Quantitative Macro-Labor: Income Processes

Professor Griffy

UAlbany

Fall 2024

Announcements

- Be sure you have done the following:
  - Setting up programming languages.
  - Set up LaTeX.
  - Look into using SSH and the campus cluster.
  - I will email IT this afternoon to make sure everyone has access to the cluster.
- Possible change to syllabus: add topic on market power?

Recap from last time

- In conducting our quantitative exercises, we want an accurate model of income and wages for calibration & evaluation.
- Basic approach:
  1. Condition on observables in data and estimate residual earnings.
  2. Use model to understand the contributions to residual earnings.
- Some considerations:
  1. Make your data look as much like the model as possible.
  2. Does your model have an hours choice, i.e., an *intensive margin*? Then you want to match hourly wages.
- Good sources (where much of this comes from):
  - Chris Tonetti's Write-up on income processes
  - Gianluca Violante's slides on income processes

Residual Earnings

- Estimate the following:

$$Y_{i,j,t} = w_t \exp(f(X_{i,j,t}) + \epsilon_{i,j,t}) h_{i,j,t} \tag{1}$$

- where

$$\begin{aligned} Y_{i,j,t} &: \text{Total income of ind. i at age j, at time t} \\ w_t &: \text{Hourly wages} \\ X_{i,j,t} &: \text{Predictable component} \\ \epsilon_{i,j,t} &: \text{Stochastic component} \\ h_{i,j,t} &: \text{Hours worked} \end{aligned}$$

- Note: if your model has an *intensive margin*, i.e., choice of hours, estimate

$$y_{i,j,t} = \frac{Y_{i,j,t}}{h_{i,j,t}} = w_t \exp(f(X_{i,j,t}) + \epsilon_{i,j,t}) \tag{2}$$

Residual Earnings

- Estimate the following:

$$ln(Y_{i,j,t}) = ln(w_t) + ln(h_{i,j,t}) + f(X_{i,j,t}) + \epsilon_{i,j,t} \qquad (3)$$

- $X_{i,j,t}$ includes covariates (heterogeneity) that is observable and not central to the research question.
- Some examples:
  - Race, gender (usually just use males), marital status
  - Year, state, age
- You might try to explain the effect of a covariate, i.e., race on labor earnings.
- Usually, write down a model to understand the residual.

Statistical Model of Income

- We can think about residual earnings as being partially transitory:
  - Temporary layoff
  - Temporary hours cut
- And persistent:
  - College degree
  - Innate ability
  - Job-specific skills
- A statistical model (i.e., one with no explicit structural interpretation) will estimate the size of these two components over the life-cycle.

Statistical Model of Income

- If you are not micro-founding the income process (i.e., writing down a search model), need alternate model of income.
- Basic approach: specify two equation transitory and persistent model of income.
- Let $\tilde{Y}_{i,j}$ be residual log earnings of individual i at age j (remember, we controlled for time in the first stage).

$$\tilde{Y}_{i,j} = z_{i,j} + \psi_{i,j} \tag{4}$$

$$z_{i,j} = \rho z_{i,j-1} + \nu_{i,j} \tag{5}$$

$$\psi_{i,a} \sim iid(0, \sigma_\psi) \tag{6}$$

$$\nu_{i,a} \sim iid(0, \sigma_\nu) \tag{7}$$

$$z_{i,0} = 0 \tag{8}$$

Identification

- Even in macro, we need to worry about identification.
- What parameters do we need to estimate?
- $\rho, \sigma_\psi, \sigma_\nu$.

$$\tilde{Y}_{i,j} = z_{i,j} + \psi_{i,j} \tag{9}$$

$$z_{i,j} = \rho z_{i,j-1} + \nu_{i,j} \tag{10}$$

$$\rightarrow Var(\tilde{Y}_{i,0}) = Var(z_{i,0}) + Var(\psi_{i,0}) + \cancel{Cov(z_{i,0}, \psi_{i,0})} \tag{11}$$

$$\rightarrow Var(\tilde{Y}_{i,0}) = \sigma_\psi^2 \tag{12}$$

$$Var(\tilde{Y}_{i,j}) = Var(z_{i,j}) + \sigma_\psi^2 \tag{13}$$

$$Var(z_{i,j}) = \rho^2 Var(z_{i,j-1}) + \sigma_\nu^2 \tag{14}$$

$$Cov(\tilde{Y}_{i,j}, \tilde{Y}_{i,j-n}) = Cov(z_{i,j}, z_{i,j-n}) \tag{15}$$

$$Cov(z_{i,j}, z_{i,j-n}) = \rho^n Var(z_{i,j-n}) \tag{16}$$

Identification II

- First, the peristence:

$$\frac{Cov(z_{i,j}, z_{i,j-2})}{Cov(z_{i,j-1}, z_{i,j-2})} = \frac{\rho^2 Var(z_{i,j-2})}{\rho Var(z_{i,j-2})} = \rho \tag{17}$$

- Initial distribution is simple: $Var(\tilde{Y}_{i,0}) = \sigma_\psi^2$
- The persistent shock:

$$Var(z_{i,j}) = \rho^2 Var(z_{i,j-1}) + \sigma_\nu^2 \tag{18}$$

$$Var(\tilde{Y}_{i,j}) - \sigma_\psi^2 = \rho^2 Var(z_{i,j-1}) + \sigma_\nu^2 \tag{19}$$

$$Var(\tilde{Y}_{i,j}) - \rho Cov(\tilde{Y}_{i,j}, \tilde{Y}_{i,j-1}) - \sigma_\psi^2 = \sigma_\nu^2 \tag{20}$$

Statistical Model of Income

- We've now identified an earnings process and can estimate it from the data.

$$\tilde{Y}_{i,j} = z_{i,j} + \psi_{i,j} \tag{21}$$

$$z_{i,j} = \rho z_{i,j-1} + \nu_{i,j} \tag{22}$$

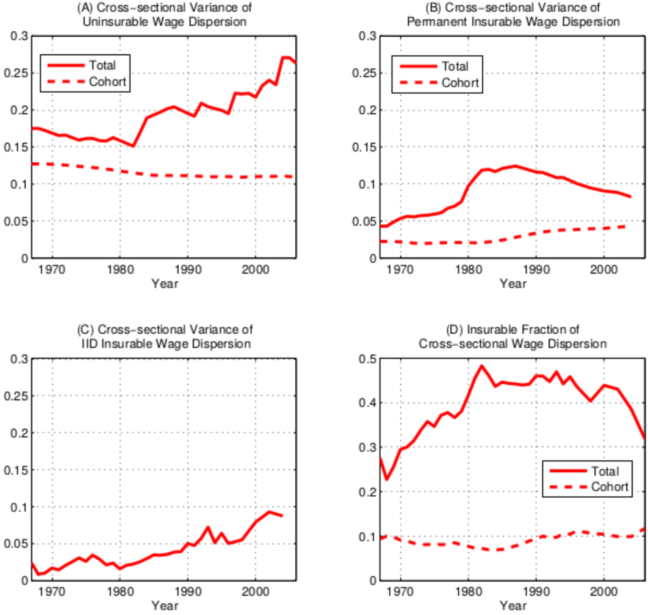
$$\psi_{i,a} \sim iid(0, \sigma_\psi) \tag{23}$$

$$\nu_{i,a} \sim iid(0, \sigma_\nu) \tag{24}$$

$$z_{i,0} = 0 \tag{25}$$

- What can we use it for?
    1. Calibration targets for a structural model.
    2. External validation of structural model.
    3. Better understanding of data.
- This also tells us roughly what fraction of income shocks are insurable (transitory) and what are not (permanent).

Statistical Model of Income



Income Profile Heterogeneity

- What have we done thus far? Specified an income process with persistent and transitory components.
- But, we've assumed that it fluctuates around some trend, in which there is no heterogeneity.
- Is this a reasonable assumption?
- Guvenen (2009) suggests that it is not.

Income Profile Heterogeneity

- (Mostly following Guvenen's notation)
- Model of log labor income of individual i at age/experience j:

$$y_j^i = \beta^i \times j + z_j^i \tag{26}$$

$$z_j^i = \rho z_{j-1}^i + \eta_j^i \tag{27}$$

- There is a deterministic trend, β^i, which is potentially heterogeneous between individuals i.
- There is a persistent shock z_j^i.
- (Full specification in subsequent discussion)

Income Profile Heterogeneity

$$y_j^i = \beta^i \times j + z_j^i \tag{28}$$

$$z_j^i = \rho z_{j-1}^i + \eta_j^i \tag{29}$$

- Two views of β^i:
  - Restricted income profiles (RIP): $\sigma_\beta = 0$, i.e., profiles fluctuate around the same trend.
  - i.e., $y_j^i = \beta \times j + z_j^i$
  - Heterogeneous income profiles (HIP): $\sigma_\beta > 0$, i.e., profiles have different individual trends.
- Why important?
  - As before, these profiles are often used as calibration targets.
  - Difference isn't an innocuous assumption.

Interpretation

$$y_j^i = \beta^i \times j + z_j^i \tag{30}$$
$$z_j^i = \rho z_{j-1}^i + \eta_j^i \tag{31}$$

- What does it mean for individuals to have heterogeneous income growth rates?
- What are some empirical implications for restricted growth rates?
- How could we use a model to think about this income process?

Guvenen (2009) Statistical Model of Earnings

- Estimate the following:

$$y^i_{j,t} = g(\theta^0_t, X^i_{j,t}) + f(\alpha^i, \beta^i, X^i_{j,t}) + z^i_{j,t} + \phi_t \epsilon^i_{j,t} \tag{32}$$

$$\begin{aligned} g(\theta^0_t, X^i_{j,t}) &: \text{Predictable common growth rate \& intercep}t \\ f(\alpha^i, \beta^i, X^i_{j,t}) &: \text{Individual specific growth rate \& intercept } (\alpha) \\ \phi_t \epsilon^i_{j,t} &: \text{Transitory shocks/Measurement error, iid} \end{aligned}$$

- Profile heterogeneity:

$$f(\alpha^i, \beta^i, X^i_{j,t}) = \alpha^i + \beta^i \times j \tag{33}$$

- Permanent shocks:

$$z^i_j = \rho z^i_{j-1} + \pi_t \eta^i_j, \quad z^i_{0,t} = 0, \eta \sim iid \tag{34}$$

- Heterogeneity uncorrelated with shocks.

Identification (ex-ante)

- As before, identification using second-order moments.
- We can denote the income residual $\tilde{y}^i_{j,t}$, i.e., the residual after estimating the polynomial g.

$$var(\tilde{y}^i_{j,t}) = [\sigma^2_\alpha + 2\sigma_{\alpha\beta} j + \sigma^2_\beta j] + var(z^i_{j,t}) + \phi^2_t \sigma^2_\epsilon \tag{35}$$

$$cov(\tilde{y}^i_{j,t}, \tilde{y}^i_{j+n,t+n}) = [\sigma^2_\alpha + \sigma_{\alpha\beta}(2j+n) + \sigma^2_\beta j(j+n)] + \rho^n var(z^i_{j,t}) \tag{36}$$

- Note: n years/experience ahead.
- Remaining parameters identified from the persistent shock:

$$var(z^i_{1,t}) = \pi^2_t \sigma^2_\eta \tag{37}$$

$$var(z^i_{j,1}) = \pi^2_1 \sigma^2_\eta \sum_{j=0}^{h-1} \rho^{2j} \tag{38}$$

$$var(z^i_{j,t}) = \rho^2 var(z^i_{j-1,t-1}) + \pi^2_t \sigma^2_\eta, t > 1, j > 1 \tag{39}$$

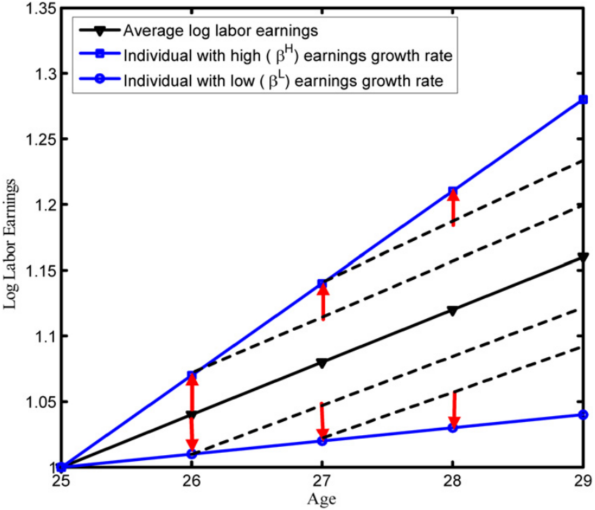
- j: experience, t: year.

Thoughts on Identification

- What is the inherent problem with identifying parameters from persistent component $z^i_{j,t}$?
- It is unobserved!
- Another name for two equation income process (and other specifications):
  - "Measurement equation": $y^i_j = \beta^i \times j + z^i_j$
  - "Transition equation": $z^i_j = \rho z^i_{j-1} + \eta^i_j$
- Learn transition equation from measurement equation observations.
- i.e., place structural restrictions on transition equation and back-out values.
- As we saw in restricted process: identification hinges on relating everything to measurement equation.

Intuitive Identification

- Heterogeneity will bias up the estimate of shock persistence.
- Good example of classical measurement error.



Sample Selection

- Data: Panel Study of Income Dynamics (PSID), 1968-1993.
- Satisfy the following criteria for at least 20 years:
  1. Male head of household;
  2. Between 20 and 64 years old;
  3. Positive labor earnings and hours;
  4. Hours each year between 520 and 5110;
  5. Wage floor and wage cap;
  6. Not in the Survey of Economic Opportunity (SEO) sample.
- This yields 1270 individuals.
- Thoughts on this selection?

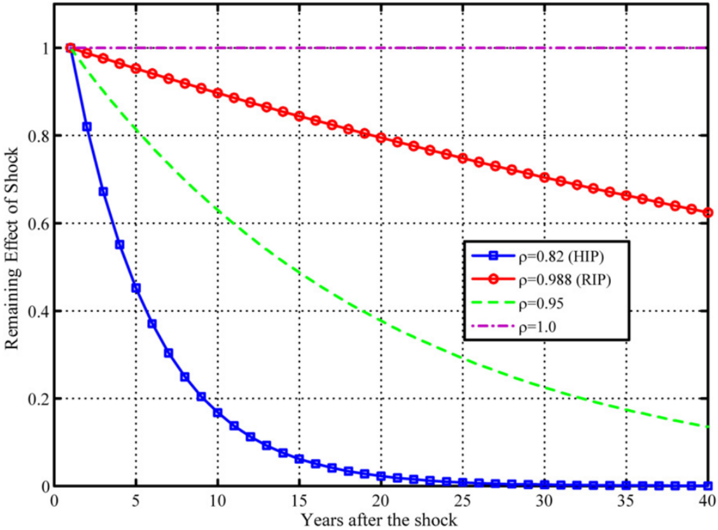
Estimated Parameters

	Sample	ρ	σ^2_α	σ^2_β	$corr_{\alpha\beta}$	σ^2_η	σ^2_ε
				Panel A: σ^2_β *restricted* to be zero (RIP process)			
(1)	All	.988	.058	–	–	.015	.061
		(.024)	(.011)			(.007)	(.010)
(2)	College	.979	.031	–	–	.0099	.047
		(.055)	(.021)			(.013)	(.020)
(3)	High-school	.972	.053	–	–	.011	.052
		(.023)	(.015)			(.007)	(.008)
				Panel B: σ^2_β *unrestricted* (HIP process)			
(4)	All	.821	.022	.00038	−.23	.029	.047
		(.030)	(.074)	(.00008)	(.43)	(.008)	(.007)
(5)	College	.805	.023	.00049	−.70	.025	.032
		(.061)	(.112)	(.00014)	(1.22)	(.015)	(.017)
(6)	High-school	.829	.038	.00020	−.25	.022	.034
		(.029)	(.081)	(.00009)	(.59)	(.008)	(.007)
(7)	All (large sample)	.842	.072	.00043	−.33	.032	.044
		(.024)	(.055)	(.00007)	(.40)	(.006)	(.008)
(8)	All (first 10 cov.)	.899	.055	.00055	−.73	.016	.047
		(.042)	(.060)	(.00013)	(.38)	(.010)	(.009)

Notes. Standard errors are in parentheses. Time effects in the variances of persistent and transitory shocks are included in the estimation in all rows. The estimated time effects for rows 1 and 4 are reported in Table 5, others are omitted to save space. The reported variances are averages over the sample period.

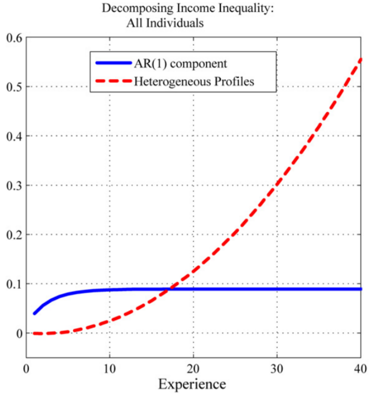
- What are some interesting features of the estimated parameters?

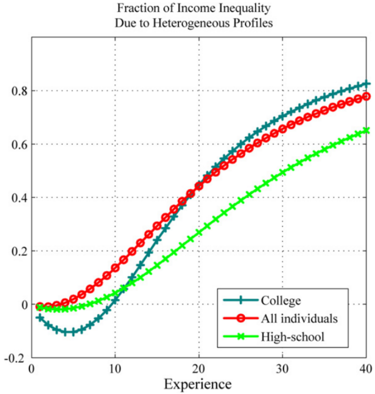
Restricted vs. Heterogeneous Persistence



- A one time shock to persistent component.

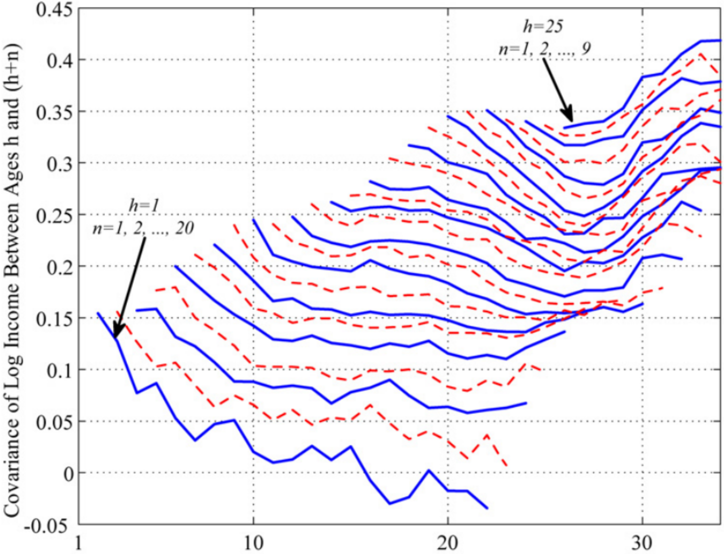
Decomposition





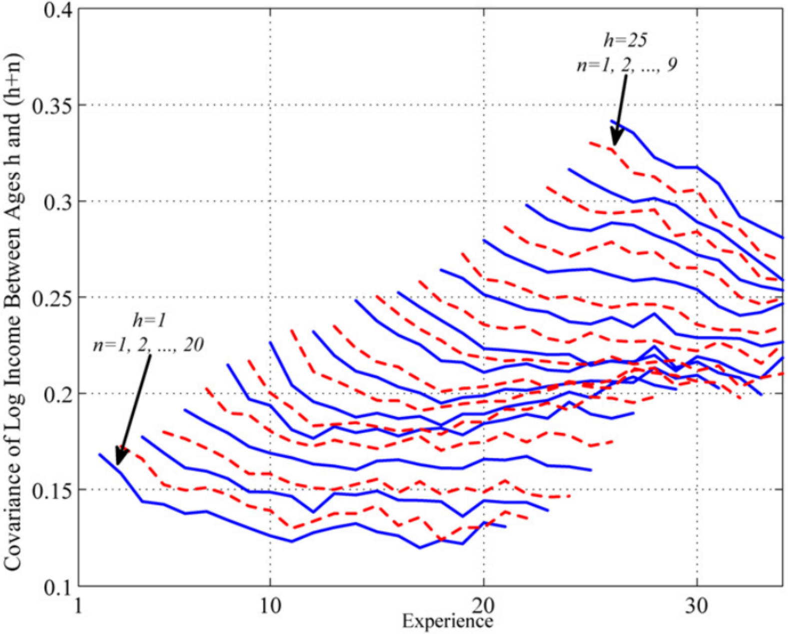
- How much income inequality is persistent shocks, i.e., AR(1), and how much is profile heterogeneity?

Covariance Structure: College

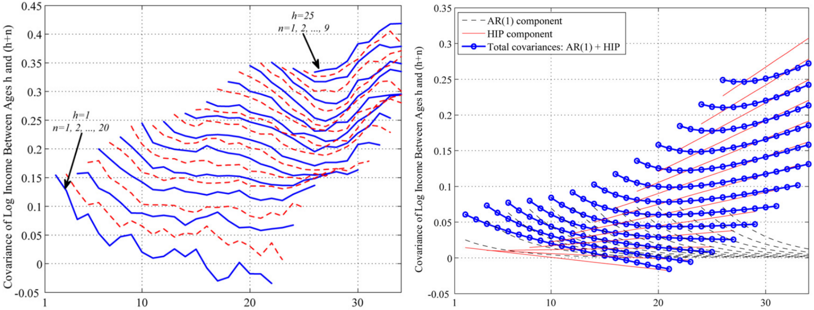


- Each line is cov. between $ln(\tilde{y})$ at exp./age h (left-most point) and h + n.

Covariance Structure: High School

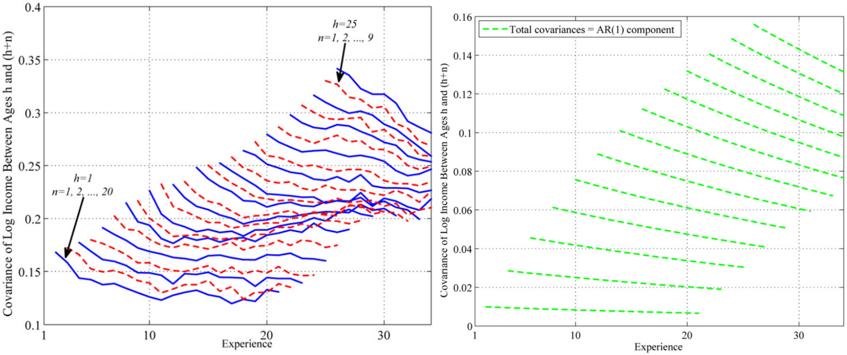


Covariance Structure: College



- College appears to have both an AR(1) component and an HIP component.
- AR(1) decreases in importance as HIP increases.
- How could this be interpreted?

Covariance Structure: High School



- High school appears to be better approximated by restricted profiles and an AR(1) shock.
- How could we interpret this?

Income Processes and Macro Models

- How do we relate these income processes to calibrating macro models?
- Straightforward in a model with competitive labor market.
- As a reminder:

$$y_j^i = \beta^i \times j + z_j^i \tag{40}$$

$$z_j^i = \rho z_{j-1}^i + \eta_j^i \tag{41}$$

Simple Model (Mostly Storesletten, Telmer, Yaron (2004))

- As a reminder

$$y_j^i = \beta^i \times j + z_j^i + \psi_j^i \tag{42}$$
$$z_j^i = \rho z_{j-1}^i + \eta_j^i \tag{43}$$

- An agent solves the following:

$$V_t(h_t, a_t, \mu) = \max_{c_t, a_{t+1}} u(c_t) + \beta E[V_{t+1}(h_{t+1}, a_{t+1}, \mu)] \tag{44}$$
$$\text{s.t. } c_t + a_{t+1} \leq (1 + r_t)a_t + w_t h_t \tag{45}$$
$$ln(h_{t+1}) = ln(h_t) + \mu + z_{t+1} + \epsilon_{t+1} \tag{46}$$
$$z_{t+1} = \rho z_t + \nu_{t+1}, \quad z_0 = 0 \tag{47}$$
$$\epsilon \sim \ iid\ N(0, \sigma_\epsilon), \quad \nu \sim \ iid\ N(0, \sigma_\nu) \tag{48}$$

- Rental rates determined competitively (exogenously, etc.):

$$w_t = F_H(K, H) \tag{49}$$
$$r_t = F_K(K, H) \tag{50}$$
$$\tag{51}$$

Mapping between our income process and the macro model

- As a reminder

$$y_j^i = \beta^i \times j + z_j^i + \psi_j^i \tag{52}$$
$$z_j^i = \rho z_{j-1}^i + \eta_j^i \tag{53}$$

- Model specified income process

$$ln(h_{t+1}) = ln(h_t) + \mu + z_{t+1} + \epsilon_{t+1} \tag{54}$$
$$z_{t+1} = \rho z_t + \nu_{t+1}, \quad z_0 = 0 \tag{55}$$
$$\epsilon \sim \ iid\ N(0, \sigma_\epsilon), \quad \nu \sim \ iid\ N(0, \sigma_\nu) \tag{56}$$

- Assume a cohort-wide rental rate and $ln(h_t)$ is proportional to log-income.
- Then, $\mu \sim N(\mu_\beta, \sigma_\beta)$;
- $\epsilon \sim N(0, \sigma_\psi)$;
- $\nu \sim N(0, \sigma_\eta)$, and ρ is identical.
- Estimate $ln(h_t)$ from the age-0 earnings distribution.

Next Time

- Review panel data methods.
- Overview of other empirical regularities in the labor market.
- To do between now and Tuesday:
  1. Install the appropriate programming languages.
  2. Start playing around with one of the panel data sources discussed in first slides (PSID, NLSY, SIPP). Or CPS (non-panel).
- I'll post helpful code next week, but I want you to check them out first.