Adding Heterogenous Earnings Processes to the OG Model

Jason DeBacker and Richard W. Evans

India Department of Revenue and World Bank
May 7, 2018

Heterogeneity in earnings

- Allows model to general a distribution of income and wealth (within cohort)
- Interesting distributional analysis across and within generations
- Since labor is endogenous, we want heterogeneity in hourly earnings (not annual earnings)

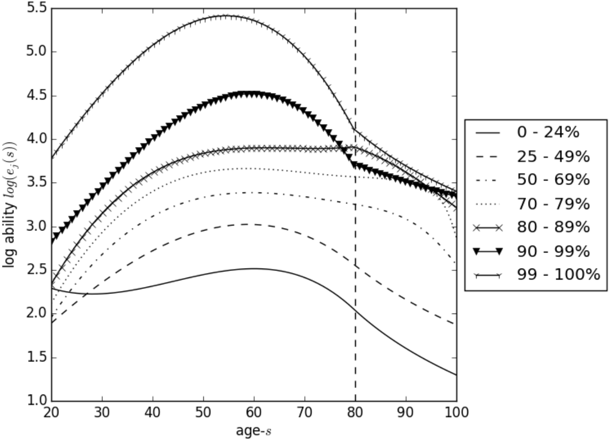
How to introduce heterogeneity in hourly earnings?

Three (of many) ways:

1. Permanent and constant differences in lifetime earnings among types
2. Differences in the lifecycle earnings profiles of the types
3. Stochastic income (perhaps with different lifecycle earnings profiles)

We go with option (2), which gives us a great deal of realism, with lower computational cost than stochastic income.

Lifecycle Profiles of Hourly Earnings



How do we estimate these profiles?

- Start with panel data that contains hourly earnings equivalent to the model unit
  - If using administrative data without measures of hours, you may need to impute hours onto those data
  - DeBacker and Ramnath (2018) provide one template for this using survey data to impute hours onto tax records
- Here we are modeling households, so we use panel data on hourly earnings of households (actually “tax filing units", but they are close)
  - Alternatively, if model units are individuals, we could like our data to have observations at the level of the individual (not household)
- Panel will allow us to get the lifecycle profiles

Defining ability groups

- Want to define ability groups
- These can't be based on total earnings, since that is endogenous (a function of labor supply)
- We use "potential lifetime income"
- Essentially, households are grouped according to what they would earning if they worked full time from ages 21 to 80
- This is important - we want to group individuals by their exogenous productivity and not by total earnings, which depend on labor supply and are thus endogenous

Defining lifetime income

Lifetime income for person i is given by:

$$LI_i = \sum_{s=21}^{80} \left(\frac{1}{1+r}\right)^{s-21} (w_{i,s} * 4000)$$

- The 4000 in this equation represents full time work for a year for two-earner household.
- Note that the discounting means that earnings that come later in life have less of an impact on this measure of lifetime income, which is a present-value measure.

Defining lifetime income

- Note that the panel data won't be balanced and we don't observe anyone from age 21 to 80 (where age defined by age of primary filer in tax unit/primary earner of household)
- So we impute wages for each year of life, s using the following regression:

$$ln(w_{i,s}) = \alpha_i + \beta_1 age_{i,s} + \beta_2 age^2_{i,s} + \beta_3 * age^3_{i,s} + \varepsilon_{i,s} \quad (1)$$

- Let $\hat{w}_{i,s}$ be the fitted values from this regression, found by using the fixed effect for the filing unit and varying ages for all $t \in [21, 80]$
- These imputed wages, $\hat{w}_{i,s}$, are used to determine potential lifetime income for each household:

Estimating lifecycle profiles by group

- We partition the households into lifetime groups based on their percentile in the distribution of lifetime income
  - We break the sample into $J = 7$ groups:

    $$[0\text{–}25\%, 25\text{–}50\%, 50\text{–}70\%, 70\text{–}80\%, 80\text{–}90\%, 90\text{–}99\%, 100\%]$$

  - One could do more/less groups, different percentiles
- Then for each lifetime income group, j, we estimate the lifecycle profile of earnings as:

$$ln(w^j_{i,s}) = \alpha + \beta^j_1 age_{i,s} + \beta^j_2 age^2_{i,s} + \beta^j_3 * age^3_{i,s} + \varepsilon_{i,s} \quad (2)$$

Descriptive statistics

Lifetime Income Category:	1	2	3	4	5	6	7	All
Percentiles	0-25	25-50	50-70	70-80	80-90	90-99	99-100	0-100
Observations	65,698	101,484	74,253	33,528	31,919	24,370	2,129	333,381
Fraction Single								
Females	0.30	0.24	0.25	0.32	0.38	0.40	0.22	0.28
Males	0.18	0.22	0.30	0.35	0.38	0.37	0.20	0.26
Fraction Married								
Female Head	0.08	0.00	0.00	0.00	0.00	0.00	0.00	0.02
Male Head	0.45	0.53	0.45	0.32	0.23	0.23	0.57	0.39
Mean:								
Age, Primary	51.72	44.15	38.05	34.09	31.53	30.79	40.17	39.10
Hourly Wage	11.60	16.98	20.46	23.04	26.06	40.60	237.80	21.33
Annual Wages	25,178	44,237	54,836	57,739	61,288	92,191	529,522	51,604
Lifetime Income	666,559	1,290,522	1,913,029	2,535,533	3,249,287	5,051,753	18,080,868	2,021,298

* CWHS data, 1991-2009, all nominal values in 2005$.

Ages 80-100

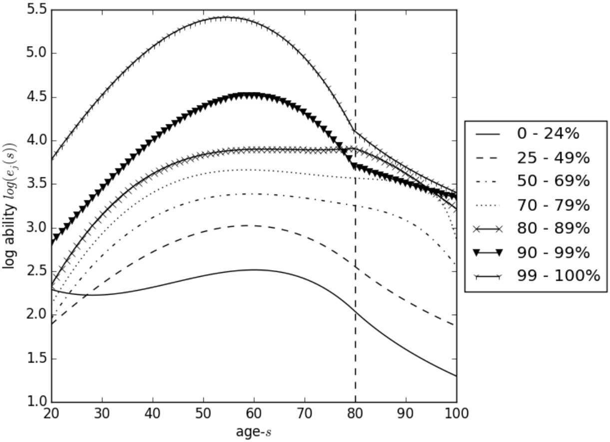
- We model individuals up until age 100, but our data contain few observations over age 80
- We thus interpolate our data out through these final 20 years
- We do this with an arctangent function:

$$y = \left(\frac{-a}{\pi}\right) * arctan(bx + c) + \frac{a}{2}$$

Ages 80-100

- In the function
$$y = \left(\frac{-a}{\pi}\right) * arctan(bx + c) + \frac{a}{2}$$
- The variable x is age
- The parameters a, b, and c are found by finding those that best fit of the function given:
    1. The value of the function should match the value of the data at age 80
    2. The slope of the arctan should match the slope of the data at age 80
    3. The value of the function should match the value of the data at age 100 times a constant.
        - This constant is .5 for all lifetime income groups, except the 2nd highest ability is .7
        - Otherwise, the 2nd highest has a lower income than the 3rd highest ability group in the last few years.

Lifecycle Profiles of Hourly Earnings



What's the model analogue to the data?

- Model thus far has labor supply (hours) and hourly wage rate
- Data are wages per hour worked
- Consider the transformation:
  - Let the wage rate be the rate per effective hour worked
  - E.g. the person with average ability earns w_t per unit of labor supplied
  - Then wages per hour in the data are $w_t \times$ effective hours

Distribution of household across types

- How many households of each type?
- The way we estimated the processes for each lifetime income group, it's clear by the percentile groups
- Divide the unit measure born in each cohort into these groups
- Let λ_j give the measure of households of type j

$$\boldsymbol{\lambda} = [0.25, 0.25, 0.20, 0.10, 0.10, 0.09, 0.01]$$

Modeling ability

- Let e_s be effective labor hours per unit of labor supplied by an age s household
- Let j represent the lifetime income group (we'll call this the ability group)
- Then $e_{j,s}$ are effective labor hours per unit of labor supplied by a household of age s and ability group j
- So total labor income in period t for this household is

$$\underbrace{w_t * e_{j,s}}_{\text{earnings per unit of labor = wages from data}} * n_{j,s,t}$$

Filling in $e_{j,s}$

- The matrix of effective labor units (ability) is $J \times S$

$$emat = \begin{bmatrix} e_{1,1} & e_{1,2} & \cdots \\ \vdots & \ddots & \\ e_{J,1} & & e_{J,S} \end{bmatrix}$$

- Recall we estimated the lifecycle earnings profiles for each type J
- Next step: Find implied values at each age, s and for each ability j
- Do this by fitting Equation 2

Filling in $e_{j,s}$

- Equation 2 gives:

$$e_{j,s} = \alpha + \beta_1^j s + \beta_2^j s^2 + \beta_3^j s^3 \tag{3}$$

- By varying $s \in [21, 80]$ and $j \in [1, J]$ we can solve for each of the *SxJ* elements of *emat*.

Normalizing $e_{j,s}$

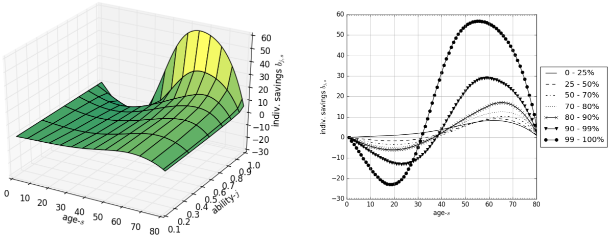
- Lastly, we do some normalization. The *emat* we use in the model is that specified above, but we:
  - Scale *emat* such that the weighted average equals 1:

$$\frac{\sum_{s=1}^{S}\sum_{j=1}^{J}\lambda_j e_{j,s}}{\sum_{s=1}^{S}\sum_{j=1}^{J}\lambda_j} = 1 \tag{4}$$

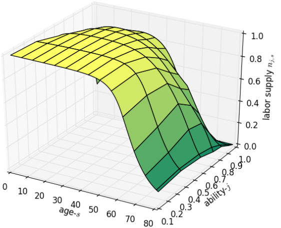
- This normalization takes out the units from the data. It will change our interpretation of the model wage, but only insofar as we now need to interepret the wage as a wage per effective labor units where the mean effective labor units per hour has been normalized to one.

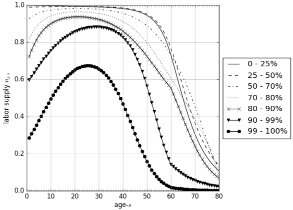
Results

Lifecycle Profiles of Savings



Lifecycle Profiles of Labor Supply





Lifecycle Profiles of Consumption