The Effect of Expected Income on Individual Migration Decisions

John Kennan and James R. Walker[1]
University of Wisconsin-Madison and NBER

July 2005

Abstract

The paper develops a tractable econometric model of optimal migration, focusing on expected income as the main economic influence on migration. The model improves on previous work in two respects: it covers optimal sequences of location decisions (rather than a single once-for-all choice), and it allows for many alternative location choices. The model is estimated using panel data from the NLSY on white males with a high school education. Our main conclusion is that interstate migration decisions are influenced to a substantial extent by income prospects. The results suggest that the link between income and migration decisions is driven both by geographic differences in mean wages and by a tendency to move in search of a better locational match when the income realization in the current location is unfavorable.

---

[1]Department of Economics, University of Wisconsin, 1180 Observatory Drive, Madison, WI 53706; jkennan@ssc.wisc.edu and walker@ssc.wisc.edu. The National Science Foundation and the NICHD provided research support. We thank Taisuke Otsu for outstanding research assistance. We are grateful to Kate Antonovics, Peter Arcidiacono, Gadi Barlevy. Phil Haile, Igal Hendel, Mike Keane, Derek Neal, John Pencavel, Karl Scholz, Marcelo Veracierto, Ken Wolpin, Jim Ziliak, and seminar and conference participants at the Chicago Federal Reserve Bank, Carnegie-Mellon, Duke, Iowa, IZA, Ohio State, Penn State, Rochester, SITE, the Upjohn Institute, Virginia, Wisconsin, and Yale for helpful comments.

1 Introduction

There is an extensive literature on migration.[2] Most of this work describes patterns in the data: for example, younger and more educated people are more likely to move; repeat and especially return migration accounts for a large part of the observed migration flows. Although informal theories explaining these patterns are plentiful, fully specified behavioral models of migration decisions are relatively scarce, and these models generally consider each migration event in isolation, without attempting to explain why most migration decisions are subsequently reversed through onward or return migration.

This paper develops a model of optimal sequences of migration decisions, focusing on expected income as the main economic influence on migration. We emphasize that migration decisions are reversible, and that many alternative locations must be considered. The model is estimated using panel data from the National Longitudinal Survey of Youth on white males with a high school education.

Structural dynamic models of migration over many locations have not been estimated before, presumably because the required computations have not been feasible.[3] A structural representation of the decision process is of interest for the usual reasons: we are ultimately interested in quantifying responses to income shocks or policy interventions not seen in the data, such as local labor demand shocks, or changes in welfare benefits. Our basic empirical question is the extent to which people move for the purpose of improving their income prospects. Work by Keane and Wolpin (1997) and by Neal (1999) indicates that individuals make surprisingly sophisticated calculations regarding schooling and occupational choices. Given the magnitude of geographical wage differentials, and given the findings of Topel (1986) and Blanchard and Katz (1992) regarding the responsiveness of migration flows to local labor market conditions, one might expect to find that income differentials play an important role in migration decisions.

We model individual decisions to migrate as a job search problem. A worker can draw a wage only by visiting a location, thereby incurring a moving cost. Locations are distinguished by known differences

[2]See Greenwood [1997] and Lucas [1997] for surveys.

[3] Holt (1996) estimated a dynamic discrete choice model of migration, but his framework modeled the move/stay decision and not the location-specific flows. Similarly, Tunali (2000) gives a detailed econometric analysis of the move/stay decision using microdata for Turkey, but his model does not distinguish between alternative destinations. Dahl (2002) allows for many alternative destinations (the set of States in the U.S.), but he considers only a single lifetime migration decision. Gallin (2004) models net migration in a given location as a response to expected future wages in that location, but he does not model the individual decision problem.

in wage distributions, amenity values and alternative income sources. A worker starts the life-cycle in some home location and must determine the optimal sequence of moves before settling down.

The decision problem is too complicated to be solved analytically, so we use a discrete approximation that can be solved numerically, following Rust (1994). The model is sparsely parameterized. In addition to expected income, migration decisions are influenced by moving costs (including a fixed cost, a reduced cost of moving to a previous location, and a cost that depends on distance), and by differences in climate, and by differences in location size (measured by the population in origin and destination locations). We also allow for a bias in favor of the home location. Age is included as a state variable, entering via the moving cost, with the idea that if the standard human capital explanation of the relationship between age and migration rates is correct, there should be no need to include a moving cost that increases with age. This idea is soundly rejected.

Our main substantive conclusion is that interstate migration decisions are indeed influenced to a substantial extent by income prospects. Although there is some evidence of a response to geographic differences in wage distributions, the results suggest that the link between income and migration decisions is largely driven by a tendency to move in search of a better locational match when the income realization in the current location is unfavorable.

More generally, the paper demonstrates that a fully specified econometric model of optimal dynamic migration decisions is feasible, and that it is capable of matching the main features of the data, including repeat and return migration. Although this paper focuses on the relationship between income prospects and migration decisions at the start of the life cycle, suitably modified versions of the model can potentially be applied to a range of issues, such as the migration effects of interstate differences in welfare benefits, the effects of joint career concerns on household migration decisions, and the effects on retirement migration of interstate differences in tax laws.[4]

2 Migration Dynamics

The need for a dynamic analysis of migration is illustrated in Table 1, which summarizes interstate migration histories of young people in the NLSY. Two features of the data are noteworthy. First, a large fraction of the flow of migrants involves people who have already moved at least once. Second, a large fraction of these repeat moves involves people returning to their original location. Simple models of isolated move-stay decisions cannot address these features of the data. In particular, a model of return

[4]See for example Kennan and Walker (2001) and Woo (2002).

migration is incomplete unless it includes the decision to leave the initial location as well as the decision to return. Moreover, unless the model allows for many alternative locations, it cannot give a complete analysis of return migration. For example, a repeat move in a two-location model is necessarily a return move, and this misses the point that people frequently decide to return to a location that they had previously decided to leave, even though many alternative locations are available.

Table 1: Interstate Migration Flows
NLSY, 1979-92

	Less than High School	High School	Some College	College
No. of people	1768	3534	1517	1435
Movers	423	771	376	469
Movers (%)	23.9%	21.8%	24.8%	32.7%
Moves Per Mover	2.0	1.8	1.7	1.6
Repeat moves (% of all moves)	50.6	45.9	41.3	35.7
Return Migration (% of all moves)				
Return - Home	24.0	24.1	17.5	13.4
Return - Else	12.4	7.2	5.9	3.3
Movers who return home (%)	48.7	44.5	29.8	20.9
Return-Home: % of Repeat	47.5	52.5	42.4	37.5

3 An Optimal Search Model of Migration

We model migration as an optimal search process. The basic assumption is that wages are local prices of individual skill bundles. We assume that individuals know the wage in their current location, but in order to determine the wage in another location, it is necessary to move there, at some cost. This assumption reflects the idea that the wage summarizes the full value of a job, taking account of working conditions, residential conditions, local amenities and so forth. Although information on some of these things can of course be collected from a distance, we view the whole package as an experience good.

The model aims to describe the migration decisions of young workers in a stationary environment. The wage offer in each location may be interpreted as the best offer available in that location.[5] Although there may be transient fluctuations in wages, the only chance of getting a permanent wage gain is to

[5]This means that we are treating local match effects as relatively unimportant: search within the current location quickly reveals the best available match.

move to a new location. One interpretation is that wage differentials across locations equalize amenity differences, but a stationary equilibrium with heterogeneous worker preferences and skills still requires migration to redistribute workers from where they happen to be born to their equilibrium location. Alternatively, it may be that wage differentials are slow to adjust to location-specific shocks, because gradual adjustment is less costly for workers and employers.[6] In that case, our model can be viewed as an approximation in which workers take current wage levels as a rough estimate of the wages they will face for the foreseeable future. In any case, the model is intended to describe the partial equilibrium response of labor supply to wage differences across locations; from the worker's point of view the source of these differences is immaterial, provided that the differences are permanent. A complete equilibrium analysis would of course be much more difficult, but our model can be viewed as a building-block toward such an analysis.

Suppose there are J locations, and individual i's income y_{ij} in location j is a random variable with a known distribution. Migration decisions are made so as to maximize the expected discounted value of lifetime utility. In general, the level of assets is an important state variable for this problem, but we focus on a special case in which assets do not affect migration decisions: we assume that the marginal utility of income is constant, and that individuals can borrow and lend without restriction at a given interest rate. Then expected utility maximization reduces to maximization of expected lifetime income, net of moving costs, with the understanding that the value of amenities is included in income, and that both amenity values and moving costs are measured in consumption units. This is a natural benchmark model, although of course it imposes strong assumptions.

There is little hope of solving this expected income maximization problem analytically. In particular, the Gittins index solution of the multiarmed bandit problem cannot be applied because there is a cost of moving.[7] But by using a discrete approximation of the wage distribution in each location, we can compute the value function and the optimal decision rule by standard dynamic programming methods, following Rust (1994).

---

[6]Blanchard and Katz (1992, p.2), using average hourly earnings of production workers in manufacturing, by State, from the BLS establishment survey, describe a pattern of "strong but quite gradual convergence of state relative wages over the last 40 years." For example, using a univariate AR(4) model with annual data, they find that the half-life of a unit shock to the relative wage is more than 10 years. Similar findings were reported by Barro and Sala-i-Martin (1991) and by Topel (1986).

[7]See Banks and Sundaram (1994) for an analysis of the Gittins index in the presence of moving costs.

3.1 The Value Function

Let x be the state vector (which includes wage information, current location and age, as discussed below). The utility flow for someone who chooses location j is specified as u(x,j) + ζ_j, where ζ_j is a random variable that is assumed to be iid across locations and across periods and independent of the state vector. Let p(x′| x,j) be the transition probability from state x to state x′, if location j is chosen. The decision problem can be written in recursive form as

$$V(x,z) = \max_j \left[v(x,j) + z_j \right]$$

where

$$v(x,j) = u(x,j) + b \sum_{x'} p(x'|x,j)\bar{v}(x')$$

and

$$\bar{v}(x) = E_z\, V(x,z)$$

and where β is the discount factor, and E_ζ denotes the expectation with respect to the distribution of ζ. We assume that ζ is drawn from the Type I extreme value distribution. In this case, using arguments due to McFadden (1973) and Rust (1987), we have

$$\exp\left(\bar{v}(x)\right) = \sum_{k=1}^{n} \exp\left(v(x,k)\right)$$

Let ρ(j,x) be the probability of choosing location j, when the state is x. Then

$$r(x,j) = \exp\left(v(x,j) - \bar{v}(x)\right) \qquad \textbf{(3)}$$

We compute V by value function iteration. Since we treat age as a state variable, it is convenient to use V ≡ 0 as the initial estimate, so that each iteration yields the value function for a person who is a year younger than the person whose value function was obtained in the previous iteration.

4 Empirical Implementation

A serious limitation of the discrete dynamic programming method is that the number of states is typically large, even if the decision problem is relatively simple. Our model, with J locations and n points of support for the wage distribution, has $J(n+1)^J$ states, for each person, at each age. Ideally, locations would be defined as local labor markets. The smallest geographical unit identified in the NLSY geocode file is a county, but we obviously cannot let J be the number of counties, since there are over 3,100 counties in the U.S. Indeed, even if J is the number of States, the model is numerically infeasible[8], but by restricting the information available to each individual an approximate version of the model can be estimated; this is explained below.

4.1 A Limited History Approximation

To reduce the state space to a reasonable size, it seems natural in our context to use an approximation that takes advantage of the timing of migration decisions. We have assumed that information on the value of human capital in alternative locations is permanent, and so if a location has been visited previously, the wage in that location is known. This means that the number of possible states increases geometrically with the number of locations. In practice, however, the number of people seen in many distinct locations is small. Thus by restricting the information set to include only wages seen in recent locations, it is possible to drastically shrink the state space while retaining most of the information actually seen in the data. Specifically, we suppose that the number of wage observations cannot exceed M, with $M < J$, so that it is not possible to be fully informed about wages at all locations. Then if the wage distribution in each of J locations has n points of support, the number of states is $(Jn)^M$, since this is the number of possible M-period histories describing the locations visited most recently, and the wages found there. For example, if J is 51 and n is 3 and M is 2, the number of states is 23,409, which is manageable.

This approximation reduces the number of states in the most obvious way: we simply delete most of them.[9] Someone who has "too much" wage information in the big state space is reassigned to a less-informed state. Individuals make the same calculations as before when deciding what to do next, and the

[8]And it will remain so: for example, if a location is a State, and the wage distribution has 3 support points, then the number of dynamic programming states is 258,600,722,446,558,797,905,327,453,896,704.

[9]Note that it is not enough to keep track of the best wage found so far: the payoff shocks may favor a location that has previously been discarded, and it is necessary to know the wage at that location in order to decide whether to go back there (even if it is known that there is a higher wage at another location).

econometrician uses the same procedure to recover the parameters governing the individual's decisions. There is just a shorter list of states, so people with different histories may be in different states in the big model, but they are considered to be in the same state in the reduced model. In particular, people who have the same recent history are in the same state, even if their previous histories were different.

4.2 State Variables and Flow Payoffs

Let $\ell = (\ell^0, \ell^1, \ldots \ell^{M-1})$ be an M-vector containing the sequence of recent locations (beginning with the current location), and let ω be the corresponding sequence containing recent wage information. The state vector x consists of ℓ, ω and age.

The flow payoff for someone whose “home” location is h is specified as

$$\tilde{u}_h(x, j) = u_h(x, j) + z_j$$

$$u_h(x, j) = a_0 y(\ell^0, w^0) + \sum_{k=1}^{K} a_k Y_k\left(\ell^0\right) + kc\left(\ell^0 = h\right) - \Delta_t(x, j)$$

Here ω^0 indexes the support points of the wage distribution, and $y(\ell^0,\omega^0)$ is the relevant point in the current location. Wage income is augmented by the nonpecuniary variables $Y_k(\ell^0)$, representing amenity values. The parameter κ is a premium that allows each individual to have a preference for their native location (χ_A is used as an indicator meaning that A is true). The cost of moving from ℓ^0 to j for a person of type τ is represented by $\Delta_\tau(x,j)$. The unexplained part of the utility flow, ζ_j, may be viewed as either a preference shock or a shock to the cost of moving, with no way to distinguish between the two.

Moving Costs

Let $D(\ell^0,j)$ be the distance from the current location to location j, and let $A(\ell^0)$ be the set of locations adjacent to ℓ^0 (where States are adjacent if they share a border). The moving cost is specified as

$$\Delta_t(x, j) = \left(g_{0t} + g_1 D(\ell^0, j) - g_2\, c\left(j \in A\left(\ell^0\right)\right) - g_3\, c(j = \ell^1) + g_4 a - g_5 n_j\right) c(j \neq \ell^0)$$

We allow for unobserved heterogeneity in the cost of moving: there are several types, indexed by τ, with differing values of the intercept γ_0. In particular, there may be a "stayer" type, meaning that there may be people who regard the cost of moving as prohibitive, in all states. The moving cost is an affine function of distance (which we measure as the great circle distance between population centroids). Moves to an adjacent location may be less costly (because it is possible to change States while remaining in the same general area). A move to a previous location may also be less costly, relative to moving to a new location. In addition, the cost of moving is allowed to depend on age, a. Finally, we allow for the possibility that it is cheaper to move to a large location, as measured by population size n_j. It has long been recognized that location size matters in migration models (see e.g. Schultz [1982]). California and Wyoming cannot reasonably be regarded as just two alternative places, to be treated symmetrically as origin and destination locations. To take one example, a person who moves to be close to a friend or relative is more likely to have friends or relatives in California than in Wyoming. One way to model this in our framework is to allow for more than one draw from the distribution of payoff shocks in each location.[10] Alternatively, location size may affect moving costs – for example, friends or relatives might help reduce the cost of the move. In practice, both versions give similar results.

The transition probabilities are as follows

$$p(x'|x,j) = \begin{cases} 1 & \text{if } j = \ell^0, \ell' = \ell, w' = w, a' = a+1 \\ 1 & \text{if } j = \ell^1, \ell' = (\ell^1, \ell^0), w' = (w^1, w^0), a' = a+1 \\ \frac{1}{n} & \text{if } j \notin \{\ell^0, \ell^1\}, \ell' = (j, \ell^0), w' = (s, w^0), s = 1,2,\ldots,n, a' = a+1 \\ 0 & \text{otherwise} \end{cases}$$

This covers several cases. First, if no migration occurs this period, then ℓ and ω are unchanged. Next, following a move to a previous location, the current and previous locations are interchanged. Finally,

[10] Suppose that the number of draws per location is an affine function of the number of people already in that location, and that migration decisions are controlled by the maximal draw for each location. This leads to the following modification of the logit function describing choice probabilities:

$$\rho(x,j) = \frac{\xi_j}{\sum_{k=1}^{J} \xi_k}; \quad \xi_k = (1 + \psi n_k) \exp[v_k(\ell,\omega)]$$

Here n_j is the population in location j, and ψ can be interpreted as the number of additional draws per person.

following a move to a new location, the current location becomes the previous location, and the location match component is drawn randomly from a distribution with n points of support. In all cases, age is incremented by one period.

4.3 Data

Our primary data source is the NLSY79; we also use data from the 1990 Census. In order to obtain a relatively homogeneous sample, we consider only white high-school graduates with no college education, using only the years after schooling is completed.[11] The sample includes only people who had completed high school by age 20, and who did not attend college. We exclude those who ever served in the military. We follow each person from age 20 to the 1992 interview, or to the first missing interview. The final sample includes 665 people, and 5,767 person-years.

4.4 Wages

The wage of individual i in location j at age a is specified as

$$w_{ij}(a) = X_i b + f(a) + m_j + u_{ij} + h_i + e_{ij}(a)$$

where $X_i \beta$ represents the effect of observed individual characteristics, μ_j is the mean wage in location j, υ is a permanent location match effect, $\phi(a)$ describes the age-earnings profile, η is an individual effect that is fixed across locations, and ε is a transient effect. We assume that η, υ and ε are independent normal random variables, and that they are identically distributed across individuals and locations. We also assume that the realizations of η and υ are seen by the individual.[12]

The incentive to migrate, for individual i, is governed by the difference between the quality of the match in the current location, measured by $\mu_j + \upsilon_{ij}$, and the prospect of obtaining a better match in another

[11] Attrition in panel data is an obvious problem for migration studies, and one reason for using NLSY data is that it minimizes this problem. Reagan and Olsen (2000, p. 339) report that “Attrition rates in the NLSY79 are relatively low ...The primary reason for attrition are death and refusal to continue participating in the project, not the inability to locate respondents at home or abroad.” Ham, Li and Reagan (2001), use NLSY data to compare wages following migration with (counterfactual) estimates of what the wage would have been if migration had not occurred, but they do not analyze the migration decision itself.

[12] An interesting extension of the model would allow for learning, by relaxing the assumption that agents know the realizations of η and υ. In particular, such an extension might help explain return migration, because moving reveals information about the wage components. Pessino (1991) analyzed a two-period Bayesian learning model along these lines, and applied it to migration data for Peru.

location k, measured by $\mu_k + \upsilon_{ik}$. The other components of wages have no bearing on migration decisions, since they are added to the wage in the same way no matter what decisions are made. The individual knows the realization of the match quality in the current location, and in the previous location (if there is one), but the prospects in other locations are random. Migration decisions are made by comparing the expected continuation value of staying, given the current match quality, with the expected continuation values associated with moving.

It is straightforward to estimate age effects and mean wages by State; since migrants are of course older following a move than they were before, it is particularly important to adjust wages for age, so as not to attribute to migration the earnings growth due to age. It is more difficult to separate the location match component from the other wage components. One problem is that even if the mean of υ_{ij} across individuals is zero in all locations, the realizations of υ found in measured wages reflect selection effects due to migration decisions. Allowing for selection effects would be difficult, and migration rates are low enough to suggest that the required effort might not be worthwhile. Another problem is that we cannot separate υ and ε using Census data, and there are not enough observations in the NLSY to get reliable estimates of wage distributions for each State.

Using 1990 Census data, we estimate State mean wages (μ_j) and the age-earnings profile (ϕ) by regressing annual earnings on a full set of State and age dummy variables.[13] This yields a set of predicted wages, for each location and age. Using the NLSY data, we regress deviations from this prediction on AFQT scores.[14] Taking the residuals from this regression (and ignoring sampling error), we have a set of individual wage histories, with observable effects removed:

$$\begin{aligned} y_{it} &= w_{it} - X_i b - f(a_{it}) - m_{j(t)} \\ &= h_i + u_{ij_{it}} + e_t \end{aligned}$$

[13]We use the 1990 Public Use Files at Minnesota and select white males using race and ethnicity definitions consistent with the coding in the NLSY. We restrict the sample to those 20-34 at the survey date who are not enrolled in school and have a GED or high school degree. Also excluded are immigrants entering the country after 1978 as these individuals are not in the sample frame of the NLSY. We define income as the sum of wage and salary, farm and business income. We drop observations with zero income but with positive hours or weeks worked reported for 1989. We remove cost of living differences using the ACCRA index, see ACCRA (2000).

[14]We include the AFQT score as a regressor in the NLSY wage equation, since it is known that it helps to explain earnings differences. We also include linear age effects in the NLSY regression just in case there is a difference between Census and NLSY age profiles (there isn't).

where a_{it} is i's age at date t, and j_{it} is the location at date t.

The next step is to extract estimates of the match components from these wage histories. We first estimate the variances of the three components, and then compute the distribution of the location match effects, conditional on the observed wage histories.

Estimation of the Variances

For each individual history (y_i), we classify the elements of the cross-products matrix $y_i y_i'$, as follows: (1) diagonal elements, (2) off-diagonal elements that refer to covariances in the same location and (3) off-diagonal elements that refer to covariances in different locations. Let A_1, A_2 and A_3 denote the sample averages of these cross-products (where the average is taken over the entire unbalanced panel). Then

$$E(A_1) = s_h^2 + s_u^2 + s_e^2$$
$$E(A_2) = s_h^2 + s_u^2$$
$$E(A_3) = s_h^2$$

Solving these equations for the three variances yields the following results (in 2005 dollars):

$$\hat{s}_h = \$11,392$$
$$\hat{s}_u = \$7,296$$
$$\hat{s}_e = \$11,928$$
$$\hat{s}_y = \$18,036$$

These magnitudes are in line with previous research indicating that the transient earnings component is responsible for about one-third of the variance of earnings.[15]

[15]See Gottschalk and Moffitt (1994) and Katz and Autor (1999).

The Signal-Extraction Problem

Consider an individual who is seen in locations 1,2,...m_i . Let T_{ij} denote the number of periods in location j, and let y_{ij} be the average wage residual in location j. In Appendix A we show that the conditional distribution of η_i, given $\{y_{ij}\}$, is normal, with mean $\hat{\eta}_i$ and variance Q_i, where

$$\hat{\eta}_i = \alpha_i \left(\frac{y_{i1}}{V_{i1}} + \frac{y_{i2}}{V_{i2}} + \cdots + \frac{y_{im_i}}{V_{im_i}} \right)$$

and where V_{ij} is the variance of the average wage in location j:

$$V_{ij} = \sigma_\upsilon^2 + \frac{\sigma_\varepsilon^2}{T_{ij}}$$

and where Q_i is given by

$$\frac{1}{Q_i} = \frac{1}{\sigma_\eta^2} + \frac{1}{V_{i1}} + \frac{1}{V_{i2}} + \cdots + \frac{1}{V_{im_i}}$$

Discrete Approximation of the Distribution of Location Match Effects

We approximate the decision problem by using a discrete distribution with n support points to represent the distribution of the location match component, and computing continuation values at these support points. This approximation works at two levels. First, we need to approximate the distribution from which the match components are drawn, in each location. Second, in order to compute the likelihood for individual i in location j, we need to approximate the conditional distribution of $\left\{ \upsilon_{ij} \right\}_{j=1}^{m_i}$, given the wage history.

For given support points, the best discrete approximation $\hat{F}$ for any distribution F assigns probabilities to the support points so as to equate $\hat{F}$ with the average value of F over each interval where $\hat{F}$ is constant. If the support points are variable, they are chosen so that $\hat{F}$ assigns equal probability to each point.[16] Thus the wage prospects associated with a move to State k are represented by an n-point distribution with equally weighted support points $a_k(r) = \hat{m}_k + \hat{u}(q_r)$, $1 \le r \le n$, where $\hat{\mu}_j$ is the estimated State effect, and $\hat{\upsilon}(q_r)$ is the q_r quantile of the distribution of υ, with

$$q_r = \frac{2r-1}{2n}$$

for $1 \le r \le n$. For example, with n = 3, the distribution of $\mu_j + \upsilon_{ij}$ in each State is approximated by a distribution that puts mass ⅓ on μ_j (which is the median of the distribution of $\mu_j + \upsilon_{ij}$), with mass ⅓ on $\mu_j \pm \sigma_\upsilon \Phi^{-1}(1/6)$, where Φ is the standard normal distribution function.

Suppose for a moment that the wage history included the individual fixed effect η_i. The location match components would then be conditionally independent, given the wage histories. Let F_{ij} be the conditional distribution function of $\mu_j + \upsilon_{ij}$, given the fixed effect and the wage history for individual i. The best approximation $\hat{F}_{ij}$ for this distribution, using the fixed support points $\{a_j(r)\}_{r=1}^{n}$, is given by

$$\hat{F}_{ij}\left(a_j(r)\right) = \frac{F_{ij}\left(a_j(r)\right) + F_{ij}\left(a_j(r+1)\right)}{2}$$

for $1 \le r \le n-1$, with $\hat{F}_{ij}(a_j(n)) = 1$. Thus the support points are the same for all individuals, but the weights on these points depend on the observed wage histories.

Of course we do not in fact have data on the fixed effects, so we integrate the approximation with respect to the distribution of η_i, conditional on the observed wages. For each individual, we draw a large sample from the conditional distribution, and for each realization we compute $\hat{F}_{ij}$ in the manner just described. Then we simply average the weights on the support points over all of the random draws of η_i.

4.5 The Likelihood Function

The likelihood of the observed history for each individual is a mixture over heterogeneous types. Let $L_{i\tau}$ be the likelihood of the history for an individual i, of type τ, and let p_τ be the probability of type τ. Then the loglikelihood is

[16] See Kennan (2004).

$$L = \sum_{i=1}^{N} \log\left(\sum_{t=1}^{K} p_t \, L_{it} \right)$$

Let θ_τ be the parameter vector, for someone of type τ. The components of the value function can be written more explicitly as

$$v_h(x, j, q_t) = u_h(x, j, q_t) + b \sum_{x'} p(x' | x, j) \bar{v}_h(x', q_t)$$

$$\bar{v}_h(x, q_t) = E_z \max_j \left[v_h(x, j, q_t) + z_j \right]$$

Then the choice probabilities are

$$r_h(x, j, q_t) = \exp\left(v_h(x, j, q_t) - \overline{v_h}(x, q_t) \right)$$

The likelihood of an individual history, for a person of type τ, is

$$L_{it} = \sum_s q(i,s) \left(\prod_{t=1}^{T_i} r_h\left(x_i(t,s), j_i(t), q_t \right) \right)$$

Here s is the profile of wage draws for individual i, and q(i,s) is the probability associated with this profile. In the case of someone who never moves, the sum has n terms (where n is the number of support points of the wage distribution). More generally, in the case of someone who visits m locations, the sum has n^m terms.

4.6 Computation

Since the parameters are embedded in the value function, computation of the gradient and hessian of the loglikelihood function is not a simple matter (although in principle these derivatives can be computed in a straightforward way using the same iterative procedure that computes the value function itself). We maximize the likelihood using a version of Newton's algorithm with numerical derivatives. We also use the downhill simplex method of Nelder and Mead, mainly to check for local maxima. This method does not use derivatives, but it is very slow.[17]

5 Empirical Results

We condition on the estimated earnings distributions for each State and maximize the partial likelihood to obtain estimates of the behavioral parameters. We set β =.95, T = 40, and M = 2. We show below that our main results are not very sensitive to these parameter settings.

Our basic results are shown in Table 2. We find that differences in expected income are a significant determinant of migration decisions for this population. There are 5,767 person-years in the data, with 213 interstate moves. This is an annual migration rate of 3.69%, and the first column in Table 2 matches this rate by setting the probability of moving to each of J-1 locations to a constant value, namely $\frac{1}{J-1}\frac{213}{5,767}$, with J = 51.[18] The next columns show that population size, distance, home and previous locations and age all have highly significant effects on migration. Local climate (represented by the annual number of heating and cooling degree-days) is also significant, although the sign of the coefficient on heating degree-days is perhaps surprising.[19] Unobserved heterogeneity in moving costs is introduced by allowing

---

[17]Given reasonable starting values (such as a fixed cost of moving that matches the average migration rate, with all other parameters set to zero), the maximal likelihood is reached by Newton's method within 60 hours, on a Pentium 4 machine. We found the Newton procedure to be well-behaved in the sense that it almost always reaches the same answer no matter what starting values are used: we have estimated hundreds of progressively more general versions of the model, and found only one local maximum; even in this case the loglikelihood and the parameter values were very close to the "true" maximum. An example of our (FORTRAN90) computer program can be found at www.ssc.wisc.edu/~jkennan/research/mbr55.f90.

[18]In other words the estimate of γ_0 solves the equation $\frac{1}{e^{\gamma_0}+J-1} = \frac{1}{J-1}\frac{213}{5,767}$; the solution is γ_0 = log(277700) - log(397).

[19]The climate variables are population-weighted monthly normals for 1931-2000, taken from Historical Climatography Series 5-1 (Heating Degree Days) and Series 5-2 (Cooling Degree Days) – see US NCDC (2002). The climate variables are measured in units of 1,000 degree-days. For example, the heating degree-day variable for Minnesota is 8.864, meaning that the difference between 65^0 and the mean daily temperature in Minnesota, summed over the days when the mean was below 65°, averaged 8,864 degree-days per year (over the years 1931-2000).

It might seem intuitively clear that the sign of the coefficient on heating degree-days should be negative. This intuition is valid for a single index of how warm the location is, but when both heating and cooling degree-days are included the intuition is not so clear, in view of the strong negative correlation between these variables. If we include only one variable or the other we get the expected result. When both variables are included we find that for two types, with one type treated as a pure stayer type (representing people with prohibitive moving costs); little is gained by introducing additional types, or by replacing the stayer type with a type with a high moving cost. The last column shows the effect of income, controlling for these other effects, using wages adjusted for cost of living differences across States.[20] These estimates are interpreted in the following subsections.

---

warmer locations (like Florida) are preferred. But our estimates also imply that among States that have the same number of cooling degree-days, there is a preference for more heating degree-days (i.e. for some cold weather). This might be interpreted as a preference for seasonal variation in climate, but it implies that States like Iowa and Nebraska have a better climate than California, which is counterintuitive.

We explored various alternative specifications of the climate amenity variables. The number of States that are adjacent to an ocean is 23. We considered this as an additional amenity variable, and also estimated models including annual rainfall, and the annual number of sunny days, but found that these variables had virtually no effect.

Given the sample size, we cannot expect to obtain precise estimates of the effects of climate variations. We include these variables mainly to ensure that the estimated effect of income is robust to the inclusion of nonpecuniary amenities.

[20]The validity of the estimates is checked in Appendix B: the estimated coefficients were used to generate simulated data sets, and the maximum likelihood procedure was applied to the simulated data. This worked well, although not perfectly: there were small discrepancies between the estimated and the true coefficients, leading to a rejection of the true DGP when the sample size was magnified by a factor of 100 or 1000.

Table 2: Interstate Migration of Young White Men				
Disutility of Moving (γ_0)	7.173	3.690	4.680	4.305
	0.000	*0.543*	*0.646*	*0.645*
Distance (γ_1)		0.325	0.275	0.278
(1000 miles, State pop centroids)		*0.138*	*0.147*	*0.148*
Adjacent Location (γ_2)		0.643	0.665	0.669
		0.153	*0.158*	*0.159*
Home Premium (κ)		0.290	0.275	0.372
		0.024	*0.021*	*0.032*
Previous Location (γ_3)		2.820	4.380	3.767
		0.235	*0.317*	*0.327*
Age (γ_4)		0.094	0.112	0.095
		0.021	*0.024*	*0.024*
Population (γ_5)		0.715	0.631	0.631
(10 million people)		*0.135*	*0.132*	*0.138*
Stayer Probability		0.481	0	0.426
		0.057	-----	*0.062*
Cooling		0.109	0.095	0.140
(1,000 degree-days)		*0.023*	*0.019*	*0.024*
Heating		0.019	0.015	0.025
(1,000 degree-days)		*0.009*	*0.008*	*0.010*
"Real" Income (α)		-------	0.466	0.552
($10,000)			*0.058*	*0.075*
Loglikelihood	-1744.88	-1309.60	-1305.44	-1287.86
Observations	5,767			
Moves	213			

5.1 Moving Costs and Payoff Shocks

Since utility is linear in income, the estimated moving cost can be converted to a dollar equivalent. Some examples are given in Table 3.

Table 3

Moving Cost Examples

	γ_0/α	Age	Distance	Adjacent	Population	Previous Location	Cost
Homogeneous Model							
Coefficients	$195,361	0.1121	0.2740	0.6654	0.6304	4.3772	
Young mover		20	1	0	1	0	$274,027
Average mover		23.5	0.715	0.338	0.759	0.300	$229,151
Move to Previous location		20	1	0	1	1	$91,490
Two-Type Model (mover type)							
Coefficients	$151,637	0.0948	0.2783	0.6693	0.6315	3.7668	
Young mover		20	1	0	1	0	$215,994
Average mover		23.5	0.715	0.338	0.759	0.300	$176,157
Move to Previous location		20	1	0	1	1	$58,911

For a 20–year-old in the homogeneous model the cost is about $274,000, in 2005 dollars. One might wonder why anyone would ever move in the face of such a cost, and in particular whether a move motivated by expected income gains could ever pay for itself. Our estimate of the standard deviation of the location match component is $7,296; the standard deviation of the State means is $2,328. A move that permanently increases both components by two standard deviations would be worth about $335,000 in present value (assuming a remaining worklife of 40 years, with $\beta = .95$). The home premium is equivalent to a wage increase of $13,090, and the cost of moving to a previous location is relatively low. Thus in some cases the expected income gains could be more than enough to pay for the estimated moving cost. Of course in most cases this would not be true, but then most people never move.

More importantly, the estimates in Table 3 do not refer to the costs of moves that are actually made, but rather to the costs of hypothetical moves to arbitrary locations. The cost is interpreted as the reduction in the expected continuation value associated with such a move, allowing for the psychic costs of moving, and also allowing for the option to return to the original location after one period. In the model, people choose to move only when the payoff shocks are favorable, and the net cost of the move is therefore much less than the amounts in Table 3. Consider for example a case in which someone is forced to move, but allowed to choose the best alternative location. The expected value of the maximum of J-1 draws from the extreme value distribution is γ+log(J-1) (where γ is Euler's constant), so if the location with the most favorable payoff shock is chosen, the expected net cost of the move is reduced by log(J-1)/α. In the homogeneous model, this is a reduction of $163,141. Moreover, this calculation refers to a move made in an arbitrary period; in the model, the individual can move later if the current payoff shocks are not sufficiently favorable, so the net cost is further reduced. Of course people actually move only if there is in fact a net gain from moving; the point of the argument is just that this can quite easily happen, despite the large moving cost estimates in Table 3.

Another way to interpret the moving cost is to consider the effect of a $10,000 migration subsidy, payable for every move, with no obligation to stay in the new location for more than one period. This can be analyzed by simulating the model with a reduction in γ_0 such that γ_0/α falls by $10,000, and with the other parameters held fixed. We estimate that such a subsidy would lead to a substantial increase in the interstate migration rate: from 3.7% to about 5%.

An Example

To understand the relationship between moving costs and prospective income gains, it is helpful to consider an example in which these are the only influences on migration decisions. Suppose that income in each location is either high or low, the difference being Δy, and suppose that the realization of income in each location is known. Then, using equation (3), the odds of moving are given by

$$\frac{1-\lambda_L}{\lambda_L} = e^{-\gamma_0}\left[J_L - 1 + J_H e^{\beta \Delta V}\right] \quad \textbf{(9)}$$

$$\frac{1-\lambda_H}{\lambda_H} = e^{-\gamma_0}\left[J_H - 1 + J_L e^{-\beta \Delta V}\right] \quad \textbf{(10)}$$

where λ_L is the probability of staying in one of J_L low-income locations (and similarly for λ_H and J_H), and where ΔV is the difference in expected continuation values between the low-income and high-income locations. This difference is determined by the equation

$$e^{\Delta V} = \frac{e^{\alpha \Delta y}\left[J_L + \left(J_H - 1 + e^{\gamma_0}\right) e^{\beta \Delta V}\right]}{J_L - 1 + e^{\gamma_0} + J_H e^{\beta \Delta V}} \quad \textbf{(11)}$$

For example, if $\beta = 0$, then $\Delta V = \alpha_0 \Delta y$, while if moving costs are prohibitive ($\exp(-\gamma_0) \approx 0$), then $\Delta V = \alpha_0 \Delta y/(1-\beta)$.

These equations uniquely identify α_0 and γ_0 (these parameters are in fact over-identified, because there is also information in the probabilities of moving to the same income level).[21] If $\gamma_0 < \beta\Delta V$, then the odds of moving from a low-income location are greater than J_H to 1, and this is contrary to what is seen in the data (for any plausible value of J_H). By making γ_0 a little bigger than $\beta\Delta V$, and letting both of these be large in relation to the payoff shocks, the probability of moving from the low-income location can be made small. But then the probability of moving from the high-income location is almost zero, which is not true in the data. In other words, if the probability of moving from a high-income location is not negligible, then the payoff shocks cannot be negligible, since a payoff shock is the only reason for making such a move.

The net cost of moving from a low-income location to a high-income location is $\gamma_0 - \beta\Delta V$, while the net cost of the reverse move is $\gamma_0 + \beta\Delta V$. The cost difference is $2\beta\Delta V$, and equations (9) and (10) show that $\beta\Delta V$ determines the relative odds of moving from low-income and high-income locations. Thus $\beta\Delta V$ is identified by the difference between λ_L and λ_H; this difference is small in the data, so $\beta\Delta V$ must be small. The magnitude of γ_0 is then determined by the level of λ_L and λ_H, and since these are close to 1 in the data, the implication is that γ_0 is large, and that it is much larger than $\beta\Delta V$. Since $\beta\Delta V$ is roughly the present value of the difference in income levels, the upshot is that the moving cost must be large in relation to income.

For example, suppose $J_L = J_H = 25$, with $\beta = .95$. In our data, the migration probability for someone in the bottom half of the distribution of wage residuals is 4.7%, and for someone in the top half it is 2.7%. If $\lambda_L = .953$ and $\lambda_H = .973$, then $\gamma_0 = 7.14$, and $\Delta V = .55$, and the implied moving cost is $\gamma_0/\alpha = 143.3\Delta y$. On the other hand if $\lambda_L = .7$, the implied moving cost is only $14.5\Delta y$. We conclude that the moving cost estimate is large mainly because the empirical relationship between current income and migration probabilities is relatively weak.

5.2 Offered and Accepted Wages

The wage distribution seen in the data is the distribution of accepted wages, but we treat it as if it were the distribution of offered wages. Although this introduces a selection bias, the distribution of accepted wages actually provides a good approximation of the offer distribution, because the migration rate is low, and because many of the moves that do occur are not motivated by wage differences.

[21] It is assumed that λ_L, λ_H, J_L, J_H, Δy and β are given. Dividing (9) by (10) and rearranging terms yields a quadratic equation in $e^{\beta\Delta V}$ that has one positive root and one negative root. Since $e^{\beta\Delta V}$ must be positive, this gives a unique solution for ΔV. Equation (9) then gives a unique solution for γ_0, and inserting these solutions into equation (11) gives a unique solution for $\alpha\Delta y$.

Using simulated data, we compared the distribution of wage offers with the distribution of accepted wages, and found that selection effects are indeed small. In the model, wage offers are uniformly distributed over three support points. The distribution of accepted wages has the same support points, but it puts more weight on higher wages. Table 4 shows that although the wage distribution shifts slightly toward the upper tail as the migration process plays out, the low support point still has 28% of the observations after 15 years (instead of 33%). The mean of the accepted wage distribution is about 2% above the mean of the offer distribution.

Table 4: Distribution of Accepted Wages

Period	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Low	0.333	0.322	0.314	0.308	0.303	0.299	0.295	0.292	0.289	0.287	0.285	0.284	0.283	0.282	0.280
Middle	0.334	0.338	0.340	0.341	0.342	0.343	0.344	0.345	0.345	0.345	0.345	0.346	0.346	0.346	0.346
High	0.333	0.340	0.346	0.351	0.355	0.358	0.361	0.364	0.366	0.368	0.369	0.371	0.372	0.373	0.374

5.3 Goodness of Fit

In order to keep the state space manageable, our model severely restricts the set of variables that are allowed to affect migration decisions. Examples of omitted observable variables include duration in the current location, and the number of moves made previously. In addition, there are of course unobserved characteristics that might make some people more likely to move than others. Thus it is important to check how well the model fits the data. In particular, since the model pays little attention to individual histories, one might expect that it would have trouble fitting panel data.

One simple test of goodness of fit can be made by comparing the number of moves per person in the data with the number predicted by the model. As a benchmark, we consider a binomial distribution with a migration probability of 3.69% (the number of moves per person-year in the data). Table 5 shows the predictions from this model: about 73% of the people never move, and of those who do move, about 15% move more than once.[22] The NLSY data are quite different: about 82% never move, and about 53% of movers move more than once. A natural interpretation of this is mover-stayer heterogeneity: some people are more likely to move than others, and these people account for more than their share of the observed moves. We simulated the corresponding statistics for the model by starting 1000 replicas of the NLSY individuals in the observed initial locations, and using the model (with the estimated parameters shown in Table 2) to generate a history for each replica, covering the number of periods observed for this

[22] Since we have an unbalanced panel, the binomial probabilities are weighted by the distribution of years per person.

individual. In the case of the homogeneous model, the fit is good, although both the proportion of people who never move, and the proportion of movers who move more than once are a bit low, relative to the data. In this respect, the observables in the model do a good job of accounting for the heterogeneous migration probabilities in the data. Allowing for unobserved heterogeneity yields a slight improvement in the ability of the model to fit the proportion of people who never move; in this model the proportion of repeat movers is a bit too high.

Table 5: Goodness of Fit

Moves	Binomial		NLSY		Homogeneous Model		Two-Type Model	
None	482.8	72.6%	544	81.80%	532083	80.01%	546251	82.14%
One	154.4	23.2%	57	8.57%	69488	10.45%	49342	7.42%
More	27.80	4.2%	64	9.62%	63429	9.54%	69419	10.44%
Proportion of movers with more than one move	15.26%		52.89%		47.72%		58.45%	
Total observations	665		665		665000		665012	

Return Migration

Table 6 summarizes the extent to which the model can reproduce the return migration patterns in the data (the statistics in the Model columns refer to the simulated data sets used in Table 5).

Table 6: Return Migration Statistics

	NLSY	Homogeneous Model	Two-Type Model
Proportion of Movers who			
Return home	33.8%	32.7%	33.7%
Return elsewhere	5.6%	7.1%	7.4%
Move on	60.6%	60.1%	58.8%
Proportion who *ever*			
Leave Home	15.3%	15.4%	14.8%
Move from not-home	41.7%	58.2%	43.3%
Return from not-home	23.6%	31.7%	28.6%

The model attaches a premium to the home location, and this helps explain why people return home. For example, in a model with no home premium, one would expect that the proportion of movers going to any particular location would be roughly 1/50, and this obviously does not match the observed return rate of 34%. The home premium also reduces the chance of initially leaving home, although this effect is offset by the substantial discount on the cost of returning to a previous location (including the home location): leaving home is less costly if a return move is relatively cheap.

The simulated return migration rates match the data reasonably well. The main discrepancy is that the homogeneous model substantially over-predicts the proportion who ever move from an initial location that is not their home location. That is, the model has trouble explaining why people seem so attached to an initial location that is not their "home". One potential explanation for this is that our assignment of home locations (the State of residence at age 14) is too crude (in some cases the location at age 20 may be more like a home location than the location at age 14). More generally, people are no doubt more likely to put down roots the longer they stay in a location, and our model does not capture this kind of duration dependence. The two-type model does a much better job of fitting this aspect of the data, although of course the introduction of unobserved heterogeneity doesn't really add much to our understanding of the underlying behavior.

5.4 Why are Younger People More Likely to Move?

It is well known that the propensity to migrate falls with age (at least after age 25 or so). Table 7 replicates this finding for our sample of high-school men. A standard human capital explanation for this age effect is that migration is an investment: if a higher income stream is available elsewhere, then the sooner a move is made, the sooner the gain is realized. Moreover, since the worklife is finite, a move that is worthwhile for a young worker might not be worthwhile for an older worker, since there is less time for the higher income stream to offset the moving cost (Sjaastad [1962]). In other words, migrants are more likely to be young for the same reason that students are more likely to be young.

Table 7 Annual Interstate Migration Rates by Age and Current Location						
	All		Not At Home[a]		At Home	
Age	N	Migration Rate	N	Migration Rate	N	Migration Rate
20	677	4.73%	74	21.62%	603	2.65%
21	637	4.87%	74	14.86%	563	3.55%
22	609	5.09%	81	19.75%	528	2.84%
23	569	3.51%	83	13.25%	486	1.85%
24	587	4.09%	83	15.66%	504	2.18%
25	533	4.69%	79	12.66%	454	3.30%
26	512	4.49%	80	17.50%	432	2.08%
27	465	1.94%	73	9.59%	392	0.51%
28	381	1.57%	57	5.26%	324	0.93%
29	307	1.63%	51	3.92%	256	1.17%
30	242	1.65%	38	7.89%	204	0.49%
31	149	2.01%	21	9.52%	128	0.78%
32	81	0.00%	12	0.00%	69	0.00%
33	18	0.00%	1	0.00%	17	0.00%
All	5,767	3.69%	807	13.38%	4,960	2.12%

[a]At Home means living now in the State of residence at age 14.

Our model encompasses this simple human capital explanation of the age effect on migration. There are two effects here. First, consider two locations paying different wages, and suppose that workers are randomly assigned across these locations at birth. Then, even if the horizon is infinite, the model predicts that the probability of moving from the low-wage to the high-wage location is higher than the probability of a move in the other direction, so that eventually there will be more workers in the high-wage location. This implies that the (unconditional) migration rate must be higher when workers are young.[23] Second, the human capital explanation says that migration rates decline with age because the

[23]One way to see this is to consider the extreme case in which there are no payoff shocks. In this case all workers born in the low-wage location will move to the high-wage location at the first opportunity (if the wage difference is big enough to offset the moving cost), and the migration rate will be zero from then on.

horizon gets closer as workers get older. This is surely an important reason for the difference in migration propensities between young adult workers and those within sight of retirement. But the workers in our sample are all in their twenties or early thirties, and the prospect of retirement seems unimportant for such workers.

The human capital model does not adequately explain the relationship between age and migration in the data. Our model includes age as a state variable, to capture the finite horizon effect just discussed. The model also allows for the possibility that age has a direct effect on the cost of migration; this can be regarded as a catch-all for whatever is missing from the human capital explanation.[24] The results in Table 2 show that this direct effect is large and significant.

5.5 Decomposing the Effects of Income on Migration Decisions

In our model, differences in wage distributions across States are due entirely to differences in State means. This raises the question of whether the estimated coefficients would be similar if wage dispersion within States is ignored, and migration decisions are modeled as responses to differences in mean wages across locations. At the other extreme, the wage distribution can be specified at the national level, with no variation across States; in this case migration is motivated only by the prospect of getting a better draw from the same wage distribution (given our assumption that location match effects are permanent). Results for these alternative specifications are shown in Table 8. The estimated income coefficient is significant even in the absence of within-State dispersion, and it is also significant even when all States share the same wage distribution.

[24] Marriage is of course a potentially important factor, but in order to deal with this we would have to double or triple the size of the state space (depending on whether we distinguished between divorced and single people). It is worth noting that if we take marital status as given, it has essentially no effect on migration in our sample, in simple logit models of the move-stay decision that include age as an explanatory variable.

Table 8: Alternative Income Specifications

	Census	State Means	National
Disutility of Moving	4.305	3.731	4.410
	0.645	*0.545*	*0.670*
Distance (1000 miles)	0.278	0.270	0.344
	0.148	*0.140*	*0.149*
Adjacent Location	0.669	0.692	0.608
	0.159	*0.153*	*0.159*
Home Premium	0.372	0.296	0.367
	0.032	*0.025*	*0.031*
Previous Location (moving cost)	3.767	2.895	4.013
	0.327	*0.236*	*0.401*
Age	0.095	0.096	0.099
	0.024	*0.022*	*0.024*
Population	0.631	0.665	0.700
	0.138	*0.135*	*0.138*
Stayer Probability	0.426	0.474	0.402
	0.062	*0.057*	*0.070*
Cooling degree-days	0.140	0.132	0.104
	0.024	*0.025*	*0.022*
Heating degree-days	0.025	0.024	0.015
	0.010	*0.009*	*0.009*
Real Income (ACCRA)	0.552	0.514	0.629
	0.075	*0.103*	*0.087*
Loglikelihood	-1287.86	-1300.24	-1294.59
N (person-years)	5,767		
Moves	213		

Notes:

The “State Means” column assumes that there is no wage dispersion within States.

The “National” column assumes that wage distributions are identical in all States.

5.6 Sensitivity Analysis

Our empirical results are inevitably based on some more or less arbitrary model specification choices. Table 9 explores the robustness of the results with respect to some of these choices. The general conclusion is that the parameter estimates are robust. In particular, the income coefficient estimate remains positive and significant in all of our alternative specifications.

The results presented so far are based on wages that are adjusted for cost of living differences across locations. If these cost of living differences merely capitalize the value of amenity differences, then unadjusted wages should be used to measure the incentive to migrate. Results for this specification are given in the second column of Table 9: the estimate of α is substantially reduced, without much effect on the other coefficients, and the likelihood is lower. Thus in practice the theoretical ambiguity as to whether wages should be adjusted for cost of living differences does not change the qualitative empirical results: either way, income significantly affects migration decisions.

The last three columns of Table 9 show the relevance of the climate variables. Including these variables gives a substantial improvement in the fit, without changing the conclusion that income significantly affects migration. If either climate variable is included on its own, the result is that warmer places are preferred.

The other specifications in Table 9 are concerned with sensitivity of the estimates to the discount factor (β), and the horizon length (T). Reducing β to .90 noticeably affects the utility flow parameters (i.e. the home premium and the income and amenity coefficients), with hardly any effect on the moving cost parameters. Although a 5% annual real interest rate is arguably more plausible than a 10% rate, the likelihood is slightly higher when β is set at .90.[25]

[25] Table 9 is a representative sample of various alternative specifications that were tried. As was mentioned earlier, size (as measured by population) may affect migration either as a scaling factor on the payoff shocks, or as a variable affecting the cost of migration. We experimented with these alternatives, and also expanded the moving cost specification to allow quadratic effects of distance and location size and climate variables; none of these experiments changed the results much.

The NLSY oversamples people whose parents were poor, and one might expect that the income process for such people is atypical, and that the effect of income on migration decisions might also be atypical. Results for this subsample were similar to those for the full sample.

Table 9: Alternative Specifications

	Baseline	No Cola	β = .9	β = .975	T = 20	No Climate	No Heating	No Cooling
Disutility of Moving	4.305	3.929	4.034	4.452	4.183	4.010	4.319	4.142
	0.645	*0.597*	*0.658*	*0.637*	*0.672*	*0.598*	*0.644*	*0.621*
Distance (1000 miles)	0.278	0.367	0.311	0.260	0.287	0.266	0.281	0.278
	0.148	*0.146*	*0.159*	*0.141*	*0.15*	*0.143*	*0.144*	*0.141*
Adjacent Location	0.669	0.635	0.72	0.638	0.67	0.653	0.691	0.681
	0.159	*0.156*	*0.17*	*0.152*	*0.16*	*0.153*	*0.158*	*0.157*
Home Premium	0.372	0.348	0.573	0.277	0.462	0.308	0.373	0.345
	0.032	*0.03*	*0.044*	*0.026*	*0.034*	*0.028*	*0.032*	*0.031*
Previous Location (moving cost)	3.767	3.929	3.667	3.809	3.617	3.378	3.751	3.556
	0.327	*0.597*	*0.323*	*0.325*	*0.313*	*0.281*	*0.316*	*0.293*
Age	0.095	0.093	0.102	0.091	0.094	0.099	0.093	0.094
	0.024	*0.022*	*0.025*	*0.023*	*0.025*	*0.023*	*0.024*	*0.024*
Population	0.631	0.638	0.685	0.598	0.64	0.648	0.610	0.565
	0.138	*0.136*	*0.144*	*0.135*	*0.139*	*0.132*	*0.133*	*0.131*
Stayer Probability	0.426	0.452	0.397	0.448	0.412	0.412	0.422	0.418
	0.062	*0.058*	*0.064*	*0.061*	*0.063*	*0.063*	*0.062*	*0.062*
Cooling degree-days	0.140	0.144	0.215	0.107	0.171	-------	0.087	-------
	0.024	*0.023*	*0.036*	*0.019*	*0.029*		*0.013*	
Heating degree-days	0.025	0.024	0.038	0.019	0.029	-------	-------	-0.024
	0.010	*0.008*	*0.015*	*0.007*	*0.012*			*0.006*
Income	0.552	0.350	0.863	0.407	0.684	0.410	0.533	0.458
	0.075	*0.063*	*0.114*	*0.056*	*0.091*	*0.070*	0.072	*0.069*
Loglikelihood	-1287.86	-1297.22	-1287.36	-1288.89	-1286.14	-1308.62	-1290.23	-1300.29

6 Spatial Labor Supply Elasticities

We use the estimated model to analyze labor supply responses to shifts in mean wages, for selected States. We are interested in the magnitudes of the migration flows in response to local wage changes, and in the timing of these responses. Since our model assumes that the wage components relevant to migration decisions are permanent, it cannot be used to predict responses to wage innovations in an environment in which wages are generated by a stochastic process. Instead, it is used to answer comparative dynamics questions: having used the data to estimate behavioral parameters, we use these parameters to predict responses in a different environment. First we do a baseline simulation, starting people in given locations, and allowing them to make migration decisions in response to the wage distributions estimated from the Census data. Then we do counterfactual simulations, starting people in the same locations, facing different wage distributions.

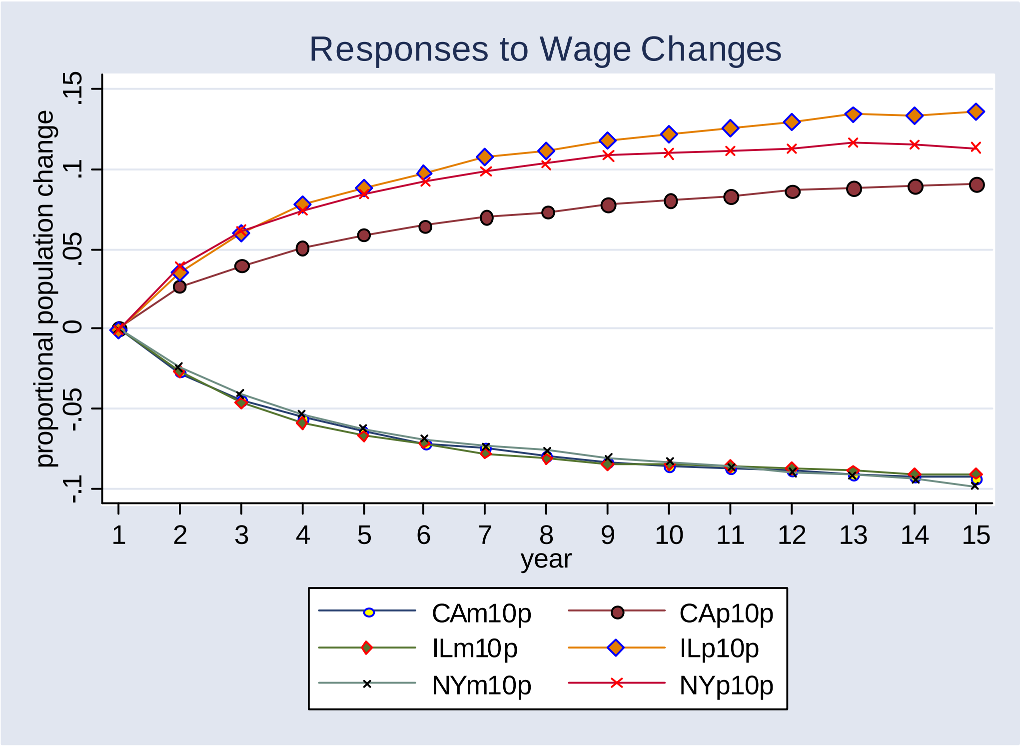
We take a set of 10,000 people, with 100 replicas of each person, distributed over States according to the 1990 Census data for white male high school graduates aged 20 to 34. We assume that each person is initially in the home State, at age 20, and simulate 15-year histories. We consider responses to 10% increases and decreases in wages, for selected States. First, we simulate baseline migration decisions using the actual Census wage distributions described previously. Then we increase or decrease the mean wage in a single State by 10%, and compare the migration decisions induced by these wages with the baseline. Supply elasticities are measured relative to the supply of labor in the baseline simulation. For example, the elasticity of the response to a wage increase in California after 5 years is computed as $(\Delta L/L)/(\Delta w/w)$, where L is the number of people in California after 5 years in the baseline simulation, and ΔL is the difference between this and the number of people in California after 5 years in the counterfactual simulation.

Figure 1 shows the results obtained from the heterogeneous model, for three large States that are near the middle of the one-period utility flow distribution. The supply elasticities are large: between 0.9 and 1.3. Adjustment is gradual, but is largely completed in 10 years.[26] Somewhat surprisingly, the responses are not symmetric, and there are noticeable differences across States.[27] Our conclusion from this exercise is that despite the low migration rate in the data, and the large migration costs implied by this rate, the

[26]Results for the homogeneous model are similar, although the elasticities are a bit bigger (between 1.1 and 1.6).

[27]These differences are much more pronounced for States at the top of the utility-flow ranking. For example, our estimates imply that the highest utility flow is achieved in Florida and Texas (mainly because of a favorable climate). A 10% wage increase in these States yields a population increase of more than 25% after 15 years. This is apparently because increasing the wage in a location that is already desirable makes this the destination of choice for anyone who is moving from a nearby location because of an adverse payoff shock in that location. The responses to wage decreases for these States are similar to the results in Figure 1. Moreover, when the climate variables are excluded from the model, the responses to wage increases are similar to the responses for the other States..

supply of labor responds quite strongly to spatial wage differences.



7 Conclusion

We have developed a tractable econometric model of optimal migration in response to income differentials across locations. The model improves on previous work in two respects: it covers optimal sequences of location decisions (rather than a single once-for-all choice), and it allows for many alternative location choices. Migration decisions are made so as to maximize the expected present value of lifetime income, but these decisions are modified by the influence of unobserved location-specific payoff shocks. Because the number of locations is too large to allow the complete dynamic programming problem to be modeled, we adopt an approximation that truncates the amount of information available to the decision-maker. The practical effect of this is that the decisions of a relatively small set of people who have made an unusually large number of moves are modeled less accurately than they would be in the (computationally infeasible) complete model.

Our empirical results show a significant effect of expected income differences on interstate migration, for white male high school graduates in the NLSY. Simulations of hypothetical local wage changes show that the elasticity of the relationship between wages and migration is roughly unity . Our results can be interpreted in terms of optimal search for the best geographic match. In particular, we find that the relationship between income and migration is largely driven by a negative effect of income in the current location on the probability of out-migration: workers who get a good draw in their current location tend to stay, while those who get a bad draw tend to leave.

The main limitations of our model are those imposed by the discrete dynamic programming structure: given the large number of alternative location choices, the number of dynamic programming states must be severely restricted for computational reasons. Goodness of fit tests indicate that the model nevertheless fits the data reasonably well. From an economic point of view, the most important limitation of the model is that it imposes restrictions on the wage process implying that individual fixed effects and movements along the age-earnings profile do not affect migration decisions. A less restrictive specification of the wage process would be highly desirable.

Appendix A: Signal Extraction

Lemma A1

Suppose $\{y_j\}_{j=1}^{m}$ is a set of random variables, with

$$y_j = \eta + u_j$$

where η and $\{u_j\}$ are independent and normally distributed, with zero mean, and variances σ_η^2 and V_j. Then the conditional distribution of η, given $\{y_j\}$, is normal, with mean $\hat{\eta}$ and variance Q_0, where

$$\hat{\eta} = Q_0 \sum_{j=1}^{m} \frac{y_j}{V_j}$$

and

$$\frac{1}{Q_0} = \frac{1}{\sigma_\eta^2} + \sum_{j=1}^{m} \frac{1}{V_j}$$

Proof

We will show that $E(\eta-\hat{\eta})y_j = 0$, for each j. Since the components are normally distributed, this implies that $\eta-\hat{\eta}$ and y_j are independent, so $E(\eta-\hat{\eta}) \mid y_j = E(\eta-\hat{\eta}) = 0$.

First, $E\eta y_j = \sigma_\eta^2$. Consider the covariance $E\hat{\eta}y_j$ for an arbitrary choice of j, say j = 1:

$$E\hat{\eta}\, y_1 = Q_0\, E\, y_1\left(\frac{y_1}{V_1} + \frac{y_2}{V_2} + \cdots + \frac{y_m}{V_m}\right)$$

$$= Q_0\left(\frac{\sigma_y^2}{V_1} + \frac{\sigma_\eta^2}{V_2} + \cdots + \frac{\sigma_\eta^2}{V_m}\right)$$

The definition of Q_0 implies that

$$Q_0\left(\frac{\sigma_\eta^2}{V_1}+\frac{\sigma_\eta^2}{V_2}+\cdots+\frac{\sigma_\eta^2}{V_m}\right)=\sigma_\eta^2-\alpha_0$$

Thus

$$\begin{aligned}E\hat{\eta}y_1 &= Q_0\left(\frac{\sigma_y^2-\sigma_\eta^2}{V_1}\right)+\sigma_\eta^2-Q_0\\ &= \sigma_\eta^2\end{aligned}$$

So $E(\eta-\hat{\eta})y_j = 0$, for each i. This proves that $\hat{\eta}$ is the conditional mean. To determine the variance, write

$$\hat{\eta}=\left(1-\frac{Q_0}{\sigma_\eta^2}\right)\eta+Q_0\sum_{j=1}^{m}\frac{\upsilon_j}{V_j}$$

Then

$$\begin{aligned}E\left(\eta-\hat{\eta}\right)^2 &= \frac{Q_0^2}{\sigma_\eta^2}+Q_0^2\sum_{j=1}^{m}\frac{1}{V_j}\\ &= \frac{Q_0^2}{\sigma_\eta^2}+Q_0^2\left(\frac{1}{Q_0}-\frac{1}{\sigma_\eta^2}\right)=Q_0\end{aligned}$$

Lemma A2

Suppose $\{y_j\}_{j=1}^{m}$ is a set of random variables, with

$$y_j=\eta+\upsilon_j+\varepsilon_j$$

where η, $\{\upsilon_j\}$ and $\{\varepsilon_j\}$ are independent and normally distributed, with zero mean, and variances σ_η^2, σ_υ^2 and τ_j, respectively. Then the conditional distribution of υ_j, given $\{y_j\}$ and η, is normal, with mean $\hat{\upsilon}_j$ and variance Q_j, where

$$\hat{u}_j = Q_j\left(\frac{y_j - h}{t_j}\right)$$

$$\frac{1}{Q_j} = \frac{1}{s_u^2} + \frac{1}{t_j}$$

Proof

This is a standard signal extraction result. It is also a special case of Lemma A1.

Appendix B: Validation of ML Estimates

The parameter estimates from Table 2 were used to generate replicas of each NLSY observation, starting from the actual value in the NLSY data, and allowing the model to choose the sequence of locations. Table 10 gives results for 10, 100, and 1,000 replicas of each NLSY observation. The last column reports the t-value testing the difference between the 1,000 replica estimates and the individual DGP parameters; the last row reports likelihood ratio tests of the hypothesis that the data were generated by the process that did in fact generate them (assuming that the simulation program works) . In all cases, the estimated coefficients are close to the true values. As the number of replicas increases, the estimated standard errors decline and the DGP is rejected. There is some evidence of a small bias (about 5%) in the estimate of the income coefficient.

Table 10: Estimates from Simulated Migration Histories

	NLSY	10 Reps	100 Reps	1000 Reps	t
Disutility of Moving (γ_0)	4.305	4.155	4.150	4.228	4.63
	0.645	*0.166*	*0.052*	*0.017*	
Distance (γ_1)	0.278	0.331	0.277	0.286	2.28
	0.148	*0.040*	*0.011*	*0.004*	
Adjacent Location (γ_2)	0.669	0.624	0.682	0.674	1.18
	0.159	*0.045*	*0.014*	*0.004*	
Home Premium (κ)	0.372	0.357	0.361	0.361	1.11
	0.032	*0.011*	*0.003*	*0.010*	
Previous Location (γ_3)	3.767	3.709	3.706	3.702	7.89
	0.327	*0.088*	*0.026*	*0.008*	
Age (γ_4)	0.095	0.098	0.100	0.097	2.34
	0.024	*0.006*	*0.002*	*0.001*	
Population (γ_5) (10 million people)	0.631	0.647	0.632	0.634	0.68
	0.138	*0.043*	*0.013*	*0.004*	
Stayer Probability	0.426	0.418	0.425	0.424	0.88
	0.062	*0.021*	*0.006*	*0.002*	
Cooling (1,000 degree-days)	0.140	0.148	0.134	0.138	2.81
	0.024	*0.009*	*0.003*	*0.001*	
Heating (1,000 degree-days)	0.025	0.029	0.022	0.024	2.55
	0.010	*0.004*	*0.001*	*0.000*	
“Real” Income (α) ($10,000)	0.552	0.516	0.524	0.526	15.54
	0.075	*0.019*	*0.005*	*0.002*	
Loglikelihood	-1287.86	-12,733.9	-124,600.7	-1,256,289.3	
Observations	5,767	580,262	576,517	5,767,065	
Moves	213	2,155	21,617	218,292	
$\chi^2(11)$		8.0	30.5	252.7	

References

American Chamber of Commerce Research Association (2000) Cost of Living Index. Arlington, VA; http://www.coli.org.

Banks, Jeffrey S. and Rangarajan K. Sundaram (1994), "Switching Costs and the Gittins Index," *Econometrica*, 62 (3): 687-694.

Barro, Robert J. and Xavier Sala-i-Martin (1991), "Convergence across States and Regions," *Brookings Papers on Economic Activity*, **1:** 107-158.

Blanchard, Olivier Jean and Lawrence F. Katz (1992), "Regional Evolutions," *Brookings Papers on Economic Activity*, **1:** 1-37.

Dahl, Gordon B., "Mobility and the Return to Education: Testing a Roy Model with Multiple Markets," *Econometrica*, Vol. 70, No. 6, pp. 2367-2420, November 2002.

Gallin, Joshua Hojvat (2004), "Net Migration and State Labor Market Dynamics," *Journal of Labor Economics*, 22, 1-21.

Gottschalk, Peter and Robert Moffitt (1994), "The Growth of Earnings Instability in the U.S. Labor Market," *Brookings Papers on Economic Activity* 2: 255-272.

Greenwood, Michael J. (1997), "Internal Migration in Developed Countries," in *Handbook of Population and Family Economics Vol. 1B*, edited by Mark R. Rosenzweig and Oded Stark. New York: North Holland.

Ham, John C., Xianghong Li and Patricia B. Reagan, "Matching and Selection Estimates of the Effect of Migration on Wages for Young Men," Ohio State University, November 2001.

Holt, Frederick (1996), "Family Migration Decisions: A Dynamic Analysis," unpublished paper, University of Virginia.

Katz, Lawrence and David Autor (1999), "Changes in the Wage Structure and Earnings Inequality," *Handbook of Labor Economics Volume 3A.* New York: Elsevier.

Keane, Michael P. and Kenneth I. Wolpin (1997), "The Career Decisions of Young Men," *Journal of Political Economy*, 105: (3), June 1997, 473-522.

Kennan, John (2004), "A Note on Approximating Distribution Functions," http://www.ssc.wisc.edu/~jkennan/research/DiscreteApprox.pdf

Kennan, John and James R. Walker (2001), "Geographical Wage Differentials, Welfare Benefits and Migration," (March); http://www.ssc.wisc.edu/~jkennan/research/jkjwPaper03-01.pdf.

Lucas, Robert E. B. (1997), "Internal Migration in Developing Countries," in *Handbook of Population and Family Economics Vol. 1B*, edited by Mark R. Rosenzweig and Oded Stark. New York: North Holland.

McFadden, D. (1973), “Conditional Logit Analysis of Qualitative Choice Behavior,” in P. Zarembka (ed.) *Frontiers in Econometrics*, New York, Academic Press.

Neal, Derek (1999), “The Complexity of Job Mobility of Young Men,” *Journal of Labor Economics*, (April): 237-261.

Pessino, Carola (1991), “Sequential migration theory and evidence from Peru,” *Journal of Development Economics* 36 (July): 55–87.

Reagan, Patricia and Randall Olsen (2000) “You Can Go Home Again: Evidence from Longitudinal Data,” *Demography* (August) 37: 339-350.

Rust, John (1987) “Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher,” *Econometrica*, 55 (5): 999-1033.

Rust, John (1994), “Structural Estimation of Markov Decision Processes,” in *Handbook of Econometrics*, Volume IV. Edited by Robert F. Engle and Daniel L. McFadden. New York: Elsevier.

Schultz, T. Paul (1982) “Lifetime Migration within Educational Strata in Venezuela: Estimates of a Logistic Model,” *Economic Development and Cultural Change* 30: 559-593.

Sjaastad, L. A. (1962), “The Costs and Returns of Human Migration,” *Journal of Political Economy* 70:80-89.

Topel, Robert H. (1986), “Local Labor Markets,” *Journal of Political Economy*, 94(3), part 2, S111-S143.

Tunali, Insan (2000), “Rationality of Migration,” *International Economic Review*, (November) 41: 893-920.

United States National Climatic Data Center/NESDIS/NOAA (2002) Historical Climatography Series 5-1 and 5-2. Asheville, NC.

Woo, Seokjin (2002), “An Empirical Analysis of the Effects of State Fiscal Policy on Retirement Migration,” University of Wisconsin-Madison, November 2002.