Earnings Expectations and the Role of Human Capital in the Migration Decision: An Empirical Analysis

Nguyen huu Nghiep, Henry W. Herzog Jr. and Alan M. Schlottmann*

INTRODUCTION

In most theories of migration, economic incentives are considered important determinants of mobility. One such theory, which draws upon the human capital literature, views migration as an investment that increases the productivity of human resources. Under this theory, an individual will migrate to some specific location if the difference between the present values of the income streams at that and the current location is both positive and of sufficient magnitude to offset relocation costs. For convenience, this theory will be termed the PVD theory, with PVD standing for "present value difference." On the other hand, an alternative theory views migration as a response to "current income differences" across space—hereafter the CID theory. This study provides an empirical test of the purported superiority of the PVD over the CID theory of migration.[1]

Our initial investigation of the literature on this topic revealed several shortcomings in the manner in which income differentials are determined for tests of the competing hypotheses. These shortcomings are discussed in the following section within the general context of earnings expectations. The claimed superiority of the PVD theory will then be examined after remedying these shortcomings. The theoretical model for this examination is developed in the third section of the paper and is followed by empirical results and conclusions.

EARNINGS EXPECTATIONS AND MIGRATION

The determination of human capital augmentation resulting from (investment in) migration requires information on current and expected future income of the potential migrant at both his present location and alternative future locations. However, previous studies in this area assume that individuals perceive future income as present income inflated by a simple productivity rate (Sjaastad, 1962; Bowman and Myers, 1967). Present income used in this conventional formulation is either average earnings of a population subgroup (Bowles, 1970) or earnings estimated on the

basis of personal characteristics and labor market variables (Kiker and Traynham, 1974, 1977). In addition, a study by Fields (1976) identified several "best" specifications for the determination of human capital proxies; however, it provided little information on either the "superiority" of the PVD theory or the manner in which earnings expectations affect migration.

The second important shortcoming of past studies concerns unobserved data. For a migrant, future income at the origin is unobservable; similarly, future income at potential destinations is unobserved for nonmigrants. These unobserved data must somehow be estimated prior to the computation of both PVD and CID for each individual in the study. Few of the past studies on this topic explicitly describe how to estimate these data. In his study on the costs and returns of human migration, Sjaastad (1962) fails to consider the question of unobserved income, stating simply that there is an increment to one's earnings stream to be obtained by moving to another place. In the Bowles' study (1970), the question of unobserved income is not addressed due, primarily, to the use of aggregate (net migration) data. Although several studies do estimate unobserved income, this determination is often based upon extremely simple techniques (Hansen, 1972; Laber, 1973).

A third general observation on past research in this area is that the majority of studies obtain empirical results from aggregate data rather than from microdata on individuals (see Davanzo, 1980a and 1980b). Only recently have migration studies begun to exploit the rich socio-economic information on individuals available on magnetic computer tapes and the binary dependent variable econometric techniques required for the use of this microdata. In this respect, a recent study by Navratil and Doyle (1977) investigated the sensitivity of empirical estimates to various levels of data aggregation. Their results demonstrate that migration response to personal characteristics is highly sensitive to the level of aggregation.

THE THEORETICAL MODEL

The model developed in this section and the empirical estimation that follows are designed to overcome the shortcomings summarized above and thus to provide a comprehensive test of the PVD and CID theories of migration. Central to this test is the computation of the PVD and CID variables. These variables will be considered by first examining three methods for the formulation of income expectations, each method in turn providing the framework for an independent test of the purported superiority of the PVD theory of migration. Second, unobserved incomes of individual migrants and non-migrants will be estimated, and the results used to compute PVD and CID values for inclusion in the migration equation.[2] Finally, the use of microdata on individuals permits a comprehensive examination of the influence of personal characteristics on the migration decision.

The income variable, X, is of crucial importance to the study since its determination presupposes knowledge of *how* individuals formulate in-

come expectations. In order to provide breadth to the empirical work that follows, three alternative formulations will be considered. Each of these will be discussed in turn, followed by a summary of the estimation procedure for X under the alternative models of income expectation.

1. *Perfect Foresight Formulation.* Under this formulation, individuals are assumed to have perfect foresight about their future income, being aware of and able to forecast future income at potential destinations and at the origin. Accordingly, actual future income will be used in the computation of X. For a future nonresidence, income is unobserved and hence must be estimated.[3]

2. *Group-Assimilating Formulation.* Under this model, each individual is assumed to expect his future income to be similar to that of individuals of the same age, education, race, sex, occupation, and state of residence (i.e., individuals can apply *ceteris paribus* conditions with respect to these variables). Expected incomes for this formulation are determined within an econometric model and provide the basis for the computation of X.

3. *Conventional Formulation.* As in Formulation 2, the individual is assumed to expect his future income to be similar to that of others of similar characteristics. However, the future income of the corresponding population subgroup is derived in this case by taking the simple average of incomes of individuals within the subgroup (individuals do not apply *ceteris paribus* conditions).

Under the perfect foresight formulation, X is computed as follows. For a migrant, future income at the destination corresponds to actual income which is available in the data source. Future income at the origin is, of course, unobservable and must be estimated. The procedure to accomplish this will be explained below. Once estimated income at the origin is obtained, income for each year will be multiplied by "k", where "k" is the ratio of *actual* income in year 1971 (prior to migration) to the *estimated* income in the same year. Expansion of estimated incomes by "k" is based upon the logic that individuals, assumed to have perfect foresight, are expected to use all available information (including relative standing among peers) to formulate future income. Similarly, incomes of a nonmigrant at potential destinations must be estimated and multiplied by "k". Incomes at origin for nonmigrants are the actual incomes which, again, are available in the data source.[4]

For the CID theory, the computation of X for each migrant under Formulation 1 can be represented as

$$X = \frac{\text{income at destination} - \text{k (estimated income at origin)}}{\text{k (estimated income at origin)}}.$$

For each nonmigrant,

$$X = \frac{\text{k (estimated income at a potential destination)} - \text{income at origin}}{\text{income at origin}}.$$

All (actual or estimated) incomes used in the above computation are for the year 1973. For the PVD theory, X is computed in a similar fashion except that present value of the income stream between 1973 and 1977 is used instead of 1973 income. As in Bowles (1970), a discount rate of six percent was used in the computation of these present values.[5]

Under Formulation 2 and the CID theory, variable X for each migrant (nonmigrant) can be represented as

$$X = \frac{\text{estimated income at (potential) destination} - \text{estimated income at origin}}{\text{estimated income at origin}}.$$

A similar computation of X is used for the PVD theory except that estimated 1973 incomes are replaced by the present values of estimated income streams between 1973 and 1977.[6]

Under Formulation 3, future incomes at both the origin and destination must also be estimated. In this case, however, these estimates comprise average income values for the respective states (rather than values derived from regression analysis with personal and state attributes employed as regressors).

As shown above, both the perfect foresight and group-assimilating models of income expectation require econometric projections of individual income (earnings). These projections were made with income equations based upon Mincer's (1970) model of schooling, experience and earnings; and take the form

$$\ln (y_t) = \ln (y_0) + \alpha_1 \ln (\text{UNION}) + \alpha_2 \ln (\text{NONWHITE}) \quad (1)$$

$$+ \alpha_3 \ln (\text{FEMALE}) + \alpha_4 \ln (\text{UNEM}) + \alpha_5 \ln (\text{EMPGR})$$

$$+ \alpha_6 \ln (\text{URBAN}) + \alpha_7 \text{EDUC} + \alpha_8 \text{EXPER} + \alpha_9 \text{ASQ}$$

$$+ \sum_{r=1}^{8} \beta_r \text{DUM}_r + \epsilon,$$

where:

y_t, y_0 are earnings of the individual under question in year t and o, respectively;

UNION, NONWHITE, FEMALE, UNEM, EMPGR, and URBAN are area (state) labor market characteristics, and are respectively the percent union, percent nonwhite, percent female, unemployment rate, employment growth rate, and the percent urban in one's state of residence;[7]

EDUC, EXPER and ASQ are respectively education (measured in years of school), working experience and age squared. Working experience is proxied by "age—years of education—6" as suggested by previous studies (Rosenzweig and Morgan, 1976);

DUM_r is a dummy variable for Census division r; and ϵ is the disturbance term.[8]

As labor earnings are expected to differ among individuals by sex, race, and occupation (Chiswick and O'Neill, 1977), equation (1) was fitted to microdata for white-males only, and for each of seven occupations: professional, technical, and kindred workers; managers, officials, and proprietors; self-employed businessmen; clerical and sales workers; craftsmen, foremen, operatives, and kindred workers; laborers and service workers; and farmers and farm managers. Parameter estimates of equation (1) for each of these seven occupations are provided in Nghiep (1981). For each individual, estimates from this "income model" were then used to quantify the economic variables — PVD and CID — in a model of the migration decision (described below).

EMPIRICAL RESULTS

The Panel Study of Income Dynamics (PSID) data from the Institute for Social Research at the University of Michigan were used to estimate both the income and migration equations. PSID is a longitudinal data set on individuals, covering the period 1968-1977.[9] Estimates were confined to white-male family heads.

As a consumer, an individual migrates from region "i" to another region to maximize utility in "consumption", the determinants of which include money-income, personal characteristics, and area characteristics. Personal characteristics include employment status, marital status, home ownership, and extent of previous migration (Ritchey, 1976). With respect to previous migration, an individual who has migrated in the recent past will, in general, have less attachment to the current location than a long-time resident, and better information about alternative locations (resulting in a higher migration propensity, *ceteris paribus*). On the other hand, factors like marriage, employment, and home ownership increase one's ties to the community and thus impede migration (Lansing and Mueller, 1967). Area characteristics such as air quality and crime are also important to the migration decision.

In compact form, the migration decision can be represented as:

$$M_{ik} = f(X_{ik}, PC_k, AC_i, U_{ik}), \qquad (2)$$

where:

M_{ik} is the probability that individual "k" in region "i" will migrate interstate. It is 1 if the individual migrates and 0 otherwise. The 48 contiguous states plus the District of Columbia comprise the regions for this study;

X_{ik} is the income differential (PVD or CID) between the (potential) destination and origin states relative to income at the origin (as perceived by individual "k");[10]

PC_k represents personal characteristics of individual "k";

AC_i represents characteristics of state "i";

U_{ik} is the disturbance term.

As the migration equation has a binary dependent variable, the logit technique was utilized for parameter estimation. Binary independent variables included in equation (2) are HSYES, MARST and EMPST, and were set equal to one (vs. zero) to represent home ownership, marriage and employment respectively. The number of prior moves is represented by TM72, while area crime and air quality indices are denoted CRIME and AIRQ respectively. Finally, all of these variables were determined as of 1972, the base year for the examination of 1972-1973 interstate migration. Estimates of equation (2) for each of the three formulations of income expectation (and CID vs. PVD) are provided in Table 1.

Given these empirical results for the migration equation, the question of superiority (PVD vs. CID) can now be addressed. For each of the three models of income expectation, the relative performance of the PVD and CID theories will be evaluated on the basis of t-statistics and the value of the likelihood function ("L" in Table 1). In addition, the performance of the income variable (X_{ik})—the key, or test, variable of the study—will be examined closely.

Under formulation 1, previous migration (TM72) and homeownership (HSYES) perform equally well in both theories while marital status (MARST) and employment status (EMPST) provide better results under the CID theory. Area characteristics—crime index (CRIME) and air quality (AIRQ)—are insignificant. Because of the greater significance of the income variable in the PVD framework and the better overall performance of the PVD theory—represented by the higher value of the likelihood function ("L" in Table 1)—the PVD theory is judged superior to the CID theory under the perfect foresight model of income expectation.

Under formulation 2 (the group-assimilating model), previous migration (TM72), homeownership (HSYES), and marital status (MARST) perform equally well in both theories. While employment status (EMPST) displays a higher level of significance in the CID theory, air quality (AIRQ) is significant only in the PVD theory. Again, the higher significance of the income variable under the PVD framework and the better overall performance of the PVD theory (based upon values of the likelihood function) provide evidence of the superiority of this theory under formulation 2. Finally, under formulation 3 (the conventional formulation), each regressor performs equally well in both theories, while the values of the likelihood function show little difference between theories.

TABLE 1
Logit Estimates of the Migration Equation

	Formulation 1		Formulation 2		Formulation 3	
	CID	PVD	CID	PVD	CID	PVD
Constant	0.9796	−0.7547	2.6644	1.5610	−2.1470	−1.7972
	**	***	**	***	***	***
X_{ik}	0.1688	0.2576	1.5052	0.8211	0.0574	0.0035
	(1.887)	(2.317)	(1.898)	(4.280)	(3.565)	(3.389)
	***	***	***	***	***	***
TM72	1.0000	1.0622	0.8714	1.080	1.0741	1.0981
	(3.037)	(3.162)	(3.464)	(4.104)	(4.309)	(4.365)
	*	*	**	**	***	***
HSYES	−0.6138	−0.6166	−0.6781	−0.6620	−0.8156	−0.8014
	(−1.365)	(−1.3562)	(−1.877)	(−1.784)	(−2.343)	(−2.303)
	***	**	***	***	***	***
MARST	−1.3586	−1.1795	−1.2670	1.2543	−1.2094	−1.2190
	(−2.229)	(−1.882)	(−2.485)	(−2.416)	(−2.426)	(−2.452)
	*		***	**		
EMPST	−1.5225	−0.0829	−2.2591	−1.8943	−0.2075	−0.2939
	(−1.336)	(−0.054)	(−2.256)	(−1.866)	(−0.263)	(−0.379)
					***	***
CRIME	−0.0123	−0.0090	−0.0035	0.0064	0.0547	0.0457
	(−0.554)	(−0.403)	(−0.196)	(0.340)	(2.248)	(1.984)
				*	***	***
AIRQ	0.0131	0.0162	0.0147	0.0230	0.0443	0.0394
	(0.767)	(0.952)	(1.133)	(1.686)	(2.870)	(2.637)
L	−83.449	−82.245	−141.648	−132.630	−140.776	−141.193

Numbers in parentheses are t-statistics.
Numbers marked with "*" are significant at the .10 level, numbers marked with "**" are significant at the .05 level, and numbers marked with "***" are significant at the .025 level.
L is the value of the likelihood function.

Based upon these results, the PVD theory of migration is judged superior to that theory based upon migration response to current income differentials, and for the following reasons: (1) the t-statistics on the income variable, the key variable of the study, are considerably higher in the PVD theory than in the CID theory under formulations 1 and 2, while these statistics are comparable within formulation 3; and (2) the overall fit of the PVD theory is better than that of the CID theory under formulations 1 and 2 and is again comparable within formulation 3.

It is interesting to note in Table 1 that the "quality of life" indicators, CRIME and AIRQ, are (highly) significant determinants of migration solely under the third, or conventional, formulation of income expectation. Given the high relative values of likelihood functions under formulations 1 and 2, the conventional formulation of income expectation appears to incorrectly attribute migration to quality of life rather than expected income differentials across space.[11]

CONCLUSIONS

This study provides additional evidence for the often claimed superiority of the "present value difference" theory of migration (over the "current income difference" theory). Thus, individuals considering geographic relocation appear to evaluate economic incentives, that vary across space, within a human investment framework. Such behavior, in turn, provides additional evidence for the human capital theory of migration.

However, these results are conditional upon access to reliable and accurate information on both age-earnings profiles and the manner in which these profiles vary over space. To the extent that income expectations follow our conventional formulation (no. 3), possibly the result of imperfect information on spatially diffused labor market conditions, geographic mobility in the United States is most likely triggered by immediate economic pressures and personal circumstances. Although our empirical results provide little information on *how* earnings expectations are formed, our comparisons of the human capital and current income difference theories of migration illustrate the inseparability of expectation theories from those concerned with economic incentives for migration.

FOOTNOTES

[1]For instance, see Bowles (1970) or Bowman and Myers (1967).

[2]Grant and Vanderkamp (1980) also estimate these incomes within an econometric model of the earnings function. However, the responsiveness of potential migrants to estimated PVD is not examined by their work; instead, comparisons are made as to the magnitude of income gains realized by migrants (and the time required to capture potential benefits to migration).

[3]The study requires an econometric model of individual income determination. This model, also required for the group-assimilating model of income expectation, is discussed below.

[4]To reduce possible heteroscedasticity in the estimated migration equations under formulations 1 and 2, X is computed as the income differential relative to origin income.

[5]The econometric results described below were examined as to sensitivity to this discount rate (and specifically to higher rates). Little variation was detected.

[6]The computation of X for the PVD theory as discussed above implicitly requires a time span of future income from the present year to the last year of the working life of each individual. However, the Panel Study of Income Dynamics (PSID) data (which will be used for this study) provide income data for ten years only:

1968-1977. Furthermore, since the base year for the study is 1972 (1968-1972 being utilized for the determination of prior mobility history), one is left with five years—from 1973 to 1977—for the computation of the present value of future income. Should one adjust for this shortcoming related to data availability? The answer is yes, as the length of the future income stream is important for the computation of present values (and differences in present values) under the PVD theory. In this respect, an adjustment to the 1973-1977 present value computations was made for each individual in the study on the basis of age. The method by which this was accomplished is described in Nghiep (1981).

The computation of X also requires that potential destination states "j" be known for nonmigrants. Information on migrant stocks was employed to define such states. Migrant stock, MS_{ij}, is defined as the number, or "stock", of individuals in state "j" who were born in state "i". For a given state of origin, one has migrant stocks (MS_{ij}'s) in 47 other states plus the District of Columbia. The bigger MS_{ij}, the more information is being sent back to potential migrants at home state "i". Because the degree to which a state "j" can qualify as a potential destination is positively related to MS_{ij}, the "n" states of highest

MS_{ij} can be chosen as potential destinations, and the income at a (hypothetical) potential state of destination, y_j, can be proxied by:

$$y_j = \sum_{z=1}^{n} MS_{iz}\, y_z \Big/ \sum_{z=1}^{n} MS_{iz}\,,$$

where y_z is the personal income in state z. An alternative method for the representation of alternative destinations in microdata-based studies of the migration decision was developed recently by Falaris (1982).

[7]For any specific occupation, unemployment and employment growth rates cannot both be included in the estimation of equation (1) due to multicollinearity between these variables.

[8]See Appendix 1 in Nghiep (1981) for a detailed explanation of this equation specification.

[9]Additional years have been added since this study.

[10]The method by which a set of potential destination states was selected for each origin state "i" is described in footnote six.

[11]This point was suggested by a referee of the journal.

REFERENCES

Bowles, Samuel. "Migration as Investment: Empirical Tests of Human Investment Approach to Geographic Mobility," *Review of Economics and Statistics,* 52 (November 1970), 356-362.

Bowman, M. J., and R. G. Myers. "Schooling, Experience, and Gains and Losses in Human Capital Through Migration," *Journal of the American Statistical Association,* 62 (September 1967), 875-895.

Chiswick, B. R., and G. A. O'Neill. *Human Resource and Income Distribution.* New York: W. W. Norton Company, Inc., 1977.

Davanzo, Julie. *Microeconomic Approaches to Studying Migration Decisions.* The Rand Corporation, Santa Monica, California: May 1980, note.

Davanzo, Julie. *Repeat Migration, Information Costs, and Location Specific Capital.* The Rand Corporation, Santa Monica, California: May 1980, The Rand Papers Series.

Falaris, Evangelos M. "Migration and Regional Wages," *Southern Economic Journal,* 48 (January 1982), 670-686.

Fields, G. S. "Labor Force Migration, Unemployment and Job Turnover," *The Review of Economics and Statistics,* 58 (November 1976), 407-415.

Grant, E. K., and Vanderkamp, J. "The Effects of Migration on Income: A Micro Study with Canadian Data: 1965-71." *The Canadian Journal of Economics,* 13, 3 (August 1980), 381-406.

Hansen, N. "Migration Centers, Growth Centers, and the Regional Commissioners: An Analysis of Expected Future Lifetime Income Gains to Migrants from Lagging Regions," *Southern Economic Journal,* 38 (April 1972), 508-517.

Kiker, B. F., and E. C. Traynham. *Return and Nonreturn Migration for the Southeast: 1960-1970.* Columbia,

South Carolina: University of South Carolina, College of Business Administration, Bureau of Business and Economic Research (Essay in Economics, No. 30, July 1974).

Kiker, B. F., and E. C. Traynham. "Earnings Differentials Among Nonmigrants, Return Migrants, and Nonreturn Migrants," *Growth and Change,* 8, 2 (April 1977), 1-7.

Laber, G. "Human Capital and Southern Migration," *Journal of Human Resources,* 8 (1973), 223-241.

Lansing, J. B., and E. Mueller. *The Geographic Mobility of Labor.* Ann Arbor, Michigan: Survey Research Center, 1967.

Mincer, J. "The Distribution of Labor Incomes: A Survey," *Journal of Economic Literature,* 8 (March 1970), 1-26.

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

Nghiep, Nguyen huu. *Earnings Expectations and the Role of Human Capital in the Migration Decision: An Empirical Analysis.* Dissertation monograph, The University of Tennessee, Knoxville (March 1981).

Ritchey, P. N., "Explanations of Migration," *Annual Review of Sociology,* 2 (1976), 112-115.

Rosenzweig, M. R., and Morgan, J. "An Exchange: On the Appropriate Specification of Human Capital Models," *The Journal of Human Resources,* 11, 1 (Winter 1976), 3-7.

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