structural analysis. Questions that the data are expected to help answer can be expressed in terms of the parameters of the model. Furthermore, the model clarifies and simplifies the estimation of spatial interaction flows. And when particular interaction effects cannot be derived from available data, they may often be calculated using other comparable data sets, (e.g., historical data on interaction). Snickars and Weibull (1977) found that migration tables of the past provide much better estimates of current accessibility than any distance measure, and since then historical data are frequently used in spatial interaction analyses to capture spatial accessibility. A drawback of using historical information is the assumption that spatial interactions somehow remain stable, i.e., that migration regimes are fixed. However, new research on matrix transformation methods and log-rate models for representing past age and spatial patterns of structural change (Rogers and Taylor-Wilson 1996; Lin 1999) provides us with a logical way to relax the strict assumption of an unchanging regime.

2.3 Numerical examples of the log-linear decomposition

The observed number of native-born migrants flowing to and from each of the four Census regions during the 1985–1990 period are set out in Table 1A below. (Comparable data on foreign-born migrants have been omitted in the interests of maintaining a degree of homogeneity among the migrants studied, and to reduce the tabular information in what is already an unusually dense numerical exercise in description. The omitted data will be analysed in a sequel.) All persons who died, were born, or left or entered the country during the analysis period are excluded. Data for the corresponding 1975–1980 period are set out in Table 1B. We now ask how we can describe and compare the migration spatial structures exhibited by these flow matrices using a spatial interaction model for each.

Table 1. U.S. native-born interregional migration flows: 1985–1990 and 1975–1980

		Origin	Destination region				
	Period	region	Northeast	Midwest	South	West	Total
A.	1985–1990	Northeast	40,262,319	336,091	1,645,843	479,819	42,724,072
		Midwest	351,029	50,677,007	1,692,687	958,696	53,679,419
		South	778,868	1,197,134	69,563,871	1,150,649	72,690,522
		West	348,892	668,979	1,082,104	37,872,893	39,972,868
		Total	41,741,108	52,879,211	73,984,505	40,462,057	209,066,881
B.	1975–1980	Northeast	38,952,034	432,155	1,608,960	672,337	41,665,486
		Midwest	327,909	49,242,680	1,764,990	1,189,854	52,525,433
		South	644,692	1,035,277	64,446,418	1,046,823	67,173,210
		West	261,830	643,100	1,036,940	33,573,888	35,515,758
		Total	40,186,465	51,353,212	68,857,308	36,482,902	196,879,887

The elements ($M(i,j)$, say) of the two 4 by 4 matrices can be expressed as follows:

$$M(i,j) = KP(i)Q(j)F(i,j), \tag{1}$$

the well-known gravity model. Such a specification of the model is consistent with that of the log-linear (multiplicative) model. Taking the natural logarithm of $M(i,j)$ results in the corresponding (linear) additive model. Model (1) has

Table 2. Saturated log-linear model parameters of U.S. native-born interregional migration flows: 1985–1990 and 1975–1980

	Period	Origin region	Destination region Northeast	Midwest	South	West	Row Effect
A.	1985–1990	Northeast	34.3181	0.2087	0.5161	0.2706	0.8381
		Midwest	0.2334	24.5432	0.4140	0.4217	1.0744
		South	0.4084	0.4572	13.4176	0.3991	1.3624
		West	0.3057	0.4270	0.3488	21.9571	0.8152
		Column Effect	0.6489	0.8908	1.7639	0.9808	2157393*
B.	1975–1980	Northeast	32.9459	0.2283	0.4443	0.2992	0.9757
		Midwest	0.2374	22.2685	0.4172	0.4533	1.1398
		South	0.4348	0.4361	14.1915	0.3716	1.2235
		West	0.2940	0.4510	0.3801	19.8385	0.7349
		Column effect	0.5723	0.9164	1.7531	1.0876	2117193*

* Overall effect.

25 parameters, considerably more than the 16 cells in the migration matrix. The parameters are not independent, however; there are only 16 independent parameters. The other parameters are obtained by adopting normalisation restrictions. The types of restrictions are determined by the coding scheme. When the coding scheme is effect coding (see below), then the product of the parameters along a particular dimension is equal to one. In the case of contrast coding the parameters associated with the reference categories are equal to one.

Table 2 displays the numerical values of the parameters of the saturated log-linear model; the parameters are not independent. Since the number of parameters exceeds the number of independent observations on migration, restrictions must be imposed on the parameter values. The restrictions imply a particular coding scheme; the coding scheme adopted in the estimation of the parameters in Table 2 is effect coding, which means that, except for the overall effect, differences from mean values are estimated. For example, consider the number of migrants moving from the Northeast to the South in 1985–1990:

$$\begin{aligned} M_{85}(1,3) &= KP(1)Q(3)F(1,3) \\ &= 2,157,393\ (0.84)\ (1.76)\ (0.52) \\ &= 1,645,843 \end{aligned}$$

The corresponding calculation for the 1975-80 period is:

$$\begin{aligned} M_{75}(1,3) &= KP(1)Q(3)F(1,3) \\ &= 2,117,193\ (0.98)\ (1.75)\ (0.44) \\ &= 1,608,960 \end{aligned}$$

The K parameter in each instance represents the overall effect; under the current coding scheme, it is equal to the geometric mean value. This value is adjusted (i.e., scaled) by the row and column marginal totals, $P(i)$ and $Q(j)$, respectively, leaving the doubly-constrained interaction term $F(i,j)$ as the influence of what Meuser (1989) calls the spatial separation component.

The method of geometric means represents the oldest procedure for estimating the parameters of the log-linear model. Proposed by Birch (1963) and since elaborated by several authors (for a good description, see e.g., Payne 1977), it provides the parameters of the log-linear model in its multiplicative specification. The parameters are exponantiations of the parameters of the conventional specification of the log-linear model, the additive specification (see e.g., Payne 1977). When a specified model is describes the data perfectly (saturated), then overall effect (K) is the geometric mean of all migration flows. Consider the migration in 1985–1990. The geometric mean of the 16 flows (4 x 4) is equal to 2,157,393 migrants. This number can now be used as a reference. The row effect is the ratio of the geometric mean of the flows originating in a particular region as well as the overall geometric mean. For instance, the geometric mean flow originating in the Northeast is 1,808,029, and the ratio of this number over the overall effect is 0.84, the row effect. The column effect is obtained analogously. The interaction effect, which represents the origin-destination interaction, is obtained as the ratio of the migration flow from i to j, the product of the overall effect, row effect and column effect. The interaction effect of the migration from the Northeast to the South in 1985–1990 is 1,645,843/[2,157,393 (0.84) (1.76)] = 0.52.

Effect coding is only one of the possible coding schemes. Different software packages use different schemes as defaults. Some packages, such as SPSS, use the last category of a variable as the reference category, of which the coding scheme is referred to as contrast coding. In that case the overall effect is migration from the West to the West (number of persons who stay in the West), which is 37,872,893.

The row effect associated with region i is the ratio of the number of migrants from i to the West over the number of stayers in the West. For instance, the row effect associated with the Northeast is 479,819/37,872,893 = 0.0127. The column effect is the ratio of the number of migrants from the West to region j and the number of stayers in the West. For instance, the column effect associated with the Northeast is 348,892/37,872,893 = 0.0092. The interaction effect is the ratio of the migration from i to j and the product of the overall effect and the main effects. For instance, the interaction effect associated with the migration from the Northeast to the South is 1,645,843 / [37,872,893*0.0127*0.0286] = 120.05. The product of the overall effect, the main effects and the interaction effect yields the observed number of migrants from the Northeast to the South, thus, in the case of the saturated log-linear model, explaining the estimation procedure.

All information used in the decomposition to arrive at the log-linear parameters is contained in the interior 4 by 4 matrix of Table 1. The row sums, column sums and the total population are simply displayed as a convention. K (= 2,157,393 for the 1985–1990 flow matrix in Table 1A) is the geometric mean. K also serves as the reference value for the other parameters. For example, the parameters $P(i)$ and $Q(j)$, reflect the ratio of the geometric mean of row i and column j, respectively, to K. The interaction parameters $F(i,j)$ captures the effects not taken in by the overall effect and the row and column effects. They represent interaction effects, i.e., interactions between rows (origins) and columns (destinations). The $P(i)$ and $Q(j)$ effects are sometimes referred to as the balancing factors. Each $P(i)$ is the geometric mean of the four elements in the ith row divided by the overall geometric mean K. $Q(j)$ is the geometric mean of the four elements in the jth column divided by the overall geometric mean K. The product of the four $P(i)$'s and the product of the four $Q(j)$'s is equal to 1.

The interior 4 by 4 matrix in Table 2 contains the $F(i,j)$ effects in Equation (1): the spatial interaction matrix F. These effects reflect the extent of deviation after controlling the overall (K), row ($P(i)$), and column ($Q(j)$) effects. The interaction effects $F(i,j)$ represent accessibilities or ease of interactions. Off-diagonal elements can also be interpreted as attractiveness measures of the jth region to migrants from the ith region. The interaction effects $F(i,j)$ are simply the observed flows divided by the product of the other parameters.

Alonso (1986) argues that the $P(i)$ effect shows the relative importance of region i as a source of migrants. The $P(i)$ effects are the row marginals in Table 2 and for the 1985–1990 matrix; for example, they show that the Northeast (0.84) and the West (0.82) have fewer people in 1985 than did the Midwest (1.07) and the South (1.36). This is because $P(i)$ is a ratio of two geometric means: the row mean to the table's overall mean. These effect measures do not always have the same relative sizes as the populations of the origins, because the geometric mean of the row elements can be larger for one region than another, even though the origin population in that region is smaller than that in the other region. We best demonstrate using the $Q(j)$ effects in Table 2. The interpretation of the $Q(j)$ effects (the column effects) shows the relative importance of the jth region measured in terms of the size of its population at the end of the interval (1990). For example, in the 1985–1990 data, the Northeast is clearly the least important region (0.65) – less important than the West (0.98) – but the total size of the Northeast after the 1985–1990 transition (41,741,108) is larger than that of the West (40,462,057). The South is the most important region at the beginning and at the end of our observation period; it has the largest diagonal value (stayers) as well as the largest off-diagonal values (migrants).

All effects taken together represent the spatial structure of migration. Since the row and column effects are ratios of geometric means, one can compare the row and column effects for each region. For example, the Northeast is more important as an origin than as destination ($0.84 > 0.65$) and the South is more important as a destination than as an origin ($1.76 > 1.36$).

An important feature of log-linear decompositions of flow matrices is the scale-free or dimensionless nature of all parameters except for K. One can re-scale the entire matrix and obtain the same row and column parameter estimates as before. However, when the same flow numbers are converted to the associated matrix of transition probabilities, the log-linear decompositions are no longer scale-free. The elements within each row retain their relationship to each other, but the relative sizes of the origins are no longer preserved. The log-linear interaction and column effects are preserved when an observed flow matrix is transformed to the corresponding transition matrix.

2.4 Bi-proportional adjustment and the method of offsets

The utility of the log-linear model extends beyond the convenient decomposition of the observed flow matrix into interpretable parameters for describing the structure of migration. The log-linear model as a generalised linear model is a powerful instrument for the study of complex data structures. Here, we demonstrate how the log-linear model can be used to predict the migration flows in one period on the basis of flows observed in a previous period. The use of historical data to capture spatial accessibility or spatial interaction hinges on the assumption that spatial interaction effects are stable over time; there is substantial support in the literature (Willekens 1983; Nair 1985; Mueser 1986). In fact, Snickars and Weibull (1977) found that migration tables of the past provide much better estimates of current accessibility than any distance measure.

In a number of different applied areas analysts have used an iterative algorithm to adjust a historical matrix to sum to new row and column marginal totals: the bi-proportional adjustment method (Bacharach 1970) effectively imposes the structure found in the historical matrix on the subsequent migration time period.

Consider for example, the "historical" flow matrix:

$$\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}.$$

Suppose that the row and column totals are doubled, then clearly:

$$\mathbf{B} = \begin{bmatrix} 2 & 4 \\ 4 & 2 \end{bmatrix}$$

is a flow matrix with the same interaction effects. But what if, instead, only the row totals are doubled and the column totals shift to 4 and 8? How do we impose, as much as possible, the spatial structure of **A** onto the set of marginals? The iterative bi-proportional adjustment method yields the matrix:

$$\mathbf{C} = \begin{bmatrix} 1.123 & 4.877 \\ 2.877 & 3.123 \end{bmatrix}.$$

Notice that the two matrices:

$$\mathbf{D} = \begin{bmatrix} 3 & 3 \\ 1 & 5 \end{bmatrix} \quad \mathbf{E} = \begin{bmatrix} 2 & 4 \\ 2 & 4 \end{bmatrix}$$

also satisfy the marginal constraints, but the spatial structure they exhibit is not bi-proportional to **A**'s spatial structure.

The interaction effects of **C** are identical to those of **A**. One may say that the interaction effect measures the preference that a migrant from region i has for region j, if one controls for the differences in the sizes of regional numbers of out-migrants and in-migrants. Migration spatial structure is therefore strongly relative to "destination preference". But if migrants select destinations that are independent of the region of origin; if the probability of selecting a particular destination is identical for all migrants, irrespective of origin, then – in that special case – spatial interaction is absent, and the flow of migrants from i to j is simply the product of total out-migrants from i times the probability of selecting destination j (which is then identical for all origins of i). The model, in which the distribution of migrants over destinations is the same irrespective of origin, is known as the migrant pool model. The migrant pool model implies the independence of origins and destinations.

The contrast between our definition, with its full specification of the spatial interaction model (9 parameters for a 2 by 2 flow matrix) and various other definitions, is that only with such a detailed specification can one reconstruct the exact flow matrix from the knowledge of its parameters. No other definition proposed thus far does that. The analogy to fully-specified model schedules in demography comes to mind. For example, the Heligman-Pollard model mortality schedule, with its 8 parameters, can be used to quite accurately reconstruct an entire schedule of age-specific probabilities of dying (Heligman and Pollard 1980). And the 13-parameter multi-exponential model migration schedule can approximate an entire schedule of age-specific probabilities of migrating (Rogers and Castro 1980). Similarly, the 9 parameter saturated log-linear model of the above matrix **A** can reconstruct, in this case exactly, the elements of that matrix.

The migrations predicted by the unsaturated log-linear model may also be obtained by the bi-proportional method; it suffices to replace the interaction term $F(i,j)$ by a matrix of ones ($F(i,j) = 1$). The bi-proportional method is also known as the method of offsets and may be expressed as follows:

$$\hat{M}(i,j) = M^O(i,j)KP(i)Q(j), \qquad (2)$$

where $M^O(i,j)$ is the matrix of offsets or, for example, an earlier historical matrix.

To illustrate the workings of the method of offsets, consider the saturated log-linear model fit of the observed 1985–1990 migration flow matrix in Table 1A. Suppose we want to keep the numerical values of the row and column marginal totals, but replace the migration interaction effects observed during that period by the interaction effects observed during the earlier 1975–1980 period using the method of offsets. What would be the corresponding set of log-linear parameters? Table 3A sets out the "predicted" flow matrix obtained by the method of offsets, and Table 3B presents the associated saturated log-linear model parameters. Note that the $F(i,j)$ of the "predicted" matrix are identical to those found for the observed 1975–1980 flow matrix, but that the other terms $(K, P(i), Q(j))$ are different and reflect the changed conditions of the 1985–1990 period.

The parameters of the model in Equation (2) are estimated in two steps. First, the 1985–1990 migration flow is predicted by bi-proportionally adjusting the 1975–1980 migration matrix to the row and column totals (total out-migrants and total in-migrants) of the 1985–1990 migration matrix. Second, the main

Table 3. Predicted U.S. native-born 1985–1990 interregional migration flows with observed 1975–1980 as the offset in a log-linear model

Period	Origin region	Northeast	Midwest	South	West	Total
		Destination region				
A. Predicted 1985–1990 migration flow matrix using 1975–1980 as offset						
1985–1990	Northeast	40,247,631	416,802	1,464,622	595,017	42,724,072
	Midwest	360,206	50,491,625	1,708,088	1,119,499	53,679,418
	South	790,480	1,184,882	69,615,788	1,099,371	72,690,521
	West	342,791	785,902	1,196,007	37,648,169	39,972,869
	Total	41,741,108	52,879,211	73,984,505	40,462,056	209,066,880
B. Saturated log-linear model parameters of the above 1985–1990 predicted data						
1975–1980	Northeast	32.9459	0.2283	0.4443	0.2992	0.8677
	Midwest	0.2374	22.2685	0.4172	0.4533	1.0777
	South	0.4348	0.4361	14.1915	0.3716	1.2912
	West	0.2940	0.4510	0.3801	19.8385	0.8282
	Total	0.6248	0.9337	1.6859	1.0169	2,253,436

effects model is specified for the ratios of the predicted migration flows and the 1975–1980 migration flows.

Finally, Table 4 presents the ratios of the two sets of parameters: (1) the ratios of the observed 1985–1990 structure to that of the observed 1975–1980 structure and (2) the ratios of the predicted 1985–1990 structure using the 1975–1980 offset to the observed 1975–1980 structure. The ratios conveniently indicate the direction of change over the decade: a ratio greater than unity indicates an increased value for the parameter; one less than unity points to a decrease.

2.5 The independence model with no interaction effects

The saturated log-linear model defined in Equation (1) has several reduced, so-called unsaturated forms, the most common of which is the case of no interaction effects. In such instances $F(i,j) = 1$ in Equation (1), resulting in the independence model:

$$M(i,j) = KP(i)Q(j) \tag{3}$$

The interregional flows in such a model depend only on origin (row) and destination (column) effects.

What is important to understand about migration flow tables in general, and is illustrated by the independence model, is the importance of the diagonals in the flow matrix (representing stayers and return migrants). The independence model of Equation (3) does not predict the data well, because the influences of the diagonal elements (and of interaction effects in general) are eliminated from the model. One way to improve the predictive capability of the independence model, while still maintaining its useful properties of consistent marginal totals, is to remove the diagonal elements, that is, to omit intra-regional migration by replacing

Table 4. Ratios of saturated log-linear parameters: Observed and predicted U.S. native-born 1985–1990 interregional migration flows to observed 1975–1980

Origin Region	Destination region Northeast	Midwest	South	West	Total
A. Observed 1985–1990 / Observed 1975–1980					
Northeast	1.0417	0.9141	1.1615	0.9042	0.8589
Midwest	0.9830	1.1022	0.9923	0.9302	0.9427
South	0.9391	1.0484	0.9455	1.0743	1.1135
West	1.0399	0.9468	0.9177	1.1068	1.1092
Total	1.1337	0.9721	1.0062	0.9018	1.0190
B. Predicted 1985–1990* / Observed 1975–1980					
Northeast	1.0000	1.0000	1.0000	1.0000	0.8893
Midwest	1.0000	1.0000	1.0000	1.0000	0.9455
South	1.0000	1.0000	1.0000	1.0000	1.0554
West	1.0000	1.0000	1.0000	1.0000	1.1269
Total	1.0916	1.0189	0.9617	0.9349	1.0644

* Predicted 1985–1990 represents the 1985–1990 main effects model with 1975–1980 as offset. The unities are there because the interaction values of the 1975–1980 flow table are also the interaction values of the predicted 1985–1990 flow table.

the diagonal elements with structural zeros. The resultant independence model predicts interregional migration flows under the condition that origin and destination are independent, and that intra-regional migrations are omitted from the data (conditional independence model). An alternative approach is to associate one or more parameters with the set of diagonal elements. In that case the size differences of the diagonal elements (intra-regional flows) may be captured by the added parameters. When the diagonal elements are replaced in the model by structural zeros the resulting predicted values under the assumption of independence are much improved. This is illustrated in Table 5, which yields R^2 values of 0.91 and 0.89 for the 1975–1980 and 1985–1990 data sets, respectively. Note that the marginal totals in Table 5A are equal to the marginal totals in Table 1, minus the diagonal elements.

3 An alternative description of origin emissiveness and destination attractiveness

Demographers have, on a number of occasions, put forward stylised standard schedules as referents against which to compare observed schedules. For example, the Coale and Trussell (1974) model of fertility incorporates two such prototype schedules: a nineteenth century Swedish age pattern of first marriage, as well as a fertility schedule representative of a society that does not practice birth

Table 5. U.S. native-born expected interregional migration flows under independence and with structural zeros: 1975–1980 and 1985–1990

Period	Origin region	Destination region Northeast	Midwest	South	West	Total
A. Migration flows						
1985–1990	Northeast	0	535,839	1,349,561	576,353	2,461,753
	Midwest	442,768	0	1,793,640	766,005	3,002,412
	South	720,681	1,159,163	0	1,246,806	3,126,651
	West	315,340	507,201	1,277,434	0	2,099,975
	Total	1,478,789	2,202,204	4,420,634	2,589,164	10,690,791
1975–1980	Northeast	0	604,551	1,381,715	727,187	2,713,452
	Midwest	434,027	0	1,866,435	982,292	3,282,753
	South	530,016	997,240	0	1,199,536	2,726,792
	West	270,388	508,742	1,162,741	0	1,941,870
	Total	1,234,431	2,110,532	4,410,890	2,909,014	10,664,867
B. Log-linear parameters						
1985–1990	Northeast	0.0000	1.0000	1.0000	1.0000	0.7786
	Midwest	1.0000	0.0000	1.0000	1.0000	1.0348
	South	1.0000	1.0000	0.0000	1.0000	1.6843
	West	1.0000	1.0000	1.0000	0.0000	0.7370
	Total	0.5458	0.8778	2.2108	0.9442	784,038
1975–1980	Northeast	0.0000	1.0000	1.0000	1.0000	0.8546
	Midwest	1.0000	0.0000	1.0000	1.0000	1.1544
	South	1.0000	1.0000	0.0000	1.0000	1.4096
	West	1.0000	1.0000	1.0000	0.0000	0.7191
	Total	0.4834	0.9096	2.0788	1.0941	777,788

control. Another example is Brass's (1971) well-known standard life table set of conditional probabilities of dying.

A useful prototype in the area of migration is the neutral migration process proposed by Liaw (1990). For every observed migration flow matrix one can define a corresponding flow matrix arising from a theoretical process that satisfies the following three conditions. First, the destination choices of out-migrants from any region do not affect the relative population distribution across other regions. Second, the total out-migration rate from any region does not alter the initial population distribution between the origin region and the rest of the system. Third, the overall migration level of the neutral migration system is the same as that of the observed migration process.

In sum, the neutral migration process is defined as one that maintains the observed overall level of migration, while ensuring the preservation of the observed geography of the population stocks. The necessary set of calculations to derive the neutral migration matrix that is associated with a particular observed matrix are set out in Liaw (1990) and need not be listed here. Our interest lies in the log-linear expressions of such matrices. Tables 8A and 8B set out the

Table 6. U.S. native-born neutral interregional migration flows: 1985–1990 and 1975–1980

	Period	Origin region	Destination region Northeast	Midwest	South	West	Total
A.	1985–1990	Northeast	40,358,666	763,325	1,033,664	568,417	42,724,072
		Midwest	763,325	50,903,206	1,298,717	714,171	53,679,419
		South	1,033,664	1,298,717	69,391,039	967,102	72,690,522
		West	568,417	714,171	967,102	37,723,179	39,972,868
		Total	42,724,072	53,679,419	72,690,522	39,972,868	209,066,881
B.	1975–1980	Northeast	39,244,885	819,146	1,047,581	553,876	41,665,488
		Midwest	819,146	49,687,417	1,320,629	698,242	52,525,434
		South	1,047,581	1,320,629	63,912,039	892,960	67,173,209
		West	553,876	698,242	892,960	33,370,680	35,515,758
		Total	41,665,488	52,525,434	67,173,209	35,515,758	196,879,889

Table 7. Saturated log-linear model parameters of U.S. native-born neutral interregional migration flows: 1985–1990 and 1975–1980

	Period	Origin region	Destination region Northeast	Midwest	South	West	Row effect
A.	1985–1990	Northeast	22.3008	0.3551	0.3824	0.3302	0.8774
		Midwest	0.3551	19.9338	0.4045	0.3493	1.0422
		South	0.3824	0.4045	17.1870	0.3762	1.3105
		West	0.3302	0.3493	0.3762	23.0443	0.8345
		Column effect	0.8774	1.0422	1.3105	0.8345	2,350,878
B.	1975–1980	Northeast	20.9493	0.3671	0.3899	0.3335	0.9014
		Midwest	0.3671	18.6985	0.4127	0.3530	1.0736
		South	0.3899	0.4127	16.5825	0.3748	1.2930
		West	0.3335	0.3530	0.3748	22.6629	0.7992
		Column effect	0.9014	1.0736	1.2930	0.7992	2,305,465

neutral migration counterparts to the observed matrices presented earlier in Tables 1A and 1B, and Tables 9A and 9B set out the associated parameters for the corresponding saturated log-linear model fits to these flow data.

The hypothetical neutral migration structures set out in Tables 8A and 8B, respectively, have analytic power because they can be compared with the observed structures to reveal the sources of population redistribution. And it turns out that the log-linear decomposition of the neutral migration structure confirms the definitional properties of neutral migration (in Table 7). The measures of each region as an origin and a destination are equal; $P(i)$ and $Q(j)$ are equal. Moreover, the matrix of interaction parameters is symmetric. This means that relative attraction from i to j is the same as from j to i. In other words, the accessibility of the Northeast to the Midwest is equal to accessibility of the Midwest to the Northeast. The log-linear decomposition clearly describes the neutral migration structure. The symmetry is pronounced both in the log-linear parameters (Table 7) and in the corresponding matrix of neutral migration flows (Table 6). The latter property gives rise to a migration efficiency index of zero for every region and measures of spatial focus (e.g., Coefficients of Variation), which are the same for both in- and out-migration streams.

Liaw and Rogers (1999) show that a comparison between an observed migration flow matrix and the corresponding neutral migration flow matrix (Liaw 1990) can illuminate a number of important attributes of the observed migration process, among which are the notions of origin emissiveness and destination attractiveness. Regions with observed departure propensities exceeding those of the associated neutral migration regime are said to exhibit strong emissiveness: a relatively weak ability to retain their own residents. And regions experiencing observed destination choice proportions exceeding those of the associated neutral migration regime are said to exhibit strong attractiveness: a relatively enhanced ability to acquire the residents of other regions.

The observed and associated neutral migration matrices of migration flows during the 1985–1990 census period for native-born persons residing in each of the four Census regions have been set earlier in Tables 1A and 8A, respectively, together with the corresponding 1985 regional populations at risk of moving. From that data we are able to establish that 2,461,752 / 42,724,072 = 0.058 (5.8%) of the Northeast’s population in 1985 was residing in one of the three other regions of the country in 1990, with 1,645,843 / 2,461,753 = 0.669 (66.9%) of that total living in the South. The corresponding fractions under the neutral migration regime according to Table 6A would be 0.055 and 0.437, respectively. Calculating these observed-to-neutral fractions, we find that the first, the emissiveness ratio, is 1.04, while the second, the attractiveness ratio, is 1.53. By comparison, the corresponding ratios for the West are 0.93 and 1.20, respectively.

The off-diagonal ratios in Table 8A represent the ratios of *ij* accessibilities in the observed system to the same accessibilities in the neutral system, or the off-diagonal elements in Table 1 divided by the off-diagonal elements in Table 6. These could be used as alternatives to the attractiveness ratios set out in Liaw-Rogers (1999), for example. The ratios of the diagonals in Table 8A are close to the value of one, and represent the slight changes between stayers in the observed and neutral migration flow matrices. Notice also that the ratios for the row marginal totals are equal to one (by design), and ratios for the column marginal totals equal the amount of growth (or decline) each region experienced during the time interval.

The measures of emissiveness and attractiveness proposed in the Liaw-Rogers (1999) article are simple to calculate, transparent in their meaning and surprisingly effective in capturing the retention and preference attributes underlying an observed migration flow pattern. The log-linear reformulation developed here is more complicated and somewhat less transparent, but it highlights several additional dimensions ignored by the simpler formulation. Both have their appropriate contexts.

The log-linear decomposition of the neutral migration structure demonstrates the analytic power of the log-linear model. The log-linear modeling perspective suggests additional interesting properties of the neutral migration process. The

Table 8. Ratios of U.S. native-born observed to neutral interregional migration flows: 1985–1990 and 1975–1980

	Period	Origin region	Destination region Northeast	Midwest	South	West	Total
A. Ratios of migration flows							
A.1	1985–1990	Northeast	0.9976	0.4403	1.5922	0.8441	1.0000
		Midwest	0.4599	0.9956	1.3034	1.3424	1.0000
		South	0.7535	0.9218	1.0025	1.1898	1.0000
		West	0.6138	0.9367	1.1189	1.0040	1.0000
		Total	0.9770	0.9851	1.0178	1.0122	1.0000
A.2	1975–1980	Northeast	0.9925	0.5276	1.5359	1.2139	1.0000
		Midwest	0.4003	0.9910	1.3365	1.7041	1.0000
		South	0.6154	0.7839	1.0084	1.1723	1.0000
		West	0.4727	0.9210	1.1612	1.0061	1.0000
		Total	0.9645	0.9777	1.0251	1.0272	1.0000
B. Ratios of saturated log-linear parameters							
B.1	1985–1990	Northeast	1.5389	0.5878	1.3496	0.8194	0.9552
		Midwest	0.6573	1.2312	1.0236	1.2073	1.0309
		South	1.0680	1.1304	0.7807	1.0609	1.0396
		West	0.9257	1.2224	0.9272	0.9528	0.9769
		Total	0.7396	0.8547	1.3460	1.1754	0.9177
B.2	1975–1980	Northeast	1.5726	0.6218	1.1396	0.8972	1.0824
		Midwest	0.6466	1.1909	1.0110	1.2842	1.0617
		South	1.1153	1.0568	0.8558	0.9914	0.9463
		West	0.8816	1.2777	1.0140	0.8754	0.9196
		Total	0.6349	0.8536	1.3559	1.3609	0.9183

neutral migration process implies a symmetric migration matrix, and as a result, row totals equal to column totals. The log-linear model of such a matrix embodies two properties: symmetry and marginal homogeneity (Agresti 1996, p. 234). The neutral migration process therefore implies a particular type of log-linear model. A further observation is that a division of the cell counts by the overall total number of migrations yields a doubly-stochastic matrix.

4 Conclusion

What do we mean when we refer to the spatial structure of migration? This expression has been used rather loosely in the literature and needs to be defined more rigorously in order to be useful as a tool for comparative analysis of flows – or in the absence of flow data, for developing indirect methods of estimating migration streams. One way to define migration spatial structure is to draw from the demographer's method of defining age structure: as the proportional distribution of the numbers of persons enumerated at each age or in each age group.

Thus, if one were to double the total population but leave the proportional distribution unchanged, one could conclude that the population increased, but that its age structure has remained unchanged. By adapting this definition for the migration structure of a region's destination-specific out-migration streams, one could define that structure to represent the proportional distribution of the total outflow across the set of alternative destinations. In that case if a doubling of the region's out-migration level were distributed in the same proportional manner over age groups and destinations, then one would conclude that the migration spatial structure had remained the same as before. But that definition only makes sense in a linear model of the phenomenon. If instead one adopts a non-linear gravity model type of formulation – a spatial interaction model representation of origin-destination-specific migration flows for example – then clearly one also needs to consider the change of the destination population and also the separation effect between each origin-destination pair of locations. The impact on spatial structure of a doubling of the migration outflow therefore needs to be considered in tandem with the change in the destination population size or size of migration inflow. For example, the impact of a tripling would be different than that of a doubling. What this implies then is that a full specification of the spatial interaction model needs to be used in the definition of migration spatial structure. But if this is true, how do we interpret the use of a historical matrix to predict current migration? We interpret it as the migration spatial structure we wish to impose on a current set of marginal totals. The row and column parameters are different for the two matrices, but the row-column interaction effects remain the same.

The various indicators of migration spatial patterns that have been popular in the literature only describe particular attributes of a particular migration spatial structure: for example, its efficiency in redistributing the multi-regional population, or its spatial focus, or, indeed, its implicit destination preferences. None of these could be used to impose a unique historical migration spatial structure onto a current situation. They allow only a partial assessment of comparative structures, and are therefore of limited use as tools of indirect estimation. But as partial indicators of different attributes of spatial patterns they can and have played a useful role in comparative studies of such patterns. The relevant literature is rich with examples of the useful findings generated by indices of migration efficiency and of spatial focus, for example. The much more recent indices outlined in Liaw and Rogers (1999) have no such history as yet, so we elaborate them further here.

Recall the assessments of relative attraction that appear as the off-diagonal elements in Table 8A. For both periods studied the observed ending populations in the Northeast and Midwest were less than the neutral regime's ending populations (which, by design, were the same as the beginning populations), whereas in the South and West, the opposite was true. However, the ratios of the observed to neutral ending populations were slightly less for the 1985–1990 period than for the 1975–1980 period, implying that less growth or decline occurred in all regions. Other patterns in the ratios of observed migration flows to neutral migration flows revealed that, in general, migrants were less likely to go to the

Northeast and Midwest and more likely to go to the South and West. The only exception to this pattern was that migrants from the Northeast were less likely to go to the West during the 1985–1990 period. Finally, the ratios in Table 8A indicate significant increases in the ratios for the migration flows to the Northeast and Midwest, along with significant declines to the South (except from the Northeast) and West (except from the South).

The ratios of emissiveness and attractiveness, defined in Liaw and Rogers (1999) rely on the observed system without the log-linear decomposition. We argue that these concepts might be better understood and measured when augmented by ratios of log-linear decompositions (Tables 2–7), as set out in Table 8B. Notice that the ratios of the total observed-to-neutral geometric means are less than one for both time periods. The intercept for the log-linear models in this article represents the geometric mean of the data and gives a rough indication of the variability in the migration flows.

Consider next the parameter ratios of the row and column marginal totals. The row marginal totals of the neutral migration process are the same as in the observed migration process (by design); but the migration flows differ, giving rise to the slight differences found in those ratios. The parameter ratios for the column marginal totals differ even more. The row and column marginal total ratios provide information on comparisons between the relative origin and destination effects between the two migration flow tables. For example, a ratio of 0.74 for the column marginal total of the Northeast during 1985–1990 implies that, as a receiving region, the Northeast was much less important than in the neutral migration table, but as a sender of migrants it was only slightly less important (0.96).

Finally, consider the ratios of the matrix of interaction effects. The interaction parameters in Table 8B show that the diagonals of the observed migration flow matrix are significantly different from the diagonals in the neutral migration flow matrix, whereas in Table 8A they are nearly identical. The reason they differ so much is that the sizes of the sending and receiving regions are now taken into account. The ratios of the interaction parameters of the migration flows also give useful insights to the neutral migration process comparison. Here one finds similar patterns as in Table 8A (high ratio values for the Northeast to South flow), but with some important differences (the South to Northeast migration flow). In Table 8A the ratio of the South to Northeast migration flow is less than one, implying a relative unattractiveness of that destination. However, in Table 8B, the ratio is higher than one, implying relative attractiveness. But remember that size is now taken into account. This is important because larger regions (e.g., the South) generally send out more migrants. The log-linear interaction parameters for this observed flow (Table 2) now shows that the South to Northeast migration flow is relatively strong compared to the other migration flows (0.43 in 1975–1980 and 0.41 in 1985–1990). In fact, the value of log-linear interaction parameter for South to Northeast migration flow is almost as large as the corresponding parameter for the Northeast to South migration flow. And since the log-linear interaction parameters for both flows are the same and lower (0.39 in 1975–1980 and 0.38 in 1985–1990) for the neutral migration process (Table 7), we naturally obtain ratio values larger than unity in Table 8B.

For reasons that are not self-evident, the ratio of attractiveness defined in Liaw and Rogers (1999) and its log-linear counterpart provide somewhat different assessments of locational preferences. We speculate that the log-linear version introduces the influence of the separation (or interaction) effect more fully, and it also seems to bring in the relative population size effects more directly. As a result, the simple coefficients of determination (R^2) of the two sets of index values for our 1975–1980 and 1985–1990 data sets are 0.40 and 0.63, respectively. The clear advantage of the ratio of attractiveness is of course its relative simplicity and transparency. The advantage of our log-linear version is that it perhaps more accurately reflects the notion of relative attractiveness. But the choice of a measure of relative attractiveness is a choice among alternative definitions of attractiveness rather than a choice among alternative ways of measuring a single construct (Allison 1978).

References

Agresti A (1996) *An Introduction to categorical data analysis*. Wiley, New York

Allison PD (1978) Measures of inequality. *American Sociological Review* 43(6): 865–881

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

Aufhauser E, Fischer MM (1985) Log-linear modelling and spatial analysis.*Environment and Planning A* 17(7): 931–951

Bacharach M (1970) *Biproportional matrices and input-output change*. Cambridge University Press, London

Bennett RJ, Haining RP (1985) Spatial structure and spatial interaction models: Modelling approaches to the statistical analysis of geographical data. *Journal of the Royal Statistical Society A* 148: 1–27

Birch MW (1963) Maximum likelihood in three-way contingency tables. *Journal of the Royal Statistical Society B* 25(1): 220–233

Brass W (1971) On the scale of mortality. In: Brass W (ed) *Biological aspects of demography*. Taylor and Francis, London

Clayton C (1977) The structure of interstate and interregional migration: 1965–1970. *Annals of Regional Science* 11(1): 109–122

Coale AJ, Trussell J (1974) Model fertility schedules: Variations in the age structures of childbearing in human populations. *Population Index* 40(2): 185–206

Haynes KE, Fotheringham AS (1984) *Gravity and spatial interaction models*. Beverly Hills, Sage

Heligman L, Pollard JH (1980) The age pattern of mortality. *Journal of the Institute of Actuaries* (Oxford), Part I, 107(434): 49–80

Knudsen DC (1992) Generalizing Poisson regression: Including a priori information using the method of offsets. *Professional Geographer* 44(2): 202–208

Liaw KL (1990) Neutral migration process and its application to an analysis of Canadian migration data. *Environment and Planning A* 22(3): 333–343

Liaw KL, Rogers A (1999) The neutral migration process, redistributional potential, and Shryock's preference indices. *The Journal of Population Studies* (published by the Population Association of Japan) 25(12): 3–14

Lin G (1999) Assessing structural change in U.S. migration patterns: A log-rate modeling approach. *Mathematical Population Studies* 7(3): 217–237

Manson GA, Groop RE (1996) Ebbs and flows in recent U.S. interstate migration. *Professional Geographer* 48(2): 156–166

Mueser P (1989) The spatial structure of migration: An analysis of flows between states in the USA over three decades. *Regional Studies* 23(3): 185–200

Nair PS (1985) Estimation of period-specific gross migration flows from limited data. *Demography* 22: 133–142

Payne C (1977) The log-linear model for contingency tables. In: Muircheartaigh CO, Payne C (eds) *The Analysis of survey data, Vol. 2: Model fitting*. Wiley, New York

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

Plane DA, Mulligan GF (1997) Measuring spatial focusing in a migration system. *Demography* 34: 251–262

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

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

Scholten H, van Wissen L (1985) A comparison of the log-linear interaction model with other spatial interaction models. In: Nijkamp P, Leitner H, Wrigley N (eds) *Measuring the unmeasurable*. Martinus Nijhoff Publishers, Dordrecht

Shryock HS (1964) *Population mobility within the United States*. Community and Family Study Center, University of Chicago, Chicago

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

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

Willekens F (1982a) Multidimensional population analysis with incomplete data. In: Land K, Rogers A (eds) *Multidimensional mathematical demography*. Academic Press, New York

Willekens F (1982b) Specification and calibration of spatial interaction models: A contingency table perspective and an application to intra-urban migration in Rotterdam. *Journal of Economic and Social Geography* 74: 239–252

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