Draft

The Effect of Expected Income on Individual Migration Decisions

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

March 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, 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 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.

---

[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.

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 is proportional to distance, 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.

Our main substantive conclusion is that interstate migration decisions are indeed influenced to a substantial extent by income prospects. On the other hand we find no evidence of a response to geographic differences in wage distributions. Instead, the results suggest that the link between income and migration decisions is 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 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.

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

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. Although there may be transient fluctuations in wages, the only chance of getting a permanent wage gain is to 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.[5] In that case, our model can be viewed as an

[5]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

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.[6] 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).

3.1 The Value Function

Let x be the state vector (which includes current location and age). The utility flow for someone who chooses location j is specified as $u(x,j) + \zeta_j$, where ζ_j is a preference shock 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

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

---

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).

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

where

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

and

$$\bar{v}(x) = E_{\zeta} V(x,\zeta)$$

and where E_ζ denotes the expectation with respect to the preference shocks. We assume that the preference shocks are drawn from the Type I extreme value distribution. In this case, using arguments due to McFadden (1973) and Rust (1987), we have

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

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

$$\rho(x,j) = \exp[v(x,j) - \bar{v}(x)] \qquad \textbf{(3)}$$

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 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, 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

When the number of locations is moderately large, the model becomes infeasible.[7] To deal with this problem 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, no matter how much time has passed. 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.[8] 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 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 (and people who have different wage information now may have the same information following a move).

[7]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.

[8]Note that it is not enough to keep track of the best wage found so far: the preference 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).

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.

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

$$u_h(x,j) = \alpha_0 y(\ell^0,\omega^0) + \sum_{k=1}^{K} \alpha_k Y_k(\ell^0) + \kappa \chi_{\{\ell^0 = h\}} - \Delta_\tau(x,j)$$

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 is represented by $\Delta(x,j)$. This is specified as

$$\Delta_\tau(x,j) = \left(\gamma_0(\tau) + \gamma_1 D(\ell^0,j) - \gamma_2 \chi(j = \ell^1) + \gamma_3 a - \gamma_4 n_j\right) \chi(j \neq \ell^0)$$

The first two terms in parentheses specify the moving cost as an affine function of the distance $D(\ell^0,j)$ from ℓ^0 to j. We allow for unobserved heterogeneity in the fixed cost of moving: there are several types, indexed by τ, with differing values of the fixed cost γ_0. In particular, we allow the possibility that there is a "stayer" type, meaning that there are people who regard the cost of moving as prohibitive, in all states. The next term allows for the possibility that it is cheaper to move to a previous location, relative to moving to a new location. The cost of moving is also allowed to depend on age, a. The last term allows 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 preference shocks in each location.[9] 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

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

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, (ℓ',ω') is obtained from (ℓ,ω) by interchanging the current and previous locations. Finally, 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.[10] The sample includes only people who had completed

---

[9]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.

[10]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

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 5767 person-years.

4.4 Wages

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

$$w_{ij}(a) = X_i\beta + \phi(a) + \mu_j + \upsilon_{ij} + \eta_i + \varepsilon_{ij}(a)$$

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, and identically distributed (across individuals and locations).

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 location, say k, which is measured by $\mu_k + \upsilon_{ik}$ (since the other components 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.

---

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.

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.[11] 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.[12] Taking the residuals from this regression (and ignoring sampling error), we have a set of individual wage histories, with observable effects removed:

$$\begin{aligned} y_{it} &= w_{it} - X_i\beta - \phi(a_{it}) - \mu_{j(t)} \\ &= \eta_i + \upsilon_{ij_{it}} + \varepsilon_t \end{aligned}$$

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. This is a simple signal extraction problem. We first generate estimates of 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

$$\begin{aligned} E(A_1) &= \sigma_\eta^2 + \sigma_\upsilon^2 + \sigma_\varepsilon^2 \\ E(A_2) &= \sigma_\eta^2 + \sigma_\upsilon^2 \\ E(A_3) &= \sigma_\eta^2 \end{aligned}$$

Solving these three equations for the three variances yields the following results (in 2004 dollars):

---

[11][explain details of Census wage measurements]

[12][more details. explain afqt, mention importance in previous wage function estimates; linear age effect included, just in case there is a difference between Census and NLSY age profiles (there isn't)].

$$\hat{\sigma}_\eta = \$11{,}181$$
$$\hat{\sigma}_\upsilon = \$7{,}161$$
$$\hat{\sigma}_\varepsilon = \$11{,}707$$
$$\hat{\sigma}_y = \$17{,}702$$

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.[13]

The Signal-Extraction Problem

Consider an individual who is seen in locations 1,2,...m_i, and let T_{ij} denote the number of periods in location j. If y_{ij} is the average wage residual in location j, then the estimated fixed effect is given by the following signal-extraction formula (which is derived in Appendix A):

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

where V_{ij} is the (estimated) variance of the average wage in location j:

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

and where α_i is given by

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

Given the wage history, the conditional distribution of the match component in location j is normal, with the following mean and variance:

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

$$E\left(\upsilon_{ij} \mid \{w_i\}\right) = \lambda_{ij}\left(y_{ij} - \hat{\eta}_i\right)$$

$$Var\left(\upsilon_{ij} \mid \{w_i\}\right) = \lambda_{ij}\frac{\sigma_\varepsilon^2}{T_{ij}} + \lambda_{ij}^2\alpha_i$$

where

$$\lambda_{ij} = \frac{\sigma_\upsilon^2}{V_{ij}}$$

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 υ_{ij}, given the wage history.

The optimal approximation for both of these problems is given in Kennan (2004). The solution is as follows. The wage prospects associated with a move to State k are represented by an n-point distribution with equally weighted support points $a_k(r) = \hat{\mu}_k + \hat{\upsilon}(q_r)$, $1 \le r \le n$, 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.

Let F_{ij} be the conditional distribution function of $\mu_j + \upsilon_{ij}$, given 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 same support points are the same for all individuals, but the weights on these points depend on the observed wage histories.

4.5 The Likelihood Function

The signal extraction procedure assigns probabilities to each support point of the wage distribution in each location j that is visited by individual i. These probabilities are used to construct the likelihood of the observed history for individual i, which 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

$$L = \sum_{i=1}^{N} \log\left(\sum_{\tau=1}^{K} p_\tau L_{i\tau}\right)$$

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

$$v_h\left(x, j, \theta_\tau\right) = u_h\left(x, j, \theta_\tau\right) + \beta \Sigma_{x'} p(x' | x, j)\, \bar{v}_h\left(x', \theta_\tau\right)$$

$$\bar{v}_h\left(x, \theta_\tau\right) = \mathrm{E}_\zeta \max_j \left[v_h\left(x, j, \theta_\tau\right) + \zeta_j\right]$$

Then the choice probabilities are

$$\rho_h(x, j, \theta_\tau) = \exp\left(v_h(x, j, \theta_\tau) - \overline{v_h}(x, \theta_\tau)\right)$$

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

$$L_{i\tau} = \sum_{s} q(i,s)\left(\prod_{t=1}^{T} \rho_h\left(x_i(t,s), j_i(t), \theta_\tau\right)\right)$$

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 an "amoeba" algorithm that implements the downhill simplex method of Nelder and Mead. This method does not use derivatives, and it seems appropriate for problems such as this in which there is no reason to expect that the loglikelihood function is concave. The method works well for our model; in particular, it is robust to large changes in the starting values of the parameters. On the other hand, the method is slow, and in practice Newton's method works equally well (given reasonable starting values), and much more quickly.[14]

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

[14]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 typically reached within 60 hours, on a Pentium 4 machine. An example of our (FORTRAN90) computer program can be found at www.ssc.wisc.edu/~jkennan/research/mbr21.f90.

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.[15] 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.[16] Unobserved heterogeneity in moving costs is introduced by allowing 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.[17] These estimates are interpreted in the following subsections.

---

[15]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 $\gamma_0 = \log(277700) - \log(397)$.

[16]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, but found that these had virtually no effect.

[17]The validity of the estimates is checked in Appendix B: the estimated coefficients were used to generate a simulated data set, and the maximum likelihood procedure successfully recovered these coefficients from the simulated data.

Table 2: Interstate Migration of Young White Men (high school)				
Disutility of Moving (γ_0)	7.173	3.480	4.450	4.091
	0.070	*0.532*	*0.624*	*0.628*
Distance (γ_1)		0.619	0.574	0.597
(1000 miles, State pop centroids)		*0.106*	*0.112*	*0.113*
Home Premium (κ)		0.289	0.282	0.382
		0.023	*0.021*	*0.032*
Previous Location (γ_2)		2.963	4.455	3.902
		0.231	*0.316*	*0.324*
Population (γ_3)		0.809	0.730	0.726
(10 million people)		*0.135*	*0.133*	*0.138*
Age		0.093	0.109	0.091
		0.021	*0.024*	*0.024*
Stayer Probability		0.485	0	0.436
		0.057		*0.061*
Cooling degree-days		0.114	0.100	0.149
		0.022	*0.019*	*0.024*
Heating degree-days		0.021	0.018	0.029
		0.009	*0.007*	*0.010*
"Real" Income (α)		-------	0.466	0.573
($10,000)			*0.059*	*0.077*
Loglikelihood	-1744.88	-1321.31	-1317.28	-1299.34
χ^2 (1) (Income)				43.948
Moving Cost (age 20, $2004)		-------	$269,159	$195,214
Observations	5,767			
Moves	213			

Notes:

Estimated asymptotic standard errors are given in italics below the coefficients.
The wage distributions have 3 points of support.
Population is measured in units of 10 million people.

5.1 Moving Costs and Preference Shocks

Since utility is linear in income, the estimated moving cost can be converted to a dollar equivalent. For a 20–year-old in the homogeneous model the cost is about $270,000, in 2004 dollars. The interpretation is that the compensation needed to just offset the cost of a move is large: other things equal, a lump-sum of $270,000 would be needed to fully compensate someone for all of the costs of a move (including the psychic costs).[18] Of course full compensation means indifference across all locations, implying a migration rate of almost 100% if the number of alternative locations is large. Consider instead 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%.

It may seem that the large moving cost is an artifact of the specification of the model. For example, in the absence of any moving cost, allowing preference shocks to be drawn randomly over J locations implies a migration probability of (J-1)/J, so that with J = 51, nearly everybody moves every period. But this just means that the moving cost must be large in relation to the preference shocks, with no implication for the magnitude of the cost in relation to income. The first column of Table 2 shows how large the fixed cost of moving has to be in relation to the preference shocks, in order to reduce the migration rate from 50/51 to the observed rate of 3.7%, when all other influences on migration are suppressed. The last column shows that the estimated moving cost is large in relation to the income coefficient, even after allowing for the effects of population and distance and the home premium and previous location.

To understand why the estimated moving cost is so big in relation to income, it is helpful to consider an example in which income differentials and moving costs are the only influences on migration decisions. Suppose that income in each location is either high or low, and let Δy be the difference between the high and low income levels. Suppose also that the realization of income in each location is known. Then, using equation (?), the odds of moving are given by

$$\frac{1-\lambda_L}{\lambda_L} = e^{-\gamma_0}\left[J_L - 1 + J_H e^{\beta \Delta V}\right] \tag{13}$$

$$\frac{1-\lambda_H}{\lambda_H} = e^{-\gamma_0}\left[J_H - 1 + J_L e^{-\beta \Delta V}\right] \tag{14}$$

---

[18] This refers to the cost of moving to a new location, ignoring the effect of population and distance. In the case of a return move, the estimated moving cost is $88,260. The estimated cost of moving 1,000 miles to a State with a population of 3 million is $283,447.

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

$$e^{\Delta V} = \frac{e^{\alpha \Delta y}\left[J_L + \left(J_H - 1 + e^{\gamma_0}\right)e^{\beta \Delta V}\right]}{J_L - 1 + e^{\gamma_0} + J_H e^{\beta \Delta V}} \tag{15}$$

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

These equations uniquely identify α 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).[19] 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 preference 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 preference shocks cannot be negligible, since a preference 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 (13) and (14) show that $\beta \Delta V$ determines the difference between the migration probabilities 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 γ_0 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.

This argument can be illustrated by the following examples. Suppose $J_L = J_H = 25$ and $\Delta y = \$3,000$, with $\beta = .9$. If $\lambda_L = .95$ and $\lambda_H = .97$, then $\gamma_0 = 7.14$, and $\Delta V = .58$, and the implied moving cost is $\gamma_0/\alpha = \$271,830$. On the other hand if $\lambda_L = .5$ and $\lambda_H = .99$, then $\gamma_0 = 7.78$, $\Delta V = 5.06$, and the implied moving cost is only \$19,636. We conclude that our moving cost estimate is large mainly because the empirical relationship between current income and migration probabilities is weak. For example, the

[19] It is assumed that λ_L, λ_H, J_L, J_H, Δy and β are given. Dividing (13) by (14) 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 (13) then gives a unique solution for γ_0, and inserting these solutions into equation (15) gives a unique solution for α.

migration rate is about 6% for those in the lowest wage category in our data, and about 3.3% for those in the highest wage category.

There are of course potentially important influences on migration decisions that are not included in our model, and one interpretation of the results is that, on average, the omitted variables strongly favor staying in the current location. If this is so, a more complete model might yield a smaller estimate of the moving cost. For example, there may be some components of wages that are known to the individual, but not included in the model. If the wage distribution is mis-specified in this way, some of the apparent gains available to a person with a low wage realization in the current location are illusory, and this tends to bias the estimate of α toward zero.

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. Our view is that this is likely to be a good approximation. The migration rate is low. Many moves are not motivated by wages. So selection effects are not likely to be big.

Using simulated data, we compared the distribution of wage offers with the distribution of accepted wages, and found that selection effects are indeed small. There is some bias in the estimated age coefficients in the wage equation, but it is not important. The wage distribution shifts slightly toward the upper tail as the migration process plays out, but the low wage bin still has 29% of the observations after 15 years (instead of 33%).

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 4.1% (the number of moves per person-year in the data). Table 3 shows the predictions from this model: about 72% of the people never move, and of those who do move, about 16%

move more than once.[20] The NLSY data are quite different: more than 80% never move, and about 44% 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 100 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 individual. In the case of the homogeneous model, using the parameter values in the penultimate column in Table 2, the results match the data very well: although the proportion of people who never move is slightly below the observed proportion, the proportion of movers who move more than once matches the data very well. In this respect, the observables in the model do a good job of accounting for the heterogeneous migration probabilities in the data. Oddly, allowing for unobserved heterogeneity (using the parameter values in the last column of Table 2) leads to a substantial deterioration in the ability of the model to fit this statistic.

Table 3: Goodness of Fit

Moves	**Binomial**		**NLSY**		**Homogeneous Model**		**Two-Type Model**	
None	482.8	72.6%	544	81.80%	52,948	79.62%	54,546	82.00%
One	154.4	23.2%	57	8.57%	6,907	10.39%	4,853	7.30%
More	27.80	4.2%	64	9.62%	6645	9.99%	7121	10.71%
Proportion of movers with more than one move	15.26%		52.89%		49.03%		59.47%	
Total observations	665		665		66500		66520	

Return Migration

Table 4 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 3).

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

Table 4: Return Migration Statistics	NLSY	Homogeneous Model	Two-Type Model
Proportion of Movers who			
Return home	33.8%	32.7%	33.8%
Return elsewhere	5.6%	6.8%	6.9%
Move on	60.6%	60.5%	59.3%
Proportion who *ever*			
Leave Home	15.3%	15.7%	15.0%
Move from not-home	41.7%	59.2%	42.5%
Return from not-home	23.6%	30.8%	28.3%

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 migration flow to any particular location would be roughly $\mu/(J-1)$, where μ is the average migration rate. Given $\mu = .0410$ and $J = 51$, 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 data 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. Although the two-type model does a much better job of fitting this aspect of the data, this improvement seems to distort the statistics on the number of moves per person, as was seen in Table 3.

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 5 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 income gain is realized. Moreover, since the worklife is finite, a move that is worthwhile for a younger 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.

There are two effects here, and both are included in our model. 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.[21] The human capital explanation also says that migration rates decline with age because the 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. This suggests that the first part of the human capital explanation must be the dominant force explaining why migration rates for 30-year-olds are substantially lower than for 25-year-olds.

The results in Table 2 indicate that the human capital model does not fully explain the relationship between age and migration. The model allows for the possibility that age has a direct effect on the cost of migration, and this effect is indeed found in the data.[22]

---

[21]One way to see this is to consider the extreme case in which there are no preference 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.

[22]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).

Table 5 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.

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 6. 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.

Table 6: Alternative Income Specifications			
	Census	State Means	National
Disutility of Moving	4.091	3.513	4.303
	0.628	*0.533*	*0.657*
Distance	0.597	0.587	0.632
(1000 miles)	*0.113*	*0.107*	*0.114*
Home Premium	0.382	0.294	0.384
	0.032	*0.024*	*0.032*
Previous Location	3.902	3.049	4.232
(moving cost)	*0.324*	*0.232*	*0.396*
Population	0.726	0.764	0.785
	0.138	*0.135*	*0.138*
Age	0.091	0.093	0.095
	0.024	*0.021*	*0.024*
Stayer Probability	0.436	0.478	0.411
	0.061	*0.057*	*0.069*
Cooling degree-days	0.149	0.136	0.108
	0.024	*0.024*	*0.023*
Heating degree-days	0.029	0.027	0.017
	0.010	*0.009*	*0.010*
Real Income	0.573	0.480	0.683
(ACCRA)	*0.077*	*0.105*	*0.089*
Loglikelihood	-1299.34	-1313.465	-1303.864
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 7 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 fourth column of Table 7: the estimate of α is reduced by about 20%, with little 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 have much effect on the empirical results: either way, income significantly affects migration decisions.

The other alternative specifications in Table 7 are concerned with sensitivity of the estimates to the discount factor (β). Reducing β to .90 has a noticeable effect on the utility flow parameters (i.e. the home premium and the income coefficient), 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.

Table 7: Alternative Specifications				
	Baseline	No Cola	β = .9	β = .975
Disutility of Moving	4.091	3.716	3.826	4.177
	0.628	*0.583*	*0.632*	*0.624*
Distance	0.597	0.653	0.650	0.5739
(1000 miles)	*0.113*	*0.111*	*0.119*	*0.110*
Home Premium	0.382	0.369	0.527	0.306
	0.032	*0.032*	*0.039*	*0.030*
Previous Location	3.902	0.697	0.757	0.660
(moving cost)	*0.324*	*0.136*	*0.143*	*0.137*
Population	0.726	0.089	0.099	0.087
	0.138	*0.022*	*0.025*	*0.023*
Age	0.091	0.162	0.205	0.124
	0.024	*0.023*	*0.035*	*0.020*
Stayer Probability	0.436	3.402	3.797	3.958
	0.061	*0.285*	*0.313*	*0.314*
Cooling degree-days	0.149	0.030	0.038	0.022
	0.024	*0.008*	*0.014*	*0.008*
Heating degree-days	0.029	0.473	0.400	0.500
	0.010	*0.054*	*0.065*	*0.057*
Real Income	0.573	0.378	0.805	0.431
(ACCRA)	*0.077*	*0.064*	*0.113*	*0.061*
Loglikelihood	-1299.34	-1308.15	-1298.99	-1301.44
N (person-years)	5,767			
Moves	213			

Notes:
The estimates in columns 2-4 are not quite final.

6 Spatial Labor Supply Elasticities

We use the estimated model to analyze short-run responses to local labor demand shocks, modeled as 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.

We take a set of 10,000 people, with 100 replicas of each person, distributed over States as in the Census data, 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, in 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.

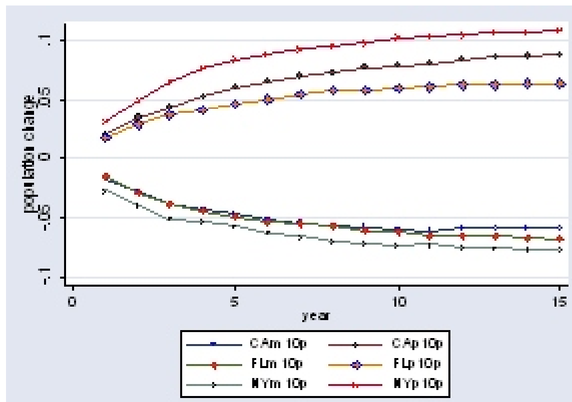

Figure XX shows the results for California, Florida and New York, using a model with no climate variables included. The supply elasticities are large: between 0.6 and 1.1. Somewhat surprisingly, they are not symmetric, and there are noticeable differences across States. The adjustment is gradual, but it is largely completed in 10 years.

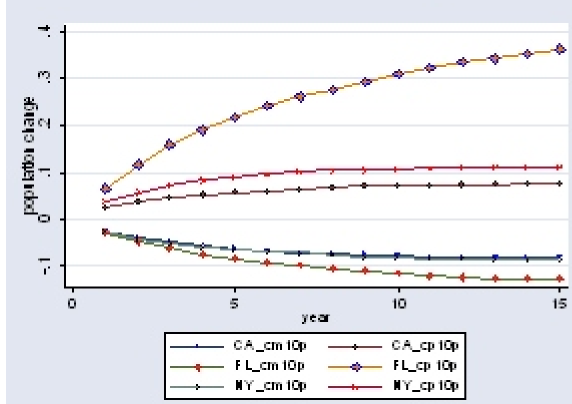

Figure YY shows the effects of introducing climate variables. This does not make much difference for California or New York, but the results for Florida change dramatically, especially for the case of a wage increase. The reason for this is not clear. One possibility is that high wages in Florida, in combination with a highly desirable climate, have the effect of making Florida the destination of choice for people who draw a large adverse shock in their current location. It may be that at the wage levels observed in the data, Florida is close to this level, and is attracting more than its share of migrants; a 10% wage increase then makes Florida really stand out, while a 10% decrease means that Florida sinks back into line with other destinations.

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 preference 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. 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. It would be very useful to extend the model to include marital status as a state variable, and this is probably feasible. 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.

References

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.

Gelbach, Jonah B. (2002), "Migration, the Lifecycle, and State Benefits: How low is the bottom?" University of Maryland, http://www.glue.umd.edu/~gelbach/papers/index.html.

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.

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.

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

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 α, where

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

and

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

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 correlation $E\hat{\eta}y_j$ for an arbitrary choice of j, say j = 1:

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

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

The definition of α implies that

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

Thus

$$E\hat{\eta}y_1 = \alpha\left(\frac{\sigma_y^2 - \sigma_\eta^2}{V_1}\right) + \sigma_\eta^2 - \alpha$$

$$= \sigma_\eta^2$$

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

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

Then

$$E(\eta - \hat{\eta})^2 = \frac{\alpha^2}{\sigma_\eta^2} + \alpha^2 \sum_{j=1}^{m} \frac{1}{V_j}$$

$$= \frac{\alpha^2}{\sigma_\eta^2} + \alpha^2\left(\frac{1}{\alpha} - \frac{1}{\sigma_\eta^2}\right) = \alpha$$

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, repectively. Then the expectation of υ_j, conditional on $\{y_k\}$, is given by

$$\hat{\upsilon}_j = \frac{\sigma_\upsilon^2}{V_j}\left(y_j - \frac{\sum_{k=1}^{m} \frac{y_k}{V_k}}{\frac{1}{\sigma_\eta^2} + \sum_{k=1}^{m} \frac{1}{V_k}}\right)$$

where $V_j = \sigma_\upsilon^2 + \tau_j$.

Proof

We will show that $E(\upsilon_j-\hat{\upsilon}_j)\, y_k = 0$, for all j and k. First, $E\, \upsilon_j\, y_k = 0$ for $k \neq j$, and $E\, \upsilon_j\, y_j = \sigma_\upsilon^2$. Next,

$$Ey_k \hat{\upsilon}_j = \frac{\sigma_\upsilon^2}{V_j}\left(Ey_k y_j - Ey_k \hat{\eta}\right)$$

Using the argument of the previous lemma, with $u_j = \upsilon_j + \varepsilon_j$, we have $Ey_k \hat{\eta} = \sigma_\eta^2$. Thus $E\, \hat{\upsilon}_j\, y_k = 0$ for $k \neq j$, and $E\, \hat{\upsilon}_j\, y_j = \sigma_\upsilon^2$.