Estimation of Income Processes

German Cubas
University of Houston

September 24, 2018

Intro

- Most of the income for most of individuals comes from the labor market.
- Understanding individual income risk is essential to model consumer behavior, to design insurance policy, etc.
- There is a big literature in labor and macro on the estimation of income processes.
- The stochastic process for labor income is a very important ingredient in macro models with heterogeneity.
- The standard assumption is that labor income is the sum of a permanent and a transitory component.

Data

- We want to model earnings dynamics.
- We require the use of panel data.
- Either consumption or income data can be used. However, we normally use income data.
- In the U.S. we normally use the PSID or the SIPP. Sometimes the NLSY can be used.

Data

- This process always starts with a data cleaning procedure.
- Only males?
- Prime age?
- Outliers? Low and high hours? Very low and very high earnings per hour?

Obtain Residual Earnings

- Earnings:

$$\tilde{Y}_{i,j,t} = w_t \exp(f(X_{i,j,t}) + u_{i,j,t})\hat{h} \tag{1}$$

In per hour terms:

$$Y_{i,j,t} = w_t \exp(f(X_{i,j,t}) + u_{i,j,t}) \tag{2}$$

Thus

$$\ln Y_{i,j,t} = y_{i,j,t} = \beta_t + f(X_{i,j,t}) + u_{i,j,t} \tag{3}$$

Structure to the Residuals

- Time invariant model of Storesletten, Telmer and Yaron (2004a)

$$u_{i,j} = \alpha_i + \eta_{i,j} + \epsilon_{i,j} \tag{4}$$

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

  where

$$\alpha \sim (0, \sigma^2_\alpha), \quad \epsilon \sim (0, \sigma^2_\epsilon), \quad \nu \sim (0, \sigma^2_\nu), \quad var(\eta_{i,-1}) = 0.$$

  and

$$\alpha_i \perp \epsilon_{i,j} \perp \nu_{i,j}, i.i.d$$

- The set of parameters to estimate is then

$$\theta = \{\rho, \sigma^2_\alpha, \sigma^2_\epsilon, \sigma^2_\nu\}$$

Cross-sectional Moments

- Let $m(\theta)_{j,n} = \mathbb{E}[u_{i,j} u_{i,j+n}]$. Then

$$\mathbb{E}[(\alpha_i + \eta_{i,j} + \epsilon_{i,j})(\alpha_i + \eta_{i,j+n} + \epsilon_{i,j+n})] =$$

$$\begin{cases} \sigma_\alpha^2 + \sigma_\epsilon^2 + \sigma_\nu^2 & if \quad j = n = 0 \\ \sigma_\alpha^2 + \rho^n \sigma_\nu^2 & if \quad j = 0, n > 0 \end{cases}$$

Identification through the Autocovariance Function

- Slope:

$$\frac{m_{03} - m_{02}}{m_{02} - m_{01}} = \frac{\sigma_\alpha^2 + \rho^3 \sigma_\nu^2 - \sigma_\alpha^2 - \rho^2 \sigma_\nu^2}{\sigma_\alpha^2 + \rho^2 \sigma_\nu^2 - \sigma_\alpha^2 - \rho \sigma_\nu^2} = \frac{\rho^2(\rho - 1)}{\rho(\rho - 1)} = \rho$$

- Difference:

$$m_{02} - m_{01} = \sigma_\nu^2 \rho(\rho - 1)$$

- Covariance

$$m_{01} = \sigma_\alpha^2 + \rho \sigma_\nu^2$$

- Variance

$$m_{00} = \sigma_\alpha^2 + \sigma_\nu^2 + \sigma_\epsilon^2$$

Estimation

- Let $\hat{m}_{j,n}$ be the empirical counterpart of $m_{j,n}$.
- The moment conditions are

$$\mathbb{E}[\lambda_{i,j,n}(\hat{m}_{j,n} - m_{j,n}(\theta))] = 0$$

where

$$\lambda_{i,j,n} = \begin{cases} 1 & if \quad i \quad is \quad present \quad at \quad j \quad and \quad j+n \\ 0 & ow \end{cases}$$

and

$$\hat{m}_{j,n} = \frac{1}{I_{jn}} \sum_{i=1}^{I_{jn}} \hat{u}_{i,j} \hat{u}_{i,j+n}$$

Estimation

- The moments can be expresses as a symmetric matrix

$$\lambda_{i,j,n} = \begin{bmatrix} m_{0,0} & m_{0,1} & \dots & m_{0,n} & \dots & m_{0,J} \\ m_{1,0} & m_{1,1} & & & & m_{1,J} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ m_{n,0} & \dots & \dots & \dots & & m_{n,J} \\ \dots & \dots & \dots & \dots & m_{J-1,J-1} & \\ m_{J,0} & \dots & \dots & m_{J,n} & \dots & m_{J,J} \end{bmatrix}$$

- Let $\bar{M} = vec(\bar{m})$ be the stacked vector of unique observations. Then θ is the solution of

$$\min_{\theta} \left([\hat{\bar{M}} - \bar{M}(\theta)]' W [\hat{\bar{M}} - \bar{M}(\theta)] \right)$$

where W is the weighting matrix.
- Optimal Weighting Matrix, Identity matrix, diagonal of optimal weighting matrix (Blundell, Pistaferri and Preston, 2008).
- Standard Errors as seen in class or bootstrap.

Issues

- Moments in levels (macro) or growth rates (labor)? See Daly, Hryshko and Manovskii (2017)
- They carry different information. Suppose an individual that appears only once. Observations surrounding missing obs are much lower than the typical ones and more volatile.
- Measurement Error: standard is assumed to be i.i.d across agents and time. Then it is included in the transitory shock.
- Put structure ($MA(q)$ model) to separate transitory shock from measurement error.

Time Varying Parameters

- Storesletten, Telmer, and Yaron (EER, 2001) allow for the conditional variance of the shocks to be different in times of expansions (σ_H^2) versus contractions (σ_L^2).

- $$u_{i,t,j} = \eta_{i,t,j} + \epsilon_{i,t} \quad (6)$$

  $$\eta_{i,t,j} = \rho\eta_{i,t-1,j-1} + \nu_{i,t} \quad (7)$$

  where

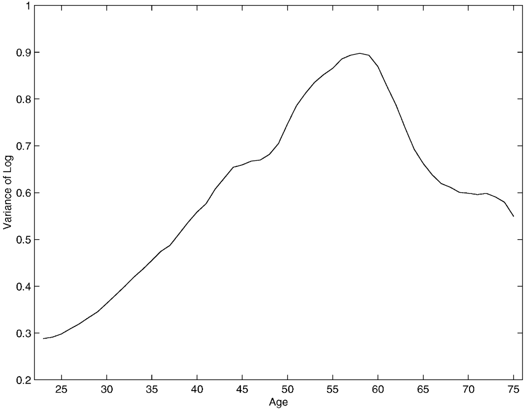
  $$\epsilon_{i,t} \sim Niid(0, \sigma_\epsilon^2), \quad \nu_{i,t} \sim Niid(0, \sigma_\nu^2(Y_t))$$

  and

  $$\sigma_\nu^2(Y_t) = \begin{cases} \sigma_H^2 & if \quad expansion \quad at \quad t \\ \sigma_L^2 & if \quad contraction \quad at \quad t \end{cases}$$

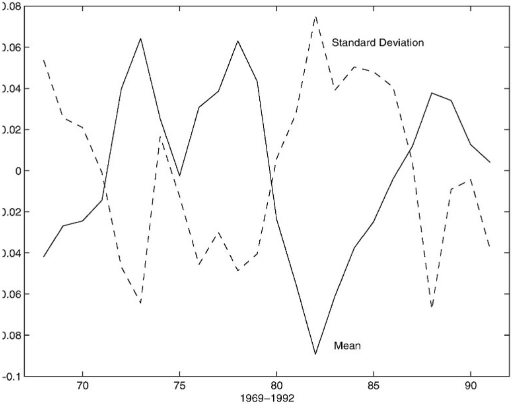
Time Varying Parameters

Over the working years earnings dispersion increases, loosely speaking, linearly ($\rho = 1$).



Time Varying Parameters

Countercyclical heteroskedasticity is a striking feature of the data. The correlation of the detrended mean and the standard deviation is -0.74.



Time Varying Parameters Identification

- Ignore the transitory shocks. Suppose that there were only three generations: Young, middle aged and old.
- Suppose also that the economy is in an expansion at the current time, but was in a recession during the previous 2 years.
- Suppose that we only observe data dated at the current time, period t.
- The population cross-sectional variances of the idiosyncratic processes, u, for each generation are

$$
\begin{aligned}
E(u_{i,t,1})^2 &= \sigma_H^2 \\
E(u_{i,t,2})^2 &= \sigma_H^2 + \rho^2\sigma_L^2 \\
E(u_{i,t,3})^2 &= \sigma_H^2 + \rho^2\sigma_L^2 + \rho^4\sigma_L^2
\end{aligned}
$$

Time Varying Parameters Estimation

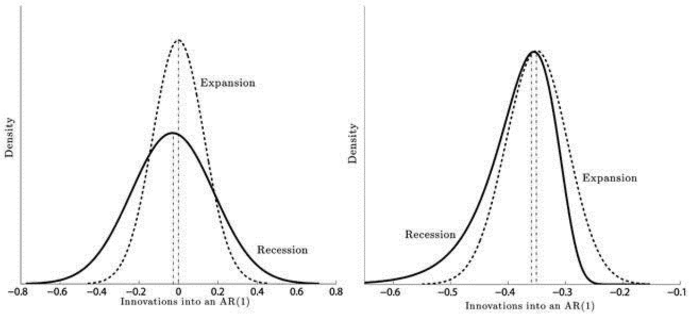
- The method relies on having many obs. on u for each generation. It does not requires to have time-series observations on individual agents.
- The key piece of information we are exploiting is how the cross-sectional variance at date t varies across age cohorts and how this interacts with what is essentially a cohort-specific macroeconomic history which is known at date t.
- Results

$$\begin{aligned}\rho &= 0.916\\ \sigma_H^2 &= 0.037\\ \sigma_L^2 &= 0.181\\ \sigma_\epsilon^2 &= 0.025\end{aligned}$$

Guvenen, Ozkan and Song (JPE, 2014)

- Variance of idiosyncratic shocks is not countercyclical.
- Instead, it is the left-skewness of shocks that is strongly countercyclical.
- During recessions, large upward earnings movements become less likely, whereas large drops in earnings become more likely
- Therefore, relative to the earlier literature that argued for increasing variance – which results in some individuals receiving larger positive shocks during recessions – these results are more pessimistic: uncertainty increases in recessions without an increasing chance of upward movements

Guvenen, Ozkan and Song (JPE, 2014)



HIP vs. RIP

- *RIP*: Restricted Income Profiles
  Individuals are subject to large and persistent income shocks but have similar life cycle income profiles.

- *HIP*: Heterogeneous Income Profiles-Guvenen (RED, 2009)
  Individuals are subject to income shocks with modest persistence, while facing individual-specific income profiles

HIP vs. RIP

- Guvenen (RED, 2009) revived Lillard and Weiss (1979)

$$u_{i,j} = \alpha_i + \beta_i j + \eta_{i,j} + \epsilon_{i,j} \tag{8}$$

$$\eta_{i,j} = \rho \eta_{i,j-1} + \nu_{i,j} \tag{9}$$

  where α_i and β_i are deterministic individual specific intercept and slope.

- For instance, the source of differences in β can come from returns to human capital accumulation. Early estimates are $0.5 < \rho < 0.7$ and $\sigma^2_\beta >> 0$

- MaCurdy (1982) cast doubt on these findings. He is not able to reject $\sigma^2_\beta = 0$. Thus, all the literature evolved assuming RIP and found very large ρ's (> 0.97).

HIP vs. RIP

- Guvenen (RED, 2009)
  - Assuming away the heterogeneity in income growth rates (as is done in the RIP process), when in fact it is present, biases the estimated persistence parameter upward.
  - Intuition: An individual with high (alternatively, low) income growth rate will systematically deviate from the average profile.
  - This fact will then lead the econometrician to interpret this systematic fanning out as the result of persistent positive (or negative) income shocks every period.
  - He provides an example of a simulation in which the persistence parameter is estimated to be about 0.90 if RIP is assumed, instead of the true value of zero.

HIP vs. RIP

- Hryshko (QE, 2012)
    - Use data on idiosyncratic labor income growth from the Panel Study of Income Dynamics.
    - Find that the estimated variance of deterministic income growth is zero, that is, the HIP model can be rejected. The RIP model with a permanent component cannot be rejected.

Age-Dependent Income Processes

- Karahan and Ozkan (RED, 2013)
  - How does the persistence of earnings change over the life cycle?
  - Do workers at different ages face the same variance of idiosyncratic shocks?

Age-Dependent Income Processes

- Intuition:
  - For young workers, job-to-job transitions might play an important role.
  - Midway through a career, settling down into senior positions as well as bonuses, promotions, or demotions may account for workers earnings dynamics.
  - Older workers are more likely to develop health problems that reduce their productivity. These changes differ in nature and, more specifically, in persistence and magnitude.

Age-Dependent Income Processes

- Specification

$$u_{i,j,t} = \alpha_i + \eta_{i,j,t} + \phi_t \epsilon_{i,j}$$

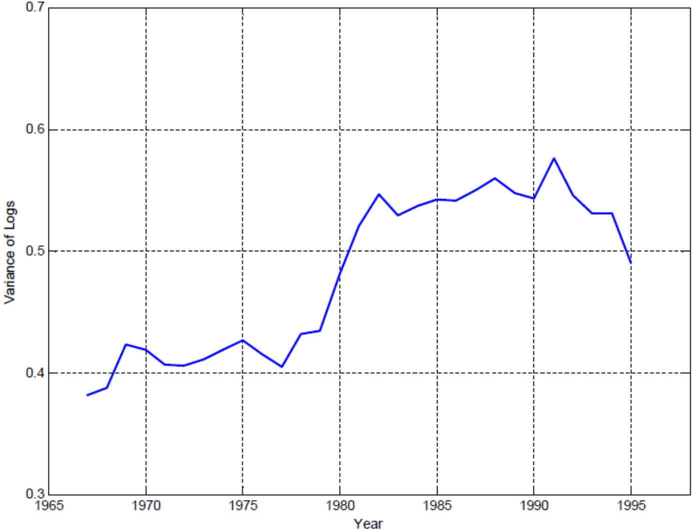
$$\eta_{i,j,t} = \rho_{j-1}\eta_{i,j-1,t-1} + \Pi_t \nu_{i,j}$$

with $\eta_{i,1,t} \sim F(0, \Pi_t^2, \sigma_{\eta_1}^2)$

- This paper: age dependent $\sigma_{\epsilon,j}^2$, $\sigma_{\nu,j}^2$ and ρ_j
- Also Π_t and ϕ_t: change in residual inequality over time
- Identification using the variance/covariance structure.

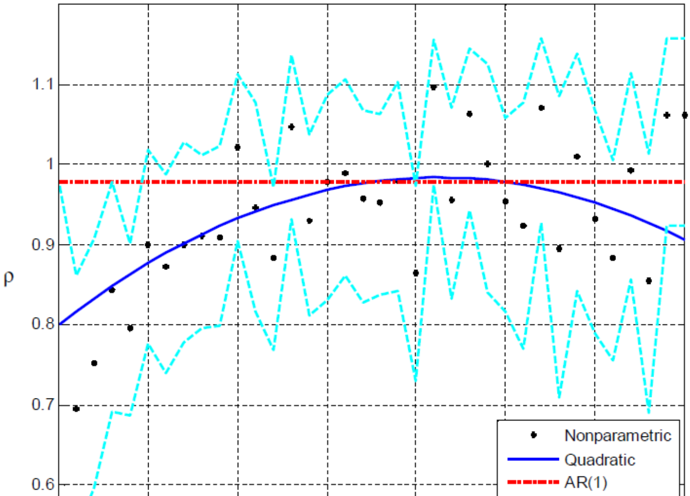
Age-Dependent Income Processes

Figure 1: Residual Inequality over Time



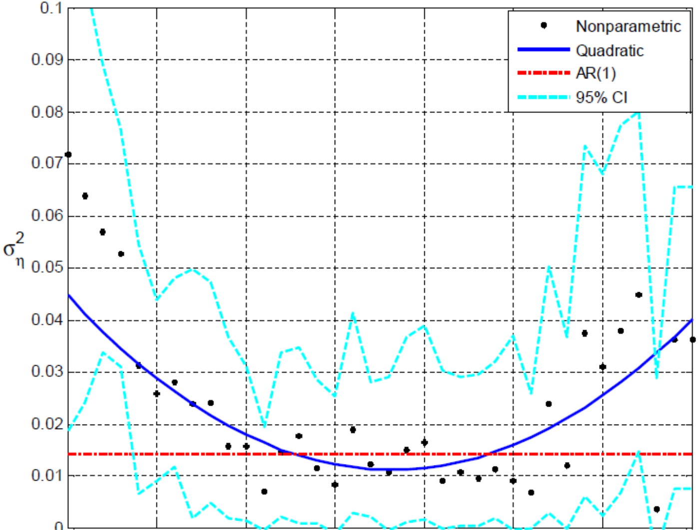
Age-Dependent Income Processes

Figure 2: Persistence Profile



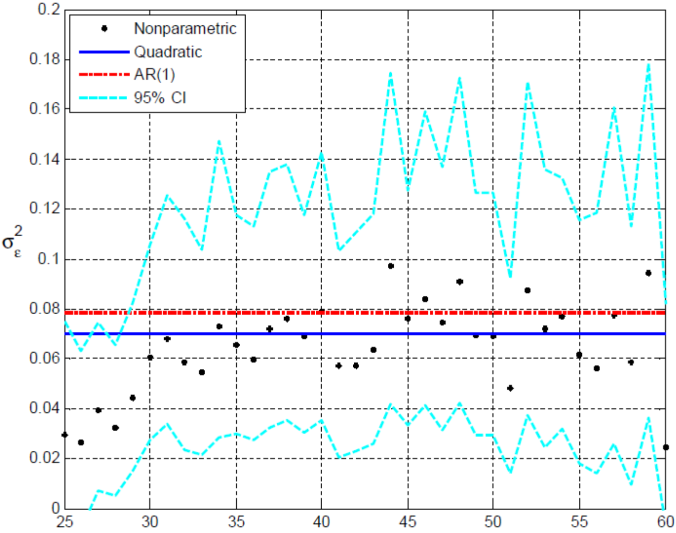
Age-Dependent Income Processes

Figure 3: Variance Profile of Persistent Shocks



Age-Dependent Income Processes

Figure 4: Variance Profile of Transitory Shocks



Age-Dependent Income Processes

Figure 6: Lifetime Profile of Residual Inequality