ANDREI ROGERS

A Regression Analysis of Interregional Migration in California

Center for Planning & Development Research: Institute of Urban & Regional Development
University of California, Berkeley

Reprint No. 29

The Center for Planning and Development Research is a unit within the Institute of Urban and Regional Development. It is devoted to the study of a wide range of activities in environmental design and urban and regional development. Researchers in the Center are engaged in investigating the processes of urban and regional growth, the consequences of alternative patterns of growth, and the scientific bases for development policy and planning.

Toward these ends, current research is directed to improving theories and methods of planning, to building models capable of simulating urban growth processes, and to formulating improved criteria for urban and regional development policies. Fields of interest to the Center include the development of new and established communities, housing and renewal programming, the effects of technological development on cities and urban life, metropolitan planning, the economics of city-building in the United States and in developing nations, and the social planning and human development issues that accompany urbanization.

The Center contains the Architectural Design Laboratory, which is devoted to research in building technology, the behavioral bases for design, and design processes.

REPRINT NO. 29

Reprinted from
THE REVIEW OF ECONOMICS AND STATISTICS
Vol. XLIX, No. 2, May 1967

A REGRESSION ANALYSIS OF INTERREGIONAL MIGRATION IN CALIFORNIA

Andrei Rogers *

Economists repeatedly have asserted that the labor market allocates workers more or less where they belong in terms of their productive ability. Comparative economic opportunity, they argue, is the driving motivational force which is manifested in interregional migration patterns. According to this thesis, employment, occupation and salary are the major considerations in any decision to move. In short, internal migration is held to be an important way by which people respond to changing economic opportunities and thereby redirect the spatial allocation of labor toward a more optimal pattern.

In this paper, we attempt to test this hypothesis by measuring statistically the degree of association between spatial variations in economic opportunities and variations in interregional migration flows in California. The migration data used for empirical testing are the United States Census 1955–1960 inter-state economic area (SEA) flows.[1] The technique of analysis is multiple regression.

Migration Streams and Economic Opportunity

Several models have been suggested for tests of the economic "push-pull" hypothesis. Of those currently available, the Lowry model appears to be conceptually most satisfactory.[2] This model is of the following form:

$$M_{ij} = k\left[\frac{U_i}{U_j}\cdot\frac{W_j}{W_i}\cdot\frac{L_iL_j}{D_{ij}}\right]$$

or in its log-transformed form:

$$\ln M_{ij} = b_0 + b_1 \ln U_i + b_2 \ln U_j + b_3 \ln W_i + b_4 \ln W_j + b_5 \ln L_i + b_6 \ln L_j + b_7 \ln D_{ij} \quad (1)$$

where

M_{ij} = number of migrants from i to j;

L_i, L_j = number of persons in the nonagricultural labor force at i and j, respectively;

U_i, U_j = unemployment as a percentage of the civilian, nonagricultural labor force at i and j, respectively;

W_i, W_j = hourly manufacturing wage (in dollars and cents) at i and j, respectively, and;

D_{ij} = airline distance (in miles) separating i and j.

For testing the statistical fit of this model to inter-standard metropolitan statistical area (SMSA) flows in California a few modifications have been made in the operational definitions of the variables.[3] We feel that these changes are desirable in

* The author wishes to acknowledge the invaluable services rendered by his research assistant, Gary Kane, who programmed the regression analyses which are analyzed in this paper and together with Howard Cherniack performed most of the data collection and processing chores.

[1] U.S. Bureau of the Census (PC(2)–2B), *U.S. Census of Population, 1960, Mobility for States and State Economic Areas* (U.S. Bureau of the Census, Department of Commerce, 1963).

[2] Ira S. Lowry, "A Model of Labor-Force Migration: A Working Paper" (unpublished manuscript, Institute of Government and Public Affairs, University of California at Los Angeles, May 1964).

[3] Data for Santa Barbara were incomplete and therefore flows from and to that SMSA are not included in the analysis. This reduces the number of observations to 72.

view of the significant role that agriculture holds in several of the regions analyzed. Thus in our version of Lowry's model:

L_i, L_j = number of persons in the civilian labor force at i and j, respectively; and
U_i, U_j = unemployment as a percentage of the civilian labor force at i and j, respectively.

Another modification is the redefinition of distance. Our D_{ij} is a measure of the shortest highway mileage separating the major county seats at i and j. In intrastate movements, one would expect this to be a more accurate reflection of the "friction" of distance.

Table 1 section A shows the estimated parameters, indicates their significance, and lists the partial correlation coefficients and coefficient of determination for the "modified" Lowry model. The fit of the model is impressive. Over 90 per cent of the variation is accounted for by the seven variables, of which four are significantly different from zero at the five per cent confidence level. Five of the seven variables have the signs which one would expect on a priori grounds. Migration from i to j is directly related to high wages at j and a large civilian labor force at either origin or destination. It is inversely related to high wages at i and increasing distance between i and j. The coefficients of unemployment rates at i and j should, on a priori grounds, be positive and negative, respectively. These are reversed in table 1. Neither is statistically significant, however.

TABLE 1. — REGRESSION STATISTICS FOR THE MODIFIED LOWRY MODEL

A Averages of 1955–1960 Values for Independent Variables			B 1955 Values for Independent Variables		
Variable	Coefficient	Partial Correlation Coefficient	Variable	Coefficient	Partial Correlation Coefficient
Constant	–11.5133		Constant	–12.7369	
log L_i	0.79578[b]	0.87806	log L_i	0.79120[b]	0.87583
log L_j	0.73608[b]	0.86157	log L_j	0.76196[b]	0.86796
log D_{ij}	–0.70688[b]	–0.78583	log D_{ij}	–0.68717[b]	–0.77657
log W_j	5.08408[a]	0.36842	log U_j	1.43190[b]	0.47725
log U_j	1.16108	0.22742	log W_j	3.72478[a]	0.28981
log U_i	–0.58726	–0.11730	log W_i	1.07927	0.08740
log W_i	–1.15274	–0.08949	log U_i	–0.15494	–0.05866
R^2	0.90680		R^2	0.90547	

[a] Significantly different from zero at the five per cent level of significance.
[b] Significantly different from zero at the 0.1 per cent level of significance.

The use of the Lowry model for forecasting purposes presents several problems. First, it requires as inputs average values over the forecasting period for most of the independent variables. In a recursive model, operating with five-year increments, this presents a problem. For example, one would need to specify the average labor force over the next five-year period in order to generate the migrant flows during the same time interval. Given a recursive structure, it is much more convenient to be able to use the outputs of the current time period to generate the flows for the subsequent interval. Thus, in place of average values, we suggest the use of values for previous points in time. More specifically, we propose the use of 1955 status variables to forecast 1955–1960 flows instead of using averages over the five-year period. This modification of Lowry's model results in section B of table 1.

It is immediately apparent that the use of 1955 values for the independent variables in the Lowry model does not seriously affect the "goodness of fit" of the model. The R^2 remains virtually unchanged. Indeed, an additional explanatory variable, the civilian unemployment rate at the destination, achieves significance. However, as before, its sign is the opposite of what one would expect from a priori reasoning.

Another problem which presents itself in any attempt to use the Lowry model as a forecasting tool is the need for forecasts of labor force totals. Within the framework of a cohort-survival projection model it is much more convenient to use the number of people in the 15 to 64 age group as a proxy. This segment of an area's population may be defined as "labor force eligibles." Moreover, to fit the Lowry model to *inter-SEA* flows in California one would require data on hourly manufacturing wage rates in nonmetropolitan subregions. This information is not currently available. As a proxy we suggest the use of per capita wages and salaries. These changes yield a model which we shall refer to as the Lowry-Rogers model:

$$M_{ij} = k\left[\frac{U_i}{U_j} \cdot \frac{WS_j}{WS_i} \cdot \frac{LF_i \cdot LF_j}{D_{ij}}\right]$$

or in its log-transformed form:

$$\ln M_{ij} = b_0 + b_1 \ln U_i + b_2 \ln U_j + b_3 \ln WS_i + b_4 \ln WS_j + b_5 \ln LF_i + b_6 \ln LF_j + b_7 \ln D_{ij} \qquad (2)$$

where

M_{ij} = number of migrants from i to j;
U_i, U_j = civilian unemployment rate at i and j, respectively;
LF_i, LF_j = labor force eligibles at i and j, respectively;
WS_i, WS_j = per capita wages and salaries at i and j, respectively, and;
D_{ij} = shortest highway mileage between the major county seats at i and j, respectively.

Fitting equation (2) to the same set of flows as were fitted by equation (1), we obtain the results shown in table 2. It appears that a higher explanatory power is achieved by the Lowry-Rogers model than by the modified Lowry model. Otherwise, it is interesting to note that the order in which the seven variables enter the step-wise regression program remains unchanged, and the same four variables are significant at the .01 per cent level. Thus one may conclude that the Lowry-Rogers model, with its "proxy" variables is a satisfactory alternative to the Lowry model.

TABLE 2. — REGRESSION STATISTICS FOR THE LOWRY-ROGERS MODEL

Variable	Coefficient	Partial Correlation Coefficient
Constant	−10.8427	
log LF_i	0.88427[a]	0.84628
log LF_j	0.74402[a]	0.80071
log D_{ij}	−0.73903[a]	−0.82692
log U_j	1.15472[a]	0.46413
log WS_j	0.66320	0.20751
log WS_i	−0.56814	−0.17879
log U_i	−0.26044	−0.11737
R^2	0.92306	

[a] Significant at the 0.1 per cent level of significance.

In the three sets of regression statistics presented so far, the coefficients of the unemployment variables at i and j have signs opposite to what one would expect on a priori grounds. Since in two instances the unemployment rate at j is highly significant, this result is disturbing. More distressing, however, is the absence of a reasonable explanation for this finding. Because of the unexplainable behavior of the unemployment variable, we have removed it from the model, leaving us with what we shall refer to as the Rogers model:

$$M_{ij} = k\left[\frac{WS_j}{WS_i} \cdot \frac{LF_i \cdot LF_j}{D_{ij}}\right]$$

or in its log-transformed form:

$$\ln M_{ij} = b_0 + b_1 \ln WS_i + b_2 \ln WS_j + b_3 \ln LF_i + b_4 \ln LF_j + b_5 \ln D_{ij} \qquad (3)$$

where

M_{ij} = number of migrants from i to j;
LF_i, LF_j = labor force eligibles at i and j, respectively;
WS_i, WS_j = per capita wages and salaries at i and j, respectively, and;
D_{ij} = shortest highway mileage between the major county seats at i and j, respectively.

In fitting the Rogers model to inter-SMSA flows (nine SMSA's, 72 observations) and then to the full matrix of inter-SEA flows (19 SEA's, 342 observations), we obtain the regression statistics presented in table 3. These indicate a very slight decline in explanatory power with the R^2 dropping from the previous high of 0.92306 to 0.89930. The extension of the model to include non-SMSA flows results in a further decline of the R^2 to 0.80451. This, however, is a direct result of an increased number of observations and the mixing of metropolitan and nonmetropolitan flows. To see this more clearly let us now turn to an analysis of migration streams differentiated by class of subregion at origin and destination.

TABLE 3. — REGRESSION STATISTICS FOR THE ROGERS MODEL

	A Fitted to 9 x 9 SMSA Flow Matrix		B Fitted to 19 x 19 SEA Flow Matrix	
Variable	Coefficient	Partial Correlation Coefficient	Coefficient	Partial Correlation Coefficient
Constant	−8.85352		−6.22950	
log LF_i	0.90427[a]	0.83042	0.77237[a]	0.67441
log LF_j	0.65533[a]	0.73382	0.68433[a]	0.62908
log D_{ij}	−0.73790[a]	−0.78883	−0.80576[a]	−0.66157
log WS_j	0.69195	0.18996	0.27159	0.06692
log WS_i	−0.57479	−0.15869	−0.27132	−0.06686
R^2	0.89930		.80451	

[a] Significant at the 0.1 per cent level of significance.

TABLE 4. — REGRESSION STATISTICS FOR FLOWS FROM METROPOLITAN SUBREGIONS

A Flows from SMSA to SMSA Subregions (10 x 10)			B Flows from SMSA to Non-SMSA Subregions (10 x 9)		
Variable	Coefficient	Partial Correlation Coefficient	Variable	Coefficient	Partial Correlation Coefficient
Constant	–9.95275		Constant	–15.6364	
log LF_i	0.92951[a]	0.76319	log LF_i	0.84771[a]	0.75224
log LF_j	0.66109[a]	0.64320	log D_{ij}	–0.90221[a]	–0.78558
log D_{ij}	–0.63601[a]	–0.64835	log LF_j	1.25055[a]	0.73883
log WS_j	0.57868	0.12368	log WS_j	1.16687[a]	0.36380
log WS_i	–0.44091	–0.09453	log WS_i	–0.81509	–0.18396
R^2	0.86651		R^2	0.85764	

[a] Significant at the 0.1 per cent level of significance.

TABLE 5. — REGRESSION STATISTICS FOR FLOWS FROM NONMETROPOLITAN SUBREGIONS

A Flows from Non-SMSA to Non-SMSA Subregions (9 x 9)			B Flows from Non-SMSA to SMSA Subregions (9 x 10)		
Variable	Coefficient	Partial Correlation Coefficient	Variable	Coefficient	Partial Correlation Coefficient
Constant	–26.5621		Constant	–8.70549	
log D_{ij}	–1.25719[c]	–0.62095	log LF_j	0.56641[c]	0.56191
log LF_j	1.29292[c]	0.62548	log D_{ij}	–0.77613[c]	–0.69778
log LF_i	0.85610[c]	0.46885	log LF_i	0.77191[c]	0.51609
log WS_j	1.25664[b]	0.32544	log WS_j	0.49537	0.10077
log WS_i	0.90810[a]	0.24136	log WS_i	0.04996	0.01489
R^2	0.78592		R^2	0.76216	

[a] Significant at the five per cent level of significance.
[b] Significant at the one per cent level of significance.
[c] Significant at the 0.1 per cent level of significance.

Migration Streams by Class of Subregion at Origin and Destination

Demographers have found considerable empirical support for the thesis that factors determining out-migration tend to vary with the type of destination as well as with the type of origin.[4] As a limited test of this hypothesis, we report in tables 4 and 5, the fit of the Rogers model to the 1955–1960 flow matrix, partitioned by metropolitan and nonmetropolitan origins and destinations.

Several interesting findings are indicated by the regression statistics. First, it is clear that the model provides a better accounting of flows originating from metropolitan subregions. This may be a reflection of the higher quality of information concerning economic opportunities that exists in metropolitan areas.[5] Second, per capita wages and salaries are significant only in flows to nonmetropolitan subregions. This suggests that people will not move to nonmetropolitan areas unless a definite improvement in per capita earnings is anticipated. Finally, it is encouraging to note that all five variables have signs which one would expect from a priori considerations. Migration from i to j is directly proportional to per capita wages at j and the product of the "masses" at i and j (as measured by labor force eligibles). It is inversely proportional to per capita wages and salaries at i and to the distance separating i and j.

Migration Streams by Class of Migrant

Having analyzed the statistical correlation between interregional migration and economic opportunity, it is appropriate to consider the effects of differentiating streams by color, age, and sex.[6]

[4] See, for example, Donald J. Bogue, Henry S. Shryock, and Siegfried A. Hoermann, *Subregional Migration in the United States: 1935–40*, I (Oxford, Ohio: Scripps Foundation Studies in Population Distribution, Miami University, 1957).

[5] See Phillip Nelson, "Migration, Real Income and Information," *Journal of Regional Science*, 1 (Spring 1959), 43–62.

[6] Color-specific flows are reported in *U.S. Census of Population, 1960, Mobility for States and State Economic Areas* (U.S. Bureau of the Census, Department of Commerce, 1963) Age-specific flows were estimated for the study by Andrei Rogers, *Projected Population Growth in California Regions: 1960–1980* (Preprint No. 12, Center for

TABLE 6. — REGRESSION STATISTICS FOR WHITE AND NON-WHITE FLOWS

A White Flows			B Non-White Flows		
Variable	Coefficient	Partial Correlation Coefficient	Variable	Coefficient	Partial Correlation Coefficient
Constant	−6.16275		Constant	−30.6021	
log LF_i	0.77124[a]	0.67062	log LF_j	2.63083[a]	0.29365
log LF_j	0.67496[a]	0.62046	log LF_i	1.83355[a]	0.20935
log D_{ij}	−0.80346[a]	−0.65721	log D_{ij}	−2.55513[a]	−0.26627
log WS_j	0.28624	0.06989	log WS_j	−1.65939	−0.04043
log WS_i	−0.28436	−0.06944	log WS_i	0.45730	0.01115
R^2	0.79983		R^2	0.26841	

[a] Significant at the 0.1 per cent level of significance.

Thus, we report below the results of introducing class-specific flows into the model defined in equation (3).

Several observations are suggested by the regression statistics presented in tables 6 through 10. First, it appears that variations in interregional non-white flows are very weakly tied to spatial variations in economic opportunities, as reflected by per capita wages and salaries. Table 6 indicates that the factors motivating non-white migration are strikingly different from those inducing white movements and, moreover, are probably largely noneconomic in character. Whereas the Rogers model accounts for 80 per cent of the total variance in the case of white flows, it explains only 27 per cent of the total variation in non-white movements.

TABLE 7. — COEFFICIENTS OF DETERMINATION FOR THE 17 AGE-SPECIFIC FLOW MATRICES

Age Group (years)	R^2
0–4	0.79626
5–9	0.78813
10–14	0.79619
15–19	0.80898
20–24	0.81940
25–29	0.81154
30–34	0.81196
35–39	0.80227
40–44	0.78311
45–49	0.51676
50–54	0.78383
55–59	0.52611
60–64	0.53807
65–69	0.45121
70–74	0.37958
75–79	0.29654
80–84	0.37596

Ever since Dorothy Thomas's exhaustive findings on migration differentials, published almost 30 years ago, demographers have consistently asserted that young adults are the most mobile segment of the population. According to statistics in tables 7, 8, and 9, young adults also are the age group most responsive to spatial variations in economic opportunities. Their flows produce the highest R^2's in the Rogers model.

Finally, complementing the demographer's hypothesis that males are more mobile than females, we find strong indications in table 10 that the economists' simple "economic opportunities" hypothesis is an inadequate framework with which to account for female migration. The Rogers model provides a satisfactory explanation of male movements but accounts for less than 45 per cent of the total variation in female flows.

This paper is a progress report on the development of a simple regression model of interregional migration in California. The statistics presented in tables 1 through 10 represent a modest attempt to assess the significance and relative weight of a set of economic variables and their differential impact on class-specific migrant streams. When considered together the results are promising. In almost every instance, variations in total flows have been accounted for by a small subset of variables. On the basis of these regression results we may reasonably conclude that comparative economic opportunity is a factor which significantly influences interregional movements in California. Moreover, the results for differentiated streams are equally encouraging. The relatively well-established hypotheses of demographers concerning variations in behavior among classes of migrants appear again in the differential response of these migrant classes to economic opportunities. Thus we find that those classes which are more mobile also tend to be more sensitive to spatial variations in economic conditions.

Planning and Development Research, University of California, Dec. 1965), 15–17. Sex-specific flows were obtained from special tabulations of a two per cent sample of reported changes of address in drivers' license registrations, supplied by the State Department of Motor Vehicles.

TABLE 8. — REGRESSION STATISTICS FOR SELECTED AGE GROUP FLOWS

A 5–9 years			B 20–24 years		
Variable	Coefficient	Partial Correlation Coefficient	Variable	Coefficient	Partial Correlation Coefficient
Constant	–6.99401		Constant	–10.4961	
log LF_j	0.68531[b]	0.61956	log LF_i	0.80592[b]	0.69428
log LF_i	0.74590[b]	0.65163	log LF_j	0.68350[b]	0.63322
log D_{ij}	–0.80669[b]	–0.65213	log D_{ij}	–0.81398[b]	–0.66982
log WS_i	–0.44112[a]	–0.10551	log WS_j	0.27630	0.06890
log WS_j	0.27036	0.06489	log WS_i	–0.03608	–0.00902
R^2	0.78813		R^2	0.81940	

[a] Significant at the five per cent level of significance.
[b] Significant at the 0.1 per cent level of significance.

TABLE 9. — REGRESSION STATISTICS FOR SELECTED AGE GROUP FLOWS

A 40–44 years			B 75–79 years		
Variable	Coefficient	Partial Correlation Coefficient	Variable	Coefficient	Partial Correlation Coefficient
Constant	–7.82173		Constant	–6.66146	
log LF_j	0.68484[b]	0.61187	log LF_i	1.60651[b]	0.34613
log LF_i	0.75827[b]	0.65052	log LF_j	1.28519[b]	0.28307
log D_{ij}	–0.80202[b]	–0.64264	log D_{ij}	–1.09556[b]	–0.22687
log WS_i	–0.44236[a]	–0.10379	log WS_i	–2.53454[a]	–0.12066
log WS_j	0.26954	0.06345	log WS_j	–0.50414	–0.02417
R^2	0.78311		R^2	0.29654	

[a] Significant at the five per cent level of significance.
[b] Significant at the 0.1 per cent level of significance.

TABLE 10. — REGRESSION STATISTICS FOR MALE AND FEMALE FLOWS

A Male Flows			B Female Flows		
Variable	Coefficient	Partial Correlation Coefficient	Variable	Coefficient	Partial Correlation Coefficient
Constant	–11.0497		Constant	–13.0820	
log LF_i	0.74863[a]	0.54917	log LF_j	0.59556[a]	0.31637
log LF_j	0.71096[a]	0.52301	log LF_i	0.73050[a]	0.37861
log D_{ij}	–0.74373[a]	–0.37681	log D_{ij}	–0.69645[a]	–0.34358
log WS_i	–0.14228	0.03140	log WS_j	0.73578	0.07711
log WS_j	0.06331	0.01145	log WS_i	–0.46286	–0.04860
R^2	0.77755		R^2	0.44484	

[a] Significant at the 0.1 per cent level of significance.