Income Process

Growth and Development

Raül Santaeulàlia-Llopis

MOVE-UAB and Barcelona GSE

Spring 2017

Estimating Income Processes in Developing Countries

- To estimate the income process we use a flexible specification that allows for permanent and transitory components, the permanent-transitory model; an "industry standard".

- The set of references in estimating income process that separate permanent and transitory shocks in poor countries is somewhat limited: Deaton and Paxson (1994), Deaton (2000), Santaeulàlia-Llopis and Zheng (2015), Meguir et al. (2016) .

- This is important because:
    - The use of estimated labor income processes is standard in macroeconomic models with heterogeneous agents where a typical source of heterogeneity is (shocks to) income.
    - A good understanding of the income process is key to study the response of consumption to income shocks.
    - **A large part of income in poor countries extends beyond labor income. For example, agricultural income represents 60% of income in the rural areas of Malawi, where 81% of the population resides. There, labor income represents a much lesser 19% of total income and non-farm businesses 10%.** (For example, see de Magalhaes and Santaeulàlia-Llopis (2015) for Malawi, Uganda and Tanzania).

Income sources in Malawi:

	Bottom(%)			Quintiles					Top(%)			All
	0-1	1-5	5-10	1st	2nd	3rd	4th	5th	10-5	5-1	1	0-100
						Income Sources (%)						
Labor	-22	22	17	20	19	17	17	19	17	26	13	19
Agriculture	24	54	61	57	60	63	66	57	66	57	41	60
Fishing	114	-0	0	-1	-0	0	0	5	1	2	16	3
Business	-2	4	3	3	4	5	6	14	8	11	29	10
Capital	-7	1	1	1	1	1	2	2	2	3	0	2
Transfers	1	2	3	3	2	1	1	0	0	-0	0	1
Food Gifts	-9	18	16	17	14	12	9	3	5	2	0	7
	100	100	100	100	100	100	100	100	100	100	100	100

Taken from De Magalhaes and Santaeulalia-Llopis (2015).

- A typical specification for the labor income process is:

$$y_{i,a,t} = \sum_t \alpha_t \mathbf{1}_t + f(a;\Theta) + g(x_{i,a,t};\Gamma) + u_{i,a,t} \quad (1)$$

where $y_{i,a,t}$ is a logged measure of income for individual i of age a at period t. Regarding the explanatory variables:

$\rightarrow$ α_t are year dummies;[1]

$\rightarrow$ and function $f(a;\Theta)$ is a deterministic function of age (e.g., a quartic polynomial).

$\rightarrow$ function $g(x_{i,a,t};\Gamma)$ may include family composition controls, sex dummies, education, race, regional dummies, and also, potentially, individual-specific fixed effects;[2,3]

Under the assumption that $E(u_{i,a,t}|x_{i,a,t}) = 0$ we can estimate (1) by OLS and work straight with the estimated residual $u_{i,a,t}$.

[1] One could run the wage equation (1) separately for each year, rather than controlling for time dummies. Running the equation separately for each year would make the shape of the functions g and f depend on time. For the same token, one could also run this wage equation separately for education groups or sex groups.

[2] The family composition controls are, potentially, and among others, a set of dummies for marital status (married, never married, widowed, divorced) together with a control for children and old dependents.

[3] The education dummies may correspond to the maximum degree attained as no schooling, primary school drop-outs, primary school, and secondary school or higher. Otherwise, schooling years is also an option.

What is $y_{i,a,t}$?

- In developed countries, the individual measure of income, $y_{i,a,t}$, that is usually implemented in the wage equation (1) is annual labor income, earnings, or simply wages.
- This might still be the case in urban areas of poor countries. BUT...
- ... in **rural areas, we need to get a good measure of agricultural income.**
- In a dynamic growing economy like China, the non-agricultural self-employed are an important proportion of the population as well.

Sample restrictions (quite different from those imposed in developing countries)

- No reason to focus only on 25-55. Education is completed much younger, look at ages starting at 20. Few household heads though. No retirement age.
- We are finding large changes in growth rates. Trim by that.
- We are finding large changes in growth rates. Practice suggests trimming income growth, consumption growth, and the interaction terms. In Tanzania, about 5% of households have growth rates of disposable income larger than 2 in absolute terms. Similar numbers in Uganda. Consumption growth lies mostly below 2 in this countries.
- We want the self-employed in. In particular, the farmers.

A Permanent-Transitory Decomposition

We assume that the estimated residual, $u_{i,a,t}$, i.e., the unobserved idiosyncratic component, consists of a persistent component $z_{i,a,t}$ and a transitory component $\epsilon_{i,a,t}$. Why?

- Residual wage inequality, i.e., the variance of $u_{i,a,t}$ (some average across inviduals of all ages), grows over time in the US. This has justified the use of nonstationary models for labor income of the type suggested in Gottschalk and Moffit (1994) in the literature.

- The Friedman's permanent income hypothesis (PIH) emphasizes the distinction between permanent and transitory shocks to income to understand the response of income. Specifically, the PIH model suggests that consumption responds one-to-one to permanent shocks but not so to transitory shocks. In a life-cycle version of the PIH model we will see that the response of consumption to a transitory shock depends on the time horizon; young individuals do not respond to transitory shocks as they still have a long road to live, but this response increases with age.

- The PT model poses the unexplained income growth (i.e. the residual) as a function of two components, a persistent component $z_{i,a,t}$ and a transitory component $\epsilon_{i,a,t}$:

$$u_{i,a,t} = z_{i,a,t} + \epsilon_{i,a,t}, \quad (2)$$

$$z_{i,a,t} = z_{i,a-1,t-1} + \eta_{i,a,t}, \quad (3)$$

with $\epsilon_{i,a,t} \sim iid(0, \sigma^2_{\epsilon_{a,t}})$ and $\eta_{i,a,t} \sim iid(0, \sigma^2_{\eta_{a,t}})$. That is, innovations to the transitory shock, $\epsilon_{i,a}$, and the innovations to the persistent shock, $\eta_{i,a}$, are iid across individuals, orthogonal to each other, and uncorrelated over age groups.[4,5]

[4] Originally Gottschalk and Moffit (1994) specify an ARMA(1,1) for the transitory component and a random walk for the permanent component.

[5] An earlier literature used error components models with relatively simple specifications for the transitory component (e.g., a first-order autoregression), and the permanent component was often assumed not to evolve over time (See Lillard and Willis (1978), and Hause (1980).

- Note that we allow the variance of the permanent and transitory shocks to depend on age and time.

- Later we will want to, perhaps, drop the dependence on age or time and focus on either $\{\sigma^2_{\epsilon_a}, \sigma^2_{\eta_a}\}$ or $\{\sigma^2_{\epsilon_t}, \sigma^2_{\eta_t}\}$. Alternative identification assumptions will apply.

Identification using the covariance structure of the residuals

The PT model

$$
\begin{aligned}
u_{i,a,t} &= z_{i,a,t} + \epsilon_{i,a,t}, \\
z_{i,a,t} &= z_{i,a-1,t-1} + \eta_{i,a,t}.
\end{aligned}
$$

with $\{\sigma^2_{\epsilon_{a,t}}, \sigma^2_{\eta_{a,t}}, \sigma^2_{z_{a,0}}\}$ has $(2 \times T \times A + A)$ parameters that we need to identify.

- To identify these paremeters we impose restrictions on the covariance structure of the PT process (Heathcote, Perri and Violante, 2010).[6]
- The literature that estimates labor income processes approaches the identification of $\{\sigma^2_{\epsilon_{a,t}}, \sigma^2_{\eta_{a,t}}\}$ using two alternative set of moments: the (auto)covariance structure of (i) income growth rates (more typical for labor economists) and (ii) log levels (more typical in macroeconomics).[7]

[6] See Hall and Mishkin 1982, Abowd and Card (1989), and Gottschalk and Moffit (1994), for earlier treatments.

[7] See a discussion in **?** and **?**.

Identification using growth rates moments

(0) Note that the growth rate of income at t is:

$$\begin{aligned}\Delta u_{i,a,t} &= u_{i,a,t} - u_{i,a-1,t-1}\\ &= z_{i,a,t} + \epsilon_{i,a,t} - (z_{i,a-1,t-1} + \epsilon_{i,a-1,t-1})\\ &= (z_{i,a-1,t-1} + \eta_{i,a,t}) + \epsilon_{i,a,t} - (z_{i,a-1,t-1} + \epsilon_{i,a-1,t-1})\\ &= \eta_{i,a,t} + \epsilon_{i,a,t} + \epsilon_{i,a-1,t-1}\end{aligned}$$

and the growth rate one period ahead,

$$\begin{aligned}\Delta u_{i,a+1,t+1} &= u_{i,a+1,t+1} - u_{i,a,t}\\ &= \eta_{i,a+1,t+1} + \epsilon_{i,a+1,t+1} + \epsilon_{i,a,t}\end{aligned}$$

These moments can be computed for the whole sample or within a group of individuals that belong to a homogenous group (i.e., same education, same cohort, same region). Here we are computing these moments for individuals that have the same age, hence, that are born the same year.

(1) We use the (auto)covariance structure of growth rates, $Cov_{a,t}(\Delta u_{i,a,t}, \Delta u_{i,a+j,t+j})$, to identify our parameters. Specifically, our model implies:

(1a) If $j > 1$,

$$Cov_{a,t}(\Delta u_{i,a,t}, \Delta u_{i,a+j,t+j}) = 0.$$

(1b) If $j = 1$,

$$Cov_{a,t}(\Delta u_{i,a,t}, \Delta u_{i,a+1,t+1}) = \sigma^2_{\epsilon_{a,t}}.$$

That is, the covariance between the growth rate from $t-1$ to t and the growth rate from t to $t+1$ identifies $\sigma_{\epsilon_{a,t}}$. That is, to identify the variance of the transitory shock at t we need individual data for three consecutive periods: $t-1$, t, and $t+1$.

(1c) If $j = 0$,

$$Cov_{a,t}(\Delta u_{i,a,t}, \Delta u_{i,a,t}) = Var_{a,t}(\Delta u_{i,a,t}) = \sigma^2_{\eta_{a,t}} + \sigma^2_{\epsilon_{a,t}} + \sigma^2_{\epsilon_{a-1,t-1}}.$$

Given, $\sigma_{\epsilon_{a,t}}$ and $\sigma_{\epsilon_{a-1,t-1}}$, the variance of the growth rate from t to $t+1$ identifies the variance of the permanent shocks $\sigma^2_{\eta_{a,t}}$. That is, to identify the variance of the permanent shock at t we need individual data from four consecutive periods: $t-2$, $t-1$, t, and $t+1$.

(2) To initiate simulations of the income process, we need to identify the initial distribution (variance) of the permanent component $Var_{a-1,t-1}(z_{i,a-1,t-1})$ for all ages. Note that the variance of the level of initial

$$\begin{aligned} Var_{a,t}(u_{i,a,t}) &= Var_{a,t}(z_{i,a,t}) + \sigma^2_{\epsilon_{a,t}} \\ &= Var_{a-1,t-1}(z_{i,a-1,t-1}) + \sigma^2_{\eta_{a,t}} + \sigma^2_{\epsilon_{a,t}} \end{aligned} \qquad (4)$$

where $Var_{a-1,t-1}(z_{i,a-1,t-1})$ is the only unknown per age a; there is one equation (7) per age.

Example: If we have available data for $a = \{18, 66\}$ and for years $t = \{1998, 2011\}$, then we will be able to identify the model for $a = \{20, 65\}$ and $t = \{2000, 2010\}$:

(i) Using the autocovariance of growth rates,

$$Cov_{a,t}(\Delta u_{i,a,t}, \Delta u_{i,a+1,t+1}) = \sigma^2_{\epsilon_{a,t}},$$

we can identify $\{\{\sigma^2_{\epsilon_{a,t}}\}_{a=19,65}\}_{t=1999}^{t=2010}$. That is, we can identify all variances of the transitory shocks except for the first and last age, and the first and last period, for which the data are available.

(ii) Given these series of transitory shocks, we can use the variance of the growth rates,

$$Var_{a,t}(\Delta u_{i,a,t}) = \sigma^2_{\eta_{a,t}} + \sigma^2_{\epsilon_{a,t}} + \sigma^2_{\epsilon_{a-1,t-1}}.$$

to identify the variance of the permanent shocks for $\{\{\sigma^2_{\eta_{a,t}}\}_{a=20,65}\}_{t=2000}^{t=2010}$.

(iii) Finally, we can identify the initial distribution (variance) of the permanent component $Var_{a-1,t-1}(z_{i,a-1,t-1})$ for each age a as well; a necessary element to initiate the simulation of the PT model. Note that the variance of the initial level

$$\begin{aligned} Var_{a,t}(u_{i,a,t}) &= Var_{a,t}(z_{i,a,t}) + \sigma^2_{\epsilon_{a,t}} \\ &= Var_{a-1,t-1}(z_{i,a-1,t-1}) + \sigma^2_{\eta_{a,t}} + \sigma^2_{\epsilon_{a,t}} \end{aligned}$$

where $Var_{a-1,t-1}(z_{i,a-1,t-1})$ is the only unknown per age a.

- If we assume the variance of the PT model does not depend on age (or any other cross-sectional partition), that is, $\{\sigma^2_{\eta_t}, \sigma^2_{\epsilon_t}, \sigma^2_{z_0}\}_{t=0}^{T}$, we identify these variances by averaging them across all ages in the sample at each period t. For example, to estimate $\sigma^2_{\epsilon_t}$, we use the moment:

$$\sum_a Cov_{a,t}(\Delta u_{i,a,t}, \Delta u_{i,a+1,t+1}) = \sigma^2_{\epsilon_t}.$$

- If we assume the variance of the PT model does not depend on time, that is, $\{\sigma^2_{\eta_a}, \sigma^2_{\epsilon_a}, \sigma^2_{z_0}\}_{a=0}^{A}$, we identify these variances by averaging them across all periods in the sample at each age a. For example, to estimate $\sigma^2_{\epsilon_a}$, we use the moment:

$$\sum_t Cov_{a,t}(\Delta u_{i,a,t}, \Delta u_{i,a+1,t+1}) = \sigma^2_{\epsilon_a}.$$

Identification using log-level moments

(1a) We use the (auto)covariance structure of log income levels to identify our parameters. Note the relationship of income for any individual i between two consecutive periods, specifically, t and $t+1$ is

$$u_{i,a,t} = z_{i,a,t} + \epsilon_{i,a,t},$$

$$u_{i,a+1,t+1} = z_{i,a+1,t+1} + \epsilon_{i,a+1,t+1} = z_{i,a,t} + \eta_{i,a+1,t+1} + \epsilon_{i,a+1,t+1}$$

that is,

$$Cov_{a,t}(u_{i,a,t}, u_{i,a+1,t+1}) = Var_{a,t}(z_{i,a,t}).$$

Further note that

$$Var_{a,t}(u_{i,a,t}) = Var_{a,t}(z_{i,a,t}) + \sigma^2_{\epsilon_{a,t}}$$

Therefore,

$$Var_{a,t}(u_{i,a,t}) - Cov_{a,t}(u_{i,a,t}, u_{i,a+1,t+1}) = \sigma^2_{\epsilon_{a,t}}$$

That is, the variance of transitory shocks for any partition (here, age) at any period t, i.e., $\sigma^2_{\epsilon_{a,t}}$, can be identified using the variance of income of individuals within that partition at period t, and the covariance of income between period t and period $t+1$. That is, to identify the variance of transitory shocks we need data on two consecutive periods, t and $t+1$.

(1b) Note the relationship of income for any individual i between two consecutive periods, specifically, t and $t-1$ is

$$\begin{aligned} u_{i,a,t} &= z_{i,a,t} + \epsilon_{i,a,t} \\ &= z_{i,a-1,t-1} + \eta_{i,a,t} + \epsilon_{i,a,t}, \\ u_{i,a-1,t-1} &= z_{i,a-1,t-1} + \epsilon_{i,a-1,t-1} \end{aligned}$$

that is,

$$Cov_{a,t}(u_{i,a,t}, u_{i,a-1,t-1}) = Var_{a,t}(z_{i,a-1,t-1}).$$

Further note that

$$Var_{a,t}(u_{i,a,t}) = Var_{a,t}(z_{i,a-1,t-1}) + \sigma^2_{\eta_{a,t}} + \sigma^2_{\epsilon_{a,t}}$$

Therefore,

$$Var_{a,t}(u_{i,a,t}) - Cov_{a,t}(u_{i,a,t}, u_{i,a-1,t-1}) = \sigma^2_{\eta_{a,t}} + \sigma^2_{\epsilon_{a,t}}$$

That is, given $\sigma^2_{\epsilon_{a,t}}$ (for which we need data on t and $t+1$), the variance of permanent shocks for any partition (here, age) at any period t, i.e., $\sigma^2_{\eta_{a,t}}$, can be identified using the variance of income of individuals within that partition at period t, and the covariance of income of that partition between period t and period $t-1$. That is, to identify the variance of permanent shocks we need data on three consecutive periods, $t-1$, t, and $t+1$.

(2) To initiate simulations of the income process, we need to identify the initial distribution (variance) of the permanent component $Var_{a-1,t-1}(z_{i,a-1,t-1})$ for all ages. Note that the variance of the level of initial

$$Var_{a,t}(u_{i,a,t}) = Var_{a,t}(z_{i,a,t}) + \sigma^2_{\epsilon_{a,t}}$$
$$= Var_{a-1,t-1}(z_{i,a-1,t-1}) + \sigma^2_{\eta_{a,t}} + \sigma^2_{\epsilon_{a,t}} \quad (5)$$

where $Var_{a-1,t-1}(z_{i,a-1,t-1})$ is the only unknown per age a; there is one equation (7) per age.

Example: If we have available data for $a = \{18, 66\}$ and for years $t = \{1998, 2011\}$, then we will be able to identify the model for $a = \{20, 65\}$ and $t = \{2000, 2010\}$:

(i) Using the covariance of log income levels between t and $t+1$,

$$Var_{a,t}(u_{i,a,t}) - Cov_{a,t}(u_{i,a,t}, u_{i,a+1,t+1}) = \sigma^2_{\epsilon_{a,t}},$$

we can identify $\{\{\sigma^2_{\epsilon_{a,t}}\}_{a=18}^{a=65}\}_{t=1998}^{t=2010}$. That is, we can identify all variances of the transitory shocks except for the first and last age, and the first and last period, for which the data are available.

(ii) Given these series of transitory shocks, we can use the covariance of log income levels between t and $t-1$,

$$Var_{a,t}(u_{i,a,t}) - Cov_{a,t}(u_{i,a,t}, u_{i,a-1,t-1}) = \sigma^2_{\eta_{a,t}} + \sigma^2_{\epsilon_{a,t}},$$

to identify the variance of the permanent shocks for $\{\{\sigma^2_{\eta_{a,t}}\}_{a=19}^{a=65}\}_{t=2000}^{t=2010}$.

(iii) Finally, we can identify the initial distribution (variance) of the permanent component $Var_{a-1,t-1}(z_{i,a-1,t-1})$ for each age a as well; a necessary element to initiate the simulation of the PT model. Note that the variance of the initial level

$$\begin{aligned} Var_{a,t}(u_{i,a,t}) &= Var_{a,t}(z_{i,a,t}) + \sigma^2_{\epsilon_{a,t}} \\ &= Var_{a-1,t-1}(z_{i,a-1,t-1}) + \sigma^2_{\eta_{a,t}} + \sigma^2_{\epsilon_{a,t}} \end{aligned}$$

where $Var_{a-1,t-1}(z_{i,a-1,t-1})$ is the only unknown per age a.

- If we assume the variance of the PT model does not depend on age (or any other cross-sectional partition), that is, $\{\sigma^2_{\eta_t}, \sigma^2_{\epsilon_t}, \sigma^2_{z_0}\}_{t=0}^{T}$, we identify these variances by averaging them across all ages in the sample at each period t. For example, to estimate $\sigma^2_{\epsilon_t}$, we use the moment:

$$\sum_a \left(Var_{a,t}(u_{i,a,t}) - Cov_{a,t}(u_{i,a,t}, u_{i,a+1,t+1}) \right) = \sigma^2_{\epsilon_t}.$$

- If we assume the variance of the PT model does not depend on time, that is, $\{\sigma^2_{\eta_a}, \sigma^2_{\epsilon_a}, \sigma^2_{z_0}\}_{a=0}^{A}$, we identify these variances by averaging them across all periods in the sample at each age a. For example, to estimate $\sigma^2_{\epsilon_a}$, we use the moment:

$$\sum_t \left(Var_{a,t}(u_{i,a,t}) - Cov_{a,t}(u_{i,a,t}, u_{i,a+1,t+1}) \right) = \sigma^2_{\epsilon_a}.$$

Identification using log-level moments (unbalanced sample)

Let's assume that while we want to estimate an annual process, data are only available biannually.

This is a typical case of an unbalanced panel data set.

(1a) We use the (auto)covariance structure of log income levels. For any individual i between two consecutive periods, specifically, t and $t+2$:

$$\begin{aligned} u_{i,a,t} &= z_{i,a,t} + \epsilon_{i,a,t}, \\ u_{i,a+2,t+2} &= z_{i,a+2,t+2} + \epsilon_{i,a+2,t+2} \\ &= z_{i,a+1,t+1} + \eta_{i,a+2,t+2} + \epsilon_{i,a+2,t+2} \\ &= z_{i,a,t} + \eta_{i,a+1,t+1} + \eta_{i,a+2,t+2} + \epsilon_{i,a+2,t+2} \end{aligned}$$

that is,

$$Cov_{a,t}(u_{i,a,t}, u_{i,a+2,t+2}) = Var_{a,t}(z_{i,a,t}).$$

Further note that

$$Var_{a,t}(u_{i,a,t}) = Var_{a,t}(z_{i,a,t}) + \sigma^2_{\epsilon_{a,t}}$$

Therefore,

$$Var_{a,t}(u_{i,a,t}) - Cov_{a,t}(u_{i,a,t}, u_{i,a+2,t+2}) = \sigma^2_{\epsilon_{a,t}}$$

That is, the variance of transitory shocks for any partition (here, age) at any period t, i.e., $\sigma^2_{\epsilon_{a,t}}$, is identified using the variance of income of individuals in that partition at period t, and the covariance of income between period t and period $t+2$. That is, to identify the variance of transitory shocks we need data on two consecutive periods (of available data), t and $t+2$.

(1b) For any individual i, residual income at two consecutive periods t and $t-2$ is

$$
\begin{aligned}
u_{i,a,t} &= z_{i,a,t} + \epsilon_{i,a,t} = z_{i,a-1,t-1} + \eta_{i,a,t} + \epsilon_{i,a,t}, \\
&= z_{i,a-2,t-2} + \eta_{i,a-1,t-1} + \eta_{i,a,t} + \epsilon_{i,a,t}, \\
u_{i,a-2,t-2} &= z_{i,a-2,t-2} + \epsilon_{i,a-2,t-2}
\end{aligned}
$$

that is,

$$Cov_{a,t}(u_{i,a,t}, u_{i,a-2,t-2}) = Var_{a,t}(z_{i,a-2,t-2}).$$

Further note that

$$
\begin{aligned}
Var_{a,t}(u_{i,a,t}) &= Var_{a,t}(z_{i,a-2,t-2}) + \sigma^2_{\eta_{a-1,t-1}} + \sigma^2_{\eta_{a,t}} + \sigma^2_{\epsilon_{a,t}} \\
&= Var_{a,t}(z_{i,a-2,t-2}) + \sigma^2_{\tilde{\eta}_{a,t}} + \sigma^2_{\epsilon_{a,t}}
\end{aligned}
$$

where I have defined $\sigma^2_{\tilde{\eta}_{a,t}} = \sigma^2_{\eta_{a-1,t-1}} + \sigma^2_{\eta_{a,t}}$. Therefore,

$$Var_{a,t}(u_{i,a,t}) - Cov_{a,t}(u_{i,a,t}, u_{i,a-2,t-2}) = \sigma^2_{\tilde{\eta}_{a,t}} + \sigma^2_{\epsilon_{a,t}}.$$

That is, given $\sigma^2_{\epsilon_{a,t}}$ (for which we need data on t and $t+2$), the variance of permanent shocks beween period $t-2$ and t (i.e., $\sigma^2_{\tilde{\eta}_{a,t}} = \sigma^2_{\eta_{a-1,t-1}} + \sigma^2_{\eta_{a,t}}$) for any partition (here, age) can be identified using the variance of income of individuals within that partition at period t, and the covariance of income between period t and period $t-2$. To identify the variance of permanent shocks we need data on three consecutive periods (of available data), $t-2$, t, and $t+2$.

(2) To initiate simulations of the income process, we need to identify the initial distribution (variance) of the permanent component $Var_{a-2,t-2}(z_{i,a-2,t-2})$ for all ages. Note that the variance of the level of initial

$$
\begin{aligned}
Var_{a,t}(u_{i,a,t}) &= Var_{a,t}(z_{i,a,t}) + \sigma^2_{\epsilon_{a,t}} \\
&= Var_{a-1,t-1}(z_{i,a-1,t-1}) + \sigma^2_{\eta_{a,t}} + \sigma^2_{\epsilon_{a,t}} \qquad (6) \\
&= Var_{a-2,t-2}(z_{i,a-2,t-2}) + \sigma^2_{\widetilde{\eta}_{a,t}} + \sigma^2_{\epsilon_{a,t}} \qquad (7)
\end{aligned}
$$

where $Var_{a-2,t-2}(z_{i,a-2,t-2})$ is the only unknown per age a; there is one equation (7) per age.

Example: If we have available data for $a = \{18, 66\}$ and for years $t = \{2006, 2008, 2010, 2012\}$, then we will be able to identify the model for $a = \{20, 64\}$ and $t = \{2008, 2010\}$:

(i) Using the covariance of log income levels between t and $t+2$,

$$Var_{a,t}(u_{i,a,t}) - Cov_{a,t}(u_{i,a,t}, u_{i,a+2,t+2}) = \sigma^2_{\epsilon_{a,t}},$$

we can identify $\{\sigma^2_{\epsilon_{a,2006}}, \sigma^2_{\epsilon_{a,2008}}, \sigma^2_{\epsilon_{a,2010}}\}_{a=18}^{a=64}$. That is, we can identify all variances of the transitory shocks the last age groups and for the last period for which the data are available.

(ii) Given these series of transitory shocks, we can use the covariance of log income levels between t and $t-2$,

$$Var_{a,t}(u_{i,a,t}) - Cov_{a-2,t-2}(u_{i,a,t}, u_{i,a-2,t-2}) = \sigma^2_{\tilde{\eta}_{a,t}} + \sigma^2_{\epsilon_{a,t}},$$

to identify the variance of the permanent shocks for $\{\sigma^2_{\tilde{\eta}_{a,2008}}, \sigma^2_{\tilde{\eta}_{a,2010}}\}_{a=20}^{a=64}$. When we report the variance of the permanent shock we do so by, following the practice in **?**, computing $.5(\sigma^2_{\eta,a} + \sigma^2_{\eta,a-1})$.

(iii) Finally, we can identify the initial distribution (variance) of the permanent component $Var_{a-2,t-2}(z_{i,a-2,t-2})$ for each age a as well; a necessary element to initiate the simulation of the PT model. Note that the variance of the initial level

$$\begin{aligned} Var_{a,t}(u_{i,a,t}) &= Var_{a,t}(z_{i,a,t}) + \sigma^2_{\epsilon_{a,t}} \\ &= Var_{a-2,t-2}(z_{i,a-2,t-2}) + \sigma^2_{\tilde{\eta}_{a,t}} + \sigma^2_{\epsilon_{a,t}} \end{aligned}$$

where $Var_{a-2,t-2}(z_{i,a-2,t-2})$ is the only unknown per age a.

- If we assume the variance of the PT model does not depend on age (or any other cross-sectional partition), that is, $\{\sigma^2_{\eta_t}, \sigma^2_{\epsilon_t}, \sigma^2_{z_0}\}_{t=0}^{T}$, we identify these variances by averaging them across all ages in the sample at each period t. For example, to estimate $\sigma^2_{\epsilon_t}$, we use the moment:

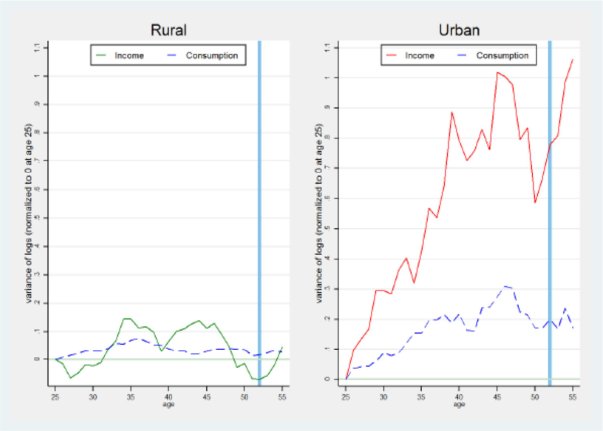
$$\sum_a \left(Var_{a,t}(u_{i,a,t}) - Cov_{a,t}(u_{i,a,t}, u_{i,a+2,t+2})\right) = \sigma^2_{\epsilon_t}.$$

- If we assume the variance of the PT model does not depend on time, that is, $\{\sigma^2_{\eta_a}, \sigma^2_{\epsilon_a}, \sigma^2_{z_0}\}_{a=0}^{A}$, we identify these variances by averaging them across all periods in the sample at each age a. For example, to estimate $\sigma^2_{\epsilon_a}$, we use the moment:

$$\sum_t \left(Var_{a,t}(u_{i,a,t}) - Cov_{a,t}(u_{i,a,t}, u_{i,a+2,t+2})\right) = \sigma^2_{\epsilon_a}.$$

Example: Residual lifecycle inequality in Poor Africa

More Insurance in Rural areas than in Urban areas



Taken from de Magalhaes et al (2016).

- The finding of increase income inequality and the fact that most of this increase is due to transitory shocks seems robust in rich countries (special 2010 RED issue) and in developing countries such as China (Santaeulàlia-Llopis and Zheng, 2015)

- Note if the increase in income inequality is mostly due to transitory shocks (e.g., increase in the instability of earnings) rather than permanent shocks (e.g., changes in the wage structure), then, through the lenses of a PILCH model consumption inequality should not increase as much as income inequality as consumers respond less to transitory shocks.

- One reason why one could observe an increase of income inequality that, although is due to transitory shocks, generates a larger increase in cosumption inequality can be explained largely by credit market imperfections that produce excess sensivity to the transitory shocks.

Permanent Income Risk in Growing China (SLZ, 2015)

		Disposable Income		Earnings + Private Transf.		Earnings + Public Transf.		Earnings Only	
		Rural	Urban	Rural	Urban	Rural	Urban	Rural	Urban
σ^2_ζ	1992-3	.2238	.0774	.2668	.1251	.2091	.0833	.2608	.0780
		(.0315)	(.0293)	(.0327)	(.0324)	(.0317)	(.0296)	(.0325)	(.0272)
	1994-7	.2620	.2456	.2777	.3331	.2103	.2610	.2812	.1616
		(.0529)	(.0646)	(.0535)	(.0846)	(.0498)	(.0708)	(.0521)	(.0487)
	1998-00	.2381	.2055	.2749	.1944	.2032	.1859	.2800	.0723
		(.0525)	(.0701)	(.0553)	(.0821)	(.0520)	(.0857)	(.0525)	(.0508)
	2001-4	.3762	.2501	.4200	.6870	.3881	.3565	.4027	.2063
		(.0549)	(.0614)	(.0585)	(.1151)	(.0494)	(.0807)	(.0483)	(.0550)
	2005-6	.3160	.2432	.3810	.5521	.2718	.1235	.2391	.1316
		(.0403)	(.0452)	(.0399)	(.0766)	(.0354)	(.0393)	(.0353)	(.0438)
$\sigma^2_{\zeta,pre97}$	(ann.)	.0913	.0488	.1050	.0718	.0814	.0521	.1037	.0396
$\sigma^2_{\zeta,post97}$	(ann.)	.1170	.0887	.1373	.1888	.1059	.0709	.1067	.0506
Obs		3,560	1,825	3,559	1,820	3,549	1,820	3,546	1,814

Notes: This table shows the estimation results from a partial insurance model where different measures of income are used, for rural and urban areas separately. The σ^2_ζ refers to the sum of the variances of permanent shocks over the period specified in the second column of the table. The σ^2_ε refers to the variance of transitory shocks in the year specified in the second column. We also provide the annualized variance of permanent (transitory) shocks over the pre-1997 period, $\sigma^2_{\zeta,pre97}$ ($\sigma^2_{\varepsilon,pre97}$), and that over the post-1997 period, $\sigma^2_{\zeta,post97}$ ($\sigma^2_{\varepsilon,post97}$). Asymptotic standard errors are in the parentheses.

Transitory Income Risk in Growing China (SLZ, 2015)

		Disposable Income		Earnings + Private Transf.		Earnings + Public Transf.		Earnings Only	
		Rural	Urban	Rural	Urban	Rural	Urban	Rural	Urban
σ_ε^2	1991	.2835	.1267	.2922	.1427	.2870	.1309	.2998	.1520
		(.0193)	(.0164)	(.0201)	(.0197)	(.0197)	(.0166)	(.0200)	(.0178)
	1993	.3874	.2581	.3511	.2103	.3974	.2564	.3544	.1981
		(.0320)	(.0302)	(.0324)	(.0332)	(.0323)	(.0310)	(.0326)	(.0276)
	1997	.4626	.2258	.4774	.3300	.4578	.2171	.4652	.2094
		(.0435)	(.0503)	(.0447)	(.0713)	(.0425)	(.0557)	(.0418)	(.0475)
	2000	.5274	.3062	.5177	.3157	.4758	.3683	.3829	.2788
		(.0460)	(.0511)	(.0465)	(.0546)	(.0431)	(.0580)	(.0355)	(.0493)
	2004	.4660	.3024	.4160	.2652	.3993	.3241	.4037	.2721
		(.0352)	(.0422)	(.0342)	(.0541)	(.0337)	(.0416)	(.0329)	(.0389)
	2006	.3917	.2219	.3658	.2478	.3800	.2576	.3795	.2412
		(.0318)	(.0369)	(.0311)	(.0511)	(.0287)	(.0337)	(.0308)	(.0426)
$\sigma^2_{\varepsilon,pre97}$	(ann.)	.3862	.1999	.3876	.2410	.3875	.1971	.3854	.1878
$\sigma^2_{\varepsilon,post97}$	(ann.)	.4504	.2704	.4205	.2710	.4103	.3076	.3884	.2608
Obs		3,560	1,825	3,559	1,820	3,549	1,820	3,546	1,814

Notes: This table shows the estimation results from a partial insurance model where different measures of income are used, for rural and urban areas separately. The σ^2_ζ refers to the sum of the variances of permanent shocks over the period specified in the second column of the table. The σ^2_ε refers to the variance of transitory shocks in the year specified in the second column. We also provide the annualized variance of permanent (transitory) shocks over the pre-1997 period, $\sigma^2_{\zeta,pre97}$ ($\sigma^2_{\varepsilon,pre97}$), and that over the post-1997 period, $\sigma^2_{\zeta,post97}$ ($\sigma^2_{\varepsilon,post97}$). Asymptotic standard errors are in the parentheses.

Misspecification Issues

- The PT model identified using growth rates of income is very different from the one identified using log income levels. While in both models the variance of transitory shocks is larger than the variance of permanent shocks, the variance of permanent shocks is about 3 times larger in growth rates than in levels, and the variance of transitory shocks is about 1.5 to 2 times larger using levels than growth rates.

  That is, identification of the PT model through growth rates delivers larger permanent shocks and, accordingly, the identification in levels delivers larger transitory shocks.[8]

- This disagreement in results suggests that the PT model is misspecified. The problem is that PT model cannot simultaneously replicate moments of the income (wages for Heathcote et al. (2010)) distribution in levels and in growth rates.

- This misspecification carries potentially large quantitative implications. For instance, the variance of permanent wage shocks is a key determinant of the size of the welfare costs of incomplete insurance against idiosyncratic risk (incomplete markets), and hence potential welfare gains from social insurance policies.

[8] When Heathcote et a. (2010) trim the top and bottom 3% of the distribution of log wage differences, they find a variance of permanent shocks roughly similar to that in levels. However, the variance of the transtiory shock is less than 1/3 smaller than its counterpart in levels. It is unclear for the authors whether the trimming eliminates genuine wage varaition or spurious outliers.

- One “reality check” proposed by Heathcote et al. (2010) for any variance estimate is to explore what implies for the growth rate of wage inequality over the life cycle.

  i. On average, the estimated variance of permanent shocks identified through growth rates is .027, which implies a rise in the variance of log wages of .94 over the 35 years (60-25). However, the observed increase in the variance of wages is of .20 (.35) when controlling for time (cohort) effects. The the variance of permanent shocks estimated with growth rates overshoot the actual variance of wages.
  ii. On average, the estimated variance of permanent shocks identified through log levels is .007, which implies a rise in the variance of log wages of .25 over the 35 years (60-25). A much more similar figure to the observed data.

  This, in principle, would favor levels against growth rates.

- However, the variance of permanent shocks identified through level moments is negative for some years, which suggests misspecification.

The case of superior information

- It is reasonable to think that consumers may know more than the econometrician; the case of superior information. For example, individuals may have information about events such as a promotion (or the opposite) that the econometrician may never hope to predict.

- One line of research finds useful to compare measures of uncertainty obtained via estimation of dynamic income processes with measures of risk recovered from subjective expectations data (see Dominitz and Manski (1998) and Barsky et al. (1997)).

Example of Trouble: Selection Bias

- **The identification problem:** If individuals leave the market because of a sudden wage drop, such as from job loss, then wage growth rates for workers (agents that remain working) will be greater than wage growth for non-workers. This problem will bias wage growth upward. At the same time, this fact will underestimate wage inequality.

- Working example: French (2005) and Olivetti and Petrongolo (2009).

- **Two remarks.**

  Panel or Cross-section? By keeping track of individuals, panel data can help us understand the selection problem better. There is an important advantage over cross-sectional data: cross-sectional estimators, such as OLS, mix the true wage growth of individuals with spurious wage growth caused by differences in the level of wages between those who enter, exit, and remain in the labor force.

  Do fixed-effects solve the selection-bias problem? NO. Fixed-effects use wage observations for workers but do not use the potential wages of non-workers. The fixed-effects estimator demeans the average level of wages for each individual in the sample and, this way, identifies the growth rate of wages of individuals while working. Because the fixed-effects estimator identifies the individual-level growth rates of workers wages, composition bias problems–i.e., the question of whether high wage or low wage individuals drop out of the labor-market–is not a problem if wage growth rates for workers and non-workers are the same... but there is no reason for this to be the case.

- **Using a structural model to solve for selection**

  Many authors have studied the extensive margin of labor supply, the decision of whether to work or not. Labor search and matching models explicitly deal with this margin. Richard Rogerson and his many coauthors have also investigated the labor supply extensively (see also Cho and Cooley (1994) and Osuna and Rios-Rull (2004)). Further, Chiappori, Blundell, and Meghir use collective models of labor supply that also consider the behavior of spouses.

Following French (2005), let's assume we take one of these models of labor supply, and to correct the selection bias we further assume that the bias in the fixed-effects wage profiles of workers will be the same in both the actual PSID and simulated data from the model. In particular, follow this iterative process:

- First, feed the estimated (and biased) fixed-effects wage profile into the model.
- Second, solve and simulate the model and estimate the fixed-effects wage profiles for both simulated workers and all simulated individuals.
- Third, compute the difference between the profiles for both simulated workers and all simulated individuals so that we can estimate the extent to which growth rates in wages are overestimated by using only simulated workers instead of all simulated individuals.
- Then use this estimate of the selection bias in the simulated wage profile to infer the extent of selection bias in the PSID data wage profile. [9]
- This iterative process is continued until a fixed point is found. Once the process converges, the estimated wage profile for all individuals is fed into the model and preference parameters are estimated using the method of simulated moments. Upon re-estimation of the model parameters, the selection bias is recomputed and the wage profiles are updated. The model parameters are then estimated again.