The Effect of Expected Income on Individual Migration Decisions

John Kennan and James R. Walker*
University of Wisconsin and NBER

February 2009

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 Introduction

There is an extensive literature on migration.[1] 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

*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 also thank the editor and referees for very detailed constructive criticism of several earlier versions of the paper. We are grateful to Joe Altonji, Kate Antonovics, Peter Arcidiacono, Gadi Barlevy. Philip Haile, Igal Hendel, Mike Keane, Derek Neal, John Pencavel, Karl Scholz, Robert Shimer, Chris Taber, 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]See Greenwood (1997) and Lucas (1997) for surveys.

are 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.[2] 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 and amenity values. We also allow for a location match component of preferences that is revealed to the individual for each location that is visited.

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 each location). We also allow for a bias in favor of the home location (measured as the State of residence at age 14). Age is included as a state variable, entering through the moving cost, with the idea that if the simplest

[2] 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. Gemici (2008) extends our framework and considers family migration decisions, but defines locations as census regions.

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.

Our main substantive conclusion is that interstate migration decisions are indeed influenced to a substantial extent by income prospects. There is evidence of a response to geographic differences in mean wages, as well as 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.[3]

2 Migration Dynamics

The need for a dynamic analysis of migration is illustrated in Table 1, which summarizes 10-year interstate migration histories for the cross-section sample of the NLSY, beginning at age 18. 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.

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 reflects the idea that people may be more productive in some locations than in others,

[3]See for example Kennan and Walker (2001) and Gemici (2008).

Table 1: **Interstate Migration**, NLSY 1979-94

	Less than High School	**High School**	**Some College**	**College**	**Total**
Number of people	322	919	758	685	2684
Movers (age 18-27)	80	223	224	341	868
Movers (%)	24.8%	24.3%	29.6%	49.8%	32.3%
Moves Per Mover	2.10	1.95	1.90	2.02	1.98
Repeat moves (% of all moves)	52.4%	48.7%	47.4%	50.5%	49.5%
Return Migration (% of all moves)					
Return - Home	32.7%	33.1%	29.1%	23.2%	28.1%
Return - not home	15.5%	7.1%	6.8%	8.6%	8.4%
Movers who return home	61.3%	56.5%	51.3%	42.8%	50.2%

The sample includes respondents from the cross-section sample of the NLSY79 who were continuously interviewed from ages 18 to 28, and who never served in the military. The home location is the State of residence at age 14.

depending on working conditions, residential conditions, local amenities and so forth. Although some information about 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.[4] Although there are 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 approximation in which workers take current wage levels as an estimate of the wages they will face for the foreseeable future. In any

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

[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 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).

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 they 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 and preference distributions, 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 wage and preference information, current location and age, as discussed below). The utility flow for someone who chooses location j is specified as $u(x,j)+\zeta_j$, 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

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

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 β is the discount factor, and E_ζ denotes the expectation with respect to the distribution of the J-vector ζ with components ζ_j. We assume that ζ_j is

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

drawn from the Type I extreme value distribution. In this case, using arguments due to McFadden (1973) and Rust (1987), we have

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

where $\bar{\gamma}$ is the Euler constant. Let $\rho\left(x,j\right)$ be the probability of choosing location j, when the state is x. Then

$$\rho\left(x,j\right) = \exp\left(v\left(x,j\right) - \bar{v}\left(x\right)\right)$$

We compute v by value function iteration, assuming a finite horizon, T. We include age as a state variable, with $v \equiv 0$ at age $T+1$, so that successive iterations yield the value functions for a person who is getting younger and younger.

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, but we obviously cannot let J be the number of labor markets; for example, there are over 3,100 counties in the U.S. Indeed, even if J is the number of States, the model is computationally infeasible,[7] 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

[7] And it will remain so: for example, if there are 50 locations, and the wage distribution has 5 support points, then the number of dynamic programming states is 40,414,063,873,238,203,032,156,980,022,826,814,668,800.

fully informed about wages at all locations. Then if the distributions of location match wage and preference components in each of J locations have n points of support, the number of states for someone seen in M locations is $J(Jn^2)^M$, the number of possible M-period histories describing the locations visited most recently, and the wage and preference components found there. For example, if J is 50 and n is 3 and M is 2, the number of states at each age is 10,125,000, 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.

Decision problems with large state spaces can alternatively be analyzed by computing the value function at a finite set of points, and interpolating the function for points outside this set, as suggested by Keane and Wolpin (1994).[9] In our context this would not be feasible without some simplification of the state space, because of the spatial structure of the states. Since each location has its own unique characteristics, interpolation can be done only within locations, and this means that the set of points used to anchor the interpolation must include several alternative realizations of the location match components for each location; allowing for n alternatives yields a set of n^J points, which is too big when $J = 50$ (even if n is small). On the other hand it is worth noting that our limited history approximation works only because we have discretized the state space. If the location match components are drawn from continuous distributions, the state space is still infinite even when the history is limited (although interpolation methods could be used in that case).

4.2 Wages

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

$$w_{ij}(a) = \mu_j + \upsilon_{ij} + G(X_i, a, t) + \eta_i + \varepsilon_{ij}(a)$$

[8] 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 abandoned, 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).

[9] For example, this method was used by Erdem and Keane (1994) to analyze the demand for liquid laundry detergent, and by Crawford and Shum (2005) to analyze the demand for pharmaceuticals. In these applications, the agents in the model do not know the flow payoffs from the various available choices until they have tried them, just as our agents do not know the location match components until they have visited the location.

where μ_j is the mean wage in location j, υ is a permanent location match effect, $G(X, a, t)$ represents a (linear) time effect and the effects of observed individual characteristics, 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 random variables that are identically distributed across individuals and locations. We also assume that the realizations of η and υ are seen by the individual.[10]

The relationship between wages and migration decisions 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 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.

4.3 State Variables and Flow Payoffs

Let $\ell = \left(\ell^0, \ell^1, \ldots, \ell^{M-1}\right)$ be an M-vector containing the sequence of recent locations (beginning with the current location), and let ω be an M-vector recording wage and utility information at these locations. The state vector x consists of ℓ, ω and age. The flow payoff for someone whose "home" location is h is specified as

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

where

$$u_h(x, j) = \alpha_0 w\left(\ell^0, \omega\right) + \sum_{k=1}^{K} \alpha_k Y_k\left(\ell^0\right) + \alpha^H \chi\left(\ell^0 = h\right) + \xi\left(\ell^0, \omega\right) - \Delta_\tau(x, j)$$

Here the first term refers to wage income in the current location. This is augmented by the nonpecuniary variables $Y_k\left(\ell^0\right)$, representing amenity values. The parameter α^H is a premium that allows each individual to have a preference for their native location (χ_A denotes an indicator meaning that A is true). The flow payoff in each location has a random permanent component ξ; the realization of this component is learned only when the location is visited. This location match component of preferences is analogous to the match component of wages (υ), except that ξ can only be inferred from observed migration choices, whereas both migration choices and wages are informative about υ. The cost of moving

[10] 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.

from ℓ^0 to ℓ^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.

4.4 Moving Costs

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

$$\Delta_\tau(x,j)=\left(\gamma_{0\tau}+\gamma_1 D\left(\ell^0,j\right)-\gamma_2\chi\left(j\in\mathbb{A}\left(\ell^0\right)\right)-\gamma_3\chi\left(j=\ell^1\right)+\gamma_4 a-\gamma_5 n_j\right)\chi\left(j\neq\ell^0\right)$$

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. For 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.[11] 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.

4.5 Transition Probabilities

The state vector can be written as $x=(\tilde{x},a)$, where $\tilde{x}=\left(\ell^0,\ell^1,x_\upsilon^0,x_\upsilon^1,x_\xi^0,x_\xi^1\right)$ and where x_υ^0 indexes the realization of the location match component of wages

[11]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:

$$\rho(x,j)=\frac{\xi_j}{\sum_{k=1}^J\xi_k}\ ;\ \xi_k=(1+\psi n_k)\exp\left(\upsilon_k(\ell,\omega)\right)$$

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

in the current location, and similarly for the other components. The transition probabilities are as follows

$$p(x' \mid x, j) = \begin{cases} 1 & \text{if} \quad j = \ell^0, \quad \tilde{x}' = \tilde{x}, \quad a' = a+1 \\ 1 & \text{if} \quad j = \ell^1, \quad \tilde{x}' = \left(\ell^1, \ell^0, x_\upsilon^1, x_\upsilon^0, x_\xi^1, x_\xi^0\right), \quad a' = a+1 \\ \frac{1}{n^2} & \text{if} \quad j \notin \left\{\ell^0, \ell^1\right\}, \quad \tilde{x}' = (j, \ell^0, s_\upsilon, x_\upsilon^0, s_\xi, x_\xi^0), \\ & \qquad (1,1) \le (s_\upsilon, s_\xi) \le (n_\upsilon, n_\xi), \quad a' = a+1 \\ 0 & \text{otherwise} \end{cases}$$

This covers several cases. First, if no migration occurs this period, then the state remains the same except for the age component. If there is a move to a previous location, the current and previous locations are interchanged. And if there is a move to a new location, the current location becomes the previous location, and the new location match components are drawn at random. In all cases, age is incremented by one period.

4.6 Identification

We can identify the influence of wages on migration decisions in our model using panel data on individuals who start out in locations with different mean wage levels, or who have different realizations from the distribution of match components in their initial location. Even if wages are irrelevant (i.e. if $\alpha_0 = 0$), the model predicts that people with unfavorable realizations of the location match preference component ξ in the initial location would be more likely to migrate. But as long as we have data on the match component of wages, we can distinguish the effect of wages from the effect of locational preferences. We illustrate this using a very simple case of the model.[12]

Suppose there are just two locations, and the payoff in location j is $\alpha_0(\mu_j + \upsilon_j) + \xi_j$. Each person knows $\{\mu_j\}$, and knows the realizations of υ and ξ in the initial location, and the distributions F_υ and F_ξ from which the match components are drawn. Assume for simplicity that only one move is possible, and that it must be made in the initial period. Then if υ and ξ have zero means, and if decisions are made to maximize expected income, the probability that someone who is born in location 1 would move to location 2 is

$$Pr\left(\alpha_0\left(\mu_1 + \upsilon_1\right) + \xi_1 + \delta < \alpha_0 \mu_2\right) = F_\xi\left(\alpha_0\left(\mu_2 - \mu_1 - \upsilon_1\right) - \delta\right)$$

[12] We do not directly see the location match component of wages, but this component can be identified by comparing the variability of wages within and across locations for each individual. One simple way to do this is as follows. Let y_i be a vector containing the wage history for individual i. For each individual history 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 A_3 is a consistent estimator for σ_η^2, $A_2 - A_3$ is a consistent estimator for σ_υ^2, and $A_1 - A_2$ is a consistent estimator for σ_ε^2. Given these variance estimates, the location match component can be estimated by solving a standard signal extraction problem. Although this heuristic method gives a transparent account of how the wage components can be extracted, in practice we estimate all of the wage parameters jointly as part of the maximum likelihood procedure, as described below.

(where $\delta = \frac{1-\beta}{1-\beta^T}\Delta$ is the flow payment that amortizes the moving cost Δ).

The data reveal the proportion of people who move at each level of wages, and the relationship between wages and migration can thus be used to predict the effect of a change in wage levels. For example, if μ_1 changes to μ_1', a person with wage w will behave in the same way as a person with wage $w + \mu_1' - \mu_1$ behaved before the change.

Here we are assuming that μ_j and ξ_j are independently distributed, and also that ξ is independently distributed across individuals. Alternatively, we can allow for unobserved amenities, represented by a component of the location match preferences that is common to all individuals. In this case the effect of differences in mean wages across locations is not identified, but the effect of income differences is nevertheless identified through the location match component of wages. The identification argument is given in detail in Appendix A.

Our basic empirical results use variation in both μ and υ to identify the effect of income differences. In the context of an equilibrium model of wage determination, this can be justified by assuming constant returns to labor in each location, so that wage differences across locations are determined entirely by productivity differences, and are thus independent of differences in amenity values. Clearly, this is a strong assumption. Accordingly, we also present estimates that control for regional differences in unobserved amenity values.

4.7 Data

Our primary data source is the National Longitudinal Survey of Youth 1979 Cohort (NLSY79); we also use data from the 1990 Census. The NLSY79 conducted annual interviews from 1979 through 1994, and changed to a biennial schedule in 1994. The location of each respondent is recorded at the date of each interview, and we measure migration by the change in location from one interview to the next. We use information from 1979 to 1994 so as to avoid the complications arising from the change in the frequency of interviews.

In order to obtain a relatively homogeneous sample, we consider only white non-Hispanic high-school graduates with no post-secondary education, using only the years after schooling is completed.[13] Appendix B describes our selection procedures. The NLSY over-samples 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. Thus we use only the "cross-section" subsample, with the poverty subsample excluded. The sample includes only people who completed high school by age 20, and who never enrolled in college. We exclude those who ever served in the military and also those who report being out of the labor force for more than one year after

[13] 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."

age 20. We follow each respondent from age 20 to the 1994 interview or the first year in which some relevant information is missing or inconsistent.

Our analysis sample contains 432 people, with continuous histories from age 20 comprising 4,274 person-years. There are 124 interstate moves (2.9 percent per annum).

In each round of the NLSY79, respondents report income for the most recent calendar year. Wages are measured as total wage and salary income, plus farm and business income, adjusted for cost of living differences across States (using the ACCRA Cost of Living Index). We exclude observations with positive hours or weeks worked and zero income.

We use information from the Public Use Micro Sample from the 1990 Census to estimate State mean effects (μ_j), since the NLSY does not have enough observations for this purpose. From the PUMS we select white high-school men aged 19-20 (so as to avoid selection effects due to migration)[14]. We estimate State mean wage effects using a median regression with age and State dummies.[15] We condition on these estimated State means in the maximum likelihood procedure that jointly estimates the remaining parameters of the wage process and the utility and cost parameters governing migration decisions.

5 Estimation

In this section we discuss the specification and computation of the likelihood function.

5.1 Discrete Approximation of the Distribution of Location Match Effects

We approximate the decision problem by using discrete distributions to represent the distributions of the location match components, and computing continuation values at the support points of these distributions. We first describe this approximation, and then describe the specification of the other components of wages.

For given support points, the best discrete approximation $\hat{F}$ for any distribution F assigns probabilities so as to equate $\hat{F}$ with the average value of F over each interval where is constant. If the support points are variable, they are chosen so

[14]The parameters governing migration decisions and the parameters of the wage process are estimated jointly to account for selection effects due to migration (although in practice these effects are empirically negligible). The State mean effects are specified as age-invariant and are estimated using wages observed at the beginning of the worklife, to minimize the potential effects of selection. We include observations for 19 year olds from the PUMS to increase the precision of the estimated State means.

[15]We measure wages as annual earnings and exclude individuals with retirement income, social security income or public assistance; we also exclude observations if earnings are zero despite positive hours or weeks worked.

that $\hat{F}$ assigns equal probability to each point.[16] Thus if the distribution of the location match component υ were known, the wage prospects associated with a move to State k could be represented by an n-point distribution with equally weighted support points $\hat{\mu}_k + \hat{\upsilon}(q_r), 1 \leq r \leq n$, where $\hat{\upsilon}(q_r)$ is the q_r quantile of the distribution of υ, with

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

for $1 \leq r \leq n$. The distribution of υ is in fact not known, but we assume that it is symmetric around zero. Thus for example with $n = 3$, the distribution of $\mu_j + \upsilon_{ij}$ in each State is approximated by a distribution that puts mass $\frac{1}{3}$ on μ_j (the median of the distribution of $\mu_j + \upsilon_{ij}$), with mass $\frac{1}{3}$ on $\mu_j \pm \tau_\upsilon$, where τ_υ is a parameter to be estimated. The location match component of preferences is handled in a similar way.

5.2 Fixed Effects and Transient Wage Components

Even though our sample is quite homogeneous, measured earnings in the NLSY are highly variable, both across people and over time. Moreover, the variability of earnings over time is itself quite variable across individuals. Our aim is to specify a wage components model that is flexible enough to fit these data, so that we can draw reasonable inferences about the relationship between measured earnings and the realized values of the location match component. For the fixed effect η, we use a (uniform) discrete distribution that is symmetric around zero, with 7 points of support, so that there are three parameters to be estimated. For the transient component ε we need a continuous distribution that is flexible enough to account for the observed variability of earnings. We assume that ε is drawn from a normal distribution with zero mean for each person, but we allow the variance to vary across people. Specifically, person i initially draws $\sigma_\varepsilon(i)$ from some distribution, and subsequently draws ε_{it} from a normal distribution with mean zero and standard deviation $\sigma_\varepsilon(i)$, with ε_{it} drawn independently in each period. The distribution from which σ_ε is drawn is specified as a (uniform) discrete distribution with four support points, where these support points are parameters to be estimated.

5.3 The Likelihood Function

The likelihood of the observed history for each individual is a mixture over heterogeneous types. Let $L_i(\theta_\tau)$ be the likelihood for individual i, where θ_τ is the parameter vector, for someone of type τ, and let π_τ be the probability of type τ. The sample loglikelihood is

$$\Lambda(\theta) = \sum_{i=1}^{N} \log\left(\sum_{\tau=1}^{K} \pi_\tau L_i(\theta_\tau)\right)$$

[16] See Kennan (2004).

For each period of an individual history two pieces of information contribute to the likelihood: the observed income, and the location choice. Each piece involves a mixture over the possible realizations of the various unobserved components. In each location there is a draw from the distribution of location match wage components, which is modeled as a uniform distribution over the finite set $\Upsilon = \{\upsilon(1), \upsilon(2), \ldots, \upsilon(n_\upsilon)\}$. We index this set by ω_υ, with $\omega_\upsilon(j)$ representing the match component in location j, where $1 \leq \omega_\upsilon(j) \leq n_\upsilon$. Similarly, in each location there is a draw from the location match preference distribution, which is modeled as a uniform distribution over the finite set $\Xi = \{\xi(1), \xi(2), \ldots, \xi(n_\xi)\}$, indexed by ω_ξ. Each individual also draws from the distribution of fixed effects, which is modeled as a uniform distribution over the finite set $H = \{\eta(1), \eta(2), \ldots, \eta(n_\eta)\}$, and we use ω_η to represent the outcome of this. And each individual draws a transient variance, from a uniform distribution over the set $\varsigma = \{\sigma_\varepsilon(1), \sigma_\varepsilon(2), \ldots, \sigma_\varepsilon(n_\varepsilon)\}$, with the outcome indexed by ω_ε.

The unobserved components for individual i are then represented by a vector ω^i with $N_i + 3$ elements: $\omega^i = \left\{\omega^i_\xi, \omega^i_\eta, \omega^i_\varepsilon, \omega^i_\upsilon(1), \omega^i_\upsilon(2), \ldots, \omega^i_\upsilon(n_\eta)\right\}$, where N_i is the number of locations visited by this individual. The set of possible realizations of ω^i is denoted by $\Omega(N_i)$; there are $n_\xi n_\eta n_\varepsilon (n_\upsilon)^{N_i}$ points in this set, and our discrete approximation implies that they are equally likely. We index the locations visited by individual i in the order in which they appear, and we use the notation κ^0_{it} and κ^1_{it} to represent the position of the current and previous locations in this index. Thus $\kappa_{it} = \left(\kappa^0_{it}, \kappa^1_{it}\right)$ is a pair of integers between 1 and N_i. For example, in the case of someone who never moves, κ^0_{it} is always 1, and κ^1_{it} is zero (by convention), while for someone who has just moved for the first time, $\kappa_{it} = (2, 1)$.

Let $\lambda_{it}\left(\omega^i, \theta_\tau\right)$ be the likelihood of the destination chosen by person i in period t. Recall that $\rho(x, j)$ is the probability of choosing location j, when the state is x. Then

$$\lambda_{it}\left(\omega^i, \theta_\tau\right) = \rho_{h(i)}\left(\ell(i,t), \omega^i_\upsilon\left(\kappa^0_{it}\right), \omega^i_\upsilon\left(\kappa^1_{it}\right), \omega^i_\xi\left(\kappa^0_{it}\right), \omega^i_\xi\left(\kappa^1_{it}\right), a_{it}, \ell^0(i, t+1), \theta_\tau\right)$$

Here the probability that i chooses the next observed location, $\ell^0(i, t+1)$, depends on the current and previous locations, the values of the location match components at those locations, the individual's home location $h(i)$, and the individual's current age. The parameter vector θ_τ includes the unknown coefficients in the flow payoff function and the support points in the sets Υ, H, Ξ , and ς.

Let $\psi_{it}(\omega_i, \theta)$ be the likelihood of the observed income for person i in period t. Then

$$\psi_{it}(\omega_i, \theta) = \phi\left(\frac{w_{it} - \mu_{\ell^0(i,t)} - G(X_i, a_{it}, \theta) - \upsilon\left(\omega^i_\upsilon\left(\kappa^0_{it}\right)\right) - \eta\left(\omega^i_\eta\right)}{\sigma_\varepsilon\left(\omega^i_\varepsilon\right)}\right)$$

where ϕ is the standard normal density function.

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

$$L_i(\theta_\tau) = \frac{1}{n_\eta n_\varepsilon n_\xi (n_\upsilon)^{N_i}} \sum_{\omega^i \in \Omega(N_i)} \left(\prod_{t=1}^{T_i} \psi_{it}(\omega^i, \theta_\tau) \lambda_{it}(\omega^i, \theta_\tau) \right)$$

5.4 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 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]

6 Empirical Results

Our basic results are shown in Table 2. We set $\beta = .95, T = 40$, and $M = 2$; we show below that our main results are not very sensitive to changes in the discount factor or the horizon length.[18] The table gives estimated coefficients and standard errors for four versions of the model that highlight both the effect of income on migration decisions and the relevance of the location match component of preferences. Unobserved heterogeneity in moving costs is introduced by allowing for two types, one of which is 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.

We find that distance, home and previous locations and population size all have highly significant effects on migration. Age and local climate (represented by the annual number of cooling degree-days) are also significant.[19] Our main

[17]Given reasonable starting values (for example, 50% type probabilities and a fixed cost for the mover type that roughly matches the average migration rate, with a unit variance for the transient component of wages, and all other parameters set to zero), the maximal likelihood is reached by Newton's method within a day or two, on a cluster of parallel CPUs, with one CPU per home location; each likelihood evaluation requires about 24 seconds. We found the Newton procedure to be well-behaved in the sense that it almost always reached the same answer no matter what starting values were used: we have estimated hundreds of different versions of the model, and found very few local maxima; even in these cases the likelihood 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/mbr87.f90.

[18]The validity of the estimates is checked in Appendix C: the estimated coefficients were used to simulate 100 replicas of each person in the data, and the maximum likelihood procedure was applied to the simulated data. The null hypothesis that the data were generated by the true DGP is accepted by a likelihood ratio test.

[19]The "cooling" variable is the population-weighted annual average number of cooling degree days (in thousands) for 1931-2000, taken from Historical Climatography Series 5-2 (Cooling Degree Days) – see US NCDC (2002). For example, the cooling degree-day variable for Florida is 3.356, meaning that the difference between 65° and the mean daily temperature in Florida, summed over the days when the mean was above 65°, averaged 3,356 degree-days per year

finding is that, controlling for these effects, migration decisions are significantly affected by expected income changes. This holds regardless of whether the location match component of preferences is included in the specification. Since the estimated effect of this component is negligible, and it enlarges the state space by a factor of about 100, we treat the specification that excludes this component as the base model in the subsequent discussion.

6.1 Wages

The estimated parameters of the wage process are summarized in Table 3, showing the magnitudes of the various components in 2008 dollars. As was mentioned above, there is a great deal of unexplained variation in wages, across people, and over time for the same person; moreover there are big differences in the variability of earnings over time from one individual to the next.[20]

The wage components that are relevant for migration decisions in the model are also quite variable, suggesting that migration incentives are strong. For example, the 90-10 differential across State means is about \$4,500 a year, and the value of replacing a bad location match draw with a good draw is about \$16,000 a year.

6.2 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 4.

For the average mover, the cost is about \$301,000 (in 2008 dollars), if the payoff shocks are ignored. 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. According to the estimates in Table 3, a move away from a bad location match would increase income by \$8,117, on average, and a move from the bottom to the top of the distribution of State means would increase income by \$9,212. A move that makes both of these changes would mean a permanent wage increase of \$17,329, or \$302,040 in present value (assuming a remaining worklife of 40 years, with $\beta = .95$). The home premium is equivalent to a wage increase of \$22,333, and the cost of moving to a previous location is relatively low. Thus in some cases the expected income gains would

(over the years 1931-2000).

We explored various alternative specifications of the climate amenity variables. Including heating degree-days had little effect on the results (see Table 10 below). 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.

[20] As indicated in Table 2, the individual characteristics affecting wages include age, AFQT score, and an interaction between the two. The interaction effect is included to allow for the possibility that the relationship between AFQT scores and wages is stronger for older workers, either because ability and experience are complementary, or because employers gradually learn about ability, as argued by Altonji and Pierret (2001).

Table 2: **Interstate Migration, Young White Men**

	$\hat{\theta}$	$\hat{\sigma}_\theta$	$\hat{\theta}$	$\hat{\sigma}_\theta$	$\hat{\theta}$	$\hat{\sigma}_\theta$	$\hat{\theta}$	$\hat{\sigma}_\theta$
Utility and Cost								
Disutility of Moving (γ_0)	4.794	0.565	4.513	0.523	4.864	0.601	4.863	0.606
Distance (γ_1) (1000 miles)	0.267	0.181	0.280	0.178	0.312	0.187	0.273	0.184
Adjacent Location (γ_2)	0.807	0.214	0.786	0.211	0.772	0.220	0.802	0.216
Home Premium (α^H)	0.331	0.041	0.267	0.031	0.332	0.048	0.340	0.045
Previous Location (γ_3)	2.757	0.357	2.542	0.300	3.080	0.449	2.826	0.417
Age (γ_4)	0.055	0.020	0.061	0.019	0.060	0.020	0.054	0.020
Population (γ_5) (millions)	0.654	0.179	0.652	0.179	0.637	0.177	0.651	0.179
Stayer Probability	0.510	0.078	0.520	0.079	0.495	0.087	0.508	0.081
Cooling (α_1) (1000 degree-days)	0.055	0.019	0.036	0.019	0.048	0.018	0.056	0.019
Income (α_0)	0.314	0.100	——		——		0.300	0.117
Location Match Preference (τ_ξ)	——		——		0.168	0.049	0.074	0.094
Wages								
Wage intercept	-5.133	0.245	-5.142	0.248	-5.143	0.248	-5.139	0.246
Time trend	-0.034	0.008	-0.032	0.008	-0.032	0.008	-0.034	0.008
Age effect (linear)	7.841	0.356	7.850	0.358	7.851	0.358	7.849	0.356
Age effect (quadratic)	-2.362	0.129	-2.377	0.129	-2.378	0.130	-2.365	0.129
Ability (AFQT)	0.011	0.065	0.012	0.066	0.012	0.066	0.012	0.065
Interaction(Age,AFQT)	0.144	0.040	0.150	0.040	0.150	0.040	0.144	0.040
Transient s.d. 1	0.217	0.007	0.218	0.007	0.218	0.007	0.217	0.007
Transient s.d. 2	0.375	0.015	0.375	0.015	0.375	0.015	0.375	0.015
Transient s.d. 3	0.546	0.017	0.547	0.017	0.547	0.017	0.546	0.017
Transient s.d. 4	1.306	0.028	1.307	0.028	1.307	0.028	1.306	0.028
Fixed Effect 1	0.113	0.036	0.112	0.035	0.112	0.035	0.112	0.036
Fixed Effect 2	0.296	0.035	0.293	0.036	0.293	0.036	0.295	0.035
Fixed Effect 3	0.933	0.016	0.931	0.017	0.931	0.017	0.933	0.017
Wage match (τ_v)	0.384	0.017	0.387	0.018	0.387	0.018	0.385	0.018
Loglikelihood	-4214.160		-4220.775		-4218.146		-4214.105	
Exclude Income: $\chi^2(1)$			13.22		8.08			
Exclude match preference: $\chi^2(1)$	0.12		5.26					
There are 4274 (person-year) observations, 432 individuals, and 124 moves.								

Table 3: **Wage Parameter Estimates** ($2008)

	AFQT percentile						
Average Wages	25			50	75		
age 20 in 1979	25,055			26,699	28,343		
age 20 in 1989	17,919			19,563	21,207		
age 30 in 1989	39,153			41,569	43,984		
	low			middle			high
Location match	-8,117			0			8,117
Fixed effect support	-19,732	-6,260	-2,390	0	2,390	6,260	19,732
State Means	low (WV)	rank 5 (OK)		Median (MO)	rank 45 (RI)		high (MD)
	12,179	13,943		16,291	18,488		21,391

Table 4: **Moving Cost Examples**

	γ_0	α_0	Age	Distance	Adjacent	Population	Previous	Cost
θ	4.794	0.314	0.055	0.267	0.807	0.654	2.757	
Young Mover			20	1	0	1	0	$371,046
Average Mover			24.355	0.664	0.427	0.727	0.371	$300,997

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 4 do not refer to the costs of moves that are actually made, but rather to the costs of hypothetical moves to arbitrary locations. 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 4. 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 $\gamma + log(J-1)$ (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 $log(J-1)/\alpha_0$. Using the estimated income coefficient, this is a reduction of $262,281. 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 unfavorable, 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 4. In section 6.3 below we analyze the average costs of moves that are actually made, allowing for the effects of the payoff shocks.

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_0 falls by $10,000, with the other parameters held fixed. We estimate that such a subsidy would lead to a substantial increase in the interstate migration rate: from 2.9% to about 4.9%.

6.2.1 Moving Costs and Payoff Shocks: 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 (5), the odds of moving are given by

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

$$\frac{1-\lambda_H}{\lambda_H} = \exp(-\gamma_0)\left[J_H - 1 + J_L e^{-\beta\Delta V}\right] \quad (2)$$

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 whereΔV is the difference in expected continuation values between the low-income and high-income locations. This difference is determined by the equation

$$e^{\Delta V} = \frac{e^{\alpha_0 \Delta y}\left(J_L + (J_H - 1 + e^{\gamma_0})\, e^{\beta\Delta V}\right)}{J_L - 1 + e^{\gamma_0} + J_H e^{\beta\Delta V}} \quad (3)$$

For example, if $\beta = 0$, then $\Delta V = \alpha_0 \Delta y$, while if moving costs are prohibitive ($e^{-\gamma_0} = 0$), then $\Delta V = \frac{\alpha_0 \Delta y}{1-\beta}$.

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 difference is $2\beta\Delta V$, and equations (1) and (2) show that $\beta\Delta V$ determines the relative odds of moving from low-income and high-income locations. Thus $\beta\Delta V$ is identified

[21] It is assumed that $\lambda_L, \lambda_H, J_L, J_H, \Delta y$ and β. Dividing (1) by (2) 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 (1) then gives a unique solution for γ_0, and inserting these solutions into equation (3) gives a unique solution for $\alpha_0\Delta y$.

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 quartile of the distribution of State mean wages is 5.5% (53 moves in 964 person-years), and for someone in the top quartile it is 2.1% (16 moves in 754 person-years). If $1-\lambda_L = 53/964$ and $1-\lambda_H 16/754$, then $\gamma_0 = 7.34$, and $\Delta V = 1.02$, and the implied moving cost is $\gamma_0/\alpha_0 = 85.3\Delta y$. Taking Δy to be the difference in the mean wages for States in the top and bottom quartiles gives $\gamma_0/\alpha_0 = \$304,670$ (in 2008 dollars). On the other hand if $\lambda_L = .7$, the implied moving cost is only $14.4\Delta y$, or \$51,449. We conclude that the moving cost estimate is large mainly because the empirical relationship between income levels and migration probabilities is relatively weak.

6.3 Average Costs of Actual Moves

Our estimates of the deterministic components of moving costs are large because moves are rare in the data. But moves do occur, and in many cases there is no observable reason for a move, so that the observed choice must be attributed to unobserved payoff shocks, including random variations in moving costs. Given this heterogeneity in moving costs, both across individuals and over time for the same individual, the question arises as to how large the actual moving costs are, conditional on a move being made.[22] Because the payoff shocks are drawn from the type I extreme value distribution, this question has a relatively simple answer.

The cost of a move may be defined as the difference in the flow payoff for the current period due to the move. Since a move to location j exchanges ζ_{ℓ^0} for ζ_j , the average cost of a move from ℓ^0 to j, given state x, is

$$\bar{\Delta}(x, j) = \Delta(x, j) - E(\zeta_j - \zeta_{\ell^0} \mid d_j = 1)$$

where d_j is an indicator variable for the choice of location j. Thus for example if a move from ℓ^0 to j is caused by a large payoff shock in location j, the cost of the move may be much less than the amount given by the deterministic cost $\Delta(x, j)$.

In logit models, the expected gain from the optimal choice, relative to an arbitrary alternative that is not chosen, is a simple function of the probability of choosing the alternative (Kennan, 2008). In the present context, this result means that the average increase in the gross continuation value, for someone

[22]See Sweeting (2007) for a similar analysis of switching costs, in the context of an empirical analysis of format switching by radio stations.

who chooses to move from ℓ^0 to j, is given by

$$E\left(\tilde{v}(x,j)-\tilde{v}\left(x,\ell^0\right)\mid d_j=1\right)=-\frac{\log\left(\rho\left(x,\ell^0\right)\right)}{1-\rho\left(x,\ell^0\right)}$$

where $\tilde{v}(x,j)$ is the continuation value when the state is x and location j is chosen, which includes the current flow payoff and the discounted expected continuation value in location j:

$$\begin{aligned}\tilde{v}(x,j) &= v(x,j)+\zeta_j \\ &= u(x,j)+\beta\sum_{x'}p\left(x'\mid x,j\right)\overline{V}\left(x'\right)+\zeta_j\end{aligned}$$

The deterministic part of the moving cost is

$$\begin{aligned}\Delta(x,j) &= u\left(x,\ell^0\right)-u(x,j) \\ &= v\left(x,\ell^0\right)-v(x,j)+\beta\sum_{x'}\left(p\left(x'\mid x,j\right)-p\left(x'\mid x,\ell^0\right)\right)\overline{V}\left(x'\right) \\ &= \tilde{v}\left(x,\ell^0\right)-\zeta_{\ell^0}-\tilde{v}(x,j)+\zeta_j+\beta\sum_{x'}\left(p\left(x'\mid x,j\right)-p\left(x'\mid x,\ell^0\right)\right)\overline{V}\left(x'\right)\end{aligned}$$

This implies that the average moving cost, net of the difference in payoff shocks, is

$$\bar{\Delta}(x,j)\equiv\Delta(x,j)-E\left(\zeta_j-\zeta_{\ell^0}\mid d_j=1\right)=\frac{\log\left(\rho\left(x,\ell^0\right)\right)}{1-\rho\left(x,\ell^0\right)}+\beta\sum_{x'}\left(p\left(x'\mid x,j\right)-p\left(x'\mid x,\ell^0\right)\right)\bar{v}\left(x'\right)$$

Since some of the components of the state vector x are unobserved, we compute expected moving costs using the conditional distribution over the unobservables, given the observed wage and migration history. Recall that the likelihood of an individual history, for a person of type τ, is

$$L_i\left(\theta_\tau\right)=\frac{1}{n_\eta n_\varepsilon n_\xi\left(n_\upsilon\right)^{N_i}}\sum_{\omega^i\in\Omega(N_i)}\left(\prod_{t=1}^{T_i}\psi_{it}\left(\omega^i,\theta_\tau\right)\lambda_{it}\left(\omega^i,\theta_\tau\right)\right)$$

Thus the conditional probability of ω^i is

$$\begin{aligned}Q\left(\omega^i\right) &= \frac{\prod_{t=1}^{T_i}\psi_{it}\left(\omega^i,\theta_\tau\right)\lambda_{it}\left(\omega^i,\theta_\tau\right)}{\sum_{\omega\in\Omega(N_i)}\left(\prod_{t=1}^{T_i}\psi_{it}\left(\omega,\theta_\tau\right)\lambda_{it}\left(\omega,\theta_\tau\right)\right)} \\ &= \frac{\prod_{t=1}^{T_i}\psi_{it}\left(\omega^i,\theta_\tau\right)\lambda_{it}\left(\omega^i,\theta_\tau\right)}{n_\eta n_\varepsilon n_\xi\left(n_\upsilon\right)^{N_i}L_i\left(\theta_\tau\right)}\end{aligned}$$

Table 5: **Average Moving Costs**

		Move Origin and Destination			
		From Home	To Home	Other	Total
Previous Location	None	-$142,150 [56]	$132,979 [1]	-$38,207 [2]	-$133,964 [59]
	Home	—	$18,127 [40]	-$120,641 [10]	-$9,627 [50]
	Other	-$144,548 [8]	$109,244 [2]	-$64,944 [5]	-$84,174 [15]
	Total	-$142,450 [64]	$24,912 [43]	-$94,038 [17]	-$77,776 [124]
Note: the number of moves in each category is given in brackets.					

The unobserved part of the state variable consists of the location match components of wages and preferences. Since the distributions of these components have finite support, there is a finite set $\Omega(N_i)$ of possible realizations corresponding to the observed history for individual i; this set is indexed by ω^i. Let $x(\omega^i)$ be the state implied by ω^i (including the location match components in the current location, and in the previous location, if any). Then if individual i moves to location j in period t, the moving cost is estimated as

$$\hat{\Delta}_{it} = \sum_{\omega^i \in \Omega(N_i)} Q\left(\omega^i\right) \bar{\Delta}\left(x\left(\omega^i\right), j\right)$$

The estimated average moving costs are given in Table 5. There is considerable variation in these costs, but for a typical move the cost is negative. The interpretation of this is that the typical move is not motivated by the prospect of a higher future utility flow in the destination location, but rather by unobserved factors yielding a higher current payoff in the destination location, compared with the current location. That is, the most important part of the estimated moving cost is $\zeta_{\ell^0} - \zeta_j$, the difference in the payoff shocks. In the case of moves to the home location, on the other hand, the estimated cost is positive; most of these moves are return moves, but where the home location is not the previous location the cost is large, reflecting a large gain in expected future payoffs due to the move.

6.4 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,

Table 6: **Goodness of Fit**

Moves	Binomial		NLSY		Model	
None	325.1	75.3%	361	83.6%	36,177	83.7%
One	91.5	21.2%	31	7.2%	2,534	5.9%
More	15.4	3.6%	40	9.3%	4,493	10.4%
Movers with more than one move	14.4%		56.3%		63.9%	
Total observations	432		432		43,204	

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 2.9% (the number of moves per person-year in the data). Table 6 shows the predictions from this model: about 75% of the people never move, and of those who do move, about 14% move more than once. The NLSY data are quite different: about 84% never move, and about 56% of movers move more than once.[23] An obvious 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. The results show that the model does a good job of accounting for the heterogeneous migration probabilities in the data. The proportion of people who never move in the simulated data matches the proportion in the NLSY data almost exactly, and although the proportion of movers who move more than once is a bit high in the simulated data, the estimated model comes much closer to this statistic than the binomial model does.

6.4.1 Return Migration

Table 7 summarizes the extent to which the model can reproduce the return migration patterns in the data (the statistics in the Model column refer to the simulated data set used in Table 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 35%. The home premium also reduces the chance of initially leaving

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

Table 7: **Return Migration Statistics**

	NLSY	Model
Proportion of Movers who		
Return home	34.7%	35.6%
Return elsewhere	3.2%	6.0%
Move on	62.1%	58.4%
Proportion who *ever*		
Leave Home	14.4%	14.0%
Move from not-home	40.0%	42.5%
Return from not-home	25.7%	32.1%

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 model over-predicts the proportion who ever return home 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.

6.5 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 8 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.

Our model encompasses this simple human capital explanation of the age effect on migration.[24] There are two effects here. First, consider two locations paying

[24]Investments in location-specific human capital might also help explain why older workers are less likely to move. Marriage might be included under this heading, for example, as in Gemici (2008). 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

Table 8: **Annual Migration Rates by Age and Current Location**

	All			Not At Home			At Home		
Age	N	Moves	Migration Rate	N	Moves	Migration Rate	N	Moves	Migration Rate
20-25	2,359	84	3.6%	244	40	16.4%	2,115	44	2.1%
26-34	1,915	40	2.1%	228	20	8.8%	1,687	20	1.2%
All	4,274	124	2.9%	472	60	13.4%	3,802	64	1.7%
At Home means living now in the State of residence at age 14.									

different wages, and suppose that workers are randomly assigned to 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 is higher when workers are young.[25] Second, the human capital explanation 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 rates 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.

We find that the simple human capital model does not fully explain the relationship between age and migration in the data. Our model includes age as a state variable, to capture the effects 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 simple human capital explanation. The results in Table 2 show that this direct effect is significant.

6.6 Decomposing the Effects of Income on Migration Decisions

Migration is motivated by two distinct wage components in our model: differences in mean wages (μ_j) across locations, and individual draws from the location match distribution (υ_{ij}). The relevance of these components can be considered separately, first by suppressing the dispersion in υ, so that wages affect migration decisions only because of differences in mean wages across locations, and alternatively by specifying the wage distribution at the national level, so that 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).

include age as an explanatory variable.

[25]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 exceeds the moving cost), and the migration rate will be zero from then on.

Consider an economy in which everyone has the same preferences over locations, and also the same productivity in each location. In a steady state equilibrium, everyone is indifferent between locations: there are wage differences, but these just equalize the amenity differences. People move for other reasons, but there are just as many people coming into each location as there are going out. There should be no correlation between wages and mobility, in the steady state. Nevertheless, if moving costs are high, at any given time one would expect to see flows of workers toward locations with higher wages as part of a dynamic equilibrium driven by local labor demand shocks. As was mentioned above (in footnote 5), there is some evidence that local labor market shocks have long-lasting effects. So in a specification that uses only mean wages in each location (with no location match effects), we should find a relationship between mean wages and migration decisions. This is in fact what we find in Table 9 (in the "State Means" column). But we also find that the exclusion of location match wage effects is strongly rejected by a likelihood ratio test.

Even if differences in mean wages merely equalize the amenity differences between locations, the model predicts a relationship between wage realizations and migration decisions, because of location match effects: if the location match component is bad, the worker has an incentive to leave. This motivates the "National Wages" column of Table 9, where it is assumed that mean wages are the same in all locations (as they would be if measured wage differences merely reflect unmeasured amenities). We find that workers who have unusually low wages in their current location are indeed more likely to move.

Finally, the "Regional Amenities" column shows that the results are robust to the inclusion of regional amenity differences. Appendix A shows that the model is identified even if each location has an unobserved amenity value that is common to all individuals. In practice, we do not have enough data to estimate the complete model with a full set of fixed effects for all 50 locations. As a compromise, we divide the States into 13 regions, and present estimates for a model with fixed amenity values for each region.[26] This has little effect on the estimated income coefficient; moreover, a likelihood ratio test accepts the hypothesis that there are no regional amenity differences.[27]

6.7 Sensitivity Analysis

Our empirical results are inevitably based on some more or less arbitrary model specification choices. Table 10 explores the robustness of the results with respect

[26] The regions are as follows: (1) Northeast (NE, ME, VT, NH, MA, RI, CT); (2) Atlantic (DE, MD, NJ, NY, PA); (3) Southeast (SE, VA, NC, SC, GA, FL); (4) North Central (NC, MN, MI, WI, SD, ND); (5) Midwest (OH, IN, IL, IA, KS, NB, MO); (6) South (LA, MS, AL, AR); (7) South Central (OK, TX); (8) Appalachia (TN, KY, WV); (9) Southwest (AZ, NM, NV); (10) Mountain (ID, MT, WY, UT, CO); (11) West (CA, HI); (12) Alaska and (13) Northwest (OR, WA).

[27] Alaska is the only region that has a significant (positive) coefficient; this is perhaps not surprising given that the model specifies the utility flow as a linear function of average temperature, and Alaska is an outlier in this respect.

Table 9: **Alternative Income Specifications**

	Base Model		**State Means**		**National Wages**		**Regional Amenities**	
	$\hat{\theta}$	$\hat{\sigma}_\theta$	$\hat{\theta}$	$\hat{\sigma}_\theta$	$\hat{\theta}$	$\hat{\sigma}_\theta$	$\hat{\theta}$	$\hat{\sigma}_\theta$
Disutility of Moving	4.794	0.565	4.567	0.532	4.754	0.568	4.765	0.610
Distance	0.267	0.181	0.254	0.183	0.270	0.183	0.355	0.236
Adjacent Location	0.807	0.214	0.810	0.213	0.804	0.213	0.775	0.234
Home Premium	0.331	0.041	0.274	0.032	0.329	0.040	0.374	0.051
Previous Location	2.757	0.357	2.554	0.299	2.728	0.347	2.735	0.424
Age	0.055	0.020	0.061	0.019	0.055	0.020	0.052	0.020
Population	0.654	0.179	0.663	0.181	0.650	0.179	0.680	0.234
Stayer Probability	0.510	0.078	0.517	0.079	0.513	0.078	0.495	0.093
Cooling	0.055	0.019	0.040	0.019	0.055	0.019	0.031	0.044
Income	0.314	0.100	0.324	0.185	0.316	0.100	0.295	0.129
Wage intercept	-5.133	0.245	-5.405	0.239	-4.019	0.270	-5.133	0.249
Time trend	-0.034	0.008	-0.050	0.005	-0.011	0.009	-0.033	0.008
Age effect (linear)	7.841	0.356	8.080	0.367	7.439	0.381	7.842	0.360
Age effect (quadratic)	-2.362	0.129	-2.318	0.134	-2.384	0.128	-2.364	0.131
Ability (AFQT)	0.011	0.065	0.062	0.059	0.020	0.064	0.011	0.066
Interaction(Age,AFQT)	0.144	0.040	0.159	0.041	0.144	0.039	0.144	0.041
Transient s.d. 1	0.217	0.007	0.231	0.007	0.217	0.007	0.217	0.007
Transient s.d. 2	0.375	0.015	0.384	0.016	0.372	0.015	0.375	0.015
Transient s.d. 3	0.546	0.017	0.559	0.018	0.544	0.017	0.546	0.017
Transient s.d. 4	1.306	0.028	1.332	0.028	1.304	0.027	1.306	0.028
Fixed Effect 1	0.113	0.036	-1.028	0.014	-0.905	0.023	0.113	0.036
Fixed Effect 2	0.296	0.035	0.252	0.013	0.167	0.041	0.297	0.036
Fixed Effect 3	0.933	0.016	0.546	0.011	0.358	0.039	0.933	0.017
Wage Match	0.384	0.017	———		0.362	0.024	0.384	0.019
Loglikelihood	-4214.16		-4267.28		-4215.81		-4202.81	
The "Regional Amenities" model includes 12 regional dummy variables (coefficients not shown).								

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 compensate for amenity differences, then unadjusted wages should be used to measure the incentive to migrate. This specification yields a slightly lower estimate of the income coefficient, without much effect on the other coefficients, and the likelihood is lower (mainly because there is more unexplained variation in the unadjusted wages). 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 other specifications in Table 10 are concerned with sensitivity of the estimates to the discount factor (β), the horizon length (T), heterogeneity in moving costs and the inclusion of a second climate variable (heating degree days).[28] Again, the effect of income is quite stable across these alternative specifications.

7 Migration and Wages

7.1 Spatial Labor Supply Elasticities

We use the estimated model to analyze labor supply responses to changes 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: we use the estimated parameters to predict responses in a different environment. First we do a baseline calculation, 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.

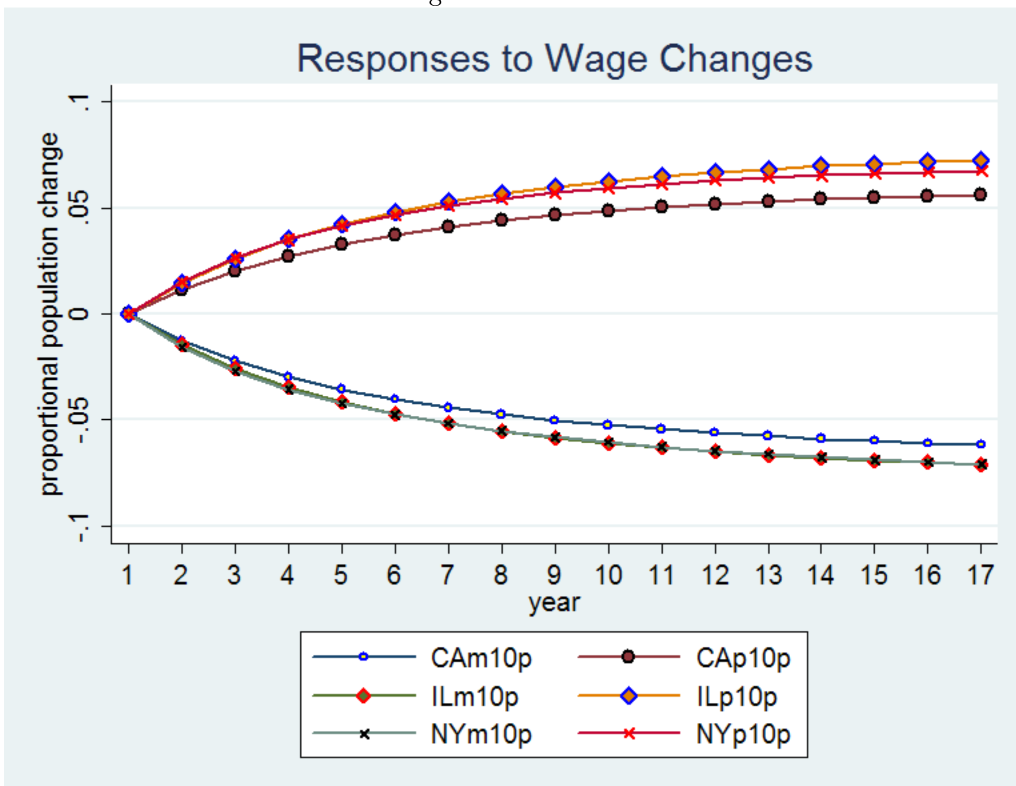
We take a set of people who are distributed over States as in the 1990 Census data for white male high school graduates aged 20 to 34. We assume that each person is initially in the home State, at age 20, and we allow the population distribution to evolve over 15 years, by iterating the estimated transition

[28] Table 10 is a sample of many 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.

Table 10: **Alternative Specifications**

	Base Model		No Cola		$\beta = .9$		$T = 40$		1 cost type		Heating	
	$\hat{\theta}$	$\hat{\sigma}_\theta$	$\hat{\theta}$	$\hat{\sigma}_\theta$	$\hat{\theta}$	$\hat{\sigma}_\theta$	$\hat{\theta}$	$\hat{\sigma}_\theta$	$\hat{\theta}$	$\hat{\sigma}_\theta$	$\hat{\theta}$	$\hat{\sigma}_\theta$
Disutility of Moving	4.794	0.565	4.704	0.555	4.697	0.578	4.495	0.613	5.282	0.559	4.765	0.556
Distance	0.267	0.181	0.283	0.182	0.299	0.198	0.267	0.186	0.264	0.182	0.276	0.193
Adjacent Location	0.807	0.214	0.797	0.215	0.838	0.232	0.81	0.224	0.78	0.214	0.794	0.221
Home Premium	0.331	0.041	0.325	0.04	0.467	0.052	0.382	0.041	0.185	0.022	0.325	0.039
Previous Location	2.757	0.357	2.709	0.367	2.809	0.348	2.811	0.333	3.377	0.321	2.771	0.365
Age	0.055	0.02	0.056	0.019	0.06	0.02	0.067	0.022	0.074	0.02	0.057	0.019
Population	0.654	0.179	0.64	0.18	0.697	0.187	0.652	0.18	0.645	0.164	0.682	0.191
Stayer Probability	0.51	0.078	0.512	0.078	0.487	0.08	0.494	0.079	0	—	0.505	0.079
Cooling	0.055	0.019	0.057	0.019	0.069	0.027	0.055	0.022	0.02	0.014	0.1	0.031
Heating	—	—	—	—	—	—	—	—	—	—	0.02	0.012
Income	0.314	0.1	0.262	0.096	0.455	0.14	0.361	0.111	0.146	0.069	0.307	0.099
Wage intercept	-5.133	0.245	-5.107	0.269	-5.131	0.245	-5.14	0.245	-5.157	0.244	-5.132	0.246
Time trend	-0.034	0.008	-0.029	0.01	-0.034	0.008	-0.034	0.008	-0.034	0.008	-0.034	0.008
Age effect (linear)	7.841	0.356	7.822	0.384	7.837	0.356	7.85	0.356	7.865	0.354	7.84	0.356
Age effect (quadratic)	-2.362	0.129	-2.379	0.127	-2.36	0.129	-2.365	0.129	-2.369	0.129	-2.362	0.129
Ability (AFQT)	0.011	0.065	0.047	0.066	0.017	0.065	0.014	0.065	0.015	0.065	0.012	0.065
Interaction(Age,AFQT)	0.144	0.04	0.132	0.039	0.14	0.04	0.142	0.04	0.146	0.04	0.145	0.04
Transient s.d. 1	0.217	0.007	0.22	0.007	0.217	0.007	0.217	0.007	0.217	0.007	0.217	0.007
Transient s.d. 2	0.375	0.015	0.38	0.016	0.375	0.015	0.375	0.015	0.375	0.015	0.375	0.015
Transient s.d. 3	0.546	0.017	0.553	0.016	0.547	0.017	0.546	0.017	0.546	0.017	0.546	0.017
Transient s.d. 4	1.306	0.028	1.322	0.029	1.308	0.028	1.306	0.028	1.308	0.028	1.307	0.028
Fixed Effect 1	0.113	0.036	0.132	0.036	0.112	0.036	0.112	0.035	0.112	0.035	0.113	0.036
Fixed Effect 2	0.296	0.035	0.307	0.037	0.295	0.035	0.295	0.035	0.296	0.036	0.295	0.035
Fixed Effect 3	0.933	0.016	0.966	0.02	0.933	0.016	0.934	0.016	0.934	0.016	0.933	0.017
Wage Match	0.384	0.017	0.401	0.019	0.384	0.017	0.384	0.017	0.382	0.017	0.384	0.018
Loglikelihood	-4214.163		-4281.937		-4213.292		-4213.288		-4231.035		-4213.357	

Figure 1:



probability matrix. We consider responses to wage increases and decreases representing a 10% change in the mean wage of an average 30-year-old, for selected States. First, we compute baseline transition probabilities using the wages that generated the parameter estimates. Then we increase or decrease the mean wage in a single State, and compare the migration decisions induced by these wage changes with the baseline. Supply elasticities are measured relative to the supply of labor in the baseline calculation. For example, the elasticity of the response to a wage increase in California after 5 years is computed as $\frac{\Delta L}{\Delta w}\frac{w}{L}$, where L is the number of people in California after 5 years in the baseline calculation, and ΔL is the difference between this and the number of people in California after 5 years in the counterfactual calculation.

Figure 1 shows the results for three large States that are near the middle of the one-period utility flow distribution. The supply elasticities are above 0.5. Adjustment is gradual, but is largely completed in 10 years. Our conclusion from this exercise is that despite the low migration rate in the data, the supply of labor responds quite strongly to spatial wage differences.

Table 11: **Migration Gains**

	Migration Rate	Mean	Match	Amenity	Total
		μ	υ	$\frac{\alpha_1 Y_1}{\alpha_0}$	
Mover type	5.27%	16705	1051	4500	22256
Stayer type	0	16678	4	4220	20902
Gain		27	1047	280	1354
Percentage Gain		0.1%	5.0%	1.3%	6.5%
Standard Deviation		1416	6633	2912	7331
Migration gains are measured in 2008 dollars.					

7.2 Migration and Wage Growth

Our model is primarily designed to quantify the extent to which migration is motivated by expected income gains. Interstate migration is a relatively rare event, and our results indicate that many of the moves that do occur are motivated by something other than income gains. This raises the question of whether the income gains due to migration are large enough to be interesting.

One way to answer this question is to compare the wages of the mover and the stayer type as time goes by, using simulated data. Table 11 shows results for a simulation that starts 1,000 people at home in each of the 50 States at age 20, and measures accumulated income and utility gains at age 34 (the oldest age in our NLSY sample)[29]. Migration increases the total utility flow by a modest but nontrivial amount. Most of the gain comes from improved location matches; even though there is considerable dispersion in mean wages across States, the estimated dispersion in the location match component of wages is much larger, and therefore a much more important source of income gains due to migration. The dollar value of the nonpecuniary gains due to (climate) amenities is also larger than the gains from moving toward high-wage States.

The importance of the home location can be seen by simulating migration decisions with the home premium parameter set to zero. The results are shown in Table 12. With no attachment to a home location, the annual migration rate increases to 6.5%, and the mover type moves about once every seven years. By age 34 the accumulated gains due to migration exceed 20% of the base utility level. Given that people are willing to forgo gains of this magnitude in order to stay in their home location, it follows that the costs of forced displacements (due to natural disasters such as hurricane Katrina, for example) are very high.

8 Conclusion

We have developed a tractable econometric model of optimal migration in response to income differentials across locations. The model improves on previous

[29]The results are weighted by the State distribution of white male high school graduates aged 19-20 from the 1990 Census

Table 12: **Migration Gains with no Home Location**

	Migration Rate	Mean	Match	Amenity	Total
Mover type	12.90%	16729	2739	5578	25046
Stayer type	0	16678	38	4221	20936
Gain		51	2701	1357	4110
Percentage Gain		0.2%	12.9%	6.5%	19.6%

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 .5 . 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 partly 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.

A Identification

In this appendix we show how the income coefficient α_0 is identified, even in the presence of unobserved amenities in each location.[30] Fix home location and

[30]Identification of dynamic discrete choice models is analyzed by Magnac and Thesmer (2002); identification of static equilibrium discrete choice models is analyzed by Berry and

age, with no previous location. Assume initially that there is no unobserved heterogeneity, and there is no location match component of preferences. Then the state consists of the current location and the location match component of wages. We will show that the model is then fully identified.

The choice probabilities are given by

$$\rho\left(\ell, \upsilon_s, j\right)=\begin{cases}\frac{\exp\left(-\Delta_{\ell j}+\beta\bar{V}_0(j)\right)}{\exp\left(\beta\bar{V}_s(\ell)\right)+\sum_{k\neq\ell}\exp\left(-\Delta_{\ell k}+\beta\bar{V}_0(k)\right)} & j\neq\ell\\ \frac{\exp\left(\beta\bar{V}_s(\ell)\right)}{\exp\left(\beta\bar{V}_s(\ell)\right)+\sum_{k\neq\ell}\exp\left(-\Delta_{\ell k}+\beta\bar{V}_0(k)\right)} & j=\ell\end{cases}$$

where $\Delta_{\ell j}$ is the cost of moving from location ℓ to location j, $\bar{V}_s\left(j\right)$ is the expected continuation value in j, given the location match component υ_s, before knowing the realization of ζ, and $\bar{V}_0\left(j\right)$ is the expected continuation value before knowing the realization of υ:

$$\bar{V}_0\left(j\right)=\frac{1}{n}\sum_{s=1}^{n}\bar{V}_s(j)$$

The probability of moving from ℓ to j, relative to the probability of staying, is

$$\frac{\rho\left(\ell,\upsilon_s,j\right)}{\rho\left(\ell,\upsilon_s,\ell\right)}=\exp\left(-\Delta_{\ell j}+\beta\left(\bar{V}_0(j)-\bar{V}_s(\ell)\right)\right)$$

Thus

$$\frac{1}{n}\sum_{s=1}^{n}\log\left(\frac{\rho\left(\ell,\upsilon_s,j\right)}{\rho\left(\ell,\upsilon_s,\ell\right)}\right)=-\Delta_{\ell j}+\beta\left(\bar{V}_0(j)-\bar{V}_0(\ell)\right)$$

and

$$\frac{1}{n}\sum_{s=1}^{n}\log\left(\frac{\rho\left(\ell,\upsilon_s,j\right)}{\rho\left(\ell,\upsilon_s,\ell\right)}\frac{\rho\left(j,\upsilon_s,\ell\right)}{\rho\left(j,\upsilon_s,j\right)}\right)=-\Delta_{\ell j}-\Delta_{j\ell}$$

This identifies the round-trip moving cost between ℓ and j.

The one-way moving costs are identified under weak assumptions on the moving cost function; for example symmetry is obviously sufficient. In the model, the round-trip moving cost between two non-adjacent locations (for someone aged a with no previous location) is given by

$$\Delta_{\ell j}+\Delta_{j\ell}=2\left(\gamma_0+\gamma_4 a+\gamma_1 D(j,\ell)\right)-\gamma_5\left(n_j+n_\ell\right)$$

Since distance and population vary independently, one can choose three distinct location pairs, such that the three moving cost equations are linearly independent; these equations identify γ_1, γ_5 and $\gamma_0+\gamma_4 a$. Then by choosing two different

Haile (2008)

ages γ_0 and $\gamma_4$are identified, and by comparing adjacent and non-adjacent pairs, γ_2 is identified.

If the continuation value in all states is increased by the same amount, then the choice probabilities are unaffected, so one of the values can be normalized to zero.[31] We assume $\bar{V}_0(J) = 0$. Then

$$\frac{1}{n}\sum_{s=1}^{n}\log\left(\frac{\rho\left(\ell, \upsilon_s, J\right)}{\rho\left(\ell, \upsilon_s, \ell\right)}\right) = -\Delta_{\ell j} - \beta\bar{V}_0(\ell).$$

This identifies $\bar{V}_0(\ell)$, since we assume that β is known, and $\Delta_{\ell j}$ has already been identified. And once $\bar{V}_0$ is identified, $\bar{V}_s(\ell)$ is identified by the equation

$$\log\left(\frac{\rho\left(\ell, \upsilon_s, j\right)}{\rho\left(\ell, \upsilon_s, \ell\right)}\right) = -\Delta_{\ell j} + \beta\left(\bar{V}_0(j) - \bar{V}_s(\ell)\right).$$

Given that the expected continuation values in all states are identified, the flow payoffs are identified by

$$\bar{V}_s(\ell) \quad = \quad \bar{\gamma} + \alpha_0\upsilon_s + A_\ell + \log\left(\exp\left(\beta\bar{V}_s(\ell)\right) + \sum_{k\neq\ell}\exp\left(\Delta_{\ell k} + \beta\bar{V}_0(k)\right)\right)$$

where A_ℓ represents amenity values and other fixed characteristics of location ℓ (both observed and unobserved), and where υ_s represents the location match component of wages. The income coefficient α_0 is identified by differencing this equation with respect to s (thereby eliminating A_ℓ), and A_ℓ is then identified as the only remaining unknown in the equation.

This identification argument uses several simplifying assumptions, but these can be relaxed. It is assumed that the distribution of υ is known, and that the realizations of υ_s are known for each person. This requires a long panel: the wage history is sufficient to identify the individual fixed effect and the location match effect. It is also assumed that there is no unobserved heterogeneity. But if the history is arbitrarily long, then each person is seen arbitrarily many times in each state. So if there is a permanent unobserved location match component of preferences, the distribution of this component is identified because there is a distribution of choice probabilities in each state – some people are more inclined to stay, and others are more inclined to leave, and this can only be because they have different realizations of the location match component. Finally, the derivation of explicit identifying equations uses the logit structure implied by the assumption that the payoff shocks ζ are drawn from the Type-I extreme value distribution, but similar (although messier) arguments can be used for other distributions.

[31]It might seem that the choice probabilities are also invariant to a rescaling of the continuation values, but we have already normalized the scale by assuming additive payoff shocks drawn from the extreme value distribution.

B The Sample

In this appendix we describe the selection rules use to construct the analysis sample of 432 respondents with 4,274 person-years.

As noted in the text we applied strict sample inclusion criteria to obtain a relatively homogenous sample. In Table 13 we report the selection rules and the number of respondents deleted by each rule. The NLSY79 contains three subsamples, a nationally representative cross section sample, a supplemental sample of minorities and economically disadvantaged youth and a sample of individuals in the military in 1979. We start with the 2,439 white non-Hispanic males in the cross-section sample. We exclude respondents who ever served in the military, and we include only those with exactly a high school education.

We assume that permanent labor force attachment begins at age 20; thus we exclude respondents who were born in 1957, and who were therefore not interviewed until they were already more than 20 years old. We drop those who are in school or report graduating from high school at age 20. Since we use the AFQT (conducted in 1980) to help explain wages, we drop individuals with missing AFQT scores. Respondents who report being out of the labor force for more than one year after age 19, due to disability, tending house, or "other", are dropped on the grounds that they are not typical of this population. We use residence at age 14 as the home location, so we drop people for whom this variable is missing; we also drop people whose location at age 20 is unknown. We dropped one person who never reported income after age 19. We also dropped four people who died in their 30s, again on the grounds that they are atypical. Finally, we dropped one individual who was incarcerated in 1993 (after reporting remarkably high incomes in earlier years). Application of these criteria produced a sample of 439 individuals and 6,585 person years.

We apply two period-level restrictions. The first is that the histories must be continuous: we follow individuals from age 20 to their first non-interview or the 1994 interview. Since a missed interview means that location is unknown, we discard all data for each respondent after the first missed interview. Finally, we delete observations before age 20 from the analysis sample. Seven respondents have information only during their teenage years.

Our final sample contains 4,274 periods for 432 men. There are 124 interstate moves, with an annual migration rate of 2.9 percent. More than a one-third of the moves (43) were returns to the home location. There are 361 people who never moved, 31 who moved once, 33 who moved twice and 7 who moved three times or more. The median age is 25, reflecting the continuous-history restriction.

C Validation of ML Estimates

The parameter estimates from Table 2 were used to generate 100 replicas of each NLSY observation, starting from the actual value in the NLSY data, and

Table 13: **Sample Selection**

	Respondents		**Person-years**	
White Non-Hispanic Males (Cross Section Sample)		2,439		39,024
Restrictions applied to respondents				
Ever in Military	-246			
High School Dropouts and College Graduates	-1,290			
Attended college	-130			
Older than age 20 at start of sample period	-134			
Missing AFQT score	-41			
Attend or graduate from high school at age 20	-87			
Not in labor force for more one year after age 19	-44			
Location at age 20 not reported	-20			
Income information inconsistent	-1			
Died before age 30	-4			
Residence at age 14 not reported	-2			
In jail in 1993	-1			
Subtotal	-2,000	439		6,585
Restrictions applied to periods				
Delete periods after first gap in history	-1		-1,104	
Delete periods before age 20	-6		-1,207	
Analysis Sample		432		4,274
Years per Person				
1		14		14
2		16		32
3		19		57
4		14		56
5		14		70
6		14		84
7		13		91
8		9		72
9		34		306
10		61		610
11		53		583
12		44		528
13		45		585
14		44		616
15		38		570
		432		4,274

Table 14: **Estimates from Simulated Migration Histories**

	Base Model		100 Reps		
	$\hat{\theta}$	$\hat{\sigma}_\theta$	$\hat{\theta}$	$\hat{\sigma}_\theta$	t
Disutility of Moving	4.794	0.565	4.775	0.058	-0.322
Distance	0.267	0.181	0.293	0.015	1.748
Adjacent Location	0.807	0.214	0.775	0.017	-1.856
Home Premium	0.331	0.041	0.328	0.004	-0.876
Previous Location	2.757	0.357	2.801	0.032	1.379
Age	0.055	0.020	0.055	0.002	0.136
Population	0.654	0.179	0.649	0.017	-0.305
Stayer Probability	0.510	0.078	0.512	0.008	0.248
Cooling	0.055	0.019	0.059	0.002	1.794
Income	0.314	0.100	0.314	0.008	0.048
Wage intercept	-5.133	0.245	-5.106	0.033	0.817
Time trend	-0.034	0.008	-0.033	0.001	0.909
Age effect (linear)	7.841	0.356	7.797	0.049	-0.892
Age effect (quadratic)	-2.362	0.129	-2.348	0.018	0.807
Ability (AFQT)	0.011	0.065	0.019	0.010	0.827
Interaction(Age,AFQT)	0.144	0.040	0.137	0.007	-0.953
Transient s.d. 1	0.217	0.007	0.217	0.001	-1.101
Transient s.d. 2	0.375	0.015	0.374	0.002	-0.297
Transient s.d. 3	0.546	0.017	0.546	0.002	0.005
Transient s.d. 4	1.306	0.028	1.309	0.004	0.809
Fixed Effect 1	0.113	0.036	0.113	0.003	0.156
Fixed Effect 2	0.296	0.035	0.296	0.003	0.139
Fixed Effect 3	0.933	0.016	0.933	0.002	-0.244
Wage Match	0.384	0.017	0.382	0.002	-1.340
Loglikelihood, $\chi^2(24)$	-4214.16		-472883.4		18.33

allowing the model to choose the sequence of locations. Table 14 gives maximum likelihood estimates using the simulated data. The last column reports the t-value testing the difference between the 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). The estimated coefficients are close to the true values, and the χ^2 test accepts the truth. We take this as evidence that our estimation and simulation programs work.

References

[1] Altonji, Joseph G. and Pierret Charles R., "Employer Learning and Statistical Discrimination," *Quarterly Journal of Economics*, February 2001, vol.

116, no. 1, pp. 313-350.

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

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

[4] Barro, Robert J. and Xavier Sala-i-Martin (1991), “Convergence across States and Regions,” Brookings Papers on Economic Activity, 1: 107-158.

[5] Berry, Steven T. and Philip A. Haile, “Nonparametric Identification of Multinomial Choice Models with Heterogeneous Consumers,” unpublished, December 2008.

[6] Blanchard, Olivier Jean and Lawrence F. Katz (1992), “Regional Evolutions,” Brookings Papers on Economic Activity, 1: 1-37.

[7] Crawford, Gregory S. and Matthew Shum, “Uncertainty and Learning in Pharmaceutical Demand,”Econometrica, Vol. 73, No. 4, pp. 1137-1173, July 2005.

[8] 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.

[9] Erdem, Tülin and Michael P. Keane (1996), “Decision-making Under Uncertainty: Capturing Dynamic Brand Choice Processes in Turbulent Consumer Goods Markets,”Marketing Science 15 (1), 1–20.

[10] Gallin, Joshua Hojvat (2004), “Net Migration and State Labor Market Dynamics,” Journal of Labor Economics, 22, 1-21.

[11] 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.

[12] Ham, John C., Xianghong Li and Patricia B. Reagan, “Propensity Score Matching, a Distance-Based Measure of Migration, and the Wages of Young Men,” University of Southern California, February 2006.

[13] Holt, Frederick (1996), “Family Migration Decisions: A Dynamic Analysis,” unpublished paper, University of Virginia.

[14] Keane, Michael P. and Kenneth I. Wolpin (1994), “The Solution and Estimation of Discrete Choice Dynamic Programming Models by Simulation and Interpolation: Monte Carlo Evidence,” Review of Economics and Statistics, Vol. 76, No. 4. (Nov., 1994), pp. 648-672

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

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

[17] Kennan, John (2008), "Average Switching Costs in Dynamic Logit Models," http://www.ssc.wisc.edu/~jkennan/research/LogitSwitchingCosts.pdf

[18] 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.

[19] 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.

[20] McFadden, D. (1973), "Conditional Logit Analysis of Qualitative Choice Behavior," in P. Zarembka (ed.) Frontiers in Econometrics, New York, Academic Press.

[21] Neal, Derek (1999), "The Complexity of Job Mobility of Young Men," Journal of Labor Economics, (April): 237-261.

[22] Pessino, Carola (1991), "Sequential migration theory and evidence from Peru," Journal of Development Economics 36 (July): 55–87.

[23] Reagan, Patricia and Randall Olsen (2000) "You Can Go Home Again: Evidence from Longitudinal Data," Demography (August) 37: 339-350.

[24] Rust, John (1987) "Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher," Econometrica, 55 (5): 999-1033.

[25] 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.

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

[27] Sjaastad, L. A. (1962), "The Costs and Returns of Human Migration," Journal of Political Economy 70:80-89.

[28] Sweeting, Andrew (2007), "Dynamic Product Repositioning in Differentiated Product Markets: The Case of Format Switching in the Commercial Radio Industry," September 2007, NBER Working Paper #13522.

[29] Magnac, Thierry and Thesmar, David (2002), "Identifying Dynamic Discrete Decision Processes," *Econometrica*, 70, 801-816.

[30] Topel, Robert H. (1986), "Local Labor Markets," Journal of Political Economy, 94(3), part 2, S111-S143.

[31] Tunali, Insan (2000), "Rationality of Migration," International Economic Review, (November) 41: 893-920.

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