The Effect of Expected Income on Individual Migration Decisions

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

March 2003

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. On the other hand we find no evidence of a response to geographic differences in wage distributions. Instead, the results suggest that the link between income and migration decisions is driven by a tendency to move in search of a better locational match when the income realization in the current location is unfavorable.

---

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

1 Introduction

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

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

Structural dynamic models of migration over many locations have not been estimated before, presumably because the required computations have not been feasible.[3] A structural representation of the decision process is of interest for the usual reasons: we are ultimately interested in quantifying responses to income shocks or policy interventions not seen in the data, such as local labor demand shocks, or changes in welfare benefits. Our basic empirical question is the extent to which people move for the purpose of improving their income prospects. Work by Keane and Wolpin (1997) and by Neal (1999) indicates that individuals make surprisingly sophisticated calculations regarding schooling and occupational choices. Given the magnitude of geographical wage differentials, and given the findings of Topel (1986) and Blanchard and Katz (1992) regarding the responsiveness of migration flows to local labor market conditions, one might expect to find that income differentials play an important role in migration decisions.[4]

---

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

[3] Holt (1996) estimated a dynamic discrete choice model of migration, but his framework modeled the move/stay decision and not the location-specific flows. Similarly, Tunali (2000) gives a detailed econometric analysis of the move/stay decision using microdata for Turkey, but his model does not distinguish between alternative destinations.

[4]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).

We model individual decisions to migrate as a job search problem. A worker can draw a wage only by visiting a location, thereby incurring a moving cost. Locations are distinguished by known differences in wage distributions, amenity values and alternative income sources. A worker starts the life-cycle in some home location and must determine the optimal sequence of moves before settling down.

The decision problem is too complicated to be solved analytically, so we use a discrete approximation that can be solved numerically, following Rust (1994). The model is sparsely parameterized. In addition to expected income, migration decisions are influenced by moving costs, including a fixed cost, a reduced cost of moving to a previous location, and a cost that is proportional to distance, and by differences in location size, measured by the population in origin and destination locations. We also allow for a bias in favor of the home location.

Our main substantive conclusion is that interstate migration decisions are indeed influenced to a substantial extent by income prospects. On the other hand we find no evidence of a response to geographic differences in wage distributions. Instead, the results suggest that the link between income and migration decisions is driven by a tendency to move in search of a better locational match when the income realization in the current location is unfavorable.

More generally, the paper demonstrates that a fully specified econometric model of optimal dynamic migration decisions is feasible, and that it is capable of matching the main features of the data, including repeat and return migration. Although this paper focuses on the relationship between income prospects and migration decisions at the start of the life cycle, suitably modified versions of the model can potentially be applied to a range of issues, such as the migration effects of interstate differences in welfare benefits, the effects of joint career concerns on household migration decisions, and the effects on retirement migration of interstate differences in tax laws.[5]

2 Migration Dynamics

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

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

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

Table 1: Interstate Migration Flows
NLSY, 1979-92

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

3 An Optimal Search Model of Migration

We model migration as an optimal search process. The basic assumption is that wages are local prices of individual skill bundles. The individual knows the wage in the current location, but in order to determine the wage in another location, it is necessary to move there, at some cost.

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

Suppose there are J locations, and individual i's income y_{ij} in location j is a random variable with a known distribution. Migration decisions are made so as to maximize the expected discounted value of lifetime utility, subject to budget constraints. Consider a person with "home" location h, who is in location ℓ this period and in location j next period. The flow of utility in the current period for such a person is specified as

$$u_h(C;\ell,j) = \alpha \frac{C^{1-\upsilon}-1}{1-\upsilon} + \kappa\chi_{\{j=h\}} - \Delta(\ell,j;h)$$

The notation is as follows. C is consumption in the current period and $\upsilon \geq 0$ is a constant relative risk aversion coefficient. There is a premium κ that allows each individual to have a preference for their native location (χ_A is used as an indicator meaning that A is true). The cost of moving from ℓ to j is denoted by $\Delta(\ell,j;h)$.

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. Suppose the marginal utility of income is constant ($\upsilon = 0$ in the specification above), and suppose 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.[6] This is a natural benchmark model, although of course it imposes strong assumptions.[7]

[6]Note that this neatly sidesteps the question of whether moving costs should be specified as "psychic" costs that directly reduce utility, or as monetary costs that reduce disposable income. With constant marginal utility of income, there is no meaningful difference between these two specifications.

[7]Even if the marginal utility of consumption is not constant, one can still compute the increase in current-period consumption needed to just offset the utility cost of moving, and use this to translate the utility cost into an income equivalent. Then the optimal migration problem can be viewed as maximization of net lifetime income, and this will be a good approximation if the compensating variation in consumption is roughly constant. But this argument rests on the assumption that the individual can borrow against future income (including income generated by a move) in order to sustain current consumption.

There is little hope of solving this problem analytically. In particular, the Gittins index solution of the multiarmed bandit problem cannot be applied because there is a cost of moving.[8] But by using a discrete approximation of the wage distribution in each location, we can compute the value function and the optimal decision rule by standard dynamic programming methods, following Rust (1994).

Let F_j be the wage distribution function in location j. We approximate this by a discrete distribution over n points, as follows. Let $a_j(s) = F_j^{-1}\left(\frac{s}{n} - \frac{1}{2n}\right)$, where s = 1,2,...,n. Then F_j is approximated by a uniform distribution over the set $\{a_j(s)\}_{s=1}^{n}$. For example, if n = 10, the approximation puts probability 1/10 on the 5th, 15th, ... 95th percentiles of the distribution F_j.

3.1 The Value Function

Consider a person currently in location ℓ, with a J-vector ω summarizing what is known about wages in all locations. Here ω_j is either 0 or an integer between 1 and n, with the interpretation that if $\omega_j = s > 0$, then the wage in location j is known to be $a_j(s)$, and if $\omega_j = 0$ then the wage in location j is still unknown, so that if the person moves to j, the wage will be $a_j(s)$ with probability 1/n, for $1 \le s \le n$. The value function for a native of location h can be written in recursive form as

$$V_h(\ell,\omega) = \begin{cases} \max_j \left[u_h\left(a_\ell(\omega_\ell);\ \ell,j \right) + \beta V_h(j,\omega) \right] & \text{if } \omega_\ell > 0 \\ \frac{1}{n} \sum_{s=1}^{n} V_h(\ell,\omega_1,\ldots,\omega_{\ell-1},s,\omega_{\ell+1},\ldots,\omega_J) & \text{if } \omega_\ell = 0 \end{cases}$$

We compute V_h by value function iteration. It is convenient to use $V_h(\ell,\omega) \equiv 0$ as the initial estimate, so that if T is the number of iterations, the result gives the optimal policy for a (rolling) T-period horizon.

4 Empirical Implementation

An important limitation of the discrete dynamic programming method is that the number of states is typically large, even if the search problem is relatively simple. If there are J locations and the wage distribution has n points of support, the number of states is $J(n+1)^J$. For example a model with J=5 and

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

n=10 has 805,255 states. Although value functions for such a model can be computed in a few hours, estimation of the structural parameters requires that the value function be computed many times. Estimation becomes infeasible unless the number of structural parameters is small.

Ideally, locations would be defined as local labor markets. The smallest geographical unit identified in the NLSY is the county, but we obviously cannot let J be the number of counties, since there are over 3,100 counties in the U.S. Indeed, even if J is the number of States, the model is numerically infeasible, but by restricting the information available to each individual an approximate version of the model can be estimated; this is explained below.

4.1 Outline of the Estimation Method

We first expand the model to allow for unobserved heterogeneity in individual payoffs. Let $\zeta = (\zeta_1,\zeta_2,\ldots,\zeta_J)$ be a vector of idiosyncratic utility adjustments that are known to the worker before the migration decision is made in each period, but not observed by the econometrician. We assume that each component ζ_j is drawn independently according to a distribution function π; also, these draws are independent across individuals and over time. The individual's value function is then given by

$$V_h(\ell,\omega,\zeta;\theta) = \begin{cases} \max_j \left[u_h\left(a_\ell(\omega_\ell);\ \ell,j \right) + \zeta_j + \beta \bar{V}_h(j,\omega;\theta) \right] & \text{if } \omega_\ell > 0 \\ \frac{1}{n} \sum_{s=1}^{n} V_h(\ell,\omega_1,\ldots,\omega_{\ell-1},s,\omega_{\ell+1},\ldots,\omega_J,\zeta;\theta) & \text{if } \omega_\ell = 0 \end{cases}$$

where θ is the vector of unknown parameters and the expected value function $\bar{V}$ is defined by

$$\bar{V}(j,\omega;\theta) \equiv \int V(j,\omega,\zeta;\theta)\, d\pi(\zeta)$$

If π is the Type 1 Extreme Value distribution then, using arguments due to McFadden (1973) and Rust (1987) we can show that the function $\bar{V}$ satisfies

$$\bar{V}_h(\ell,\omega;\theta) = \begin{cases} \log\left(\sum_{j=1}^{J} \exp\left[v_{hj}(\ell,\omega;\theta) \right] \right) & \text{if } \omega_\ell > 0 \\ \frac{1}{n} \sum_{s=1}^{n} \bar{V}_h(\ell,\omega_1,\ldots,\omega_{\ell-1},s,\omega_{\ell+1},\ldots,\omega_J;\theta) & \text{if } \omega_\ell = 0 \end{cases}$$

where

$$v_{hj}(\ell,\omega;\theta) \;=\; u_h\Big(a_\ell(\omega_\ell)\,;\; \ell,j \Big) \;+\; \beta\,\bar{V}_h(j,\omega;\theta)$$

This gives the probability, $\Pr[d(j) = 1 \,|\, h, \ell,\omega]$, that a native of h in location ℓ with information ω will move to location j:

$$\Pr[d(j)=1 \,|\, h,\ell,\omega;\theta] \;=\; \frac{\exp[v_{hj}(\ell,\omega;\theta)]}{\sum_{\tau=1}^{J} \exp[v_{h\tau}(\ell,\omega;\theta)]}$$

4.2 A Limited History Approximation

When the number of locations is moderately large, the model becomes computationally infeasible.[9] This is a common problem with discrete dynamic programming models, and various devices have been proposed to deal with it. In our context it seems natural to use an approximation that takes advantage of the timing of migration decisions. We have assumed that information on the value of human capital in alternative locations is permanent, and so if a location has been visited previously, the wage in that location is known, no matter how much time has passed. This means that the number of possible states increases geometrically with the number of locations. In practice, however, the number of people seen in many distinct locations is small. Thus by restricting the information set to include only wages seen in recent locations, it is possible to drastically shrink the state space while retaining most of the information actually seen in the data. Specifically, we suppose that the number of wage observations cannot exceed M, with M < J, so that it is not possible to be fully informed about wages at all locations. Then if the wage distribution in each of J locations has n points of support, the number of states is $(Jn)^M$, since this is the number of possible M-period histories describing the locations visited most recently, and the wages found there. For example, if J is 50 and n is 5 and M is 2, the number of states is 62,500, which is manageable.

This approximation reduces the number of states in the most obvious way: we simply delete most of them.[10] Someone who has “too much” wage information in the big state space is reassigned to a less-

[9]And it will remain so: for example, if a location is a State, 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.

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

informed state. Individuals make the same calculations as before when deciding what to do next, and the econometrician uses the same procedure to recover the parameters governing the individual's decisions. There is just a shorter list of states, so people with different histories may be in different states in the big model, but they are considered to be in the same state in the reduced model. In particular, people who have the same recent history are in the same state, even if their previous histories were different (and people who have different wage information now may have the same information following a move).

In order to obtain the likelihood using this approximation, it is convenient to redefine notation. Let $\ell = (\ell^0, \ell^1, ... \ell^{M-1})$ be an M-vector containing the sequence of recent locations (beginning with the current location), and let ω be the corresponding sequence containing recent wage information. Then the probability that an individual in state (ℓ,ω) will move to location j can again be written in the form

$$\Pr[d(j)=1 \mid h,\ell,\omega;\theta] = \frac{\exp[v_{hj}(\ell,\omega;\theta)]}{\sum_{\tau=1}^{J} \exp[v_{h\tau}(\ell,\omega;\theta)]}$$

where v_j is now defined as

$$v_{hj}(\ell,\omega;\theta) = u_h\left(a_{\ell^0}(\omega_{\ell^0});\ \ell^0, j \right) + \beta \bar{V}_h\left((j,\ell^0),(\omega^j,\omega^0);\theta \right)$$

with

$$\bar{V}_h(\ell,\omega;\theta) = \begin{cases} \log\left(\sum_{j=1}^{J} \exp[v_{hj}(\ell,\omega;\theta)] \right) & \text{if } \omega^0 > 0 \\ \frac{1}{n} \sum_{s=1}^{n} \bar{V}_h(\ell,(s,\omega^1);\theta) & \text{if } \omega^0 = 0 \end{cases}$$

4.3 Population Effects

It has long been recognized that location size matters in migration models (see e.g. Schultz [1982]). California and Wyoming cannot reasonably be regarded as just two alternative places, to be treated symmetrically as origin and destination locations. To take one example, a person who moves to be close to a friend or relative is more likely to have friends or relatives in California than in Wyoming. A convenient way to model this in our framework is to allow for more than one draw from the distribution of preference shocks in each location. Specifically, we assume 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 migration probabilities:

$$\Pr[d(j)=1 \mid h,\ell,\omega;\theta] = \frac{\xi_j}{\sum_{\tau=1}^{J} \xi_\tau}$$

where

$$\xi_\tau = (1+\psi n_\tau)\exp[v_{h\tau}(\ell,\omega;\theta)]$$

Here n_j denotes the population in location j, and the (nonnegative) parameter ψ can be interpreted as the number of additional draws per person.

4.4 Moving Costs

The cost of moving is specified as

$$\Delta(\ell,j;h) = \left[\gamma_0 + \gamma_1 D(\ell^0,j) - \gamma_2 \chi(j=\ell^1) - \gamma_3 n_j\right]\chi(j \neq \ell^0)$$

The notation is as follows. The first two terms specify the moving cost as an affine function of the distance $D(\ell^0,j)$ from ℓ^0 to j. The next term allows for the possibility that it is cheaper to move to a previous location, relative to moving to a new location (χ denotes the indicator function). The last term is an alternative specification of the effect of location size, allowing for the possibility that it is cheaper to move to a large location, as measured by population size n_j. One motivation for this is that a larger location is more likely to contain friends or relatives who would help reduce the cost of the move.

4.5 Computation

Since the parameters are embedded in the value function, computation of the gradient and hessian of the loglikelihood function is not a simple matter (although in principle these derivatives can be computed in a straightforward way using the same iterative procedure that computes the value function itself). We maximize the likelihood using an "amoeba" algorithm that implements the downhill simplex method of Nelder and Mead. This method does not use derivatives, and it seems appropriate for problems such as this in which there is no reason to expect that the loglikelihood function is concave. In practice the method works well for the models we have estimated so far; in particular, it is robust to large changes in the starting values of the parameters. On the other hand, the method is slow, and so we also use gradient methods to speed up the computations, particularly when doing sensitivity analysis.[11]

5 Empirical Results

We analyze the migration decisions of men aged 20-35, using the non-military subsample of the NLSY79, observed over the period 1979-1992. In order to obtain a relatively homogeneous sample, we consider only white high-school graduates with no college education, using only the years after schooling is completed.[12]

5.1 Age Adjustment of Earnings

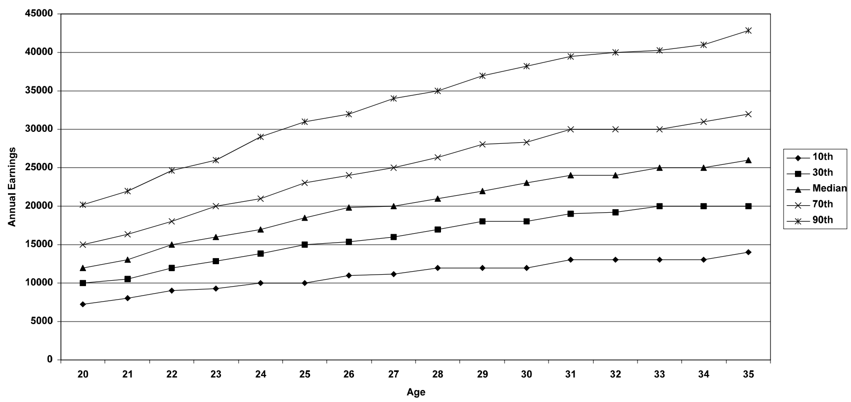
The model assumes that wages are stationary, and that each individual draws a level of permanent income in each location that is visited. In the data, however, wages vary systematically with age, and there are also substantial transient wage variations. Figure 1 shows the age-earnings profiles (by quintile) for white high school graduates in the NLSY79. These profiles are steep: earnings more than double between the ages of 20 and 35. Since migrants are necessarily older following a move than they were before the move, we must make some adjustment for age, so as not to attribute to migration the earnings growth due to age. We assume that wage components are additively separable, that the age-earnings profile is the same across all locations, and that the transient wage component is drawn from the same distribution in all locations. This implies that neither the transitory component nor the earnings profile is relevant for migration decisions.

---

[11] Given reasonable starting values (such as a fixed cost of moving that matches the average migration rate, with all other parameters set to zero), the maximal likelihood is typically reached within 24 hours, on a Pentium 4 machine. An example of our (FORTRAN90) computer program can be found at www.ssc.wisc.edu/~jkennan/research/mbr21.f90.

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

Figure 1: Annual Earnings - White Male HS Graduates (1990 PUMS)



Suppose that the wage of individual i in location j at age a is

$$w_{ij}(a) \; = \; \mu_j + \phi(a) + \eta_i + u_{ij} + \varepsilon_{ij}(a)$$

where μ_j is a known constant, $\phi(a)$ is a known age-earnings profile, η_i is an individual effect that is fixed across locations (and known to the individual), u_{ij} is a permanent match effect, and ε is a transient effect. We assume that η, u and ε are independent, and that u and ε are identically distributed across locations. In order to implement the model, we first need to estimate the distribution of u. One problem is that even if the mean of u_{ij} across individuals is zero in all locations, the realizations of u found in measured wages reflect selection effects due to migration decisions. Allowing for selection effects would be difficult, and migration rates are low enough to suggest that the required effort might not be worthwhile. Another problem is that we cannot separate u and ε using Census data, and there are not enough observations in the NLSY to get reliable estimates of wage distributions for each State. We deal with this by appealing to results from previous research indicating that the transient earnings component is responsible for about one-third of the variance of earnings.[13]

The wage distribution in State j is modeled as a 3-point approximation with support points $a_j(s) = \hat{\mu}_j + \hat{u}(q_s)$, s $\in$ {1,2,3}, where $\hat{\mu}_j$ is the estimated State effect, and $\hat{u}(q_s)$ is the q_s quantile of the estimated distribution of u, with $\{q_s\} = \left\{\frac{1}{6}, \frac{1}{2}, \frac{5}{6}\right\}$. This wage distribution refers to earnings at some standard age, which we take to be 30. PUMS data from the 1990 Census are used to estimate wage

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

distributions for each State (because the sample size in the NLSY data is not sufficient for this purpose). Wages are adjusted for cost of living differences using the ACCRA index. These State wage distributions are tabulated in Appendix A.

Consider an individual who is in the same location for m years. The average wage over this period, excluding State and age effects, is

$$\tilde{w}_{ij} = \frac{1}{m}\sum_{t=1}^{m}\left(w_{ij}(a_t) - \phi(a_t)\right) - \mu_j = \eta_i + u_{ij} + \frac{1}{m}\sum_{t=1}^{m}\varepsilon_i(a_t)$$

This implies

$$\mathrm{Var}\left(\tilde{w}_{ij}\right) = \sigma_\eta^2 + \sigma_u^2 + \frac{\sigma_\varepsilon^2}{m}$$

Suppose that one-third of the total variance is due to the transient component, and let ρ be the proportion of the remaining variance accounted for by η. Then $\sigma_\varepsilon^2 = ½(\sigma_\eta^2 + \sigma_u^2)$ and $\sigma_\eta^2 = \rho(\sigma_\eta^2 + \sigma_u^2)$, so the match component u can be estimated by the signal-extraction formula

$$\hat{u}_{ij} = \frac{\sigma_u^2}{\sigma_\eta^2 + \sigma_u^2 + \frac{1}{m}\sigma_\varepsilon^2}\tilde{w}_{ij} = \frac{1-\rho}{1+\frac{1}{2m}}\tilde{w}_{ij}$$

We use this estimate to determine individual i's position in the 3-point wage distribution for location j. If û falls in the top third of the distribution, then the high wage is assigned; if û falls in the bottom third then the low wage is assigned, and otherwise the median wage is assigned. In other words, for each State in which we have earnings data on an individual, we approximate that individual's permanent wage in that State as one of three wage numbers in the appropriate row of Appendix Table A1.

5.2 Partial Likelihood Estimates

We condition on the estimated earnings distributions for each State and maximize the partial likelihood to obtain estimates of the behavioral parameters. We set β =.9, T = 40, and ρ = ½. We show in section 5.8 below that our main results are not very sensitive to these parameter settings.

Table 2: Interstate Migration of Young White Men 12 Years of Schooling			
Disutility of Moving (γ_0)	7.0642	7.0108	6.4083
	0.0513	*0.0513*	*0.1111*
Distance (γ_1) (1000 miles)			0.5210
			0.0760
Home Premium (κ)			0.3554
			0.0175
Previous Location (γ_2)			3.1624
			0.1492
Population (γ_3) (moving cost)			0.8284
			0.0871
"Real" Income (α) ($10,000)		0.1824	0.2477
		0.0353	*0.0572*
Loglikelihood	-3209.87	-3193.78	-2471.87
$\chi^2(1)$		32.186	
Moving Cost		$384,363	$258,727
Observations	9,682		
Moves	397		

Notes:
Estimated asymptotic standard errors are given in italics below the coefficients.
The length of the horizon is 40 years, with discount factor $\beta = .9$
The wage distributions have 3 points of support.
Distances are measured between State population centroids (in thousands of miles).
Population is measured in units of 10 million people.

Table 2 shows that differences in expected income are a significant determinant of migration decisions for this population. There are 9,682 person-years in the data, with 397 interstate moves. This is an annual migration rate of 4.1%, and the first column in Table 2 matches this rate by setting the probability of moving to each of J-1 locations to a constant value, namely $\frac{1}{J-1}\frac{397}{9{,}682}$, with J = 51.[14] The next columns show that population size, distance, and home and previous locations all have highly significant effects on migration.[15] The last column shows the effect of income, controlling for these other effects, using wages adjusted for cost of living differences across States.[16] These estimates are interpreted in the following subsections.

5.3 Moving Costs and Preference Shocks

Since utility is linear in income, we can translate the estimated moving cost into a dollar equivalent. This gives $\delta_0/\alpha =$ \$258,727 (using the estimates in the last column of Table 2), with the interpretation that the compensation needed to just offset the cost of a move is very large: other things equal, a lump-sum of about \$250,000 would be needed to fully compensate someone for the costs of a move.[17] Another way to interpret our result is to note that a \$10,000 migration subsidy (modeled as a reduction in δ_0 such that δ_0/α falls by \$10,000 with the other parameters held fixed) would increase the interstate migration rate from 4.1% to 5.75% .

It may seem that th large moving cost is an artifact of the specification of the model. For example, in the absence of any moving cost, allowing preference shocks to be drawn randomly over J locations implies a migration probability of (J-1)/J, so that with J = 51, nearly everybody moves every period. The first column of Table 2 shows how large the fixed cost of moving has to be in relation to the preference shocks, in order to reduce the migration rate from 50/51 to the observed rate of 4.10%, when all other influences on migration are suppressed. The last column shows that the estimated moving cost is large in relation to the income coefficient, even after allowing for the effects of population and distance and the home premium and previous location.

---

[14]In other words the estimate of δ_0 solves the equation $\frac{1}{e^{\delta_0}+J-1} = \frac{1}{J-1}\frac{397}{9{,}682}$; the solution is $\delta_0 = \log(464250) - \log(397)$.

[15]The χ^2 statistics in the table are for likelihood ratio tests of the form $2\log(L^U/L^R) \sim \chi^2(r)$, where r is the number of restrictions embodied in L^R relative to L^U.

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

[17]This refers to the cost of moving to a new location, ignoring the effect of population and distance. In the case of a return move, the estimated moving cost is \$131,048. The estimated cost of moving 1,000 miles to a State with a population of 3 million is \$269,726.

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

$$\begin{aligned} \frac{1-\lambda_L}{\lambda_L} &= e^{-\delta_0}\left[J_L - 1 + J_H e^{\beta \Delta V}\right] \\ \frac{1-\lambda_H}{\lambda_H} &= e^{-\delta_0}\left[J_H - 1 + J_L e^{-\beta \Delta V}\right] \end{aligned} \qquad \textbf{(24)}$$

where λ_L is the probability of staying in a low-income location and J_L is the number of such locations, and similarly for λ_H and J_H, and where ΔV is the difference in expected continuation values between the low-income and high-income locations, which is determined by the equation

$$e^{\Delta V} = \frac{e^{\alpha \Delta y}\left[J_L + \left(J_H - 1 + e^{\delta_0}\right) e^{\beta \Delta V}\right]}{J_L - 1 + e^{\delta_0} + J_H e^{\beta \Delta V}} \qquad \textbf{(25)}$$

For example, if $\beta = 0$, then $\Delta V = \alpha \Delta y$, and if moving costs are prohibitive ($\exp(-\delta_0) \approx 0$), then $\Delta V = \alpha \Delta y/(1-\beta)$.

These equations identify α and δ_0 (these parameters are in fact over-identified, because there is also information in the probabilities of moving to the same income level). If $\delta_0 < \beta \Delta V$, 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 δ_0 a little bigger than $\beta \Delta V$, and letting both of these be large in relation to the preference shocks, the probability of moving from the low-income location can be made small. But then the probability of moving from the high-income location is almost zero, which is not true in the data. In other words, if the probability of moving from a high-income location is not negligible, then the preference shocks cannot be negligible, since a preference shock is the only reason for making such a move.

By making both δ_0 and $\beta \Delta V$ large, the ratio can be made arbitrarily close to 1 while preserving a fixed difference. But making them large kills the effect of the preference shocks. There must be a positive difference to explain why there is a strong tendency to stay in low-income locations. The upshot is that the moving cost has to be a relatively large multiple of the difference in continuation values. For example, if Δy = \$5,000, then $\beta \Delta V/\alpha$ could be as large as \$45,000, and the moving cost must be higher than this. But this doesn't explain why the estimate is over \$250,000.

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

5.4 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 age, duration in the current location, and the number of moves made previously. In addition, there are of course unobserved characteristics that might make some people more likely to move than others. Thus it is important to check how well the model fits the data. In particular, since the model pays very little attention to individual histories, one might expect that it would have trouble fitting panel data.

One simple test of goodness of fit can be made by comparing the number of moves per person in the data with the number predicted by the model. As a benchmark, we consider a binomial distribution with a migration probability of 4.1% (which is the number of moves per person-year in the data). Table 3 shows the predictions from this model: about 72% of the people never move, and of those who do move, about 16% move more than once.[18] The NLSY data are quite different: more than 80% never move, and about 44% of movers move more than once. A natural interpretation of this is mover-stayer heterogeneity: some people are more likely to move than others, and these people account for more than their share of the observed moves. We simulated the corresponding statistics for the model by starting 100 replicas of the NLSY individuals in the observed initial locations, and using the model (with the estimated parameters shown in Table 2) to generate a history for each replica, covering the number of periods observed for this individual. The results match the data very well: although the proportion of people who

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

never move is slightly below the observed proportion, the proportion of movers who move more than once matches the data very closely. In this respect, the observables in the model do a good job of accounting for the heterogeneous migration probabilities in the data.[19]

Table 3: Goodness of Fit						
Moves	**Binomial**		**NLSY**		**Model**	
None	887.20	72.48%	986	80.56%	96229	78.62%
One	282.99	23.12%	133	10.87%	14685	12.00%
More	53.81	4.40%	105	8.58%	11486	9.38%
Proportion of movers with more than one move	15.98%		44.12%		43.89%	
Total observations	1224		1224		122400	

Return Migration

Table 4 summarizes the extent to which the model can reproduce the return migration patterns seen in the data (the statistics in the Model column refer to the simulated data set used in Table 3).

[19]We have not estimated models with unobserved heterogeneity, because even the simplest specification doubles the size of the state space and introduces a difficult initial conditions problem, and because there is no particular reason to believe that our main results are sensitive to unobserved heterogeneity. As a rough check, we simulated migration histories for a heterogeneous population and estimated our (mis-specified) model on these data. Heterogeneity was introduced by mixing two sub-samples with different moving costs. We tried several experiments along these lines, with similar results: neglecting unobserved heterogeneity in moving costs introduces a negligible bias in the estimated coefficients. Most importantly, we find that the estimated effect of income is, if anything, slightly underestimated, indicating that models with unobserved heterogeneity are likely to strengthen our conclusion that migration decisions are sensitive to differences in income prospects.

Table 4: Return Migration Statistics		
	NLSY	**Model**
Proportion of Movers who		
Return home	34.3%	31.3%
Return elsewhere	6.5%	5.0%
Move on	59.2%	63.6%
Proportion who *ever*		
Leave Home	15.5%	14.5%
Move from not-home	45.1%	59.3%
Return from not-home	33.3%	27.3%

The model attaches a premium to the home location, and this helps explain why people return home. For example, in a model with no home premium, one would expect that the migration flow to any particular location would be roughly $\mu/(J-1)$, where μ is the average migration rate. Given $\mu = .0410$ and $J = 51$, this obviously does not match the observed return rate of 34%. The home premium also reduces the chance of initially leaving home, although this effect is offset by the substantial discount on the cost of returning to a previous location (including the home location): leaving home is less costly if a return move is relatively cheap.

The return migration in the simulated data matches the actual data reasonably well. The main discrepancy is that the model substantially over-predicts the proportion who ever move from an initial location that is not their home location. That is, the model has difficulty 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 presumably more likely to put down roots the longer they stay in a location, and our model cannot capture this kind of duration dependence.

5.5 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 5 replicates this finding for our sample of high-school men. A standard human capital explanation for this age effect is that migration is an investment: if a higher income stream is available elsewhere, then the sooner a move is made, the sooner the income increase is realized. Moreover, since the worklife is finite, a move that is worthwhile for a younger worker might not be worthwhile for an older worker, because there is less time for the higher income stream to offset the moving cost (Sjaastad [1962]). In other words, migrants are more likely to be young for the same reason that students are more likely to be young.

Table 5

Annual Interstate Migration Rates by Age and Current Location

	All		Not At Home[a]		At Home	
Age	N	Migration Rate	N	Migration Rate	N	Migration Rate
20	817	0.050	101	0.228	716	0.025
21	907	0.052	102	0.206	805	0.032
22	931	0.048	116	0.207	815	0.026
23	915	0.043	131	0.153	784	0.024
24	942	0.051	146	0.171	796	0.029
25	895	0.053	142	0.141	753	0.036
26	888	0.045	154	0.175	734	0.018
27	862	0.034	136	0.125	726	0.017
28	706	0.035	123	0.114	583	0.019
29	595	0.018	109	0.064	486	0.008
30	483	0.021	84	0.083	399	0.008
31	340	0.026	51	0.118	289	0.010
32	228	0.018	37	0.027	191	0.016
33	130	0.015	12	0.000	118	0.017
34	43	.000	5	0.000	38	0.000
All	9,682	0.041	1,449	0.146	8,233	0.022

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

This explanation for age effects has two parts, and our model deals with the first part, but not the second. We assume an infinite horizon, so that the decision problem is stationary. This assumption is made for tractability: in a finite-horizon model, age is a state variable, and so the size of the state space increases dramatically.[20] Given workers of different ages who otherwise have the same migration and wage histories, the infinite-horizon model makes the same prediction: the age difference is irrelevant, according to the model. Nevertheless, the model can potentially explain why younger workers are more likely to move. For example, consider two locations paying different wages, and suppose that workers are randomly assigned across these locations at birth. Then 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, and the migration rate will be unrelated to age. This implies that the migration rate must be higher when workers are young.[21]

The second part of 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 propensities between young adult workers and those within sight of retirement. But the workers in our sample are all in their twenties or early thirties, and the prospect of retirement seems unimportant for such workers. Indeed, that is why the infinite-horizon assumption seems like a reasonable approximation for the population that we are studying. This suggests that the first part of the human capital explanation must be the dominant force explaining why migration rates for 30-year-olds are substantially lower than for 25-year-olds. In other words, if the human capital explanation is correct, our infinite-horizon model should be able to capture the relationship between migration rates and age.

One way to examine this question is to ask whether our model fits equally well for younger and older workers. Table 6 shows that it does not: although there are no dramatic differences in the parameter estimates for younger and older workers, a likelihood ratio test decisively rejects the hypothesis that the parameters are equal. In principle, this difference might be explained by the difference in horizon, but the last column of the table shows that reducing the horizon of the older subsample by 10 years has a negligible effect on the results. This suggests that the human capital model does not give an adequate explanation of the relationship between age and migration rates.[22]

---

[20]Marriage is another important factor, but in order to deal with this we would have to double or triple the size of the state space (depending on whether we distinguished between divorced and single people).

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

[22]To analyze this, the model must be extended to include age as a state variable. Although this is beyond the scope of the current paper, the extension might not be as difficult as it seems. We compute the infinite horizon model by iterating on the value function, starting from zero, and continuing through T iterations. This algorithm is known to converge to the infinite-horizon value function, so if T is large, additional iterations leave the value function approximately unchanged. In practice, T = 40 is large enough for this purpose. What this means it that the

Table 6: Age Differences in Migration Rates

White Men, High School Education

	All Ages	20-25	26-34	26-34 (T=30)
Disutility of Moving	6.4083	6.3112	6.5155	6.5164
	0.1111	*0.1376*	*0.2031*	*0.2029*
Distance (1000 miles)	0.5210	0.4247	0.7435	0.7434
	0.0760	*0.0930*	*0.1375*	*0.1379*
Home Premium	0.3554	0.2702	0.3310	0.3333
	0.0175	*0.0300*	*0.0269*	*0.0269*
Previous location (moving cost)	3.1624	3.3394	3.1564	3.1601
	0.1492	*0.1916*	*0.2645*	*0.2647*
Population	0.8284	0.8328	0.7322	0.7320
	0.0871	*0.1063*	*0.1610*	*0.1607*
Real Income (ACCRA)	0.2477	0.3230	0.2723	0.2736
	0.0572	*0.0771*	*0.0966*	*0.0972*
Loglikelihood	-2471.870	-1625.881	-821.327	-821.342
		-2447.209		
$\chi^2(6)$ p-value	49.32	0.00000		
Moving cost	$258,713	$195,393	$239,244	$238,172
N (person-years)	9,682	5,407	4,275	4,275
Moves	397	267	130	130
Migration rate	0.0410	0.0494	0.0304	0.0304

algorithm computes the value function for someone who has 40 years left before retirement, but as a by-product it also computes the value functions for someone with t years to retirement, for any t (the first iteration gives the value function with one year left, and the t^{th} iteration gives the value function with t years left).

5.6 Decomposing the Effects of Income on Migration Decisions

In our model, differences in wage distributions across States are due entirely to differences in State means. This raises the question of whether the estimated coefficients would be similar if wage dispersion is ignored, and migration decisions are modeled as responses to differences in mean wages across locations. The results of this exercise are shown in Table 7, 9, 9. Surprisingly, the estimate of the income coefficient (α) is insignificant in this specification. Going to the other extreme, we specified the wage distribution at the national level, with no variation across States. This restores the positive estimate of α. Evidently, our results are not driven by differences in mean wages across States. We turn next to an analysis of how the data manage to generate a significant income coefficient, even when the variation in incomes across States is suppressed.

Table 7: Alternative Income Specifications			
	Census	State Means	National
Disutility of Moving	6.4083	6.3999	6.4324
	0.1111	*0.1099*	*0.1112*
Distance	0.5210	0.5338	5.3347
(1000 miles)	*0.0760*	*0.0771*	*0.0758*
Home Premium	0.3554	0.3506	0.3595
	0.0175	*0.0170*	*0.0177*
Previous Location	3.1624	3.0311	3.1947
(moving cost)	*0.1492*	*0.1449*	*0.1506*
Population	0.8284	0.8731	0.8559
(moving cost)	*0.0871*	*0.0878*	*0.0868*
Real Income	0.2477	0.0482	0.3270
(ACCRA)	*0.0572*	*0.0880*	*0.0666*
Loglikelihood	-2471.87	-2481.074	-2470.069
N (person-years)	9,682		
Moves	397		

Notes:
The "State Means" column assumes that there is no wage dispersion within States. The "national" column assumes that wage distributions are identical in all States.

5.7 Movers and Stayers

A useful decomposition of the likelihood can be obtained by separating the decision on whether to move from the decision on where to go, conditional on moving. The likelihood that location j is chosen when the current location is ℓ can be written as

$$p_j = Pr(d(j)=1 \mid h,\ell,\omega;\theta) = \frac{\xi_j}{\sum_{\tau=1}^{J} \xi_\tau}$$

Then p_ℓ is the probability of staying, and $1\text{-}p_\ell$ is the probability of moving. The probability of choosing location j can be factored as

$$p_j = \tilde{p}_j \left[1 - p_\ell \right]$$

where $\tilde{p}_j$ is the probability of choosing j, conditional on moving:

$$\tilde{p}_j = \frac{\xi_j}{\sum\limits_{\tau \neq \ell} \xi_\tau}$$

The parameters governing migration decisions can be estimated using only the move/stay probabilities, and they can also be estimated using only the conditional destination choice probabilities. This helps pin down the source of the results. Each observation adds $\log(p_j)$ to the full loglikelihood, where j is the chosen location. In the case of a move, the loglikelihood in the mover-stayer model is counted as $\log(1\text{-}p_\ell)$, while the destination choice model counts the loglikelihood of each observation as $\log(p_j) - \log(1\text{-}p_\ell)$. Thus, for given parameter values, the sum of the loglikelihoods for the mover-stayer model and the destination choice model must be the same as the loglikelihood of the full model.[23]

The results of this decomposition are shown in Table 8. Since the destination choice data contain very little information on the fixed cost of moving, this parameter was held fixed in the destination choice model. Table 8 shows that the positive income coefficient appears in the decision to move, but not in the choice of destination. The move/stay model shows that the probability of moving is higher when the income realization in the current location is bad. The result for the destination choice model indicates that high-wage States are *not* more likely to be chosen.

[23]Note that the move-stay model accounts for the full set of alternative destination choices: there is no need to choose a “representative” alternative, as in Gelbach (2002), for example. The continuation value for each alternative location is evaluated using location-specific data, and a move occurs if the continuation in *some* alternative location beats the value of the incumbent location (given the current realization of the vector of preference shocks).

Table 8: Movers and Stayers White Men, High School Education				
		Full Model	Move-Stay	Destination Choice
Disutility of Moving		6.4083	7.6490	6.4083
		0.1111	*0.2822*	*fixed*
Distance (1000 miles)		0.5210	-0.1742	0.6685
		0.0760	*0.1033*	*0.0846*
Home Premium		0.3554	0.2702	0.4573
		0.0175	*0.0300*	*0.0286*
Previous location (moving cost)		3.1624	5.5976	2.3031
		0.1492	*0.3251*	*0.1671*
Population		0.8284	-0.2629	1.1064
		0.0871	*0.1444*	*0.1027*
Real Income (ACCRA)		0.2477	0.3504	-0.0768
		0.0572	*0.0805*	*0.0881*
Loglikelihood		-2471.870	-1429.159	-995.805
			-2424.965	
$\chi^2(6)$	p-value	93.81	0.00000	
N (person-years)		9,682		
Moves		397		

Another result is that the model fails the specification test associated with the decomposition of the likelihood into move-stay and conditional destination choice components: the parameter estimates differ significantly across these two components. In particular, the home premium seems to have a bigger effect on destination choices than on decisions about whether to stay in the home location.

5.8 Sensitivity Analysis

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

The results presented so far are based on wages that are adjusted for cost of living differences across locations. If these cost of living differences merely capitalize the value of amenity differences, then unadjusted wages should be used to measure the incentive to migrate. Results for this specification are given in the fourth column of Table 9: the estimate of α is reduced by about 20%, with little effect on the other coefficients, and the likelihood is lower. Thus in practice the theoretical ambiguity as to whether wages should be adjusted for cost of living differences does not have much effect on the empirical results: either way, income shows up as a significant determinant of migration decisions.

Table 9 also shows that the results are not sensitive to variations in how distance and location size affect migration. As was discussed in Section 4, size (as measured by population) may affect migration either as a scaling factor on the preference shocks, or as a variable affecting the cost of migration. The results in the last column of Table 9 show that allowing population to enter as a scaling factor on the preference shocks adds virtually nothing to the basic specification. We also expanded the moving cost specification to allow quadratic effects of distance and location size; this has little effect on the results.

The other alternative specifications in Table 9 are concerned with sensitivity of the estimates to the discount factor (β), the horizon length (T) and the proportion of the residual permanent wage variance attributed to individual effects that are fixed across locations (ρ). Increasing β to .95 has a noticeable effect on the utility flow parameters (i.e. the home premium and the income coefficient), with hardly any effect on the moving cost parameters. Although a 5% annual real interest rate is arguably more plausible than the 10% rate assumed in our baseline specification, the likelihood when β is set at .95 is substantially lower.[24] Reducing T from 40 to 20 has very little effect (as might be expected with β =.9). Large changes in ρ lead to modest changes in α: increasing the relative importance of location match effects (i.e. decreasing ρ) yields some improvement in the likelihood, and a somewhat lower estimate of α.[25]

---

[24]The maximum likelihood is estimate of β is around .84, but β = .9 is easily accepted by a likelihood ratio test.

[25]In principle, ρ can be estimated using the NLSY data, because the autocovariance of wages includes σ_u^2 for stayers, but not for movers. The best estimate of ρ obtained from the wage covariogram is about .75, but this estimate is fragile, and a smaller value of ρ gives a higher value of the likelihood in the migration model. Although joint estimation of ρ and the other parameters is feasible, we have not pursued this because the results in Table 9 suggest that it would not be very informative.

Table 9: Alternative Specifications									
	Base	β =.95	β =.85	No Cola	Quadratic	ρ =.25	ρ =.75	T=20	AltPop
Disutility of Moving	6.4083	6.4185	6.4128	6.3957	6.4427	6.4124	6.4009	6.4158	6.3663
	0.1111	*0.1093*	*0.1073*	*0.1104*	*0.1566*	*0.1112*	*0.1107*	*0.1111*	*0.1197*
Distance	0.5210	0.4796	0.4474	0.5211	0.5897	0.5230	0.5210	0.5213	0.5112
(1000 miles)	*0.0760*	*0.0696*	*0.0654*	*0.0766*	*0.2079*	*0.0758*	*0.0765*	*0.0766*	*0.0763*
Squared Distance	------	------	------	------	-0.0373	------	------	------	------
					0.0920				
Home Premium	0.3554	0.2482	0.1962	0.3513	0.3543	0.3581	0.3510	0.3668	0.3553
	0.0175	*0.0127*	*0.0104*	*0.0172*	*0.0177*	*0.0177*	*0.0173*	*0.0179*	*0.0181*
Previous Location	3.1624	3.0699	2.9518	3.0843	3.1521	3.1815	3.1095	3.1797	3.1743
(moving cost)	*0.1492*	*0.1451*	*0.1396*	*0.1466*	*0.1507*	*0.1498*	*0.1474*	*0.1499*	*0.1495*
Population	0.8284	0.7678	0.7300	0.8431	1.0042	0.8313	0.8337	0.8342	0.7678
(moving cost)	*0.0871*	*0.0827*	*0.0794*	*0.0876*	*0.2582*	*0.0869*	*0.0874*	*0.0869*	*0.1012*
Squared Population	------	------	------	------	-0.0829	------	------	------	-------
					0.1127				
Population	------	------	------	------	------	------	------	------	0.0193
(preference shocks)									*0.0172*
Income	0.2477	0.1672	0.3293	0.1988	0.2444	0.1963	0.2690	0.2568	0.2408
	0.0572	*0.0383*	*0.0770*	*0.0703*	*0.0572*	*0.0422*	*0.0785*	*0.0595*	*0.0582*
Loglikelihood	-2471.87	-2476.90	-2471.34	-2477.73	-2471.48	-2470.42	-2475.74	-2472.13	-2471.23

Notes: The base specification assumes β = .9, T = 40 and ρ = .5.

6 Conclusion

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

Our empirical results show a significant effect of expected income differences on interstate migration, for white male high school graduates in the NLSY. On the other hand we find little evidence of migration in response to wage differentials across States. Instead, 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 driven primarily 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.

Our estimates indicate that moving costs are very large. For example, if we ignore differences due to distance and location size and the home location effect, we estimate that about a quarter of a million dollars would be needed to fully compensate for the costs of an interstate move. But if moving costs were fully compensated, nearly everyone would move all the time. Perhaps a more informative statement is that we estimate that a $10,000 migration subsidy would increase the interstate migration rate by about 40%.

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. The main discrepancy between the model and the data arises from a stationarity assumption that precludes the use of age as a state variable. The development of a model that relaxes this assumption is a promising area for further research.

References

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Appendix A: Wage Distributions

Table A1 shows the three-point approximation to the state earnings distributions derived from the PUMS data. Earnings are expressed in 1983 dollars, adjusted for cost of living differences using the ACCRA index (http://www.coli.org/). We used unweighted city averages within States from one quarter in 1979, 1981, 1987, and 1990 to calculate the index. The second column reports the number of observations in the PUMS for each State. Earnings values for the low, medium and high cells appear in the 16%, 50% and 83% columns. The ⅓ and ⅔ quantile values define the cell boundaries.

Table A1: State Earnings Distributions
White Male High School Graduates

		Percentiles of Earnings Distribution				
State	Nobs	16%	33%	Median	67%	83%
Alabama	5127	15548	16712	17674	18733	20166
Alaska	1016	14784	15948	16910	17969	19402
Arizona	3986	13747	14910	15872	16932	18365
Arkansas	3969	13771	14935	15897	16957	18389
California	27983	15635	16799	17761	18821	20254
Colorado	4828	14421	15585	16547	17606	19039
Connecticut	4881	15932	17096	18058	19117	20550
Delaware	1083	15820	16984	17946	19005	20438
DC	90	12011	13174	14136	15196	16629
Florida	17080	14446	15609	16571	17631	19064
Georgia	9109	16143	17307	18269	19328	20761
Hawaii	1063	11666	12830	13792	14851	16284
Idaho	1544	14427	15591	16553	17612	19045
Illinois	16658	16060	17224	18186	19245	20678
Indiana	11491	15998	17162	18124	19183	20616
Iowa	5108	14275	15438	16400	17460	18893
Kansas	4142	14775	15939	16901	17960	19393
Kentucky	6987	14433	15597	16558	17618	19051
Louisiana	5411	15260	16424	17386	18446	19878
Maine	2869	15249	16412	17374	18434	19867
Maryland	6671	17539	18703	19665	20724	22157
Massachusetts	9882	15824	16988	17950	19009	20442
Michigan	15702	14834	15998	16960	18020	19453
Minnesota	8103	14627	15790	16752	17812	19245
Mississippi	2589	14585	15749	16711	17770	19203
Missouri	9285	14918	16081	17043	18103	19536
Montana	1188	13190	14354	15316	16376	17809
Nebraska	2572	13766	14929	15891	16951	18384
Nevada	2157	15756	16920	17882	18941	20374
New Hampshire	2292	13749	14913	15875	16934	18367
New Jersey	11183	16588	17751	18713	19773	21206
New Mexico	1949	12508	13672	14634	15694	17127
New York	23896	15329	16492	17454	18514	19947
North Carolina	10021	14817	15981	16943	18002	19435
North Dakota	930	12766	13930	14892	15951	17384
Ohio	20932	15363	16526	17488	18548	19981
Oklahoma	4617	13623	14786	15748	16808	18241
Oregon	4117	14246	15410	16372	17431	18864
Pennsylvania	25366	14851	16015	16977	18036	19469
Rhode Island	1602	15336	16499	17461	18521	19954
South Carolina	4230	15585	16749	17711	18770	20203
South Dakota	1077	12699	13862	14824	15884	17317
Tennessee	8052	14802	15966	16928	17987	19420
Texas	20624	14588	15751	16713	17773	19206
Utah	2208	15825	16989	17951	19010	20443
Vermont	1254	13766	14930	15892	16951	18384
Virginia	9097	16124	17287	18249	19309	20742
Washington	7417	16018	17182	18144	19203	20636
West Virginia	3524	13279	14443	15405	16464	17897
Wisconsin	10503	16109	17273	18235	19294	20727
Wyoming	806	15947	17111	18073	19132	20565

Appendix B: Validation of ML Estimates

Table B: ML Estimates Using Simulated Data		
	NLSY	Simulated
Disutility of Moving	6.4083	6.4040
	0.1111	*0.0121*
Distance (1000 miles)	0.5210	0.5277
	0.0760	*0.0073*
Home Premium	0.3554	0.3560
	0.0175	*0.0021*
Previous Location	3.1624	3.1807
	0.1492	*0.0174*
Population (moving cost)	0.8284	0.8322
	0.0871	*0.0087*
"Real" Income ($10,000)	0.2477	0.2497
	0.0572	*0.0056*
Loglikelihood	-2471.870	-247440.93
Moving Cost	$258,727	$256,440
Observations	9,682	968,200
Moves	397	39,700

Explanation:
The ML 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 allowing the model to choose the sequence of locations. Two alternative starting points were used when estimating the parameters from the simulated data: the actual parameter values used to generate the data, and the one-parameter estimate from the first column of Table 2 (with all other parameters set to zero). For both starting points the estimates converged to the values shown in the last column above (this required 2 Newton steps starting from the truth, and 9 Newton steps starting from the one-parameter estimate).