Migration and Income: The Question of Self-Selection*

ROBERT A. NAKOSTEEN
Tennessee Valley Authority

MICHAEL ZIMMER
University of Evansville

I. Introduction

In recent years economists have become increasingly interested in the apparent self-selection exhibited by economic agents as they participate in market processes. Embodied in this concept is the notion that agents choose among competing alternatives at least in part on the basis of anticipated incremental returns. Rationality dictates that persons choosing a given alternative do so because they have some tangible basis for perceiving a more favorable return than those who choose otherwise. The result is that persons selecting a particular course of action tend to be non-randomly distributed within the population as a whole. As a consequence there is inherent "selectivity bias" in data which report relative returns to competing alternatives. This problem is recognized as a complicating factor in attempts to estimate returns to schooling, labor force participation, unionization, and migration, to mention a few [4;12].

Researchers have recently addressed some of these questions in the context of econometric models which explicitly account for selectivity bias in wage/income comparisons. The problem of labor force participation is considered by Heckman [5] and Nelson [19]. The effects of unionization on wage levels have been analyzed by Lee[10], while Roberts, Maddala and Enholm[22] have examined problems associated with behavior of regulated firms. Returns to college education are reported by Kenny et.al.[8], and earnings effects of military occupational training are discussed by Trost and Warner[24]. The purpose of this paper is to describe and estimate a model of returns to migration which explicitly accounts for self-selection of migrants from the working population. The essence of the problem as it pertains to migration is summarized in a recent survey:

* 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.

The authors benefitted from the helpful comments of G.S. Maddala, Robert Trost, Henry W. Herzog, Jr., and an anonymous refree. Programming assistance was provided by Ronald L. Henry. Responsibility for any remaining errors is our own.

> Presumably, by examining the earnings of otherwise comparable individuals who do not migrate, we take into account what an individual would have earned had he not moved. However, the fact that individual A migrates, while otherwise comparable B does not, suggests that an important difference does exist between the individuals. These differences may be in the way they view benefits. Individual A, for example, may be more highly motivated to invest in human capital formation, not only in migration, but in other forms as well. If such were the case, the earnings of the remaining cohort from which the migrant is drawn may provide a lower bound for the earnings the migrant would have received in the absence of migration[3, 409].

The present study is intended as a point of departure for more fruitful analysis, so that returns to migration need not be couched in terms of a mere "lower bound" of migrant earnings.

The remainder of the paper is organized as follows: Section II describes a simple model of migrant behavior in which the decision to migrate, viewed in the context of investment in human capital, results as the solution to an optimal control problem. Section III provides details of a simultaneous equations model which incorporates the decision to migrate, returns to migration and self-selection. Results of estimation are presented in Section IV, based on a large random sample of individual wage earners taken from the Continuous Work History Sample (CWHS) of the Social Security Administration. Section V summarizes and concludes the study.

II. Background

Recent advances in migration literature have conformed to the general framework of Sjaastad[23] 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 an earnings differential as well as the direct costs attendant to moving:

$$PV(t) = \int_0^T \Big[Y_{mt} - Y_{nt} \Big] e^{-\rho t} dt - C_{mn} \qquad (1)$$

where the Y's denote earnings in areas m and n respectively at time t, C_{mn} is the cost of moving from region n to region m, ρ is the implicit discount rate and T represents the time during which the individual will remain in the labor force. In this simple formulation individuals respond to positive values in expression (1) by changing location; otherwise no migration occurs.

Refinements of the simple Sjaastad model include a model recently described by Polachek and Horvath[21], who propose an optimal control model of life cycle locational change. The individual is assumed to maximize his present value of lifetime earnings, where available controls include: (1) investment in human capital, (2) a strategy of search for attractive wage opportunities in other locations, and (3) mobility investment in the form of location change. The investment is carried out at each stage of the life cycle, where "location" is viewed as a composite of locational attributes. Inclusion of search time as an available control introduces a feature of periodicity in migration over the life cycle, since migrants realize depreciation in their accumulated stocks of information subsequent to moving. The model, which is based on a household utility function, accounts for externalities among family members arising from labor force participation of family members other than the head of household and from the presence of school-age children.

Polacheck and Horvath point out that under certain conditions the model gives rise to a conventional "bang-bang" solution, wherein the optimal control switches from its upper to lower bound without assuming intermediate values. The important feature of this dichotomy is that the decision to migrate may be modeled in part by means of a binary variable representing "move" or "not move" for each individual. Most recent attempts to model migration decisions have relied on such a device.[1] The model presented in the following sections adheres to a similar convention, while allowing the migration decision and returns to migration to be determined simultaneously and paying explicit attention to the problem of self-selection.

III. Econometric Model: Specification and Estimation

Specification

The model described in this section is one of a class of models characterized by Maddala and Nelson [14] as switching regression models with endogenous switching. It consists of two income equations (one for migrants, the other for non-migrants) as well as an equation describing the dichotomous decision to migrate. The sample observations may be thought of as falling into one of two mutually exclusive regimes, with the decision equation serving as an endogenous selectivity criterion which determines the appropriate regime (migrant versus non-migrant).

The task at hand is to utilize the sample observations to estimate the parameters of the decision function and the income equations. If consistent estimates of the earnings equations can be obtained, then fitted values from the income equations may be used to estimate the parameters of the decision function.

To simplify the discussion of Section II, assume that at any point in time individual i elects to migrate if the percent gain in moving exceeds the associated total costs. Thus a person chooses to migrate if

$$(Y_{mi} - Y_{ni})/Y_{ni} > B_i \quad (2)$$

where B_i represents direct and indirect costs, as a proportion of income, incurred by individual i in moving from region m to region n. Further, suppose that proportionate costs may be represented as a function of one or more personal characteristics (X), one or more attributes of the origin locality (Z), and a random disturbance term:[2]

$$B_i = g(X_i, Z_i) + \epsilon_i \quad (3)$$

Regional attributes are included in (3) to reflect indirect costs of moving from areas which offer attractive opportunities for growth in employment and income.

Expressions (2) and (3) suggest, as a general proposition, that the migrant selectivity cri-

1. Other examples include Kaluzny [7] and Navratil and Doyle [18].

2. It can be argued that attributes of the potential destination and origin-to-destination distance also contribute to total migration costs. For example, migration into regions with substantial shortages of housing may entail additional search costs. However, inclusion of potential destination attributes raises substantial definitional and empirical problems in models of the type described in this study. Consequently only origin attributes are used.

terion is a function of gains in earnings along with regional and personal attributes. In this study the criterion is modeled as a linear combination of these variables which, taken together, explain an individual's propensity to migrate. Specifically, individual i chooses to migrate if

$$I_i^* > 0 \tag{4}$$

and doesn't migrate if

$$I_i^* \leqq 0 \tag{5}$$

where

$$I_i^* = \alpha_0 + \alpha_1 \,[(Y_{mi} - Y_{ni})/Y_{ni}] + \alpha_2 X_i + \alpha_3 Z_i - \epsilon_i \tag{6}$$

The model is completed by specifying income equations for migrants and non-migrants respectively:

$$Y_{mi} = \theta_{m0} + \theta_{m1} X_i + \theta_{m2} Z_i + \epsilon_{mi} \tag{7}$$

$$Y_{ni} = \theta_{n0} + \theta_{n1} X_i + \theta_{n2} Z_i + \epsilon_{ni}. \tag{8}$$

The vectors of explanatory variables in (7) and (8) do not necessarily consist of the same elements as those appearing in (6); the disturbance terms ϵ_{mi} and ϵ_{ni} are assumed to be normally distributed with variances σ_m^2 and σ_n^2, respectively.

Expressions (6)–(8) comprise the basic structural form of the model. The endogenous variables are I^*, Y_m and Y_n. We do not observe I^*, but only

$$I_i = 1 \text{ if } I_i^* > 0$$
$$I_i = 0 \text{ if } I_i^* \leqq 0.$$

In addition, for migrants we observe only the destination wage, and for non-migrants we observe only the origin wage; i.e., we observe

$$Y = Y_m \text{ when } I_i = 1$$
$$Y = Y_n \text{ when } I_i = 0.$$

What is required in (6)–(8) is some suitable measure of income. In this study the natural logarithm of annual earnings is chosen as the dependent variable in the income equations. Since $(Y_{mi} - Y_{ni})/Y_{ni}$ is approximated by $\log Y_{mi} - \log Y_{ni}$, the latter variable is inserted in (6). Use of this approximation simplifies the estimation procedure and remains consistent with the choice mechanism outlined above. With these modifications, the model to be estimated becomes

$$I_i^* = \alpha_0 + \alpha_1 \,[\log Y_{mi} - \log Y_{ni}] + \alpha_2 X_i + \alpha_3 Z_i - \epsilon_i \tag{6$'$}$$

$$\log Y_{mi} = \theta_{m0} + \theta_{m1} X_i + \theta_{m2} Z_i + \epsilon_{mi} \tag{7$'$}$$

$$\log Y_{ni} = \theta_{n0} + \theta_{n1} X_i + \theta_{n2} Z_i + \epsilon_{ni} \tag{8$'$}$$

The binary nature of the observed dependent variable in (6)′ suggests that the parameters of the decision equation may be estimated by maximum likelihood probit or logit techniques. The income equations could be estimated by ordinary least squares (OLS) and the resulting fitted values of log-income could be inserted into (6)′ to obtain consistent estimates of the decision equation. An analagous procedure is employed by Polachek and Horvath [21] to estimate a similar model. The problem with this procedure is that OLS is inappropriate for the income equations, since it effectively fails to reflect the presence of self-selection in migration. This can be seen by noting that the conditional means of the income disturbance terms are non-zero and not constant for all observations:

$$E(\epsilon_{mi}|I_i = 1) = \sigma_{m\epsilon^*} [-f(\psi_i)/F(\psi_i)] \tag{9}$$

$$E(\epsilon_{ni}|I_i = 0) = \sigma_{n\epsilon^*} [f(\psi_i)/1-F(\psi_i)] \ , \tag{10}$$

where $\sigma_{m\epsilon^*}$, $\sigma_{n\epsilon^*}$ and ψ_i are defined below, and $f(\cdot)$ and $F(\cdot)$ are the standard normal density and distribution functions, respectively. Expressions (9) and (10) are based on well-known conditional formulae for the truncated normal distribution [6]. The argument ψ_i in (9) and (10) is obtained as follows. Substituting (7)′ and (8)′ into (6)′ gives the reduced form of the decision equation:

$$I_i^* = \beta_0 + \beta_1 X_i' + \beta_2 Z_i' - \epsilon_i^* \tag{11}$$

where the vectors X_i' and Z_i' consist of all exogenous variables in the model. If we assume that the disturbance term is normally distributed with unit variance, (11) may be estimated by maximum likelihood probit methods. Define the expression

$$\psi_i = \beta_0 + \beta_1 X_i' + \beta_2 Z_i' \ .$$

Probit estimation yields fitted values $\hat{\psi}_i$, which are to be used as estimates of the arguments in (9) and (10).

The coefficients $\sigma_{m\epsilon^*}$ and $\sigma_{n\epsilon^*}$ in (9) and (10) are elements of the covariance matrix of the disturbances:

$$\text{COV}\ (\epsilon_m \epsilon_n \epsilon^*) = \begin{bmatrix} \sigma_m^2 & \sigma_{mn} & \sigma_{m\epsilon^*} \\ & \sigma_n^2 & \sigma_{n\epsilon^*} \\ & & 1 \end{bmatrix} . \tag{12}$$

Expressions (9) and (10) summarize the selectivity bias which results from OLS estimation of the income equations; as a consequence, OLS estimates are inconsistent and lead to biased estimates of returns to migration. It is in this respect that previous empirical studies of migration have failed to fully account for the endogenous nature of the decision to migrate. Most studies entail estimation of earnings equations which include (exogenous) dummy variables distinguishing migrants from non-migrants. The model described in this section explicitly recognizes the endogenous nature of the migration decision and thus formally accounts for the problem of migrant self-selection. A more general discussion of econometric models of self-selection, including problems of identification and estimation, is found in Maddala [13], and Maddala and Roberts [15].

Estimation

Procedures for estimating the parameters in (6)′–(8)′ and (12) are developed by Lee [9]. The suggested procedure is to modify the income equations by incorporating the appropriate "selectivity variables" and to add error terms with zero means. A two stage estimation procedure is then employed to estimate all the parameters in the model. The corrected income equations may be written:

$$\log Y_{mi} = \theta_{m0} + \theta_{m1} X_i + \theta_{m2} Z_i + \sigma_{m\epsilon^*} [-f(\psi_i)/F(\psi_i)] + \eta_{mi} \tag{13}$$

$$\log Y_{ni} = \theta_{n0} + \theta_{n1}X_i + \theta_{n2}Z_i + \sigma_{n\epsilon^*}[f(\psi_i)/1-F(\psi_i)] + \eta_{ni} \tag{14}$$

where

$$E(\eta_{mi}|I_i = 1) = 0$$

and

$$E(\eta_{ni}|I_i = 0) = 0 .$$

Stage one entails probit estimation of the reduced form decision equation (11). Fitted values obtained from stage one denoted $\hat{\psi}_i$, are used to construct variables

$$u_{1i} = [-f(\hat{\psi}_i)/F(\hat{\psi}_i)]$$

and

$$u_{2i} = [f(\hat{\psi}_i)|1-F(\hat{\psi}_i)] .$$

In stage two, u_{1i} and u_{2i} are inserted into the appropriate income equations and these are estimated by OLS. Estimates obtained by this procedure are known to be consistent. Further discussion of the estimation procedure, its theoretical properties and conditions and identification appears in Lee [9].

In the following section this procedure is used to obtain estimates of the model. The major issues at hand include (1) testing for the significance of the "truncation variable" in seeking support for the selectivity hypothesis, and (2) obtaining consistent parameter estimates of the structural form of the migration decision equation.

IV. Empirical Results

Sample Description and Explanatory Variables

The data used in this study are based on a random selection of 9,223 employed persons drawn from the Social Security Administration Continuous Work History Sample (CWHS) for whom earnings records were available for both 1971 and 1973 (the decision-to-migrate interval). For each individual the data furnish information on earnings (Y), age, race, sex, state of employment, along with the industry in which the person is employed (measured at the two-digit Standard Industrial Classification (SIC) level) and a dummy variable indicating whether the person is self-employed (SE). In addition, such regional variables as rate of growth of state employment (ΔEMP) and state per capita income (ΔPCI) have been added to each record in the sample. Migrant status is defined as follows: the individual is a migrant if state of employment changes from 1971 to 1973; he is a non-migrant otherwise.

Use of the CWHS as a data base introduces shortcomings into the specification of the model. In particular, it is not possible to explicitly account for the effects of education, labor force behavior of other family members, stability of the family unit, and the presence of children on the decision to migrate and subsequent earnings. It is maintained, however, that these disadvantages are offset by incorporation of endogenous self-selection into the model. Indeed, the influence of these variables is likely to be effectively embodied in the self-selection process. Moreover, the two-stage procedure employed in this study results in estimates possessing the usual desirable asymptotic properties; there are facilitated by use of the CWHS, which permits arbitrarily large samples.

As noted previously, the structural form of the model consists of a decision equation and earnings equations for migrants and nonmigrants. The model is specified by stating the exogenous variables included in each equation:

$$I_i = \alpha_0 + \alpha_1 (\log Y_{mi} - \log Y_{ni}) + \alpha_2 AGE + \alpha_3 RACE + \alpha_4 SEX + \alpha_5 SE + \alpha_6 \Delta PCI + \alpha_7 \Delta EMP + \epsilon_i \quad (15)$$

$$\log Y_{mi} = \Theta_{m0} + \Theta_{m1}\, \Delta SIC + \Theta_{m2} SE + \epsilon_{mi} \quad (16)$$

$$\log Y_{ni} = \Theta_{n0} + \Theta_{n1}\, \Delta SIC + \Theta_{n2} SE + \epsilon_{ni} \quad (17)$$

The variable *AGE* is included in the decision equation to reflect the widely held notion that the probability of migration declines with age; thus the coefficient of *AGE* is expected to be negative. *RACE* is a dummy variable which assumes the value one for non-whites and zero for whites. There is considerable ambiguity in the literature regarding the effect of race on migration. Greenwood [2], for example, reports that non-whites are more responsive than whites to opportunities for income growth, while Persky and Kain [20] suggest that whites are more responsive to the availability of job opportunities. Thus, there is little a priori basis for the sign of the *RACE* coefficient. The dummy variable *SEX* is zero for males and one for females. Its coefficient is expected to be negative to reflect the effect of family ties on the mobility of females.[3] The dummy variable *SE* is one for self-employed persons and zero otherwise. A negative coefficient is expected, since self-employed persons are less susceptible to location change resulting from job transfer and promotion. Two regional variables are included in the decision equation to measure attributes of the individual's origin locality. Growth in total employment (ΔEMP) is expected to have a negative impact on the probability of migration, since individuals are less likely to migrate from areas of rapidly growing employment opportunities. Although growth in per capita income (ΔPCI) would seem to exert a similar deterrent effect on migration, its actual impact is the subject of extensive debate.[4] Consequently, the income coefficient is not unambiguously negative.

The earnings equations include only exogenous variables that are thought to exert an impact on earnings distinct from their impact on the decision to migrate. Inclusion of all exogenous variables in both the decision and earnings equations introduces a collinearity problem in the second stage of the estimation procedure. In this case, those included are the self-employment dummy variable (*SE*) and an industry change variable (ΔSIC) which is one if the individual changes industry of employment between 1971 and 1973 and zero otherwise. Self-employment is included due to the presence of self-employed farmers in the sample. Its coefficient is expected to be negative, reflecting the familiar phenomenon of lower reported earnings of owner-operators in agriculture. The industry change dummy should have a negative impact on earnings, since individuals who change industries of employment are likely to experience depreciation in specific human capital during periods immediately following the change.[5]

3. A thorough discussion of this point is found in Mincer [17].
4. For a discussion of this issue, see Greenwood [3, 400].
5. A discussion of this point is found in Galloway [1].

Results of Estimation

Substitution of earnings equations (16) and (17) into the decision equation (15) results in a reduced form decision equation which includes as explanatory variables all the exogenous variables in the model. Maximum likelihood estimates of this equation are presented in Table I. Examination of the results reveals that the signs of parameter estimates generally conform to a priori expectations. The probability of migration decreases with age and decreases for self-employed persons. The dummy variable corresponding to inter-industry migration is positive and highly significant, indicating that locational change may occur in many cases simultaneously with occupational or industry change. The sex dummy is significantly negative, as expected, though the coefficient on the race dummy variable is not significantly different from zero; regional variables describing growth in origin employment and origin per capita income are also significant. The sign of the coefficient on the employment growth variable is negative, as expected. The result suggests that workers in a state with rapidly growing employment opportunities would be less likely to migrate. The positive sign on the per capita income growth variable implies that workers are more likely to leave areas of high or growing income, or perhaps that rapid growth in per capita income is used to finance interstate mobility.[6] The product of −2 and the log-likelihood ratio is distributed as Chi-Square with seven degrees of freedom, where the restricted form of the likelihood function constrains all coefficients to zero. Its large value in this case attests to the overall explanatory power of the reduced form decision equation.

In the next step, fitted values from the reduced form probit model are used to construct selectivity variables, one for migrants and the other for non-migrants, in accordance with expressions (9) and (10). When these variables are appended to the corresponding earnings equations, the resulting "corrected" equations may be correctly estimated by OLS. Estimates of the earnings equations are presented in Tables II and III. As noted previously, the earnings equations are specified to include only those variables which are thought to influence earnings in a manner distinct from their impact on the decision to migrate. This is borne out by the estimates, which indicate that the self-employment and industry change variables are significant and negatively related to earnings, as expected. Of particular interest are the estimated coefficients of the selectivity variables. In the migrant equation this coefficient is not significantly different from zero, while it is positive and significant in the non-migration equation. This result lends support to the hypothesis of self-selection, at least as it pertains to non-migrants from the population. It may be interpreted in support of the notion that non-migrants in the population choose their status because they fail to perceive more favorable returns elsewhere. These findings corroborate the results of earlier studies in which observed *levels* of migrant earnings are found to be lower than those of non-migrants.[7]

An additional point of interest involves the combined effect of the two selectivity variables on *unconditional* earnings. Clearly the combined truncation effect should be positive, so that the process of self-selection serves to enhance unconditional expected earnings. To see this, note [13] that unconditional expected earnings for individual i may be written

$$E(Y_i) = E(Y_i|I_i = 1) \cdot P(I_i = 1) + E(Y_i|I_i = 0) \cdot P(I_i = 0)$$

6. This result, though perhaps surprising on a priori grounds, is not unprecedented. Similar results are inferred by Vanderkamp [25, 1,022] and Miller [16, 400].

7. See, for example, Polachek and Horvath [21, 127].

Table I Probit Estimation of the Reduced Form Decision Equation

Variable	Coefficient	t-value
constant	-1.509	
SE	-0.708	-5.72
ΔEMP	-1.488	-2.60
ΔPCI	1.455	3.14
AGE	-0.008	-5.29
RACE	-0.065	-1.17
SEX	-0.082	-2.14
ΔSIC	0.948	24.15

Number of observations = 9,223
Number of migrants = 1,078
Number of non-migrants = 8,145
−2(log-likelihood ratio) = 104.95

so that

$$E(Y_i) = (\Theta'_m X_{mi} - \sigma_{m\epsilon} \cdot (f(\Psi_i)/F(\Psi_i))) \cdot F(\Psi_i) + (\Theta_n' X_{ni} + \sigma_{n\epsilon}\, (f(\Psi_i)/1 - F(\Psi_i))) \cdot (1 - F(\Psi_i)),$$

where X_{mi} and X_{ni} refer to all exogenous variables in the migrant and non-migrant earnings equations, respectively. Rewriting,

$$E(Y_i) = (\Theta_m' X_{mi}) \cdot F(\Psi_i) + (\Theta_n' X_{ni}) \cdot [1 - F(\Psi_i)] + (\sigma_{n\epsilon} - \sigma_{m\epsilon}) \cdot f(\Psi_i).$$

Table II Migrant Earnings Equation: Second Stage Estimates

Variable	Coefficient	t-value*
constant	9.041	
ΔSIC	-0.790	-2.24
SE	-4.104	-9.54
SELECTIVITY	0.212	0.50

Number of observations = 1,078
R^2 = .160
F = 118.54
*The t-values are slightly biased [11].

Table III Non-Migrant Earnings Equation: Second Stage Estimates

Variable	Coefficient	t-value*
constant	8.593	
ΔSIC	-0.927	-9.35
SE	-4.161	-57.71
SELECTIVITY	0.863	2.84

Number of observations = 8,145
$R^2 = .456$
F = 2,233.89
*The T-values are slightly biased [11].

The third term represents the combined effect of self-selection on expected earnings. Based on the estimates from Tables II and III, we have

$$\hat{\sigma}_{n\epsilon} - \hat{\sigma}_{m\epsilon} = .652,$$

indicating that the combined effect on earnings is positive.

The final step in the estimation procedure entails probit estimation of the structural form of the migrant status equation (6)′. Consistent estimates of the parameters in the earnings equations are used to obtain fitted values of log-earnings, which together with appropriate exogenous variables are then inserted into the structural decision equation. It can be shown [9] that the parameters of the structural decision equation are identified if the earnings equations contain at least one exogenous variable which does not appear in the structural decision equation. In this model a variable measuring industry change (ΔSIC) appears in each earnings equation but not in the decision equation; thus the condition for identification is satisfied.

The resulting maximum likelihood estimates are presented in Table IV. Of particular interest is the estimated coefficient of the earnings variable ($\Delta \log \hat{Y}$). The estimates reveal that the most significant factor determining migrant status is the migrant-non-migrant earnings differential. Its relatively large positive value lends strong support to the essential hypotheses of the conventional human capital model of migration. Specifically, the effect of expected monetary gains is to significantly increase the probability of migration. An additional point of interest is that the magnitudes and standard errors of the other coefficients are, after rounding, virtually unchanged from their counterparts in the reduced form equation. Exceptions are the constant and the self-employment dummy variable. The likelihood function is globally concave, so that the very large sample used in this study facilitates rapid convergence of estimates to their maximum likelihood values. The estimates reveal, as before, that self-employment and increasing age are deterrents to migration. Also, the probability of migration increases with the rate of growth of income in the origin area, and decreases with its rate of growth in employment. In addition, the probability of migration is higher if the potential migrant is male, reflecting the position of the male wage earner as head of a typical household. Finally, the race dummy is negative but not significant, indicating that the white-non-white dichotomy has no significant effect on the probability of migration.

Table IV Probit Estimation of the Structural Form Decision Equation

Variable	Coefficient	t-value
constant	-4.791	
SE	-1.146	-9.05
ΔEMP	-1.488	-2.60
ΔINC	1.455	3.14
AGE	-.008	-5.33
RACE	-.065	-1.17
SEX	-.082	-2.14
$\Delta \log \hat{Y}$	7.293	2.43

Number of observations = 9,223
Number of migrants = 1,078
Number of non-migrants = 8,145
−2 (log-likelihood ratio) = 881.00

V. Summary and Conclusions

This study represents the first attempt to incorporate endogenous selectivity into a model of migration and income. The structural model comprises a decision equation along with earnings equations for migrants and non-migrants. Estimation of the model is conducted by a two stage procedure which has proven effective in other empirical studies of self-selection. Results of estimation lead to the major findings of the study: (1) there is strong evidence of self-selection in the earnings of non-migrants; (2) the effect of the self-selection process on unconditional expected earnings is positive; and (3) the structural decision equation has among its explanatory variables expected changes in earnings, thus supporting the traditional Sjaastad framework for the human capital approach to migration.

The results of this study indicate that viable opportunities for research exist in areas of estimating returns to migration. The presence of self-selection necessitates certain refinements in the manner that potential migrants are hypothesized to view their own expected gains from migration. Research in these areas will contribute to the further refinement of human capital models of locational choice.

References

1. Galloway, L. E., "The Effect of Geographic Labor Mobility on Income: A Brief Comment." *Journal of Human Resources*, Winter 1969, 103–109.
2. Greenwood, M. J., "A Simultaneous-Equation Model of White and Nonwhite Migration and Urban Change." Mimeograph, 1973.

3. ———, "Research on Internal Migration in the United States: A Survey." Journal of Economic Literature, June 1975, 397–433.

4. Gronau, Reuben, "Wage Comparisons: A Selectivity Bias." *Journal of Political Economy*, November 1974, 1119–43.

5. Heckman, James, "Shadow Prices, Market Wages and Labor Supply." *Econometricia*, July 1974, 679–94.

6. Johnson, N. L. and S. Kotz. *Distributions in Statistics: Continuous Multivariate Distributions*. New York: Wiley, 1972.

7. Kaluzny, R., "Determinants of Household Migration: A Comparative Study by Race and Poverty Level." *Review of Economics and Statistics*, August 1975, 269–74.

8. Kenny, Lawrence., Lung-fei Lee, G. S. Maddala, and Robert P. Trost, "Returns to College Education: An Investigation of Self-Selection Bias Based on the Project Talent Data." *International Economic Review*, forthcoming.

9. Lee, Lung-fei. *Estimation of Limited Dependent Variable Models by Two State Methods*. Unpublished Ph.D. dissertation, University of Rochester, 1976.

10. ———, "Unionism and Wage Rates: A Simultaneous Equations Model with Qualitative and Limited Dependent Variables." *International Economic Review*, forthcoming.

11. Lee, Lung-fei, G. S. Maddala, and R. P. Trost, "Instrumental Variable Estimation of Simultaneous Equation Systems with Selectivity." Discussion paper, University of Florida, October 1977.

12. Lewis, H. Gregg, "Comments on Selectivity Biases in Wage Comparisons." *Journal of Political Economy*, November 1974, 1145–56.

13. Maddala, G. S., "Self-Selectivity Problems in Econometric Models," in Applications of Statistics, edited by P. R. Krishniah. Amsterdam: North Holland Publishing Company, 1977.

14. Maddala, G. S. and Forest D. Nelson, "Switching Regression Models with Exogenous and Endogenous Switching." *Proceedings of the Business and Economic Statistics Section*, American Statistical Association, 1975, 423–26.

15. Maddala, G. S. and R. Blaine Roberts, "Estimation of Econometric Models Involving Self-Selection." Paper from European Meetings of the Econometric Society in Vienna, September 1977.

16. Miller, E., "Is Out-Migration Affected by Economic Conditions?" *Southern Economic Journal*, January 1973, 396–405.

17. Mincer, J., "Family Migration Decisions." *Journal of Political Economy*, October 1978, 749–73.

18. Navratil, Frank J. and James J. Doyle, "The Socioeconomic Determinants of Migration and the Level of Migration." *Southern Economic Journal*, April 1977, 1547–59.

19. Nelson, Forest D., *"Estimation of Economic Relationships with Censored, Truncated, and Limited Dependent Variables."* Unpublished Ph.D. dissertation, University of Rochester, 1975.

20. Persky, J. J. and Kain, J. F., "Migration, Employment, and Race in the Deep South." *Southern Economic Journal*, January 1970, 268–76.

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

22. Roberts, R. Blaine, G. S. Maddala, and Gregory Enholm, "Determinants of the Requested Rate of Return and the Rate of Return Granted in a Formal Regulatory Process." *Bell Journal of Economics*, Fall 1978, 611–21.

23. Sjaastad, L., "The Costs and Returns of Human Migration." *Journal of Political Economy*, October 1962, 80–93.

24. Trost, Robert P. and J. T. Warner, "The Effects of Military Occupational Training on Civilian Earnings: An Income Selectivity Approach." Presented to the meetings of the Southern Economic Association, Washington, D.C., November 1978.

25. Vanderkamp, J., "Migration Flows, Their Determinants and the Effects of Return Migration." *Journal of Political Economy*, September/October 1971, 1012–31.