PREDICTING INDIVIDUAL MIGRATION BEHAVIOR*

Wen Lang Li, Ohio State University

Migration prediction is an important and well-attended topic. The literature in migration studies is overwhelmed by this type of research. But, despite their diverse formulations, the basic of migration prediction can be summarized in the so-called "gravity model" (Shaw, 1975, Ch. 3). It was proposed by Zipf (1946) as follows:

$$M_{ij} = \frac{P_i \; P_j}{D_{ij}} K$$

where M_{ij} = net or gross migration between areas i and j

P_i = population in area i

D_{ij} = distance between areas i and j

K = constant term.

This type of prediction model is important because it attempts to estimate the volume of population flows among regions, which are indispensable for the projection of regions' populations. But, this type of prediction method is inadequate because it assumes that we have already known the regions' populations, P_i and P_j. In fact, the regions' population should be predicted, not as predictive variables. Demographers have come to realize that perhaps our understanding of migration behavior is not proper enough for predicting regional populations (Bogue, 1959).

There are some other models which attempt to "explain" migration behavior. The major difference of these models and gravity models is that their focus is on migration behavior, instead of the amount of migration. A migration rate is constructed based on whatever direction of movement: immigration, outmigration, or net migration. The units of analysis are mostly geographic: counties, states, or census tracts. Some socioeconomic characteristics of the areas are also collected. Then a multivariate analysis is performed, which generally is a linear regression, to derive the most acceptable prediction model of migration behavior.

Because of the units of analysis are geographic, the observations derived from the analysis have to be confined in the regional context. The results of migration research are hardly generalizable. Consequently, the majority of migrations is meaningful only in the historical sense. The traditional "push-pull" models have yielded little predictive results except to depict the regional socioeconomic correlates of migration. It would be truly an "ecological fallacy" if these observations are translated into migration predictions (Robinson, 1950).

Although it has been acclaimed as one of the meaningful explanations of migration behavior (Herrick, 1965), the push-pull model is at most a relabelling of the migration process. That people moves itself implies a departure from the area of distress to the areas of attraction. Numerous amounts of studies have endeavored to delineate the complexity of the "pull" and "push" forces, but the model is not able to take into account the behavioral components of the migrants. As Wolpert (1965:161) pointed out:

> Attempts at model building in migration research have largely focused on variables and surrogates such as distance and ecological characteristics of places exerting 'push' and 'pull' forces, to the exclusion of behavioral parameters of the migrants.

Migration as an Individual Behavior

Migration is basically an individual behavior. It is the individual who decides whether to move or to stay. Even in an area of natural disaster many people choose to stay. The importance of considering the decision mechanisms in the migration process is clear. Perhaps the precise mechanisms involved in the process varies from case to case. But, it is possible that certain regularities can be detected and generalized.

Cost-benefit model was proposed as an approach to study the migration decision process (Sjaastad, 1962). It attempts to calculate the monetary and employment payoffs as affecting migration behavior. Nevertheless, the classic utility concepts may not be useful in migration explanations. Many non-monetary forces are important and yet difficult to be directly measured. Obviously, a person's decision to migrate is dictated by a variety of socioeconomic constraints, for example, his stage of life cycle, his employment status, or his social contacts with other communities. Before a cost-benefit model can be fully developed, it is perhaps more proper that the study of migration behavior should begin with the theories of migration differentials.

Unfortunately, the theories of migration differentials are relatively underdeveloped. As aforementioned, model constructions in migration studies are predominantly in aggregate level. One of their distinctive features is to treat a population as homogeneous. The differential aspects of interregional population movement are not well attended in previous research (Li, 1970). The census data from which most generalizations of migration differentials are drawn do not have much behavioral measurements. The aggregate data make it impossible to extend further the theories of migration differentials. Except some scanty attempts (Beshers and Nishiura, 1961; Lee, 1966), many researchers seem to contend that a search for universal generalizations would be fruitless. Bogue (1959) concludes as follows:

> A little reflection convinces one that the search for universal migration differentials not only is doomed to failure but also fails to appreciate the reasons for migration selectivity. (Bogue, 1959: 504).

* This paper has been written with the partial support of the National Institute for Child Health and Human Development, HD-07094-01.

Such pessimistic contention may not only be due to the nature of migration data. The analytical methods employed to study migration differentials may also be a contributing factor. They are generally so primitive that an extensive study of the migration propensity is impeded. Multivariate analysis techniques have been used by previous researchers (Hamilton, 1959; Bogue, et. al., 1953; Tarver, 1961). In most cases the method is multiple linear regression which requires stringent assumptions and particular units of analysis. Yet migration is basically a binary variable: to move or not to move. Although the analytical techniques for this type of problems have been proposed in 1930's (Fisher, 1937), it is only in recent years that the method has received extensive attention (Cox, 1970).

The purpose of this paper is to introduce a multivariate analysis technique to migration studies. The technique is useful in analyzing migration behavior in the micro (individual) level. Its application will predict individual as well as aggregate propensity to migrate. The subsequent portion of this paper is divided into three parts: first, the structure of the model; next, the nature of the data and selected predictive variables; and finally, the research findings and conclusions.

The Model

The first step of the model is to use the linear discriminant function as proposed by R. A. Fisher (1936). We are given two groups of persons: migrants and nonmigrants. Every person is measured by k numbers of socioeconomic characteristics, or variables. It is assumed that, in the population from which the groups are drawn, the characteristics have a common multivariate normal distribution. The discriminant coefficients (which is analogous to regression coefficients) can be computed by:

$$b = V^{-1} d \tag{2}$$

where V^{-1} is the inverse of the pooled covariance matrix and d the vector of differences between the pairs of means of the two groups. The constant term of the discriminant function is estimated by:

$$a = -\tfrac{1}{2}\sum_{i}^{k} b_i \, (\bar{x}_{i0} + \bar{x}_{i1}) - \ln(n_0/n_1) \tag{3}$$

where 0 and 1 denote respectively migrant and non-migrant groups; $\bar{x}_i$ the ith variable's mean; and n the number of persons in each group. In other words, the constant term is a function of discriminant coefficients, the average values of the variables, and the proportion of migrants.

Based on the assumptions of homoscedasticity and multivariate normality, it is possible to show that a multiple logistic function can be derived from the linear discriminant function (Cornfield, 1967). And the predicted outcome corresponds to the probability of migration, such as:

$$p = 1/(1+ e^{-(a+\sum b_i x_i)}) \tag{4}$$

where e is the base of natural logarithms. The multiple logistic function can be used for screening highly mobile individuals in the general population.

But the assumptions involved in this formulation are too stringent. Migrants' or nonmigrants' characteristics are rarely normally distributed, nor their covariance matrices are identical. To avoid such restrictions Walker and Duncan (1967) suggest a maximum likelihood approach. For a group of n individuals, the likelihood function of migration is:

$$L = \prod_{j}^{n} p_j^{\,y_j} \, (1-p_j)^{1-y_j} \tag{5}$$

where y_j equals to one or zero depending on whether the jth individual is a migrant or not; and $\prod$ is the product sign. The function is to be maximized. Newton-Ralphson procedure is chosen. The coefficients from the linear discriminant function are used as initial values. Through successive approximation some refined estimates of the coefficients are obtained. Tests on the goodness of fit, such as maximum likelihood ratio, can determine whether the model outcomes are agreeable with the empirical data.

Data and Variables

The public use sample data of the 1970 census are used to test the model. From one-in-10,000 U.S. population, individual's information on socioeconomic activities in both 1970 and five years earlier in 1965 was collected. A question was asked about the place of residence in 1965. Together, these data make it possible to identify the migration and socioeconomic characteristics of every individual. Those whose states of residence differ between 1965 and 1970 are grouped as migrants; otherwise, nonmigrants. Of the total 20,196 persons in the sample, our analysis includes only the household's chief income recipients. Although they are not strictly the household heads, the chief income recipients are perhaps the major decision-makers of migration behavior. Other household members are excluded. Thus, the sample size for this study is reduced to 7,124 persons; roughly one person for one household in the U.S.

Three predictive variables of migration behavior are selected as predicated by the nature of the data. These variables were also shown to be highly related to migration in previous studies. The first variables is age, which is a clear determinant of migration behavior. Age is almost a perfect indicator for the stage of life cycle. It is also closely related to the duration of residence. Both concepts have been used as the predominant explanations for why people move (Rossi, 1955; Morrison, 1967). Thomas' classic report on migration differentials shows a definitively established generalization: young adults are more mobile than older persons (Thomas, 1938). As presented in Figure 1 our data clearly confirm this generalization. Migration behavior is indeed a decreasing function of the aging process.

Years of schooling are used as the second predictive variable of migration behavior. Many previous studies have observed that migration is positively related to education. The number of years one spends in schooling undoubtedly expand his information contacts with outside world beyond

Figure 1: Migration Rate by Age

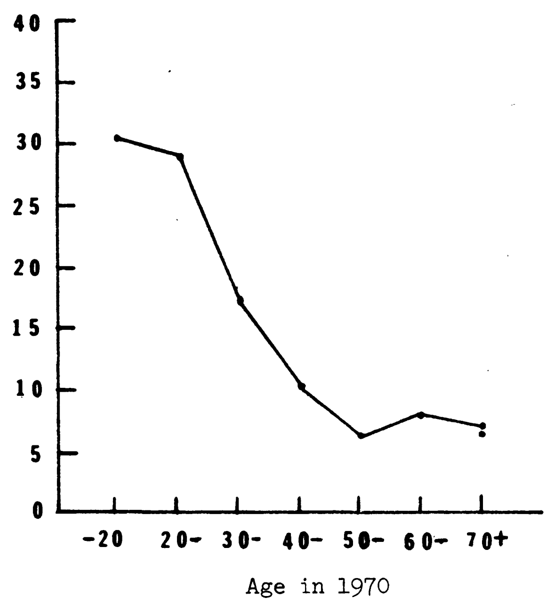
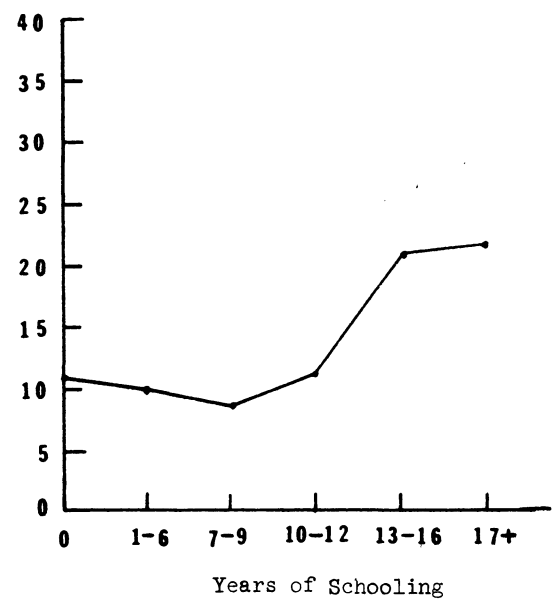


Figure 2: Migration Rate by Education



his locality. Wolpert (1965) has called our attention to the importance of "information fields" in regulating people's search behavior. In addition, Bogue et. al. (1957) conclude that the two factors that seem to contribute most to the mobility of the population are above average educational training and employment in white collar occupations. Again, our data clearly support this observation. Figure 2 shows that migration behavior is determined by the number of years of schooling.

The last predictive variable is employment status before migration occurs. The 1970 census provides the possibility of testing this relationship. A question was asked to every sample individual about his economic activity in 1965. As shown in Table 1, the results from the census data strongly suggest that migration and employment status are related as observed in many previous studies. Both Lansing and Mueller (1967) and Li (1976) have found that unemployed persons are much more likely to migrate than are employed persons, and that persons in schooling or in military services are perhaps the most mobile of all.

Empirical Results

The three predictive variables of migration behavior are denoted respectively by x_1, x_2, and x_3. Both age and educational attainment use

Table 1: Household Chiefs by Economic Activity and Migration Status: U. S., 1970

Economic Activity	Migrants	Nonmigrants	Percent Migrant
Employed	401	4,146	8.8
Unemployed	412	1,636	20.1
Services & Schooling	220	309	41.6
Total	1,033	7,124	35.2

straightforward measurements. Employment status is measured by: 0 for the employed, 1 for the unemployed, and 2 for being in colleges or military services. Premilinary results of the linear discriminant function, as described in Equation (2) and (3), are shown in Table 2. The constant term has a value of -1.691, which serves as a reference point for the comparison of discriminant scores among the sample individuals. The discriminant coefficients for the three predictive variables are -0.031, 0.046, and 1.176 respectively. The signs of the coefficients are consistent with what have been observed in the literature. Following the procedure described in Equation (5), a more refined version of the model is obtained. Compared to the results of the preliminary version, the values of the constant term and the discriminant coefficients decrease slightly The value of a is -1.436, whereas the coefficients for x_1, x_2, and x_3 respectively are -0.029, 0.038, and 0.803.

To assess the statistical significance of the discriminant coefficients, the following formulation is used to compute the standard errors of the estimates:

$$s = w \cdot m \tag{6}$$

where w is the diagonal vector of V^{-1}, and m is $(1/n_0+1/n_1)$. The results of s vector are also shown in Table 2. Then t-scores are computed and used to determine the significance of the coefficients. It is noted that the three predictive variables are all statistically significant at 5 percent level. In other words, the results are quite consistent with the theories which were previously presented.

Table 2: Results of Model Implementation

Variable	b Coefficients	Standard error	t Score
Preliminary Version			
x_1	-0.031	0.002	-15.4
x_2	0.046	0.010	4.6
x_3	1.176	0.055	21.2
Constant a	-1.691		
Refined Version			
x_1	-0.029	0.002	-12.8
x_2	0.038	0.011	3.4
x_3	0.803	0.050	16.0
Constant a	-1.436		

The refined version of the model can be used to predict an individual's migration probability. Assume that a person is 25 years of age, college-graduated and employed, the model predicts that his migration probability would be 0.173, or 17.5 percent. If he happens to be unemployed, then the probability increases drastically to 31.9 percent. And, if he is in military services, his chance to move across state boundries is 48.9 percent.

Table 3 presents the simulated migration probabilities for a person who is 25 years of age in the U. S. The probabilities are calculated with various assumptions about his educational attainment and employment status. A general pattern emerges. It is noted that an increase of his educational level will result in an increase of roughly 2 percent points in migration probability. For example, among the employed the migration probability is 0.153 for a high-school graduate and 0.173 for college graduate. The difference is exactly 0.02. On the other hand, a change of the employment status, from being employed to unemployed, will almost double his migration probability; say, for an uneducated to change from 0.103 to 0.204. The results appear to indicate that employment status weights much more than educational attainment in determining migration behavior.

Table 3: Simulated Migration Probability for a Person at Age 25.

Educational Attainment	Employment status: Employed	Employment status: Unemployed	Schooling or Services
None	0.103	0.204	0.364
6th Grade	0.126	0.243	0.417
9th Grade	0.139	0.264	0.445
12th Grade	0.153	0.287	0.473
College Graduate	0.173	0.319	0.489

The predicted probability can be used as an instrument to screen the potential migrants. Note that the overall rate of migration is 14.5 percent as estimated from our sample. In other words, about 15 out of every 100 Americans are expected to migrate across state boundaries during 5-year period. Undoubtedly, any individuals whose predicted migration probability is higher than this figure would be judged as potentially active migrants. The average adult American is not a potential active migrant. Statistically speaking, he is about 46 years of age, received 11 years of schooling, and currently employed. His predicted probability of migration is only 8.7 percent, which is far less than 14.5 percent as expected. Nevertheless, if he happens to be unemployed, his migration probability would nearly be doubled. It increases to 17.5 percent. In this case, he would be a potentially active migrant.

An important use of the model is for regional population estimation and projection. As every individual's migration behavior can be predicted, so is the total population in an area. It is a simple case of summation. If a person's migration probability is 0.3, whereas another person is 0.7, then it is expected that one out of these two persons will be a migrant. Table 4 shows that through such aggregation the model can yield fairly reliable estimation of migrants for a region. As a case of illustration, the division of East South Central has 60 migrants in the sample; our model estimates it has 62. The migration rate is almost identical between the observed and the predicted.

Table 4: Observed and Predicted Migrants and Migration Rates by Division, 1970

Division	Number of Migrants		Migration Rate (%)	
	Observed	Predicted	Observed	Predicted
New England	71	65	16.4	15.1
Middle Atlantic	141	179	10.8	13.7
East North Central	151	188	10.9	13.7
West North Central	62	78	11.0	13.9
South Atlantic	199	159	18.6	14.9
East South Central	60	62	14.1	14.5
West South Central	94	96	14.4	14.8
Mountain	80	48	26.2	15.8
Pacific	175	157	17.7	16.0
Total	1,033	1,033	14.5	14.5

It is not only that a region's amount of migrants can be estimated, but also the demographic characteristics of the migrants can be ascertained. Given enough sample size, the model can yield an estimation of migrants by age, sex, and race for each region. This type of information is urgently needed in the projection of a region's population. As shown in Table 5, migrants are estimated by age for each division. Migration rates are then computed. The predicted migration rates are compared with the observed migration rates. In many cases the goodness of fit seems quite acceptable.

Some Cautious Remarks

A micro-predictive model as proposed in this paper has its merits as well as demerits. Although we have presented some pleasing results, they are by no means overwhelmingly satisfactory. The success of a prediction model depends on many factors. The most important one is perhaps the choice of "right" predictive variables. Obviously, the choice should be based on both theoretical and methodological considerations. Kendall (1966) has suggested a so-called "ratio-statistic method," which is unfortunately not usable because of the nature of our data. What this paper relies on is strictly theoretical considerations. It is therefore suggested that more behavioral variables need to be taken into account. And more sophisticate methods of variable-selection should be explored.

Though we have attempted to aggregate individual migration probabilities into a region's migration estimation, the major thrust of this paper is to present a micro model. No serious attempt has been made to relate the behavioral prediction with locational characteristics. As Lee (1966) correctly pointed out, a proper model of migration should simultaneously consider four dimensions: characteristics of origin, of destination, of intervening obstacles, and of migrants. This paper chooses to expand upon Lee's paradign through investigating first the migrants' characteristics as determinants of the migration decision. One may argue that the decision to move is a direct reflection of the locational characteristics. However, both the micro and the macro aspects of migration are indeed difficult to be separated. It is imperative that a behavioral model as presented in this paper should be integrated with a location model.

References

Beshers, James M. and Eleanor N. Nishiura
1961 "A Theory of Internal Migration Differentials," Social Forces 39: 214-218.

Bogue, Donald J.
1959 "Internal Migration," Pp. 486-505 in P. M. Hauser and O. D. Duncan (eds.) The Study of Population. Chicago, Ill. University of Chicago Press.

Bogue, Donald J., Henry S. Shryock, and Siegfried A. Hoermann
1957 "Subregional Migration in the United States, 1935-40," Vol. 1, Streams of Migration Between Subregions. Oxford, Ohio: Scripp Foundation, Miami University.

Cox, D. R.
1970 The Analysis of Binary Data. London: Methuen.

Cornfield, Jerome
1967 "Discriminant Functions," Review of the International Statistical Institute 35 (2): 142-153.

Fisher, R. A.
1936 "The Use of Multiple Measurements in Taxonomic Problems," Annals of Eugenics 7: 179-188.

Hamilton, C. Horace
1951 "Population Pressure and Other Factors Affecting Net Rural-Urban Migration," Social Forces 30 (December): 209-215.

Herrick, Bruce H.
1965 Urban Migration and Economic Development in Chile. Cambridge, Mass.: MIT Press.

Kendall, M. G.
1966 "Discrimination and Classification," Pp. 165-185 in P. R. Krishnaiah (ed.) Multivariate Analysis, N. Y.: Academic Press.

Lee, Everett S.
1966 "A Theory of Migration," Demography 3: 47-57.

Li, Wen Lang
1970 "A Differential Approach to the Study of Migration Streams," Rural Sociology 35 (December): 534-538.

1976 "A Note on Migration and Employment," Demography 13 (November): forthcoming.

Morrison, Peter A.
1967 "Duration of Residence and Prospective Migration: The Evaluation of a Stochastic Model," Demography 4: 553-561.

Robinson, W. S.
1950 "Ecological Correlations and the Behavior of Individuals," American Sociological Review 25: 352-357.

Rossi, Peter H.
1955 Why Families Move. Glencoe, Ill.: The Free Press.

Shaw, R. Paul
1975 Migration Theory and Fact. Philadelphia, Pa.: Regional Science Research Institute. Bibliography Series, No. 5.

Sjaastad, L. A.
1962 "The Costs and Aeturns of Human Migration," Journal of Political Economy 70 (Part II): 80-93.

Taeuber, Karl E.
1964 "White Migration and Socio-Economic Differences Between Whites and Suburbs," American Sociological Review 29(October): 718-729.

Tarver, James D.
1961 "Predicting Migration," Social Forces 39(March): 207-213.

Thomas, Dorothy S.
1938 Research Memorandum on Migration Differentials. New York: Social Science Research Council.

Walker, S. H. and D. B. Duncan
1967 "Estimation of the Probability of an Event as a Function of Several Independent Variables," Biometrika, 54(1): 167-179.

Zipf, G. K.
1947 "The P_1P_2/D Hypothesis: On the Inter-City Movement of Persons," American Sociological Review 11: 677-686.