Hryshko & Manovskii: Greatest Hits and All-Time Favorites (& Some That Will Be)

Iourii Manovskii

University of Pennsylvania

Guest Lecture for
Mariacristina's Course at the University of Minnesota

14 October 2020

Plan I

- There is an extensive literature estimating stochastic processes for individual earnings.
    - Recently, using large administrative data
- Typically estimated on the data and then used as a source of risk in household budgets in the model.
- Large controversy about the nature and the size of risk.
    - Striking discrepancy between the estimates based on targeting the data moments in levels and differences.
- "Improving the Measurement of Earnings Dynamics" uncovers the source of the discrepancy and shows how to correct for it. Easy-peasy.

Plan II

- Blundell, Pistaferri and Preston (2008) found that consumption of U.S. families is much better insured against fluctuations in income than predicted by incomplete markets models and found a very large insurance role of the tax and transfer system.

- We will skim through "Income Dynamics and Consumption Insurance" and see that getting income process right is crucial for:

  (a) estimating of consumption insurance against permanent shocks to family earnings and to net family incomes.

  (b) estimating of the role of the tax and transfer system in mitigating the impact of shocks to family earnings on consumption.

Plan III

- Most of what is known to economists about joint income and consumption dynamics is based on data from the Panel Study of Income Dynamics (PSID).
  - For example, Blundell, Pistaferri and Preston (2008) is based on these data.

- In "How Much Consumption Insurance in the U.S.?", we
  1. identify a very puzzling feature of PSID data:
     - two subsets of PSID families differ dramatically in family income and consumption dynamics
     - this is surprising given the PSID design;
  2. explain the sources of the difference;
  3. reassess excess insurance puzzle.

Improving the Measurement of Earnings Dynamics

Moira Daly, Copenhagen Business School

Dmytro Hryshko, University of Alberta

Iourii Manovskii, University of Pennsylvania

Motivation

- There is an extensive literature estimating stochastic processes for individual earnings.
  - Recently, using large administrative data
  - Typically unbalanced panels (select workers with strong labor-market attachment, no outliers, etc.)
- Crucial ingredient of many quantitative labor and macro models with incomplete insurance markets.
- Typically estimated on the data and then used as a source of risk in household budgets in the model.
- Despite its importance, there's still controversy about the nature and the size of risk.
  - Large administrative data don't settle the controversy.

Permanent-transitory decomposition

A typical earnings process:

$$
\begin{aligned}
y_{it} &= \alpha_i + p_{it} + \tau_{it}, & \alpha_i &\sim \text{iid}(0, \sigma_\alpha^2) \\
p_{it} &= \phi_p p_{it-1} + \xi_{it}, & \xi_{it} &\sim \text{iid}(0, \sigma_\xi^2) \\
\tau_{it} &= \theta(L)\epsilon_{it}, & \epsilon_{it} &\sim \text{iid}(0, \sigma_\epsilon^2)
\end{aligned}
$$

y_{it} is individual i's log-earnings (residuals) at time t;

p_{it} is the permanent component (random walk if $\phi_p = 1$);

τ_{it} is the transitory component: MA(1), ARMA(1,1), AR(1), or iid;

α_i is an individual fixed effect.

$\sigma_\xi^2 =?$ $\sigma_\epsilon^2 =?$ No consensus!

Estimated inc. processes in levels vs. diffs.: a puzzle

- (Minimum distance) estimation typically targets either:

  Moments in levels : $E[y_{it} y_{it \pm j}], \quad j = 0, 1, 2, \ldots$

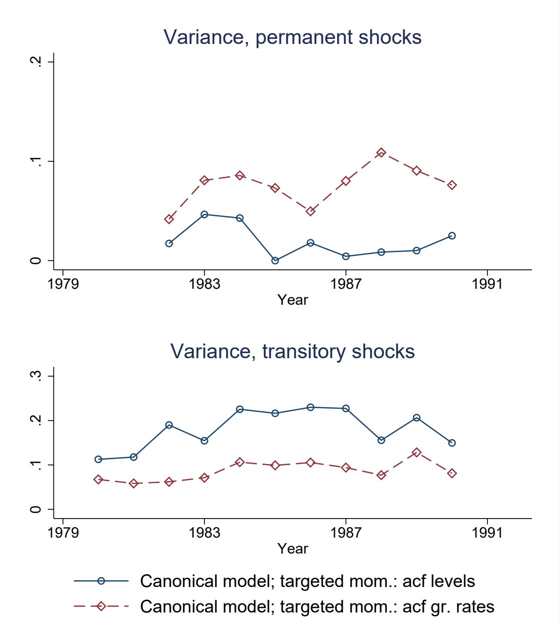
  Moments in differences : $E[\Delta y_{it} \Delta y_{it \pm j}], j = 0, 1, 2, \ldots$

- Commonplace finding (data for various countries, various datasets):

  Permanent shock variance σ_ξ^2 (levels) $\ll$ σ_ξ^2 (differences)

  Transitory shock variance σ_ϵ^2 (levels) $\gg$ σ_ϵ^2 (differences)

Estimates of the variances in the PSID

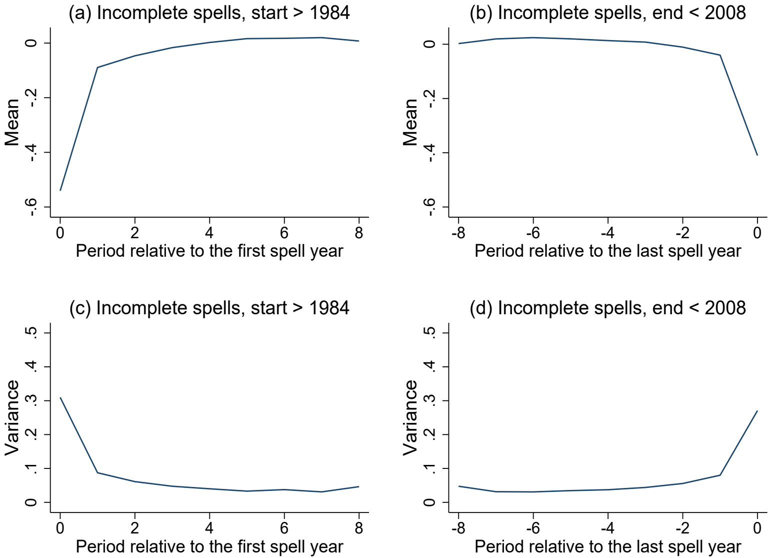


Misspecified income process

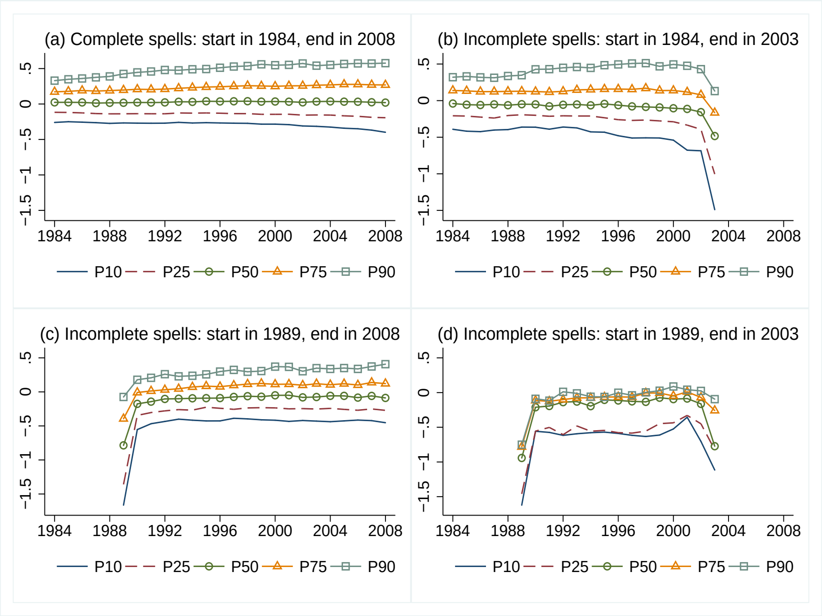
- Heathcote, Storesletten, Violante (2010):
  Income process is misspecified.

- How?

Properties of incomplete spells. German data: 1984–2008



Various earnings percentiles. German data



Why??? Elementary, Watson.

- Incomplete spells are by definition preceded or followed by a missing observation.
- Many reasons for missing obs. in administrative data: Self employment, work for the government, studying, unemployment, undocumented work, time spent abroad, extended maternity/paternity leave, death, early retirement (e.g., due to disability or wealth windfall), removed earnings outliers or observations with very low earnings/hours, etc.
- Individuals do not generally time these events to begin on Jan. 1 and end on Dec. 31.
- On average such spells are expected to begin or end in a middle of the year, leading to significantly lower earnings. But also to high variance as some individuals begin spells in e.g., March while others in e.g., November.

Identification. Random walk & iid trans. component

Differences:

$$\sigma^2_{\xi,t} = E[\Delta y_{it}\Delta y_{it-1}] + E[\Delta y_{it}\Delta y_{it}] + E[\Delta y_{it}\Delta y_{it+1}] \quad \text{(D1)}$$

$$\sigma^2_{\epsilon,t} = -E[\Delta y_{it}\Delta y_{it+1}] \quad \text{(D2)}$$

Levels:

$$\sigma^2_{\xi,t} = E[y_{it}y_{it+1}] - E[y_{it-1}y_{it+1}] - E[y_{it}y_{it-2}] + E[y_{it-1}y_{it-2}] \quad \text{(L1)}$$

$$\sigma^2_{\epsilon,t} = E[y_{it}y_{it}] - E[y_{it}y_{it+1}] - E[y_{it}y_{it-1}] + E[y_{it-1}y_{it+1}] \quad \text{(L2)}$$

- (L1) is an expansion of (D1), and (L2) is an expansion of (D2).
- The moments should deliver the same estimates in a sample of individuals whose earnings are nonmissing for the periods $t-2$ through $t+1$.

Identification in unbalanced panels I

$$y_{it} = \alpha_i + p_{it} + \epsilon_{it} + \nu_{it}, \quad \nu_{it} \text{ rare shock} \sim \text{iid}(\mu_\nu, \sigma_\nu^2)$$

Ex. $\nu_{it} \neq 0$ if $t = t_0^i \neq t_0$ (t_0 first sample year)

ν_{it} uncorr. with α_i, ξ_{is}, ϵ_{is} $\forall$ t, s.

Levels:

$$\sigma^2_{\epsilon,t} = \underbrace{E[y_{it}y_{it}]}_{\substack{\text{extra var.:} \\ \mu_\nu^2 + \sigma_\nu^2}} - E[y_{it}y_{it+1}] - E[y_{it}y_{it-1}] + E[y_{it-1}y_{it+1}]$$

Differences:

$$\sigma^2_{\xi,t+1} = E[\Delta y_{it+1}\Delta y_{it}] + \underbrace{E[\Delta y_{it+1}\Delta y_{it+1}]}_{\substack{\text{extra var.:} \\ \mu_\nu^2 + \sigma_\nu^2}} + E[\Delta y_{it+1}\Delta y_{it+2}]$$

Identification in unbalanced panels II

$$y_{it} = \alpha_i + p_{it} + \epsilon_{it} + \nu_{it}, \quad \nu_{it} \text{ rare shock} \sim \text{iid}(\mu_\nu, \sigma^2_\nu)$$

Ex. $\nu_{it} \neq 0$ if $t = t_0^i \neq t_0$ (t_0 first sample year)

ν_{it} uncorr. with α_i, ξ_{is}, ϵ_{is} $\forall$ t, s.

Permanent shock:

- Levels: no biases
- Differences: $\hat{\sigma}^2_{\xi,t+1} - \sigma^2_{\xi,t+1} = s_{t,t+1}(\mu^2_\nu + \sigma^2_\nu) > 0$

$s_{t,t+1}$ is the share of individuals who start (incomplete) earnings spells at time t, with nonmissing earnings at times t and $t+1$, in the total number of individuals observed at t and $t+1$.

Transitory shock:

- Levels: $\hat{\sigma}^2_{\epsilon,t} - \sigma^2_{\epsilon,t} = s_t(\mu^2_\nu + \sigma^2_\nu) > 0$
- Differences: no biases

Administrative data

- ⋆ Administrative data on annual earnings from the 1981–2006 tax registers for more than 99.9% of Danish residents between the ages of 15 and 70.
  - Earnings include all earned labor income, taken from the tax records.
- ⋆ Administrative data from the IABS, a 2% random sample of German social security records for the years 1974–2008.
  - Since 1984, wages+bonuses recorded; use the data for 1984–2008.

Data. Sample selection

Danish data

- Males born in 1951–1955, no immigrants.
- Never self-employed during the period 1981–2006; finished school.
- Drop annual records for those who have worked less than 10% of the year as a full-time employee, or whose income is nonpositive.

German data

- Males born in 1951–1955, from West Germany, not in apprenticeship.
- Daily wages are right-censored at the highest level subject to SS contributions; impute daily wages in the upper tail using a Pareto distribution.
- Drop records when the combined duration of job spells within a year is below 35 calendar days.

Samples used

We use 3 different samples for each dataset:

1. 9 or more consecutive earnings observations (e.g., Meghir and Pistaferri 2004; Browning, Ejrnæs, and Alvarez 2010). Danish data: 102,825 individuals. German data: 22,791 individuals.

2. 20 or more, not necessarily consecutive, earnings observations (e.g., Guvenen 2009). Danish data: 90,668 individuals. German data: 17,621 individuals.

3. Balanced sample (all 26 observations). Danish data: 67,008 individuals. German data: 12,274 individuals.

Estimated earnings process

$$y_{it} = \alpha_i + p_{it} + \tau_{it}, \;\; \alpha_i \sim \text{iid}(0, \sigma_\alpha^2)$$

The permanent component:

$$p_{it} = \phi_p p_{it-1} + \xi_{it}, \;\; \xi_{it} \sim \text{iid}(0, \sigma_\xi^2)$$

The transitory component:

$$\tau_{it} = \epsilon_{it} + \theta \epsilon_{it-1}, \;\; \epsilon_{it} \sim \text{iid}(0, \sigma_\epsilon^2)$$

Minimum-distance estimation using the optimal weighting matrix (large samples, clean data).

Unbalanced panels: 9 or more consec. obs.

	German data		Danish data	
	Levs. (1)	Diffs. (2)	Levs. (3)	Diffs. (4)
$\hat{\phi}_p$	0.976	0.992	0.955	0.987
$\hat{\sigma}_\xi^2$	0.008	0.019	0.008	0.013
$\hat{\theta}$	0.129	0.153	0.204	0.209
$\hat{\sigma}_\epsilon^2$	0.024	0.009	0.019	0.012
$\hat{\sigma}_\alpha^2$	0.024	—	0.020	—

All coefficients significant at the 1% level.

Unbalanced panels: 20 not nec. consec. obs.

	German data		Danish data	
	Levs. (1)	Diffs. (2)	Levs. (3)	Diffs. (4)
$\hat{\phi}_p$	0.999	0.991	0.964	0.982
$\hat{\sigma}_\xi^2$	0.0048	0.009	0.007	0.012
$\hat{\theta}$	0.119	0.192	0.137	0.217
$\hat{\sigma}_\epsilon^2$	0.016	0.009	0.022	0.013
$\hat{\sigma}_\alpha^2$	0.027	—	0.023	—

All coefficients significant at the 1% level.

Balanced samples

	German data		Danish data	
	Levs. (1)	Diffs. (2)	Levs. (3)	Diffs. (4)
$\hat{\phi}_p$	1	0.998	0.969	0.970
$\hat{\sigma}^2_\xi$	0.0031	0.0033	0.005	0.005
$\hat{\theta}$	0.278	0.258	0.212	0.209
$\hat{\sigma}^2_\epsilon$	0.008	0.0078	0.009	0.009
$\hat{\sigma}^2_\alpha$	0.024	—	0.017	—

All coefficients significant at the 1% level.

Residual earnings. Panel regressions. German data

	9 or more consec.	**20 not nec. consec.**
Year obs.: first, dummy	–0.57	–0.65
Year obs.: last, dummy	–0.43	–0.47
1 year before earn. miss., dummy		–0.27
1 year after earn. miss., dummy		–0.39
No. obs.	379,080	330,748
No. indiv.	18,130	13,635

- Residual earnings are lower than average in the (few) first and last periods of the earnings spell; and in the (few) periods before and after missing. Only dummies for the observations before and after missing records are included here (see paper for the full regression results).

Squared res. earnings. Panel regressions. German data

	9 or more consec.	20 not nec. consec.
Year observed: first, dummy	0.23	0.29
Year observed: last, dummy	0.20	0.25
1 year before earn. missing, dummy		0.15
1 year after earn. missing, dummy		0.23
No. obs.	379,080	330,748
No. indiv.	18,130	13,635

- Residual earnings are more volatile in the (few) first and last periods of the earnings spell; and in the periods before and after a year of missing. Only dummies for the observations before and after missing records are included here (see paper for the full regression results).

Earnings, daily wages, and days worked residuals. German data

	9 or more consec.			20 not nec. consec.		
	Earn. (1)	Days (2)	Wages (3)	Earn. (4)	Days (5)	Wages (6)
Year obs.: first	–0.57	–0.43	–0.14	–0.67	–0.49	–0.17
Year obs.: last	–0.43	–0.38	–0.05	–0.48	–0.38	–0.09
1 year before miss.				–0.27	–0.23	–0.04
1 year after miss.				–0.39	–0.27	–0.12
No. obs.	379,080	379,080	379,080	330,748	330,748	330,748
No. indiv.	18,130	18,130	18,130	13,635	13,635	13,635

All coefficients significant at the 1% level.

What to do?

1. Drop earnings observations around missing records
2. Model those observations (by adding an extra transitory component to the earnings process)

Model outliers: extended earnings process

$$y_{it} = \alpha_i + p_{it} + \tau_{it} + \chi_{it}, \quad t = t_0, \ldots, T$$

$$p_{it} = \phi_p p_{it-1} + \xi_{it}$$

$$\tau_{it} = \epsilon_{it} + \theta \epsilon_{it-1}$$

$$\chi_{it+j} = \begin{cases} \nu_{it} & \text{if } y_{it-k} \text{ or } y_{it+k} \text{ is miss. and } t-k \geq t_0,\, t+k \leq T, j = 0 \\ \theta \nu_{it} & j = 1 \\ 0 & \text{otherwise} \end{cases}$$

9 or more. German data

	Full sample		Drop first & last 3 obs.		Model outliers first & last obs.		Model outliers first & last 3 obs.	
	Levs. (1)	Diffs. (2)	Levs. (3)	Diffs. (4)	Levs. (5)	Diffs. (6)	Levs. (7)	Diffs. (8)
$\hat{\phi}_p$	0.976	0.992	0.982	0.994	0.981	0.996	0.982	0.999
$\hat{\sigma}_\xi^2$	0.0078	0.019	0.006	0.005	0.007	0.005	0.006	0.004
$\hat{\theta}$	0.129	0.153	0.197	0.186	0.135	0.145	0.168	0.203
$\hat{\sigma}_\epsilon^2$	0.024	0.009	0.010	0.009	0.01	0.01	0.01	0.01
$\hat{\sigma}_\alpha^2$	0.024	—	0.019	—	0.013	—	0.017	—

All coefficients significant at the 1% level.

20 or not nec. consec. German data

	Full sample		Drop first & last 3 obs.		Model outliers first & last obs.		Model outliers first & last 3 obs.	
	Levs. (1)	Diffs. (2)	Levs. (3)	Diffs. (4)	Levs. (5)	Diffs. (6)	Levs. (7)	Diffs. (8)
$\hat{\phi}_p$	0.999	0.991	0.992	0.995	0.999	0.995	0.999	0.996
$\hat{\sigma}^2_\xi$	0.0048	0.009	0.0047	0.0046	0.0045	0.0055	0.004	0.005
$\hat{\theta}$	0.119	0.192	0.204	0.190	0.194	0.171	0.214	0.208
$\hat{\sigma}^2_\epsilon$	0.016	0.009	0.009	0.008	0.009	0.009	0.009	0.008
$\hat{\sigma}^2_\alpha$	0.027	—	0.021	—	0.021	—	0.027	—

All coefficients significant at the 1% level.

Simulated "German" data

	9 consec.				20 not nec. consec.			
	Full sample		Drop		Full sample		Drop	
	Levs. (1)	Diffs. (2)	Levs. (3)	Diffs. (4)	Levs. (5)	Diffs. (6)	Levs. (7)	Diffs. (8)
$\hat{\phi}_p$	0.979	0.988	0.980	0.980	0.997	0.995	0.999	0.999
$\hat{\sigma}_\xi^2$	0.008	0.016	0.008	0.008	0.005	0.009	0.005	0.005
$\hat{\theta}$	0.133	0.143	0.170	0.170	0.152	0.189	0.20	0.20
$\hat{\sigma}_\epsilon^2$	0.018	0.009	0.01	0.01	0.014	0.01	0.01	0.01
$\hat{\sigma}_\alpha^2$	0.025	—	0.025	—	0.024	—	0.024	—

Implications for calibrating the earnings process

- Use the moments in levels to estimate the variance of permanent shocks and moments in differences to estimate the variance of transitory shocks.
- Estimate the earnings process on the data that do not include the observations surrounding the missing ones.
- Incorporate transitory shocks at the start and the end of contiguous earnings histories into the estimation.

  The mean and variance of these shocks can be identified from the mean and variance of earnings in those periods.

- What to do about missing observations themselves?
  (a) Typically treated either as missing at random or
  (b) missing incomes due to nonparticipation caused by an adverse shock

Conclusion

- A puzzle: the moments in levels and differences deliver different estimates of permanent and transitory risk.
- We find a source of difference: large deviations of earnings at the start and end of continuous individual earnings spells.
- Finally, we can estimate the process that can be used as an input into our models.
- But there is more. Next we will see that getting income process right is crucial for:
  (a) estimating of consumption insurance against permanent shocks to family earnings and to net family incomes.
  (b) estimating of the role of the tax and transfer system in mitigating the impact of shocks to family earnings on consumption.

Income Dynamics and Consumption Insurance

Dmytro Hryshko, University of Alberta

Iourii Manovskii, University of Pennsylvania

Introduction

- Empirical measures of the extent and sources of consumption insurance are used widely:
  - as the key benchmarks for assessing the performance of incomplete markets models (e.g., Kaplan and Violante 2010, Guvenen et al. 2016, De Nardi et al. 2020)
  - as inputs into models evaluating the optimality of the tax and transfer system (e.g., Heathcote et al. 2017, Wu and Krueger 2020)
- We provide such measurement building on Blundell, Pistaferri, Preston (2008, BPP) methodology
- For measuring the importance of taxes and transfers, it is superior to widely used alternatives in the presence of measurement error in incomes
- We provide a correction crucial for BPP-based methodology to reveal accurate measures of insurance in unbalanced panels

Findings

Benchmark BPP:

- ≈ 36% of permanent shocks to net family incomes do not pass through to consumption
- ≈ 63% of shocks to family earnings do not pass through to net family incomes

Our findings:

- Smaller role of assets in insuring permanent shocks to net family incomes

  ≈ 12% of permanent shocks to net family incomes do not pass through to consumption

- Smaller role of the tax and transfer system in insuring the shocks to family earnings

  ≈ 35% of shocks to family earnings do not pass through to net family incomes

BPP methodology

$$y_{it} = \alpha_i + p_{it} + \tau_{it}$$
$$p_{it} = p_{it-1} + \xi_{it},$$
$$\tau_{it} = \epsilon_{it} + \theta\epsilon_{it-1}$$

$$\Delta c_{it} = \phi\xi_{it} + \psi\epsilon_{it} + \zeta_{it} + \Delta u_{it}$$

- y_{it} is log net family income or family earnings
- ϕ and ψ measure transmission of permanent and transitory income shocks to consumption

  $1 - \phi$ and $1 - \psi$ are measures of insurance

- Estimate ϕ, ψ, σ^2_ϵ, σ^2_ξ using data from the PSID on (imputed) nondurable consumption and incomes for married couples, 1979–1993

The tax and transfer scheme

Heathcote et al. (2014), Blundell et al. (2016), etc.:

$$\text{Net fam. income}_{it} = \kappa \cdot (\text{Family Earnings}_{it})^{1-\gamma}$$

γ measures the share of shocks (perm. and trans.) to family earnings that do not pass through to net family incomes

Three approaches to measuring γ in the data

1. Running a regression (in levels/differences/using FE)

$$\log(\text{Net fam. inc.})_{it} = \text{const} + (1-\gamma)\cdot\log(\text{Fam. earnings})_{it} + \text{error}_{it}$$

2. Using estimated transmission coefficients:

$$\hat{\gamma}_{\phi} = 1 - \frac{\hat{\phi}_{[\text{using earn.}]}}{\hat{\phi}_{[\text{using net fam. inc.}]}} \qquad \hat{\gamma}_{\psi} = 1 - \frac{\hat{\psi}_{[\text{using earn.}]}}{\hat{\psi}_{[\text{using net fam. inc.}]}}$$

3. Using estimated variances of shocks:

$$\hat{\gamma}_{\sigma_\xi} = 1 - \frac{\hat{\sigma}_{\xi,[\text{using net fam. inc.}]}}{\hat{\sigma}_{\xi,[\text{using earn.}]}} \qquad \hat{\gamma}_{\sigma_\epsilon} = 1 - \frac{\hat{\sigma}_{\epsilon,[\text{using net fam. inc.}]}}{\hat{\sigma}_{\epsilon,[\text{using earn.}]}}$$

Biases in γ due to measurement error in incomes

- All measures of γ should recover truth if there's no meas. error in family earnings and net family incomes
- Measurement error in the family earnings data (HRS linked to administrative tax data)
  - the variance is large (and also goes down with age)
  - is not persistent
  - is negatively correlated with true (administrative) earnings
- Due to meas. error, the regression estimates of γ, and the estimates of γ_{σ_ϵ} and γ_ψ based on the var. of trans. shocks and the cons. insur. of trans. shocks might be biased

The effect of meas. error in model generated data

- Use data from a lifecycle model with incomplete insurance markets from Wu and Krueger (AEJM 2020)

Model true γ=0.146

Implied insurance due to taxes & transfers, $\hat{\gamma}$

	No meas. error	Meas. error
From regression in levels	0.146	0.196
From regression in levels, FE	0.146	0.270
From regression in diffs.	0.146	0.430
Using trans. cons. insur., $\hat{\gamma}_{\psi}$	0.146	0.490
Using var. trans. shocks, $\hat{\gamma}_{\sigma_{\epsilon}}$	0.146	0.330
Using perm. cons. insur., $\hat{\gamma}_{\phi}$	0.146	0.149
Using var. perm. shocks, $\hat{\gamma}_{\sigma_{\xi}}$	0.146	0.148

Empirical evaluation

- Married couples with male heads of ages 30–65 observed within the 1979–1993 period in the PSID (we follow BPP)
- Income measures: net family income and the combined earnings of the head and wife
- 2,430 families, unbalanced panel

Full-sample results. PSID

	Family earnings	Net family income	Implied role of tax & transf., $\hat{\gamma}$
ϕ, transm. of perm. shock ξ	0.23 (0.04)	0.63 (0.08)	63%
ψ, transm. of trans. shock ϵ	0.09 (0.03)	0.06 (0.04)	
σ_{ξ}^2, var. perm. shock (avg.)	0.06 (0.006)	0.02 (0.002)	38%
σ_{ϵ}^2, var. trans. shock (avg.)	0.07 (0.005)	0.03 (0.002)	
Age (avg.)	43	43	
Wealth (median)	50,200	50,200	
No. househ.	2,430	2,430	

Balanced-sample results. PSID

	Family earnings	Net family income	Implied role of tax & transf., $\hat{\gamma}$
ϕ, transm. of perm. shock ξ	0.68 (0.17)	0.83 (0.18)	17%
ψ, transm. of trans. shock ϵ	0.09 (0.03)	0.08 (0.06)	
σ_ξ^2, var. perm. shock (avg.)	0.02 (0.004)	0.01 (0.002)	27%
σ_ϵ^2, var. trans. shock (avg.)	0.06 (0.005)	0.04 (0.003)	
Age (avg.)	46	47	
Wealth (median)	60,222	61,626	
No. househ.	478	516	

Unbalanced panels

- The results from balanced and unbalanced panels are very different. Why?

Daly, Hryshko, Manovskii (2016):

(i) showed that earnings observations in ubalanced panels around missing values of incomplete spells are systematically different (lower and more volatile)
   - About 43% of families are married after 1978.
   - About 36% of marriages ended in widowhood/divorce before 1992.

(ii) failure to account for this feature biases up the estimated variance of permanent shocks using the BPP methodology

Canonical income process and consumption insurance

$$y_{it} = \alpha_i + p_{it} + \epsilon_{it}, \qquad p_{it} = p_{it-1} + \xi_{it}$$
$$\Delta c_{it} = \phi \xi_{it} + \psi \epsilon_{it} + \zeta_{it}$$

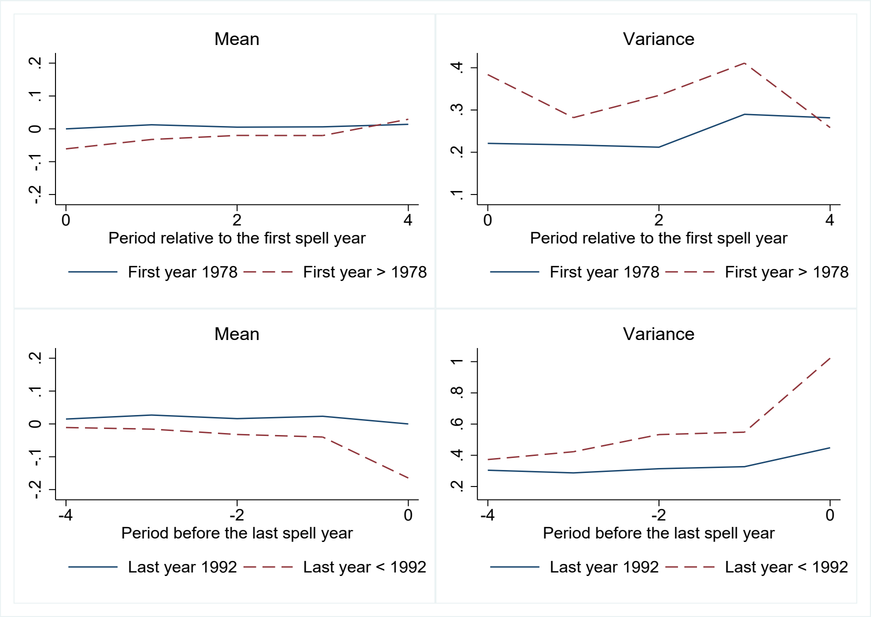
Variance permanent shock:

$$\sigma^2_{\xi,t} = E[\Delta y_{it} \Delta y_{it-1}] + E[\Delta y_{it} \Delta y_{it}] + E[\Delta y_{it} \Delta y_{it+1}]$$

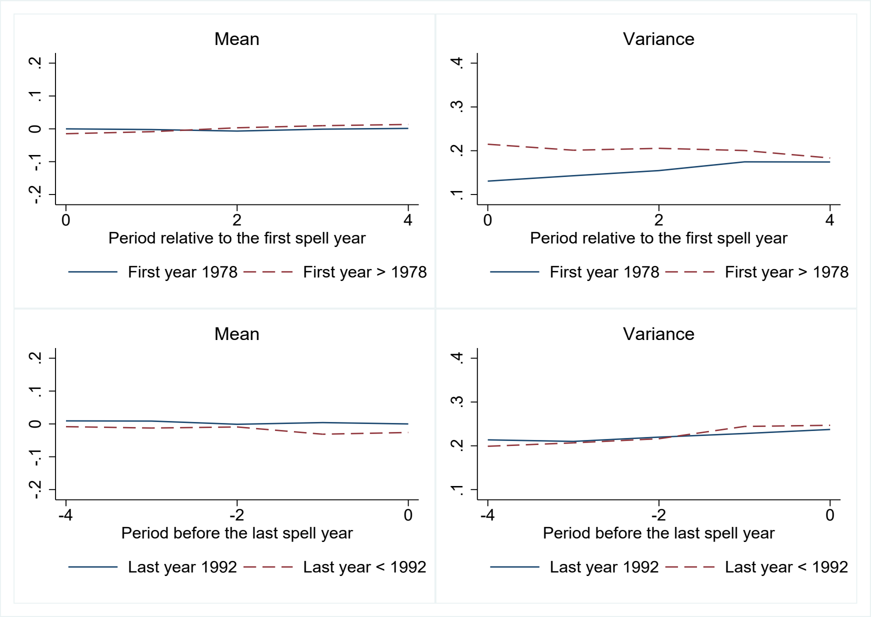
Permanent insurance:

$$1 - \phi_t = 1 - \frac{E[\Delta c_{it} \Delta y_{it-1}] + E[\Delta c_{it} \Delta y_{it}] + E[\Delta c_{it} \Delta y_{it+1}]}{E[\Delta y_{it} \Delta y_{it-1}] + E[\Delta y_{it} \Delta y_{it}] + E[\Delta y_{it} \Delta y_{it+1}]}$$

Family earnings spells, PSID data



Net income spells, PSID data



Net fam. income and earn. res. around miss. values

Income measure:	Panel A: Fam. earn.				Panel B: Net fam. inc.			
Dep. var.:	Means		Var.		Means		Var.	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
1 yr after	–0.15***	–0.03	0.29***	–0.06***	–0.02	–0.00	0.04**	–0.06***
2 yrs after	–0.09***	–0.02	0.11***	–0.09***	–0.00	–0.00	0.02	–0.05***
3 yrs after	–0.07***	–0.02	0.09**	–0.07***	0.01	–0.01	0.03	–0.04***
3 yrs before	–0.03	–0.01	0.24***	0.03	0.00	–0.00	0.05***	0.03***
2 yrs before	–0.03	–0.01	0.28***	0.07***	–0.02	0.00	0.08***	0.04***
1 yr before	–0.14***	–0.03	0.64***	0.18***	–0.02	–0.00	0.08***	0.05***
Const.	0.04***		0.25***		0.00		0.19***	
No. obs.	20,465		20,465		21,076		21,076	
No. indiv.	2,420		2,420		2,429		2,429	

***[**](*) significant at 1% [5%] (10%) level

Extended income process

$$y_{it} = \alpha_i + p_{it} + \tau_{it} + \chi_{it}, \quad t = t_0, \ldots, T$$

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

$$\tau_{it} = \epsilon_{it} + \theta\epsilon_{it-1}$$

$$\chi_{it+j} = \begin{cases} \nu_{it} & \text{if } y_{it-1} \text{ or } y_{it+1} \text{ is missing and } t-1 \geq t_0,\ t+1 \leq T, j = 0 \\ \theta\nu_{it} & j = 1 \\ 0 & \text{otherwise.} \end{cases}$$

$$\Delta c_{it} = \phi_t \xi_{it} + \psi_t \epsilon_{it} + \psi_t^{\nu} \nu_{it} + \zeta_{it}$$

Biases in the estimated consumption insurance

1. Consecutive unbalanced samples:
   - Start late, at $t \in (t_0, T)$:

$$(1-\hat{\phi}_{t+1}) - (1-\phi_{t+1}) = \underbrace{\frac{s_{t,t+1}(\mu_\nu^2+\sigma_\nu^2)}{s_{t,t+1}(\mu_\nu^2+\sigma_\nu^2)+\sigma_{\xi,t+1}^2}}_{=\lambda_{t+1}} \phi_{t+1} > 0$$

   - Exit early, at $t \in (t_0, T)$:

$$(1-\hat{\phi}_t) - (1-\phi_t) = (\phi_t - \psi_t^\nu)\lambda_t > 0$$

2. Nonconsecutive unbalanced samples (earn. missing at t):

$$(1-\hat{\phi}_{t-1}) - (1-\phi_{t-1}) = (\phi_{t-1} - \psi_{t-1}^\nu)\lambda_{t-1} > 0$$
$$(1-\hat{\phi}_{t+2}) - (1-\phi_{t+2}) = \phi_{t+2}\lambda_{t+2} > 0$$

Biases in the estimated insur. role of taxes & transfers

- Using estimated transmission coefficients:

$$\hat{\gamma}_{\phi} = 1 - \frac{1}{1-\gamma} \frac{\sigma^2_{\xi,e}(1-\gamma)^2 + s_{ni}(\mu^2_{\nu,ni} + \sigma^2_{\nu,ni})}{\sigma^2_{\xi,e} + s_e(\mu^2_{\nu,e} + \sigma^2_{\nu,e})}.$$

- Using estimated variances of shocks:

$$\hat{\gamma}_{\sigma_\xi} = 1 - \left(\frac{(1-\gamma)^2\sigma^2_{\xi,e} + s_{ni}(\mu^2_{\nu,ni} + \sigma^2_{\nu,ni})}{\sigma^2_{\xi,e} + s_e(\mu^2_{\nu,e} + \sigma^2_{\nu,e})} \right)^{\frac{1}{2}},$$

where s_{ni} (s_e) is the share of individuals with incomplete family income (family earnings) spells in a typical year, $\mu_{\nu,ni}$ ($\mu_{\nu,e}$) is the mean and $\sigma^2_{\nu,ni}$ ($\sigma^2_{\nu,e}$) is the variance of family income (family earnings) records surrounding the missing ones.

Consumption insurance estimates

	Panel A: Family earnings				Panel B: Net family income			
	Full	Bal.	Drop 1st & last 3	Ext. model	Full	Bal.	Drop 1st & last 3	Ext. model
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ϕ, transm. of perm. shock ξ	0.23 (0.04)	0.68 (0.17)	0.58 (0.12)	0.64 (0.17)	0.63 (0.08)	0.83 (0.18)	0.88 (0.14)	0.99 (0.13)
ψ, transm. of trans. shock ϵ	0.09 (0.03)	0.09 (0.05)	0.08 (0.04)	0.08 (0.03)	0.06 (0.04)	0.08 (0.06)	0.06 (0.04)	0.07 (0.04)
σ_ξ^2, var. perm. shock (avg.)	0.06 (0.006)	0.02 (0.004)	0.03 (0.005)	0.02 (0.003)	0.02 (0.002)	0.01 (0.002)	0.01 (0.002)	0.01 (0.001)
σ_ϵ^2, var. trans. shock (avg.)	0.07 (0.005)	0.06 (0.005)	0.06 (0.005)	0.07 (0.006)	0.03 (0.002)	0.04 (0.003)	0.04 (0.002)	0.04 (0.002)
Age (avg.)	43	46	43	43	43	47	43	43
Wealth (median)	50,200	60,222	50,679	50,200	50,288	61,626	51,231	50,288
No. househ.	2,430	478	2,430	2,430	2,430	516	2,429	2,430

Implied insurance due to taxes and transfers

Min. dist. est./Samp.	Full sample	Drop first & last 3 obs.	Extended model
Use perm. cons. insur., $\hat{\gamma}_\phi$	0.63	0.34	0.35

Another consequence of irregular income observations

- BPP methodology provides two measures for the insurance role of taxes and transfers, γ
  - γ_{ϕ} based on comparing transmission coeffs. for permanent shocks to net incomes and earnings
  - $\gamma_{\sigma_{\xi}}$ based on comparing variances of permanent shocks to net incomes and earnings
- The estimates differ dramatically: $\hat{\gamma}_{\phi} = 0.63 > \hat{\gamma}_{\sigma_{\xi}} = 0.38$
- Why?
- The difference is induced by the irregular nature of income observations around missing ones

Conclusion

- Empirical measures of the extent and sources of consumption insurance are the key:
  - benchmarks for assessing the performance of incomplete markets models
  - inputs into models evaluating the optimality of the tax and transfer system
- We provide such measurement building on BPP methodology
  - It is critical to account for the irregular nature of observations around missing ones
- Insurance of permanent shocks to net family incomes of 12% – no excess insurance
- Limited insurance role of the tax and transfer system – 35% of the shocks to family earnings are insured

How Much Consumption Insurance in the U.S.?

Dmytro Hryshko
University of Alberta

Iourii Manovskii
University of Pennsylvania

Introduction

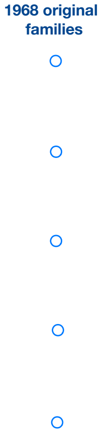
- The Panel Study of Income Dynamics (PSID) is key for research in social sciences.
- E.g., most of what is known to economists about joint income and consumption dynamics is based on data from the PSID.
- In this paper, we
  1. identify a very puzzling feature of PSID data:
     - two subsets of PSID families differ dramatically in family income and consumption dynamics
     - this is surprising given the PSID design;
  2. explain the sources of the difference;
  3. reassess excess insurance puzzle.

Sample and nonsample PSID members

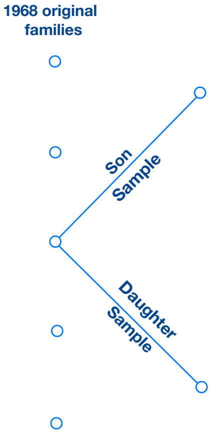
- PSID started in 1968 with a nationally representative cross-sectional sample.
- Since then, it follows this branch of the U.S. demographic tree.

- All members of families interviewed by the PSID in 1968 and their descendants (children, grandchildren, etc.) are called "sample" PSID members. Always tracked by the PSID.
- Those who marry PSID sample members after 1968 are called "nonsample" PSID members; tracked only during the marriage.
- We call a family "sample" or "nonsample" if the husband is a sample or nonsample PSID member, respectively.

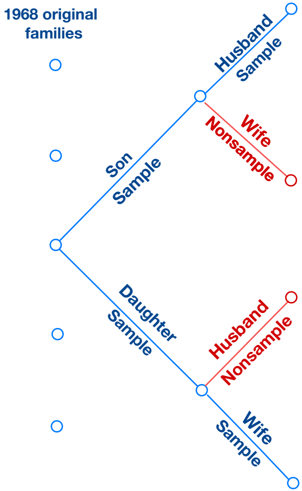
Sample and nonsample PSID members, I



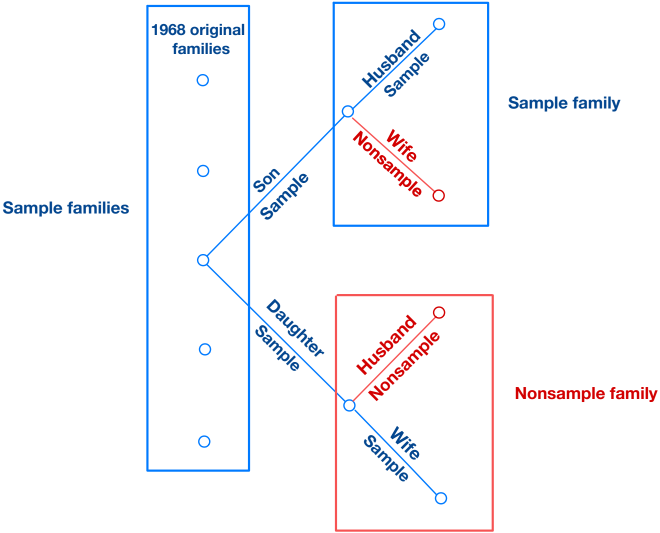
Sample and nonsample PSID members, II



Sample and nonsample PSID members, III



Sample and nonsample PSID families



Measuring income dynamics and consumption insurance

Blundell, Pistaferri, and Preston (BPP, 2008) methodology:

$$y_{it} = \alpha_i + p_{it} + \tau_{it}$$
$$p_{it} = \rho p_{it-1} + \xi_{it}, \qquad \rho = 1$$
$$\tau_{it} = \epsilon_{it} + \theta \epsilon_{it-1}$$

$$\Delta c_{it} = \phi \xi_{it} + \psi \epsilon_{it} + \zeta_{it} + \Delta u_{it}.$$

- ϕ and ψ measure transmission of permanent and transitory income shocks to consumption, respectively.

  $1 - \phi$ and $1 - \psi$ are measures of insurance.

- BPP estimated ϕ and ψ using the minimum-distance method, imputed **nondurable consumption** and **net family income** for married couples from the PSID.

Baseline insurance estimates

	Combined (1)	Sample (2)	Nonsample (3)
ϕ, transmission of perm. shock	0.6436 (0.0858)	0.9430 (0.1508)	0.4303 (0.0950)
ψ, transmission of trans. shock	0.0291 (0.0436)	–0.0108 (0.0469)	0.1014 (0.1009)

(Standard errors in parentheses.)

p-value for test of equal ϕ (ψ) between sample and nonsample families equals 0.4% (31%).

Composition of couples in the BPP dataset

- BPP objective: select couples with male heads of ages 30–65 with no marital status change during 1978–1992.

- In practice, many sample females divorcing and remarrying during 1978–1992 are left in the data (nonsample families). This doesn't drive the estimated difference in insurance.

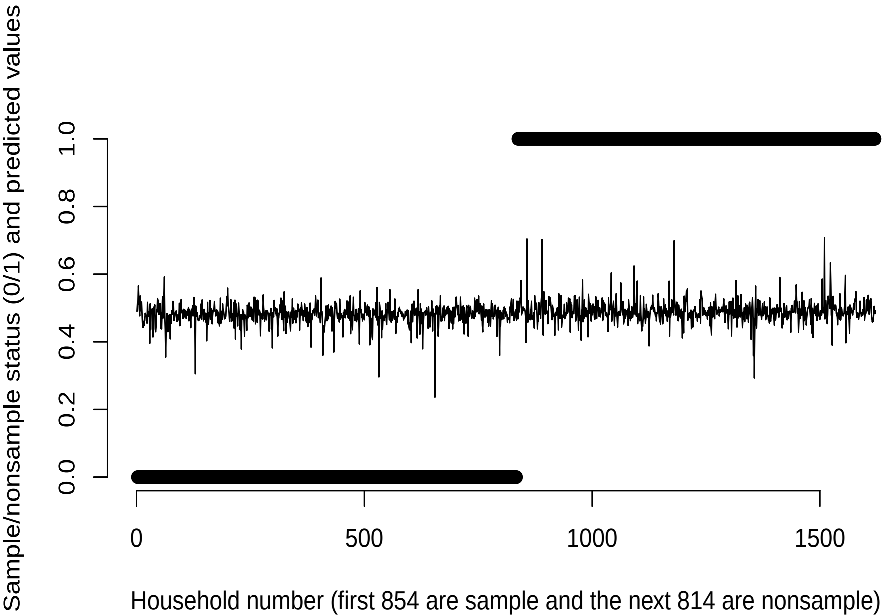
Families:	Formed <1978	Formed >=1978	Total
Sample	885	80	965
Nonsample	373	437	800

- We symmetrically add sample males remarrying and divorcing during 1978–1992 (i.e, keep data for all families).

Cross-sectional means

	Samp. orig. (1)	Samp. sons (2)	Nonsamp. (3)	p-value test: (2)=(3)
Head's age	51.149	38.54	38.834	43%
Wife's age	48.775	35.758	35.431	37%
White	0.912	0.936	0.916	16%
Region grew: foreign country	0.035	0.012	0.013	41%
Nondurable consumption	44015	24734	25650	43%
Net family income	41845	40389	40498	93%
Head's earnings	27547	27952	28038	94%
Wife's earnings	7074	10254	10042	68%
Transfers, family	2746	1449	1560	52%
Assets	132032	89005	90656	83%
If family owns house	0.923	0.823	0.810	43%
Head's hours	1933	2162	2181	55%
Wife's hours	949	1205	1200	90%
If head changed occupation	0.331	0.352	0.34	44%
If respondent wife	0.227	0.167	0.414	0%

LASSO prediction of sample/nonsample status



Insurance estimates (all families during 1978–1992)

	Sample sons	Non-sample	Sample original	Sample all	Sibling pairs Sons	Sibling pairs Daughters
	(1)	(2)	(3)	(4)	(5)	(6)
ϕ, transmission	0.87	0.46	1.09	0.90	1.07	0.32
perm. shock	(0.19)	(0.10)	(0.20)	(0.15)	(0.34)	(0.16)
ψ, transmission	0.07	0.12	0.04	0.05	0.15	–0.12
trans. shock	(0.08)	(0.09)	(0.04)	(0.04)	(0.13)	(0.18)

Consumption insurance by gender of the respondent

	Head respondent		Wife respondent	
	All	Nonsample	All	Nonsample
ϕ, transmission	0.5920	0.2149	0.6080	0.3812
perm. shock	(0.1030)	(0.1233)	(0.2630)	(0.0913)
ψ, transmission	0.0092	0.3899	0.1294	0.0845
trans. shock	(0.0541)	(0.1986)	(0.1090)	(0.2352)

- No evidence of different consumption insurance for male and female respondents.
- HRS linked to SSA earnings records: no difference in the dynamics of income measurement error for male and female respondents.
- Same patterns in 5 datasets from other countries modeled on the PSID, where individuals respond for themselves (discussed below).

Other measures of consumption and income

	Sample orig. (1)	Sample sons (2)	Non-samp. (3)	Sample orig. (4)	Sample sons (5)	Non-samp. (6)
	Panel A. Nondur. cons., total earnings			Panel B. Nondur. cons., male earninigs		
ϕ, transm.	0.45	0.50	0.15	0.66	0.57	0.14
perm. shock	(0.11)	(0.11)	(0.06)	(0.07)	(0.12)	(0.04)
ψ, transm.	0.03	0.07	0.17	–0.02	0.05	0.04
trans. shock	(0.03)	(0.05)	(0.07)	(0.06)	(0.03)	(0.06)
	Panel C. Food, total earnings			Panel D. Food, male earnings		
ϕ, transm.	0.31	0.35	0.12	0.43	0.37	0.09
perm. shock	(0.08)	(0.09)	(0.05)	(0.12)	(0.09)	(0.04)
ψ, transm.	0.02	0.03	0.11	0.03	–0.02	0.04
trans. shock	(0.03)	(0.04)	(0.05)	(0.05)	(0.03)	(0.05)

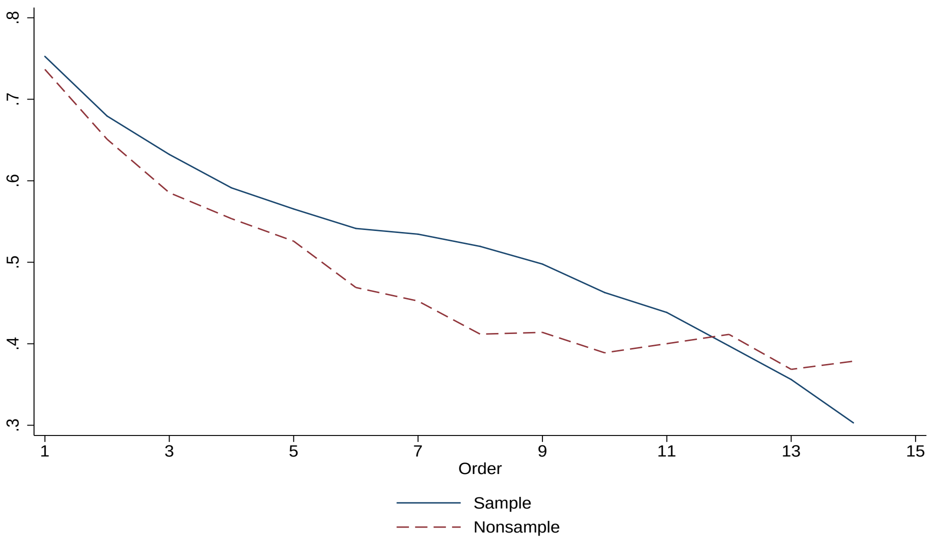
The role of the income process

- So far we relied on the standard assumption that the RW-MA(1) income process is the same for sample and nonsample households.

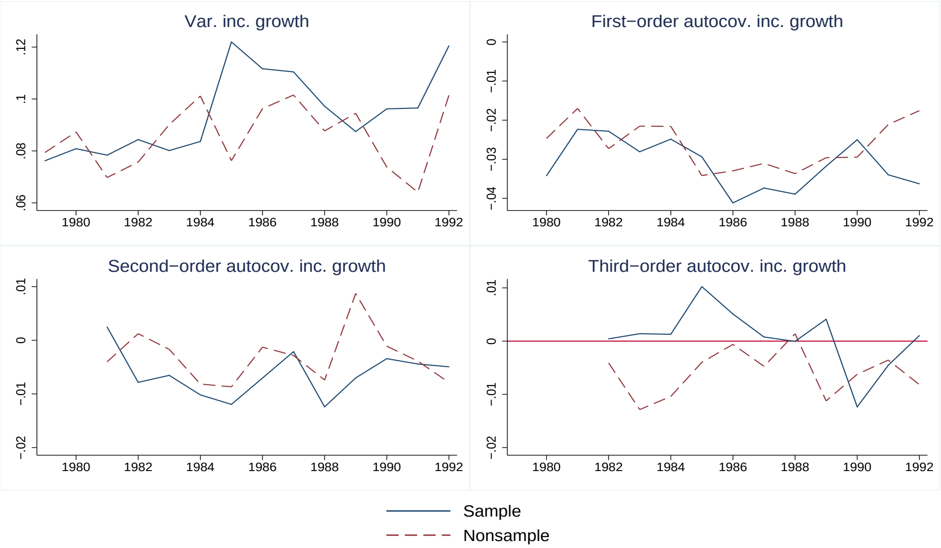
- While sample and nonsample households were compared cross-sectionally by, e.g., Becketti et al. (1988), the dynamic properties of their incomes were never compared.

- The measurement and the interpretation of consumption insurance depends on the income process.

Autocorrelation function of net family incomes



Data moments



Diff. income process: Evidence from inc. growth rates

Under the null of RW+MA(1) as in BPP:

$$E[\Delta y_{it} \Delta y_{it+k}] = 0, \qquad k \geq 3.$$

- **Sample**: p-value=42%. Fail to reject the null.

  Can maintain RW+MA(1).

- **Nonsample**: p-value<2%. Reject the null.

  Modify to AR(1)+MA(1).

GMM estimates of persistence

- GMM does not rely on fitting the autocovariance function of growth rates.

- Under the assumption $y_{i1} = m + \alpha_i + \xi_{i1} + \epsilon_{i1} + \theta\epsilon_{i0}$, α_i independent of shocks, rewrite the income process for $t > 1$ as

$$y_{it} - \rho y_{it-1} = (1-\rho)\alpha_i + \xi_{it} + \epsilon_{it} - (\rho - \theta)\epsilon_{it-1} - \rho\theta\epsilon_{it-2}.$$

- The following set of orthogonality conditions will hold:

$$E\left[(y_{it} - \rho y_{it-1})\Delta y_{it-j}\right] = 0, t = 1982, \ldots, 1992, j \geq 3.$$

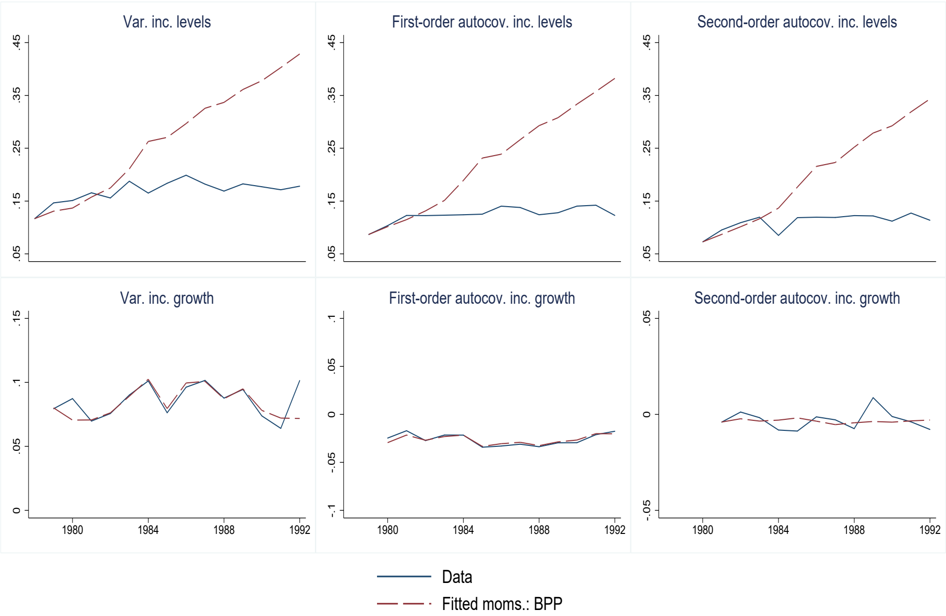
GMM estimates of persistence, ρ

	Overall data			Sibling pairs	
	Sample original	Sample sons	Non-sample	Sons	Daughters
	(1)	(2)	(3)	(4)	(5)
ρ, persistence	0.94	0.96	0.82	0.92	0.79
perm. shock	(0.03)	(0.04)	(0.05)	(0.04)	(0.07)

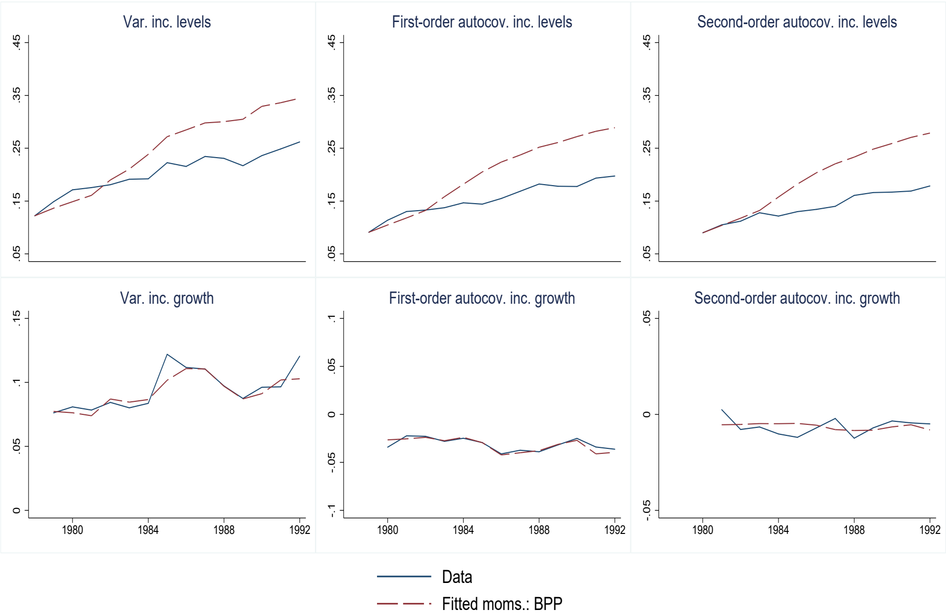
Refining the estimates of income processes

- Incomes of nonsample families appear less persistent.
- We will use a model to relate income and consumption dynamics for the two groups.
  - Getting persistence right is essential for measuring consumption insurance.
  - Getting variances of the shocks right is key for replicating income and wealth distributions.
- How does the standard RW-MA(1) income process fit the data?

Model Fit. RW income process. Nonsample families



Model Fit. RW income process. Sample families



Poor fit of the standard RW model

Poor fit can be due to:

1. misspecified persistence
   - allow for $\rho < 1$
2. systematically different income observations in the beginning and end of incomplete spells, with low mean and high variance (Daly, Hryshko, and Manovskii 2016).
   - allow for an extra transitory component χ_{it}

The effect of misspecified pers. of “permanent” shocks

Heathcote, Perri, Violante (2010). Identification for RW+iid:

Differences :

$$\sigma^2_{\xi,t,\text{diffs}} = E[\Delta y_{it}\Delta y_{it-1}] + E[\Delta y_{it}\Delta y_{it}] + E[\Delta y_{it}\Delta y_{it+1}]$$
$$\sigma^2_{\epsilon,t,\text{diffs}} = -E[\Delta y_{it}\Delta y_{it+1}].$$

Levels :

$$\sigma^2_{\xi,t,\text{levs}} = E[y_{it}y_{it+1}] - E[y_{it}y_{it-1}]$$
$$\sigma^2_{\epsilon,t,\text{levs}} = E[y_{it}y_{it}] - E[y_{it}y_{it+1}].$$

Should deliver the same estimates if the true process is RW+iid.

BPP used moments for growth rates.

Biases. Misspecified persistence

If the permanent component is an AR(1) process instead ($p_{it} = \rho p_{it-1} + \xi_{it}$),

$$\sigma^2_{\xi,t,\text{diffs}} - \sigma^2_{\xi,t,\text{levs}} = (1-\rho)(\rho + \rho^3)\text{var}(p_{t-2}) > 0$$
$$\sigma^2_{\epsilon,t,\text{levs}} - \sigma^2_{\epsilon,t,\text{diffs}} = \rho(1-\rho)\text{var}(p_{t-1}) > 0.$$

Accounting for misspecification. Modified inc. process

$$y_{it} = \alpha_i + p_{it} + \tau_{it} + \chi_{it}, \quad t = t_0, \ldots, T$$

$$p_{it} = \rho p_{it-1} + \xi_{it}$$

$$\tau_{it} = \epsilon_{it} + \theta \epsilon_{it-1}$$

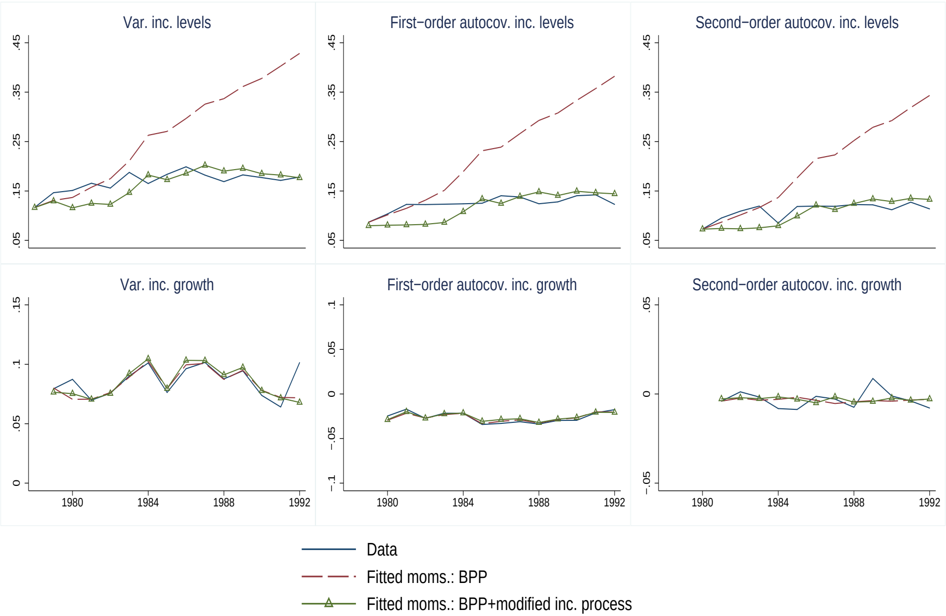
$$\chi_{it+j} = \begin{cases} \nu_{it} & \text{if } y_{it-1} \text{ or } y_{it+1} \text{ is missing and } t-1 \geq t_0,\ t+1 \leq T, j = 0 \\ \theta \nu_{it} & j = 1 \\ 0 & \text{otherwise.} \end{cases}$$

SMM estimates: BPP+modified income process

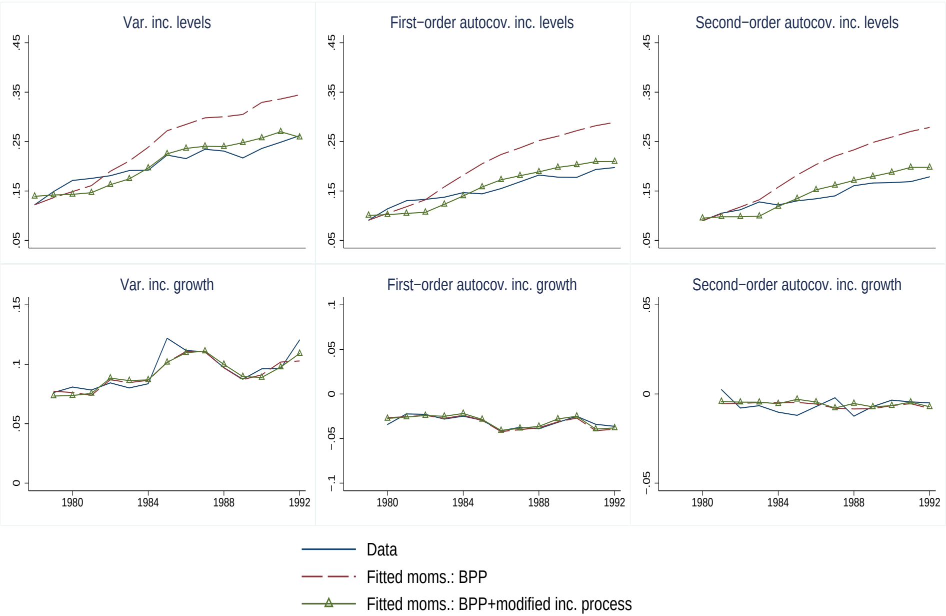
	Sample (1)	Nonsample (2)
ρ, AR coefficient	0.9956	0.9003
σ_ξ^2, variance perm. shock (avg.)	0.015	0.029
θ, MA coefficient	0.1315	0.0456
σ_ϵ^2, variance trans. shock (avg.)	0.041	0.027
ϕ, transmission perm. shock	0.9903	0.4798
ψ, transmission trans. shock	0.0922	0.0997

Income for nonsample households is less persistent and insurance of persistent shocks is much higher.

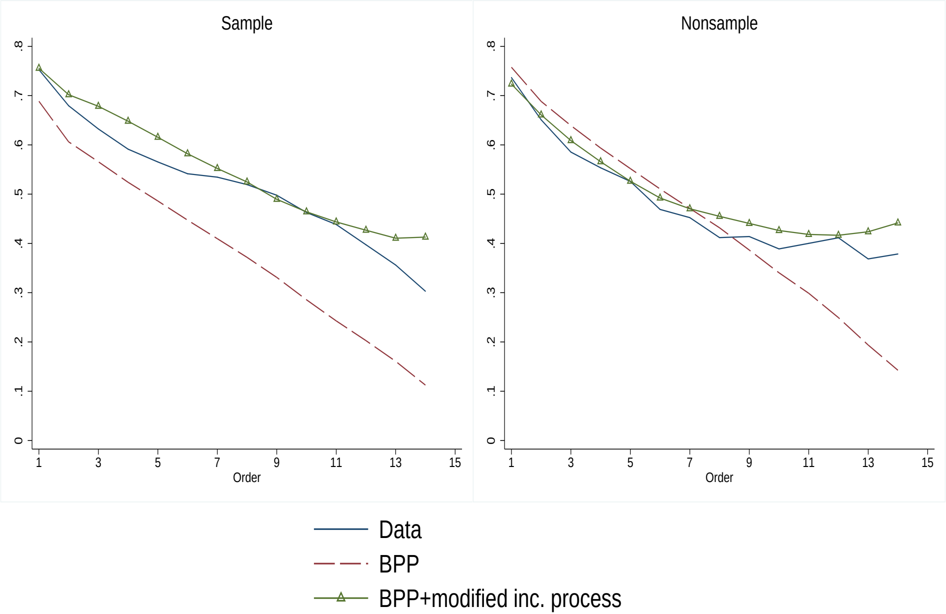
Model Fit. Modified inc. process. Nonsample families



Model Fit. Modified inc. process. Sample families



Model fit. Acf of income levels



Income and consumption dynamics in a lifecycle model

- We found different consumption insurance for sample and nonsample families.
- We found different variances and the persistence of income shocks for sample and nonsample families.
- We need a model to interpret the relationship between income and consumption dynamics.

Model

$$\max_{\{C_{it}\}_{t=t_0}^{T}} E_{i,t_0} \sum_{t=t_0}^{T} \beta^{t-t_0} s_t \frac{C_{it}^{1-\gamma} - 1}{1-\gamma},$$

subject to

$$\begin{aligned}
W_{it+1} &= (1+r)(W_{it} + Y_{it} - C_{it}), \\
Y_{it} &= \mu_t P_{it} V_{it}, \quad t = t_0, \ldots, t_R \\
P_{it} &= P_{it-1}^{\rho} \exp(\xi_{it}) \\
V_{it} &= \begin{cases} \exp(\epsilon_{it}), & \text{with prob. } 1-\pi \\ 0, & \text{with prob. } \pi \end{cases} \\
Y_{it} &= \kappa P_{it_R}, \quad t = t_R + 1, \ldots, T \\
W_{it} &\geq 0, \; t = t_0, \ldots, T.
\end{aligned}$$

Calibration and simulations

1. Calibration
   - Target the data for nonsample families only.
   - Use their estimated income process.
   - Calibrate γ, β, and π by fitting percentiles of their wealth distribution.
2. Simulations
   - Fix the calibrated parameters.
   - Simulate data for the families of sons and daughters, using their respective income processes.

Model calibration

	Sons		Daughters	
	Data (1)	Model (2)	Data (3)	Model (4)
Various income percentiles, in ’000s				
P10	16.9	20.4	16.7	22.0
P25	27.9	28.6	27.7	27.9
P50	35.9	35.5	36.3	37.1
P75	49.3	44.4	49.6	48.8
P90	66.6	64.5	66.5	63.3
Various wealth percentiles, in ’000s				
P10	4.7	11.9	4.3*	3.9
P25	19.7	26.4	18*	14.9
P50	54.2	57.7	48*	47.9
P75	119.7	118.2	125.4*	129.8
P90	218.4	220	254.2*	265.4
Internally calibrated parameters				
Time disc. factor, β			0.969**	
Coeff. RRA, γ			0.405**	
Prob. of zero inc. state, π			0.006**	

Notes: * indicates calibration targets; ** marks calibrated parameter values.

Consumption insurance in simulated and PSID data

	Sons		Daught.		Comb. Sons & Daught.		
	PSID	Model	PSID	Model	PSID	Model	RW Model
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
ϕ, transm.	0.87	0.84	0.46	0.45	0.57	0.62	0.92
perm. shock	(0.19)	(0.13)	(0.10)	(0.06)	(0.09)	(0.06)	(0.12)
ψ, transm.	0.07	0.06	0.12	0.14	0.09	0.08	0.06
trans. shock	(0.08)	(0.05)	(0.09)	(0.06)	(0.06)	(0.04)	(0.06)

Why different income dynamics?

- We compare PSID data to administrative earnings data linked to the 1998 HRS (PSID sample individuals born before 1948 and observed in 1998).

	Family earnings		Male earnings	
	PSID (1)	HRS-SSA (2)	PSID (3)	HRS-SSA (4)
ρ, persistence	0.98	0.93	1.02	0.96
perm. shock	(0.04)	(0.02)	(0.02)	(0.02)
No. ind./fam.	508	1822	520	2628

- Initial PSID sample is representative cross-sectionally but possibly not dynamically.

GMM estimates of persistence. Various datasets.

Dataset:	PSID		GSOEP		BHPS		HILDA		KLIPS		SHP	
Country:	U.S.A.		Germany		U.K.		Australia		Korea		Switz.	
	S	NS	S	NS	S	NS	S	NS	S	NS	S	NS
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
ρ, persist.	0.95	0.83	0.93	0.89	0.85	0.80	0.96	0.88	0.92	0.79	0.91	0.82
No. fam.	1593	889	2044	423	2467	554	3286	949	2181	516	1625	258

"NS" and "S" stand for nonsample and sample families, respectively.

- In all of the datasets, incomes of sample families are more persistent than incomes of nonsample families.
- It's unlikely that all of them oversample more persistent families in their original selection.

Attrition rates

Dataset:	PSID	GSOEP	BHPS	HILDA	KLIPS	SHP
Country:	U.S.A.	Germany	U.K.	Australia	Korea	Switzerland
	(1)	(2)	(3)	(4)	(5)	(6)
Men	51.1	63.2	56.3	58.4	67.2	77.8
Women	44.5	58.5	48.1	56.2	61.9	75.7

- In all of the datasets, women are less likely to attrit than men.

GMM estimates of persistence by attrition. PSID

	Sample		Nonsample	
	Non-attritors (1)	Attritors (2)	Non-attritors (3)	Attritors (4)
ρ, persistence perm. shock	0.96 (0.02)	0.78 (0.05)	0.89 (0.04)	0.72 (0.07)
No. families	1156	174	585	74

- Attriting families have a lower persistence of the permanent income component.

GMM estim. of persist. by attrition. Various datasets

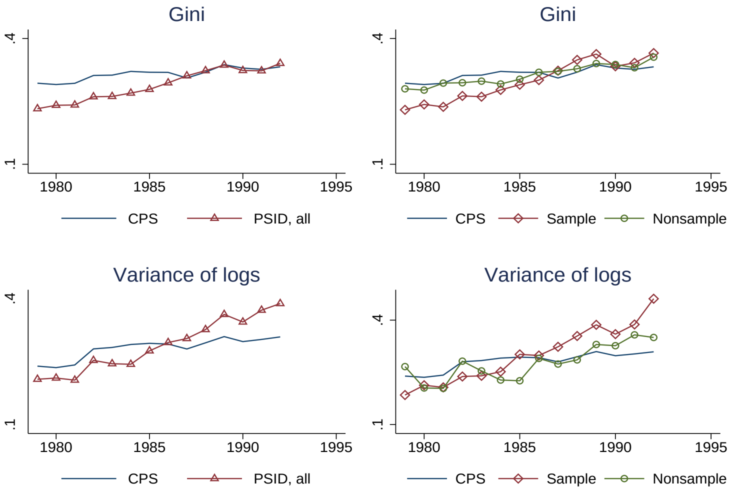
Dataset:	PSID		GSOEP		BHPS		HILDA		KLIPS		SHP	
Country:	U.S.A.		Germany		U.K.		Australia		Korea		Switz.	
	NA	A	NA	A	NA	A	NA	A	NA	A	NA	A
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
ρ, persist.	0.91	0.81	0.99	0.75	0.88	0.83	0.87	0.66	0.87	0.56	0.96	0.66
No. fam.	573	56	627	101	724	133	836	67	654	86	449	53

"NA" and "A" stand for families of non-attritors and attritors, respectively.

Selective attrition and its effects

1. Attrition is not random.
2. Males are more likely to attrit than females, and so sample families are more likely to attrit than nonsample families.
3. Families of attritors have lower persistence.
4. Thus, the remaining sample families have higher income persistence than the remaining nonsample families.
   - This also explains a higher persistence of PSID vs. HRS-SSA.
5. Nonsample families are a less selected set, and so are a better guide to the income dynamics and consumption insurance of a typical U.S. family.

Inequality in the PSID and CPS, net family incomes



For the same cohorts, income inequality trends for nonsample families are more aligned with the CPS.

Conclusion

- Nearly everything known about the joint behavior of consumption and income in the U.S. is based on the PSID.
- We find striking differences across "sample" and "nonsample" families:
  - in the insurance of permanent shocks,
  - in the persistence of permanent shocks.
- Estimated insurance in each subgroup is consistent with the incomplete-markets theory.
  - No apparent excess insurance/excess smoothness puzzle.
- The differences in income dynamics between sample and nonsample households can be explained by selective attrition on gender and income persistence.
- Families of daughters of the original PSID members are a less selected set and are a better guide to the income dynamics and consumption insurance available to a typical U.S. household.