DETERMINANTS OF NET MIGRATION AND MIGRATION EFFICIENCY IN SOUTHERN STATE ECONOMIC AREAS

William J. Serow, Julia H. Martin, Michael A. Spar
Tayloe Murphy Institute, University of Virginia

Introduction

In a recent paper by Galle and Williams (1972) migration efficiency rates for 1955 to 1960 were investigated using large Standard Metropolitan Statistical Areas (SMSAs) as the unit of analysis. The present study extends that investigation by analyzing characteristics of State Economic Areas (SEAs) within the East South Central and South Atlantic Census Divisions in relation to migration efficiency rates calculated for the 1965 to 1970 time period. One major objective of this study therefore, is to determine whether specific SEA characteristics are associated with migration efficiency; another is to locate and describe methodological problems arising when a number of the independent variavles used to explain the phenomenon of migration efficiency--or indeed any similar phenomenon--prove to be closely interrelated.

SEAs are used as the unit of analysis in this research because unlike SMSAs they permit coverage of rural and nonmetropolitan urban as well as metropolitan areas and populations while retaining relative homogeneity of social and economic characteristics. The Bureau of the Census defines two types of SEA, nonmetropolitan and metropolitan. The latter were first defined in 1950 as standard metropolitan areas with total 1940 populations of 100,000 or more. In 1960 additional metropolitan SEAs were defined due to compositional changes in some SMSAs. At that time metropolitan SEAs were defined as 1960 SMSAs with central cities of 50,000 or more and total populations of 100,000 or more. The East South Central and South Atlantic Divisions contain 143 SEAs, of which 54 are classified as metropolitan.

Following Galle and Williams (1972) and Shryock (1964), migration efficiency is defined as the quotient of net migration to an SEA (inmigrants minus outmigrants) divided by gross migration, the sum of all moves centered on the SEA (inmigrants plus outmigrants). It follows from this definition that the larger the absolute value of this quotient the more "efficient" the migration. For example, an SEA with net inmigration of 1,000 produced by 1,000 inmigrants and zero outmigrants would have a migration efficiency rate of 1.0, while the same volume of net migration produced by 5,500 inmigrants and 4,500 outmigrants would have an efficiency rate of 0.1. Despite the fact that rates of +1.0 and -1.0 denote the same degree of efficiency, these rates are qualitatively different, being achieved through different configurations of the basic rates. It is therefore appropriate to analyze separately those SEAs which are characterized by positive efficiency rates and those with negative efficiency rates since the pattern and weight of variables which are associated with either may be expected to differ.

Sources of Data and Methods of Analysis

Migration efficiency rates were calculated from data contained in Migration Between State Economic Areas (U.S. Bureau of the Census, 1972). Twenty-two independent variables reflecting general social and economic characteristics of the SEAs were derived from State Economic Areas (U.S. Bureau of the Census, 1972). To avoid artificial inflation of the explanatory power of these variables, no independent variable concerned with population change was included in the analysis. The twenty-two variables and their assigned mnemonic codes are listed in Table 1.

Three statistical techniques were used in the research. An initial zero-order correlation matrix was obtained in order to suggest the amount of multicollinearity present among the initial 22 independent variables. Factor analysis was then used to reduce multicollinearity, permitting the construction of factor scales. Both the factor scales and the unfactored variables were then used as independent variables in separate regression analyses, yielding a comparison of the explanatory power of each as well as the relative importance of the factor scales in "explaining" migration efficiency rates.

Methodology

In the first stage of the research the universe of 143 SEAs was divided into four groups, based on the criteria of metropolitan status (metropolitan or nonmetropolitan SEA) and the sign of the migration efficiency rate (positive or negative). A correlation matrix of the independent variables was then obtained for each group. Inspection of these matrices revealed that in each case approximately 20 percent of the correlations were greater than $|.50|$, a figure judged to indicate a high degree of multicollinearity.

A major problem associated with the use of interrelated independent variables in regression analysis was pointed out by Blalock (1963): as the degree of correlation between independent variables increases, the standard error of the estimates of the slope becomes quite large, decreasing the accuracy of the estimates. Consequently, it becomes difficult or impossible to use the beta coefficients as indicators of the relative importance of the explanatory variables, although R^2, the coefficient of multiple determination, may still be used to indicate the total

TABLE 1. INDEPENDENT VARIABLES AND ASSIGNED MNEMONIC CODES

Mnemonic	Variable Description
POP17UND	Percent of SEAs population age zero to seventeen
POP65OVR	Percent of SEAs population age 65 years and over
DORMIESS	Percent of SEAs population residing in either barracks or dormitories
CHILDFAM	Percent of all families in SEA with own children age zero to five
NONWHITE	Percent of SEAs population nonwhite
URBANPOP	Percent of SEAs population classified as urban by the Bureau of the Census
BORNOUTS	Percent of SEAs population born out-of-state
MEDYEDML	Median years of education for the male population over age 25
MEDYEDFL	Median years of education for the female population over age 25
CHLDBORN	Number of children ever born per 1,000 married women age 35 to 44
LFPRMALE	Labor force participation rate, males age 16+
LFPRFALE	Labor force participation rate, females age 16+
UNMPLYED	Percent of labor force unemployed
LFPRWCU6	Labor force participation rate of women with children under age six
AGRCULTR	Percent of labor force in agricultural occupations
BLUECLAR	Percent of labor force in blue collar occupations
WHITECLR	Percent of labor force in white collar occupations
FEDERALS	Percent of labor force in government occupations
MEDFAMIN	Median family income
PERCAPIN	Per capita income
WLFREFAM	Percent of families receiving welfare payments
PVRTYFAM	Percent of families with earnings less than poverty level

amount of variance accounted for by all independent variables taken together.

In the present study, however, a second major problem, also noted by Blalock (1960: 357) would make the R^2's unreliable, since artificially large multiple correlations may be obtained if the number of variables in the estimated regression equation begins to approach the number of cases analyzed, a situation which did in fact occur when the original 143 SEAs were divided into four groups. Therefore, within each SEA group the 22 independent variables were subjected to a factor analysis, the factor matrices being rotated via varimax rotation to final solution. Varimax rotation is a type of orthogonal rotation procedure that attempts to obtain factors which are maximally independent of one another. In the ideal case each independent variable will load significantly on only one factor, with factor loadings near zero on the other factors, though in practice some multiple significant loadings usually occur. A successful factor analysis would, however, provide a solution to the two problems described above by both reducing the number of variables and maximizing their independence.

Having obtained the rotated factor matrix, it is necessary to transform the results into a form usable as a new set of independent variables, thus the construction of factor scale scores for each case. In constructing these scales, three decisions must be made: (1) which of the independent variables should be included in the scale; (2) what should be done with variables which load significantly on more than one factor; and (3) how should the variables be weighted in the scale. The scales used in the present study were constructed as follows: first, only those variables with factor loadings greater than $|.30|$ were considered significant. This significance level was chosen because the square of the factor loading represents the amount of variation in that variable explained by the factor. A factor loading of .30 is therefore equivalent to explaining about 10 percent of the variance in the variable. Second, variables loading significantly on more than one factor were eliminated from the analysis in order to maximize independence of the factor scales, reducing multicollinearity. Third, in cases where these criteria led to the elimination of all but one variable from a factor, the remaining variable was retained as an independent variable in standardized form. Fourth, where these criteria led to the elimination of all variables from a factor, the factor was discarded. The scales were then computed by multiplying the square of the factor loading of each variable selected by the value of the standardized variable and summing the results.

The results of these procedures yielded a new set of independent variables, greatly reduced in number and in the degree of their interrelationship. It remained to determine their explanatory power for migration efficiency and to compare this result with the power of the original, unfactored independent variables. As the final step, therefore, the factor scales and the original, unfactored variables were used as independent variables in separate multiple regression analyses with migration efficiency rates as the dependent variable in both cases.

Results

Factor Analysis

The results of the factor analyses of the 22 independent variables are presented in Table 2.

TABLE 2. FACTOR LOADINGS FOR TWENTY-TWO INDEPENDENT VARIABLES, BY SEA GROUP

Independent variables	Factor 1: High SES*				Factor 2: Young Families			
	NM(+)	NM(-)	M(+)	M(-)	NM(+)	NM(-)	M(+)	M(-)
POP17UND					.92	.79	.93	.93
POP65OVR					-.90	-.75	-.93	-.58
DORMIESS					.35			
CHILDFAM					.97	.91	.95	.75
NONWHITE	.38				.40	.36		
URBANPOP	.64	.63	.51	.57		.38	-.34	
BORNOUTS	.85	.66	.45					
MEDEDML	.94	.91	.86	.92				
MEDEDFL	.92	.91	.86	.91				
CHLDBORN	.37				.56	.67	.42	.73
LFPRMALE		.41			.77		.86	.58
LFPRFALE	-.35						.35	
UNMPLYED								
LFPRWCU6	-.35						-.31	
AGRCULTR								
BLUECLAR	-.91	-.32	-.97	-.92				
WHITECLR	.95	.77	.87	.88				
FEDERALS	.65		.60	.68	.47		.55	
MEDFAMIN	.31	.73	.56	.40	.35		.35	
PERCAPIN	.51	.76	.61	.45				
WLFREFAM		-.42				.43		
PVRTYFAM		-.65	-.41					

	Factor 3: Female Employment				Factor 4: Agricultural Employment			
	NM(+)	NM(-)	M(+)	M(-)	NM(+)	NM(-)	M(+)	M(-)
POP17UND						.41		
POP65OVR		-.40						
DORMIESS								
CHILDFAM								
NONWHITE	.32	.42		.79		.66		.32
URBANPOP					-.43		.68	.55
BORNOUTS			.68					
MEDEDML								
MEDEDFL								
CHLDBORN			.50			.56	-.31	
LFPRMALE		.83						
LFPRFALE	.84	.93	-.81	.91				
UNMPLYED	-.77	-.65	.63	-.41				
LFPRWCU6	.86	.87	-.71	.96				
AGRCULTR					.85	.77	-.85	-.79
BLUECLAR						-.83		
WHITECLR		-.41						
FEDERALS		-.66		.43	.30	.32		
MEDFAMIN		.39				-.49		
PERCAPIN		.34				-.41		
WLFREFAM		-.31				.56		.59
PVRTYFAM		-.37				.56		

	Factor 5: Poverty Families				Factor 6: Institutionalized Population			
	NM(+)	NM(-)	M(+)	M(-)	NM(+)	NM(-)	M(+)	M(-)
POP17UND		Factor Not Present				-.32	Factor Not Present	
POP65OVR								-.65
DORMIESS			.72		.80	.88		.85
CHILDFAM								.53
NONWHITE	.64		.75	.33				
URBANPOP					.42			
BORNOUTS			-.30			.36		.80
MEDEDML								
MEDEDFL			-.31					
CHLDBORN	.60		.54	.49				
LFPRMALE	-.31							.65
LFPRFALE								
UNMPLYED	.34		.57	.62				

TABLE 2. FACTOR LOADINGS FOR TWENTY-TWO INDEPENDENT VARIABLES, BY SEA GROUP (Continued)

Independent variables	Factor 5: Poverty Families				Factor 6: Institutionalized Population			
	NM(+)	NM(-)	M(+)	M(-)	NM(+)	NM(-)	M(+)	M(-)
LFPRWCU6		Factor Not Present	.31		Factor Not Present			
AGRCULTR								
BLUECLAR								
WHITECLR			-.34	-.32				
FEDERALS						.37		.31
MEDFAMIN	-.78		-.68	-.86				
PERCAPIN	-.77		-.72	-.69				
WLFREFAM	.84		.80	.55				
PVRTYFAM	.92		.88	.90				

* NM(+) = Nonmetropolitan SEAs, Positive Migration Efficiency Rate
NM(-) = Nonmetropolitan SEAs, Negative Migration Efficiency Rate
M(+) = Metropolitan SEAs, Positive Migration Efficiency Rate
M(-) = Metropolitan SEAs, Negative Migration Efficiency Rate

Of the six factors described in this table, four are present in all four SEA groups. We have labelled these high socioeconomic status (factor 1), young families (factor 2), female employment (factor 3) and agricultural employment (factor 4). It should be noted, however, that despite the fact that the substantive content of these factors is sufficiently similar so as to warrant identical labelling, there are differences in factor composition among the groups. The case of the variable NONWHITE on factor 2 is an example: it has moderately high positive loadings in both nonmetropolitan groups but is insignificant for both metropolitan groups. Similarly, the variables FEDERALS and MEDFAMIN have moderately high positive loadings for SEAs with positive rates of migration efficiency but are insignificant in both SEA groups characterized by negative efficiency rates.

In addition to these four "common" factors, two other factors were also generated. These are labelled poverty families (factor 5) and institutional population (factor 6) and are presented in the bottom section of Table 2. The poverty families factor is not present in the nonmetropolitan SEA group with negative efficiency rates; instead, the variables which load on this factor are found on the agricultural employment factor.

Factor 6, institutionalized population factor, is not present in the metropolitan SEA group with positive efficiency rates; variables which "belong" on this factor appear mainly on factors 2 and 5. A list of the variables used in constructing the final factor scales, together with their factor loadings, is contained in Table 3.

These factor scales, while not completely independent, display a pattern of intercorrelations of less magnitude than the independent variables. Data in Table 4 show that the highest correlation between any two factor scales is $|.38|$, whereas approximately 20 percent of the correlations among the unfactored variables were above $|.50|$. The use of factor analytic techniques has thus significantly reduced the degree of multicollinearity between the independent variables.

Multiple Regression Analysis

Tables 5 and 6 present results of the multiple regression analysis; Table 5 contains results for the 22 independent variables and Table 6 results for the factor scales. For each SEA group values of R^2, the coefficient of multiple determination, are given, as well as values of $\hat{R}^2$, an unbiased estimate of R^2. Our use of $\hat{R}^2$ is necessary due to the large number of independent variables in the regression function. Comparison of the two tables, and especially the values for $\hat{R}^2$ indicate that the explanatory power of the factor scales is much less than that of the unfactored variables. For the unfactored variables, R^2 ranges from .55 to .76, for the factor scales the range is much lower, from .04 to .48. The reduction in $\hat{R}^2$ values is difficult to explain, although we hypothesize that the situation is such that each of the independent variables explains some small proportion of the variance in the efficiency rate, independent of the contribution made by the others. In the aggregate this results in fairly high values of R^2 when all independent variables are used to estimate the regression function. The ommission of most of these variables in constructing factor scales leads directly to low values of R^2.

Despite the low values of R^2 produced when factor scales are used to fit a least squares equation, the following discussion will focus on the relationships between migration efficiency rates and the factor scales, because of the greater reliability of the beta coefficients derived from the regression of the factor scales.

For metropolitan SEAs with net inmigration, efficiency rates increase with increases in both the proportion of the labor force engaged in agricultural employment and with the SES level of the population. Conversely, efficiency is reduced as the proportion of either young families or poor

TABLE 3. FACTOR LOADINGS OF INDEPENDENT VARIABLES USED IN CONSTRUCTING FACTOR SCALES

Independent variables	Factor 1: High SES*				Factor 2: Young Families			
	NM(+)	NM(-)	M(+)	M(-)	NM(+)	NM(-)	M(+)	M(-)
POP17UND					.92		.93	.93
POP65OVR					-.90		-.93	
CHILDFAM					.97	.91	.95	
BORNOUTS	.85							
MEDYEDML	.94	.91	.86	.92				
MEDYEDFL	.92	.91		.91				
LFPRMALE							.86	
LFPRFALE	-.35						.35	
LFPRWCU6	-.35							
BLUECLAR	-.91		-.97	-.92				
WHITECLR	.95							
	Factor 3: Female Employment				**Factor 4: Agricultural Employment**			
	NM(+)	NM(-)	M(+)	M(-)	NM(+)	NM(-)	M(+)	M(-)
LFPRFALE	Factor Dropped	.93	Factor Dropped	.91				
UNMPLYED		-.65						
LFPRWCU6		.87		.96				
AGRCULTR					.85	.77	-.85	-.79
	Factor 5: Poverty Families				**Factor 6: Institutionalized Population**			
	NM(+)	NM(-)	M(+)	M(-)	NM(+)	NM(-)	M(+)	M(-)
DORMIESS		Factor Not Present	.72		Factor Dropped	.88	Factor Not Present	.85
NONWHITE			.75					
BORNOUTS								.80
UNMPLYED	.34		.57					
WLFREFAM	.84		.80					
PVRTYFAM	.92			.90				

* NM(+) = Nonmetropolitan SEAs with Positive Migration Efficiency Rates
NM(-) = Nonmetropolitan SEAs with Negative Migration Efficiency Rates
M(+) = Metropolitan SEAs with Positive Migration Efficiency Rates
M(-) = Metropolitan SEAs with Negative Migration Efficiency Rates

TABLE 4. ZERO-ORDER CORRELATION COEFFICIENTS FOR FACTOR SCALES, FOR SEA GROUPS

Factor Scale	M(+) Above Diagonal, NM(+) Below Diagonal			
	High SES	Young Families	Poverty Families	Agricultural Employment
High SES	---	.12	-.34	-.05
Young Families	.06	---	.05	-.24
Poverty Families	-.21	.05	---	-.22
Agricultural Employment	.05	-.26	.23	---

	M(-) Above Diagonal, NM(-) Below Diagonal				
	Female Employment	High SES	Agricultural Employment	Young Families	Institutionalized Population
Female Employment	---	.14	.05	.10	.24
High SES	.18	---	-.38	-.13	.35
Agricultural Employment	.20	-.29	---	.30	-.22
Young Families	.13	.05	-.03	---	-.01
Institutionalized Population	.19	.25	.03	-.01	---

families in the SEA increases. The proportion of young or poor families has a marginally greater association with migration efficiency than either of the other two factor scales. Thus, a change of one standard deviation unit in the young or poor family factors is associated with a decrease of .36 and .38 standard deviation units in the efficiency rate, compared to an increase of .24 and .27 deviation units in the efficiency rate when the high SES or agricultural employment factors increase by one unit.

It is suggested that the presence of large numbers of either poor or young families decreases migration efficiency precisely because it is these family types that are migration-prone,

TABLE 5. BETA COEFFICIENTS AND COEFFICIENTS OF MULTIPLE DETERMINATION FOR TWENTY-TWO INDEPENDENT VARIABLES, BY SEA GROUP

Independent variables	Beta Coefficients for Independent Variables M(+)*	NM(+)	NM(-)@	M(-)@
POP17UND	-2.82	1.73	-1.01	1.19
POP65OVR	.06	1.81	-.51	-1.19
DORMIESS	-1.32	.34	-.42	-.31
CHILDFAM	1.29	-.22	-.49	-1.15
NONWHITE	-.24	.46	.00	.63
URBANPOP	-.92	.48	.27	-.82
BORNOUTS	**	-.08	.08	**
MEDYEDML	-1.07	.06	-.51	**
MEDYEDFL	.32	-1.09	.17	-.57
CHLDBORN	1.42	-.10	1.18	.20
LFPRMALE	1.44	-.30	.51	-.36
LFPRFALE	.19	.41	-.29	.22
UNMPLYED	.79	-.51	.36	.48
LFPRWCU6	.36	-.68	-.17	-.64
AGRCULTR	-1.01	1.37	-.46	-.44
BLUECLAR	-3.41	3.84	-.22	-2.98
WHITECLR	-2.80	3.51	-.05	-2.04
FEDERALS	**	.80	-.30	.22
MEDFAMIN	-.57	-1.19	-.89	-2.56
PERCAPIN	.72	.17	.87	1.09
WLFREFAM	-.99	-.18	-.05	-.43
R^2	.91	.85	.84	.94
$\hat{R}^2$	.55	.62	.72	.76
N	27	37	52	27

** Indicates independent variables not included in regression analysis due to insufficient tolerance level in computations.

TABLE 6. BETA COEFFICIENTS AND COEFFICIENTS OF MULTIPLE DETERMINATION FOR FACTOR SCALES, BY SEA GROUP

Factor Scales	Beta Coefficients for Factor Scales M(+)*	NM(+)	M(-)@	NM(-)@
High SES	.24	.48	.14	-.29
Agricultural Employment	.27	.06	-.09	.22
Young Families	-.36	-.26	-.31	-.05
Poverty Families	-.38	-.11	.45	**
Institutionalized Population	**	**	-.25	-.28
Female Employment	**	**	-.24	-.43
R^2	.54	.34	.26	.53
$\hat{R}^2$	.46	.26	.04	.48

** No factor scale constructed.
@ Beta coefficients multiplied by -1.

* M(+) = Metropolitan SEAs with Positive Migration Efficiency Rates
NM(+) = Nonmetropolitan SEAs with Positive Migration Efficiency Rates
M(-) = Metropolitan SEAs with Negative Migration Efficiency Rates
NM(-) = Nonmetropolitan SEAs with Negative Migration Efficiency Rates

and further that these family types have a relatively high turnover rate, moving both into and out of metropolitan SEAs in large numbers. The presence of high SES families and high proportions of agricultural workers within the SEA is seen on the other hand to promote efficient migration. It is probable that metropolitan SEAs are, in fact, attracting large numbers of high SES families, and losing relatively few. The positive association with agricultural employment is less explicable, although one possibility is that agricultural employment levels act as a proxy for an areas suburbanization potential--that is, those metropolitan SEAs with high agricultural employment are those that included substantial rural areas in 1960, providing a necessary precondition for future suburbanization.

The situation just dipicted also holds, with minor exceptions, for nonmetropolitan SEAs with net inmigration. For these SEAs the high SES factor is strongly associated with migration efficiency, while the agricultural employment factor and the poor families factor have weaker associations.

The relationship between factor scales and migration efficiency rates for nonmetropolitan SEAs with net outmigration is quite complex. For this group of SEAs the agricultural employment factor was positively associated with the efficiency rate while all other beta coefficients were negative. It is especially interesting to note the negative association for the high SES factor. A possible explanation of this finding is that the outmigration stream tends to be composed of young families with relatively high SES characteristics, whereas the inmigration stream tends to be composed of older families with relatively high SES characteristics. This hypothesis is supported (but by no means proven) when we look at the age distribution of the migrants moving into and out of rural SEAs. The mean percent of outmigrants between the ages of 20 and 34 is 44 percent, compared to a mean percent of 37 in this age category for inmigrants. Individuals over age 65 constitute 4.1 percent of all outmigrants, but 5.0 percent of all inmigrants. Thus, the age distribution of the in and outmigrants is in the expected direction. A more rigorous test of this hypothesis would be to relate SES characteristics and age for both in and outmigrants, but this data is not available.

REFERENCES

Blalock, H. M., Jr., "Correlated Independent Variables: The Problem of Multicollinearity," Social Forces, 42, 2: 223-237, 1963.

_________, Social Statistics, New York, 1963.

Galle, Omer R. and Max W. Williams, "Metropolitan Migration Efficiency," Demography, 9, 4: 655-664, 1972.

Shryock, Henry S., Jr., Population Mobility Within the United States, Chicago: University of Chicago Community and Family Study Center, 1964.

U.S. Bureau of the Census, State Economic Areas, Final Report, PC(2)-10B, 1972.

_________, Migration Between State Economic Areas, Final Report PC(2)-2E, 1972.