"The Effect of Expected Income on Individual Migration Decisions" J. Kennan, J. Walker (2011)

Dan Beemon, John Stromme, Anna Trubnikova, Anson Zhou

October 15, 2017

Motivation

Fact 1: Large fraction of movers are 'repeat movers'

Fact 2: Large fraction of movers return home.

TABLE I

INTERSTATE MIGRATION, NLSY 1979–1994[a]

	Less Than High School	High School	Some College	College	Total
Number of people	322	919	758	685	2,684
Movers (age 18–27)	80	223	224	341	868
Movers (%)	24.8%	24.3%	29.6%	49.8%	32.3%
Moves per mover	2.10	1.95	1.90	2.02	1.98
Repeat moves (% of all moves)	52.4%	48.7%	47.4%	50.5%	49.5%
Return migration (% of all moves)					
Home	32.7%	33.1%	29.1%	23.2%	28.1%
Not home	15.5%	7.1%	6.8%	8.6%	8.4%
Movers who return home	61.3%	56.5%	51.3%	42.8%	50.2%

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

Literature

Previous attempts were not able to model the complexity of the migration decision

- Holt (1996) & Tunali (2000) - Only modeled move-stay decisions, do not distinguish between different destinations
- Dahl (2002) - Many destinations, but only a single life time migration decision
- Gallin (2004) - Modeled net migration as a response to wages, but does not model individual decision problem

General setup

Finite-period discrete choice Bellman equation for individual i

$$V(x, \epsilon, \zeta) = \max_j \left(u_j(x, \epsilon) + \zeta_j + \beta \sum_{x', \epsilon'} \bar{V}(x', \epsilon') f_j(x', \epsilon' | x, \epsilon) \right)$$

- State:
  - observable x: current location l, previous location l^{-1}, age
  - constant parameters: h - home location, τ - type
  - ζ_j - preference or moving costs shock $\sim$ type I EV,
  - ϵ - other unobservables (more on this later)
- Choice:
  - j - new location ($d_{jt}^{(n)} = 1$ in lecture notations)

Conditional independence doesn't hold, because some unobservables in ϵ are persistent over time.

But as $\zeta_j \perp \epsilon$, iid over time, we can get rid of it by using CCP

$$\rho_j(x, \epsilon) = \exp(0.57 + v_j(x, \epsilon) - \bar{V}(x, \epsilon))$$

Specification of flow payoff

- Flow payoff: $u_j(x,\epsilon) = \alpha_0 w_{ilt} + \alpha^H \mathbb{I}_{l=h} + \xi_{il} + \text{amenities}_l - \Delta_\tau(x,j)$
    - ξ_{il} - utility fixed effect of location (agent knows after visit)

- Wage equation: $w_{ilt} = \mu_l + \nu_{il} + \eta_i + \text{deterministic trend} + \varepsilon_{it}$
    - μ_l - mean wage at location (from data)
    - η_i - individual fixed effect (agent knows ex ante)
    - ν_{il} - permanent location match parameter (agent knows after visit)
    - ε_{it} - random shock (can be inferred by agent)

- Moving costs: (only if person moves: $j \neq l$)

$$\Delta_\tau(x,j) = \gamma_{0\tau} + \gamma_1 \text{distance}(l,j) - \gamma_2 \mathbb{I}_{j \text{ is adjacent to } l} - \gamma_3 \mathbb{I}_{j=l^{-1}} + \gamma_4 \text{age} - \gamma_5 \text{pop-n}_j$$

    - Intercept differs w.r.t. types τ: *movers* and *stayers* (prohibitive cost of moving in all states).

Identification strategy

Main idea

Parameters are identified using the variation in mean wages across locations or by using the variation in the location match component of wages

Key assumption

Wage components $(\eta_i, \nu_{il}, \varepsilon_{it})$ and the location match component of preferences ξ_{il} are all i.i.d. across individual and states, and ε_{it} is i.i.d. over time

Identification steps:

1. Identify the CCP function
2. Identify other parameters by exploiting variations

Identification of CCP function: simple example

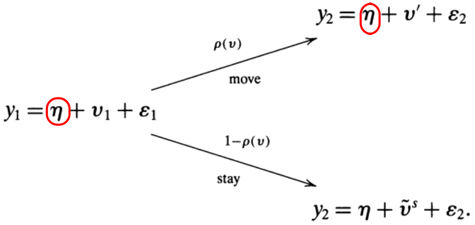
- Consider just two observations for each person, wage residual for i in period t in location $l(t)$ is

$$y_{it} = w_{ilt} - \mu_l - G(X_i, a, t) = \eta_i + \nu_{il(t)} + \varepsilon_{it}$$

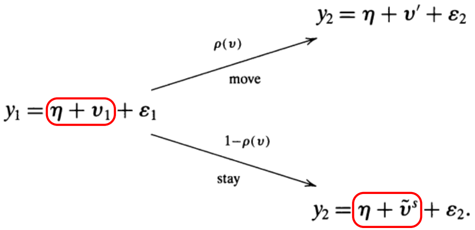
- Since (η, ν, ε) are independent, the probability of moving (in the first period) only depend on $\nu_{il(t)}$, denoted as $\rho(\nu)$

Kotlarski's Lemma

Suppose one observes the joint distribution of two noisy measurements $(Y_1, Y_2) = (M + U_1, M + U_2)$ of a random variable M, where random U_1 and U_2 are measurement errors. When (M, U_1, U_2) are mutually independent, $\mathbb{E}(U_1) = 0$, and the characteristic functions of M, U_1, U_2 are non-vanishing, then the distributions of M, U_1 and U_2 are identified.



- For movers, $y_1 = \eta + \tilde{\nu}^m + \varepsilon_1$ and $y_2 = \eta + \nu' + \varepsilon_2$ where $\tilde{\nu}^m$ is the censored random variable ν by discarding the people who stay, and ν' is a new draw (independent of $\tilde{\nu}^m$)
- Apply Kotlarski's Lemma, the distributions of $\eta, \tilde{\nu}^m + \varepsilon_1$ and $\nu + \varepsilon_2$ are identified



- For stayers, $y_1 = \eta + \tilde{\nu}^s + \varepsilon_1$ and $y_2 = \eta + \tilde{\nu}^s + \varepsilon_2$ where $\tilde{\nu}^s$ is the censored random variable ν by discarding the people who move
- Apply Kotlarski's Lemma, the distributions of $\eta + \tilde{\nu}_s, \varepsilon_1$ and ε_2 are identified

- Since we can identify the distributions of $\eta, \tilde{\nu}^m + \varepsilon_1$,$\nu + \varepsilon_2$, $\eta + \tilde{\nu_s}, \varepsilon_1$ and ε_2, the distributions of $\eta, \nu, \varepsilon_1, \varepsilon_2, \tilde{\nu}^m$, and $\tilde{\nu}^s$ are all identified (either directly or by deconvolution)
- The conditional choice probabilities $\rho(\nu)$ are identified by Bayes theorem

$$f_{\tilde{\nu}^m}(\nu) = \frac{\rho(\nu) f_\nu(\nu)}{\text{Prob(move)}}$$

- The shape of $\rho(\nu)$ shows the effect of income on migration decisions

Identification of Income Coefficients

- In the model, CCP is given by

$$\rho_j(l, \nu_s) = \begin{cases} \frac{\exp(-\Delta_{lj} + \beta \bar{V}_0(j))}{\exp(\beta \bar{V}_s(l)) + \sum_{k \neq l} \exp(-\Delta_{lk} + \beta \bar{V}_0(k))} & j \neq l \\ \frac{\exp(\beta \bar{V}_s(l))}{\exp(\beta \bar{V}_s(l)) + \sum_{k \neq l} \exp(-\Delta_{lk} + \beta \bar{V}_0(k))} & j = l \end{cases}$$

  where Δ_{lj} is cost of moving from l to j, $\bar{V}_s(j)$ is expected continuation value after knowing ν_s but before knowing ζ, and $\bar{V}_0(j)$ is expected continuation value before knowing ν
- We are able to identify the CCP function, but not the CCPs themselves since ν_s is unobserved! (Consequence of relaxing the CIA)
- Need to normalize a payoff, let $\bar{V}_0(J) = 0$

Suppose ν_s is known, the identification can proceed:

- Use round-trip to cancel out $\bar{V}_0$ and $\bar{V}_s$, we have

$$\frac{1}{n}\sum_{s=1}^{n}\log\left(\frac{\rho_j(l,\nu_s)}{\rho_l(l,\nu_s)}\frac{\rho_l(j,\nu_s)}{\rho_j(j,\nu_s)}\right) = -\Delta_{lj} - \Delta_{jl}$$

  where left-hand-side is identified

Suppose ν_s is known, the identification can proceed:

- Use round-trip to cancel out $\bar{V}_0$ and $\bar{V}_s$, we have

$$\frac{1}{n}\sum_{s=1}^{n}\log\left(\frac{\rho_j(l,\nu_s)}{\rho_l(l,\nu_s)}\frac{\rho_l(j,\nu_s)}{\rho_j(j,\nu_s)}\right) = -\Delta_{lj} - \Delta_{jl}$$

  where left-hand-side is identified

- With parametrization of Δ, for two nonadjacent locations,

$$\Delta_{lj} + \Delta_{jl} = 2(\gamma_0 + \gamma_4 a + \gamma_1 D(j,l)) - \gamma_5(n_j + n_l)$$

  1. By choosing three distinct location pairs, we can get variation on $D(j,l)$ and $n_j + n_l$ so as to identify γ_1, γ_5 and $\gamma_0 + \gamma_4 a$
  2. By choosing different a, γ_0 and γ_4 are identified
  3. The remaining parameter γ_2 (coefficient of adjacent dummy) is identified by comparing adjacent and nonadjacent pairs

- All coefficients in Δ_{ij} are identified

1. Normalize $\bar{V}_0(J) = 0$, $\bar{V}_0(I)$ is identified:

$$\frac{1}{n}\sum_{s=1}^{n} \log\left(\frac{\rho_J(I,\nu_s)}{\rho_I(I,\nu_s)}\right) = -\Delta_{Ij} - \beta\bar{V}_0(I)$$

2. $\bar{V}_s(I)$ is also identified:

$$\log\left(\frac{\rho_j(I,\nu_s)}{\rho_I(I,\nu_s)}\right) = -\Delta_{Ij} + \beta(\bar{V}_0(j) - \bar{V}_s(I))$$

3. Coefficient of wage in determining utility, α_0 is identified by differencing the equation

$$\bar{V}_s(I) = \bar{\gamma} + \alpha_0\nu_s + A_I + \log\left(\exp(\beta\bar{V}_s(I) + \sum_{k\neq I}\exp(\Delta_{Ik} + \beta\bar{V}_0(k))\right)$$

4. Amenity values A_I is identified as the remaining term

Maximum Likelihood Estimation

- Full Maximum Likelihood $\Lambda(\theta) = \sum_{i=1}^{N} \log \sum_{\tau=1}^{2} \pi_\tau \mathtt{L}_i(\theta_\tau)$
  - π_1 - probability of stayer, π_2 - probability of mover
- Location is a choice/state, wage is not, but we want to use extra data:

$$\mathtt{L}_i(\theta_\tau) = \mathtt{P}(\{\text{data}_i\}_1^T | l_1) = \int_\epsilon \prod_{t=1}^{T-1} \mathtt{H}(x_{t+1}^{(i)}, \epsilon_{t+1} | x_t^{(i)}, \epsilon_t) \cdot$$

$$\cdot \prod_{t=1}^{T} \mathtt{P}(w_t^{(i)} | l_t^{(i)}, \epsilon_t)\, g(\epsilon_1 | l_1^{(i)})\, d\epsilon$$

- As in lectures, $\mathtt{H}(\cdot)$ is a probability of new state, conditional on optimal choice $l(t+1)$ and previous state:

$$\mathtt{H}(x_{t+1}, \epsilon_{t+1} | x_t, \epsilon_t) = \rho_{l(t+1)}(l_t, l_{t-1}, \epsilon_t) f_{l(t+1)t}(x_{t+1}, \epsilon_{t+1} | x_t, \epsilon_t)$$

- Assume $\varepsilon_{it} \sim N(0, \sigma_\varepsilon^2(i))$:

$$\mathtt{P}(w | l, \epsilon) = f_{\varepsilon_i}(w - \mu_l - \nu_{il} - \eta_i - \text{trend} | \nu_{il}, \eta_i)$$

ML: more on integration over unobservables

The goal is to simplify integration over unobservables.

- Unobservables ϵ include:
  - wage-location effect ν_{il}: 3 point uniform, symmetric around 0 $\rightarrow$ 1 prm
  - utility-location effect ξ_{il}: 3 point uniform, symmetric around 0 $\rightarrow$ 1 prm
  - individual fixed effect η_i: 7 point uniform, symmetric around 0 $\rightarrow$ 3 prm
  - variance of wage error $\sigma_\varepsilon(i)$: 4 point uniform $\rightarrow$ 4 prm

Since all ϵ are independent of time, authors can draw $\eta_i, \sigma_\varepsilon(i)$ once beforehand, as well as $\{\nu_{il}, \xi_{il}\}_{l=1}^{N_i}$ for all visited locations for this individual.

- Then calculate CCP and probability of observed wage for all combinations of unobservables and average across:

$$\mathrm{L}_i(\theta_\tau) = \frac{1}{(3 \cdot 3)^{N_i} \cdot 7 \cdot 4} \sum_{\epsilon} \prod_{t=2}^{T-1} \rho_{l^{(i)}(t+1)}(l_t^{(i)}, l_{t-1}^{(i)}, \epsilon_t) \prod_{t=1}^{T} \mathrm{P}(w_t^{(i)} | l_t^{(i)}, \epsilon_t)$$

Goodness of fit – Fact 1

Model does a much better job of predicting moves than benchmark binomial distribution

- Started 100 replicas of NLSY individuals in initial locations and generate histories from (estimated) model
- Compare to a binomial distribution with migration probability of 2.9%
- Model overpredicts number who move more than once, but much closer to data than binomial

TABLE VI

GOODNESS OF FIT

Moves	Binomial		NLSY		Model	
None	325.1	75.3%	361	83.6%	36,257	83.9%
One	91.5	21.2%	31	7.2%	2,466	5.7%
More	15.4	3.6%	40	9.3%	4,478	10.4%
Movers with more than one move	14.4%		56.3%		64.5%	
Total observations	432		432		43,201	

Matching return migration – Fact 2

The model does a reasonably good job of reproducing return migration patterns in the data

- Overpredicts return home from initial non-home location
- Model does not capture duration dependence

TABLE VII

RETURN MIGRATION STATISTICS

Movers	NLSY	Model
Proportion who		
Return home	34.7%	37.5%
Return elsewhere	4.8%	6.2%
Move on	60.5%	61.9%
Proportion who ever		
Leave home	14.4%	13.7%
Move from not-home	40.0%	42.5%
Return from not-home	25.7%	32.3%

New insights from empirical results

Substantial result

Model allows for estimation of moving costs, which are not directly observable from the data

Migration decisions significantly impacted by expected income changes

- Moving cost is about $312,000
  - Move away from bad location match worth $8,366
  - Move from bottom to top of state means worth $9,531
- Describes avg. value of hypothetical moves, not moves actually made
  - Justifies that most people never move
- Note: continuation values normalized, so dollar values not necessarily reliable

Returning home more favorable

- Home premium worth wage increase of $23,106
- Cost of moving to previous location relatively low

Conclusions

Model improves on previous work in two respects

1. Covers optimal sequences of location decisions
2. Allows for many alternative location choices

This makes analysis of return migration decisions feasible

- Return migration frequently seen in the data

Migration partly driven by negative effect of current income

- Good draws tend to stay, bad draws tend to leave, independent of distribution in new location