Education, Job Search and Migration

Emek Basker*

University of Missouri - Columbia

September 2002

Abstract

Job-search and migration behavior differ across educational groups. In this paper, I explore several differences between the migration and search behavior of workers with different levels of education, both theoretically and empirically. I start with two stylized facts. First, the propensity to migrate increases with education. Second, conditional on migration, the probability that a worker moves with a job in hand (rather than moving to search for a job in the new location) also increases with education. I present a simple individual optimization problem that captures these facts and generates a number of predictions about differential sensitivity of migration to observed variables by education. These predictions, including a non-monotonicity of migration elasticities with respect to business-cycle conditions by educational group, and less-educated groups' higher sensitivity to local economic conditions in the migration decision, are verified using CPS data.

*Comments welcome to emek@missouri.edu. I thank Daron Acemoglu and Olivier Blanchard for helpful advice and guidance. I also benefited from conversations with Saku Aura, Hoyt Bleakley, Al Nucci, and participants in the MIT labor lunch. Pegah Ebrahimi and Amy Mok helped with typing.

"If you don't already have a job in your new location, your number one priority probably will be to find one. Looking for a job long distance can be more fun than looking in your existing location.... No matter how thrilling the prospects, however, you have to be certain you can afford a long-distance search."
– *Steiner's Complete How-to-Move Handbook*

1 Introduction

Between two and three percent of Americans move their residence across state lines every year, many of them for work-related reasons. Some move to search for work in a new location, though most move to take jobs they have already secured. Not much is known about this process of job-search and migration: Why do some people move first, before they have found a job? How do employment outcomes vary by the type of move? How sensitive are movers to local and overall economic conditions? The purpose of this paper is, therefore, to explore the interaction between job search and migration, both theoretically and empirically. I focus on the ways in which education changes workers' incentives, and therefore the migration process.[1]

This much is known: The propensity to migrate decreases with age and increases with education (see Greenwood 1975 and 1993). There is also evidence that the unemployed are more likely to migrate than the employed (Schlottmann and Herzog 1981), and that the unemployed are more sensitive than the employed to the overall unemployment rate in their migration decision (DaVanzo 1978; Bartel 1979). Since the incidence of unemployment is higher among less-educated workers, this effect may mitigate the direct positive effect of education on migration. Table 1 shows some summary statistics on the differential rates of migration across education categories, computed from the March Current Population Survey (CPS) from 1981-2000.[2] While the overall rates of migration are small, the differences between groups are striking: college educated workers are

[1] I use the terms "migrant" and "mover" interchangeably throughout the paper. I also use the terms "high-skill", "high-education" and "high-wage" interchangeably.
[2] See Section 3 for a description of the data.

82% more likely to migrate in any given year than are high-school dropouts.[3]

In recent years the CPS questionnaire has solicited information about movers' main reason for their move. Approximately half of all migrants over the period 1997-2000 moved for job-related reasons. Of these, 90% moved to take a new job or for a job transfer, and the remaining 10% moved to search for work.[4,5] This fraction varies substantially by education level, as shown in Table 2. The probability that a migrant is moving to take a job increases monotonically with her education level; the probability that she moves to look for work or for non-job related reasons decreases monotonically with her education. Strikingly, of the high-school dropouts who moved either to take a job or to look for a job, nearly a third moved to search for a job. Fewer than 3% of college graduates who moved for one of these two reasons (to take a new job or search for a job) moved for the purpose of searching.[6,7]

The focus of this paper is on the interaction of the migration decision with job-search, and the ways in which this interaction depends on a worker's level of education. I start with a simple consumer-choice model in which workers have the choice of searching for a job locally, searching for a job globally (in multiple regions simultaneously), or moving to another region and searching for a job there. I derive conditions under which each strategy dominates the others; these depend on the worker's expected wage (a proxy for her education

[3] Mauro and Spilimbergo (1999) redo the classic analysis of Blanchard and Katz (1992) using Spanish data, but are able to obtain separate estimates by education level of the population. They find that, following an adverse regional employment shock, adjustment for highly-educated workers occurs quickly via out-migration, whereas adjustment for low-education workers is much slower and involves high unemployment and low participation rates for a prolonged period. The implication of these findings is that highly-educated workers migrate away in response to a negative idiosyncratic regional shock, whereas low-education workers do not.

[4] Unfortunately, the *ex-ante* labor-force status (employed, unemployed, not in labor force) of these workers is not known.

[5] Since this question was only asked in the last four years of a long economic expansion, results may not generalize. For example, a larger fraction of moves may be for the purpose of looking for a job in leaner years.

[6] The order of job search and migration has implications for employment outcomes: workers who move to take a job they have already found are up to 13% more likely to be employed the following March than workers who move first and search later.

[7] The educational differences in migration motives are larger than differences in migration motives along other demographic dimensions, such as race, sex and age.

Table 1: Education and Migration Statistics

	Fraction of Population	Propensity to Migrate	Fraction of Migrants[a]
All	100%	2.69%	100%
HS Dropouts	12.52%	2.03%	9.45%
HS Graduates	35.06%	2.16%	28.23%
Some College	25.18%	2.70%	25.27%
College Grads +	27.24%	3.66%	37.06%

[a] May not add to 100% due to rounding

Source: Author's calculations from CPS, 1981-2000

Table 2: Main Reason for Migration by Education

Main Reason for Move[a]	All Movers[b]	HS Dropouts	HS Grads	Some College	College Grads +
		Panel A: Full Sample			
New job / job transfer	46.43%	26.76%	32.30%	39.29%	60.46%
Looking for work / lost job	5.47%	12.81%	9.28%	6.79%	1.78%
Other job-related reason[c]	8.13%	7.71%	7.99%	8.84%	7.84%
Non-job related reason[d]	39.97%	52.72%	50.44%	45.09%	29.92%
		Panel B: Men Only			
New job / job transfer	49.70%	31.58%	33.29%	38.92%	66.02%
Looking for work / lost job	6.26%	15.48%	11.21%	8.79%	1.21%
Other job-related reason[c]	8.81%	7.73%	9.41%	9.81%	8.15%
Non-job related reason[d]	35.22%	45.21%	46.10%	42.49%	24.62%

[a] May not add to 100% due to rounding

[b] Includes only movers whose moving status and reason for moving are not imputed

[c] Includes retirement, easier commute, and miscellaneous job-related reasons

[d] Includes family reasons (e.g., move for spouse), health reasons, etc.

Source: Author's calculations from CPS, 1997-2000 or skill) as well as on economic conditions, both aggregate and local. I find that, as expected, high-skilled workers are more likely to search globally (and therefore to migrate for job-related reasons) than are low-skilled workers; high-skilled workers are also least likely to engage in labor-market arbitrage in the sense of moving from high-unemployment states to low-unemployment states. I also find that, while migration is expected to be pro-cyclical, the cyclicality of migration should be greatest for workers with intermediate skill (education) level.

I next turn to testing these hypotheses using US data. Using pooled cross-section individual-level data, spanning two decades, from the CPS, I estimate individual migration equations. I find that, as predicted, migration is pro-cyclical in the aggregate and counter-cyclical with respect to state-level conditions: in other words, workers tend to move out of high-unemployment states. When I allow the effect of economic variables to differ across education categories, I find that high-school graduates are more sensitive to aggregate business-cycle conditions than are workers with both higher and lower education levels. Workers with low education levels are, however, the most sensitive to arbitrage opportunities in unemployment rates across states.

Existing models of job search assume either that migration must precede search or that search must precede migration. McCall and McCall (1987) assume that migration (between cities) must precede job search. Coulson, Laing, and Wang (2001) model search in a single metropolitan area, and allow agents to search in either the central business district or the suburbs, but not both simultaneously (though they may search – and work – in either market with a commuting cost); they argue that global search will never occur. Spilimbergo and Ubeda (2002a) develop a model of migration in which they focus on double matching in the labor market and social setting. They assume that unemployed workers can find a job with certainly upon migration (and with probability less than 1 locally), so there is no job search following a move. In contrast to these papers, I derive the conditions under which an unemployed worker will search locally and globally, and the corresponding conditions under which a worker will move to search for a job or to take a job.

The remainder of the paper is organized as follows. Section 2 presents the model. Section 3 describes the data used in the analysis, and Section 4 discusses the empirical strategy. Results are presented in Section 5. Section 6 concludes.

2 Model

2.1 Setup

Consider a worker, who, if employed, produces output y and earns a wage w. If unemployed, the worker earns zero.

There is one period. At the beginning of the period, the worker is unemployed and searches for a job. The worker is risk-neutral and seeks to maximize expected income, less search costs. She has several alternative technologies for job search. Local search in the worker's region of residence (Region 1) is costless, but yields a job with lower probability than global search. Global search, in all regions simultaneously, yields a job with higher probability, but at a cost (c); if the job is found in the other region (Region 2), moving costs need also to be incurred. Finally, the worker may choose to move preemptively to Region 2, incurring the moving cost with certainty, but searching only locally once she arrives. A worker who moves incurs a cost of moving m, whether to search for a job or to take a job found in a global search.

Let h be the probability that the worker finds a job anywhere; h proxies for global business-cycle conditions. Let ph be the probability that a worker finds a job in Region 1, and let the probability of finding a job in Region 2 be $(1-p)\,h$. For a worker in Region 1, searching locally involves a cost of zero and a probability of getting a job of ph. Searching globally involves a cost of c and a probability of getting a job of h, with an additional cost of moving m with probability $(1-p)\,h$. Moving to Region 2 and searching there involves a moving cost m with certainty, and a probability of finding a job $(1-p)\,h$. The worker's expected utility, conditional on each of these three actions, are:

$$\begin{aligned} U^L &= phw \qquad (1) \\ U^G &= hw - c - (1-p)\,hm \qquad (2) \\ U^M &= (1-p)\,hw - m. \qquad (3) \end{aligned}$$

where U^L is her expected utility from conducting a local search, U^G is her expected utility from conducting a global search, and U^M is her expected utility from moving to Region 2 and conducting a local search there.

For simplicity, we start by focusing on the case where $h = 1$ (i.e., global search yields a job with certainty, and local search yields a job with probability p). In that case

$$\begin{aligned} U^L &= pw \qquad (4) \\ U^G &= w - c - (1-p)\,m \qquad (5) \\ U^M &= (1-p)\,w - m. \qquad (6) \end{aligned}$$

Given these expected utility functions, workers in Region 1 will choose

$$A = \begin{cases} G & \text{if} \quad w > \overline{w} = \max\left\{\frac{1}{1-p}c + m, \frac{1}{p}c - m\right\} \\ L & \text{if} \quad w < \underline{w} = \min\left\{\frac{1}{1-p}c + m, \max\left\{\frac{1}{1-2p}m, 0\right\}\right\} \\ M & \text{if} \quad w \in [\underline{w}, \overline{w}] \end{cases} . \qquad (7)$$

The probability that a worker moves from Region 1 is therefore

$$P\,(\text{move}) = \begin{cases} 0 & \text{if} \quad w < \underline{w} \\ 1 & \text{if} \quad w \in [\underline{w}, \overline{w}] \\ 1-p & \text{if} \quad w > \overline{w} \end{cases} . \qquad (8)$$

Figure 1 shows the decision space for the worker in (p, w) space if $m = c = 2$ and $h = 1$. For low w, searching locally dominates for sufficiently low p.

Searching globally dominates for sufficiently high wages regardless of p. For an intermediate set of wages and sufficiently low probability of finding a job locally, moving to Region 2 to search there dominates.

Extrapolating to a population of workers facing similar problems (perhaps with idiosyncratic values of w, corresponding to ability, and m, corresponding to local attachment or the "psychic cost" of moving), we can make the following predictions. When local conditions are very bad – p is very low – all but the lowest-skilled workers migrate to search for work elsewhere. As p increases, and local conditions improve, high-wage workers turn to global search, which decreases the probability that they will migrate, and low-wage workers turn to local search. At very high levels of p, when local conditions are very favorable, high-wage workers too eventually switch to local search. As p increases, therefore, the probability of migration decreases for two reasons. First, for high wage workers, the nature of the search changes discretely: from certain or possible migration (if M or G dominate) to certain non-migration. Second, in the region where G dominates, the probability of migration decreases as p increases since the probability that a global search will result in a job outside the region declines with p.

Note that the above results do not depend on the one-period specification, but would carry through (with obvious modifications) to a dynamic setting in which jobs may be lost and the decision to move may be revisited. The key assumption driving the results is that the costs of search and moving (c and m, respectively) are fixed, whereas the wage increases with ability. A high-skilled worker has a high opportunity cost associated with unemployment, and is therefore willing to spend more resources – in the form of c and m – to increase the probability of finding a job. In contrast, a worker whose wage, and therefore opportunity cost of unemployment, is low, will not spend as many resources on job search.

2.2 Comparative Statics

As m decreases (i.e., the workers becomes less migration-averse), all thresholds shift down, so that both M and G dominate on larger regions. Figure 2 shows decision space for the case of $m = h = 1$ and $c = 2$; as $m \to 0$ local search exists only when $p \geq \frac{1}{2}$.

As the cost of global search, c, decreases, global search becomes relatively more attractive. Figure 3 shows decision for $m = 2$ and $c = h = 1$; at the limit as $c \to 0$, global search always dominates moving to search, so the relevant choice margin is between local search and global search.

Finally, consider the case where the probability of finding a job anywhere is $h < 1$. For $m = c = 2$ and $h = \frac{1}{2}$, the regions are then as shown in Figure 4. As $h \to 0$, L dominates in an ever-increasing region; at the limit, all workers search locally since there is no expected return to a global search or move.

2.3 Multiple Regions

Though they were derived in a 2-region setting, these results hold for the case in which there are $n > 2$ regions as well. Continue to assume that the worker's home (or origin) region is Region 1, where the probability of finding a local job is $p_1 h$. WLOG let $p_2 \geq p_3 \geq \cdots \geq p_n$ (where $\sum_{i=1}^{n} = 1$) and consider the case of $h = 1$ as a baseline. By construction, the only region to which the worker would move in order to search is Region 2 (since it is a more attractive destination *ex ante* than any other region).[8,9] The decision problem of the worker is therefore:

$$U^L = p_1 w \quad (9)$$

$$U^G = w - c - (1 - p_1)\, m \quad (10)$$

$$U^{M2} = p_2 w - m \quad (11)$$

[8]Note that we have not restricted p_1 relative to p_2, though if $p_1 > p_2$ the worker will never move in order to search, regardless of m.

[9]The model can be amended to accomodate simultaneous migration to multiple regions by adding heterogeneous individual preferences for different regions, or by building a spatial regional structure and allowing the cost of moving to depend on distance.

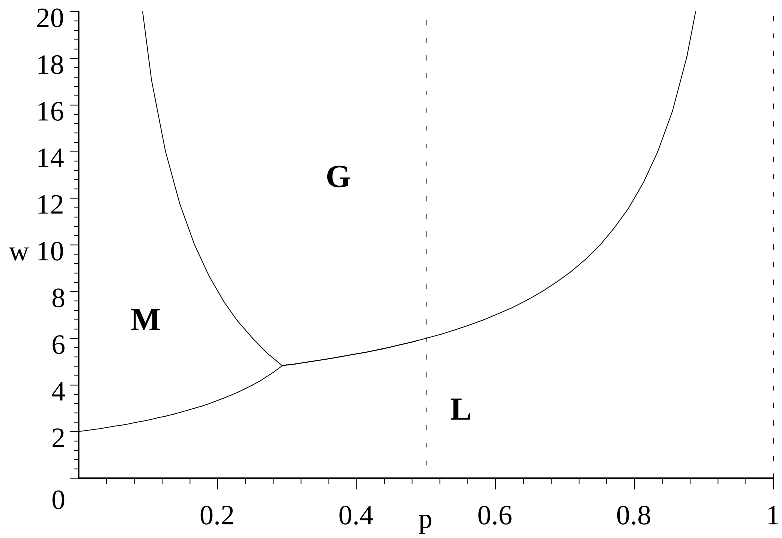

Figure 1: Decision Space for $m = c = 2, h = 1$

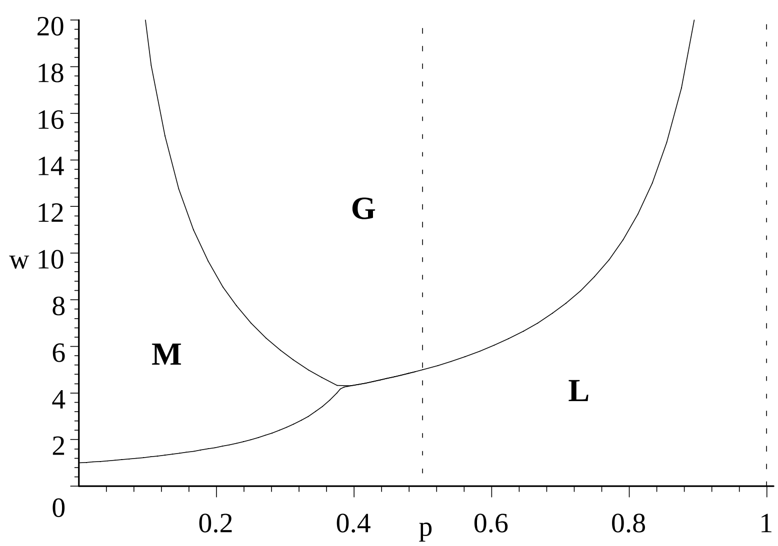

Figure 2: Decision Space for $m = 1, c = 2, h = 1$

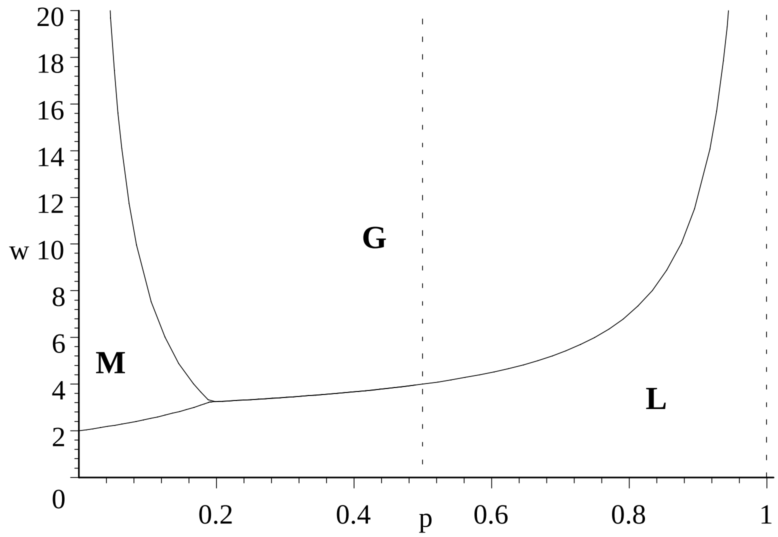

Figure 3: Decision Space for $m = 2, c = 1, h = 1$

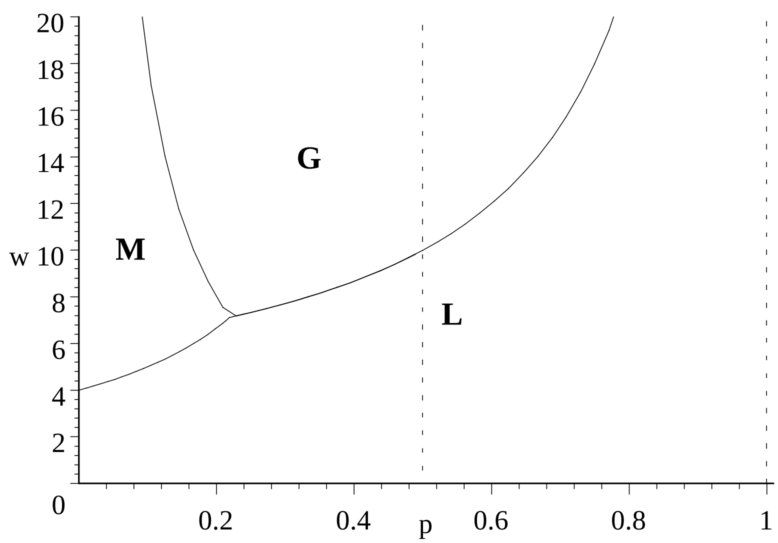

Figure 4: Decision Space for $m = c = 2, h = \frac{1}{2}$

where U^{M2} is her utility of moving to Region 2 and searching locally from there.

Area "A" in Figure 5 shows the domain (p_1, p_2) for the case where $n = 3$: given $p_1 \in [0, 1]$, we restrict $p_2 \in \left[\frac{1-p_1}{2}, 1 - p_1\right]$.[10] As the number of regions increases, the domain of p_2 increases to $\left[\frac{1-p_1}{n-1}, 1 - p_1\right]$; at the limit, the domain includes area "B" as well as area "A". Figure 6 shows the decision space on the domain ("A"+"B") in (p_1, p_2, w) space. It is easy to see that the comparative static results derived above for the 2 region case remain intact in the more general case, with slight notational modifications.

2.4 Implications

The model yields the following testable empirical implications:

1. High-skilled workers are more likely to migrate than are low-skilled workers. Although the propensity to migrate is, in general, not monotonic in skill – at low p workers with intermediate wages may migrate with higher probability than high-wage workers (because of the discreteness of the M vs G choice) – it is monotonic when evaluated at the average region's conditions (i.e., at $p = \frac{1}{2}$).

2. The propensity to migrate may be non-monotonic in skill for sufficiently depressed regions, where intermediate-skilled workers will migrate at higher rates than high-skilled workers. This situation corresponds to the case where p very close to zero in the model.

3. Migration will tend to arbitrage unemployment-rate differentials:

   (a) Workers in states with bad economic conditions will be more likely to move than those living in states with good economic conditions.

   (b) Destinations will have better economic conditions than origin states.

[10] The upper bound on p_2 is to ensure that the 3 probabilities do not exceed 1. The lower bound is necessary and sufficient to ensure that the assumption $p_2 \geq p_3$ holds.

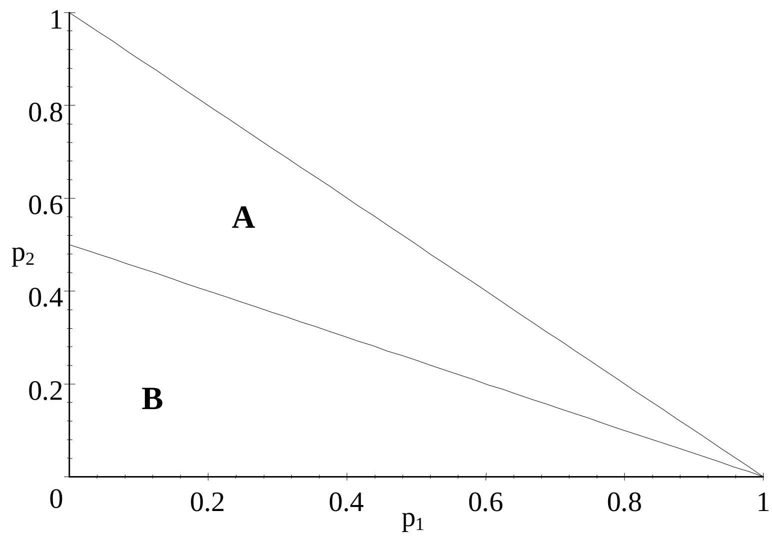

Figure 5: Decision Space Domain for $n \geq 3$

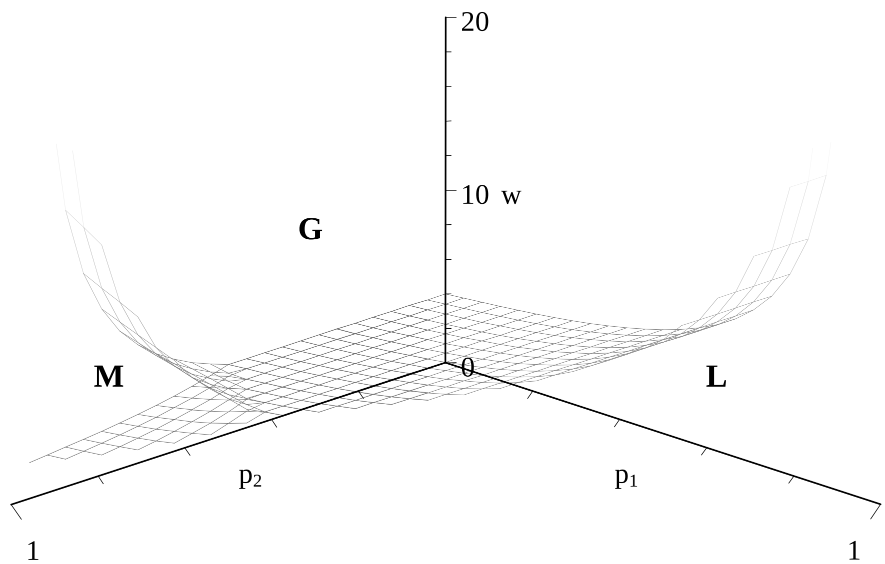

Figure 6: Decision Space for n-Region Case given $m = c = 2, h = 1$

4. Low-skilled workers' destinations will, on average, represent a larger improvement over their origins than is the case for high-skilled workers. This is because destination economic conditions figure directly in the selection of a location for low-skilled workers who move first and search later ("search migrants"), but only indirectly – by affecting the probability that a job is found – in the location choice of workers who search first and move later ("job migrants").

5. Migration is pro-cyclical: as h falls, migration decreases.

6. The effect of fluctuations in aggregate economic conditions will not be uniform across skill groups. Low-skilled workers, never very likely to move, and high-skilled workers, who search globally for a wide range of local conditions, will change their behavior only slightly. The marginal workers will be intermediate-skilled; they are likely to be most sensitive to such business-cycle fluctuations. In other words, we expect to find a non-monotonic relationship between skill and elasticity of migration with respect to business-cycle conditions.

7. For given state-level economic conditions (p and h), migrants who move before finding a job should be less skilled than those who move after finding a job. This implication, of course, is part of the motivation for the model (see Table 2), but needs to be confirmed in a regression with controls for other observable characteristics.

8. Among migrants, even after controlling for skill, the probability of being employed is higher for those who searched globally and moved only after they found a job than for "searching migrants" who move first and search for work later. This is a direct consequence of the fact that global searchers move only if they find a job in the destination state, whereas searching migrants move in order to search.

These hypotheses are tested in Section 5 below.

3 Data

I use March Current Population Survey (CPS) data from 1982-2001 (excluding 1985 and 1995). For many variables, including migration, I attribute the variable values to the previous calendar year: for example, the 1982 survey supplies 1981 data. I therefore distinguish between the *survey year* (the calendar year in which the survey was administered) and the *reference year* (the year preceding the survey year). Respondents have been asked whether they moved in the last year (and where from), almost every year since 1982; exceptions are the 1985 and 1995 surveys. Since 1997 they have also been asked for the main reason for their move.

The Census Bureau selects residential addresses (dwelling units), not their occupants, for inclusion in the CPS; each dwelling unit is included in the March CPS twice, one year apart. By construction, then, non-movers are interviewed twice, whereas movers are interviewed only once – the address is visited twice but two different households respond to the survey. While the CPS questionnaire is quite thorough, most variables – labor force status (employed, unemployed, or not in the labor force), marital status, student status (full- or part-time student), and homeownership status – are available only on a current basis (survey year) but on a lagged basis (reference year).[11]

To eliminate as many non-labor-market-related moves as possible, I limited my sample to civilian adults ages 25-60 in the reference year (thereby eliminating as much as possible retirement-related moves) who were not students during the survey year.[12] This restriction provides me with approximately 60,000 observations per year.[13]

As with many surveys, data accuracy is a concern. Questions that are not answered during a survey are replaced by imputed ("allocated") values, which

[11] For further information about CPS design and methodology, see Current Population Survey (2000).

[12] In 1981-1983, I could not eliminate students due to incomplete data.

[13] Because the interpretation of interstate migration is ambiguous for individuals living in Washington D.C., I omit both current D.C. residents and individuals who moved out of D.C. in the past year. None of the results reported here are sensitive to this omission.

are generated from other ("donor") records. Allocations can be common for some variables and can have a large effect on mean values of some variables, notably migration status. Unfortunately, records with altered or imputed data were not flagged by the Census Bureau until the 1996 survey (referring to 1995 migration data). Since the time series for which allocated values are properly flagged is very short (and migration is a rare event), the analysis presented here cannot be repeated using only allocated values. More details on CPS allocation procedures are presented in Appendix A.1.[14]

4 Empirical Methodology

4.1 Migration Regressions

Because migration is a relatively rare event, even a large sample such as the CPS contains only a small number of migrant observations in any given year. I therefore use pooled cross-section data to estimate the individual-level migration equation

$$\mathbb{P}\left(\text{migrate}_{it}\right) = \Phi\left(\alpha + \boldsymbol{\beta} x_{it} + \sum_{s} \sigma_s \text{state}_{ist} + \sum_{t} \delta_t \text{year}_t\right) + \text{u}_{it} \qquad (12)$$

where $\mathbb{P}\left(\text{A}\right)$ is the probability that event A occurs, migrate_{it} is an indicator equal to 1 if individual i moved between years t and $t+1$ (and 0 otherwise), $\Phi\left(\cdot\right)$ is the standard normal CDF, state_{ist} is an indicator equal to 1 if individual i lived in state s in year t, year_t is a year indicator, and x_{it} is a vector of additional explanatory variables. For regressions with only individual-level demographic variables, the error term u_{it} are clustered at the household level, allowing correlation between the migration decisions of spouses, as well as a across the two interviews of each dwelling unit. For regressions where the coefficient of interest varies only by state and year, u_{it} is clustered at the state

[14]To complicate matters further, allocation procedures changed in 1988. More details on this change are in Appendix A.2.

level.[15,16]

When the vector x_{it} includes economic variables, such as the unemployment rate in individual i's state of residence in year t, the coefficients β are of direct interest in answering the first question posed. The sign, magnitude, and statistical significance of these coefficients gives us a measure of the importance of the explanatory variables in the migration decision.

Year fixed effects are included in some, but not all, of the regressions. Year-to-year fluctuations in the migration rate due to unobservable changes will be captured by these year fixed effects, when included. Year fixed effects are omitted when aggregate economic conditions are included in the regression. I assume, critically, that the effect of specific characteristics, such as education, on the propensity to migrate is not time-varying.

Unless otherwise noted, I report the derivatives $\frac{\partial\Phi(x,z)}{\partial x}|_{(\overline{x},\overline{z})}$ (where z is the vector of all explanatory factors excluding x) rather than probit coefficients. The numbers reported may therefore be interpreted as the effect of an infinitesimal change in the variable of interest, x, on the probability of migration where the independent variables are evaluated at sample means. In cases where x is a binary variable (such as an indicator for race, sex, or education), I report instead the change in the probability of migration associated with a discrete change in x:

$$\Phi\left(x=1;\overline{z}\right)-\Phi\left(x=0;\overline{z}\right).$$

Most regressions use all available observations, with the exception of a few focusing only on movers. Because women are more likely than men to move for reasons other than work (specifically, for a spouse or other family member), I have also estimated results using a male-only sample. The results tend to

[15]For completeness, I also estimate the equivalent OLS regression
$\text{migrate}_{it}=\alpha+\boldsymbol{\beta}x_{it}+\sum\limits_{s}\sigma_s\text{state}_{st}+\sum\limits_{t}\delta_t\text{year}_t+\text{u}_{it}$.
Results from these regressions are uniformly extremely similar to the probit results, and are therefore not reported.

[16]The individual characteristics for which I control are sex, race, age and education. Though the workers' *ex ante* employment status (employed or unemployed) is not elicited in CPS questionnaires, I attempt to control for it indirectly, as explained in Section 4.3 below.

be qualitatively similar though most effects are magnified because of men's higher sensitivity to labor-market conditions.[17] I report the results for men only when they are sufficiently different from the results for the full sample to be of independent interest.

4.2 Unemployment Rates

I compute state-level unemployment rates, as well as state unemployment rates by education category (high-school dropouts, high-school graduates, some college, and college degree or beyond), using records of male non-migrants ages 25-60.[18]

The arbitrage motive in migration, by the model, depends on the difference between the unemployment rate in a worker's current region of residence and the unemployment rate in other regions to which the worker could potentially move. I compute the "target unemployment rate" for a worker in state s as

$$\text{targetue}_{st} = \sum_{m \neq s} \omega_{sm} \text{unemp}_{mt}$$

where unemp_{mt} is the unemployment rate in state m in year t, and the weights ω_{sm} represent the share of movers from state s who moved to state m over the period 1981-2000:

$$\omega_{sm} = \frac{\text{migrants}_{sm}}{\sum_{k \neq s} \text{migrants}_{sk}}. \tag{13}$$

The target unemployment rate is therefore the unemployment rate in the "typical" state of destination for state-s out-migrants.

The relevant variable is then the difference between the worker's current state's unemployment rate and the unemployment rate in the target area, eval-

[17]Of course, since the sample is about half the size, some significance is lost despite the (absolutely) larger estimated coefficients

[18]The unemployment rate by state and education is measured with more error than the average unemployment rate in the state due to the relatively small number of observations in every state*year*education cell. 258 of 3600 state*year*education cells – more than half of them for the highest education category – have no observations of unemployed workers. These cells are assigned an unemployment rate of zero.

uated before migration:

$$\text{targetdiff}_{st} = \text{unemp}_{st} - \text{targetue}_{st}. \tag{14}$$

One issue that warrants mention is measurement error. The predications of the model in Section 2 have all to do with the *individual's* employment prospects, not with the average employment opportunity in her state. In the empirical analysis, the state unemployment rate is used to proxy for individual employment prospects, but as such it is measured with error. Attenuation bias in the coefficient on the differential unemployment rates is therefore to be expected, both because the unemployment rate is itself measured with error (especially when broken down by education category, where very few observations inform each calculation), and because it is an imperfect proxy.

4.3 Personal Unemployment Status

As noted in Section 3, the employment status of workers (employed vs. unemployed) is only available *ex post*, that is, for the survey year, and is not known for the reference year. Unfortunately, it is their *ex ante* employment status – at the time of the decision to stay or migrate – that we want to control for in the migration equations estimated in this paper.

Since more-educated workers are less likely to be unemployed, and since the unemployed may be more likely to migrate, controlling for education without controlling for labor-force status risks biasing the results reported here. In the baseline regressions reported in Section 5.1, omitting workers' employment status may bias the measured effect of education on migration downwards.

On the other hand, in regressions in which workers' education levels are interacted with business-cycle indicators – as in Section 5.3 – the omission is likely to bias *upwards* the differential effect of business cycles by education on migration. To see this, note that the less-educated tend to be more adversely affected by recessions than the more-educated. The aggregate unemployment rate variable, intended to capture business-cycle conditions, is therefore correlated more strongly with the probability that a high-school dropout is unemployed than with the probability that a college graduate is unemployed, and a spurious differential relationship between aggregate conditions and migration, by education, may be estimated as a result.

I attempt to address this concern using two proxies for individual unemployment status in the reference year: the number of weeks worked in the previous year, or – alternatively – an indicator for a worker having been employed at least 50 of the last 52 weeks.[19] These proxies should correlated with workers' past employment status (prior to their move, or their decision to stay put), but they are muddied by their equally-strong correlation with the workers' current employment status. More important than the measurement error in these proxies is their possible endogeneity, since the number of weeks worked last year is – in the model presented here as well as in numerous others – endogenous to the migration decision.

When these controls are added to the regressions, the estimated coefficients increase or decrease in the predicted direction. An open – and unanswerable – question is how much the results presented below would change if less-noisy and truly exogenous controls for last year's employment status were used.

5 Results

5.1 Baseline Regression

In this section I test hypotheses (1) and (2) from Section 2.4. Recall, these are:

1. Evaluated at average regional conditions, high-skilled workers are more likely to migrate than are low-skilled workers.

2. The propensity to migrate may be non-monotonic in skill for sufficiently depressed regions.

[19]The second proxy is simply a discretization f the first.

To test these hypotheses, the baseline regression includes only individual-level explanatory variables: age, sex, race, and education, as well as state and (for some regressions) year fixed effects. I report the coefficients on these demographic variables from probit regressions in Table 3. Coefficients on age fixed effects from Column (1) are plotted in Figure 7. Three education categories are reported (high school diploma, some college, college degree or beyond); the omitted education category is less than high school. As expected, migration increases monotonically with education, and decreases with age.

When the number of weeks worked last year is included in the regression (either as a continuous variable or as an indicator for 50 or more weeks worked), the estimated effect of education increases. Since more-educated workers are less likely to be unemployed, and since the unemployed may be more likely to migrate, controlling for education without controlling for labor-force status biases downward the estimated coefficient on education. Controlling for the number of weeks worked last year therefore increases the measured effect of education on migration. Since the number of weeks worked last year is itself endogenous to the migration decision, however, results with these controls should be interpreted with care. The coefficients on weeks of work should be interpreted with caution, if at all.[20]

Next, I test the prediction that propensity to migrate may not be monotonic in education for regions with sufficiently high unemployment rates. Let bad_{st} be an indicator for bad economic conditions in the state relative to the "target area" for potential migrants:

$$\text{bad}_{st} = \begin{cases} 1 & \text{if targetdiff}_{st} > 0 \\ 0 & \text{otherwise,} \end{cases}$$

[20]To see why the number of weeks worked is endogenous, note that (1) moving takes time, and (2) workers who move in order to search for work are expected to be unemployed for a time in their new location. The number of weeks worked is also a noisy proxy for the worker's labor-force status the previous year, which is the (unavailable) control variable of interest. While the endogeneity biases the coefficient upwards (in absolute terms), the measurement error biases it towards zero; whether the true effect of having been employed one year ago is larger or smaller than the one estimated is impossible to say.

Table 3: Probability of Migration: Baseline Regression Estimates

Variable	(1)	(2)	(3)	(4)	(5)	(6)
Male	0.174	0.483	0.499	0.175	0.489	0.505
	(0.027)	(0.027)	(0.026)	(0.027)	(0.027)	(0.026)
White	-0.055	-0.009	-0.041	-0.053	0.004	-0.026
	(0.065)	(0.061)	(0.061)	(0.064)	(0.061)	(0.060)
High School	0.001	0.273	0.293	-0.004	0.263	0.279
Diploma Exactly	(0.067)	(0.067)	(0.066)	(0.067)	(0.067)	(0.066)
Some College,	0.474	0.832	0.862	0.475	0.818	0.841
No Degree	(0.076)	(0.079)	(0.078)	(0.076)	(0.079)	(0.078)
College Degree	1.452	1.964	1.993	1.446	1.941	1.961
or Beyond	(0.086)	(0.092)	(0.091)	(0.086)	(0.091)	(0.090)
Weeks Worked		-0.075			-0.075	
Last Year (Number)		(0.001)			(0.001)	
Worked 50+ Weeks			-3.209			-3.238
Last Year (Indicator)			(0.060)			(0.061)
χ^2 test for equality of	793	1,100	1,145	790	1,085	1,123
education coefficients	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Age Fixed Effects	Y	Y	Y	Y	Y	Y
Origin State FE	Y	Y	Y	Y	Y	Y
Year Fixed Effects	N	N	N	Y	Y	Y

Notes: 944,061 observations used. Standard errors are clustered at the household level. All coefficients are multiplied by 100.

where targetdiff_{st} is as defined in Equation (14). Table 4 presents coefficients on *interactions* of education and bad economic conditions, to test whether the effect of education is monotonic in the good regime (when $\text{bad}_{st} = 0$) but non-monotonic in the bad regime ($\text{bad}_{st} = 1$). Each regression is presented in two columns, the first showing coefficients on education for state-year cells with the bad regime, and the second showing the coefficients for state-year cells with the good regime.

The first two columns show coefficients from a regression with no year fixed effects; the last two columns repeat this exercise with year fixed effects. In each column, I show first the coefficients on the interaction terms, followed by a χ^2 test for equality of the coefficients by education for each regime. In every case, equality of the coefficients on education *within* the regime is clearly rejected.

Below this first χ^2 test statistic I show the χ^2 statistic of interest: testing whether changes in the probability of migration change with education *differently* across the two regimes. The prediction that intermediate-skilled workers will migrate more than both low- and high-skilled workers out of states with bad economic conditions clearly fails: there is no statistical difference between the elasticity of migration with respect to education in the bad regime and that in the good regime.

5.2 Arbitrage

In this section I test hypotheses (3) and (4) from Section 2.4:

3. (a) Workers in states with bad economic conditions will be more likely to move than those living in states with good economic conditions.
   (b) Destinations will have better economic conditions than origin states.
4. Low-skilled workers' destinations will, on average, represent a larger improvement over their origins than is the case for high-skilled workers.

Before turning to regression analysis, I first present a simple tabulation of the average improvement in migrants' state-level unemployment rate. Table 5

Figure 7: Coefficients on Age Indicators ($\beta_{60} \equiv 0$)

Table 4: Baseline Regression Estimates with *bad* interaction

	(1)		(2)	
Variable	bad = 1	bad = 0	bad = 1	bad = 0
High-School Dropout	(dropped)	-0.277 (0.111)	(dropped)	-0.226 (0.122)
High School Diploma Exactly	-0.107 (0.094)	-0.181 (0.099)	-0.071 (0.093)	-0.164 (0.097)
Some College, No Degree	0.308 (0.109)	0.327 (0.114)	0.369 (0.108)	-0.342 (0.112)
College Degree or Beyond	1.358 (0.134)	1.320 (0.137)	1.439 (0.131)	1.316 (0.135)
χ^2 test for equality of education coefficients	364.55 0.0000	445.53 0.0000	398.28 0.0000	411.25 0.0000
χ^2 test for significance of *bad* interactions	6.54 0.1626		5.84 0.2113	
Year Fixed Effects	N		Y	

Notes: 944,061 observations used. Regressions include all controls from Table 3. Standard errors are clustered at the household level. All coefficients are multiplied by 100.

tabulates the fraction of migrants whose destination-state unemployment rate (measured in year t, before the move) is lower than their year-t origin-state unemployment rate, by education. I show results using both average state unemployment rates (Column 1) and education-specific state unemployment rates (Column 2), both constructed from March CPS files using male non-movers ages 25-60.

The table confirms hypotheses (3b) and (4). On average, migrants move to states with lower unemployment rates than their states of origin *for their education category*, though not necessarily to states with lower overall unemployment rates. As the model predicts, the fraction of migrants whose destinations have lower unemployment rates than their origins decreases monotonically with education.

To test hypothesis (3a), I add the variable targetdiff_{st} (the difference between the state unemployment rate and the unemployment rate in the state's "target region") to the regression presented in Table 3. Table 6 shows the estimated effect of targetdiff_{st} on migration. Column (1) shows results using the average target unemployment-rate differential. Columns (2) and (3) control for weeks worked last year (as a continuous variable and as a discrete variable, respectively); Columns (4)-(6) repeat these regressions with year fixed effects. All other controls from Table 3 are included in all regressions.

When no controls for employment status are included, the effect of targetdiff_{st} is positive and significant: the higher is a state's unemployment rate, relative to the region to which its residents are likely to migrate, the higher is the probability that they will move. For the regressions with year fixed effects, this result continues to hold when weeks of work are included in the regression; the results are marginally-significant (significant at the 10% confidence level) when no year fixed effects are included. As above, it is important to interpret the results of regressions which control for weeks of work with care, since the number of weeks worked is endogenous to the migration decision.

5.3 Cyclical Patterns

I now turn to testing the hypotheses regarding the cyclical behavior of migration:

5. Migration is pro-cyclical.

6. The cyclicality of migration is strongest for workers with intermediate skills (education).

Table 7 shows coefficients on the U.S. unemployment rate **when interacted with education variables**. In Column (1), the US unemployment rate is interacted with individuals' education categories. In Columns (2) and (3), controls are added for weeks worked last year (continuously and discretely, respectively). Column (4) shows results similar to Column (1) with the difference that individuals' education categories are interacted with the unemployment rate *by education category*. Columns (5) and (6) repeat this specification again controlling for weeks worked last year. Standard errors are clustered at the education-category level. Note that education enters into these regressions directly as well as through the interaction terms.

The χ^2 statistic shows a test for equality of the effect of business cycles conditions across educational categories. In all cases a χ^2 test rejects equality of the coefficients across education groups. The estimated effect of the business cycle is non-monotonic: the migration rate of high-school graduates is more sensitive to business-cycle conditions that that of high-school dropouts. This nonmonotonicity is strong in some specifications and almost invisible in others.

5.4 Reasons for Migration

In this section I test whether the reasons for migration are sensitive to observable variables. Since questions about the main reason for migration were not asked before 1997, only a small set of observations is available to answer this question, and some variables (specifically, cyclical ones) cannot be used on the right-hand side. Moreover, the model presented in this paper predicts that workers who

Table 5: Unemployment Rate Arbitrage by Migrants' Education

	Fraction Moving to States with Lower...	
	Unemployment Rate	Education*UE Rate
All Movers	48.98%	52.51%
HS Dropouts	53.69%	54.59%
HS Graduates	49.78%	54.20%
Some College	48.65%	52.81%
College Grads +	47.40%	50.49%

Source: Author's calculations from CPS, 1981-2000

Table 6: Arbitrage Regression Estimates

Variable	(1)	(2)	(3)	(4)	(5)	(6)
targetdiff$_{st}$	5.431	3.812	3.572	5.849	4.345	4.156
	(2.410)	(2.297)	(2.205)	(2.010)	(1.905)	(1.839)
Weeks Worked		-0.075			-0.075	
Last Year (Number)		(0.002)			(0.002)	
Worked 50+ Weeks			-3.206			-3.234
Last Year (Indicator)			(0.096)			(0.093)
Year FE	N	N	N	Y	Y	Y

Notes: 944,061 observations used. Regressions include all controls from Table 3. Standard errors are clustered at the state level. All coefficients are multiplied by 100.

Table 7: Cyclicality of Migration by Education Category

Variable	(1)	(2)	(3)	(4)	(5)	(6)
Unemp Variable		US Average			US Average by Education	
High School	-8.889	-14.531	-15.869	-4.370	-7.750	-8.553
Dropout	(0.525)	(0.635)	(0.600)	(0.317)	(0.396)	(0.362)
High School	-10.491	-14.925	-16.104	-9.063	-12.690	-13.588
Graduate	(0.580)	(0.696)	(0.550)	(0.453)	(0.548)	(0.431)
Some College,	-1.769	-6.056	-7.409	-4.328	-9.095	-10.540
No Degree	(0.820)	(0.948)	(0.847)	(0.854)	(0.987)	(0.874)
College Degree	1.315	-1.107	-1.686	5.059	-0.062	-2.062
or Beyond	(0.786)	(0.856)	(0.830)	(1.309)	(1.467)	(1.397)
Weeks Worked		-0.076			-0.076	
Last Year (Number)		(0.003)			(0.003)	
Worked 50+ Weeks			-3.249			-3.247
Last Year (Indicator)			(0.159)			(0.161)
χ^2 test for equality	441,286	26,992	149,981	7,208	17,500	8,788
of interaction terms	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Notes: 944,061 observations used. Regressions include all controls from Table 3. Standard errors are clustered at the education-category level. All coefficients are multiplied by 100.

search globally migrate with probability $(1-p)\,h < 1$, implying that only a fraction of global searchers will be included in the migrant sample.

With these caveats in mind, I first show that search migrants are more likely to arbitrage unemployment rate differences across states than are job migrants. Table 8 shows the average arbitrage in unemployment rates by type of migration for 1997-2000 (the years for which type of migration is solicited). Strikingly, a statistically-significant 64% of search movers move to states with lower unemployment rates, while only 52% of job movers do so (and for them the fraction is not statistically different from half); the numbers are slightly smaller, but the pattern the same, when education-specific unemployment rates are used.

I next turn to testing the model's predictions about differences between "job movers" and "search movers":

7. For given state-level economic conditions (p and h), search migrants will be less skilled than job migrants.

8. Among migrants, even after controlling for skill, the probability of being employed is higher for job migrants than for search migrants.

To verify that the stylized fact presented in Table 2 – that workers with higher education are more likely to be job-movers than search-movers, with the converse being true of less-educated workers, I limit my sample to job movers and search movers (removing all other movers, as well as all non-movers), and run the regression

$$\mathbb{P}\left(\text{jobmove}_{it} \mid \text{migrate}_{it}\right) = \Phi\left(\alpha + \boldsymbol{\beta} x_{it} + \sum_t \delta_t \text{year}_t\right) + \text{u}_{it} \tag{15}$$

where jobmove_{it} is an indicator equal to 1 if the migrant is a "job mover", and x_{it} are demographic variables – sex, race, age, and education – on the remaining subsample. The results are shown in Table 9. No state-of-origin fixed effects are included due to the small sample size. Column (1) shows results using a full set of age fixed effects; because of the small sample, Column (2) uses fixed effects for 6-year age groups. Columns (3) and (4) repeat Columns (1) and (2) with the addition of year fixed effects. The results show that, even after controlling for age, the probability that a worker is a job-mover increases with education.

Finally, we turn to outcomes. How important is the type of migration to employment outcomes? To answer this question, I regress current employment status on type of move, controlling for the same household characteristics:

$$\mathbb{P}\left(\text{employed}_{i,t+1} \mid \text{migrate}_{it}\right) = \Phi\left(\alpha + \boldsymbol{\beta} x_{it} + \kappa \text{jobmove}_{it} + \sum_t \delta_t \text{year}_t\right) + \text{u}_{it} \tag{16}$$

where $\text{employed}_{i,t+1}$ is an indicator which equals 1 if the worker is employed in year $t+1$, and zero otherwise (i.e., for both the unemployed and non-participants). The results are shown in Table 10. Columns (1) and (2) show results for the full sample of movers, without and with year fixed effects, respectively; Columns (3) and (4) repeat the analysis for men only.

These results suggest that, conditional on migration, male workers who moved to take jobs are 13% more likely to be employed the following March than male workers who moved to search for work. Interestingly, controlling for type of move, education does not significantly change the probability of being employed. Unfortunately, these results are probably driven, at least in part, by an unaddressed endogeneity problem: workers who move to search for a job were probably unemployed for a while before they moved, whereas workers who move to take jobs (job movers) are more likely to have engaged in on-the-job search before the move. This implies that being a job-mover is probably correlated with other, unobserved, characteristics that make the worker more employable in any location. It is therefore hard to judge how much of the increased probability of being employed is due to the type of migration; the 13% figure (16% for both sexes) should be taken as an upper bound.

Table 8: Unemployment Rate Arbitrage by Migration Type

	Fraction Moving to States with Lower...	
	Unemployment Rate	Education*UE Rate
All Movers	53.15%	52.28%
Job Movers	52.34%	50.88%
Search Movers	64.26%	58.86%

Source: Author's calculations from CPS, 1997-2000

Table 9: Regression Results for Job-Mover Characteristics

Variable	(1)	(2)	(3)	(4)
Male	-1.144	-1.543	-1.042	-1.488
	(1.103)	(1.174)	(1.112)	(1.185)
White	5.510	5.696	5.668	5.891
	(2.911)	(3.140)	(2.953)	(3.183)
High School	4.052	3.552	4.094	3.580
Diploma Exactly	(1.812)	(2.007)	(1.798)	(2.001)
Some College,	6.602	6.116	6.655	6.179
No Degree	(1.711)	(1.873)	(1.703)	(1.867)
College Degree	23.388	22.460	23.271	22.322
or Beyond	(3.819)	(3.680)	(3.815)	(3.664)
χ^2 test for equality of	108.71	94.47	106.51	94.07
education coefficients	0.0000	0.0000	0.0000	0.0000
Age Fixed Effects	36	6	36	6
Year Fixed Effects	N	N	Y	Y
Observations	1830	1845	1830	1845

Notes: Standard errors are clustered at the household level.
All coefficients are multiplied by 100.

Table 10: Employment Probability for Migrant Sub-Sample

Variable	All Movers (1)	(2)	Men Only (3)	(4)
Job Mover	16.091	16.098	13.229	12.973
	(4.369)	(4.319)	(4.237)	(4.228)
Male	18.280	18.379		
	(1.947)	(1.942)		
White	-1.221	-1.159	-0.372	-0.210
	(2.724)	(2.792)	(2.631)	(2.649)
High School Diploma Exactly	0.652	0.467	0.243	1.647
	(3.703)	(3.752)	(2.940)	(2.915)
Some College, No Degree	1.184	1.105	1.422	1.299
	(3.749)	(3.761)	(2.672)	(2.654)
College Degree or Beyond	3.482	3.306	3.239	3.053
	(3.877)	(3.879)	(3.190)	(3.110)
χ^2 test for equality of education coefficients	2.26	2.16	2.51	2.39
	0.5203	0.5400	0.4734	0.4959
Age Fixed Effects	6	6	6	6
Year Fixed Effects	N	Y	N	Y
Sex Composition	M & F		M only	
Observations	1861	1861	1088	1088

Notes: Standard errors are clustered at the household level.
All coefficients are multiplied by 100.

6 Conclusion

This paper presents a very simple, one-period, consumer-choice model of migration and job search intended to capture two stylized facts: the positive relationship between education and propensity to migrate, and the tendency of more-educated workers to find work before they move, while less-educated workers move first and search for work second. Several predictions about aggregate migration behavior emerge from the model, and are tested empirically, with very good results.

First, I test whether the monotonic relationship between education and migration holds both in below-average and above-average states. I find that the relationship is somewhat more stable in states with worse-than-average economic conditions than in states with better-than-average conditions. The strong prediction of the model, that intermediate-skilled workers will have the highest out-migration rates from worse-than-average states, fails empirically. At the same time, the difference between the out-migration rates of the highly-educated and the low-educated is smaller in these worse-than-average states, as the model predicts.

Second, I show that, *conditional on moving*, low-skilled workers are more sensitive to relative unemployment rates than are high-skilled workers. I also look at the sensitivity of different educational groups to business-cycle conditions, and find, as the model predicts, that intermediate-skilled workers are most sensitive to business-cycle conditions in their migration decision.

Finally, I show that search movers are substantially less likely to be employed following their move than are job movers.

The model presented here is not intended to capture all aspects of the migration decision; as Table 2 (in the Introduction) shows, nearly half of all migrants give reasons other than work for their decision to move. And while the model imposes identical preferences and identical search-and-migration technologies on all workers (allowing them to differ along a single dimension – wages, assumed to increase monotonically with skill), in reality there are many other differences between low- and high-skilled workers. On the preference dimension, workers may care differentially about their career. If skill is acquired, workers who are "career-minded" may choose to acquire skill and, concurrently, be more willing to migrate even when the expected gain is small. In the terminology of the model presented here, this would imply a correlation between the (psychic) cost of moving and skill: $\rho(m, w) < 0$. On the technology dimension, skilled workers may face lower global-search costs, so that c may decrease with skill. Such modifications to the model would increase the migration rate of the high-skilled relative to the low-skilled, and further decrease the sensitivity of high-skilled workers to cyclical patterns.

References

[1] Bartel, Ann P. (1979). "The Migration Decision: What Role Does Job Mobility Play?" *American Economic Review* 69:5.

[2] Blanchard, Olivier J. and Lawrence F. Katz (1992). "Regional Evolutions." *Brookings Papers on Economic Activity*.

[3] Coulson, N. Edward, Derek Laing, and Ping Wang (2001). " Spatial Mismatch in Search Equilibrium." *Journal of Labor Economics* 19:4.

[4] DaVanzo, Julie (1978). "Does Unemployment Effect Migration? Evidence from Micro Data." *Review of Economics and Statistics* 60:4.

[5] Gabriel, Stuart A., Janice Shack-Marquez, and William L. Wascher (1993). "Does Migration Arbitrage Regional Labor Market Differentials?" *Regional Science and Urban Economics* 23:2.

[6] Greenwood, Michael J. (1975). "Research on Internal Migration in the United States: A Survey." *Journal of Economic Literature* 13:2.

[7] Greenwood, Michael J. (1993). "Migration: A Review." *Regional Studies* 27:4.

[8] Mauro, Paolo, and Antonio Spilimbergo (1999). "How Do the Skilled and the Unskilled Respond to Regional Shocks?: The Case of Spain." *IMF Staff Papers* 46:1.

[9] Pissarides, Christopher A., and Jonathan Wadsworth (1989). "Unemployment and the Inter-regional Mobility of Labour." *Economic Journal* 99:397.

[10] Schlottmann, Alan M., and Henry W. Herzog, Jr. "Employment Status and the Decision to Migrate." *Review of Economics and Statistics* 63:4.

[11] Spilimbergo, Antonio, and Luis Ubeda (2002a). "A Model of Multiple Equilibria in Geographic Labor Mobility." IMF Working Paper WP/02/31.

[12] Spilimbergo, Antonio, and Luis Ubeda (2002b). "Family Attachment and the Decision to Move by Race." IMF Working Paper WP/02/83.

[13] Steiner, Clyde and Shari (1999). *Steiner's Complete How-to-Move Handbook*. San Francisco: Independent Information Publications.

[14] U.S. Department of Commerce, Bureau of the Census (1991). *Current Population Survey: Annual Demographic File, 1988 Rewrite Supplement*. Washington, D.C.: U.S. Department of Commerce, Bureau of the Census.

A Data Appendix

A.1 Allocated Values

As mentioned in Section 3, missing data in the CPS are replaced by allocated values, which are generated from other ("similar") records. Unfortunately, records with altered or imputed data were not properly flagged by the Census Bureau before 1995. One variable that is particularly susceptible to allocation is the migration variable. Table 11 lists the number of records with an allocated migration status for each year, their relative weight in the sample, and the propensity of allocated records to be coded as migrations. Beginning with the 1996 survey, over a thousand observations annually are allocated, and the fraction of these observations that are assigned migrant status increases sharply over time, to nearly 60% by the 2001 survey.[21]

A.2 1988 Processing Changes

Following a change in the CPS processing system in 1988, the 1988 survey data were re-released, having been processed using the new system. There are therefore two files containing 1988 data, the first of which was used by the Census Bureau to produce their reports, and the second, known as the "bridge" file (or, alternatively, as the 1988 rewrite file or the 1988B file), intended to facilitate comparisons to subsequent years. Since the input data – the pool of respondents, the survey questions and answers – are identical across the two 1988 files, one should in principle be able to use either one for analysis. Unfortunately, these processing changes were not completely benign. Among the changes made in the re-processing were changes to the imputation procedures for missing data (Bureau of the Census 1991).

While the demographic characteristics of respondents (age, sex, occupation, marital status, and race) are statistically indistinguishable across the two files, as seen in Table 12, migration data are disconcertingly different in these two

[21] Allocation flags were not included in the data prior to the 1988 bridge file.

surveys, by a statistically significant margin. Note that the p-value is given here only as a reference, since the differences between the two data sets are all due to imputation. As these are not actually two separate samples, it is not clear what interpretation, if any, a t-test can be given in this case.

The change in the migration estimation is due to 445 records which were coded as movers in one file but as non-movers in the other. These include 116 that were coded as movers in the original file (but not in the bridge file), and another 329 that were coded as movers in the bridge file (but not in the original file). A frightening 684 additional records are coded as migrants in both files, but their state of origin differs across files. According to the Census Bureau, these changes had the effect of "making migration recodes more consistent with residence fields" (US Census Bureau 1991). In the data analysis I rely exclusively on the bridge file for 1988 data (1988B), in the hope that the processing changes improved the data quality.

Table 11: Allocations and Migration in CPS Data

Survey	Observations	Migration Allocations	Allocation Weight	Allocated Migration[a]
1988B	66,828	0	0	n/a
1989	62,477	8	<0.001	0
1990	68,121	5	<0.001	0
1991	68,341	6	<0.001	0
1992	67,613	0	0	n/a
1993	67,179	0	0	n/a
1994	63,822	5	<0.001	0
1996	55,000	1,125	0.023	0.213
1997	55,666	1,217	0.025	0.228
1998	56,259	1,052	0.020	0.209
1999	56,524	1,160	0.023	0.409
2000	56,718	1,149	0.022	0.561
2001	54,754	1,003	0.021	0.596

[a] Fraction of allocated observations that are assigned migrant status

Table 12: Summary Statistics for 1988 Surveys

Variable (I=Indicator)	Mean 1988[a]	Mean 1988B	p-Value for Equality
Interstate Migration (I)	0.025	0.028	0.002
Age	39.75	39.74	0.789
Male (I)	0.487	0.486	0.845
White (I)	0.858	0.859	0.797
High School Dropout (I)	0.178	0.178	0.922
High School Graduate (I)	0.378	0.378	0.921
Some College, No Degree (I)	0.215	0.215	0.905
College Degree or Beyond (I)	0.229	0.229	0.930
Observations	66,504	66,828	

Notes: Means include past and present DC residents. All means are weighted. Means are reported for non-student civilian adults ages 25-60. Hypothesis tests assume equal variance across surveys.