NBER WORKING PAPERS SERIES

SELF-SELECTION AND INTERNAL MIGRATION IN THE UNITED STATES

George J. Borjas

Stephen G. Bronars

Stephen J. Trejo

Working Paper No. 4002

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
February 1992

This paper is part of NBER's research program in Labor Studies. Any opinions expressed are those of the authors and not those of the National Bureau of Economic Research.

NBER Working Paper #4002
February 1992

SELF-SELECTION AND INTERNAL MIGRATION IN THE UNITED STATES

ABSTRACT

Within the conceptual framework of the Roy model, this paper provides an empirical analysis of internal migration flows using data from the National Longitudinal Surveys of Youth. The theoretical approach highlights regional differences in the returns to skills: regions that pay higher returns to skills attract more skilled workers than regions that pay lower returns. Our empirical results suggest that interstate differences in the returns to skills are a major determinant of both the size and skill composition of internal migration flows. Persons whose skills are most mismatched with the reward structure offered by their current state of residence are the persons most likely to leave that state, and these persons tend to relocate in states which offer higher rewards for their particular skills.

George J. Borjas
Department of Economics
University of California
at San Diego
La Jolla, CA 92093
and NBER

Stephen G. Bronars
Department of Economics
University of California
at Santa Barbara
Santa Barbara, CA 93106

Stephen J. Trejo
Department of Economics
University of California
at Santa Barbara
Santa Barbara, CA 93106

I. Introduction

The population of the United States is highly mobile, with about 3 percent of the population moving across state lines in any given year, and almost 10 percent of the population changing states over a five-year period [15, p. 51]. As fertility rates remain at low levels following the baby boom, internal migration has become an increasingly important source of regional demographic change and a major determinant of concurrent shifts in regional economic growth.

The theoretical and empirical study of internal migration has a long history in economics [10,11]. Guided by the income-maximizing models of Hicks [14] and Sjaastad [23], early empirical research focused on explaining the size and direction of migration flows, as well as determining why certain groups of individuals, such as the highly educated, are more likely to migrate than others [5,9,22]. Later work fruitfully emphasized the role of the family in migration decisions [6,17]. However, despite this extensive early literature and the central role played by internal migration in the operation of a competitive economy, we believe it is fair to conclude that the study of internal migration has not been at the forefront of research in mainstream labor economics over the past decade.[1]

This recent disinterest is not symptomatic of a more general indifference to the study of geographic mobility. By contrast, the analysis of international migration has flourished in recent years [1,4]. The immigration literature suggests that embedding the Hicks-Sjaastad income-maximizing approach within Roy's [21] self-selection model generates important new insights [3].

The current paper presents an application of this more general approach to the analysis of internal migration flows within the United States. We argue that the Hicks-Sjaastad framework is too restrictive for studying internal migration. Its key predictions are that persons migrate from low-income regions to high-income regions and that increases in mobility costs deter migration. In turn, these predictions focus empirical work almost exclusively on the size and direction of population flows across regions. Although the data generally support these predictions, there are many other interesting and important questions left unaddressed.

The Hicks-Sjaastad model emphasizes the fact that mean income levels differ across regions, and these income differentials (net of migration costs) generate unidirectional migration flows. Conversely, the theoretical approach suggested by the Roy model stresses regional differences in the returns to skills (as well as regional differences in mean income). These skill-price differentials determine the skill composition of migration flows. Regions that pay higher returns to skills attract more skilled workers than regions that pay lower returns. Because the economic impact of migration depends on which people move as well as on how many people move, we believe that these issues are as important as those that have dominated the literature.

Of course, earlier studies have recognized that migrants are self-selected. The development of econometric techniques to account for selection bias [12] led to several applications of this methodology to the analysis of migrant earnings. Nakosteen and Zimmer [18] and Robinson and Tomes [20] report the standard selectivity-corrected earnings functions in the mover and stayer samples. These estimates, however, are based on a conceptual framework which accommodates only one potential destination, and the studies fail to test for the presence of the equilibrium sorting predicted by the economic theory of selection.

Within the conceptual framework of the Roy model, we provide an empirical analysis of internal migration flows using data from the National Longitudinal Survey of Youth. As shown in Table 1, internal migration is quite prevalent among these young men and women.[2] More than a quarter of the sample currently (i.e., in 1986) resides in a state other than the state of birth, and about 18 percent are living in a different state than at age 14. The data also indicate that roughly 6 percent of the sample moves across state boundaries in any given year.

Our empirical analysis indicates that interstate differences in the returns to skills are a major determinant of both the size and skill composition of internal migration flows. Persons whose skills are most mismatched with the reward structure offered by their current state of residence are the persons most likely to leave that state, and these persons tend to relocate in states which offer higher rewards for their particular skills. These results suggest that the Roy model provides a useful framework for analyzing internal migration flows in the United States.

II. Theory

Consider a country partitioned into k distinct geographic regions, indexed by $i=1,\ldots,k$. To simplify the exposition, we initially assume that there are no costs of relocating across regions. Individuals compare their earnings opportunities in the various regions and move to the location that maximizes their earnings. We also assume that the <u>initial</u> distribution of individual skills is the same in all regions. At the time of birth, therefore, individuals are randomly allocated across regions in terms of their skills. The population log earnings distribution in region i is given by:

$$\log w_i = \mu_i + v_i, \qquad i=1,\ldots,k \tag{1}$$

where μ_i is the mean income that would be observed in region i in the absence of any internal migration, and v_i is a random variable with mean zero and variance σ_i^2 that measures person-specific deviations from mean income in region i.

The assumption that the initial skill distributions are identical across regions makes the income distributions in (1) independent of initial conditions: the same earnings generation process applies to all individuals currently residing in a given region, regardless of where these individuals were born.[3] However, because of regional differences in natural resources, physical capital, and aggregate economic conditions, mean incomes μ_i and the distributions of the random variables v_i will vary across regions.

An income-maximizing individual chooses to reside in region j whenever:

$$\log w_j > \max_{i \neq j}[\log w_i] \tag{2}$$

The characteristics of the sorting generated by (2) cannot be described without additional restrictions on the distribution of the random variables $v_1, \ldots, v_k$. A simplifying assumption which allows a complete characterization of the equilibrium sorting is that individual earnings are perfectly correlated across regions, so that $\mathrm{Corr}(v_i, v_j)=1$ for all i,j. The population income distribution of region i can then be written as:

$$\log w_i = \mu_i + \eta_i v, \qquad i=1,\ldots,k \tag{1'}$$

This specification implies that the same random variable v determines an individual's potential earnings in each of the various regions. Thus, v indexes an individual's ability or skills, and equation (1') assumes that the earnings determination process can be characterized by a one-factor model of ability. The coefficient η_i can be interpreted as a factor-loading parameter, or more generally as the "rate of return" to skills in region i. It is convenient to label the regions such that they are ranked in terms of η with $\eta_1<\eta_2<\ldots<\eta_k$. We assume that v is a continuous random variable with mean zero and a range defined over the real number line.

Although the assumption that earnings are perfectly correlated across regions is quite strong (because it implies that the ranking of individuals by skill level is the same in all regions), it enables us to derive a number of testable implications from a multi-region selection model. Furthermore, this framework may provide a reasonably accurate representation of earnings opportunities across regions within the United States, given the relatively strong regional similarities in economic, legal, and social institutions.

Note that the random variable v need not be observed by the researcher.

Individuals sort themselves across regions on the basis of all of their skills, not just those that happen to be econometrically convenient. By using a one-factor model of ability, we assume that the relative prices of all skills are the same across regions, and so the composite commodity theorem allows us to focus on a single skill that is being "sold" across regions. It is possible to analyze the migration decision within the context of a multi-factor model of ability, but we do not pursue this generalization because it detracts from the main points that we make in this paper.

Using equations (1′) and (2), region j is preferred to region i whenever $(\eta_j-\eta_i)v > (\mu_i-\mu_j)$. This implies that $v > (\mu_i-\mu_j)/(\eta_j-\eta_i)$ for j>i, and that $v < (\mu_i-\mu_j)/(\eta_j-\eta_i)$ for j<i (because $\eta_j-\eta_i$ has the same sign as j-i). Region j is chosen only if it is preferred to all other regions. As a result, individuals sort themselves across regions according to the following inequalities:

$$\text{Choose region 1:} \qquad v < \min_{i=2,\ldots,k} \left[\frac{\mu_1-\mu_i}{\eta_i-\eta_1}\right] \tag{3a}$$

$$\underset{(1<j<k)}{\text{Choose region j:}} \qquad \max_{i=1,\ldots,j-1} \left[\frac{\mu_i-\mu_j}{\eta_j-\eta_i}\right] < v < \min_{i=j+1,\ldots,k} \left[\frac{\mu_j-\mu_i}{\eta_i-\eta_j}\right] \tag{3b}$$

$$\text{Choose region k:} \qquad v > \max_{i=1,\ldots,k-1} \left[\frac{\mu_i-\mu_k}{\eta_k-\eta_i}\right] \tag{3c}$$

So long as the range of v extends far enough, equations (3a) and (3c) imply that some individuals will choose to reside in the "extreme" regions 1 and k (i.e., the regions with the lowest and highest values of η). It is less apparent, however, that all of the interior regions will be populated. Depending upon the parameter values, it is possible to have regions for which equation (3b) is never satisfied, and hence no individuals locate in these regions.

Because unpopulated regions are of no interest empirically, we restrict our attention to regions where some individuals choose to reside. Equation (3b) implies that a necessary condition for region j to be inhabited is:

$$\frac{\mu_{j-1}-\mu_j}{\eta_j-\eta_{j-1}} < \frac{\mu_j-\mu_{j+1}}{\eta_{j+1}-\eta_j} \tag{4}$$

This can be rewritten as:

$$\mu_j > \frac{(\eta_{j+1}-\eta_j)\mu_{j-1} + (\eta_j-\eta_{j-1})\mu_{j+1}}{(\eta_{j+1}-\eta_{j-1})} \tag{5}$$

We assume that equation (5) is satisfied for all regions j (j=2,...,k-1). Equation (5) then defines the Existence Condition that mean income in region j must satisfy in order for region j to attract and retain a population. This condition greatly simplifies the characterization of the equilibrium sorting. In particular, repeated use of the Existence Condition to make pairwise comparisons of the arguments in the min(·) and max(·) expressions in (3) yields the following inequalities:

Choose region 1: $v < \frac{\mu_1 - \mu_2}{\eta_2 - \eta_1}$ (6a)

Choose region j: (1<j<k) $\frac{\mu_{j-1} - \mu_j}{\eta_j - \eta_{j-1}} < v < \frac{\mu_j - \mu_{j+1}}{\eta_{j+1} - \eta_j}$ (6b)

Choose region k: $v > \frac{\mu_{k-1} - \mu_k}{\eta_k - \eta_{k-1}}$ (6c)

Figure 1 illustrates the nature of the equilibrium skill sorting when there are five regions. The least skilled workers move to the region with the lowest rate of return to skills, while the most skilled workers choose the region with the highest rate of return. Persons with intermediate levels of skills move to intermediate regions, with the more skilled workers choosing regions with higher rates of return. In effect, income-maximizing behavior induces a positive correlation between the average skill level of a region's inhabitants and the region's rate of return to skills:

$$E(v|\text{choose } i) > E(v|\text{choose } j) \text{ if and only if } \eta_i > \eta_j \qquad (7)$$

The assumption that earnings are perfectly correlated across regions implies that individuals who rank highly in the income distribution of one region would also rank highly in the income distribution of any other region. Skilled workers, therefore, are attracted to high-η regions because these workers can then enjoy a more generous return on their superior skills. In contrast, unskilled workers choose regions with less income inequality because this minimizes the economic penalty for lacking human capital. In essence, skill prices play an important allocative role in the internal migration decision.

This insight helps explain the economic content of the Existence Condition. In order for region j to be inhabited, the inequality in (5) requires that mean earnings in region j exceed a weighted average of mean earnings in the "neighboring" regions j-1 and j+1. Note that these neighboring regions need not be geographically adjacent, but are instead neighbors in an economic sense. Because neighboring regions offer relatively similar rewards for the skills of potential migrants, these are the regions which compete with region j in attracting human capital.

Suppose that mean earnings in region j are below mean earnings in both neighboring regions. The Existence Condition is not satisfied and no individuals choose to locate in j. For some persons to reside in region j, mean earnings in j must exceed mean earnings in either region j-1 or region j+1, or both. Because these neighboring regions offer either a lower or a higher rate of return to skills than region j, they hold a natural advantage over j in attracting residents. In other words, for the same mean earnings, skilled individuals ($v>0$) prefer the region with a higher rate of return to skills, while unskilled individuals ($v<0$) prefer the region which least penalizes their lack of skills. Therefore, if mean earnings were equal in all three of these regions, or if mean earnings in j were lower than mean earnings in both of the neighboring regions, region j does not make a competitive offer to potential migrants. In contrast, a sufficiently higher mean income in region j than in either of its neighboring regions compensates potential migrants for region j's relative disadvantage and attracts a population.

The Existence Condition imposes a specific pattern of economic opportunities across populated regions. In effect, the Existence Condition rules out the case where the relationship between μ and η is U-shaped and the case where μ is constant across regions. Consider any three regions that are adjacent in terms of the rate of return to skills they offer. All three regions can be populated if μ is monotonically increasing or decreasing in η, or if μ and η are related in an inverted-U shape. However, if the relationship between μ and η is flat or U-shaped for any three neighboring regions, then the middle region would not be able to compete with its neighbors and would fail to attract any residents.

This discussion suggests an important avenue for future research. Regions can attract migrants only if they make competitive offers. In a more general model, the parameters that summarize regional income distributions are themselves endogenous, and the equilibrium income distributions are determined simultaneously with the equilibrium skill sorting of workers across regions. This general equilibrium model would also introduce the role played by the prices of fixed factors, such as land. Although research on this topic is in its infancy (see, for instance, [13] and [19]), it is clear that this type of analysis will provide a much deeper characterization of spatial equilibrium.

The discussion also highlights a feature of the k-region selection model that is shared by the standard two-region Roy model. In both models, the ranking of skill prices across regions completely determines where a region ranks in terms of the average skill level of its inhabitants. As long as the Existence Condition is satisfied, mean incomes play no role in determining the skill ranking of regions. Of course, mean incomes do affect the size and skill composition of the population that chooses to reside in any given region, and in this way mean incomes influence the average skill level of a region's inhabitants. It is obvious from Figure 1, for example, that a region attracts more residents when its mean income rises.

Our approach not only raises a number of new substantive issues regarding the internal migration process, but also simplifies the empirical analysis of internal migration flows. Earlier work has been hampered by the fact that there are k(k-1) possible migration flows in a k-region model, and the size and composition of each of these flows depend on all of the parameters of the model. Given the Existence Condition, equation (6) implies that the size and skill composition of the population choosing region j can be completely determined from the parameters of the income distributions for that region and the two neighboring regions, greatly diminishing the number of parameters that influence migration flows into any given region. In fact, our framework implies that internal migration can be analyzed using an ordered qualitative choice model.

Our theory also implies that region j can be both a source and a destination for migrants. As long as skill prices differ across regions, the spatial missorting of individuals at the time of birth is likely to be substantial. Skilled individuals, for instance, may be born in low-η regions and subsequently move to high-η regions, while less able workers move in the opposite direction. Two-way population flows occur naturally as the mismatches caused by being born in the wrong region are corrected.

The introduction of migration costs does not alter any of the key results. For concreteness, consider the migration decisions of persons born in region i. Migration to region j ($j \neq i$) takes place whenever:

$$\log w_j - C_{ij} > \max_{r \neq j}[\log w_r - C_{ir}] \tag{8}$$

where C_{ij} is a time-equivalent measure of the costs of migrating from region i to region j, with $C_{ii}=0$.[4] For simplicity, we assume that migration costs C_{ij} are the same for all persons currently residing in region i.[5] Of course, we do allow migration costs to vary when moving from region i to different destination regions (i.e., $C_{ir} \neq C_{is}$ for $r \neq s$).

Assuming initially that every region receives at least one migrant from region i (a restriction analogous to the Existence Condition), the equilibrium sorting of individuals born in region i can be derived:

$$\text{Choose region 1:} \quad v < \frac{\mu_1 - \mu_2 - (C_{i1} - C_{i2})}{\eta_2 - \eta_1} \tag{9a}$$

$$\underset{(1<j<k)}{\text{Choose region j:}} \quad \frac{\mu_{j-1} - \mu_j - (C_{i,j-1} - C_{ij})}{\eta_j - \eta_{j-1}} < v < \frac{\mu_j - \mu_{j+1} - (C_{ij} - C_{i,j+1})}{\eta_{j+1} - \eta_j} \tag{9b}$$

$$\text{Choose region k:} \quad v > \frac{\mu_{k-1} - \mu_k - (C_{i,k-1} - C_{ik})}{\eta_k - \eta_{k-1}} \tag{9c}$$

Figure 2 illustrates the sorting of workers born in region three when there are five regions and it is costly to move. It is apparent that the equilibrium sorting resembles that obtained when mobility is costless: skilled workers move to high-η regions and unskilled workers move to low-η regions. The introduction of migration costs, however, alters the cutoff points determining who moves to which region. These thresholds now depend on mean incomes net of migration costs. This fact obviously implies that fewer persons will leave their region of birth.

A simple parameterization of migration costs reveals exactly which interregional flows are most likely to be affected by the fact that internal migration is costly. Suppose that the costs of moving from region i to region j are $C_{ij}=\bar{C}$ for $i \neq j$, and 0 otherwise. In this specification, migration costs are simply the fixed costs of moving that do not depend on the distance of the move or any other factors which vary with the precise identities of the origin and destination. From Figure 2, it is clear that these fixed costs cancel out of all of the cutoff points except those bordering the region of origin.

Consider an increase in the fixed costs of moving. Obviously, this increases the fraction of region i residents who decide not to migrate. For "small" changes in $\bar{C}$, the pool of individuals who previously would have migrated but now decide to remain in i is drawn entirely from those individuals who would have moved to neighboring regions (i.e., regions with neighboring values of η). Therefore small changes in fixed migration costs do not alter the size or skill composition of the outflow to non-neighboring regions. Put differently, small changes in fixed migration costs only change the incentives of "marginal" migrants.

Of course, the larger the increase in fixed migration costs, the more likely it is that no one will move to the neighboring regions i-1 and i+1. Sufficiently high levels of migration costs make it unlikely that all regions are destinations for persons originating in region i. Moreover, those regions which fail to attract migrants from region i are those which most resemble region i in terms of the payoff to skills.

We noted above that our model generates two-way migration flows without resorting to imperfect information or informational asymmetries among participants in the marketplace. The existence of migration costs adds further substance to this insight. As fixed migration costs increase, fewer people move to neighboring regions from any region of origin. There is a tendency, therefore, for persons to migrate to relatively "extreme" regions (i.e., regions with high or low levels of η). But these are precisely the regions where the costs of being mismatched are largest, and hence extreme regions will also be the origin of sizable migration flows. This suggests that extreme regions simultaneously experience large inflows and outflows of migrants. Of course, the the magnitude of these flows depends on the exact distribution of skills, on the levels of fixed and variable migration costs, and on the parameters of the income distributions in each of the k regions. It is therefore difficult to quantify the importance of this tendency without additional restrictions on the model.

III. Empirical Analysis

The theory developed in the previous section generates sharp empirical predictions about the relationship between regional differences in the returns to skills and such diverse factors as the migration propensities of individuals, the direction and composition of migration flows, and the spatial distribution of skills. To test these predictions, we analyze the 1979-1986 waves of the National Longitudinal Survey of Youth (NLSY). Because young workers have not yet accumulated a great deal of job- and location-specific human capital and because they have a long working life remaining over which to collect returns, they should be especially responsive to economic incentives for migration (recall the high rates of geographic mobility reported in Table 1).

NLSY respondents are between the ages of 14 and 22 at the time of the first interview, and the subsequent annual interviews provide a detailed history of each individual's labor market activity and geographic mobility. In order to focus exclusively on internal migration, we exclude individuals born outside the United States or ever observed to reside abroad. In order to mitigate the impact of extraneous factors on migration flows, we also exclude individuals who left school after 1984 or who were ever members of the military.

The NLSY reports each individual's state of residence at age 14 and his state of residence at the time of each of the eight interviews.[6] Because tracking geographic location is central to our analysis, we exclude individuals for whom this information is incomplete. We define as movers those individuals who reside in a different state in 1986 than at age 14, regardless of where they lived during the intervening years. Similarly, non-movers are those who reside in the same state in 1986 as at age 14, even if they lived elsewhere in between.

We use state at age 14 as the place of origin for two reasons. First, this maximizes the sample size because alternative definitions such as state of residence when the individual first enters the labor market are not available for those older respondents who had already joined the labor force when the survey began in 1979. Second, and more importantly, state of residence at age 14 is likely to be exogenously determined by parental location decisions. Alternative definitions introduce endogeneity because the location at the time of labor market entry already reflects the individual's initial decision as to where he would like to live and work. Our definition instead exploits the fact that optimal location decisions for parents and children need not coincide, so spatial mismatches can arise at age 14. We focus on the subsequent migration decisions made by young workers as they attempt to correct these mismatches.

Although SMSAs may better approximate local labor markets, the preceding considerations led us to adopt states as the geographic unit of analysis, because the NLSY does not report SMSA of residence at age 14. In addition, state boundaries are stable over time and create an exhaustive partition of the United States, whereas SMSAs do not share these features.

We construct four alternative measures of a worker's skills. The first measure is the number of years of completed education (as of 1986). The second measure is based on aptitude test scores. Between July and October 1980, the Armed Services Vocational Aptitude Battery (ASVAB) was administered to about 94 percent of the NLSY respondents. The ASVAB consists of 10 tests that measure knowledge and skills in areas ranging from word knowledge and arithmetic reasoning to mechanical comprehension and electronics information. The military sums the scores of four of these tests (word knowledge, arithmetic reasoning, paragraph comprehension, and half of the score in numeric operations) to create the Armed Forces Qualification Test (AFQT). The AFQT is a general measure of aptitude, and its score is standardized so that the population distribution has mean zero and a standard deviation of one.

The final two measures of worker skills are based on an individual's average hourly wage, defined as the ratio of annual earnings to annual hours of work. We exploit the panel aspect of the NLSY to estimate individual-specific fixed effects. As a result, the wage-based skill measures are available only for the subsample of workers who have at least two years of wage data.[7] These wage observations need not be in consecutive years.

The wage-based skill measures are constructed as follows. Consider the earnings function:

$$\log w_{ijt} = \beta_1 STATE_{it} + \beta_2 YEAR + \beta_3 X_{it} + \epsilon_{ijt} \qquad (10)$$

where w_{ijt} is individual i's hourly wage in state j in year t. The wage is a function of a vector of dummy variables indicating the current state of residence (STATE), a vector of dummy variables indicating the year of the observation (YEAR), and a vector of control variables (X). The control variables include age, age squared, years of completed education, job tenure, union status, marital status, health status, metropolitan residence, industry, and occupation.

The error term in (10) depends on the state/worker match. We assume that $\epsilon_{ijt}=\eta_j(v_i+u_{it})$. This decomposes the wage residual into the product of the stock of (unobservable) person-specific human capital (v_i) and the state-specific return to human capital (η_j), plus a random error term ($\eta_j u_{it}$). Differences in the return to human capital across states generate heteroskedasticity in the earnings function.

Because we have at least two observations on hourly wage rates for each individual in the sample, we can compute estimates of the v_i. If all states paid the same return to skills, the v_i could be estimated by simply adding person-specific intercepts to the regression. However, because skills are rewarded differently across states, we add separate intercepts for each worker/state pair in the sample. The error specification implies that these worker/state intercepts are proportional across states (i.e., that ϵ_{ijt} is proportional to η_j). Note also that the vector of control variables (X) should include only those characteristics that vary over the sample period. Differences in earnings due to person-specific factors that are not time-varying (such as race and sex) are captured by the v_i.

The first wage-based measure of worker skills is obtained by estimating (10) using data differenced from person-specific means, and then calculating for each individual his average residual (appropriately weighted for state of residence).[8] This procedure yields point estimates that are identical to those that would result from adding thousands of worker/state intercepts to the regression. This skill measure, which we call the "standardized wage," represents the number of standard deviations that a worker's hourly wage is above or below the mean wage for workers with similar demographic characteristics.

The second wage-based measure of skills does not control for demographic characteristics. This "unstandardized wage" is calculated as described above, except that the control vector X is omitted from equation (10). The unstandardized wage measures the number of standard deviations that a worker's wage is above or below the state average, without controlling for any observable factors.

Given our sample selection criteria, some measure of worker skills is available for a total of 6666 individuals in the NLSY. However, not all individuals report each of the four different skill measures. AFQT scores are available for 6510 individuals, and only 5182 individuals report the necessary information to construct standardized and unstandardized wages. The empirical results presented below are based on the maximum sample size possible for each skill measure. Virtually identical results were obtained when the empirical analysis was performed on the subsample of 5064 individuals who reported all four skill measures.

Table 2 presents means and standard deviations of the four skill measures and selected demographic characteristics. The summary statistics are provided for the entire sample, as well as separately for movers and non-movers. On average, movers are more skilled than non-movers, regardless of how skills are measured.

Table 3 reveals that, although the four skill measures represent different aspects of ability, they are highly correlated across individuals. The top panel presents the matrix of correlation coefficients for these variables, and the bottom panel presents partial correlations that first control for age, sex and race. The correlations among the skill measures are uniformly strong and positive, and the correlations are not appreciably weaker within age/race/sex groups.

Before proceeding with more formal statistical tests, Tables 4 and 5 display the patterns of interstate migration that emerge from the NLSY data. Table 4 describes out-migration from selected states (the 25 states with the largest sample sizes). The first column gives the number of persons in our sample who resided in each state as of age 14, and the second column reports the fraction of these "natives" who left the state by 1986. The remaining columns describe how the skills of movers differ from the skills of those who remained in the state. In order to facilitate interstate comparisons, we normalize the mean skill level in each state to be zero. Thus a weighted average of the skills of movers and stayers equals zero in each state.

There is substantial interstate variation in the abilities of out-migrants. For example, Massachusetts exports its most able workers. Compared to the overall state average, migrants from Massachusetts score about .4 higher on the AFQT, have completed more than an additional year of schooling, and command higher wage rates. On the other hand, states such as Minnesota and West Virginia export young workers who are below-average in all skill measures.

Table 5 provides a complimentary description of in-migration. The first column gives the number of persons residing in each state as of 1986, and the second column reports the fraction of these individuals who lived elsewhere at age 14. The remaining columns present the average skill levels of in-migrants, measured relative to the mean skill level of natives in the migrant's state of origin.[9]

As was the case with out-migrants, the data reveal substantial variation in the skills of in-migrants. For instance, the average person migrating to Connecticut has one more year of education than the average native in the migrant's home state (as well as higher AFQT scores and wage rates). This does not necessarily mean that Connecticut is importing highly skilled workers in an absolute sense, but rather that Connecticut attracts young workers who are more able than the average native in those states that export workers to Connecticut.

The model presented in Section II implies that the equilibrium sorting of skills across states is largely determined by the parameters η_j, the state-specific returns to skills. In the context of the model, relative skill prices are proportional to the extent of earnings inequality, and so we use the standard deviation of the wage distribution within each state to measure the returns to skills. Because of the small sample sizes for some states in the NLSY data, we use samples of male, private sector workers from the 5/100 1980 Census microdata to estimate wage dispersion within each state.

We construct two measures of wage dispersion. The first is the standard deviation of the log hourly wage, which we call the "unstandardized" dispersion in wages. The alternative "standardized" measure of dispersion is the root mean square error from state-specific log wage regressions. This measure represents an estimate of the residual wage variation that remains after controlling for observable demographic characteristics. The control variables in these wage regressions include education, age, age squared, and dummy variables indicating marital status, immigrant status, and metropolitan residence.

The state-specific measure of unstandardized wage dispersion ranges from .579 (Maine) to .745 (Alaska), with a mean of .654 and a standard deviation of .032. The standardized measure of wage dispersion has a mean of .339 and a standard deviation of .040. The standardized and unstandardized measures of wage dispersion are highly correlated across states, with a correlation coefficient of .895. Southern and western states tend to display greater wage dispersion than the rest of the country.

According to the self-selection model, migration decisions are motivated by an initial mismatch between workers and states. The larger the initial mismatch for a given worker, the more likely he is to leave his native state. Skilled workers are more likely to leave states where skill prices are relatively low, even though mean earnings in the state are sufficiently high to retain other workers. Conversely, unskilled workers are more likely to leave states where wage dispersion, and hence skill prices, are relatively high. The theory thus predicts that the correlation between skill levels and out-migration rates should be more positive in states with little earnings inequality than in states with a large amount of dispersion.

To test this implication, we estimate probit models where the dependent variable is a dummy identifying those individuals who eventually left their native state, and the independent variables include a measure of skills as well as dummy variables for race (white, black, or Hispanic) and sex.

Separate probits are estimated for each of the four alternative skill measures. In order to allow the magnitude of the initial mismatch to influence the probability of out-migration, we segregate the sample according to state of origin. In particular, we divide the sample into four approximately equal-sized partitions, with the grouping based on the rank order of the unstandardized wage dispersion in each individual's native state.[10] Through the use of interaction terms, we allow the effect of skills on out-migration rates to differ according to which state group the individual resided in at age 14.

Table 6 reports the resulting estimates. For each skill variable in each of the sample partitions, the table presents the probit coefficient and the implied effect of a one standard deviation change in skills on the out-migration rate, computed at sample means. The first column reports these statistics for the quarter of the sample who, at age 14, resided in states ranking lowest in unstandardized wage dispersion. The remaining columns present the same information for individuals originating in states with progressively more wage dispersion.

The positive coefficients indicate that out-migration rates are higher for the more skilled, regardless of the origin state. However, the results also suggest that this correlation is stronger in states with less wage dispersion (columns 1 and 2) than in states with more wage dispersion (columns 3 and 4). For example, a one standard deviation increase in education raises by about 4.8 percentage points the probability that a worker leaves a state from the lowest wage dispersion group, while the same change in education has a much smaller impact (1.5 percentage points) on the out-migration rate of individuals from states in the highest dispersion group. The same pattern emerges for the other skill measures. Because low levels of wage dispersion indicate low returns to skills, the results confirm the theoretical prediction that high ability workers are more likely to leave states with relatively low skill prices.

The theoretical model also predicts that skilled workers move to states with greater wage dispersion and unskilled workers move to states with less wage dispersion. We test this implication by viewing the change in wage dispersion between the native (age 14) state and the current (1986) state as a choice variable. For individuals currently residing in their native state, this change is zero. Among movers, the mean change in the unstandardized wage dispersion is .012, while the mean change in the standardized dispersion is .013. Therefore, on average, the young workers in the NLSY migrate to states with greater wage dispersion. The self-selection model predicts that changes in wage dispersion should be positively related to skill levels.

We begin testing this hypothesis by estimating least-squares regressions of the change in wage dispersion on race and sex dummies and, in separate regressions, the four alternative measures of skills. Table 7 presents the estimated coefficients of the skill variables. For readability, the coefficients have been multiplied by 100. The regressions reported in the first two columns were run on the entire sample, including non-movers. Regardless of whether the dependent variable is defined using the unstandardized or standardized measure of wage dispersion, there is a strong and statistically significant positive relationship between each of the skill variables and the change in wage dispersion.

The last two columns of Table 7 present similar estimates for the subsample of movers. The coefficients are uniformly larger than those obtained from the full sample. Evidently, skills endowments have an important influence on the direction of internal migration flows. Moreover, these effects are economically important. For instance, a one standard deviation increase in schooling raises the change in wage dispersion by about 25 percent, and a comparable increase in the AFQT score yields an even larger rise.

An alternative way of investigating patterns of internal migration is to model the direction but not the magnitude of the change in wage dispersion. Each migrant has two choices: move to a state with less wage dispersion than the native state, or move to a state with greater dispersion. This discrete representation of migration patterns may be superior to the continuous dependent variable used in Table 7 if our measures of wage dispersion are plagued by substantial measurement error.

Table 8 presents the results of this alternative specification. Probit models are estimated on the subsample of movers, with the dependent variable identifying those workers who moved to states with greater wage dispersion than their native state. The estimated coefficients confirm the results of Table 7 in that there is a strong positive relationship between skill levels and the probability of moving to a state with increased wage dispersion.

The theoretical model also implies that persons locate across states according to a rank ordering of their abilities. In the absence of mobility costs, the state with the highest returns to skills attracts the most able workers, whereas the state with the second highest returns to skills attracts workers who are less able than those in the first state but more able than those in the state with the third highest returns to skills, and so on. Put differently, ranking states by the average skills of their residents should produce the same outcome as ranking states by the extent of wage dispersion. Allowing for mobility costs dampens but does not fundamentally alter this pattern.

An empirical test of this implication can be conducted by estimating an ordered probability model of an individual's state of residence in 1986. Before doing this, however, it is important to determine whether the initial distribution of persons across states is random with respect to skills. Table 9 reports correlation coefficients between skill levels and wage dispersion in the native state. There is a strong negative correlation between the average skills of natives and wage dispersion. In other words, states with more wage dispersion tend to start out with less skilled youth. If mobility costs are sizable, it is unlikely that migrant self-selection will offset this initial distribution of skills and yield a positive relationship between skill levels and wage dispersion in the destination state. To control for the nonrandom initial distribution of skills, we normalize mean skills among natives in each state to be zero. By construction, these normalized skill measures are uncorrelated with wage dispersion in the native state (or, for that matter, with any other characteristic of the native state).

In order to ease the computational burden, we once again divide the sample into four approximately equal-sized partitions, but this time the grouping is based on the rank order of the unstandardized wage dispersion in each individual's 1986 state of residence.[11] The state groups are defined so that group 1 has the lowest wage dispersion and group 4 has the highest. In the absence of mobility costs, individuals sort perfectly into these four groups on the basis of their labor market skills. Let v_i represent a latent variable measuring individual i's skill level. The self-selection model generates the following equilibrium sorting of workers by skill level:

$$\text{Choice of location in 1986} = \begin{cases} 1 \text{ if } v_i < \bar{v}_1 \\ 2 \text{ if } \bar{v}_1 < v_i < \bar{v}_2 \\ 3 \text{ if } \bar{v}_2 < v_i < \bar{v}_3 \\ 4 \text{ if } \bar{v}_3 < v_i \end{cases} \tag{11}$$

As can be seen from equation (6) in Section II, the cutoff values $\bar{v}_1$, $\bar{v}_2$, and $\bar{v}_3$ depend on the parameters of the regional earnings distributions.

Let Z_i denote an observable proxy for the skills of person i, and let F_i be a vector of race and sex dummy variables. Parameterize each individual's overall skill level as $v_i = \alpha_1 Z_i + \alpha_2 F_i + e_i$, where e_i represents skills not captured by our skill measures. If unobserved skills are normally distributed, then the equilibrium sorting is described by an ordered probit model.[12] Maximum likelihood estimation yields estimates of α_1 and α_2 as well as two of the three cutoff values (one of the thresholds is arbitrarily normalized to zero). The self-selection model predicts that more skilled workers locate in regions with higher returns to skills, and this implies that the coefficient α_1 should be positive.

Table 10 reports the estimation results. The skill measures used in the first and third columns are normalized relative to the state of origin (in order to control for the initial skill distribution), and the coefficients on these variables are positive and statistically significant. These results confirm the prediction of the self-selection model. However, empirical support for the theory disappears when the skill measures are not normalized, as in the second and fourth columns of Table 10. This reflects the dominant role played by the initial nonrandom distribution of skills across states.

IV. Conclusion

This paper has analyzed the internal migration of young workers in the United States. Our research is motivated by the realization that migrants are not randomly selected from the population. We therefore adapt the Roy model of self-selection in order to study internal migration. This approach generates new theoretical insights, raises questions ignored by previous research, and simplifies the empirical analysis of multi-directional migration flows.

The self-selection model provides a framework for simultaneously analyzing questions related to the size, direction, and skill composition of internal migration flows. Income-maximizing behavior generates an equilibrium sorting of skills in which regions offering high rewards for skills attract skilled workers and unskilled workers move to regions with low skill prices. Because skilled workers currently residing in regions with low skill prices and unskilled workers living in regions with high skill prices are mismatched spatially, these workers are likely to migrate.

Migrants are expected to relocate in regions where the returns to skills are more compatible with their skill endowments. Our model extends the earlier Hicks-Sjaastad framework by emphasizing the role of skill prices in allocating workers across regions.

Our analysis of data from the National Longitudinal Survey of Youth suggests that the self-selection model provides useful insights into the internal migration process. Individuals are more likely to migrate the greater is the mismatch between their skill endowments and the returns paid to skills in their native state. Moreover, the direction and skill composition of internal migration flows seem to be guided by comparative advantage. Skilled workers tend to move to states with greater wage dispersion than their native state, whereas unskilled workers are more likely to move to states with less dispersion.

FOOTNOTES

*This research was funded by the U.S. Department of Labor. We thank Dan Hamermesh for helpful comments.

1. We do not mean to imply that internal migration research ceased altogether. A large number of studies have appeared since the mid-1970s which provide useful extensions of the empirical literature (e.g., [2,7,8]).

2. The sample will be described in greater detail in Section III.

3. This assumption ignores the possibility that persons born in region i are distinctly different, on average, from persons born in region j, and therefore the income distributions should also be subscripted for region of birth. The main results of our model are unaffected by this complication.

4. If the dollar costs of moving from region i to region j are given by D_{ij}, then time-equivalent costs are given by the ratio D_{ij}/w_i. We assume that this ratio is "small" in deriving equation (8).

5. It is not difficult to allow mobility costs to vary across individuals. In the simpler two-region model with a normal distribution of skills, it can be shown that introducing variable mobility costs does not alter any of the results if earnings and mobility costs are uncorrelated, or if the variance of mobility costs is small relative to the variance of earnings.

6. The District of Columbia is considered to be a separate state.

7. In constructing the wage-based skill measures, we restricted the sample as follows. Observations with computed hourly wage rates of less than $.50 or greater than $100 were considered outliers and excluded. We also excluded observations for which any of the following variables were

missing: industry, occupation, job tenure, health status, years of completed education, school enrollment status, marital status, union status, and whether the respondent resided in a metropolitan area.

8. We estimate the parameters in (10), including the v_i, by a two-step procedure that corrects for heteroskedasticity across states. In the first step, equation (10) is estimated by ordinary least squares and state-specific estimates of wage dispersion are calculated from the residuals. The second step uses these estimated variances to reestimate (10) by generalized least squares.

9. We measure the skills of in-migrants relative to the state of origin because there exist sizable regional differences in mean AFQT scores, years of education, and earnings.

10. The partitions are roughly but not exactly the same size because all individuals from the same native state were grouped into the same partition. Grouping states according to their standardized (rather than unstandardized) wage dispersion produces similar empirical results.

11. This contrasts with Table 6, where the grouping was based on wage dispersion in each individual's native state. The aggregation of states into four groups insures that adequate samples are observed in each region and also simplifies estimation of the ordered probit model.

12. For a discussion of ordered response models, see Maddala [16, pp. 46-49].

REFERENCES

1. J. M. Abowd and R. B. Freeman, Eds., "Immigration, Trade, and the Labor Market," University of Chicago Press, Chicago (1991).

2. A. P. Bartel, The migration decision: What role does job mobility play?, American Economic Review, 69, 775-786 (1979).

3. G. J. Borjas, Self-selection and the earnings of immigrants, American Economic Review, 77, 531-553 (1987).

4. G. J. Borjas, "Friends or Strangers: The Impact of Immigrants on the U.S. Economy," Basic Books, New York (1990).

5. S. Bowles, Migration as investment: Empirical tests of the human capital approach to geographic mobility, Review of Economics and Statistics, 52, 356-362 (1970).

6. J. DaVanzo, "Why Families Move: A Model of the Geographic Mobility of Married Couples," Report R-1972-DOL, Rand Corporation, Santa Monica (1976).

7. J. DaVanzo, Repeat migration in the United States: Who moves back and who moves on?, Review of Economics and Statistics, 65, 552-559 (1983).

8. G. S. Fields, Place to place migration: Some new evidence, Review of Economics and Statistics, 61, 21-32 (1979).

9. M. J. Greenwood, An analysis of the determinants of geographic labor mobility in the United States, Review of Economics and Statistics, 51, 189-194 (1969).

10. M. J. Greenwood, Research on internal migration in the United States: A survey, Journal of Economic Literature, 13, 397-433 (1975).

11. M. J. Greenwood, Human migration: Theory, models, and empirical studies, Journal of Regional Science, 25, 521-544 (1985).

12. J. J. Heckman, Sample selection bias as a specification error, Econometrica, 47, 153-161 (1979).

13. J. J. Heckman and G. Sedlacek, Heterogeneity, aggregation, and market wage functions: An empirical model of self-selection in the labor market, Journal of Political Economy, 93, 1077-1125 (1985).

14. J. R. Hicks, "The Theory of Wages," Macmillian, London (1932).

15. L. Long, "Migration and Residential Mobility in the United States," Russell Sage Foundation, New York (1988).

16. G. S. Maddala, "Limited Dependent and Qualitative Variables in

Econometrics," Cambridge University Press, New York (1983).

17. J. Mincer, Family migration decisions, Journal of Political Economy, 86, 749-775 (1978).

18. R. A. Nakosteen and M. Zimmer, Migration and income: The question of self-selection, Southern Economic Journal, 46, 840-851 (1980).

19. J. Roback, Wages, rents, and amenities: Differences among workers and regions, Economic Inquiry, 26, 23-42 (1988).

20. C. Robinson and N. Tomes, Self-selection and interprovincial migration in Canada, Canadian Journal of Economics, 15, 474-502 (1982).

21. A. D. Roy, Some thoughts on the distribution of earnings, Oxford Economic Papers, 3, 135-146 (1951).

22. A. Schwartz, Interpreting the effect of distance on migration, Journal of Political Economy, 81, 1153-1169 (1973).

23. L. A . Sjaastad, The costs and returns of human migration, Journal of Political Economy, 70, 80-93 (1962).

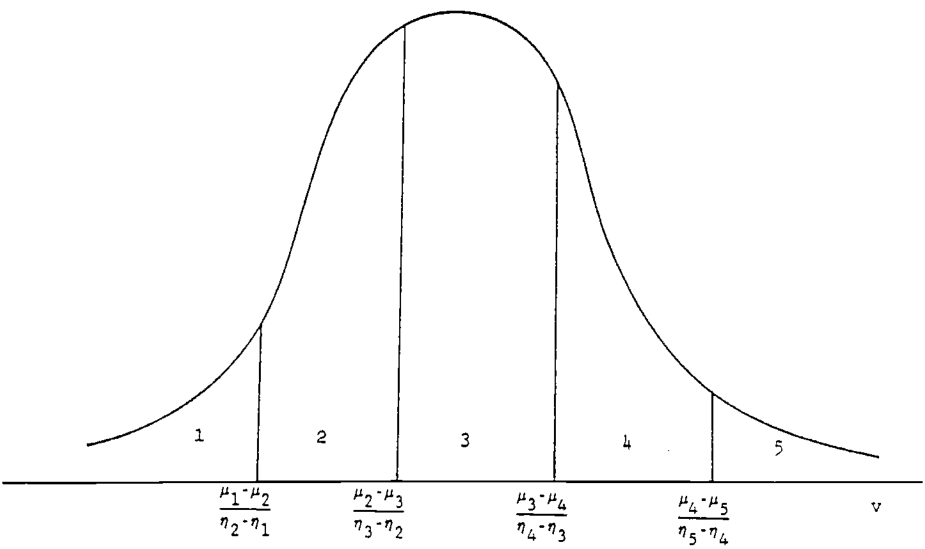

Figure 1. Residential location in a 5-region model, with zero mobility costs.

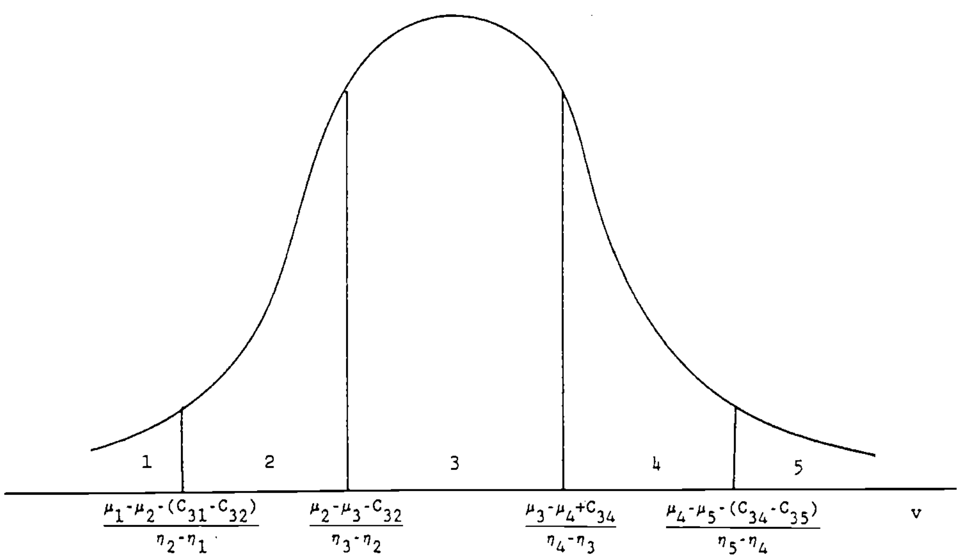

Figure 2. Residential location in a 5-region model for persons born in region 3, with positive mobility costs.

TABLE 1

INTERSTATE MIGRATION RATES
NATIONAL LONGITUDINAL SURVEY OF YOUTH

	Whites		Blacks		Hispanics	
	Males	Females	Males	Females	Males	Females
% Living in Different State Than at Birth	27.70	32.22	25.86	24.89	20.45	26.74
% Living in Different State Than at Age 14	18.79	23.18	15.11	15.63	12.66	12.67
% Living in Different State Than in 1979	18.06	20.60	16.07	13.65	8.65	10.21
% Moving Across States Per Year:						
1979-1980	7.56	7.73	6.95	5.73	3.59	3.78
1980-1981	6.55	7.20	7.79	4.79	4.85	5.67
1981-1982	8.70	6.68	5.88	4.48	4.85	4.16
1982-1983	6.04	7.40	6.00	5.73	3.38	4.54
1983-1984	5.19	5.77	3.36	4.06	2.95	2.84
1984-1985	5.30	4.82	4.20	4.17	3.38	3.21
1985-1986	6.26	9.25	5.16	5.10	3.38	3.59
Sample Size	1772	2097	834	960	474	529

TABLE 2

MEANS OF VARIABLES
(standard deviations in parentheses)

	Overall Sample	Movers	Non-Movers
Age	24.85 (2.23)	25.13 (2.21)	24.79 (2.23)
Percent Female	53.80	57.53	52.96
Percent Black	26.91	22.59	27.88
Percent Hispanic	15.05	10.39	16.09
Percent Movers	18.33	100.00	0.00
Length of Time in Current State (yrs)	5.90 (2.20)	3.27 (2.70)	6.50 (5.59)
Average Real Wage (1986 dollars)	5.63 (3.10)	5.79 (3.24)	5.59 (3.07)
AFQT Score	-0.40 (0.89)	-0.21 (0.93)	-0.44 (0.88)
Years of Education (in 1986)	12.03 (1.98)	12.36 (2.18)	11.95 (1.92)
Unstandardized Wage	-0.10 (0.78)	-0.06 (0.79)	-0.11 (0.78)
Standardized Wage	-0.19 (0.90)	-0.10 (0.89)	-0.21 (0.91)
Sample Size	6666	1222	5444

TABLE 3

CORRELATION OF SKILL MEASURES
(asymptotic standard errors in parentheses)

	AFQT Score	Education	Unstandardized Wage	Standardized Wage
AFQT Score	1.00			
Education	0.584 (.008)	1.00		
Unstandardized Wage	0.355 (.013)	0.307 (.013)	1.00	
Standardized Wage	0.408 (.012)	0.331 (.012)	0.820 (.005)	1.00

PARTIAL CORRELATIONS OF SKILL MEASURES
CONTROLLING FOR AGE, SEX, RACE
(asymptotic standard errors in parentheses)

	AFQT Score	Education	Unstandardized Wage	Standardized Wage
AFQT Score	1.00			
Education	0.586 (.008)	1.00		
Unstandardized Wage	0.318 (.013)	0.289 (.013)	1.00	
Standardized Wage	0.318 (.013)	0.257 (.013)	0.940 (.002)	1.00

NOTE: Partial correlations are obtained by computing simple correlations between the residuals from regressions of each skill measure on age, sex, and race variables.

TABLE 4

OUT-MIGRATION RATES AND AVERAGE SKILLS OF OUT-MIGRANTS RELATIVE TO NATIVES IN THEIR STATE OF ORIGIN, SELECTED STATES

State	N	Rate	AFQT	Education	Unstand. Wage	Stand. Wage
Alabama	246	.191	0.089	0.071	-.011	-.100
California	661	.097	-.013	-.049	-.116	-.079
Colorado	122	.254	0.120	0.233	-.152	-.006
Connecticut	119	.193	0.180	0.841	0.163	0.145
Florida	240	.204	0.151	-.058	0.020	0.085
Georgia	283	.113	0.264	0.386	0.045	0.474
Illinois	195	.221	0.376	0.678	0.160	0.116
Indiana	118	.271	0.298	1.092	0.250	0.237
Massachusetts	113	.124	0.434	1.303	0.180	0.389
Michigan	318	.255	0.161	0.465	-.007	0.098
Minnesota	148	.223	-.111	-.181	-.205	-.307
Missouri	177	.254	0.074	0.021	-.111	-.040
New Jersey	258	.233	0.370	0.521	0.079	-.112
New York	412	.201	0.122	-.030	0.009	-.038
North Carolina	264	.167	0.159	1.000	-.073	0.167
Ohio	411	.180	0.088	0.207	0.081	0.131
Oklahoma	103	.155	-.245	0.919	-.101	-.209
Pennsylvania	309	.133	0.122	0.670	0.071	0.060
South Carolina	171	.082	0.394	1.091	0.306	0.609
Tennessee	140	.136	-.001	-.062	0.034	0.391
Texas	456	.096	0.307	0.196	0.060	0.023
Virginia	160	.138	0.383	0.675	0.102	0.040
Washington	84	.214	0.234	0.040	-.134	-.139
West Virginia	114	.228	-.212	-.162	-.143	-.172
Wisconsin	256	.125	0.219	0.355	0.114	0.025
ALL STATES	6666	.183	0.144	0.282	0.027	0.062

NOTE: N equals the number of respondents residing in a given state at 14 years of age. The out-migration rate equals the fraction of respondents who were living in a different state in 1986 than they were at age 14. The reported means of AFQT, education and the standardized and unstandardized wages in this table are the differences between the mean of each skill variable for out-migrants and the overall mean of the skill variable for all residents of the state at age 14.

TABLE 5

IN-MIGRATION RATES AND AVERAGE SKILLS OF IN-MIGRANTS RELATIVE TO NATIVES IN THEIR STATE OF ORIGIN, SELECTED STATES

State	N	Rate	AFQT	Education	Unstand. Wage	Stand. Wage
Alabama	216	.078	-.014	-.557	0.053	0.289
California	709	.158	0.145	0.385	0.070	0.167
Colorado	135	.326	0.462	0.317	-.038	0.085
Connecticut	112	.143	0.429	1.177	0.467	0.539
Florida	314	.392	0.115	-.165	0.026	0.088
Georgia	309	.188	0.327	0.991	0.316	0.180
Illinois	190	.200	-.095	0.214	0.026	0.087
Indiana	101	.149	0.103	-.502	0.026	0.128
Massachusetts	108	.083	0.094	0.600	0.247	0.035
Michigan	258	.081	0.212	0.818	0.192	0.169
Minnesota	131	.122	-.001	0.466	0.078	0.236
Missouri	162	.185	0.012	0.282	0.130	0.182
New Jersey	230	.139	-.062	0.335	-.111	-.020
New York	386	.148	0.174	0.696	-.230	-.066
North Carolina	258	.147	0.020	-.216	0.007	-.038
Ohio	366	.079	-.114	-.412	-.477	-.284
Oklahoma	108	.194	0.671	1.157	-.185	-.008
Pennsylvania	314	.146	0.157	0.246	-.203	-.089
South Carolina	176	.108	0.418	0.323	0.344	0.272
Tennessee	135	.104	-.048	-1.186	0.511	0.678
Texas	534	.228	0.051	0.201	0.075	0.111
Virginia	182	.242	-.215	-.055	-.024	-.040
Washington	89	.258	0.251	0.343	-.176	-.169
West Virginia	97	.093	0.145	-.073	0.620	0.672
Wisconsin	249	.100	0.302	0.316	-.300	-.008
ALL STATES	6666	.183	0.144	0.282	0.027	0.062

NOTE: N equals the number of respondents residing in a given state in 1986. The in-migration rate equals the fraction of respondents who were living in a different state at age 14 than they were in 1986. The reported means of AFQT, education and the standardized and unstandardized wages in this table are the differences between the mean of each skill variable for in-migrants and the overall mean of the skill variable for all residents of their state of origin (state at age 14).

TABLE 6

THE EFFECT OF SKILLS ON MIGRATION RATES
IN MOVER/NON-MOVER PROBIT MODELS
(asymptotic t-statistics in parentheses)

Skill Measure	Unstandardized wage dispersion in state of residence at age 14: First Quartile	Second Quartile	Third Quartile	Fourth Quartile
AFQT Score	.1172 (2.65)	.1874 (4.80)	.0700 (1.71)	.0976 (1.92)
Effect of One S.D. Change in AFQT on Migration Rate	.0283	.0465	.0166	.0234
Education	.0868 (4.55)	.0636 (3.86)	.0090 (0.55)	.0288 (1.33)
Effect of One S.D. Change in Education on Migration Rate	.0482	.0347	.0047	.0152
Untandardized Wage	.1113 (2.16)	.0355 (0.74)	.0114 (0.23)	.0252 (0.40)
Effect of One S.D. Change in Un. Wage on Migration Rate	.0241	.0075	.0024	.0053
Standardized Wage	.0987 (2.18)	.0876 (2.10)	.0181 (0.41)	.0857 (1.62)
Effect of One S.D. Change in St. Wage on Migration Rate	.0247	.0218	.0044	.0213

NOTE: The probit model also includes race and sex dummy variables.

TABLE 7

THE EFFECT OF SKILLS ON THE CHOICE OF DESTINATION
(t-statistics in parentheses)

Dependent Variable: (Difference in wage dispersion between state of residence in 1986 and at age 14)

Skill Measure	Entire Sample: Change in Unstand. Dispersion	Entire Sample: Change in Stand. Dispersion	Movers Only: Change in Unstand. Dispersion	Movers Only: Change in Stand. Dispersion
AFQT Score/100	.1224	.1651	.4197	.5845
	(4.52)	(4.88)	(3.08)	(3.45)
Education/100	.0479	.0580	.1454	.1720
	(4.41)	(4.27)	(2.81)	(2.68)
Unstandardized Wage/100	.0862	.0860	.3779	.3725
	(2.62)	(2.09)	(2.34)	(1.85)
Standardized Wage/100	.0735	.0732	.2983	.2743
	(2.63)	(2.09)	(2.06)	(1.55)

NOTE: The regressions also include race and sex dummy variables

TABLE 8

THE EFFECT OF SKILLS ON THE CHOICE OF DESTINATION (PROBIT MODELS)
(asymptotic t-statistics in parentheses)

Dependent Variable: Dummy variable indicating whether wage dispersion increased between state of residence at age 14 and state of residence in 1986

Movers Only

Skill Measure	Change in Unstandardized Dispersion	Change in Standardized Dispersion
AFQT Score	.0755 (1.69)	.0995 (2.23)
Education	.0377 (2.21)	.0418 (2.45)
Unstandardized Wage	.2159 (3.98)	.1134 (2.13)
Standardized Wage	.1346 (2.86)	.0846 (1.82)

NOTE: These probit models also include race and sex dummy variables

TABLE 9

CORRELATION BETWEEN SKILLS AND WAGE DISPERSION IN STATE OF ORIGIN
(asymptotic standard errors in parentheses)

	Unstandardized Dispersion	Standardized Dispersion
AFQT Score	-.2072 (.0131)	-.2162 (.0121)
Education	-.0782 (.0122)	-.0884 (.0122)
Unstandardized Wage	-.0328 (.0139)	-.0367 (.0139)
Standardized Wage	-.0835 (.0138)	-.1040 (.0138)

TABLE 10

THE EQUILIBRIUM SORTING OF SKILLS ACROSS STATES
(asymptotic t-statistics in parentheses)

Dependent Variable: Quartile of state in 1986, where states are ordered by their wage dispersion

Skill Measure	Unstandardized Dispersion		Standardized Dispersion	
AFQT score, differenced from mean score in state of origin	.1652 (9.62)	.	.1464 (8.58)	.
AFQT score	.	-.0922 (-5.48)	.	-.1127 (-6.69)
Education, differenced from mean in state of origin	.0294 (4.18)	.	.0179 (2.61)	.
Education	.	-.0209 (-3.09)	.	-.0124 (-1.80)
Unstandardized Wage, differenced from mean in state of origin	.0730 (3.67)		.0638 (3.18)	
Unstandardized Wage	.	.0409 (2.07)		.0254 (1.28)
Standardized Wage, differenced from mean in state of origin	.0523 (3.05)	.	.0314 (1.82)	.
Standardized Wage	.	-.0281 (-1.67)	.	-.0827 (-4.90)

NOTE: These coefficients are obtained from ordered probit models that also include race and sex dummy variables.