RIP to HIP: The Data Reject Heterogeneous Labor Income Profiles

Dmytro Hryshko

Econ 350, Spring 2022

Idiosyncratic labor income

Consider the following model for labor income of individual i with h years of labor market experience at time t:

$$Y_{iht} = \exp(\alpha_t + \gamma_t' X_{iht}) \exp(y_{iht})$$
$$\log(Y_{iht}) = \alpha_t + \gamma_t' X_{iht} + y_{iht}.$$

X_{iht}: education and a polynomial in age/potential experience. Observable controls normally explain about 30% of variation in individual labor incomes.

I model the *idiosyncratic* component of labor income, y_{iht}.

Heterogeneity and labor income risk

Idiosyncratic income, y_{iht}, comprises heterogeneity and individual-specific shocks to incomes.

Heterogeneity:

- in initial incomes (e.g., due to abilities);
- in income profiles (idiosyncratic growth rates due to differential human capital investment).

Shocks differ in their “durability”:

- persistent/permanent shocks (e.g., disability, promotion, demotion, displacement);
- transitory shocks (e.g., short unemployment spells, bonuses).

Heterogeneity versus labor income risk

- The importance of shocks (uncertainty) versus initial conditions (heterogeneity) for the *life-cycle* profiles of earnings and welfare inequality (e.g., Huggett et al., 2007), and consumption inequality (Guvenen 2007).
- The choice of an appropriate model of the variation in individual and household idiosyncratic incomes used in macro models with heterogenous agents and uninsurable labor income risks.

Heterogeneity versus labor income risk

- Understanding insurability of shocks.
- Matters for policy. If most of the variation is due to heterogeneity target the initial conditions (e.g., education for disadvantaged). If most of the variation is due to shocks invest into insurance policies, or educate about insurance markets.
- Permanent labor income risk may affect economic growth (Krebs 2003), and make the costs of business cycles sizable (De Santis 2007).

Early ideas on labor income models

Friedman and Kuznets (1954): labor income can be modeled as the sum of permanent, quasi-permanent and purely transitory components.

They were not very specific on the model of a permanent component—could be heterogeneity or shocks.

In modern time series language, purely transitory component is an i.i.d. shock; quasi-permanent component is a mean-reverting stochastic process—normally AR(1), MA(1), or ARMA(1,1).

HIP: Heterogeneous Income Profiles

$$y_{iht} = \underbrace{\alpha_i + \beta_i h}_{\text{heterogeneity}} + \underbrace{\tau_{iht}}_{\text{risk}} + \underbrace{u_{iht,me}}_{\text{meas. error}}$$

$$\tau_{iht} = \theta(L)\epsilon_{iht}$$

h—labor market experience;

β_i—individual i's growth rate of income;
α_i—individual i's initial level of income;

$\theta(L)$—a moving average polynomial in L;
τ_{iht}—the (transitory) stochastic component of income;
ϵ_{iht}—a mean-zero shock to the transitory component;

$u_{iht,me}$—a mean-zero measurement error+purely transitory shock.

RIP: Restricted Income Profiles

$$y_{iht} = \underbrace{\alpha_i}_{\text{heterogeneity}} + \underbrace{p_{iht} + \tau_{iht}}_{\text{risk}} + \underbrace{u_{iht,me}}_{\text{meas. error}}$$

$$p_{iht} = p_{iht-1} + \xi_{iht}$$

$$\tau_{iht} = \theta(L)\epsilon_{iht}$$

h—labor market experience;

p_{iht}—the permanent stochastic component of income;

ξ_{iht}—a mean-zero shock to the permanent component;

$u_{iht,me}$—a mean-zero measurement error+purely transitory shock;

Encompassing model

$$y_{iht} = \alpha_i + \beta_i h + p_{iht} + \tau_{iht} + u_{iht,me}$$

- **HIP**: $p_{iht} = 0$, all t.

  Baker (1997), Guvenen (2008), Haider (2001), Hause (1980), Lillard and Weiss (1979).

- **RIP**: $\beta_i = 0$.

  Abowd and Card (1989), Carroll and Samwick (1997), MaCurdy (1982), Meghir and Pistaferri (2004), Moffitt and Gottschalk (1995).

Findings in the RIP/HIP studies

- **HIP** studies: find a moderate persistence of the stochastic component and substantial and significant growth-rate heterogeneity.
- **RIP** studies with a permanent random walk component: find a significant variance of permanent shocks and a strong mean-reverting component in earnings.
- Why does it matter?

Modeling consumption dynamics

Consider the PIH: infinite horizon, quadratic utility, saving and borrowing at the risk-free rate r.

- Under RIP with a permanent random walk component and an MA(1) transitory component:

$$y_{it} = p_{it} + \tau_{it}$$
$$p_{it} = p_{it-1} + \xi_{it}$$
$$\tau_{it} = \epsilon_{it} + 0.30\epsilon_{it-1}$$

$$\Delta c_{it} = \alpha_P \xi_{it} + \alpha_T \epsilon_{it} = \xi_{it} + \frac{r(1+r+\theta)}{(1+r)^2}\epsilon_{it}.$$

$r = 0.02$, $\theta = 0.30$, $\underline{\alpha_T = 0.025}$, $\underline{\alpha_P = 1}$.

$$var^i(c_{it}) = var^i(c_{it-1}) + \sigma^2_{\xi_t} + 0.025^2\sigma^2_{\epsilon_t} \approx var^i(c_{it-1}) + \sigma^2_{\xi_t}.$$

- Under HIP with an AR(1) component and individual's perfect knowledge of β_i:

$$y_{it} = \beta_i t + \tau_{it}$$
$$\tau_{it} = 0.80\tau_{it-1} + \epsilon_{it}$$

$$\Delta c_{it} = \alpha_T \epsilon_{it} = \frac{r}{1+r-\phi}\epsilon_{it}.$$

$r = 0.02$, $\phi = 0.80$, $\underline{\alpha_T = 0.09}$.

$$var^i(c_{it}) = var^i(c_{it-1}) + 0.09^2\sigma^2_{\epsilon_t} \approx var^i(c_{it-1}).$$

Is it possible to identify a model that encompasses the important features of HIP and RIP using *just* income data?

Main findings

It is possible to identify the growth-rate heterogeneity, the variance of permanent shocks, the persistence of the mean-reverting component, and the variance of shocks to it.

Using data on income growth rates from the Panel Study of Income Dynamics (PSID), HIP model can be rejected. RIP model with a permanent random walk and a transitory component cannot be rejected. That is, the data favors RIP.

THE UNIVERSITY OF CHICAGO

Rest of Paper

- A Monte Carlo Study exploring identification of income processes in unbalanced panels.
- Identification arguments.
- Estimations using household heads' labor income data from the PSID.

Link to Appendix

Identification

$$E[\Delta y_{it}\Delta y_{it+k}] = \sigma_\beta^2 \mathbf{1}, \quad k = 3, \ldots, T-t, \;\; t = 1, \ldots, T-k,$$

where $\mathbf{1}$ is a vector of ones of the row dimension $(T-3)(T-2)/2$.

$$
\begin{aligned}
\sigma_\xi^2 =& E(\Delta y_{it}\Delta y_{it}) \\
&+ E(\Delta y_{it}\Delta y_{it+1}) + E(\Delta y_{it}\Delta y_{it-1}) \\
&+ E(\Delta y_{it}\Delta y_{it+2}) + E(\Delta y_{it}\Delta y_{it-2}) \\
&- 5\sigma_\beta^2
\end{aligned}
$$

With an MA(1) transitory component, it is possible to identify two out of the other three parameters: σ_ϵ^2, $\sigma_{u,me}^2$, θ.

Similar identification arguments apply to models with an AR(1)/ARMA(1,1) persistent components. If ARMA(1,1), the variance of meas. error is not separately identified.

True=estimated models:

$$y_{iht} = \alpha_i + \beta_i h + p_{iht} + \tau_{iht} + u_{iht,me}$$

$$\alpha_i = \sqrt{0.03} * iidN(0,1)$$
$$\beta_i = \sqrt{0.0004} * iidN(0,1)$$
$$p_{iht} = p_{ih-1t-1} + \sqrt{0.02} * iidN(0,1)$$

$$\tau_{iht} = 0.50\tau_{ih-1t-1} + \epsilon_{iht} - 0.20\epsilon_{ih-1t-1} \quad \textbf{if ARMA(1,1)}$$
$$\tau_{iht} = 0.50\tau_{ih-1t-1} + \epsilon_{iht} \quad \textbf{if AR(1)}$$
$$\tau_{iht} = \epsilon_{iht} + 0.50\epsilon_{ih-1t-1} \quad \textbf{if MA(1)}$$

$$\epsilon_{iht} = \sqrt{0.04} * iidN(0,1)$$
$$u_{iht,me} = \sqrt{0.02} * iidN(0,1).$$

Estimates of HIP with R.W. Simulated Data

Parameters/Trans. comp.	ARMA(1,1)	AR(1)	MA(1)
Heterog. growth, $\hat{\sigma}_{\beta}^2$	0.0004 (0.0001)	0.0004 (0.0001)	0.0004 (0.00008)
Var. perm. shock, $\hat{\sigma}_{\xi}^2$	0.02 (0.002)	0.02 (0.002)	0.02 (0.001)
AR, $\hat{\rho}$	0.494 (0.097)	0.496 (0.05)	— —
MA, $\hat{\theta}$	–0.287 (0.03)	— —	0.50 (0.01)
$\hat{\sigma}_{\epsilon}^2$	0.061 (0.004)	0.04 (0.002)	0.04 (0.001)
$\sigma_{u,me}^2$	0.00 —	0.02 (0.002)	0.02 —
Median χ^2[d.f.] Rejection rate at 1%	566.97 [430] 91%	554.70 [430] 95%	558.54 [431] 96%

Data in first differences. Main finding

If idiosyncratic incomes contain both the growth-rate heterogeneity, the random walk and transitory components, these components should be precisely recovered from estimations utilizing data on idiosyncratic labor income growth.

MaCurdy's test

Guvenen (2008): tests of the HIP (e.g., MaCurdy 1982) rely on the significance of higher-order autocovariances of income data in first differences. Even in the presence of HIP, these higher-order autocovariances are not significantly different from zero. The test lacks power against growth-rate heterogeneity alternative.

True for a model with HIP and random walk. But identification depends on the *size* of higher-order autocovariances, and there is *additional information* about HIP in the autocovariance matrix besides that contained in higher-order autocovariances.

Autocovariances for income processes with growth-rate heterogeneity and a random walk component

Order	σ_β^2=0.0004, σ_ξ^2=0.02 τ_{iht} ~MA(1), $\theta = 0.50$	σ_β^2=0.0004, σ_ξ^2=0.02 τ_{iht} ~AR(1), $\phi = 0.50$
0	0.12014	0.11361
	(0.00077)	(0.00078)
1	–0.02962	–0.03302
	(0.00051)	(0.00056)
2	–0.01956	–0.00629
	(0.00061)	(0.00056)
3	0.00039	–0.00295
	(0.00063)	(0.00056)
4	0.00039	–0.00126
	(0.00065)	(0.00058)
5	0.00038	–0.00046
	(0.00064)	(0.00061)
10	0.00039	0.00038
	(0.00082)	(0.00077)
15	0.00043	0.00047
	(0.00111)	(0.00102)
20	0.00035	0.00038
	(0.0017)	(0.00163)

Identification. Misspecified HIP

If $\sigma^2_\beta = 0$ and τ_{iht} is an MA(1)/AR(1)/ARMA(1,1), the variance of the permanent shock to income can be identified from the following moment condition:

$$\lim_{T\to\infty} E\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Delta y_{it}\right]^2 = \sigma^2_\xi .$$

Cochrane (1988) uses this moment to identify the size of the random walk in U.S. GNP, $\sigma^2_\xi/\sigma^2_{\Delta y_t}$.

For data with a finite T and an MA(1) transitory component:

$$E\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Delta y_{it}\right]^2 = \sigma_\xi^2 + \frac{2}{T}\left[\sigma_\epsilon^2(1+\theta^2)+\sigma_{u,me}^2\right].$$

If, instead, the HIP is estimated using data of length T (i.e., the random walk component is ignored):

$$E\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Delta y_{it}\right]^2 = T\hat{\sigma}_\beta^2 + \frac{2}{T}[\hat{\sigma}_\epsilon^2(1+\hat{\theta}^2)+\hat{\sigma}_{u,me}^2].$$

Misspecified HIP

Thus, if the random walk is ignored and the HIP is estimated instead:

$$\hat{\sigma}_{\beta}^{2} \approx \frac{1}{T}\sigma_{\xi}^{2}.$$

As an example, if $\sigma_{\beta}^{2} = 0.00$, $\sigma_{\xi}^{2} = 0.02$, and $T = 30$ ($T = 29$ for income growth rates), $\hat{\sigma}_{\beta}^{2} \approx 0.0007$.

True models (RW+trans. component):

$$y_{iht} = \alpha_i + p_{iht} + \tau_{iht} + u_{iht,me}$$

$$\alpha_i = \sqrt{0.03} * iidN(0,1)$$
$$p_{iht} = p_{ih-1t-1} + \sqrt{0.02} * iidN(0,1)$$
$$\tau_{iht} = 0.50\tau_{ih-1t-1} + \epsilon_{iht} - 0.20\epsilon_{ih-1t-1} \quad \textbf{if ARMA(1,1)}$$
$$\tau_{iht} = 0.50\tau_{ih-1t-1} + \epsilon_{iht} \quad \textbf{if AR(1)}$$
$$\tau_{iht} = \epsilon_{iht} + 0.50\epsilon_{ih-1t-1} \quad \textbf{if MA(1)}$$
$$\epsilon_{iht} = \sqrt{0.04} * iidN(0,1)$$
$$u_{iht,me} = \sqrt{0.02} * iidN(0,1).$$

Estimated misspecified models (HIP+trans. component):

$$y_{iht} = \alpha_i + \beta_i h + \tau_{iht} + u_{iht,me}$$

Misspecified HIP. Simulated Data

Parameters/Trans. comp.	ARMA(1,1)	AR(1)	MA(1)
Heterog. growth, $\hat{\sigma}^2_\beta$	**0.00053** **(0.00006)**	**0.00056** **(0.00006)**	**0.0007** **(0.00006)**
AR, $\hat{\rho}$	**0.776** **(0.017)**	**0.68** **(0.015)**	— —
MA, $\hat{\theta}$	**–0.34** **(0.014)**	— —	**0.474** **(0.008)**
$\hat{\sigma}^2_\epsilon$	**0.09** **(0.0008)**	**0.054** **(0.001)**	**0.052** **(0.0005)**
$\sigma^2_{u,me}$	**0.00** —	**0.024** **(0.001)**	**0.0173** —
Median χ^2[d.f.]	**627.79 [431]**	**656.98 [431]**	**1597.38 [432]**

Data in first differences. Main finding.

If the stochastic component of idiosyncratic earnings consists of a random walk and a mean-reverting component, and there is no growth-rate heterogeneity and an econometrician estimates the HIP, the estimated persistence can be modest and the variance of the deterministic growth-rate heterogeneity can be substantial and significant—as is found in the HIP studies.

True models (HIP+trans. component):

$$y_{iht} = \alpha_i + \beta_i h + \tau_{iht} + u_{iht,me}$$

$$\alpha_i = \sqrt{0.03} * iidN(0,1)$$
$$\beta_i = \sqrt{0.0004} * iidN(0,1)$$
$$\tau_{iht} = 0.50\tau_{ih-1t-1} + \epsilon_{iht} - 0.20\epsilon_{ih-1t-1} \quad \textbf{if ARMA(1,1)}$$
$$\tau_{iht} = 0.50\tau_{ih-1t-1} + \epsilon_{iht} \quad \textbf{if AR(1)}$$
$$\tau_{iht} = \epsilon_{iht} + 0.50\epsilon_{ih-1t-1} \quad \textbf{if MA(1)}$$
$$\epsilon_{iht} = \sqrt{0.04} * iidN(0,1)$$
$$u_{iht,me} = \sqrt{0.02} * iidN(0,1).$$

Estimated misspecified models (R.W.+HIP+trans. component):

$$y_{iht} = \alpha_i + \beta_i h + p_{iht} + \tau_{iht} + u_{iht,me}$$

Misspecified RIP. Simulated Data

Parameters/Trans. comp.	ARMA(1,1)	AR(1)	MA(1)
Heterog. growth, $\hat{\sigma}^2_\beta$	0.0004	0.00038	0.00038
	(0.00007)	(0.00007)	(0.00005)
Var. perm. shock, $\hat{\sigma}^2_\xi$	0.0007	0.00046	0.0002
	(0.0015)	(0.0015)	(0.0009)
AR, $\hat{\rho}$	0.464	0.487	—
	(0.084)	(0.041)	—
MA, $\hat{\theta}$	–0.270	—	0.502
	(0.059)	—	(0.009)
$\hat{\sigma}^2_\epsilon$	0.06	0.04	0.052
	(0.002)	(0.002)	(0.0005)
$\sigma^2_{u,me}$	0.00	0.02	0.02
	—	(0.002)	—
Median χ^2[d.f.]	600.34 [430]	654.19 [430]	573.51 [431]

Empirical data

- Income and demographic data from the 1968–1997 waves of the PSID.
- Select male household heads of age 25–64.
- The measure of income: head's labor income from all sources, inclusive of the labor part of farm and business income.
- No Latino, SEO, and Immigrant Samples; drop those with a spell of self-employment; drop income outliers.
- The measure of idiosyncratic labor income growth for each year: the residual from a cross-sectional regression of head's income growth on a third-order polynomial in age, education dummies, and interactions between education dummies and the age polynomial.
- The main sample contains data for 1,916 heads with at least 9 consecutive observations on labor income (29,753 person-year observations).

Estimates of income processes. PSID data in first differences

	(1) HIP	(2) add RW	(3) est. pers.	(4) chang. perm./ trans. var.	(5) use only 1st 10 acfs
$\hat{\sigma}^2_\beta$	0.0004	0.00	0.00	0.00	0.00
	(0.00004)	(0.00006)	(0.001)	—	(0.0002)
$\hat{\sigma}^2_\xi$	0.00	0.015	0.016	0.017	0.015
	—	(0.002)	(0.002)	(0.005)	(0.003)
$\hat{\rho}$	0.712	0.367	0.343	0.357	0.369
	(0.029)	(0.115)	(0.194)	(0.114)	(0.138)
$\hat{\theta}$	–0.187	–0.091	–0.081	–0.105	–0.092
	(0.024)	(0.08)	(0.113)	(0.086)	(0.088)
$\hat{\sigma}^2_\epsilon$	0.046	0.028	0.027	0.027	0.028
	(0.001)	(0.002)	(0.005)	(0.005)	(0.003)
$\hat{\rho}_{RW}$	0.0	1.0	0.992	1.0	1.0
	—	—	(0.158)	—	—

PSID data. Sample split by education.

	High school grad. or less		Some college or more	
	(1)	(2)	(3)	(4)
$\hat{\sigma}^2_\beta$	0.0003	0.00	0.0004	0.00
	(0.00006)	(0.00008)	(0.00007)	(0.0001)
$\hat{\sigma}^2_\xi$	0.00	0.012	0.00	0.02
	—	(0.002)	—	(0.003)
$\hat{\rho}$	0.588	0.335	0.848	0.385
	(0.050)	(0.141)	(0.029)	(0.209)
$\hat{\theta}$	–0.165	–0.073	–0.179	–0.084
	(0.041)	(0.099)	(0.025)	(0.143)
$\hat{\sigma}^2_\epsilon$	0.048	0.035	0.044	0.020
	(0.002)	(0.003)	(0.002)	(0.003)

Figure 1: The Variance of Log Labor Income by Year

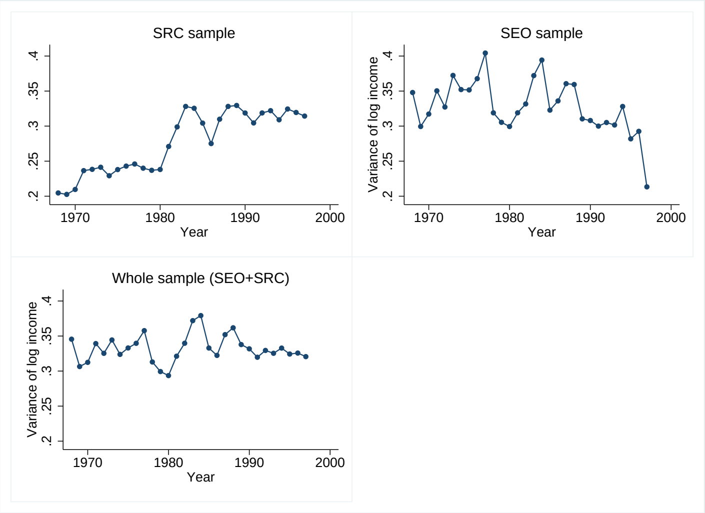
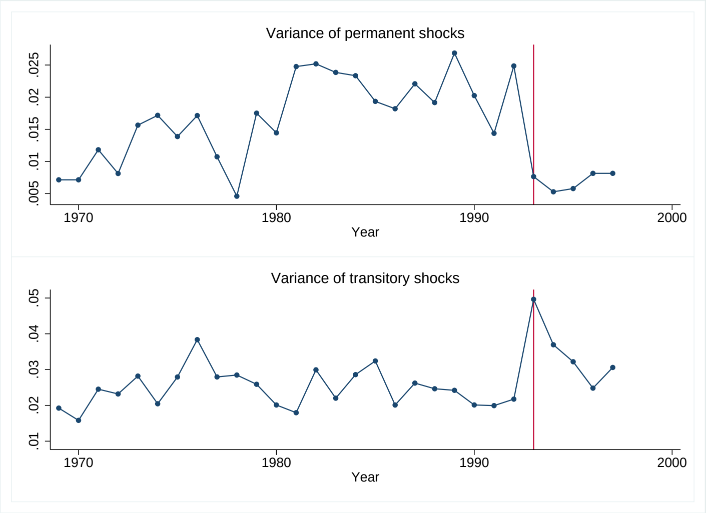


Figure 2: The Variance of Shocks to Labor Income by Year



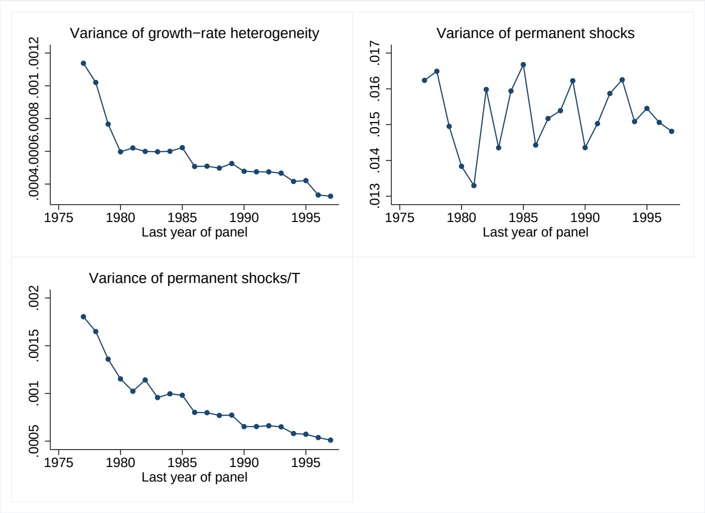
The Variance of Growth-Rate Heterogeneity and the Time Dimension of the Sample Size

- For the same individuals, the estimated variance of growth-rate heterogeneity should not depend on the time dimension of the sample size.

- Previous arguments: if the true model contains a random walk component and no deterministic growth-rate heterogeneity, but the model is estimated as (misspecified) HIP, $\hat{\sigma}^2_\beta$ should be smaller for larger T.

- First take PSID data for heads with at least 5 consec. income obs. during 1968–1977, estimate σ^2_β; add one more year of inc. observations, estimate σ^2_β, etc. until the time span is 1968–1997. Plot $\hat{\sigma}^2_\beta$. Same individuals, but clearly lower $\hat{\sigma}^2_\beta$ for larger T [the leftmost graph in Figure 3].

Figure 3: The Variance of Growth-rate Heterogeneity and T



Conclusion

- In the PSID, I find that the estimated variance of the deterministic growth-rate heterogeneity is zero, i.e., the HIP model can be rejected.
- The RIP model, with a permanent random walk and mean-reverting components, cannot be rejected. The estimated variance of the (stochastic) permanent component is significant and substantial.

THE UNIVERSITY OF CHICAGO

Conclusion

- Implications for structural modeling of wage/labor income dynamics. Some shocks do affect productivity of individuals permanently (e.g., disability).
- Internal propagation mechanism of non-persistent (iid) shocks? (Postel-Vinay and Thuron 2009).
- What are the reasons behind the time series pattern of the variances of transitory and permanent shocks?

Appendix

Monte Carlo simulations

$$y_{iht} = \alpha_i + \beta_i h + p_{iht} + \tau_{iht} + u_{iht,me}$$

- Heterogeneity:

  $(\alpha_i, \beta_i) \sim iidN(0, \Omega)$, $\Omega_{11} = \sigma_\alpha^2$, $\Omega_{22} = \sigma_\beta^2$, $\Omega_{12} = \Omega_{21} = \sigma_{\alpha\beta}$.

- Uncertainty:

  Perm. shock—$\xi_{iht} \sim iidN(0, \sigma_\xi^2)$.

  Trans. shock—$\epsilon_{iht} \sim iidN(0, \sigma_\epsilon^2)$.

  τ_{iht} is AR(1)/MA(1)/ARMA(1,1).

- Measurement Error:

  $u_{iht,me} \sim iidN(0, \sigma_{u,me}^2)$.

THE UNIVERSITY OF CHICAGO

Simulation details

- Simulate individual incomes "observed" for at most 30 periods.
- In the first year: a cross section of households whose heads' experience ranges from 1 to 30 years, 70 of each type. Year 1: Heads with 1 year of experience→30 obs. towards the final sample, heads with 30 years of experience→1 observation only.
- Keep only those who contribute at least 9 consecutive observations towards the final sample (this selection criterion will be followed in the empirical part and is similar to Meghir and Pistaferri 2004).

Simulation details, contd.

- For each estimated income model, I report the results based on 100 simulated samples.

- The models identified by fitting the theoretical autocovariances, $\Gamma(\Theta)$, to the autocovariances in the simulated data, $\hat{\Gamma}^s_T$. Estimation by the minimum distance method, with the identity weighting matrix (EWMD).

- Some elements of $\hat{\Gamma}^s_T$ for data in first differences: $E[\Delta y_{i2}\Delta y_{i2}]$, $E[\Delta y_{i2}\Delta y_{i3}]$, ..., $E[\Delta y_{i2}\Delta y_{iT}]$.

Monte Carlo results. Data in first differences

$$\Delta y_{it} = \beta_i + \xi_{it} + \theta(L)\Delta\epsilon_{it} + \Delta u_{it,me}$$

If τ_{iht} is MA(1), i.e., $\theta(L) = 1 + \theta L$, the auto-covariance moments are:

$$E[\Delta y_{it}\Delta y_{it}] = \gamma_0 = \sigma^2_\xi + \sigma^2_\beta + (1 + (1-\theta)^2 + \theta^2)\sigma^2_\epsilon + 2\sigma^2_{u,me}$$

$$E[\Delta y_{it}\Delta y_{it+1}] = \gamma_1 = \sigma^2_\beta \underbrace{-(\theta - 1)^2\sigma^2_\epsilon - \sigma^2_{u,me}}_{\texttt{mean reversion}}$$

$$E[\Delta y_{it}\Delta y_{it+2}] = \gamma_2 = \sigma^2_\beta \underbrace{-\theta\sigma^2_\epsilon}_{\texttt{mean reversion}}$$

$$E[\Delta y_{it}\Delta y_{it+k}] = \gamma_k = \sigma^2_\beta, \quad k \geq 3.$$