Lecture VIII

Income Process: Facts, Estimation, and Discretization

Gianluca Violante

New York University

Quantitative Macroeconomics

Estimation of income process

- Competitive model: unique market price per efficiency unit $\rightarrow$ wage distribution reflects heterogeneity in efficiency units:

$$Y_{iat} = P_t \cdot \exp\left(\tilde{Y}_{iat}\right)$$

- Compute residuals y_{iat} from regression in logs

$$\log Y_{iat} = \mathbf{D}_t + \beta' \mathbf{X}_{iat} + y_{iat}$$

  time dummies $\mathbf{D}_t = [\log P_1, \log P_2, ...]$, and $\mathbf{X}_{iat}$ are fixed observables that capture predictable part of $\tilde{Y}_{iat}$: age, edu, race

- Gender: be aware of of selection. Typically use data on men.

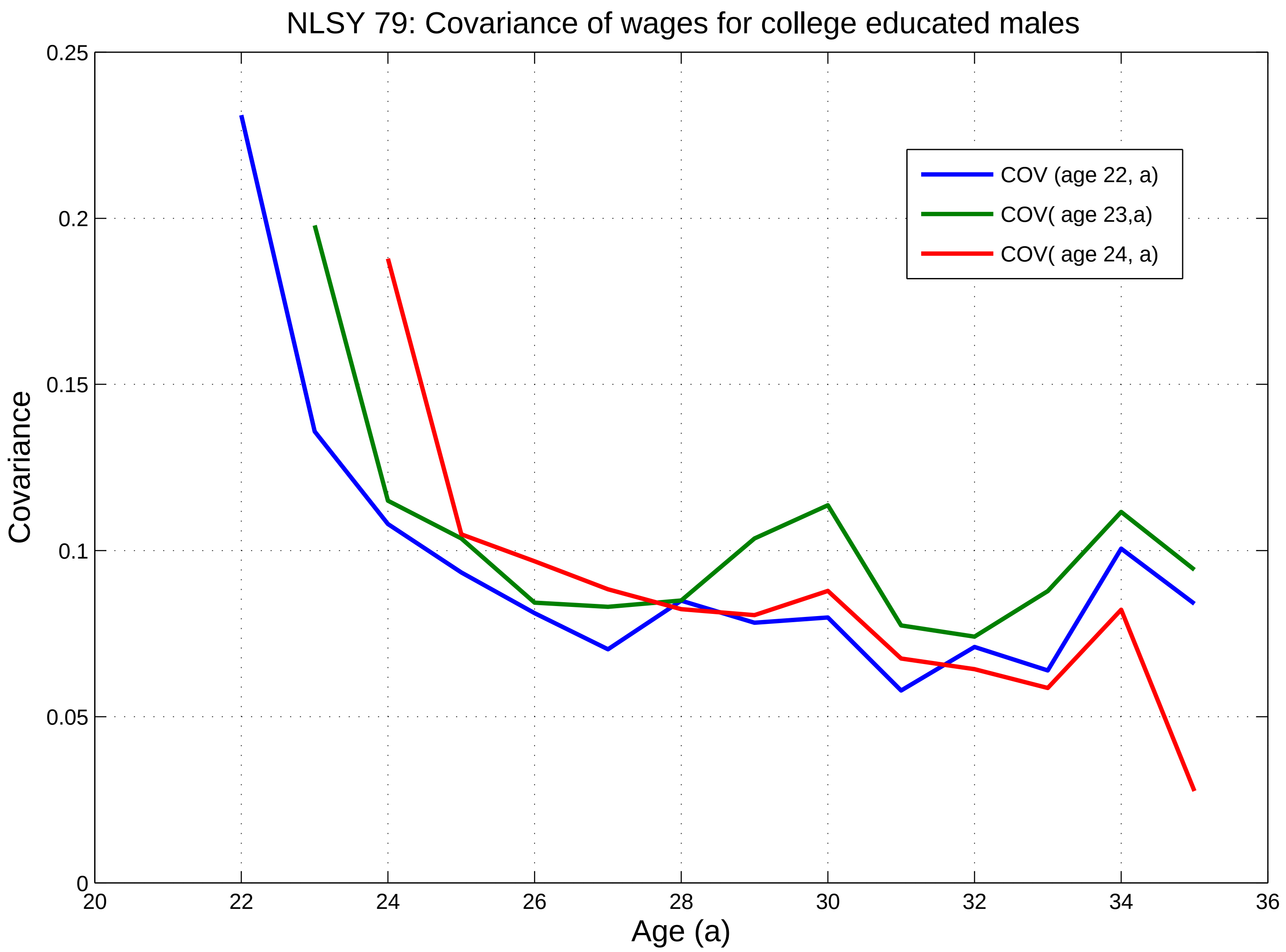
- Important to distinguish between wages (endogenous labor supply), individual earnings (two-earner hh, inelastic labor supply), hh earnings (one earner hh, inelastic labor supply), hh income (endowment economy)

Step 1: choose statistical model

- Assume process is stationary: when you compute moment conditions, do it t by t, but then average across t.
- Identification/estimation in nonstationary process is more complicated, see Heathcote-Storesletten-Violante (2010).
- We choose following specification for log earnings process:

$$
\begin{aligned}
y_{ia} &= z_{ia} + \varepsilon_{ia} \\
z_{ia} &= \rho z_{i,a-1} + u_{ia} \\
& u_{ia} \overset{iid}{\sim} (0, \sigma_u) \\
& z_{i0} \overset{iid}{\sim} (0, \sigma_{z_0}) \\
& \varepsilon_{ia} \overset{iid}{\sim} (0, \sigma_\varepsilon)
\end{aligned}
$$

- Why persistent + transitory?



NLSY 79: Covariance of wages for college educated males
Covariance
Age (a)
0.25
0.2
0.15
0.1
0.05
0
20
22
24
26
28
30
32
34
36
COV (age 22, a)
COV(age 23,a)
COV(age 24, a)

Identification in levels

$$\begin{aligned} y_{ia} &= z_{ia} + \varepsilon_{ia} \\ z_{ia} &= \rho z_{i,a-1} + u_{ia} \end{aligned}$$

$$\begin{aligned} var\left(y_{i0}\right) &= \sigma_{z0} + \sigma_{\varepsilon} && \text{for } a = 0 \\ var\left(y_{ia}\right) &= var\left(z_{ia}\right) + \sigma_{\varepsilon} && \text{for } a > 0 \\ var\left(z_{ia}\right) &= \rho^2 var\left(z_{i,a-1}\right) + \sigma_u \\ cov\left(y_{ia}, y_{i,a-j}\right) &= cov\left(z_{ia}, z_{i,a-j}\right) && \text{for } j > 0 \\ cov\left(z_{ia}, z_{i,a-j}\right) &= \rho^j var\left(z_{i,a-j}\right) \end{aligned}$$

- Identification of ρ from the slope of the covariance at lags > 0:

$$\frac{cov\left(y_{ia}, y_{i,a-2}\right)}{cov\left(y_{i,a-1}, y_{i,a-2}\right)} = \frac{\rho^2 var\left(z_{i,a-2}\right)}{\rho var\left(z_{i,a-2}\right)} = \rho$$

Identification in levels

- Identification of σ_ε from the difference between variance and covariance

$$var\left(y_{i,a-1}\right)-\rho^{-1}cov\left(y_{ia},y_{i,a-1}\right)=var\left(z_{i,a-1}\right)+\sigma_\varepsilon-var\left(z_{i,a-1}\right)$$

- Given σ_ε, obtain σ_{z0} residually from $var\left(y_{i,0}\right)$

- Finally, identification of σ_u:

$$\begin{aligned} var\left(y_{i,a-1}\right)-cov\left(y_{ia},y_{i,a-2}\right)-\sigma_\varepsilon &= \\ \rho^2 var\left(z_{i,a-2}\right)+\sigma_u+\sigma_\varepsilon-\rho^2 var\left(z_{i,a-2}\right)-\sigma_\varepsilon &= \sigma_u \end{aligned}$$

- Identification achieved with two lags: $(a-2,a-1,a)$.

- Typically, with long panels ($T\simeq 10-15$) model is largely overidentified, and parameters tightly estimated because N is large (thousands of observations) but if nonstationary...

Identification in first differences (for $\rho = 1$)

$$\begin{aligned}
\Delta y_{ia} &= u_{ia} + \varepsilon_{ia} - \varepsilon_{i,a-1} \\
\Delta y_{i,a-1} &= u_{i,a-1} + \varepsilon_{i,a-1} - \varepsilon_{i,a-2} \\
var\left(\Delta y_{ia}\right) &= \sigma_u + 2\sigma_\varepsilon \\
cov\left(\Delta y_{ia}, \Delta y_{i,a-1}\right) &= -\sigma_\varepsilon
\end{aligned}$$

- Of course, you can use moments in first differences as additional moments, together with those in level, to estimate the parameters of $AR(1)$
- Not-so-much-fun fact: if you estimate a permanent-transitory process in levels and first-differences, you obtain very different results. Why?

Heterogeneous Income Profiles

- A common alternative specification to the persistent-transitory representation of idiosyncratic income risk is:

$$
\begin{aligned}
y_{ia} &= \alpha_i + \beta_i a + z_{ia} \\
\\
z_{ia} &= \rho z_{i,a-1} + u_{ia} \\
u_{ia} &\overset{iid}{\sim} (0, \sigma_u) \\
(\alpha_i, \beta_i) &\overset{iid}{\sim} (0, \Sigma)
\end{aligned}
$$

- (α_i, β_i) are, respectively, constant and slope of earning profile, heterogeneous across individuals

- Typically, z_{it} estimated to be a lot less persistent

- HIP difficult to distinguish, empirically, from persistent-transitory

Step 2: Estimation

- You have an unbalanced panel of individuals aged $a = 1, ..., A$.
- For every individual $i = 1, ..., I$ define a $(A \times 1)$ vector

$$\mathbf{d}_i = \begin{pmatrix} d_{i1} \\ ... \\ d_{iA} \end{pmatrix}$$

  where $d_{ia} \in \{0, 1\}$ is an indicator variable for whether individual i is present at age a in the sample.
- Analogously to $\mathbf{d}_{ia}$ define an $(A \times 1)$ vector

$$\mathbf{y}_i = \begin{pmatrix} y_{i1} \\ ... \\ y_{iA} \end{pmatrix}$$

  since the panel is unbalanced, must set missing elements to zero

Step 2: Estimation

- The covariance of earnings is then a $(A \times A)$ symmetric matrix computed as

$$\mathbf{C} = \frac{\mathbf{Y}}{\mathbf{D}}$$

  This is an element by element division, where $\mathbf{Y}$ and $\mathbf{D}$ are $(A \times A)$ matrices given by

$$\mathbf{Y} = \sum_{i=1}^{I} \mathbf{y}_i \mathbf{y}_i' \qquad \mathbf{D} = \sum_{i=1}^{I} \mathbf{d}_i \mathbf{d}_i'$$

- Each element of $\mathbf{y}_i \mathbf{y}_i'$ is product of earnings at ages (p, q) for i
- Each element of $\mathbf{d}_i \mathbf{d}_i'$ is 1 only if i is present at both ages (p, q).
- Each entry of $\mathbf{C}$ is the cross-sectional covariance of earnings at ages (p, q)

Step 2: Estimation

- Take the upper triangular portion of $\mathbf{C}$ and vectorize it into an $(A(A+1)/2, 1)$ vector.

$$\mathbf{m} = vech\left(\mathbf{C}^{UT}\right)$$

- Let $\mathbf{f}(\Theta)$ be the comforming $(A(A+1)/2, 1)$ vector of empirical moments.

- Estimation based on a minimum distance estimator that minimizes the distance btw the model covariance matrix and the empirical covariance matrix (Chamberlain, 1984) — family of GMM.

- The estimator that solves the following minimization problem

$$\min_{\Theta} \left[\mathbf{m} - \mathbf{f}(\theta)\right]' \Omega \left[\mathbf{m} - \mathbf{f}(\theta)\right]$$

  where Ω is a weighting matrix and θ is a $(n, 1)$ vector of parameters.

Weighting matrix

- Define, comformably with $\mathbf{m}$, the vector $\mathbf{m}_i$ that contains the distinct elements of the cross-product matrix $\mathbf{y}_i\mathbf{y}_i'$.
- Similarly, define the vector $\mathbf{s} = vech\left(\mathbf{D}^{UT}\right)$. Chamberlain proved that $\mathbf{m}$ has asymptotic variance

$$\mathbf{V} = \frac{\sum_{i=1}^{I}\left(\mathbf{m}-\mathbf{m}_i\right)\left(\mathbf{m}-\mathbf{m}_i\right)'}{\mathbf{s}\mathbf{s}'}$$

  where $\mathbf{V}$ is a $\left(A\left(A+1\right)/2, A\left(A+1\right)/2\right)$ matrix

- Chamberlain shows that the optimal weighting matrix Ω is $\mathbf{V}^{-1}$
- Altonji and Segal (1996) find, by MC simulations, that there is substantial small sample bias in the estimates of θ and recommend using (i) $\Omega = I$ (equally weighted estimator) or $\Omega = diag\left(\mathbf{V}^{-1}\right)$ (diagonally weighted estimator)

Step 3: S.E. and Testing

- Standard errors of θ are then computed as

$$\left(\mathbf{G}'\Omega\mathbf{G}\right)^{-1}\mathbf{G}'\Omega\mathbf{V}\boldsymbol{\Omega}\mathbf{G}\left(\mathbf{G}'\Omega\mathbf{G}\right)^{-1}$$

  where $\mathbf{G}$ is the $\left(A\left(A+1\right)/2, n\right)$ gradient matrix $\partial\mathbf{f}\left(\theta\right)/\partial\theta$ evaluated at θ^*.

- Finally, the chi-squared goodness of fit statistic is

$$\left[\mathbf{m}-\mathbf{f}\left(\theta^*\right)\right]\mathbf{R}^{-}\left[\mathbf{m}-\mathbf{f}\left(\theta^*\right)\right]' \sim \chi_q^2$$

  where $\mathbf{R}^{-}$ is the generalized inverse of $\mathbf{R}=\mathbf{W}\mathbf{V}\mathbf{W}'$ where $\mathbf{W}=\mathbf{I}-\mathbf{G}\left(\mathbf{G}'\Omega\mathbf{G}\right)^{-1}\mathbf{G}'\Omega$, and $q=A(A+1)/2-n$, are the the degrees of freedom.

Higher moments of income distribution

Three approaches to think about skewness and excess kurtosis with a continuous process:

1. Assume innovations drawn from skewed and leptokurtic distributions and add higher moments to estimation

2. Clark (ECMA, 1973): subordinated stochastic processes

   - $Y(T(t))$, where $T(t)$ is the directing process that sets number of times Y changes in period t
   - Both $Y(t)$ and $T(t)$ have independent increments
   - If ΔY is normal and ΔT lognormal, $Y(T(t))$ displays kurtosis.

3. Kou (Management Science, 2002): combination of diffusion plus jump process, where jump is drawn from an asymmetric double-exponential distribution

Step 4: Discretization

- Two dominant methodologies:

  1. Tauchen method (Tauchen EL, 1986)

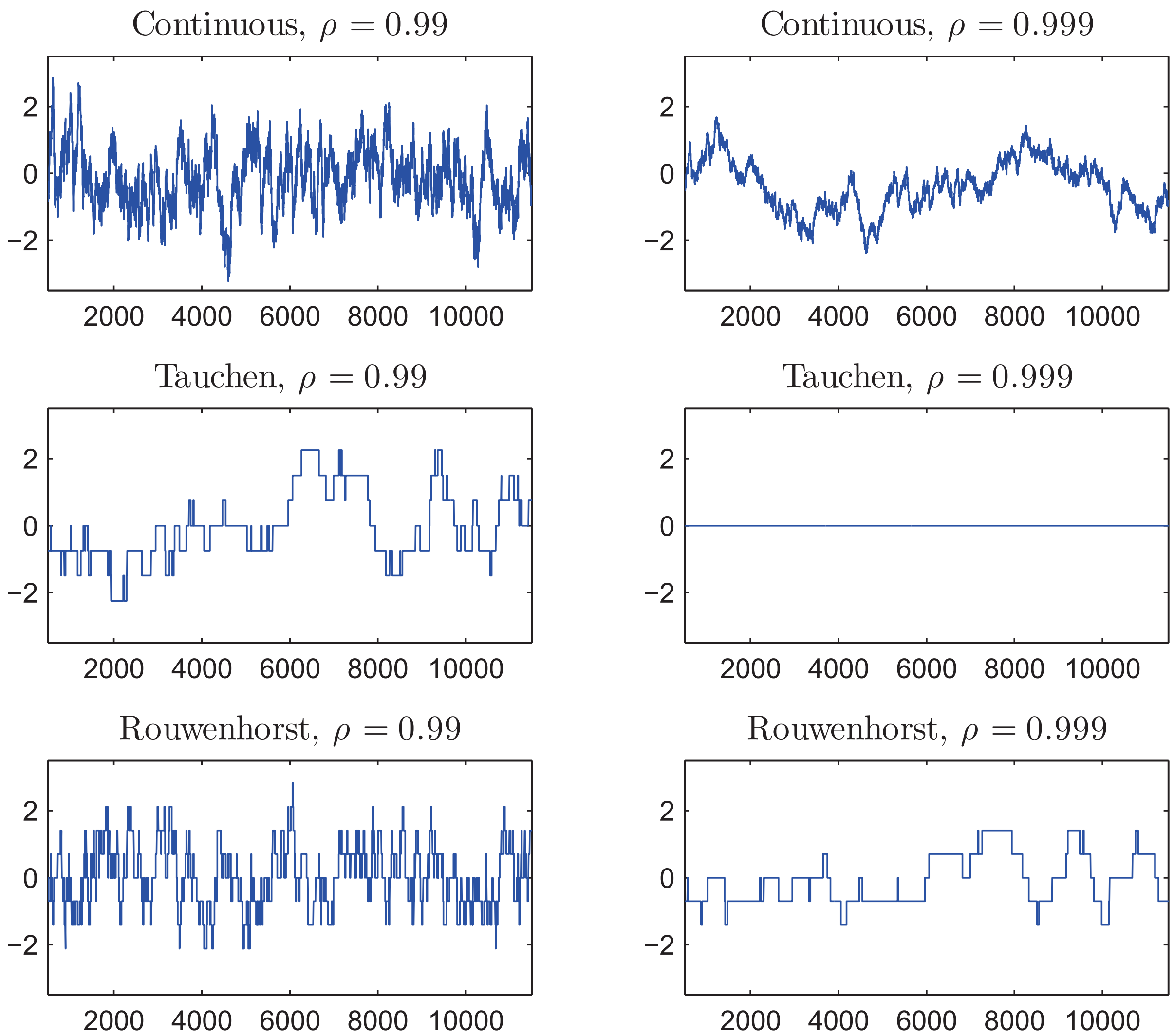
  2. Rowenhorst method (article in “Frontiers of business cycle”, (Tom Cooley, editor, 1995)

     - Rowenhorst method better as $\rho \to 1$, plus one can derive first four moments in closed form

     - Two approaches to derive moments in closed form: Kopecki-Suen (RED, 2010) and Lkhagvasuren (2012)

     - Multivariate version: Terry-Knotek (EL, 2011), Galindev-Lkhagvasuren (JEDC, 2010)

Tauchen vs Rowenhorst (9 points, $\sigma_y = 1$)



Rowenhorst method to discretize AR(1)

- Consider the following two-state Markov chain $x \in \{0, 1\}$ where

$$\begin{aligned} \Pr\left(x' = 0 | x = 0\right) &= p & \Pr\left(x' = 1 | x = 0\right) = 1 - p \\ \Pr\left(x' = 0 | x = 1\right) &= 1 - q & \Pr\left(x' = 1 | x = 1\right) = q \end{aligned}$$

- Compute the invariant distribution

$$\begin{bmatrix} 1-\alpha & \alpha \end{bmatrix} \begin{bmatrix} p & 1-p \\ 1-q & q \end{bmatrix} = \begin{bmatrix} 1-\alpha & \alpha \end{bmatrix}$$

A system of 2 equations in 2 unknowns with with solution

$$\alpha \equiv \Pr(x = 1) = \frac{1-p}{2-p-q}$$

Moments in closed form

- Bernoulli invariant distr. $\rightarrow$ we can compute moments analytically:

$$
\begin{aligned}
E(x) &= \alpha = \frac{1-p}{2-p-q} \\
Var(x) &= \alpha(1-\alpha) = \frac{(1-p)(1-q)}{(2-p-q)^2} \\
Skew(x) &= \frac{1-2\alpha}{\sqrt{\alpha(1-\alpha)}} = \frac{p-q}{\sqrt{(1-p)(1-q)}} \\
Ekurt(x) &= -6 + \frac{1}{\alpha(1-\alpha)} = -2 + \frac{(p-q)^2}{(1-p)(1-q)} \\
Corr(x, x') &= p + q - 1
\end{aligned}
$$

Generalization to n-state process

- Consider the auxiliary process

$$X_n = \sum_{i=1}^{n-1} x_i$$

where each x is an independent 2-state Markov chain

- Therefore $X_n \in \{0, 1, ..., k, ..., n-1\}$. Define the process:

$$y = \sum_{i=1}^{n-1} (a + bx_i)$$

and choose a and b so that y takes values in $[m - \Delta, m + \Delta]$
where m is the mean and 2Δ the range of the state space.

Generalization to n-state process

- By setting $y_{\min} = m - \Delta$ and $y_{\max} = m + \Delta$, it is easy to see that

$$\begin{aligned} a &= \frac{m - \Delta}{n - 1} \\ b &= \frac{2\Delta}{n - 1} \end{aligned}$$

- It follows that the kth point of the grid $(0 < k < n - 1)$ is:

$$y_k = m - \Delta + \frac{2\Delta}{n - 1} k$$

- We have the grid for y as a function of three parameters (n, m, Δ)

Moments in closed form for n-state process

- Then, it is easy to see that:

$$
\begin{aligned}
E(y) &= (n-1)\left[a + bE(x)\right] = m - \Delta + 2\Delta \frac{1-p}{2-p-q} = m + \Delta \frac{q-p}{2-p-q} \\
Var(y) &= (n-1)\, b^2 Var(x) = \frac{4\Delta^2}{n-1} \frac{(1-p)(1-q)}{(2-p-q)^2} \\
Corr(y, y') &= Corr(x, x') = p + q - 1
\end{aligned}
$$

- Note: don't need to set $q = p$ for the process to have mean zero.
- Less easy to show that:

$$
\begin{aligned}
Skew(y) &= \frac{Skew(x)}{\sqrt{(n-1)}} = \frac{p-q}{\sqrt{(n-1)(1-p)(1-q)}} \\
Ekurt(y) &= \frac{Ekurt(x)}{n-1} = \frac{-2}{n-1} + \frac{(p-q)^2}{(n-1)(1-p)(1-q)}
\end{aligned}
$$

therefore, important constraint: $Ekurt(y) = -2/(n-1) + Skew(y)^2$

Exact moment matching

- 5 moments $\{E(y), Var(y), Corr(y,y'), Skew(y), Ekurt(y)\}$ and 5 parameters $\{m, \Delta, n, p, q\}$

- Moment matching:

$$\begin{aligned}(p,q,n) &\rightarrow Corr(y,y'), Skew(y), Ekurt(y)\\ (m,\Delta) &\rightarrow E(y), Var(y)\end{aligned}$$

- In the data (SCF log household earning residuals, ages 22-60):

  $$E(y) = 0, Var(y) = 0.8, Corr(y,y') = 0.95, Skew(y) = -1.1, Ekurt = 4$$

  so one must give up either on the skewness or on the kurtosis...

- Homework: estimate parameters to fit those moments. Simulate process. Can you replicate the var, skewness and kurtosis of the one year changes? Can you derive the moments in first differences in closed form like the ones in level?

Stationarizing the stochastic growth model

- The representative household solves:

$$\max_{\{I_t, C_t, N_t\}} \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \frac{\left[C_t^{\phi}\left(1-N_t\right)^{1-\phi}\right]^{1-\sigma}}{1-\sigma}$$

$$s.t.$$

$$\begin{aligned} C_t + I_t &= \left(1-\tau_t\right) W_t N_t + r_t K_t + \Phi_t \\ K_{t+1} &= \left(1-\delta\right) K_t + I_t \\ Z_t &= G^t z_t, \qquad z_t \text{ stationary,} \quad G = 1+g \end{aligned}$$

- The representative firm solves (F is CRS):

$$\max_{\{N_t, K_t\}} F\left(K_t, Z_t N_t\right) - W_t N_t - r_t K_t$$

- The government balances its budget: $\Phi_t = \tau_t W_t N_t$

Stationarizing the stochastic growth model

- Define a blue new set of stationary variables:

$$c_t \equiv C_t/G^t, k_t \equiv K_t/G^t, i_t \equiv I_t/G^t, \phi_t \equiv \Phi_t/G^t, w_t \equiv W_t/G^t, z_t \equiv Z_t/G^t$$

- Rewrite the household problem as:

$$\max_{\{i_t, c_t, N_t\}} \mathbb{E}_0 \sum_{t=0}^{\infty} \tilde{\beta}^t \frac{\left[c_t^{\phi} (1 - N_t)^{1-\phi}\right]^{1-\sigma}}{1-\sigma}$$

$$s.t.$$

$$\begin{aligned} c_t + i_t &= (1 - \tau_t) w_t N_t + r_t k_t + \phi_t \\ k_{t+1} G &= (1 - \delta) k_t + i_t \end{aligned}$$

where $\tilde{\beta}^t = \left(\beta G^{\phi(1-\sigma)}\right)^t$

Stationarizing the stochastic growth model

- By CRS, the firm problems becomes

$$\max_{N_t,k_t} F\left(k_t, z_t N_t\right) - w_t N_t - r_t k_t$$

- The government budget constarint becomes

$$\phi_t = \tau_t w_t N_t$$

- Thus all the stationarized variables grow at rate $(G-1)$

- Labor (btw 0 and 1) and interest rate (Y/K) do not grow

- The model has a balanced growth path

- Key: preferences s.t. inc. effect = subst. effect on labor supply

Stationarized Euler equation

$$
\begin{aligned}
c_t^{\phi(1-\sigma)-1}(1-N_t)^{(1-\phi)(1-\sigma)} &= \lambda_t \\
\tilde{\beta}^t \lambda_t G &= \tilde{\beta}^{t+1} \lambda_{t+1} \left[1 + F_k\left(k_{t+1}, z_{t+1} N_{t+1}\right) - \delta\right]
\end{aligned}
$$

which implies:

$$
\beta^t G^{\phi(1-\sigma)t+1} \lambda_t = \beta^{t+1} G^{\phi(1-\sigma)(t+1)} \lambda_{t+1} \left(1 + r_{t+1} - \delta\right)
$$

$$
G^{1-\phi(1-\sigma)} \left(\frac{c_{t+1}}{c_t}\right)^{1-\phi(1-\sigma)} \cdot \left(\frac{1-N_t}{1-N_{t+1}}\right)^{(1-\phi)(1-\sigma)} = \beta \left[1 + F_k\left(k_{t+1}, z_{t+1} N_{t+1}\right) - \delta\right]
$$

- We could have written the FOC from the model with trends and stationarized it (note that F_k is homogenous of degree zero). What is preferable? It depends how you solve the model!