AgEcon SEARCH
RESEARCH IN AGRICULTURAL & APPLIED ECONOMICS

The World's Largest Open Access Agricultural & Applied Economics Digital Library

This document is discoverable and free to researchers across the globe due to the work of AgEcon Search.

Help ensure our sustainability.

Give to AgEcon Search

AgEcon Search
http://ageconsearch.umn.edu
aesearch@umn.edu

*Papers downloaded from **AgEcon Search** may be used for non-commercial purposes and personal study only. No other use, including posting to another Internet site, is permitted without permission from the copyright owner (not AgEcon Search), or as allowed under the provisions of Fair Use, U.S. Copyright Act, Title 17 U.S.C.*

RSP 10(2): 32-42. ©1980 MCRSA. All rights reserved.

DETERMINANTS OF INTERINDUSTRY AND INTERREGIONAL MIGRATION*

Robert A. Nakosteen and Michael A. Zimmer*

Introduction

A sizable body of literature exists as evidence of interest among economists in determinants and consequences of labor mobility. This interest is generated in part by the role of migration in restoring wage and price equilibria among spatially dispersed markets. Traditionally, interest has focused primarily on regional migration, while relatively few studies have examined migration among industries.[1] Surprisingly little attention has been paid to the joint nature of these two aspects of migration, in spite of tentative indications that they are significantly related. Gallaway [4], for example, infers that emvloyed persons who migrate regions are more likely than nonmigrants to change industries of employment, and the two groups exhibit noticeable income differentials. More recently, studies by Bartel [1] and DaVanzo [2] have examined the role of unemployment status on locational choice. In each case estimates of migration decision equations based on micro data confirm the hypothesis that job mobility plays a significant role in locational choice.

The objective of this study is to examine joint determinants of interregional and interindustry migration. It is based on a large set of individual administrative data taken from the Social Security Administration's 1 Percent Continuous Work History Sample. The sample contains longitudinal data on employed persons, describing state of residence, industry of employment, and earnings, as well as limited demographic information. These data make it possible to observe each individual's labor force behavior through time. In particular, each individual may be observed at two distinct points in time, and may be categorized into one of four mutually exclusive classes of migrant behavior; each person is observed to have changed region while remaining in the same industry, changed industry while remaining in the same region, changed neither region nor industry, or changed both. The purpose of this paper is to determine whether available data are capable of systematically differentiating among these classifications. Empirical testing is based on discriminant analysis of the sample data. Estimation of the linear discriminant function permits delineation of explanatory variables which significantly separate the various classifications. In addition, the predictive power of the estimated discriminant function may be assessed by comparing actual sample classifications with predicted classifications generated by the estimated function. Effective separation of the classifications may be interpreted as evidence of interdependence between interregional and interindustry migration. This paper, therefore, concerns itself with empirical results from the discriminant

*Certain data used in this study were derived from computer tapes furnished by the Social Security Administration. The authors did not at any time have access to nor did they receive any information relating to specific individuals or reporting units.

**Economist, Tennessee Valley Authority and Assistant Professor, Department of Economics, University of Evansville, respectively.

[1]A survey of studies in migration is found in Greenwood [7].

model. Its remaining sections are organized as follows. Section two briefly describes the background of the study, while section three details the sample and defines variables used in the model. Section four describes the discriminant model technique used in the study. Section five presents results of estimation, and section six summarizes and concludes the study.

Background

Recent research efforts in migration have conformed to the general framework of Sjaastad [14] in viewing migration as one means of investing in human capital. In this framework it is assumed that potential migrants behave as though they seek to maximize the present value of net gains resulting from locational change. The individual's objective function reflects a regional earnings differential and the direct costs attendant to moving. In cases where the discounted earnings gain exceeds moving costs, the individual responds by migrating to the more attractive region; otherwise no locational change occurs.

A variation on the theme is suggested by Gallaway [3], whose analysis of interindustry mobility rests on the traditional microfoundations of labor supply; the individual is assumed to possess a well-ordered utility function, with leisure and real income as arguments, which is maximized subject to a time constraint. Individual labor force behavior implied by the model entails migration to industries in which real wages are highest in relation to corresponding total costs of employment.

The implication of these remarks is that models which purport to explain migration decisions necessarily include appropriate measures of the costs and gains of location or industry change. As a consequence, considerable empirical research has focused on identifying the appropriate explanatory variables. One of the most widely held notions is that the propensity to migrate declines with age. In this respect age is viewed as one index of the cost of migration. The deterrent effect of age on regional migration is recently documented, for example, by Polachek and Horvath [13], and its role in interindustry mobility is examined by Gallaway [3].

In addition to age, race and sex of potential migrants are hypothesized to have significant impacts on mobility. There is, however, considerable ambiguity in the literature regarding the effect of race on migration. Greenwood [6], for example, reports that non-whites are more responsive than whites to opportunities for income growth, while Persky and Kain [12] suggest that whites are more responsive to the availability of job opportunities. Nakosteen and Zimmer [10], on the other hand, report that race has an insignificant impact on the probability of regional migration.

The sex of potential migrants assumes a significant role in view of the effect of family ties on the mobility of females. This point is discussed by Mincer [9], who reports that households including employed females are less likely to migrate regions. Furthermore, married female migrants have lower labor force paticipation rates at destination than origin localities.

Attempts to measure gains from migration customarily focus on growth in employment opportunities and income in the origin or potential destination. A region with rapidly growing employment opportunities is likely to experience a net inflow of migrants from less attractive areas. Although growth in income in a particular region would seem to exert a similar deterrent effect on outmigration, its actual impact is the subject of some debate. For example, Vanderkamp

[15] and Miller [8] present evidence suggesting that workers in regions with high or growing income are more likely to migrate. These results could be interpreted to mean that rapid growth in income is used to finance migration. In any case, it is clear that variables describing general income or employment opportunities are appropriate in models of interregional or interindustry mobility.

Description of Sample and Data

One feature of recent empirical studies in migration appears to be the increasing reliance on individual microdata to estimate models describing the decision to migrate and returns to migration. Examples include Navratil and Doyle [11], Polachek and Horvath [13], Nakosteen and Zimmer [10], Bartel [1], and Da Vanzo [2]. The present study continues along these lines, utilizing a subset of observations from the Social Security Administration's 1 Percent Continuous Work History Sample (CWHS). The data are drawn from a random selection of 9,223 employed persons for whom earnings records are available for both 1971 and 1973 (the assumed decision-to-migrate interval). For each individual the data furnish information on earnings, age, race, sex, state of principal employment, industry of principal employment (measured at the two-digit SIC level), and a dummy variable (SE) which is one for self-employed persons and zero otherwise. In addition, such regional variables as rate of growth of state employment (GEMR) and state per capita income growth (GPCI), and such industry variables as national rate of growth of industry employment (GSIC) and growth in average hours worked per employee (GAVW) have been added to each record in the sample. The variables GEMR and GPCI are measured in terms of the origin (1971) state of employment. Industry variables GSIC and GAVW are measured in terms of the 1971 industry of employment.[2] Measures of earnings consist of 1971 earnings (Y71) and rate of growth in earnings between 1971 and 1973 (GY). Migrant status is defined as follows: the individual is a regional (industry) migrant if state (industry) of principal employment changes between 1971 and 1973; he is, in each case, a non-migrant otherwise.

Use of the CWHS as a data base introduces certain shortcomings into the analysis. In particular it is not possible to account for the effects of education, labor force behavior of other household members, stability of the family unit, and the presence of children on the decision to migrate. These disadvanatages are partially offset by the presumably high degree of accuracy in reported earnings and other data, and by the very large sample sizes which are facilitated by the CWHS. Indeed, the CWHS is necessary for studies of the type discussed here because it permits samples large enough to generate enough observations, in each of the four migrant categories, for reliable estimation and inference.

The Discriminant Model

The issue of interdependence between regional migration and employment status has recently been addressed by Bartel [1] and DaVanzo [2]. Each study

[2]In studies of the present kind where individual administrative data are used, substantial empirical and conceptual problems are introduced by the inclusion of destination rates of growth. The same is true of the distance variable so commonly used as a proxy measure of costs of migration. With micro data distance assumes a value of zero for the majority of observations (non-migrants) in the sample. Consequently, in this study only origin rates of growth are used, and distance is not included among the explanatory variables.

presents estimates of alternative models of the decision to migrate in which the migration decision, represented by a dichotomous dependent variable, is a linear function of selected explanatory variables. Estimates of the coefficients and their standard errors are then used to make inferences concerning the effects of employment status on migration.

The approach taken in this study, as described earlier, is to view each sample observation as falling into one of four mutually exclusive classes of migrant behavior. Discriminant analysis, applied to the four classes of data, permits examination of the extent to which the data significantly separate the migrant categories. Significant separation is offered as evidence of dependence between industry and regional mobility. For example, if the data distinguish those who migrate regions but not industries from those who migrate both regions and industries, then it can be inferred that the population of regional migrants is partitioned with respect to industry migration. It may then be argued that the strong distinguishing feature of industry migration is too important to be ignored in models of locational choice. It should be noted that the discriminant model is not intended to depict the migration decision, but rather to determine whether a set of explanatory variables exists which significantly distinguishes among the migrant categories.

The purpose of discriminant analysis is to construct a linear composite of explanatory variables, creating a one-dimensional index for purposes of classifying sample observations:

$$(1) \qquad I_i = k_0 + k_1 X_{1i} + k_2 X_{2i} + \ldots + k_p X_{pi} .$$

Estimates of the discriminant weights k_j, $j = 0, \ldots, p$, are based on the observed values of the explanatory variables as well as the actual classification of each sample observation. In order to describe the procedure, define the following:

G = number of groups

N = total sample size

n = number of observations in each sample group, assumed to be equal for ease of exposition

e_n = n-dimensional column vector of ones

e_N = N-dimensional column vector of ones

H = $e_n \times I_G$, where I_G is a G-dimensional identity matrix and x denotes the Kronecker product

X = matrix of sample observations on the explanatory variables, of dimension $N \times (P + 1)$, where $P + 1$ is the number of discriminant weights to be estimated.

The computation proceeds by first computing the matrix of sample means by group, $\bar{X}_g$ and the vector of overall sample means, $\bar{X}$:

(2) $\bar{X}_g = (H' H)^{-1} H' X$

(3) $\bar{X}' = \bar{N}^{1} e'_n X$.

These are used to form the matrix of within-group deviations, P, and the matrix of between-group deviations, Q:

(4) $P = X - H\bar{X}_g$

(5) $Q = H\bar{X}_g - e_n \bar{X}^{1}$.

The results of (4) and (5) are used to form the within-groups and between-groups cross products matrices, W and B respectively:

(6) $W = P' P$

(7) $B = Q' Q$

Finally, the total sample cross products matrix is the sum of (6) and (7):

(8) $T = W + B$

The vector of discriminant weights is obtained by selecting values of k which minimize the ratio:

$$\lambda = \frac{k' B k}{k' W k}$$

Differentiating with respect to k and setting equal to zero yields the characteristic equation:

$$(W^{-1} B - \lambda I) k = 0,$$

leading to a solution for k.

The stepwise descriminant routine proceeds in a manner analogous to conventional stepwise regression programs. At each step in the procedure explanatory variables are chosen for inclusion in the model on the basis of their contribution to separation of the sample groups: variables currently included in the model may be deleted if their contributions are rendered insignificant by the entry of other variables. The F-statistics upon which these changes are based provide a convenient means of testing for the significance of individual explanatory variables.

Each step in the procedure results in a matrix of F-statistics, with one statistic corresponding to each distinct pair of groups. Each may be used to

test the null hypothesis that the variables thus far included in the model fail to significantly separate the two groups in question. The pairwise F's are based on Mahalanobis' D^2 statistic; for groups a and b, with observations n_a and n_b, respectively, D^2 is given by

$$(\bar{X}_a - \bar{X}_b)' C_w^{-1} (\bar{X}_a - X_b),$$

where $\bar{X}_a$ and $\bar{X}_b$ represent the respective vectors of sample means, and C_w is the pooled within-groups covariance matrix. A simple transformation of D^2, given by

$$[n_a n_b (n_a + n_b - p-1) \;/\; p(n_a + n_b)(n_a + n_b - 2)] \cdot D^2$$

results in a variate which is distributed as F with degrees of freedom (p, $n_a + n_b$-1), under the null hypothesis that groups a and b are not significantly separated. Consequently, the pairwise F-statistics yield information on which combinations of groups are separated by the data.

The final step in the procedure results in a classification matrix of the sample data. Sample observations, for which actual classifications are known, are assigned to groups on the basis of estimated classification functions. Comparison of the predicted and actual classifications provides additional information on the validity of the discriminant model. A model which "fits" the data well is expected to exhibit a satisfactory degree of accuracy in classification.

A point of principal interest in this study concerns the dimensionality of the discriminant space. As a general proposition, in multiple discriminant analysis involving G groups and n explanatory variables, the *maximum* number of discriminant functions obtainable is the smaller of G-1 and n. It may happen that fewer discriminant functions are sufficient to effectively separate the sample observations. In such a case at least one group displays substantial overlap with another group, so that the two could not be viewed as distinct. In this study, as noted previously, the sample is partitioned into four groups and it is hypothesized that the groups are distinct in the sense that they can be significantly separated. Thus it is of interest to determine the number of discriminant functions obtained in the estimation procedure. Significant separation obviously requires that at least one discriminant function be obtained; if the maximum of three functions is obtained, then each group may be viewed as significantly separated from every other group.

Estimation of the model provides a basis for determining the appropriate number of discriminant functions. As the functions are obtained in sequence, successive values of the test statistic, Wilks' lambda, measure the inverse of discriminating power contained in the original explanatory variables but not yet removed by existing discriminant functions:

$$\Lambda = |W|/|T| .$$

A simple transformation of Λ results in Chi-square variate, useful in testing the null hypothesis that the most recently obtained discriminant function does not significantly improve group separation.

Empirical Analysis

Results described in this section were obtained by means of a stepwise discriminant program, utilizing the various measures outlined in the previous section as explanatory variables. The purpose of this section is to determine the extent to which the explanatory variables discriminate among the four sample groups, and to assess the capacity of the model for accurately classifying the sample observations.

As noted previously, data used in this study consist of 9,223 observations for the years 1971 and 1973. Sample means of selected explanatory variables along with definitions of the four sample groups are presented in Table 1. Not surprisingly, the majority of sample observations occurs in group 1, consisting of nonmigrants. The smallest number of observations is found in group 2, consisting of persons who migrate regions while remaining in the same industry. The sample means indicate that industry migrants (groups 3 and 4) display the lowest earnings levels along with the highest earnings growth. Also, as expected, migrants display lower age averages than nonmigrants.

TABLE 1. Selected Sample Means*

Variable	Group 1	Group 2	Group 3	Group 4
Y71	6155.39	7269.04	4236.93	5000.21
GY	.163	.235	.516	.472
AGE	41.5	35.6	32.1	30.6
GEMR	.069	.079	.074	.075
GPCI	.041	.042	.037	.041
GSIC	.051	.048	.053	.052
GAVW	.017	.016	.016	.017
Sample Size	5622	267	2523	811

***Group Definitions**
Group 1: changed neither region nor industry
Group 2: changed region but not industry
Group 3: changed industry but not region
Group 4: changed both industry and region

Results of the stepwise discriminant analysis are presented in Table 2. As noted previously, the final-step statistics result from sequential selection of explanatory variables on the basis of their contributions to group separation. In this study the selection criterion at each step is to choose the explanatory variable which maximizes the smallest F- statistic between pairs of groups. Table 2 presents standardized coefficients of three discriminant functions obtained through this procedure. They are obtained by transforming the original data to standard form (zero mean, unit standard deviation), and are analogous to beta coefficients often cited in applied regression analysis. Use of standardized coefficients permits assessment of the relative importance of the explanatory variables in group separation. From the table it can be seen that all three functions reflect the importance of earnings. In addition, the first function indicates an important role for age, the second function reflects the influence of the self-employment variable, and the third shows the effects of

TABLE 2. Standardized Discriminant Function Coefficients and Tests of Significance.

Variable	F*	Function 1	Function 2	Function 3
SE	9.52	.017	.512	.615
GY	3.40	.068	.187	.272
GEMR	10.07	.121	-.258	.734
GPCI	3.73	-.051	-.235	-.507
AGE	223.21	-.689	.334	-.200
SEX	26.99	-.254	.123	.151
GSIC	4.80	.035	-.347	-.138
GAVW	4.57	-.080	.176	-.322
Y71	142.33	-.599	-.462	.561
Wilks' Lambda		.833	.987	.998
Chi-Square		1680.9**	121.6**	17.9**

*Approximate critical values: 1.88 (.05); 2.41 (.01)

**Significant at the .01 level

self-employment along with regional growth in employment and per capita income.

Of particular importance in the table are the successive Wilks' lambda statistics and their associated Chi-square values. Each statistic is significant at the approximate 1 percent level, indicating that the sample data justify the maximum of three discriminant functions. In view of previous remarks, this result is interpreted to mean that there exist significant differences among all the migrant categories.

Also presented in the table are F- statistics for individual explanatory variables. They are to be interpreted as testing the null hypothesis that the variable in question does not add significantly to group separation, given the inclusion of all other explanatory variables in the model. Examination of the results reveals that each variable in the table is significant at the approximate 1 percent significance level. In fact, of the original explanatory variables, only the race variable is not significant. These results imply that the variable set used in this study is effective in distinguishing among the categories of migrant behavior.

Additional insight into the separation of groups may be gained by examining Table 3, which contains pairwise F tests among groups. Each statistic tests the null hypothesis that the explanatory variables have identical vectors of means for the two groups in question. The table results show that each pairwise F is significant at the 1 percent level, providing further evidence of the separation of classes of migrant behavior.

A final consideration is the predictive accuracy of the model, summarized in Table 4. Classification is carried out through a series of classification functions, with one function corresponding to each group. The classification function is a linear combination of the explanatory variables; each sample

TABLE 3. F Statistics Between Pairs of Groups.

Group	Group 1	2	3
2	9.92*		
3	156.77*	16.99*	
4	69.85*	10.80*	6.51*

*Significant at the .01 level

TABLE 4. Classification Matrix.

Actual Group	Predicted Group 1	2	3	4	Number of Observations
1	5115	0	506	1	5622
2	227	0	40	0	267
3	1576	0	947	0	2523
4	515	0	294	2	811

Percent correctly classified: 65.75.

observation is evaluated for each classification function and is assigned to the group corresponding to the highest "classification score."

Under the assumption that the vector of explanatory variables is multivariate normal with common covariance matrix across groups, the classification scores can be converted to probabilities of membership in each group. An equivalent assignment procedure is then to assign each observation to the most likely group. A Bayesian adjustment of probabilities is made possible by the use of some appropriate measure of prior probability of group membership. The most reasonable set of priors is the group distribution of sample proportions. Accordingly, each observation is classified in the group displaying the highest posterior probability.

Proceeding in this manner, the classification matrix in Table 4, summarizes predicted and actual group membership. It can be seen that the model's predictive capabilities leave much to be desired. In particular, the model fails to assign any observations to group 2, and assigns only three to group 4. Actual members of these groups are erroneously assigned, for the most part, to group 1. Also, only about 38 percent of group 3 observations are correctly assigned, the remainder assigned to group 1.

The overall accuracy of predictions, 66 percent, is largely attributable to the success of the model in correctly identifying nonmigrants (group 1).

The general failure of the model to assign observations to groups 2 and 4 may be due in part to the small prior probabilities associated with those groups (.03 for group 2, .09 for group 4). To check this, computations were repeated using equal prior probabilities. The resulting classification matrix showed a

more uniform distribution of predictions although, as expected, its overall predictive accuracy was substantially lower than that of Table 4.

Conclusions

This study presents results of discriminant analysis of migrant behavior by employed wage earners between regions and industries. Its principal conclusion is that all four classes of migrant behavior are significantly separated by explanatory variables commonly found in models embodying the human capital approach to migration: earnings and earnings growth, regional and industrial employment and income opportunities, age, sex, and self-employment status. Tests of significance of Wilks' lambda indicate that the maximum number of three discriminant functions is necessary to separate the sample groups, further implying distinct differences among all classes of migrant behavior.

The major shortcoming in the study is the poor predictive performance of the discriminant model, which displays a bias in favor of predicting observations to be nonmigrants. It is likely that improvements could be attained by the inclusion of additional data. Missing variables include education, attributes of other household members, and presence and age of children in the household. Unfortunately, these are not available in the CWHS. Increasing utilization and sophistication of panel data, however, offer possibilities for further refinement of human capital models of interregional and interindustry migration. In any case, available evidence indicates that researchers should remain cognizant of the interdependence between these two dimensions of labor mobility.

REFERENCES

1. Bartel, A. P., "The Migration Decision: What Role Does Job Mobility Play?" *American Economic Review*, 69, (5), December 1979. 775-786.

2. DaVanzo, J., "Does Unemployment Affect Migration? — Evidence From Micro Data," *Review of Economics and Statistics*, LX, (4), November 1978, 504-514.

3. Gallaway, L. E., *Interindustry Labor Mobility in the United States, 1957 to 1960*. Social Security Administration, Office of Research and Statistics, Research Report No. 18, Washington D. C.; U.S.G.P.O, 1967.

4. _____, "The Effect of Geographic Labor Mobility on Income: A Brief Comment," *Journal of Human Resources*, 4, (1), Winter, 1969, 103-109.

5. Green, P. E., *Analyzing Multivariate Data*, Hinsdale, Ill.: Dryden Press, 1978.

6. Greenwood, M. J., "A Simultaneous-Equation Model of White and Non-white Migration and Urban Change," mimeograph, 1973.

7. _____, "Research on Internal Migration in the United States: A Survey," *Journal of Economic Literature*, 13, (2), June, 1975, 397-433.

8. Miller E., "Is Out-Migration Affected by Economic Conditions?" *Southern Economic Journal*, 39, (3), January, 1973, 396-405.

9. Mincer, J., "Family Migration Decisions," *Journal of Political Economy*, 86, (5), October, 1978, 749-773.

10. Nakosteen, R. A., and M. A. Zimmer, "Migration and Income: The Question of Self-Selection," *Southern Economic Journal*, 46, (3), January 1980.

11. Navratil, F. J. and J. J. Doyle, "The Socioeconomic Determinants of Migration and the Level of Migration" *Southern Economic Journal*, 43, (4), April, 1977, 1547-1559.

12. Persky, J. J. and J. F. Kain, "Migration, Employment and Race in the Deep South," *Southern Economic Journal*, 36, (3), January, 1970, 268-276.

13. Polachek, S. W. and F. W. Horvath, "A Life Cycle Approach to Migration: Analysis of the Perspicacious Perigrinator," in *Research in Labor Economics* edited by R. Ehrenberg, Greenwich, Conn.: JAI Press, 1977.

14. Sjaastad, L. "The Costs and Returns of Human Migration," *Journal of Political Economy*, 70, (5), October, 1962, 80-93.

15. Vanderkamp, J., "Migration Flows, Their Determinants and the Effects of Return Migration," *Journal of Political Economy*, 79, (5), September/October, 1971, 1012-1031.

16. _____, "Industrial Mobility: Some Further Results," *Canadian Journal of Economics*, 10, (3), August, 1977, 462-472.