Reconciling Estimates of Earnings Processes in Growth Rates and Levels

Moira Daly, Copenhagen Business School
Dmytro Hryshko, University of Alberta
Iourii Manovskii, University of Pennsylvania

ECB
December 18, 2015

Motivation

- There is a large literature estimating stochastic processes for individual earnings.
  - Recently, using large administrative data
  - Typically unbalanced panels (select strong labor-market attachment, no outliers, etc.)
- Crucial ingredient of many quantitative 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 issue.

The 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 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!

Levels vs. differences. Question

RED 2010, Cross-sectional facts for macroeconomists:

The estimates using the moments of log-earnings in levels, $E[y_{it}y_{it+j}]$, are substantially different from the estimates using the moments of log-earnings in differences, $E[\Delta y_{it}\Delta y_{it+j}]$.

$$\text{Permanent variance,} \quad \sigma_\xi^2 \text{ (levels)} \ll \sigma_\xi^2 \text{ (differences)}$$
$$\text{Transitory variance,} \quad \sigma_\epsilon^2 \text{ (levels)} \gg \sigma_\epsilon^2 \text{ (differences)}$$

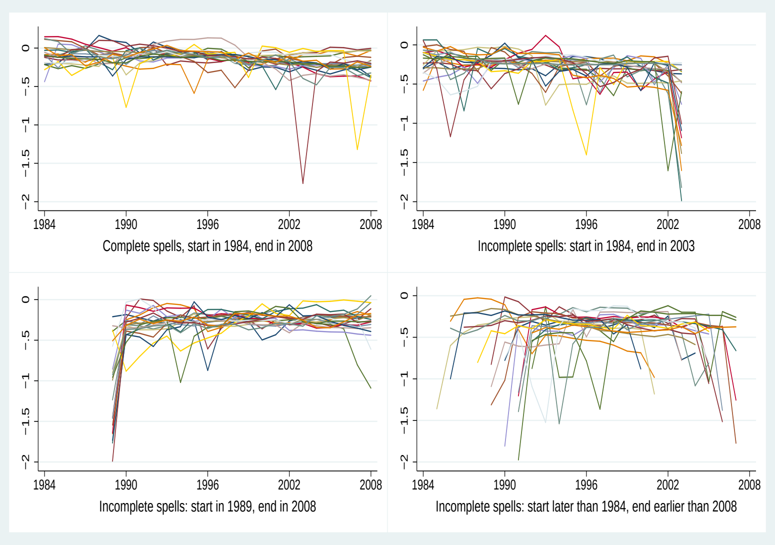
Why?

Heathcote, Storesletten, Violante (2010): misspesification.

Our contribution

- Using administrative Danish (1981–2006) and German (1984–2008) data, we find the source of discrepancy.
- Earnings are lower on average and are more volatile in the few first and last periods of the incomplete contiguous earnings spells. Fig.
- Those "outlying" observations contribute differently to the moments in levels and growth rates, and create divergence in the levels/growth estimations.
- These spell characteristics matter in recovering parameters of the earnings process and the extent of consumption insurance against permanent shocks.

Random earnings paths. German data.



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 are expected to deliver the same estimates of perm. (trans.) variance in a sample of individuals whose earnings are nonmissing for the periods $t-2$ ($t-1$) through $t+1$ ($t+1$).
- What if individual earnings spells have missing observations, or have an earnings spell starting at $t-2$ and/or ending at $t+1$?

Identification.

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

Ex. $\nu_{it} \neq 0$ if $t = t_0^i$ and $t_0^i \neq t_0$

ν_{it} uncorr. with $\alpha_i, \xi_{is}, \epsilon_{is},$ all t, s

Levels:

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

Differences:

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

Identification.

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

Ex. $\nu_{it} \neq 0$ if $t = t_0^i$ and $t_0^i \neq t_0$

ν_{it} uncorr. with $\alpha_i, \xi_{is}, \epsilon_{is},$ all t, s

Permanent shock:

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

$s_{t,t+1}$ 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: $\hat{\sigma}^2_{\epsilon,t} - \sigma^2_{\epsilon,t} = s_t(\mu^2_\nu + \sigma^2_\nu)$
- 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 non-positive.

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 data set (2 unbalanced and one balanced):

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.

The estimated earnings process

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

The permanent component:

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

The transitory component:

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

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

Unbalanced panels. German data

	9 or more cons.		20 not nec. cons.	
	Levs.	Diffs.	Levs.	Diffs.
$\hat{\phi}_p$	0.976	0.992	0.999	0.991
Var. perm. shocks, $\hat{\sigma}_\xi^2$	**0.0078**	**0.019**	**0.0048**	**0.009**
$\hat{\theta}$	0.129	0.153	0.119	0.192
Var. trans. shocks, $\hat{\sigma}_\epsilon^2$	**0.024**	**0.009**	**0.016**	**0.009**
$\hat{\sigma}_\alpha^2$	0.024	—	0.027	—

All coefficients significant at the 1% level.

Balanced sample. German data

	Levs.	Diffs.
$\hat{\phi}_p$	0.998	0.994
Var. perm. shocks, $\hat{\sigma}^2_\xi$	**0.0033**	**0.0036**
$\hat{\theta}$	0.193	0.185
Var. trans. shocks, $\hat{\sigma}^2_\epsilon$	**0.007**	**0.007**
$\hat{\sigma}^2_\alpha$	0.023	—

All coefficients significant at the 1% level.

Missing records in administrative data

- There's no (random) attrition in administrative data; there must be reasons for missing earnings records (besides cleaning restrictions such as working less than 35 days).
- For example, an individual is in transition from the pool of unemployed/out-of-labor-force into the pool of employed in the first period of the spell.

Mean residual earnings

Residual earnings are lower than average in the (few) first and last periods of the earnings spell (if not the first sample year/the last sample year); and in the (few) periods before and after a year of missing earnings (non-consecutive sample).

Consistent with working less, and lower wages in the first period of the spell (e.g., Altonji and Shakotko (1987) find high wage growth in the first year on the job).

Residual earnings. Panel regressions. German data

	9 or more consec. (1)	**20 not nec. consec.** (2)
Year obs.: first	–0.57	–0.65
Year obs.: last	–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

All coefficients significant at the 1% level.

Volatility of residual earnings

Residual earnings are more volatile in the (few) first and last periods of the earnings spell (if not the first sample year/the last sample year); and in the periods before and after a year of missing earnings (non-consecutive sample).

Squared residual earnings. Panel regressions. German data

	9 or more consec. (1)	**20 not nec. consec.** (2)
Year observed: first	0.23	0.29
Year observed: last	0.20	0.25
1 year before earn. miss., dummy		0.15
1 year after earn. miss., dummy		0.23
No. obs.	379,080	330,748
No. indiv.	18,130	13,635

All coefficients significant at the 1% level.

Earnings, wage, and hours residuals. German data.

	9 or more cons.			20 not nec. consec.		
	Earn. (1)	Hours (2)	Wages (3)	Earn. (4)	Hours (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 earn. miss.				–0.27	–0.23	–0.04
1 year after earn. 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.

Spell years, unemp. and job mobility. German data, consec. sample.

	Changed occ., %	Changed industry, %	Unemp. %
Year obs.: first, first spell year = 1984	8.32	5.96	4.71
Year obs.: first, first spell year ≠ 1984	36.89	35.32	22.51
Year obs.: last, last spell year = 2008	2.05	1.80	1.96
Year obs.: last, last spell year ≠ 2008	7.34	6.38	26.02
Overall avg.	5.21	3.95	2.96

What to do?

1. Drop earnings observations around missing records;
2. model those observations.

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
$\hat{\sigma}^2_\xi$	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
$\hat{\sigma}^2_\epsilon$	0.024	0.009	0.010	0.009	0.01	0.01	0.01	0.01
$\hat{\sigma}^2_\alpha$	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.

Implications for Measuring Consumption Insurance Against Shocks to Earnings

$$\Delta c_{it} = \phi_t \xi_{it} + \psi_t \epsilon_{it}.$$

Can we recover the true insurance coefficients against permanent and transitory shocks, $1 - \phi_t$ and $1 - \psi_t$?

Insurance Coefficients. Blundell, Pistaferri, Preston (2008) moments

$$\Delta c_{it} = \phi_t \xi_{it} + \psi_t \epsilon_{it},$$

ξ_{it}, and ϵ_{it} are permanent and transitory shocks to male earnings.

Perm. ins.: $1 - \hat{\phi}_t = 1 - \frac{E[\Delta c_{it} \Delta y_{it-1}] + E[\Delta c_{it} \Delta y_{it}] + E[\Delta c_{it} \Delta y_{it+1}]}{E[\Delta y_{it} \Delta y_{it-1}] + E[\Delta y_{it} \Delta y_{it}] + E[\Delta y_{it} \Delta y_{it+1}]}$

Trans. ins.: $1 - \hat{\psi}_t = 1 - \frac{E\left[\Delta c_{it} \Delta y_{it+1}\right]}{E\left[\Delta y_{it} \Delta y_{it+1}\right]}$

Biases in Estimated Insurance Coefficients Due to Presence of Rare Shocks

$$\Delta c_{it} = \phi_t \xi_{it} + \psi_t \epsilon_{it} + \psi_{t,\text{rare}} \nu_{it}, \quad \psi_{t,\text{rare}} = \psi_t, \quad \nu_{it} \sim (\mu_\nu, \sigma_\nu^2)$$

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

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

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

$$(1 - \hat{\phi}_t) - (1 - \phi_t) = (\phi_t - \psi_t)\lambda_t$$

2. Non-consecutive unbalanced samples (earnings missing at t):

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

Application to the PSID: BPP male earnings and (imputed) household nondurable consumption data: 1978–1992.

BPP data: 1978–1992. Male earnings residuals, PSID

Dependent variable:	Residuals		Squared residuals	
	(1)	(2)	(3)	(4)
Year observed: first	–0.11***	–0.01	0.09**	–0.18***
Year observed: last	–0.07*	0.00	0.57***	0.18***
1 year before earn. miss.	–0.32**	–0.33**	1.56***	1.57***
1 year after earn. miss.	–1.11***	–1.11***	1.35***	1.40***
No. obs.	16,496	16,496	16,496	16,496
No. indiv.	1,741	1,741	1,741	1,741

The earnings process as in BPP

1. Full sample:

$$\text{Levels}: \quad \hat{\sigma}^2_\xi = 0.017 \qquad \hat{\sigma}^2_\epsilon = 0.20$$
$$\text{Differences}: \quad \hat{\sigma}^2_\xi = 0.071 \qquad \hat{\sigma}^2_\epsilon = 0.095 \qquad \hat{\phi} = 0.26$$

74% of the permanent shocks to male earnings are insured.

2. Drop outlying obs. [drop random obs.]:

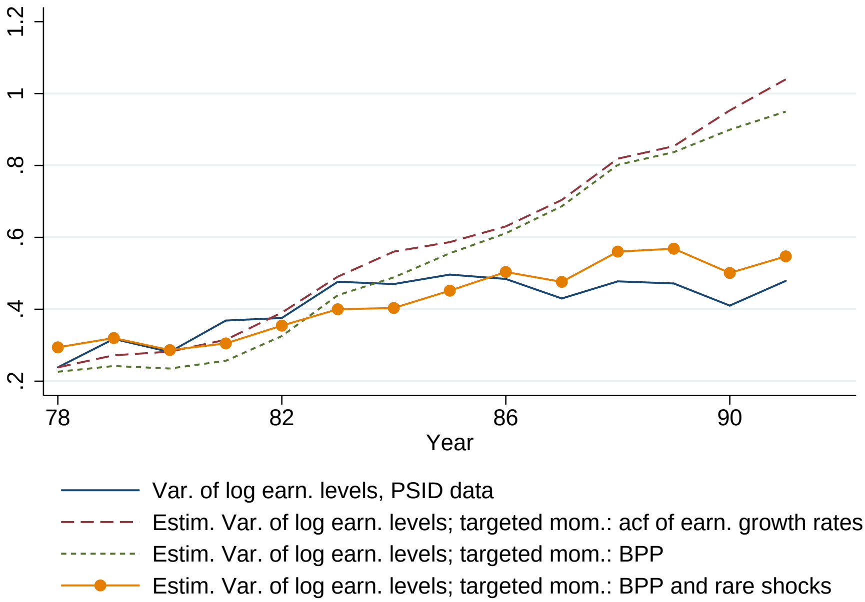
$$\text{Levels}: \quad \hat{\sigma}^2_\xi = 0.015 \qquad \hat{\sigma}^2_\epsilon = 0.108$$
$$\text{Differences}: \quad \hat{\sigma}^2_\xi = 0.022 \qquad \hat{\sigma}^2_\epsilon = 0.084 \qquad \hat{\phi} = 0.64 \quad [\text{mean } 0.27]$$

36% of the permanent shocks to male earnings are insured.

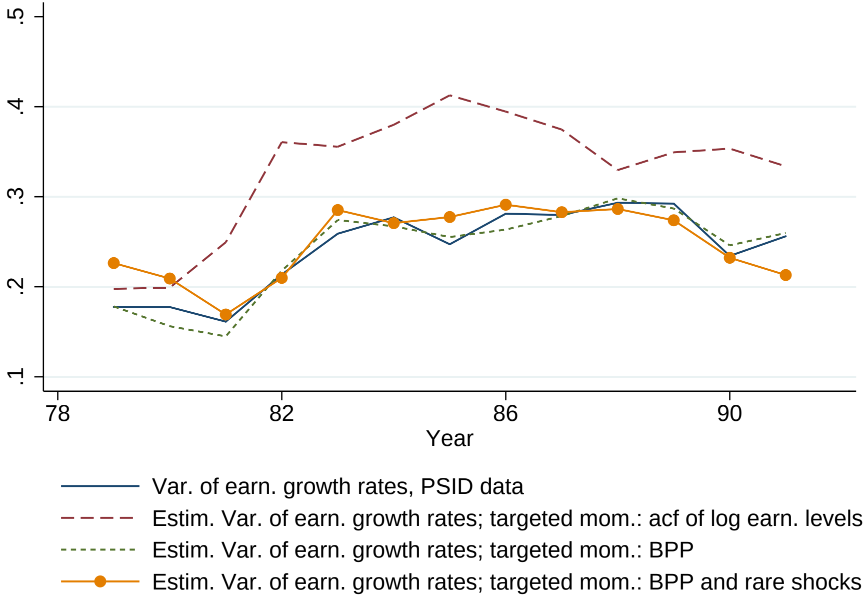
3. Estimate $\hat{\phi}$, $\hat{\psi}$, and $\hat{\psi}_{\text{rare}}$, and the earn. process using BPP moments, and the data mom. on outlying obs. $\hat{\phi} = 0.57$

43% of the permanent shocks to male earnings are insured.

Fit to the variances of log earnings in levels



Fit to the variances of log earnings in differences



Conclusion

- A puzzle: the moments in levels and differences deliver different estimates of permanent and transitory risk.

- We find a source of difference in administrative data: large deviations of earnings at the start and end of continuous individual earnings spells. Also true for PSID male earnings data.

- They also induce a substantial upward bias in the estimates of the insurance against permanent shocks.

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 additional transitory shocks at the beginning and the end of contiguous earnings histories into the estimation—the mean and the variance of these shocks can be identified from the mean and the variance of earnings in those periods.

Insurance coefficients for male earnings shocks

	(1) BPP moms.	(2) BPP moms.+ rare shocks
ϕ	0.2629	0.5719
(Partial ins. perm. shock)	(0.0549)	(0.1989)
ψ	0.0364	–0.0285
(Partial ins. trans. shock)	(0.0295)	(0.0369)
ψ, rare shock	—	0.3679
(Partial ins. rare trans. shock)	—	(0.1028)