Preliminary and Incomplete

Geographical Wage Differentials, Welfare Benefits and Migration

John Kennan and James R. Walker[1]

University of Wisconsin-Madison and NBER

July 2000

[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 (Kennan) and the NICHD (Walker) provided research support.

1 Introduction

Geographical wage differentials are large and persistent, despite large migration flows. There are also large geographical differences in welfare benefits, and policy-makers express concerns that these differences might create "welfare magnets" in some locations. We are interested in the extent to which migration flows can be explained by differentials in wages and welfare benefits.

We model individual decisions to migrate as a job search problem in which welfare benefits or other alternative sources of income act as a floor, insuring workers against bad job search outcomes. This differs from the standard job search model in which unemployment benefits are treated as a subsidy received while search continues. In our model, welfare provides a safety net in case the search fails. A worker can draw a wage only by visiting a location, thereby incurring a moving cost. Locations are distinguished by known differences in mean wages, 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. There is a two-dimensional ranking of locations, *ex ante*: some places have high mean wages, and others have attractive fallback options (both adjusted for amenity values). In addition we allow for a bias in favor of the home location.

The decision problem is too complicated to be solved analytically (although the well-known Gittins index results can be extended to obtain a partial characterization of the optimal policy). We proceed by using a discrete approximation that can be solved numerically, following Rust (1994). The parameters of the model include a discount factor, a risk aversion coefficient and a home premium summarizing individual preferences; moving costs, including a fixed cost and a cost that is proportional to distance; means and variances of wages in each location; a relative variance parameter governing the extent to which individual wage offers are correlated across locations and a persistence parameter governing the relative importance of permanent and transitory components of wages. The model is used to interpret migration patterns found in the NLSY, with particular attention to return and repeat migration in the early stages of the life cycle.

2 Geographical Wage Differentials and Migration Flows

We begin with some descriptive statistics on wage differentials across metropolitan areas in the United States, using data for white workers from the Census Public Use Microdata Series (PUMS)[2]. We use a 3-level grouping of education (high school, some college, and college plus); individuals still enrolled in school and those with less than a high-school education are excluded. Because we are primarily interested in migration decisions of workers early in the life-cycle, we use 10-year age groups for each educational level (high school: ages 19-28; some college: ages 21-30; and college plus: ages 23-32). We also impose two restrictions on labor supply: (i) average hours per week must be more than 35 hours; and (ii) weeks worked must be fifty or more. We have yet not made any adjustment for cost of living differences between cities. Given that these differences to some extent merely capitalize differences in amenity values, the appropriate adjustment is not obvious, but no adjustment is surely not the right answer. One possibility, following the recommendations of the National Research Council (1995), is to adjust for differences in housing costs, on the grounds that these differences account for a large proportion of the variation in overall living costs, because the geographic variation in housing prices is large relative to the variation in other prices, and because about one third of the typical worker's income is spent on housing.

The SMSA wage distributions are summarized in Figures 1-3 and in Table 1. The dispersion of wages across locations is large.[3] Real wages fell for high-school graduates

---

[2] These data can be found at http://www.ipums.umn.edu. There are obvious sample size trade-offs among the definition of each locality and the fineness of the subpopulation group under consideration. The finest geographical unit recorded in the 1980 PUMS is the *county group* — a region with a population of 100,000 or more. Approximately 350 counties are uniquely identified in the 1980 PUMS as county groups. Yet the relationship between most counties and their county groups is more complicated: county groups may include several complete counties or parts of several counties, and may even cross state boundaries. To make matters worse, the geographical unit of the 1990 Census, the PUMA, although conceptually similar, is definitionally quite different from the county group. Connecting geographical regions below the state level between the 1980 and 1990 Censuses is a nontrivial task. Because of these complications we analyze wages across SMSAs.

[3] The dispersion of wages adjusted for differences in living costs is presumably less than the unadjusted dispersion, but it is unlikely that this would change the conclusion that geographical wage differentials are large. For example, the National Research Council (1995) computed housing cost indices for 9 regions and 5 city size classes, using the 1990 Census, and found that the index for the most expensive locations (metropolitan areas of more than 2.5 million people in the New England states or the Pacific states) was about 20% above the mean level,

between 1979 and 1989, while real wages for college graduates increased slightly.

Table 1: Summary Statistics

Median SMSA Wages by Education, 1979 and 1989

	Obs	Median	90/10	Min	10%	25%	75%	90%	Max
1980 Census									
High school	93	8.20	1.37	4.83	6.66	7.39	8.46	9.11	12.39
Some College	52	9.68	1.28	7.54	8.35	8.86	10.26	10.70	11.96
College	58	12.35	1.25	10.68	11.17	11.68	13.14	13.98	15.26
1990 Census									
High school	68	7.27	1.42	4.81	6.23	6.73	8.43	8.84	9.89
Some College	68	8.77	1.44	7.12	7.69	8.17	9.62	11.06	12.02
College	65	12.50	1.31	10.34	11.31	12.02	13.46	14.84	16.19

The table includes only those SMSAs with more than 50 observations in the relevant education group. The 1979 wages are converted to 1989 dollars using the CPI.

Figure 2 shows that places that have high wages for one education group generally have high wages for the other groups as well. Figure 3 shows that there is a strong positive correlation between location-specific wages in 1989 and in 1979, especially for college graduates. This persistence is important because in our model, migration decisions are made in response to known and stable differences in wage distributions across locations. Such a model is open to the obvious theoretical objection that if wages adjust rapidly, the driving force behind individual migration

---

and the index for the least expensive location (areas with less than 250,000 people in Mississippi and adjacent states) was about 20% below the mean level.

decisions will not be seen in the data. Our view is that wage adjustment is in fact slow enough so that the cross-section wage measures are adequate.[4]

2.1 Migration Patterns in the NLSY

Given that the census data show large and persistent geographical differences in wages, we turn next to panel data on migration decisions and wage outcomes. The basic empirical question is the extent to which people actually move for the purpose of improving their wage 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, we would expect to find that wage differentials play an important role in individual migration decisions.

The National Longitudinal Survey of Youth (NLSY) is one of the primary sources of longitudinal information on income, program participation and migration. Initially fielded in 1979, annually through 1994, and biannually since 1994, the NLSY was designed to be a nationally representative sample of American youth, ages 14 to 22 as of January, 1979. The survey contains an oversample of minorities and economically disadvantaged white youth living in the United States as well as a subsample of individuals serving in the military. Each wave obtained detailed information on earnings, income and assets, household composition, employment, marital history, training, educational status and attainment, and geographic residence. The survey records county and state (or country if outside the United States) for each residence at birth, at age 14, and for each residence for the first three waves of the NLSY (1979-1982), and annually (at the time of the survey) thereafter. We use the residence at the interview date to define migration.

Using data from the 1979-92 waves, we pick up respondents as they first leave school. We impose few selection rules for inclusion in our sample. For obvious reasons we exclude

---

[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. This suggests that our results should not be too sensitive to the use of a point-in-time wage measure. Similar findings were reported earlier by Barro and Sala-i-Martin (1991) and by Topel (1986).

observations from the military subsample. Otherwise, the primary reason for exclusion was missing information on either race or sex or geographic location in the two years following exit from school (those who never report leaving school or who leave school after 1990 are also excluded). Our sample contains 8,257 individuals.

Interstate migration flows are summarized in Table 2, which reports the number of moves in the first 5 years after leaving school, and also the number of moves in the full sample period. Clearly, young people move a lot, and college-educated people move a lot more than others.[5] But the main point of the table for our purposes is the importance of return migration, and especially migration back to the original location. Consider a person who has already moved, and who is moving again: what is the probability that such a person is moving back to a previous location, and in particular to the home location? The answer (in the bottom row of Table 2) for the home location (state) is at least 37%. We argue below that these flows are difficult to explain in a behavioral model. An interesting feature is that this homing instinct is much weaker for college graduates than for those with less education: for example, the proportion of high school graduates returning home is above 50%.

[5]Some people are merely returning home after leaving college: we have not yet tabulated these moves.

Table 2: Interstate Migration Flows, NLSY

	Less than High School		High School		Some College		College	
Horizon (years)	5	13	5	13	5	13	5	13
No. of people	1768		3534		1517		1435	
No. of movers	334	423	598	771	327	376	441	469
Repeat moves	239	434	313	653	167	264	196	261
Percent Movers	18.9	23.9	16.9	21.8	21.6	24.8	30.7	32.7
Moves Per Mover	1.7	2.0	1.5	1.8	1.5	1.7	1.4	1.6
Return Migration (% of all moves)								
Return - Home	22.9	24.0	20.6	24.1	16.0	17.5	12.4	13.4
Return - Else	5.4	12.4	2.7	7.2	2.8	5.9	2.5	3.3
Movers who return home (%)	39.2	48.7	31.4	44.5	24.2	29.8	17.9	20.9
Return-Home: % of Repeat	**54.8**	**47.5**	**60.1**	**52.5**	**47.3**	**42.4**	**40.3**	**37.5**

Finally, Table 3 provides some evidence on the effects of personal characteristics and demographic variables on the probability of moving within the first two years following school. In this table migration is defined as a move of more than 50 miles in either of the first two years, with return migrants counted as stayers. As one would expect, marriage and children exert strong negative influences on migration, especially for women.

Table 3: Logit Estimates

of the probability of moving more than 50 miles from the original location in the first two years after leaving school; white workers, at least high-school

	Men		**Women**	
	Coefficient	SE	Coefficient	SE
Married	-0.134	0.204	-0.392	0.169
Number of children	-0.350	0.270	-0.406	0.229
Weeks worked	-0.023	0.005	-0.016	0.004
Some College	0.900	0.191	0.586	0.177
College graduate (16+ years)	1.743	0.177	1.033	0.170
Postgraduate (17+ years)	1.691	0.223	0.834	0.269
Longitude	0.171	0.059	0.172	0.064
Latitude	-0.083	0.013	-0.049	0.013
Longitude squared ($/10^2$)	-0.001	0.030	-0.084	0.033
Constant	-9.607	3.002	-8.101	3.213
N	1841		1868	
$\chi^2(9)$	139.57		77.89	

Note: Individuals who did not live in an SMSA when they left school are excluded.

Thus the migration patterns in the NLSY can be summarized as follows: (1) young people move a lot, with more educated individuals moving more; (2) many of the repeated moves are return moves to the home location and (3) being married or having children are significant retardants to interstate migration, especially for women. The prevalence of repeat and return migration suggests that static analyses of migration will be flawed. In the next section we present a theoretical framework for analyzing these dynamic patterns.

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 not in other locations, and in order to determine the wage at each location, it is necessary to move there, at some cost. In each location there is also a fallback option, such as welfare or family support, and the value of this is known.

Suppose there are J locations, and the random part of individual i's wage W_{ij} in location j is the product of two components:

$$\log(W_{ij}) \equiv w_{ij} = \mu_j + \eta_i + \varepsilon_{ij}$$

where μ_j is the mean log wage in location j, η_i is an individual effect that is fixed across locations, and ε_{ij} is a matching effect reflecting the relative demand price of i's skill bundle in location j. We assume that η_i and $\{\varepsilon_{ij}\}_{j=1}^{J}$ are independent normal random variables with zero mean. The variance of η_i is σ_η^2, and the variance of ε_{ij} is σ_ε^2, for all j. The unconditional distribution of W_{ij}, before any wages have been drawn, is denoted by F_j (this is the above lognormal distribution with $\eta_i = 0$). The fallback option is b_j, and thus income in location j is $y_{ij} = \max\,[W_{ij}, b_j]$. In each period there is a chance that the current wage will be lost, and in that case the wage in the next period is a new draw from the distribution of η_i and ε_{ij}. We assume that the survival probability for the wage, ρ, is constant across wages and locations.

We specify the moving cost as an affine function of distance: $\Delta = \delta_0 + \delta_1 D(j,j')$ for $j' \neq j$, where $D(j,j')$ is the distance from j to j'. We assume that it is not possible to borrow money to finance a move (e.g. lenders are not keen to advance money that will be used to put borrowers out of their reach). We also rule out the possibility of accumulating money over several periods to finance a move. This means that the only way to finance a move from i to j is to reduce consumption in the current period by c_{ij}: with diminishing marginal utility, this means that the utility cost of moving is high at low income levels, as we shall see.

Migration decisions are made so as to maximize the expected discounted value of lifetime utility, with discount factor β. The flow of utility in location j is

$$u_j(y) = \left(1 + \kappa_j\right) \frac{(y - \Delta)^{1-\gamma} - 1}{1 - \gamma}$$

Thus we are assuming constant relative risk aversion, with coefficient $\gamma \geq 0$. The parameter κ_j captures some of the nonpecuniary differences across locations; in particular, by setting $\kappa_i > 0$ for some home location i, and $\kappa_j = 0$ otherwise, we can allow each individual to have a preference for their native location. The model is summarized by the eight parameters $(\gamma,\beta,\rho,\kappa,\delta_0,\delta_1,\sigma^2_\eta,\sigma^2_\varepsilon)$.

There is little hope of solving this problem analytically. In particular, the Gittins index solution of the multiarmed bandit problem cannot be applied because (1) there is a cost of moving and (2) wages are correlated across locations, and the Gittins index method does not work when either of these features is present. 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).

First we approximate the unconditional distribution F_j in each location by a discrete distribution over n points, as follows. Let $A^j_i = F^{-1}_j(i/n)$ and $a^j_i = F^{-1}_j(i/n - 1/2n)$. Then F_j is approximated by a uniform distribution over the set $\{a^j_i\}^n_{i=1}$. In practice, we set $n \leq 10$ so as to keep the state space tractable. For example, if n = 10, the approximation puts probability 1/10 on the 5^{th}, 15^{th}, ... 95^{th} percentiles of the distribution F_j. Next, we fix the support of the discrete approximation and vary the probabilities as new information comes in, so that the state space remains fixed over time. Specifically, if G_j is the conditional distribution over wages in location j after wages in some other locations have been observed, the discrete approximation puts probability $G_j(A^j_i) - G_j(A^j_{i-1})$ on each point a^j_i.

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 $\omega_j = i > 0$, then the wage in location j is known to be a^j_i , 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_i with probability $G_j(A^j_i) - G_j(A^j_{i-1})$, for $1 \leq i \leq n$.

In each period t, the wage history includes $M_t \leq t$ wage draws from different locations. Using the assumption of lognormality, the estimate of η is given by

$$\hat{\eta} = \frac{\lambda}{M_t} \sum_{k=1}^{M_t} \left[w_k - \mu_k \right] ; \quad \lambda \equiv \frac{1}{1 + \dfrac{\sigma_\varepsilon^2}{M_t \sigma_\eta^2}}$$

and the variance of the next wage draw is given by

$$\tau^2 = \left[1 + \frac{\lambda}{M_t} \right] \sigma_\varepsilon^2 = \frac{1}{\dfrac{M_t}{\sigma_\varepsilon^2} + \dfrac{1}{\sigma_\eta^2}} + \sigma_\varepsilon^2$$

with $\tau_0 = \sigma_\varepsilon^2 + \sigma_\eta^2$ initially. The posterior wage distribution puts probability p_s on the point $a_s = G^{-1}\left(\frac{2s-1}{2n} \right)$, where G is the lognormal distribution function with mean $\hat{\eta}$ and variance τ^2, and the probabilities are as follows

$$p_s = \text{Prob}\left[W_{ij} = e^{\mu_j + \tau_0 \Phi^{-1}\left(\frac{2s-1}{2n} \right)} \right] = \Phi\left(\frac{\tau_0 \Phi^{-1}\left(\frac{s}{n} \right) - \hat{\eta}}{\tau} \right) - \Phi\left(\frac{\tau_0 \Phi^{-1}\left(\frac{s-1}{n} \right) - \hat{\eta}}{\tau} \right)$$

where Φ is the standard normal distribution function.

The value function can be written in recursive form as

$$V(\ell,\omega) = \begin{cases} \displaystyle\sum_{s=1}^{n} p_s V(\ell, \omega_1, \ldots, \omega_{\ell-1}, s, \omega_{\ell+1}, \ldots, \omega_J) & \text{if } \omega_\ell = 0 \\ \max\limits_j \left[(1+\kappa_\ell) u\left(y_\ell(\omega_\ell) - \Delta_{\ell j} \right) + \beta\rho V(j,\omega) + \beta(1-\rho) V(j,0) \right] & \text{if } \omega_\ell > 0 \end{cases}$$

where $y_j(\omega) = \max[b_j, a_\omega^j]$. The optimal decision rule is

$$\alpha(\ell,\omega) = \underset{\mathrm{j}}{\text{argmax}}\, \hat{V}_j(\ell,\omega)$$

We compute V by value function iteration. It is convenient to use $V(\ell,\omega) \equiv 0$ as the initial estimate, so that if T is the number of iterations, the result gives the optimal policy for a T-period horizon; thus if T represents the length of the life-cycle, there is no need to check whether the iterations have converged to the infinite-horizon solution.

4 Moving Costs and the Gittins Index

We model migration decisions as the solution of a bandit problem that is extended to include moving costs and correlation across locations. The basic tool for the analysis of bandit problems without these extensions is the Gittens Index, which reduces the complicated dynamic programming problem discussed above to a simple calculation of reservation wages in each location, as if the other locations did not exist; the optimal policy is to move to the location with the highest reservation wage, and stop when a wage is found that beats the reservation wage in every location. Miller (1984) applied this to the study of occupational choice; McCall and McCall (1987) applied it to migration, but as they point out, the method works only if there is a zero cost of moving back to a previously visited location. Banks and Sundaram (1994) show that if every move involves some cost, there is no generally valid way to define an index with the Gittins properties. We ask a more relaxed question, and obtain a positive answer: can the Gittens index be used to simplify the calculation of the value function?

We present a preliminary result, and offer conjectures about generalizations. Suppose there are three locations, including a current location labeled 0, and two alternative locations i and j that have not yet been visited. A move to a new location incurs a cost M, and a return to a previous location incurs a cost m. Income in location 0 is known: call this y_0. Incomes in i and j are independent random variables. Moving to another location takes one period, and the discount factor is β per period.

Define the function $\psi_j(x)$ as follows:

$$\psi_j(x) \equiv \beta E \max\left(y_j, x\right) - M$$

This is the value of moving to search location j, with x as a reservation level: the interpretation is that x is available for sure, in case y_j turns out to be low. Note that $\psi_j'(x) = \beta F(x) < 1$, and $\lim_{x \to -\infty} \psi_j(x) = \beta E y_j - M$, so $\psi_j(x) > x$ for x sufficiently small, and $\psi_j(x) < x$ for x sufficiently big; thus the function ψ_j has a unique fixed point, ξ_j. Also $\psi_j(x) \geq x$ for $x \leq \xi_j$ and $\psi_j(x) \leq x$ for $x \geq \xi_j$. This is the standard Gittins index calculation. A similar calculation for location 0 would yield βy_0 - M, but instead define $\xi_0 = \beta y_0$ - m.

Lemma:

Suppose $\xi_0 = \max(\xi_0, \xi_i, \xi_j)$. Then it is optimal to remain in location 0.

Proof:

Suppose not. Without loss of generality say that it is optimal to move to location i initially. It will be shown that for any realization of y_i, the option of moving again to location j is dominated (either by the value of moving back to 0, or by the value of staying in i).

First, $\xi_0 \geq \xi_j$ implies

$$\xi_0 \geq \beta E \max\left(y_j, \xi_0\right) - M$$

This means that searching j with ξ_0 as the reservation level is dominated by ξ_0, which can be realized by an immediate return to location 0. Thus if it is optimal to move on to j, the reservation level must be the value of returning to location i, which is βy_i - m, and this must be larger than ξ_0, because in order for the move to be optimal it is necessary that $\psi_j(\beta y_i - m) \geq \xi_0$. Thus $\beta y_i - m \geq \xi_0 \geq \xi_j$. But this means that the value of moving to j with reservation level βy_i - m is dominated by βy_i - m, and staying in location i yields y_i, which is even better. Thus regardless of whether ξ_0 or βy_i - m is used as the reservation level, the option of searching location j is dominated (given an initial move to location i).

Now consider the initial move to i. Since j is dominated (by the argument just given), the result of a move to i will be either y_i or ξ_0. Since $\xi_0 \geq \xi_i$, the value of the move to i is dominated by ξ_0, and this in turn is dominated by y_0. Thus it is optimal to remain in location 0.

Conjectures

The above logic can potentially provide the basis for a systematic analysis of the optimal search policy in the presence of moving costs; the analysis might also be extended to allow some correlation in wages across locations, subject to the restriction that all locations are equally informative about wages in other locations (as is true in the simulations discussed in Section 4 above). We have the following conjectures.

4.1 If $\xi_i = \max(\xi_0, \xi_i, \xi_j) > y_0$, then it is optimal to move to location i.

4.2 The argument in the Lemma remains valid under risk aversion (just redefine everything in terms of utilities).

4.3 The argument can be extended by induction to cover an arbitrary number of locations.

5 Simulation Results

In order to make an initial assessment of whether the dynamic programming model is likely to be useful, we have computed optimal migration decisions using monthly wage and benefit data for five large states (California, Florida, Illinois, New York and Texas). The wage and benefit data are roughly the actual 1980 data for black women (all education levels), although these numbers are used only to illustrate how the model works on realistic data for a low-wage population. The percentiles shown are computed from lognormal distributions with a common variance and different means.[6]

Table 4 shows an optimal sequence of moves for a risk-neutral individual who begins in Florida, and who would take a large wage cut rather than move elsewhere ($\kappa = .5$). In this example, the wage has no common component ($\eta = 0$). Despite the large home premium, the optimal decision rule specifies a move unless the home wage is in the top three deciles, followed by onward migration unless the wage is in the top two deciles in the new location. This decision rule illustrates the difficulty of matching the homeward migration patterns in the NLSY (as shown in Table 2) when there is no correlation in wages across locations. The parameters are chosen to produce a large migration flow, even though the home location is 50% better than elsewhere. Moreover, since there are only 5 locations, the chances of a return to the origin are artificially high. Still, in a simulation of 50,000 cases using this decision rule, only 10% of all moves are associated with returning home.

[6] Mean log wages are (CA,FL,IL,NY,TX) = (6.16, 6.0, 6.296, 6.242 , 6.077), and the log standard deviation is .7766.

Table 4: Optimal Migration Decisions: Linear Utility, Large Home Premium

	5%	15%	25%	35%	45%	55%	65%	75%	85%	95%	Mean	b	Emax	Eu
FL (**Home**)	114	182	240	300	366	444	543	678	896	1432	520	347	575	78.4
	IL	IL	IL	IL	IL	IL	IL	Stay	Stay	Stay				
IL	152	243	322	402	492	598	731	915	1211	1940	701	412	754	82.8
FL=543	NY	NY	NY	NY	NY	NY	NY	NY	Stay	Stay				
NY	144	230	305	381	466	566	692	866	1146	1834	663	486	753	83.9
IL=915	CA	CA	CA	CA	CA	CA	CA	CA	Stay	Stay				
IL=731	CA	CA	CA	CA	CA	CA	CA	CA	Stay	Stay				
CA	131	211	280	351	429	522	639	801	1061	1705	613	542	746	84.6
IL=915, NY=866	IL	IL	IL	IL	IL	IL	IL	IL	Stay	Stay				
IL=731, NY=866	TX	TX	TX	TX	TX	TX	TX	TX	Stay	Stay				
TX	125	199	261	325	396	479	583	726	956	1515	557	277	581	76.7
IL=731, NY=866	NY	NY	NY	NY	NY	NY	NY	NY	Stay	Stay				
TX	125	199	261	325	396	479	583	726	956	1515	557	277	581	76.7
IL=731, NY=692	FL	FL	FL	FL	FL	FL	FL	FL	Stay	Stay				

Explanation: This is a (small) piece of the optimal decision rule (with a 40-period horizon), showing a sequence of optimal responses to wage draws in successive locations, for a native of Florida. The cost per move is 250. the discount factor is $\beta = .9$, and the persistence parameter is $\rho = .975$. The home premium is 50%.

5.1 The Welfare Trap

The effects of risk aversion on migration decisions are illustrated in Table 5, which shows the initial decision for each wage decile in each location. The most surprising result is the nonmonotonicity of the migration decision. The most attractive location is California: anyone who moves goes to California, and no one leaves there. The least attractive location is Texas, but those who draw wages above the 70th percentile of the Texas wage distribution find it optimal to stay (although they would ultimately leave because wages are not permanent, since the persistence parameter ρ is set to .975). This is as expected. But those who draw the lowest wages in Texas or Florida also find it optimal to stay, even though they would leave if they had drawn a higher wage.

Table 5: Optimal Migration Decisions: The Welfare Trap

	5%	15%	25%	35%	45%	55%	65%	75%	85%	95%	Mean	b	Emax	Eu
CA	131	211	280	351	429	522	639	801	1061	1705	613	542	746	84.6
	Stay	Stay	Stay	Stay	Stay	Stay	Stay	Stay	Stay	Stay				
FL	114	182	240	300	366	444	543	678	896	1432	520	347	575	78.4
	Stay	Stay	Stay	Stay	Stay	**CA**	**CA**	Stay	Stay	Stay				
IL	152	243	322	402	492	598	731	915	1211	1940	701	412	754	82.8
	CA	**CA**	**CA**	**CA**	**CA**	Stay	Stay	Stay	Stay	Stay				
NY	144	230	305	381	466	566	692	866	1146	1834	663	486	753	83.9
	CA	**CA**	**CA**	**CA**	**CA**	Stay	Stay	Stay	Stay	Stay				
TX	125	199	261	325	396	479	583	726	956	1515	557	277	581	76.7
	Stay	Stay	Stay	Stay	**CA**	**CA**	**CA**	Stay	Stay	Stay				

Explanation: This shows the initial migration decision when a wage has been drawn in the home location, and wages in other locations are unknown. The utility function has constant relative risk aversion, with $\gamma = 2$. The cost per move is 300, with $\beta = .9$, $\rho = .975$ and no home premium.

Although at first sight this result may look wrong, there is a simple explanation. The model does not allow borrowing, so in order to pay the cost of moving (set to \$300 in this example), consumption must be reduced in the current period. Since we have assumed that the marginal utility of consumption is diminishing, the utility cost of any given consumption change is higher when consumption is low. The utility function is assumed to be of the form $u(y) = -1/y$, and for convenience we report utilities as $U(y) = 100(1 - 100/y)$. In these units, the cost of moving from Florida for a welfare recipient is $U(347) - U(47) = 183.9$, while the gain from being on welfare in California is $U(542) - U(347) = 10.4$. Thus even though a one-time moving cost of \$300 is small in relation to a permanent gain of \$195 (542 - 347), the cost-benefit calculation looks much different in utility terms. But with a wage of \$444, the cost of moving is reduced to $U(347) - U(144) = 40.6$, and then the present value of the gain exceeds the cost.

A remarkable implication of this result is that in some circumstances an increase in welfare benefits actually causes emigration. In fact, the simulation results show that no one chooses to stay on welfare in New York or in Illinois, because welfare in California is more attractive, and because the welfare benefits in New York and in Illinois are high enough to cover the utility cost of moving. In Texas and Florida, on the other hand, welfare recipients are caught in a kind of trap: in order to finance a move, they would have to reduce consumption to such a low level that the present value of the gains from moving would not cover the immediate cost. If the welfare benefit in these states were to be increased to the Illinois level, however, it would be optimal to leave.

5.2 Learning and Return Migration

The basic difficulty in developing a forward looking model with return migration is that individuals must decide to leave before they can decide to return to their native location. A strong preference for the native location provides a rationale for return migration *after* leaving the native location, but of course it makes the initial move much less likely. When wage draws in different locations have no common component and when there is no risk aversion, return migration occurs only after an individual has sampled nearly all other locations.

Correlation in wages across locations can potentially explain large return migration flows. When the wage in each location has one component that is specific to the individual and another that is specific to the location, a relatively high wage in a low-wage location indicates that the individual-specific component may be large, implying that a move to a high-wage location is likely to pay off. Then if the wage drawn in the new location is relatively low, the inference is that the wage in the initial location reflected a good match in that location, so a return move is indicated.

Table 6 shows simulated return migration statistics with and without risk aversion for various specifications of the common wage component (parameterized by $\lambda_1 = \frac{\sigma_\eta^2}{\sigma_\eta^2 + \sigma_\varepsilon^2}$). When there is no common component, there is no tendency to return to the home location, relative to other locations. But as the common component becomes more important, the return migration pattern seen in Table 2 emerges: return moves are common, with a strong tendency to return to the initial location, rather than an intermediate location. This table also shows that risk aversion has a strong effect on migration decisions. In the risk-neutral case, the migration rate is decreasing in λ_1, because fewer moves are needed to learn about η. The same effect is present in the risk-averse case, but the migration rate is lower by an order of magnitude. Moreover, the migration rate is a U-shaped function of λ_1, because when λ_1 is close to 1, there is very little uncertainty about wages in other locations, so that even a risk-averse individual is willing to move to the high-wage location.

Table 6: Spatial Correlation, Risk Aversion and Return Migration

Simulation of 50,000 Five-Period Migration Histories

	Risk-Averse (γ = 2)				Risk-Neutral			
Common Component Variance Share (λ_1)	35%	55%	75%	95%	35%	55%	75%	95%
Moves	9,785	7,582	14,999	19,506	113,142	103,821	85,385	62,125
Return home	0	0.7%	23.5%	27.4%	7.2%	11.1%	15.2%	20.4%
Return, not home	0	0.0%	0.3%	2.0%	6.9%	6.7%	5.8%	3.1%
Onward	100%	99.3%	76.1%	70.6%	85.8%	82.2%	79.1%	76.5%

Explanation: the table summarizes optimal migration flows for alternative values of the variance share $\lambda_1 = \frac{\sigma_\eta^2}{\sigma_\eta^2 + \sigma_\varepsilon^2}$. The wage distributions and parameter values are as in Table 5.

6 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 discrete approximation of the wage distribution has n points of support, the number of states is $J(n+1)^J$. In the simulation results presented above, J=5 and n=10, yielding 805,255 states. Although the value functions for these simulations were computed in a few hours, estimation of the basic structural parameters (such as the coefficient of risk aversion, and the premium for the home location) requires that the value function be computed many times. Estimation becomes infeasible unless the number of structural parameters is small.

The large number of locations poses another computational challenge. Ideally, locations would be defined as local labor markets. The smallest geographical unit identified in the NLS geocode file is the county, but we obviously can't let J be the number of counties, since there are over 3,100 counties in the U.S. Indeed, even restricting J to the number of states still far exceeds current computational capabilities. To aggregate locations beyond the state level (e.g. Census Regions) is uninterpretable; for example, we lose the ability to identify the effects of state benefit systems. Consequently, we define locations as states, but restrict the information available to each individual.

6.1 Outline of the Estimation Method

We expand the model presented in Section 3 above to allow for unobserved heterogeneity in individual payoffs. Let $\zeta = (\zeta_1,\zeta_2,...,\zeta_J)$ 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

$$V(\ell,\omega,\zeta;\theta) = \begin{cases} \displaystyle\sum_{s=1}^{n} p_s V(\ell,\omega_1,...,\omega_{\ell-1},s,\omega_{\ell+1},...,\omega_J,\zeta;\theta) & \text{if } \omega_\ell = 0 \\ \max\limits_{j}\left[\left(1+\kappa_\ell\right)u\left(y_\ell(\omega_\ell) - \Delta_{\ell j}\right) + \zeta_j + \beta\rho\,\bar{V}(j,\omega;\theta) + \beta(1-\rho)\,\bar{V}(j,0;\theta)\right] & \text{if } \omega_\ell > 0 \end{cases}$$

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 we assume that π is the Type 1 Extreme Value distribution[7] then, using arguments due to McFadden (1973) and Rust (1987) we can show that the function $\bar{V}$ satisfies

[7]A random variable X is exponentially distributed if exp(-X) is uniformly distributed on [0,1]. Repeating this yields an extreme value distribution: Y has the extreme value distribution if exp(-Y) is exponentially distributed.

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

where

$$v_j(\ell,\omega;\theta) = \left(1+\kappa_\ell\right) u\left(y_\ell(\omega_\ell) - \Delta_{\ell j}\right) + \beta\rho\,\bar{V}(j,\omega;\theta) + \beta(1-\rho)\,\bar{V}(j,0;\theta)$$

This gives a closed form for the probability, Pr(d(j)=1 | ℓ,ω), that an individual in location ℓ with information ω will move to location j:

$$Pr(d(j)=1 \mid \ell,\omega;\theta) = \frac{\exp\left(v_j(\ell,\omega;\theta)\right)}{\sum_{\tau=1}^{J} \exp\left(v_\tau(\ell,\omega;\theta)\right)}$$

The individual contribution to the likelihood function for t=1,..,T periods is then

$$\prod_{t=1}^{T} \prod_{j} Pr(d_t(j) \mid \boldsymbol{x_t},\theta)^{d_t(j)} \, R(\boldsymbol{x_t} \mid \boldsymbol{x_{t-1}},\boldsymbol{d_{t-1}},\theta)$$

where $\mathbf{x_t}$ is the observed state (ℓ, ω, location and wage history), $d_t(j)$ is an indicator of whether location j is selected in period t, and R is the state transition probability. Note that although there is a large number of states, R has a simple structure, since it merely tracks the information encoded in ω.

Although Rust (1994) showed that the maximum likelihood estimator of θ is consistent and asymptotically efficient, it requires the simultaneous solution of the value function (for given θ) and zeros of the score function (for a given value function). Consequently we use Rust's (1994) multi-step estimation procedure, which is also consistent and efficient.[8] This method splits the parameter vector θ into sub-vectors (θ_1,θ_2), such that θ_1 enter only through R. The first step consistently estimates θ_1; this is done independently of the dynamic programming algorithm. The second step uses this estimate and the dynamic programming algorithm to jointly estimate the

[8]See van der Klaauw (1996) for another application that successfully applies a variant of this estimator.

value function and θ_2. Step three uses estimates from the first two stages to perform Newton steps on the full likelihood function, yielding an estimator that is asympotically equivalent to the maximum likelihood estimator. In our application θ_1 includes ρ, σ_ε, σ_η and μ_j, and θ_2 contains κ, γ, β, m_0, and $m_{1.}$

6.2 A Limited Memory Approximation

When the number of locations is moderately large, the model becomes computationally infeasible (and will remain so, even if computers improve: for example, if a location is a State, and the wage distribution has 5 points of support, then the number of dynamic programming states is 4041406387323820303215698002282681466880 0). This is a common problem with numerical 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. So far, 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. Suppose instead that migration decisions are made only once a year, and that wage information becomes worthless after M years (because local labor market conditions change over time). Then if the number of locations exceeds M, it is not possible to be fully informed about wages at all locations. This means that the number of dynamic programming states is limited. If there are J locations, and the wage distribution in each location has n points of support, then 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. Then if J is 50 and n is 5 and M is 2, the number of states is 62,500, which is manageable.

Note that we are reducing the number of states in the most obvious way: we simply delete most of them. Someone who has “too much” wage information in the big state space is reassigned to a less-informed state. Individuals makes 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 two people who have 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, two people who have the same recent history are in the same state, even if their previous histories were different (and two people who have different wage information now may have the same wage information following a move).

In order to compute the likelihood function 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. For now, we suppress the common component of wages, and we set $\rho = 1$ (since the wage information becomes irrelevant after M periods in any case).

The probability that an individual in state (ℓ,ω) will move to location j can again be written in the form

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

where v_j is now defined as

$$v_j(\ell,\omega;\theta) = \big(1+\kappa_{\ell^0}\big)u\Big(y_{\ell^0}(\omega_{\ell^0}) - \Delta(\ell^0,j)\Big) + \beta\,\bar{V}\Big((j,\ell^0,\ell^1,\ldots,\ell^{M-2}),(\omega^j,\omega^0,\omega^1,\ldots,\omega^{M-2});\theta\Big)$$

with

$$\bar{V}(\ell,\omega;\theta) = \begin{cases} \sum_{s=1}^{n} p_s\,\bar{V}(\ell,(s,\omega^1,\omega^2,\ldots,\omega^{M-1});\theta) & \text{if } \omega^0 = 0 \\ \log\left(\sum_{j=1}^{J} \exp\big[v_j(\ell,\omega;\theta)\big]\right) & \text{if } \omega^0 > 0 \end{cases}$$

6.3 Migration and Welfare

We will analyze the migration decisions of low income women at risk to receive AFDC. This is a natural application of our model, because (a) location-specific benefits in the model are most directly related to welfare benefits (AFDC and Food Stamps) within each state and (b) we believe that our imperfect capital market assumptions provide a reasonable approximation for this group.

The recent literature on welfare-induced migration is summarized by Meyer (1999). While the consensus view from earlier work reviewed by Moffitt (1992) was that differences in welfare benefits across states had a significant effect on migration decisions, subsequent studies by Levine and Zimmerman (1995) and by Walker (1994) found little or no effect. Meyer argued that by paying careful attention to the determinants of welfare participation, the ambiguity in these results can be resolved in favor of a significant (but small) effect of welfare on migration. All of these studies relied on simple heuristic models of migration decisions, and we believe that a more systematic analysis is warranted. The finding in Section 5 above regarding the perverse effects of benefits levels in the case of the welfare trap illustrates the potential gains from the use of a fully specified model to interpret the data.

6.2.1 Definition of the Estimation Sample

We restrict the estimation sample to women from the non-military subsample of the NLSY79 with twelve or fewer years of education. The observational window begins in the year the woman is first single with a dependent child. To be included in the estimation sample, information on residence must be observed for at least two periods. We follow these respondents either until the end of their single parenthood, the end of the sample period (1992) or the first wave they are not interviewed. There were 1,704 people satisfying these restrictions, and we have data on 12,051 location decisions (i.e. person-years). The overall interstate migration rate is 2.29% per year (276 moves).

6.2.3 Monthly Earnings and Welfare Benefits

For each respondent, monthly earnings equals the sum of annual wage and salary income for all years residing in that state divided by total weeks worked (again for all periods in the state) times 4. Monthly benefits correspond to the combined 1980 AFDC and Food Stamp benefits for a family of 3. The Consumer Price Index for Urban Workers 1983-1984 = 100.0 is used to deflate nominal earnings and benefits into real terms.

Table 8: Wages and Benefits, by State

Single Women with Children, NLSY, $1980

	Observations	Benefits	Mean Wage	85-th Percentile
Alabama	456	4,068	4,605	7,576
Alaska	80	9,864	11,273	18,548
Arizona	113	5,124	9,046	14,884
Arkansas	238	4,692	4,794	7,888
California	1,306	7,896	5,777	9,505
Colorado	164	6,024	8,032	13,216
Connecticut	316	7,908	9,067	14,919
Delaware	11	5,784	4,823	12,097
Dc	127	5,988	8,835	14,537
Florida	493	5,052	6,752	11,109
Georgia	777	4,728	5,349	8,801
Hawaii	5	10,896	4,823	5,820
Idaho	4	6,360	4,823	611
Illinois	392	6,000	5,651	9,298
Indiana	135	5,664	5,102	8,394
Iowa	39	6,744	1,953	3,213
Kansas	85	6,588	6,317	10,394
Kentucky	64	4,980	3,661	6,024
Louisiana	159	4,560	4,908	8,076
Maine	0	5,927	4,823	
Maryland	148	5,820	8,136	13,387
Massachusetts	243	6,936	6,420	10,563
Michigan	502	7,404	5,429	8,932
Minnesota	138	7,308	5,649	9,294
Mississippi	199	3,744	5,841	9,611
Missouri	426	5,592	3,748	6,167
Montana	79	5,712	4,858	7,993
Nebraska	33	6,228	4,716	7,759
Nevada	7	5,736	4,823	10,363
New Hampshire	0	6,597	4,823	
New Jersey	371	6,744	8,253	13,579
New Mexico	93	5,316	5,392	8,871
New York	739	7,080	5,350	8,803
North Carolina	484	5,028	6,679	10,989
North Dakota	0	6,466	4,823	
Ohio	754	5,748	6,372	10,484
Oklahoma	96	5,940	7,823	12,872
Oregon	45	6,480	4,424	7,279
Pennsylvania	382	6,456	7,764	12,774
Rhode Island	13	6,540	4,823	10,949
South Carolina	485	4,224	6,377	10,493
South Dakota	17	6,336	4,823	8,030
Tennessee	208	4,116	4,991	8,212
Texas	893	4,032	6,975	11,477
Utah	29	6,744	4,823	4,839
Vermont	43	8,088	7,400	12,176
Virginia	284	6,228	5,418	8,915
Washington	145	7,728	5,532	9,102
West Virginia	88	5,172	3,431	5,645
Wisconsin	404	7,584	5,545	9,124
Wyoming	1	6,276	4,823	45
All States	12,313	6,162	5,818	

For this stage we relax the restriction that a respondent had to be observed in two consecutive waves to obtain as many observation as possible to estimate the parameters of the distributions in each state. Table 8 reports real monthly benefit levels and descriptive statistics on annual earnings distributions. A notable feature of these data is that less than 50% of single women with children earn more than the benefit level.

The model assumes that individuals who receive wage offers less than the monthly benefit do not work, but accept the benefits. As Table 9 reveals, a large number of respondents have earnings that fall short of the monthly benefit. While the difference could reflect measurement errors in either benefits or earnings, the large gap suggests the difference is real and are consistent with program take-up rates far below 100 percent. Future work will investigate this difference, but for this set of estimates we maintain the full take-up of benefits. We estimate the parameters of the earnings distribution by assuming lognormality and using the upper tail of the distribution to estimate the mean and variance. The state specific means are shown in Table 8; the standard deviation of log wages is assumed constant across states, with an estimated value of 0.7566. We use these estimates in the second stage estimation of the preference and moving cost parameters.

6.2.4 Partial Likelihood Estimates (Stage 2)

We condition on the estimated parameters of the log earnings distribution for each state and the state-specific welfare benefits, and estimate the partial likelihood to recover parameter estimates of the fixed cost of moving (δ_0) and the curvature of the single-period utility function (γ). We fix beta at 0.9 and set the per-mile moving cost δ_1 the home premium κ to zero. To avoid an initial conditions problem, we assume there is no common component or learning.[9]

Table 9 reports preliminary parameter estimates for a 3-point approximation

[9] In the presence of learning, the estimation procedure must recognize that the first location observed in the observational window is not necessarily the first location the individual has worked (and received a wage draw). With access to panel data, we can use information prior to the observational window to correctly estimate the parameters of the subjective wage distribution at the start of the observational window.

Table 9
Partial Likelihood Estimates
Single Women with a Dependent Child

	3-Point approximation of earnings distribution
Fixed cost of moving δ_0	\$3,742.72
Risk aversion γ	0.54
Partial Loglikelihood value	-1575.4981
Number of observations	10,347

Fixed parameters: $\beta = 0.9$, $\rho = 1.0$, $\eta = 0$

The estimated fixed cost of moving is just below the value of benefits in Mississippi. This is the upper limit for the fixed cost parameter for this choice set; since any larger value would yield negative utility and zero predicted probability of moving for a Mississippi welfare recipient, and in fact three welfare recipients did move from Mississippi. This indicates that the model is too tightly specified. Indeed, simulations using the estimated parameters predict migration rates that are much higher than those seen in the data. The model is forced to interpret the low migration rate as the result of drawing utility shocks from the tail of the extreme value distribution; when more likely shocks are used in the simulations, the migration rate is much higher. This indicates that it is necessary to allow for heterogeneity in moving costs, with the interpretation that those who moved despite having low income were people who found a cheap way to move. Unfortunately, unobserved heterogeneity in this form cannot be accommodated within the extreme value framework, so that an additional level of numerical integration is needed to obtain the choice probabilities.

Bibliography

Archdeacon, Thomas J. *Becoming American: An Ethnic History*. New York: The Free Press. 1983

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

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

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

Blank, Rebecca M. (1988) "The Effect of Welfare and Wage Levels on the Location Decisions of Female-Headed Households," *Journal of Urban Economics* 24, 186-211.

Borjas, George J., "Immigration and Welfare Magnets," *Journal of Labor Economics*, 17 (4), Part 1, October 1999, 607-637.

Borjas, George J., Stephen G. Bronars and Stephen J. Trejo, "Self-Selection and Internal Migration in the United States," *Journal of Urban Economics*, 32, September 1992, 159-185.

Center for Human Resource Research, *NLS Handbook 1995*: The National Longitudinal Surveys. The Ohio State University, Columbus, Ohio.

DaVanzo, Julie S. and Peter A. Morrison (1981) "Return and Other Sequences of Migration in the United States," *Demography* 18: 85-101.

El-Gamal, Mahmoud A., "A Dynamic Migration Model with Uncertainty," *Journal of Economic Dynamics and Control*, 18(1994), 511-538.

French, Eric, "Essays on the Dynamics of Hours Worked, Wages, and Health," unpublished Ph.D. thesis, University of Wisconsin-Madison.

Grogger, Jeff and Charles Michalopoulos, "Welfare Dynamics Under Term Limits," NBER Working Paper No. W7353, September, 1999.

Heckman, James J., and James R. Walker, "Forecasting Aggregate Period-Specific Birth Rates: The Time Series Properties of a Microdynamic Neoclassical Model of Fertility,' *Journal of the American Statistical Association*, 84 (December, 1989): 958-965

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

Levine, Phillip B., and David J. Zimmerman "An Empirical Analysis of the Welfare Magnet Debate Using the NLSY," NBER Working Paper No. W5264, September 1995

McCall, B. P. and J. J. McCall (1987) "A Sequential Study of Migration and Job Search," *Journal of Labor Economics* 5 (October 1987): 452-476.

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

McFadden, D. (1981) “Econometric Models of Probablistic Choice,” in *Structural Estimation of Discrete Data with Econometric Applications.* Edited by C. Manski and D. McFadden. Cambridge, MA: MIT Press.

Meyer, Bruce D. “Do the Poor Move to Receive Higher Welfare Benefits?” Northwestern University, October 1999.

Miller, Robert A., “Job Matching and Occupational Choice”, *Journal of Political Economy*, December 1984, vol. 92, no. 6, 1086-1120.

Moffitt, Robert (1992), "Incentive Effects of the United States Welfare System: A Review," *Journal of Economic Literature* 30 (March): 1-61.

National Research Council, (1995) *Measuring Poverty: A New Approach.* Edited by Constance F. Citro and Robert T. Michael. Washington D.C.: National Academy Press. 1995.

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

Roberts, Bryan R. “Socially Expected Durations and the Economic Adjustment of Immigrants,” in *The Economic Sociology of Immigration: Essays on Networks, Ethnicity, and Entrepreneurship*. Edited by Alejandro Portes. New York: Russell Sage Foundation. 1995.

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

Rust, John (1996), "Numerical Dynamic Programming," in *Handbook of Computational Economics*, North Holland, H. Amman, D. Kenrick, and J. Rust eds.

Rust, John and Christopher Phelan (1997) “How Social Security and Medicare Affect Retirement Behavior in a World of Incomplete Markets,” *Econometrica* 65 (4), July 1997, 781-832.

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

Van der Klaauw, Wilbert (1996) “Female Labor Supply and Marital Status Decisions: A Life-Cycle Model,” *Review of Economic Studies* 63: 199-235.

Walker, James R. (1994) "Migration Among Low-Income Households: Helping the Witch Doctors Reach Consensus," Discussion Paper, #94-1032, Institute For Research on Poverty.

Weizman, Martin L., “Optimal Search for the Best Alternative”*Econometrica*, Vol. 47, No. 3. (May, 1979), 641-654.

Figure 1 SMSA Wage Distribution 1980 and 1990 Pums

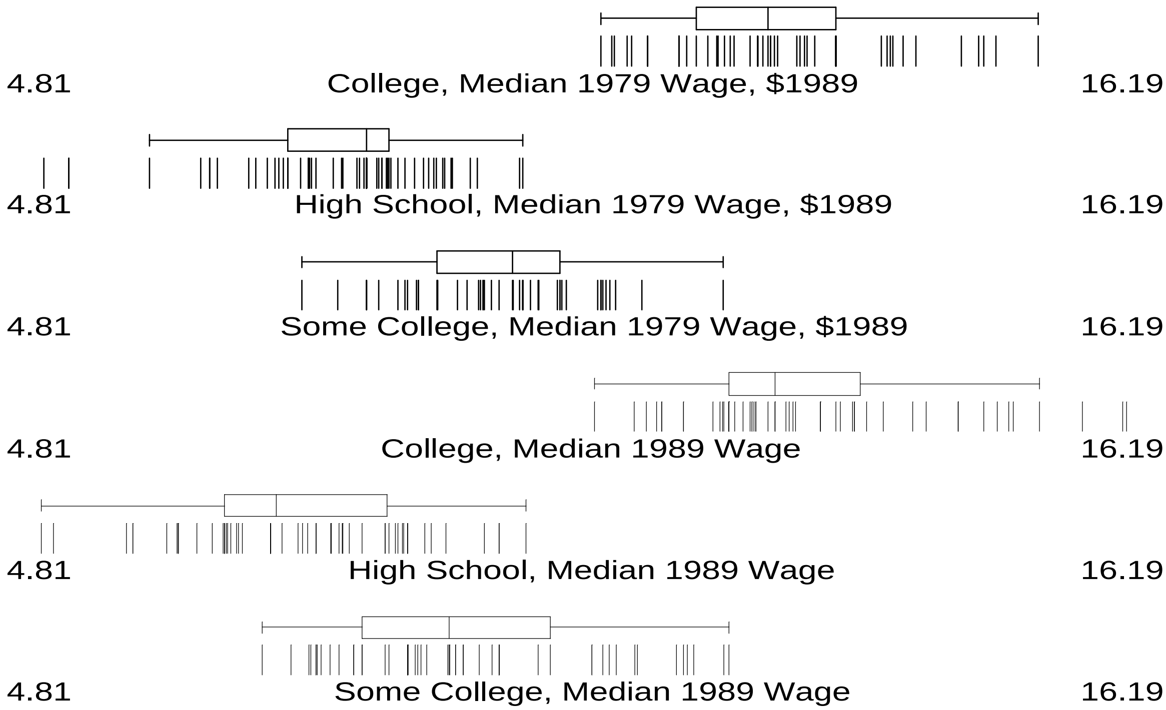

Figure 2 High-Wage and Low-Wage Markets by Education

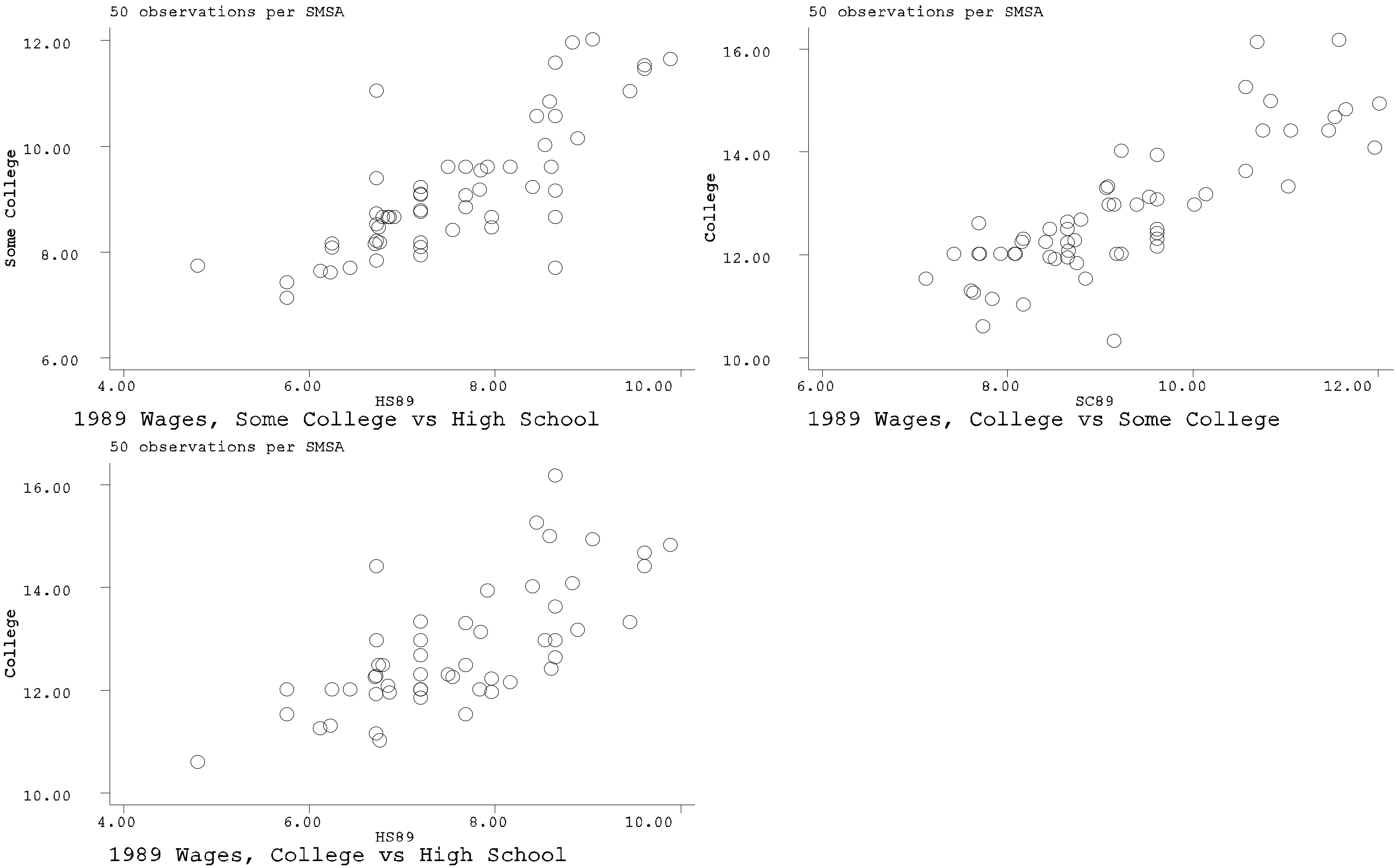

Figure 3 Persistence of Geographical Wage Differentials