Faculty Working Papers

A RECURSIVE MODEL OF THE SPATIAL ALLOCATION OF MIGRANTS: SOME EMPIRICAL RESULTS

James B. Kau and C. F. Sirmans

#341

College of Commerce and Business Administration
University of Illinois at Urbana-Champaign

FACULTY WORKING PAPERS

College of Commerce and Business Administration

University of Illinois at Urbana-Champaign

October 7, 1976

A RECURSIVE MODEL OF THE SPATIAL ALLOCATION OF MIGRANTS: SOME EMPIRICAL RESULTS

James B. Kau and C. F. Sirmans

#341

A RECURSIVE MODEL OF THE SPATIAL ALLOCATION OF MIGRANTS: SOME EMPIRICAL RESULTS

I. Introduction

The model specification to explain the internal migration decision has been the subject of controversy in the literature.[1] This controversy has centered on the influence of past migration on current migration patterns. Following the seminal work of Nelson (1959), recent studies (Greenwood [1969]; Laber [1972]; Levy and Wadycki [1973]; Renshaw [1974]; and Langley [1974]) have dealt with this aspect of the migration model. These authors hypothesized that migration will flow toward those destinations inhabited by earlier migrants from the same origin and have attempted to account for this effect by including a migrant stock variable in a regression model. Their results lend support to the importance of information flow from previous migrants.

Laber (1972) suggests that the migrant stock variable is a partial adjustment mechanism which introduces multicolinearity and thus may not reflect the impact of information flow. Greenwood's reply (1972) asserts that multicolinearity between the migrant stock variable and the other independent variables is not a serious problem. Greenwood also states that Laber does not allow adequately

for return migration and remigration and therefore uses an approximation of net migration rather than gross migration rates. Renshaw (1974) provides insight on these issues and suggests that caution should be exercised when interpreting the migrant-stock-coefficient estimates because other forces in addition to previous migrants could be reflected by the migrant stock variable. Renshaw (1974) recommends various procedures to isolate the various forces affecting current migration which may be reflected in migrant stocks. In response Dunlevy and Gemery (1975) use a Koyck lag form of a partial adjustment model of immigration to analyze whether migrant stock is merely a proxy for a lagged adjustment process or a measure of information flow from previous migrants. Their results tentatively support the Nelson and Greenwood hypothesis, however, many of the coefficients are either insignificant or have the wrong sign.

This paper is the first to estimate a recursive model of internal migration in the United States. The data for the gross interstate migration flows for the periods 1940, 1950, 1960, and 1970 are expressed in a block recursive system of equations. This recursive model helps to clarify the controversy over the relationship between past migration and current migration flows and reduces the problems associated with multicolinearity. Section II develops a theoretical framework of the migration decision based on an integration of utility

maximization and investment behavior into a discrete decision-making process. Section III outlines the methodology of the recursive model to be estimated in this paper. The empirical model, data, and results with the relevant comparisons between recursive and non-recursive models are discussed in Section IV. Finally, Section V contains a summary and some brief concluding remarks.

II. Theoretical Framework

The theoretical framework of migration behavior in this study rests on the assumption that migrants move in order to maximize their utility subject to the constraints of income and prices at each possible destination. It is further assumed that the migration decision is an individual decision-making process based on the available information. The migration process is an attempt by the migrant to incorporate both investment opportunities, such as greater return on his human capital, and utility maximization into one objective function. This framework can be expressed in a discrete decision-making process as follows.[2]

$$(1)\qquad U = U(x_1 \ldots x_n) - \sum_{j=1}^{m} \left\{ \lambda_j [\delta_j (I_j - \sum_{k=1}^{n} P_{kj} X_{kj} - M_j)] \right\}$$

where:

$U(x_1 \ldots x_n)$ = utility achieved from the consumption goods;

n = number of consumption goods being considered;

m = number of possible destinations;

λ_j = discrete choice parameter which has a value of 1 for the selected destination and a value of 0 for all other non-selected destinations. When λ_j = 1, its function is that of a Lagrange multiplier;

δ_j = the uncertainty associated with the expected value of I_j and P_{kj} for each destination;

I_j = expected income for each destination;

P_{kj} = set of prices for each destination; and

M_j = pecuniary moving costs for each destination.

Migration takes place as a result of the desire to maximize the individual migrant's utility function.

The migration decision is thus viewed as a discrete decision since only one destination can be selected at any point in time. The constraint is binding only at the selected destination. The value of λ_j is 1 for the chosen destination and 0 for all other possible destinations in which the constraint is not binding on the objective function. This framework combines utility maximization, in the selection of the optimum mixture of consumption goods, with investment behavior, in the selection of greatest economic returns, into one theoretical model of the migration decision.

It can be seen from equation (1) that the decision to move is based on certain trade-offs. Income, costs, uncertainty, and availability of goods all enter into the migration process. The availability of various goods (X_{kj}) is reflected by the relative costs (P_{kj}) of consuming a

particular commodity. Each possible destination for the individual represents a set of characteristics indicating alternative consumption choices and expected net income potential. The existence of a population with differing preferences and the availability of alternative consumption bundles is the reason why individuals with the same net income potential move to different areas.

Even if the market for human capital is perfectly competitive, the information available and the corresponding uncertainty may vary with each migrant. Therefore, access to information may dominate the determinants of migration flows. Nelson (1959) hypothesized that a migrant is more likely to receive information about a region to which friends and relatives have previously migrated. Recent literature on migration has been concerned with the influence of past migration flows on current migration patterns. The hypothesis is that past migrants influence the locational choice of current migrants by providing information as well as reducing the costs associated with moving.

The introduction of migrant stock (a measure of past migration) into the specification of the model is used to capture the information flow from friends and relatives. The migrant stock variable is a measure of total past migration and hence a function of the same variables affecting current migration. This leads to certain specification problems concerning the form of the migration

model to be estimated. In such a model, time determines the structure of the estimating equations; therefore, a traditional migration model such as Greenwood's (1969), where current migration for the 1965-70 time span is a function of 1960 migrant stock, would not be correctly specified. The residuals over time from this traditional OLS single equation model would probably be highly correlated. In essence, the traditional model ignores past structural relationships which may produce residual correlations. This would imply that migrant stock is an endogenous variable and that the traditional model should be estimated using a full structural simultaneous system. Yet it is probably unnecessary to use a full simultaneous equation model since some of the endogenous coefficients will be insignificant. Current flows, for example, have no effects on past migration. Therefore, while the traditional single equation model is too simplistic, the full simultaneous model is not appropriate. Thus, what is needed is a model which lies between the two. One such model is the recursive system developed by Wold (1954).

III. Methodology

It is possible to restrict the full structural simultaneous equation model. One set of restrictions leads to a special case of the full structural model known as a recursive system. In this system the structural equations are ordered such that the first equation has only one endogenous variable, the second equation has two

endogenous variables, and so forth. The system could be represented as:

$$(2) \qquad M_1 = \beta_{11}X_{11} + \beta_{12}X_{12} + \ldots + \beta_{1m}X_{1m} + \varepsilon_1$$

$$(3) \qquad \alpha_{21}M_1 + M_2 = \beta_{21}X_{21} + \beta_{22}X_{22} + \ldots + \beta_{2m}X_{2m} + \varepsilon_2$$

$$(4) \qquad \alpha_{n1}M_1 + \alpha_{n2}M_2 + \ldots + M_n = \beta_{n1}X_{n1} + \beta_{n2}X_{n2} + \ldots + \beta_{nm}X_{nm} + \varepsilon_n$$

where α represents the coefficients on the n endogenous variables, β represents the coefficients on the m exogenous variables and ε represents the disturbance term for the equations. The M's represent migration flows in the model while the X's represent economic or migrant related variables. There would be T periods of observations on M and X.

If the disturbances of the equations in this type of model are independent, then concerning random components, each equation is unrelated to the preceding equations. There are no problems with the first equation since there is only one endogenous variable. The dependent variable in the second equation is determined by the exogenous variables, M_1 and ε_2. The random component of M_1 is ε_1, which is assumed independent of ε_2, thus M_1 may be regarded as predetermined with respect to M_2. Similar reasoning can be continued for the rest of the equations in the system.

Hence, all of the variables in any particular equation except the dependent variable can be treated as being predetermined. The ordinary least squares estimator gives unbiased estimates under these conditions. It is thus unnecessary to use a two-stage least squares estimator,

which is important due to several difficulties involved in using a two-stage least squares approach. The replacing of endogenous variables with their expectations in the second stage is one such difficulty. A great deal of inefficiency is introduced into the estimation in the second stage if these expectations are poor predictors. Another difficulty with the two-stage least squares estimator stems from multicolinearity in the data as well as multicolinearity associated with the estimation process. The first stage regresses the endogenous variables on all of the exogenous variables in the system. In a migration model, it is likely the exogenous variables are highly related. Another source of multicolinearity comes from the second stage when the endogenous variables are replaced by their expectations. These expectations are linear combinations of the exogenous variables.

A model which lies between the single equation regression model and the full structural simultaneous equation model is the recursive system. The simple estimation technique of the single equation model with parts of the simultaneity of the full structural simultaneous equation model is combined in the recursive system. The recursive system unlike the full structural simultaneous equation model does require an ordering of the dependent variables. The most logical ordering procedure would be an ordering based on time. If the dependent variables were determined at different times, they could be ordered to let

the dependent variable, which is determined first, be in the first equation. This procedure could be followed until the dependent variable which could be expected to be determined last would be in the last equation. Fortunately, it is easy to justify an ordering based on time in a migration model. The specification of the migration model as a recursive system provides information on the determinants of past and current migration as well as the influence of past migration on current flows.

The migration model is recursive over time but each time-period equation contains income variables. Income might be more the result than the cause of migration; therefore, it is treated as endogenous. The problems of inefficiency and multicolinearity that exist with endogenous migration variables are not encountered with income. The first-stage estimations are good predictors and the correlation between income variables is low. Therefore, with income being non-recursive in structure, a two-stage regression model was used for determining the predictive values of income for each equation within the recursive system.

IV. Empirical Model, Data, and Results

The theoretical framework based on an integration of utility maximization with investment behavior outlined in Section II can be combined with the block-recursive technique defined in Section III into an operational set of equations as follows:

$$(5)\quad M_{ij}^{40} = b_o + b_1D_{ij} + b_2\hat{I}_i^{40} + b_3\hat{I}_j^{40} + b_4T_i + b_5T_j + b_6A_i^{40} + b_7E_i^{40} + b_8MS_{ij}^{30} + e_{ij}$$

$$(6)\quad M_{ij}^{50} = c_o + c_1D_{ij} + c_2\hat{I}_i^{50} + c_3\hat{I}_j^{50} + c_4T_i + c_5T_j + c_6A_i^{50} + c_7E_i^{50} + c_8MS_{ij}^{30} + c_9M_{ij}^{40} + e_{ij}$$

$$(7)\quad M_{ij}^{60} = e_o + e_1D_{ij} + e_2\hat{I}_i^{60} + e_3\hat{I}_j^{60} + e_4T_i + e_5T_j + e_6A_i^{60} + e_7E_i^{60} + e_8MS_{ij}^{30} + e_9M_{ij}^{40} + e_{10}M_{ij}^{50} + e_{ij}$$

$$(8)\quad M_{ij}^{70} = d_o + d_1D_{ij} + d_2\hat{I}_i^{70} + d_3\hat{I}_j^{70} + d_4T_i + d_5T_j + d_6A_i^{70} + d_7E_i^{70} + d_8MS_{ij}^{30} + d_9M_{ij}^{40} + d_{10}M_{ij}^{50} + d_{11}M_{ij}^{60} + e_{ij}$$

where:

M_i^{70} = the migration rate: the number of migrants who were residing in state i in 1965 and had migrated to state j by 1970, divided by total population at origin i in 1965;

M_{ij}^{60}, M_{ij}^{50}, M_{ij}^{40} = migration rates similarly defined for the given year[3];

MS_{ij}^{30} = past migrants (migrant stock) who were born in state i (origin) and enumerated in state j (destination), 1930;

$\hat{I}_i$, $\hat{I}_j$ = mean family income in state i and j for the given year; the hats indicate two-stage estimates;

T_i, T_j = absolute deviation of mean yearly temperature from 65 in state i and j. T_i and T_j are assumed constant over time;

E_i = median years of school completed at state (origin) i for the given year;

A_i = median age of population at state i for the given year;

D_{ij} = highway mileage between the city with the greatest population at state i to that of state j; and

e_{ij} = a random error term assumed to have a lognormal distribution.

The sample used for an empirical test of the recursive system is the set of cross-section gross migration census data for the years 1930, 1940, 1950, 1960, and 1970.[4] Cross-section migration data are likely to have inter-dependencies which is a relevant consideration since the recursive system will give better estimates than the traditional single equation migration model only if there are important inter-migration relationships which could not be explained by each migration-equation relationship with the current explanatory variables. The empirical results are discussed in terms of the traditional model, the lagged migration model and finally the relationship between the migrant stock concept and spatial information flows.

Non-Recursive Results: The Traditional Model

For comparative purposes, estimates of the traditional non-recursive equation model with migrant stock were made. The results from the non-recursive regressions for migration rates in Table I indicate that the migrant stock variable is highly significant for each year, thus supporting the information flow hypothesis. The percentage of variation explained by each equation was approximately 65.0%.

The distance variable is used in the analysis as a proxy for time, psychic cost and direct money costs of

Table I

TRADITIONAL MODEL: NON-RECURSIVE MIGRATION EQUATIONS for 1970, 1960, 1950, and 1940[a]

Variable	Migration Equation By Year			
	M_{ij}^{70}	M_{ij}^{60}	M_{ij}^{50}	M_{ij}^{40}
Constant	-1.828 (9.16)	-11.47 (8.82)	-10.98 (9.38)	-24.05 (21.11)
Distant (D_{ij})	-0.379 (12.95)	-0.476 (16.18)	-0.492 (14.97)	-0.718 (20.72)
Income Destination ($\hat{I}_j$)	1.959 (11.64)	1.632 (12.87)	0.856 (7.01)	1.177 (15.38)
Income Origin ($\hat{I}_i$)	-1.539 (6.69)	-3.541 (9.33)	-2.779 (8.13)	-2.240 (13.87)
Age (A_i)	-3.088 (8.68)	0.107 (0.22)	-0.193 (0.32)	0.521 (1.08)
Education (E_i)	7.095 (13.45)	5.135 (17.92)	8.111 (24.67)	10.96 (30.48)
Temperature Origin (T_i)	0.125 (4.39)	.127 (4.71)	0.117 (3.86)	0.006 (1.89)
Temperature Destination (T_j)	-.394 (15.48)	-0.500 (18.54)	-0.500 (16.61)	-0.511 (17.06)
Migrant Stock - 1960 MS_{ij}^{60}	0.451 (32.74)			
Migrant Stock - 1950 MS_{ij}^{50}		0.434 (34.01)		
Migrant Stock - 1940 MS_{ij}^{40}			0.422 (31.74)	
Migrant Stock - 1930 MS_{ij}^{30}				0.471 (3.61)
R^2	0.628	0.681	0.642	0.700

[a]The absolute value of the t-ratios is in parentheses; each equation is based on 2230 observations. This is less than 2256 (48 x 47) because in some instances the migration flows were zero. These observations were eliminated from the sample. Equations were estimated in double-log form, hence the estimated coefficients are directly interpretable as elasticities.

moving. Greater distance may also increase the cost of acquiring information which in turn increases uncertainty.[5] All the above factors lead to the expectation that migration will be negatively related to distance. This was confirmed with the non-recursive results with distance being negative and significant in all cases, but declining in importance over time.[6] This suggests that as transportation and communications facilities improve, the impact of distance will continue to decline.

Income is used to reflect the economic opportunities in a state. It is generally assumed in studies on internal migration that higher income at the origin will deter migration and higher income possibilities at the destination will attract migration. Migrants consistently respond to economic conditions as reflected by the significant coefficients on the income variables. The destination income elasticity is greater than one for all years except 1950 when it was .856. The origin income elasticity is greater than minus two in all cases except in 1970 when it was a -1.54. The size of the income elasticities reflect the willingness of migrants to move toward expected superior economic opportunities.

Recent studies have suggested several explanations for the influence of education on migration.[7] Education may increase the ability of a person to obtain more information about destination areas. The educated may also face lower risk when moving since they are more

adaptable to changing environment and job opportunities. These factors indicate that educated persons are more likely to migrate. The results in Table I indicate that education has had a significant positive impact in all years on the propensity to migrate.

Age at the origin (A_i) may be a proxy for job experience which could substitute for education (Schwartz [1962]). In this respect, age could have a positive effect on migration. However, a more likely possibility, especially since education is included in the analysis, is that older populations may have less of a tendency to migrate since older persons have a shorter expected working life over which to capture the advantages of migrating.[8] The corresponding investment and the rate of return from migration is lower. For the non-recursive results in Table I the coefficient for age is negative and significant in 1970 but insignificant in all other cases.

Moderate temperatures are more attractive and tend to possibly reduce the cost of living. The temperature variable included in this study represents a departure from that used by Greenwood (1969) and Cebula and Vedder (1973). The temperature variable is viewed as the absolute deviation of the mean temperature from 65°F which measures the preference for a temperate climate. This definition of temperature allows for the expected negative impact on migration of extreme variations at both ends of the scale. It is expected that the temperature variables at the origin

(T_i) would be positively related to migration while (T_j) would have a negative influence. The temperature coefficients for both origin and destination are significant and the correct sign in all cases.

Thus the traditional model seems to correctly specify the migration process. Laber (1972) and Renshaw (1974), however, have suggested specification problems related to the migrant stock variable. The criticisms are that either it is a partial adjustment mechanism which induces multicolinearity or it is biased by left-out variables. In order to investigate the specification problems surrounding the migrant stock variable, the spatial allocation of migrants over time is estimated in a block-recursive system of equations.

Recursive Results: The Lagged Migration Model

The results for the estimation of the recursive system are in Table II. The table indicates many strong intermigration relationships not accounted for by the migrant stock variable. For example, the t-values indicate a strong relationship between migration rates in 1960 and rates in 1970 and for rates between 1950 and 1960. These results suggest that the influence of migrant stock is decomposed into its component parts. In other words, the migrant stock variable reflects the combined influence of recent and past migration patterns. If these patterns were not consistent over time, then the migrant stock variable may not reflect accurate information concerning the impact of past migration

Table II

RECURSIVE MODEL: LAGGED MIGRATION
for 1970, 1960, 1950, and 1940[a]

Variable	Migration Equation By Year			
	M_{ij}^{70}	M_{ij}^{60}	M_{ij}^{50}	M_{ij}^{40}
Constant	1.249 (1.29)	0.206 (0.28)	-7.223 (6.69)	-24.05 (21.11)
Distance (D_{ij})	-.054 (3.51)	-0.042 (2.32)	-0.302 (9.88)	-0.718 (20.72)
Income Destination (I_j)	0.016 (0.20)	0.556 (7.58)	0.673 (6.11)	1.177 (15.38)
Income Origin (I_i)	-0.449 (3.94)	-1.035 (4.38)	-3.361 (10.64)	-2.240 (13.87)
Age (A_i)	-0.189 (1.12)	0.590 (2.04)	0.612 (1.10)	0.521 (1.08)
Education (E_i)	1.174 (4.33)	0.810 (2.81)	6.856 (23.02)	10.96 (30.48)
Temperature Origin (T_i)	.037 (2.74)	0.017 (1.14)	0.249 (8.82)	0.006 (1.89)
Temperature Destination (T_j)	-0.017 (1.38)	-0.131 (8.49)	-0.347 (12.30)	-0.511 (17.06)
Migrant Stock - 1930 (MS_{ij}^{30})	-0.0008 (0.14)	-0.007 (0.99)	-0.072 (3.34)	0.471 (3.61)
M_{ij}^{40}	-0.046 (3.06)	0.187 (11.12)	0.589 (25.41)	
M_{ij}^{50}	0.107 (5.41)	0.654 (37.47)		
M_{ij}^{60}	0.857 (4.40)			
R^2	0.916	0.903	0.706	0.700

[a]See footnote a in Table I.

on future migration rates. The recursive results separate these influences over time and indicate that recent migrants represent the dominant influence on future migrants. For example, 1970 migration rates were insignificantly influenced by 1930 migrant stock. The 1940 rates exerted a negative influence of -0.042 which suggests that 1970 migrants were not following 1940 migration patterns. The coefficients in the 1970 equation for 1950 and 1960 migration rates were .097 and .859, respectively, with corresponding t-values of 7.85 and 53.67, thus demonstrating the importance of recent past migrants in providing prospective current migrants with information.

The coefficient of determination after being adjusted for the degrees of freedom for the equations in the recursive system is higher than in the traditional model in all cases. For example, the R^2 for the 1970 equation increased from .633 in the traditional equation to .916 in the recursive system. In general, the R^2 will increase if and only if the student-t of the additional independent variable is larger than one. In any event, when the R^2 increases from .633 to .916 this strongly indicates that the explanatory power of the model has been increased.

An examination of the degree of correlation among the residuals can give some insight into the specification of the model. The correlation of the residuals for each equation with the residuals for all of the other equations for both the traditional equation model and the recursive system are in Table III. For example, the correlation between the residuals

Table III

RESIDUAL CORRELATION MATRIX FOR THE TRADITIONAL AND RECURSIVE MODEL[a]

Equation	M_{ij}^{40}	M_{ij}^{50}	M_{ij}^{60}	M_{ij}^{70}
M_{ij}^{40}		(.110)	(.124)	(.048)
	XX	.680[b]	.713[b]	.632[b]
M_{ij}^{50}			(.050)	(.019)
		XX	.728[b]	.589[b]
M_{ij}^{60}				(.017)
			XX	.768[b]
M_{ij}^{70}				
				XX

[a]The correlation coefficients for the recursive model are in parentheses.

[b]Significant at the one percent level for a two-tailed test.

for 1970 and 1960 migration equations drop from 0.768 to .017. The recursive system significantly decreases residual correlations among all migration equations.

The significant residual correlations in the traditional model violate the assumptions of the migration model. The reduction in the correlation of residuals in the recursive system provides evidence of the suitability of the model. Residuals close to zero make the recursive system estimates essentially equal to the full information maximum likelihood estimates.

The other variables in the recursive system are generally stable with respect to sign but many of the coefficients decrease in magnitude. Thus, the recursive specification suggests that the traditional model under-estimates the magnitude and importance of information flows between migrants and overemphasizes the impact of the other explanatory variables. The problem with this conclusion is, for example, that migration rates for 1970 are regressed on migration rates for 1960. If migration patterns have remained rather stable between two periods of time then the lagged migration rate may become a dominant variable or at least exhibit some of the symptoms (downward bias) which reduce the reliability of the estimates. Thus an alternative functional form which captures the essence of information flow without the problem of being a lagged dependent variable would possibly be more desirable.

Information Flow and the Migrant Stock Variable

An alternative recursive formation would be to use differential migrant stock variables instead of lagged migration rates. For example, the 1970 migration equation would have four migrant stock variables, the total migrant stock for 1930 (MS_{ij}^{30}), the migrant stock for 1940 minus 1930 ($MS_{ij}^{40}-MS_{ij}^{30}$), migrant stock for 1950 minus 1940 ($MS_{ij}^{50}-MS_{ij}^{40}$), and migrant stock for 1960 minus 1950 ($MS_{ij}^{60}-MS_{ij}^{50}$). The correlation between migration rate 1970 and 1960 is 0.95, whereas for the differential migrant stock variable and the migration rate in 1970, the correlation is only 0.48, thus eliminating the dominant variable phenomenon and yet preserving the impact of information flow from previous migrants.

The differential migrant-stock equations for 1970, 1960, and 1950 in Table IV indicate structural variations regarding the impact of past migrants on current migration. In the 1970 equation, the migrant stock 1930 variable represents the influence of migrants two generations old and/or is a measure of the influence of long-run migration patterns. For the 1970 and 1960 equations, the former seems the most probable since information flow from two-generation-old migrants would be sparce.

Comparisons across migration equations reveal a decreasing migrant-stock-1930 coefficient reflecting reduced information flow from older migrants. The 1970 equations and 1960 migration equations reveal significant insights into the

Table IV

RECURSIVE MODEL: LAGGED MIGRANT STOCK
for 1970, 1960, 1950, and 1940[a]

Variable	Migration Equation By Year M_{ij}^{70}	M_{ij}^{60}	M_{ij}^{50}	M_{ij}^{40}
Constant	-19.416 (9.52)	-8.914 (6.636)	-10.55 (8.90)	-24.05 (21.11)
Distance (D_{ij})	-.463 (15.51)	-0.536 (17.45)	-.492 (14.91)	-0.718 (20.72)
Income Destination ($\hat{I}_j$)	2.174 (12.41)	1.496 (11.05)	0.812 (6.43)	1.177 (15.38)
Income Origin ($\hat{I}_i$)	-1.525 (6.34)	-3.850 (9.07)	-2.544 (7.34)	-2.240 (13.87)
Age (A_i)	-2.701 (7.55)	0.566 (1.08)	-0.509 (0.83)	0.521 (1.08)
Education (E_i)	6.897 (12.58)	8.211 (16.21)	7.81 (23.30)	10.96 (30.48)
Temperature Origin (T_i)	0.121 (4.27)	0.143 (5.20)	0.121 (3.97)	0.006 (1.89)
Temperature Destination (T_j)	-.402 (15.47)	-0.469 (16.46)	-0.478 (15.30)	-0.511 (17.06)
Migrant Stock - 1930 (MS_{ij}^{30})	0.231 (20.89)	0.270 (23.95)	0.382 (30.43)	0.471 (3.61)
Migrant Stock - 1940-30 ($MS_{ij}^{40} - MS_{ij}^{30}$)	0.00891 (1.21)	0.015 (2.08)	0.057 (8.38)	
Migrant Stock - 1950-40 ($MS_{ij}^{50} - MS_{ij}^{40}$)	0.048 (4.62)	0.137 (16.05)		
Migrant Stock - 1960-50 ($MS_{ij}^{60} - MS_{ij}^{50}$)	0.128 (14.97)			
R^2	.633	0.676	0.640	0.700

[a]See footnote a in Table I.

structure of information flow over time. The most recent past-migrants ($MS_{ij}^{60}-MS_{ij}^{50}$) have a significant positive influence on current migration patterns. The 1950-1940 addition to migrant stock is also significant and positive but the coefficient is significantly smaller reflecting their decreased influence. The 1940-1930 migrant stock has no significant influence on current migration. The MS_{ij}^{30} variable has the largest coefficient thus reflecting the importance of either long-run migration patterns or the influence from 1930 of all past migrants. The 1960 equation has the same pattern. For the 1950 equation migrant stock 1930 has a much larger impact than the differential migrant stock 1940-1930 variable. The 1940 equation represents the traditional model with a migrant stock variable. The migrant stock variables for the migration equations in Table III provide additional knowledge of the differential impact of migrants over time. The differential migrant stock approach as compared to the lagged migration rate seems to eliminate the dominant variable effect resulting from introducing an explanatory variable almost identical to the dependent variable. Whereas the lagged migration equations "seem" to reduce the significance of other explanation variables, the differential migrant stock equation in Table III produces approximately the same significant coefficients on the variables of income, age, education, distance, and temperature as in the traditional model while still providing for possible information flow from friends and relatives.

V. Summary and Conclusions

The purpose of this paper has been to present some empirical evidence on the determinants of the spatial allocation of migrants in the United States over time. The specification of the migration model has been inadequate in the treatment of the impact of past migration flows on future migration probabilities. The paper has explored this aspect of the migration model using data on gross interstate migration flows in the United States for the 1940, 1950, 1960, and 1970 time periods.

A block-recursive system has been offered as a compromise between the simplicity of the traditional migration model and the theoretically eloquent full structural simultaneous equation model. The results from an empirical test comparing the recursive and traditional model support the existence of significant residual correlations, thus casting doubt on the usefulness of the traditional migration model.

The main contribution of the recursive system is its decomposition of the influence of past migrants on current migration flows. This decomposition is important and useful since the traditional migration model does not discriminate sources of change over time. The effect of information flow from past migrants decreases rapidly over time with long-run patterns of migration being rather stable. The recursive system for migration allows for a better explanation of the effect of past migration on current flows.

The recursive system does have a tendency when dealing with migration to create possible dominant variable problems. This problem was eliminated by changing the lagged dependent variable into a differential migrant stock variable. This allows for a more accurate interpretation of the influence of past migrants with coefficients of other explanatory variables remaining unbiased.

Footnotes

[1]For a review of the literature on internal migration in the United States, see Greenwood (1975).

[2]See Baumol (1965) for a discussion of integer programming and economics.

[3]The data for 1950 was reported as migration over only a one-year period of time. Thus the migration rate was divided by population in 1950. This smaller time span, while decreasing the absolute number of migrants, probably does not decrease the importance of the direction of the migration flows.

[4]The migration data were taken from the Census of Population for the respective years. Likewise, the income, age, and education data were from the Census. Data on interstate flows were first collected in the 1940 Census data. The data on mean temperature were from the U. S. Statistical Abstract while distance was based on the Rand-McNally Road Atlas. Temperature and distance were assumed to be constant over time. A detailed appendix containing data sources is available from the authors.

[5]The role of distance in the migration decision has been explained by three hypotheses: diminishing information hypothesis, intervening opportunities hypothesis, and increasing costs hypothesis. Three recent empirical studies have attempted to interpret the influence of distance on migration with varying conclusions. See Miller (1972), Levy and Wadycki (1974), and Schwartz (1973).

[6]F-tests were performed which indicated that the coefficients were significantly different across each of the equations.

[7]For a discussion of the influence of education on migration, see Levy and Wadycki (1974), Sahota (1968), Greenwood (1969), Beals, Levy, and Moses (1967), and Bowles (1970). The paper by Levy and Wadycki (1974) provides an empirical test of the various hypotheses concerning the influence of education on migration of three migration flows classified by education levels for Venezuela.

[8]One of the limitations on previous studies of internal migration has been the failure to control for differences in the propensity to migrate caused by increased age and education. See the study by Langley for some results of differing migration behavior for various age groups.

References

Baumol, W. J., Economic Theory and Operations Analysis (New York: Prentice-Hall, 1965).

Beals, Ralph E., M. B. Levy, and L. N. Moses, "Rationality and Migration in Ghana," Review of Economics and Statistics (1967), 480-486.

Bowles, S., "Migration as Investment: Empirical Tests of the Human Investment Approach to Geographical Mobility," Review of Economics and Statistics (1970), 356-362.

Cebula, Richard J. and Richard K. Vedder, "A Note on Migration, Economic Opportunity and the Quality of Life," Journal of Regional Science (August 1973), 205-211.

Dunlevy, J. A. and H. A. Gemery, "A Comparison of Migrant Stock and Lagged Migration as Explanatory Variables in Migration Studies," Presented at Western Economic Association meetings (June 1975).

Greenwood, Michael J., "An Analysis of the Determinants of Geographic Labor Mobility in the United States," Review of Economics and Statistics (May 1969), 189-194.

_______________, "Lagged Response in the Decision to Migrate," Journal of Regional Science (December 1970), 375-38 .

_______________, "Lagged Response in the Decision to Migrate: A Reply," Journal of Regional Science (August 1972), 311-324.

_______________, "Research on Internal Migration in the United States: A Survey," Journal of Economic Literature (1975), 397-433.

Johnston, J., Econometric Methods (New York: McGraw-Hill, 1972).

Klein, L. R. and N. Mitsugu, "Singularity in the Equation Systems of Econometrics: Some Aspects of the Problem of Multicolinearity," International Economic Review (1962), 274-289.

Laber, Gene, "Lagged Response in the Decision to Migrate: A Comment," Journal of Regional Science (August 1972), 307-310.

Langley, Paul C., "The Spatial Allocation of Migrants in England and Wales: 1961-1966," Scottish Journal of Political Economy (November 1974), 259-276.

Levy, Mildred B. and Walter J. Wadycki, "The Influence of Family and Friends on Geographic Labor Mobility: An International Comparison," Review of Economics and Statistics (May 1973), 198-203.

______________, "What is the Opportunity Cost of Moving? Reconsideration of the Effects of Distance on Migration," Economic Development and Cultural Change (January 1974), 198-214.

______________, "Education and the Decision to Migrate: An Econometric Analysis of Migration in Venezuela," Econometrica (March 1974), 377-388.

Miller, Edward, "A Note on the Role of Distance in Migration: Costs of Mobility Versus Intervening Opportunities," Journal of Regional Science (December 1972), 475-478.

Nelson, Philip, "Migration, Real Income and Information," Journal of Regional Science (1959), 43-74.

Renshaw, Vernon, "A Note on Lagged Response in the Decision to Migrate," Journal of Regional Science (August 1974), 273-280.

Sahota, Gian S., "An Economic Analysis of Internal Migration in Brazil," Journal of Political Economy (1968), 218-245.

Schwartz, Aba, "Interpreting he Effect of Distance on Migration," Journal of Political Economy (September/ October 1973), 1153-1169.

Sjaastad, Larry A., "The Costs and Returns of Human Migration," Journal of Political Economy (October 1962), 80-93.

U. S. Census of Population, Lifetime and Recent Migration, Subject Report 2D (1973), Table 11.

______________, Mobility for the States and the Nation, Subject Report 2B (1973), Table 44.

Wold, H., "Causality and Econometrics," Econometrica (1954), 162-177.