Micro Data: Some Sources and Tools

Quantitative Macroeconomics [Econ 5725]

Raül Santaeulàlia-Llopis

Washington University in St. Louis

Spring 2016

Micro Data

- We next turn to discuss some cross-sectional and panel data facts using individual and household-level sources of data.
- Then, we discuss the estimation of the labor income process decomposing permanent and transitory components. We also study two standard problems that typically arise in these estimations:

  1. The identification of age/cohort/time effects and,
  2. Selection issues.

Current Population Survey (CPS)

- Primary source for labor force statistics in the U.S.
- About 50,000 households are interviewed monthly based on their areas of residence and represent the U.S. household as a whole, individual states and othe specified areas.
- Eight rotatiton groups are interviewed a total of 8 months. Groups are rotated consecutively for 4 months, and for 4 consecutive months again after resting 8 months. That is, if they are interviewed Jan-April 2011, they will be interviewed again Jan-April 2012.

- A typical unit of observation is individuals within households; however the March series also has family and household observations (the March Supplement).

- It collects data on several dimensions:

  - Employment for the week prior to the survey: employment status, occupation, industry of adults.
  - Demographic information: age, sex, race, marital status, family structure.
  - Educational attainment.
  - Periodically we also have data on health, income, previous work experience...

- The CPS sample attempts to reprsent the civilian, noninstitutional popluation of the U.S. by using a probability sample to select housing units.

- The large sample allows for accurate analyses at a high degree of disaggregation That is, we can consider partitions by education, age, sex, race, etc...

Figure: **Skill Wage Premium and Relative Quantity of Skills, U.S. 1976-2010 (College vs Non College)**

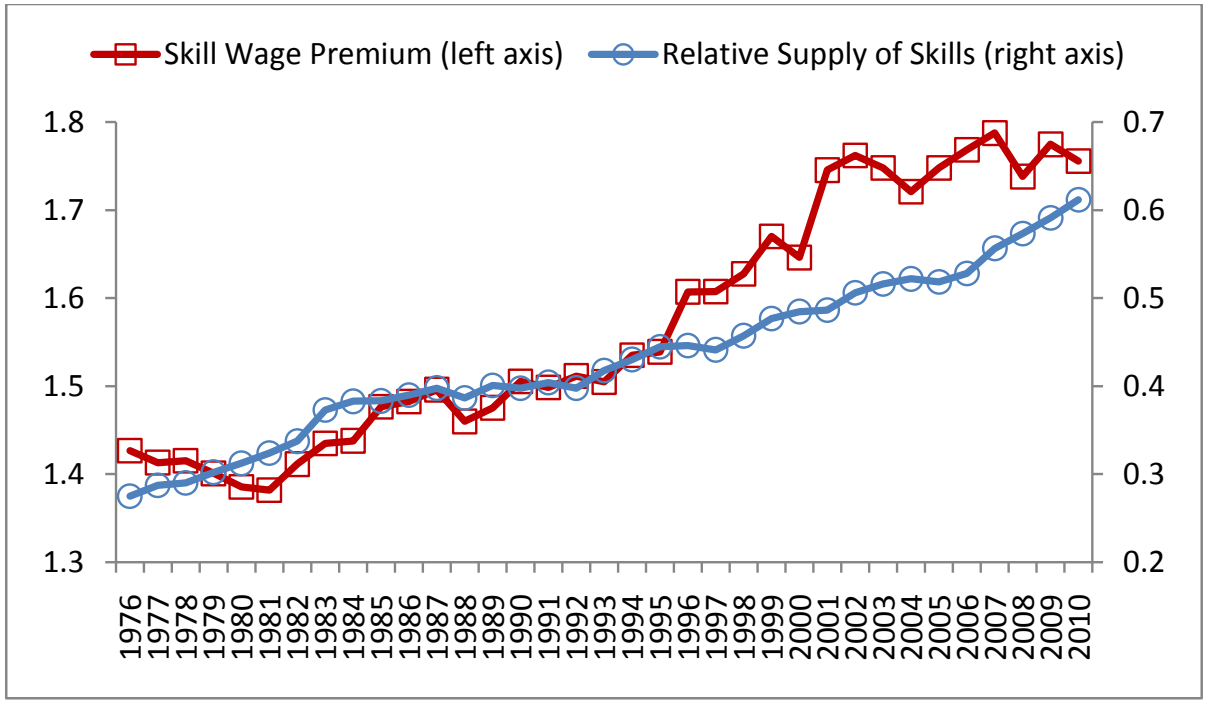
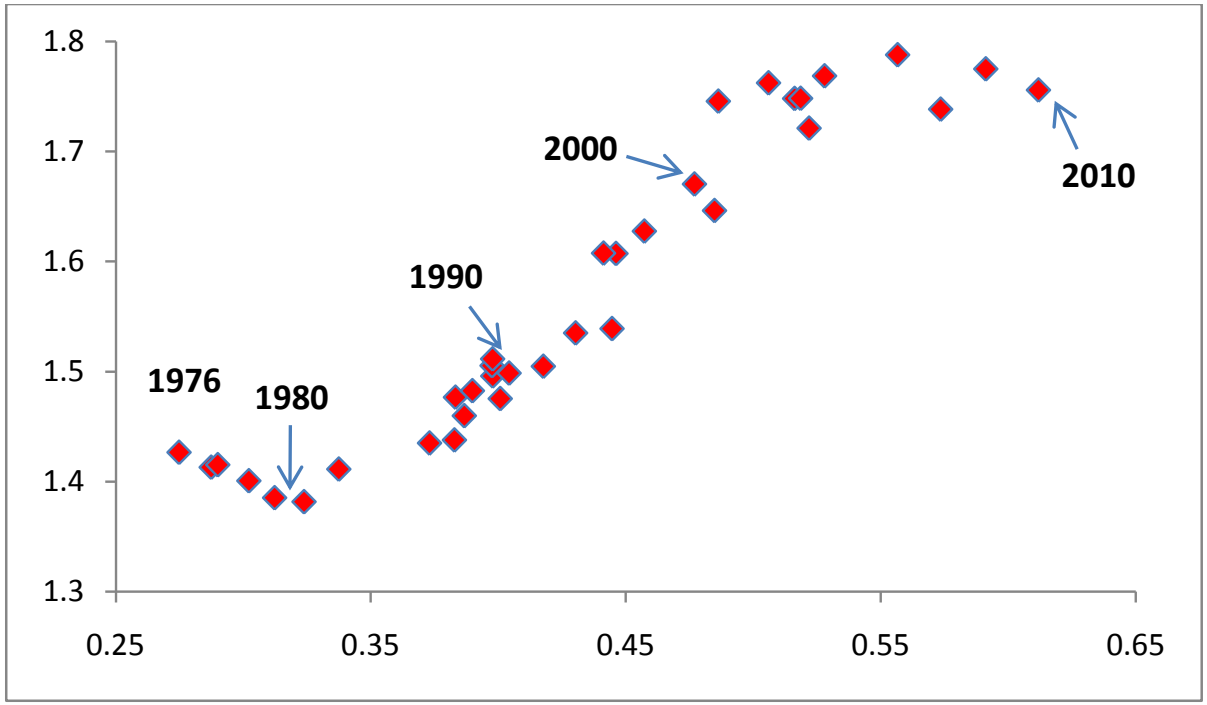


Figure: **Scatterplot: Skill Wage Premium (vertical axis) and Relative Quantity of Skills(Horizontal Axis)**



What accounts for the rise of the skill premium?

- A prominent explanation for the rise of wage inequality is skilled biased technical change (SBTC), see Katz and Murphy (1992), Katz and Autor (1999), Acemoglu (2002), Hornstein et al. (2005) and Acemoglu and Autor (2011).
- We can associate SBTC to a demand shift towards information and communications technology.
- One natural mechanism behind SBTC is the skilled labor-capital equipment complementarity formalized by Krusell et al. (2000).
- Alternative mechanisms introduce the role of deunionization as in Acemoglu et al. (2001), and trade-induced skilled biased technical change as in Bloom et al. (2010).

Survey Consumer Finances (SCF)

- This is a representative U.S. household survey of the balance sheet, pension, income and other demographic variables.
- The survey oversamples the rich. This is very convenient, as it is this group that holds most of the wealth in the U.S.
- The data are cross-sectional, and collected every three years since 1983 to 2010. There was also an initial panel attempt in 1961-1962.
- It also provides information on the use of financial institutions.

- Working example: Budria et al. (2002) and the update Díaz-Giménez et al. (2011).

- They document the distributional properties of earnings, income, and wealth.

- They do so by several partitions: educational attainment, employment status groups and by household composition.

- A large set of the incomplete markets literature seeks to replicate those facts to do policy. See, for instance, Aiyagari (1994), Huggett (1996), Castañeda et al. (2003), Quadrini (1999), and De Nardi (2004). A goal of this course is to learn how to solve these type of models.

Figure: U.S. Economic Inequality: Distribution of Income (Gini=.57)

	Bottom (%)			Quintiles					Top (%)			All
	0–1	1–5	5–10	1st	2nd	3rd	4th	5th	90–95	95–99	99–100	0–100
						Averages (x 10^3 2007 USD)						
Earnings	0.0	1.9	2.9	4.2	18.2	36.4	64.6	195.6	144.6	264.6	1,111	63.8
Income	−7.6	7.0	10.5	11.7	28.2	47.1	76.6	254.4	169.6	330.8	1,753	83.6
Wealth	490	82.0	53.2	102.8	139.4	211.3	377.3	1,946	1,195	3,174	14,407	555.4
						Shares of Total Sample (%)						
Earnings	0.0	0.1	0.2	1.3	5.7	11.4	20.3	61.3	11.3	16.6	17.4	100.0
Income	−0.1	0.3	0.6	2.8	6.7	11.3	18.3	60.9	10.2	15.9	21.0	100.0
Wealth	0.9	0.6	0.5	3.7	5.0	7.6	13.6	70.1	10.8	22.9	25.9	100.0
						Income Sources (%)						
Labor	32.2	25.2	26.1	35.6	60.8	72.6	77.8	60.5	75.8	60.6	39.0	64.3
Capital	−105.5	1.5	1.0	−1.9	1.8	1.9	2.9	15.5	7.7	11.9	30.4	10.3
Business	−37.1	1.5	1.3	0.7	4.6	5.5	7.6	19.0	10.9	22.5	28.3	13.9
Transfers	9.9	66.0	64.2	59.9	30.3	18.4	10.5	4.3	4.7	4.0	2.1	10.3
Other	−0.5	5.9	7.4	5.7	2.5	1.6	1.2	0.7	0.9	1.1	0.3	1.2

Source: Díaz-Giménez et al. (2011)

Figure: U.S. Economic Inequality: Distribution of Wealth (Gini=.81)

	Bottom (%)			Quintiles					Top (%)			All
	0–1	1–5	5–10	1st	2nd	3rd	4th	5th	90–95	95–99	99–100	0–100
						Averages (x 10^3 2007 USD)						
Earnings	35.5	31.9	15.7	22.1	34.4	47.4	62.0	153.2	104.6	254.1	764.3	63.8
Income	38.4	37.8	21.8	27.5	40.5	56.5	74.2	219.2	137.9	347.6	1,323	83.6
Wealth	−79.0	−13.6	−0.9	−5.3	29.7	123.6	312.3	2,316	1,233	3,710	18,653	555.4
						Shares of Total Sample (%)						
Earnings	0.6	2.0	1.2	6.9	10.8	14.9	19.4	48.0	8.2	15.9	12.0	100.0
Income	0.5	1.8	1.3	6.6	9.7	13.5	17.8	52.5	8.3	16.6	15.8	100.0
Wealth	−0.1	−0.1	−0.0	−0.2	1.1	4.5	11.2	83.4	11.1	26.7	33.6	100.0
						Income Sources (%)						
Labor	85.6	83.5	72.4	78.9	81.2	78.6	77.1	51.4	58.6	54.7	30.2	64.3
Capital	0.0	−0.0	0.0	0.1	0.5	1.0	2.7	18.3	7.9	17.8	33.7	10.3
Business	8.1	1.2	−0.3	1.9	4.2	6.2	7.5	21.4	20.1	21.4	32.0	13.9
Transfers	3.7	12.1	22.3	15.5	12.0	12.4	12.1	8.2	12.6	5.5	3.6	10.3
Other	2.7	3.3	5.5	3.7	2.0	1.8	0.7	0.7	0.9	0.7	0.6	1.2

Source: Díaz-Giménez et al. (2011)

	Micro Data				Macro Data
	Income		Wealth		Income (p.c.)
Countries	Top 1%	Top 10%	Top 1%	Top 10%	WDI
Rich:					
US	20%	48%	34%	71%	48,377
Britain	15%	42%	28%	70%	38,363
France	9%	33%	24%	62%	40,706
Sweden	7%	28%	20%	59%	52,076
Emerging:					
India	12%	*n.a.*	*n.a.*	*n.a.*	1,417
Indonesia	13%	*n.a.*	*n.a.*	*n.a.*	2,946
China	11%	*n.a.*	*n.a.*	*n.a.*	4,433
South Africa	17%	*n.a.*	*n.a.*	*n.a.*	7,175
Argentina	17%	*n.a.*	*n.a.*	*n.a.*	11,460
Colombia	20%	*n.a.*	*n.a.*	*n.a.*	6,179
SSA:					
Malawi:					359
Rural	14%	43%	17%	49%	–
Urban	25%	62%	32%	73%	–

Note: The figures for rich and emerging countries are retrieved from Piketty (2014). All numbers refer to 2010. SSA figures are retrieved from deMagalhaes and Santaeulàlia-Llopis (2015).

What is **wealth**? We measure net worth as assets minus debt.

One taxonomy for assets and debt is as follows:

- Assets:
  - Financial: Liquid, Bonds, Stocks and Miscellanous.
  - Non-Financial: Privately Held Business and Durables

- Debt:
  - Collateralized Debt
  - Other Debt

Wealth I: A Taxonomy of Assets

Financial Assets	
Liquid	Checking Accounts Savings Accounts Money Market Accounts Brokerage Accounts Certificates of Deposit
Bonds	U.S. Government Saving Bonds Mortgage-Backed Bonds Treasury Bills State, Municipal or other tax free Bonds Foreign Bonds Corporate Bonds Other Saving Bonds
Stocks	Stocks Publicly Traded Corporations, Mutual Funds Life Insurance Retirement Accounts Job Pensions Other Managed Assets: Annuities and Trusts
Miscellaneous Assets	
Non Financial Assets	
Privately Held Business	Actively Managed Business Non Actively Managed Business
Durables	Principal Residence Investment Real Estate Vehicles

Wealth II: A Taxonomy of Debt

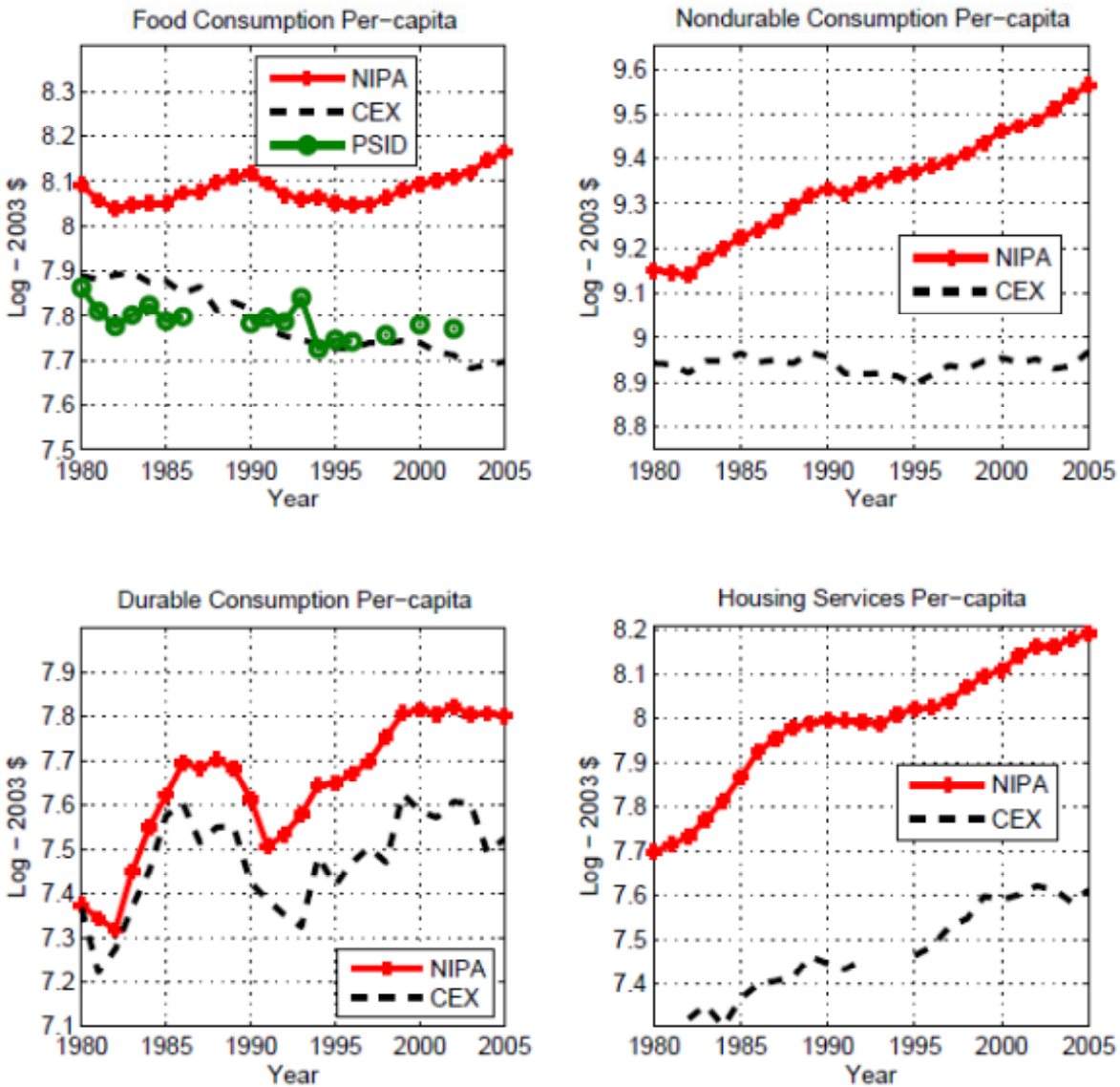
Debt	
Collateralized Debt	Mortgages, Home Equity Loans Home Equity Lines of Credit Debt on Investment Real Estate Debt on Vehicles
Other Debt	Other Loans to Purchase Principal Residence Debt on Life Insurance Lines of Credit Credit Card Home Improvement, Student Loans, Other Consumer Loans

Notes: Retirement Accounts, Job Pensions, Annuities and Trusts are also partly invested to liquid and/or bonds.

Consumer Expenditure Survey (CEX)

- This is a representative panel of cross-sectional data of U.S. households. These data are used to construct the CPI.
- Data are collected on many items of consumption (food and other nondurables, durables, housing services, etc.). About 90% to 95% of total consumption expenditures are included.
- It also provides measures of income. These data set has been extensively used to study the response of consumption to income shocks.
- Working example: Krüger and Perri (2006) and Heathcote et al. (2010).

Figure: Evolution of consumption per capita over time



Source: Heathcote, Perri, and Violante (2010)

Savings Rate as Income Increases

	Saving Rates by Income deciles									
	1	2	3	4	5	6	7	8	9	10
U.S. CEX	-1.3	-0.63	-0.66	-0.12	0.00	0.05	0.10	0.19	0.26	0.36
Malawi Rural	-5.8	-2.49	-1.61	-1.25	-1.01	-0.75	-0.46	-0.30	-0.06	0.43
Malawi Urban	-19.7	-4.42	-2.53	-1.73	-1.39	-1.09	-0.99	-0.31	-0.23	0.47

The average savings rate is 0.12 for the US and -0.23 and -0.15 for respectively rural and urban Malawi.
For the US we use Sabelhaus and Groen (2000) and for Malawi we use deMagalhaes and Santaeulàlia-Llopis (2015)

Earnings Inequality over Time

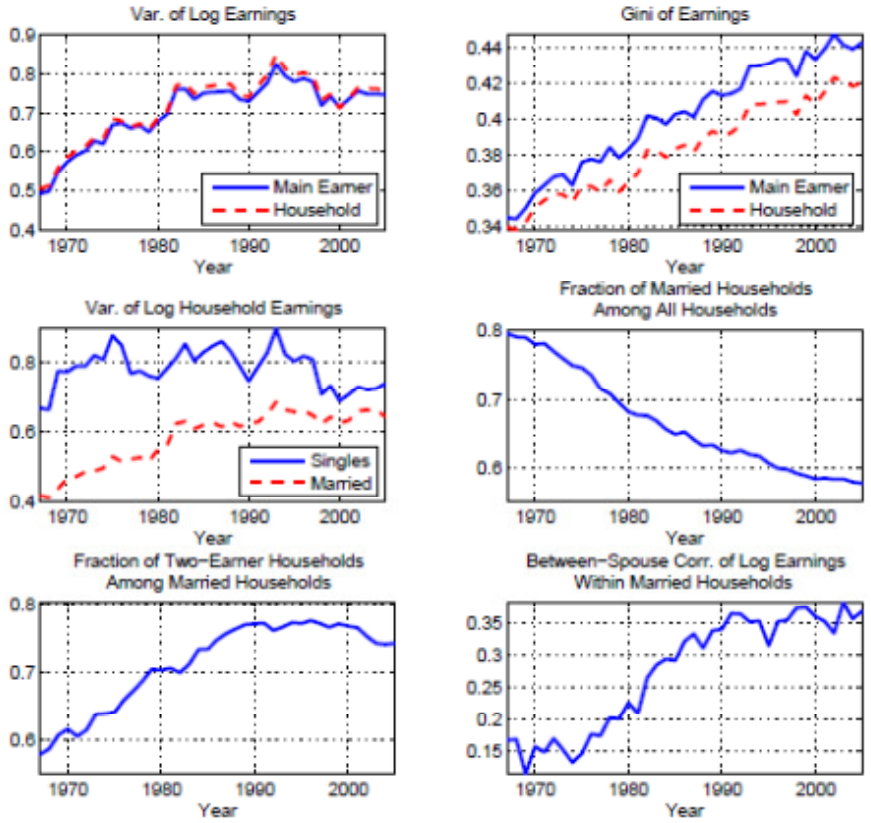


Figure 10: Understanding the role of the family for earnings inequality (CPS)

Source: Heathcote, Perri, and Violante (2010)

Disposable Income and Consumption Inequality over Time

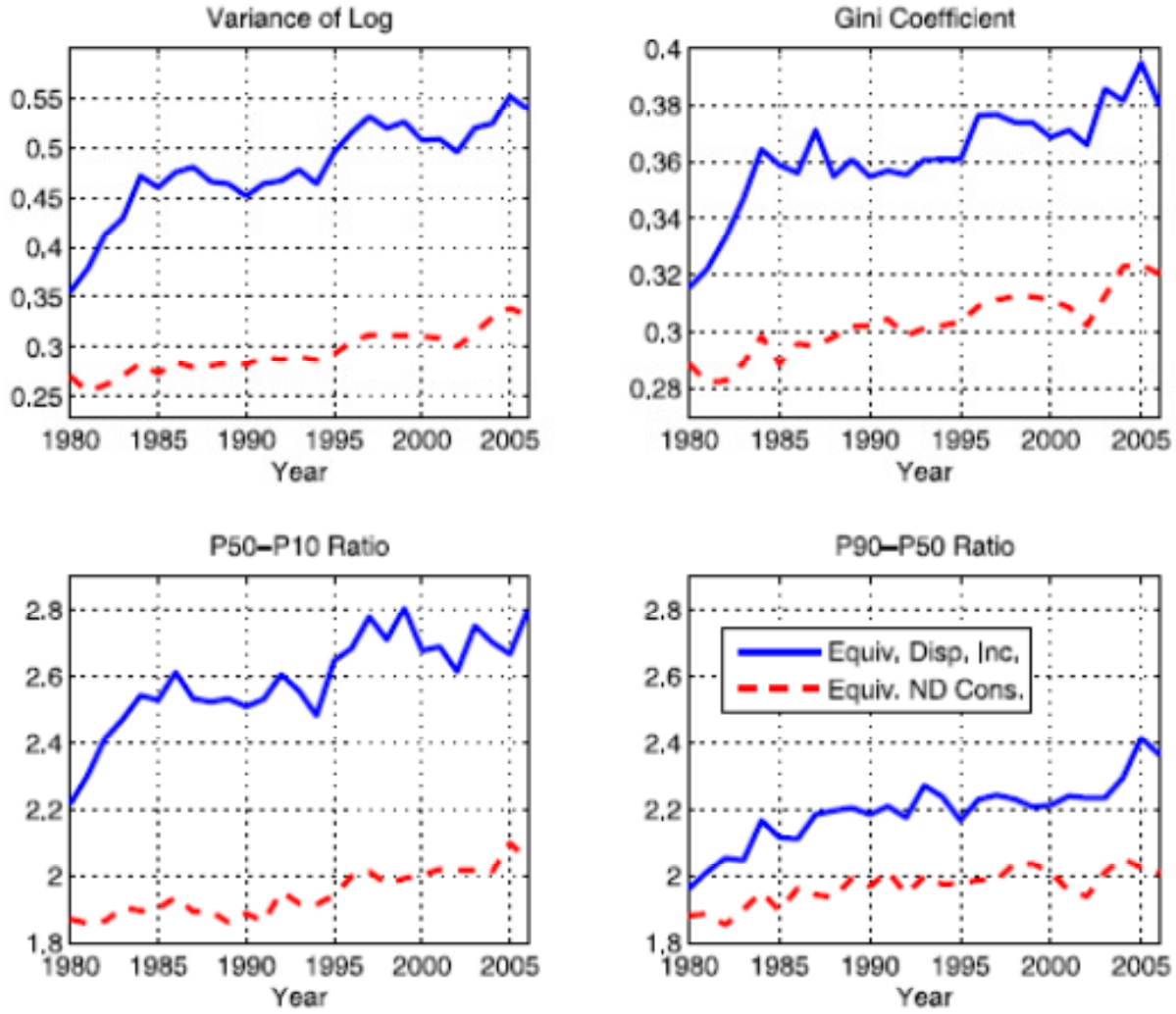


Fig. 13. From disposable income to consumption (CEX).

Source: Heathcote, Perri, and Violante (2010)

Labor Market Inequality

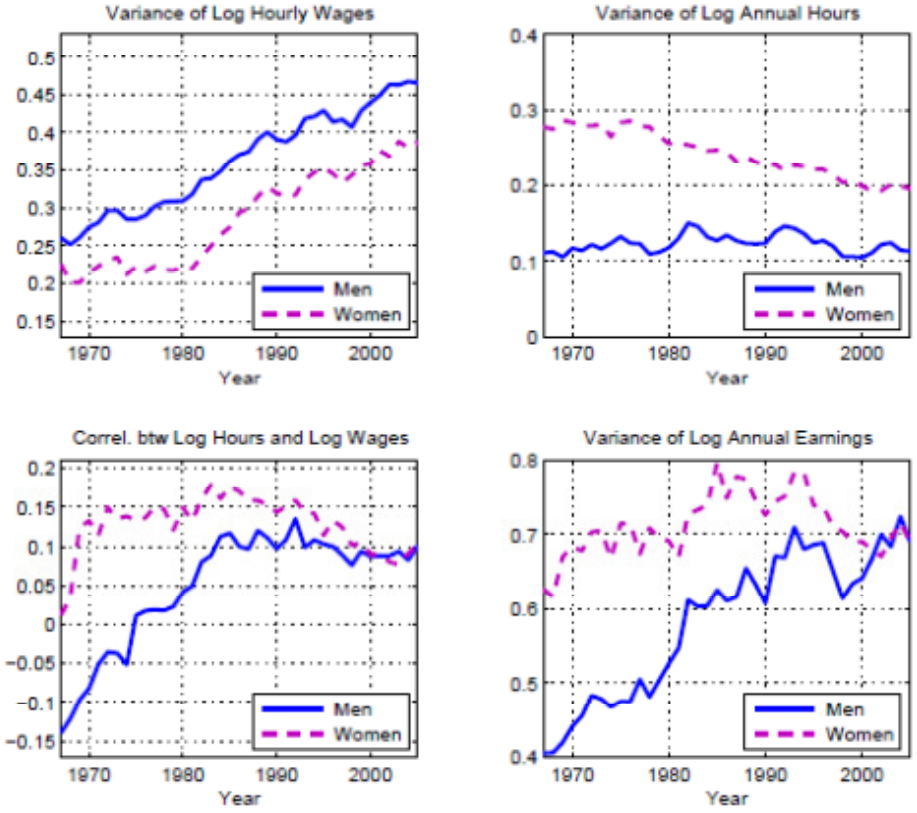


Figure 6: Inequality in labor supply and earnings of men and women (CPS)

Source: Heathcote, Perri, and Violante (2010)

Panel Survey of Income Dynamics (PSID)

- This is a representative panel data set of U.S. households.
- Data are collected at the household level (e.g. housing) and also at the individual level (e.g. age, education, earnings). The most detailed data is on the household 'head'.
- It focuses on income sources and amounts, employment and family composition changes and resident locations. Data no wealth is not as rich as in SCF.
- Very useful to understand aspects of economic mobility.

Economic Mobility in Rich and Poor Countries

		Fraction (%) of Households that Left the Quintile				
		1st	2nd	3rd	4th	5th
Income:						
US		23	37	40	39	23
Uganda:	▷ Rural	71	71	74	72	51
	▷ Urban	70	69	72	71	46
Wealth:						
US		25	39	41	37	20
Uganda:	▷ Rural	58	69	71	66	45
	▷ Urban	53	66	69	64	39

Current Population Survey, Merged Outgoing Rotation Groups, (CPS-MORG)

- This is the merged outgoing ration group of the CPS.
- Individuals in this group are also asked about their labor income
- Working example: Castro and Coen-Pirani (2007).
- They document a large change in the volatility of labor input across skill groups before and after 1984. Different results than those documented by Kydland in the first half of the 1980s.

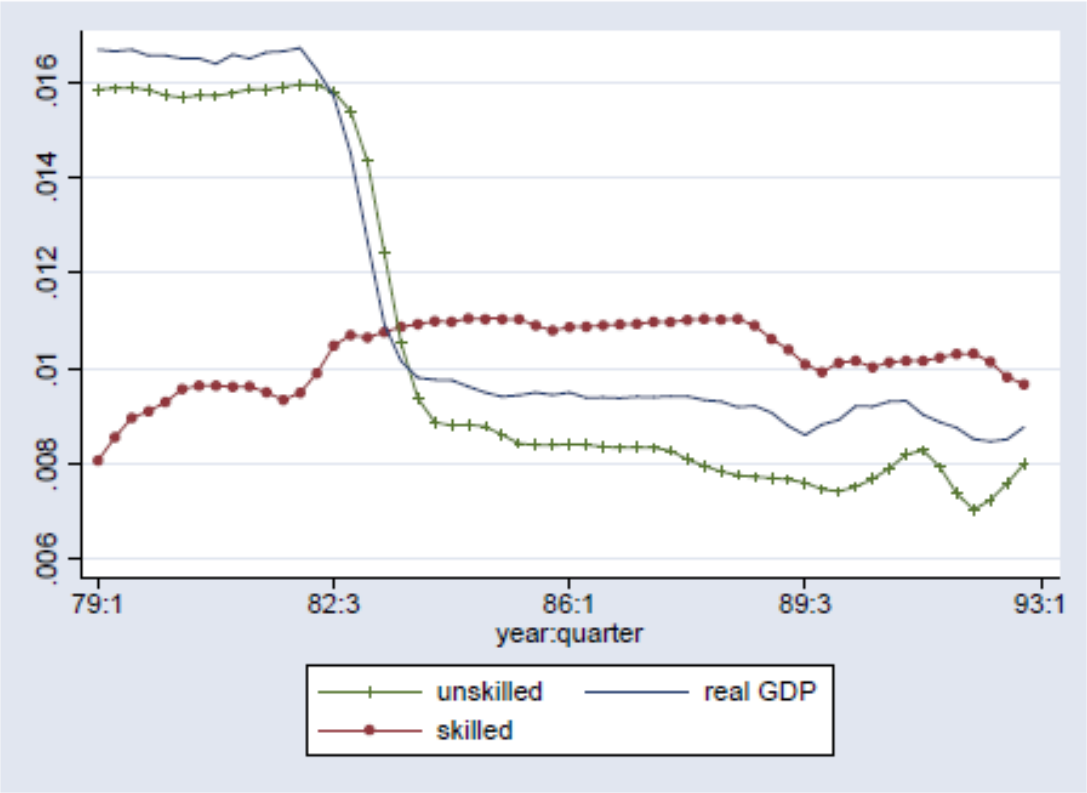


Figure 3: Rolling Standard Deviations (40 quarters ahead) of GDP, Unskilled and Skilled Hours

Source: Castro and Coen-Pirani (2007)

	Variable	Relative volatility			Comovement		
		skilled	unskilled	aggregate	skilled	unskilled	aggregate
1979:1-1983:4	Total hours	0.37	0.97	0.73	0.61^a	0.88^a	0.91^a
	Employment	0.32	0.82	0.67	0.25	0.86^a	0.87^a
	Average weekly hours of work	0.18	0.17	0.11	0.78^a	0.90^a	0.71^a
1984:1-2003:4	Total hours	1.04	0.90	0.81	0.71^a	0.69^a	0.80^a
	Employment	0.93	0.81	0.73	0.66^a	0.69^a	0.76^a
	Average weekly hours of work	0.30	0.26	0.22	0.45^a	0.28^b	0.47^a

Notes: a, b, c denote correlations significant at 1, 5, and 10 percent level respectively.

Table 1: Volatility and co-movement of total hours, employment and average weekly hours per skill group (Household Survey)

	Variable	Relative volatility			Comovement		
		skilled	unskilled	aggregate	skilled	unskilled	aggregate
1979:1-1983:4	Hourly wage	0.75	0.57	0.63	0.29	0.41^c	0.41^c
	Skill premium			0.26			-0.11
1984:1-2003:4	Hourly wage	0.96	0.66	0.74	0.35^a	0.27^b	0.29^a
	Skill premium			0.60			0.15

Notes: a, b, c denote correlations significant at 1, 5, and 10 percent level respectively.

Table 2: Volatility and co-movement of the skill premium and wages per skill group and in the aggregate (Household Survey)

Source: Castro and Coen-Pirani (2007)

Estimating Income Processes

- 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 is large. See the recent comprehensive reviews in Meghir and Pistaferri (2010) and Krueger et al. (2010).[1]

- 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. See Imrohoroglu (1989), Hugget (1993), Aiyagari (1994), Ríos-Rull (1995), Storesletten et al. (2004a), Storesletten et al. (2004b), and Krüger and Perri (2006).

  - A good understanding of the income process is key to study the response of consumption to income shocks.

[1] See also Guvenen (2009).

- A typical specification for the labor income process is:

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

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

→ α_t are year dummies;[2]
→ and function $f(a;\Theta)$ is a deterministic function of age (e.g., a quartic polynomial).
→ 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;[3,4]

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 $u_{i,a,t}$.

[2] 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.

[3] 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.

[4] 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 $y_{i,a,t}$?

- 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.

  One can argue that wages are the most 'primitive' measure of income as it does not explicitly depend on endogenous labor choices.[5] In principle, how correct that statement is depends on whether we disregard (i) on-the-job learning (or other human capital) arguments that suggest that labor choices today affect future wages and (ii) efficiency wages conditions where contemporaneous wages depend on current effort.

  Some times, as it is the case with CPS data, individual wages are not directly available from the survey. Data may, however, come in the form of individual labor income, wh_i, and individual hours, h_i. If so, we can compute wages as

$$w_i = \frac{wh_i}{h_i}.$$

  This is not free of problem, as in the computation itself we have forced individual wages to be an explicit function of self-reported individual hours.[6]

[5] One can also explicitly study the extend to what changes in labor income are due to changes in hours. See Abowd and Card (1989) and Low, Meghir, and Pistaferri (2009).

[6] If we are after average wages within education (or other) groups, we could, in principle, use data for wh_i from survey A, and data for h_i from survey B.

Sample restrictions

- More often than not, labor income estimations focus on the main earner (usually, the head of the household). This avoids dealing with (though not necessarily overcomes) explicit endogenous selection problems associated with labor supply choices of spouses.[7]

- Further, the literature usually focuses on employed individuals only. That is, the self-employed and individuals that have not worked for the past year are dropped from the sample. This removes from the analysis the endogenous selection of self-employed individuals bsaed on risk preferences (see Skinner, 1987; Guiso, Jappelli and Pistaferri, 2002; Fuchs-Schuendeln and Schuendeln, 2005).

[7] Blundell, Chiappori, and Meguir (2006) explicitly deal with this with a collective model of labor supply where husbands always work and choose how many hours (i.e., the intensive margin) and wives decide whether to work or not (i.e., the extensive margin).

- Also typically we look at individuals with completed education so that there are no problems of endogenous selection that may arise if we estimate g and f separately for each education group.

- Typically, we stop the estimation at the retirement age. However, if $y_{i,a,t}$ is household income, then it is hard to believe that income fluctuations stop at retirement age. Income shocks (not labor income shocks, but other shocks) may still occur after 65. For example, spouses may die (life insurance), there are transfers towards (or from) children or relatives, (this may include intervivos transfers or bequests), loss (or gains) in the value of assets (shocks to stocks and retirement accounts), etc.

  Retirement is potentially endogenous. For this reason some restrict the sample to few years before retirement. If retirement occurs on average at 65, it is typical to restrict the sample to individuals below 60 years of age.

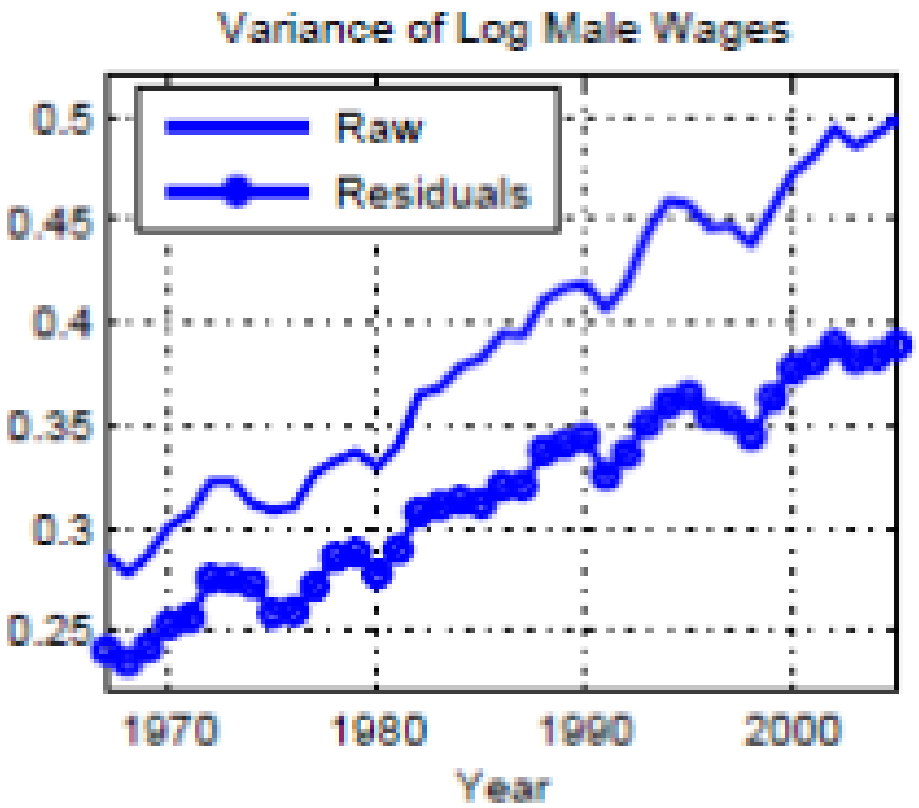
The Permanent-Transitory Decomposition of the Residual

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 (see Figure 6). 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.

Figure: Residual wage inequality over time



Source: Heathcote, Perri, and Violante (2010)

- 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.[8,9]

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

[9] 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 will impose restrictions on the covariance structure of the PT process; we will follow Heathcote et al. (2010).[10]

- 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).[11]

[10] See Hall and Mishkin (1982), Abowd and Card (1989), and Gottschalk and Moffit (1994), for earlier treatments; and Blundell et al. (2008), Jappelli and Pistaferri (2010), and Meghir and Pistaferri (2010), for more general treatments.

[11] See a discussion in Guvenen (2009) and Hryshko (2010).

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

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

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 of that partition 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

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

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 to identify our parameters. Note the relationship of income for any individual i between two consecutive periods, specifically, t and $t+2$ is

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

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., $\sigma^2_{\epsilon_{a,t}}$, 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 $t+2$. That is, to identify the variance of transitory shocks we need data on two consecutive periods (of available data), t and $t+2$.

(1b) Note the relationship of income for any individual i between two consecutive periods, specifically, 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

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

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

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

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., $\sigma^2_{\widetilde{\eta}_{a,t}} = \sigma^2_{\eta_{a-1,t-1}} + \sigma^2_{\eta_{a,t}}$) 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 of that partition between period t and period $t-2$. That is, 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

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

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 $t-2$,

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

to identify the variance of the permanent shocks for $\{\sigma^2_{\widetilde{\eta}_{a,2008}}, \sigma^2_{\widetilde{\eta}_{a,2010}}\}_{a=20}^{a=64}$. When we report the variance of the permanent shock we do so by, following the practice in Heathcote et al. (2010), 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

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

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:

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

Interpreting the results

- Next we discuss the results in Heathcote et al. (2010) that conduct a PT estimation of wages for the U.S. The U.S. PSID data sampled biannually to deal with the fact that recently the PSID moved from collecting data every year to every two years.

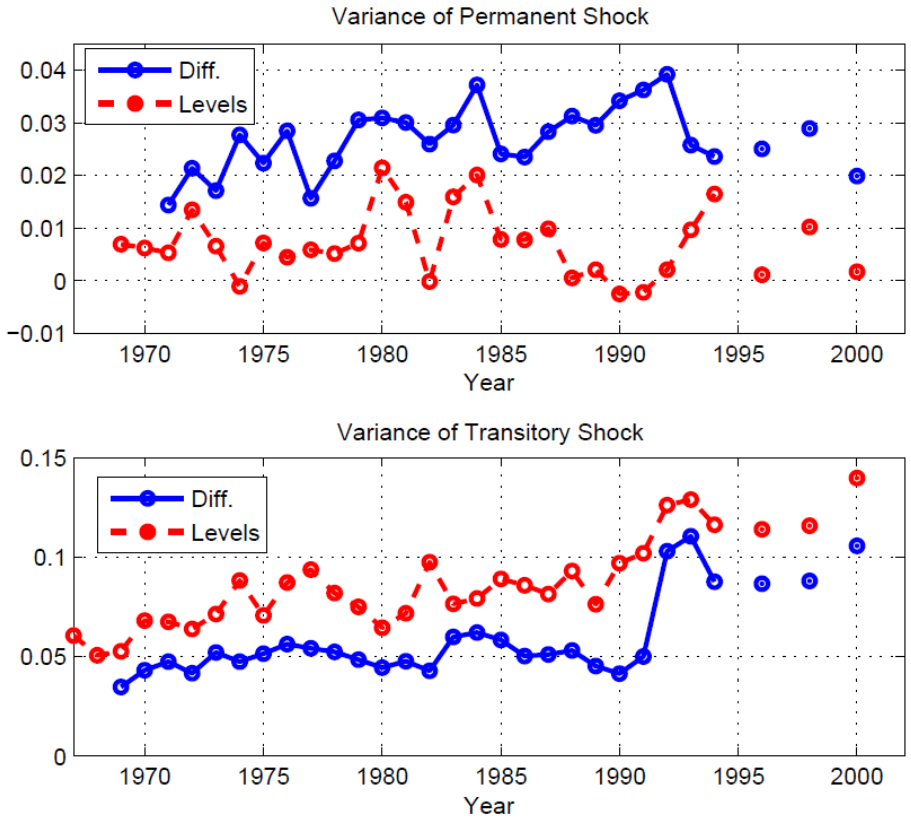


Figure 18: Estimates of the variances of the transitory and permanent components (PSID)

Source: Heathcote, Perri, and Violante (2010)

Findings:

1. The variance of permanent shocks remains fairly constant in levels and trends, while the variance of transitory shocks increases over time since late 1960s in differences, and largely since early 1990s in levels. That is, the increase in residual variance discussed earlier can be largely attributed to a rise in the variance of transitory shocks; in particular, the 1990s and after show an episode of really high transitory shocks.

   This suggests that the increase in residual inequality can be (in principle) easily insurable. This is consistent with inequality in consumption increasing less than inequality of income over this period (a feature of the U.S. data that we have seen earlier).

2. The variance of the transitory shocks is larger than the variance of the permanent shocks:[12]

   i. On average, the variance of the transitory shocks identified in log levels (about .09) is about 9-10 times larger than the variance of the permanent shocks (about .007),
   ii. On average, the variance of the transitory shocks identified in growth rates (about .07) is between 3 and 4 times larger than the variance of the permanent shocks (about .027),

[12] The greater the share of the permanent variance in the total variance of (log) earnings, the greater the persistence of earnings differentials. Sources of changes in the permanent variance often referred to are changes in the returns to different levels of skills, or changes in institutional arrangements such as changes in industrial structure or collective bargaining mechanisms. The greater the share of the transitory variance, the greater year-to-year mobility of workers there is within the earnings distribution.

- The finding of increase income inequality and the fact that most of this increase is due to transitory shocks seems robust across European countries as well.

- 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. This is what we actually observe in the U.S, Italy, and many other countries.

- 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.

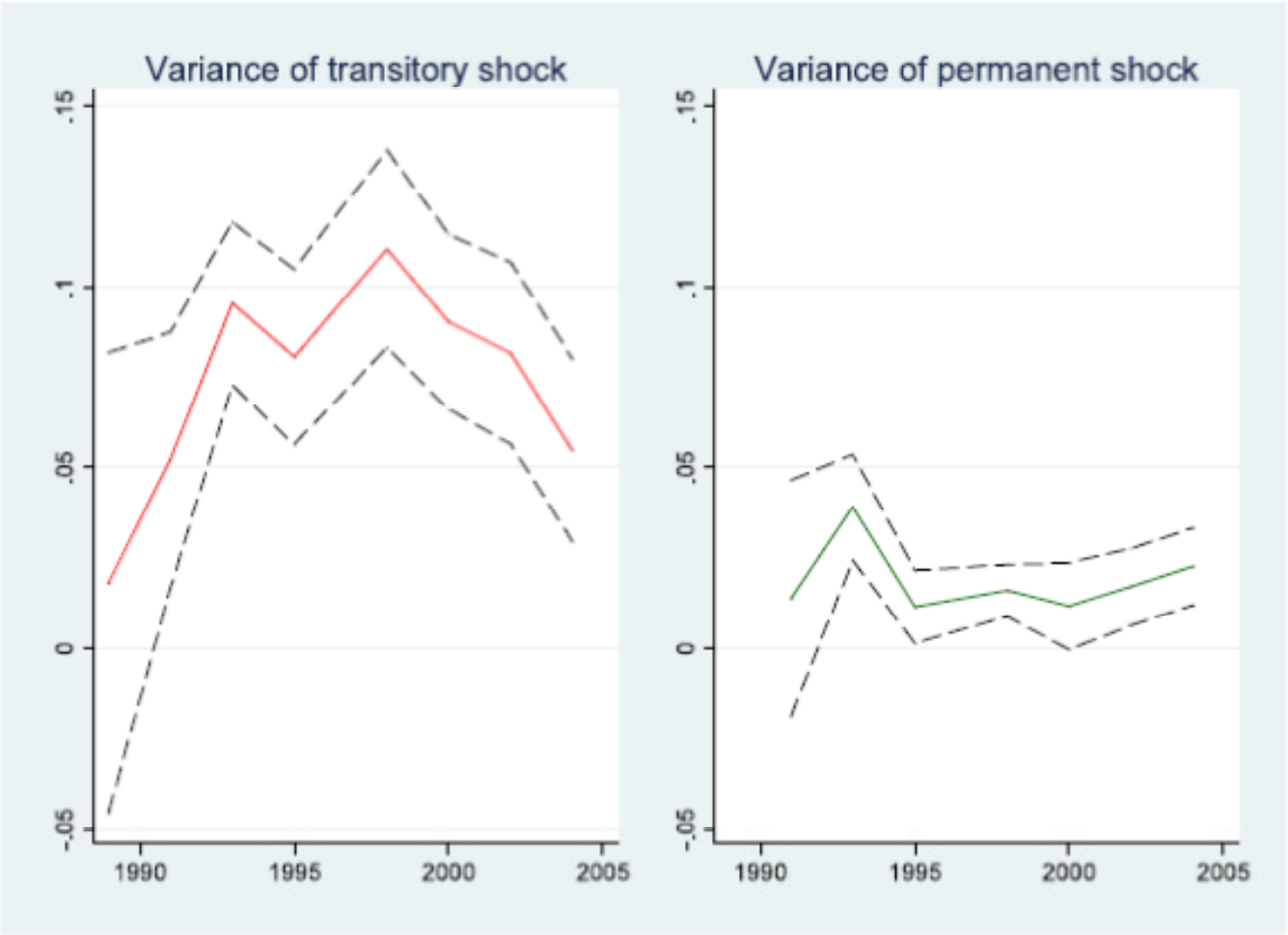
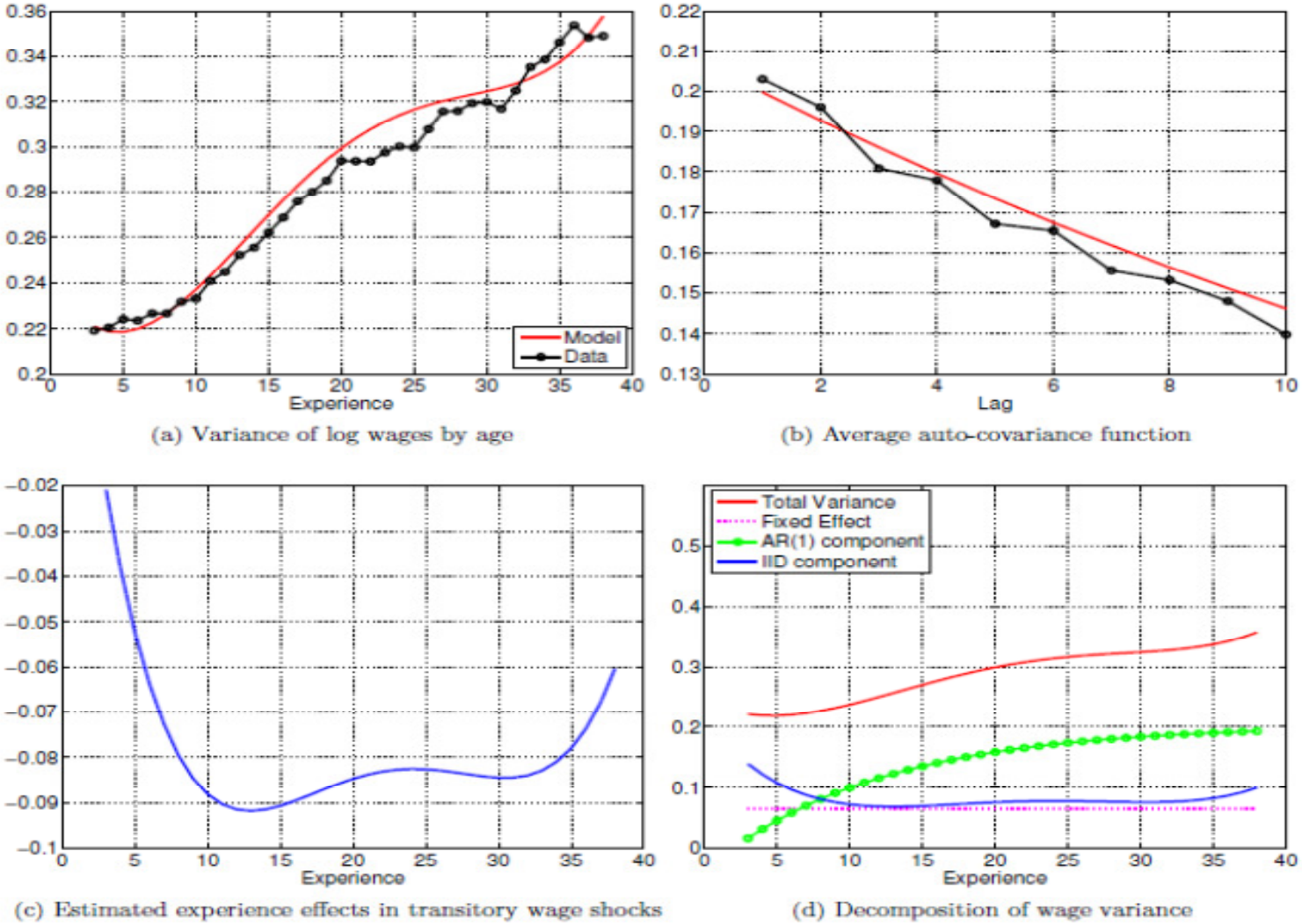


Fig. 20. Variances of transitory and permanent income shocks. *Note.* Variances of the two shocks are computed using the panel section of the SHIW. Income is defined as disposable income net of financial assets. The lines represent interpolation of the original data points using locally weighted OLS.

Source: Jappelli and Pistaferri (2010) [Italy]



(a) Variance of log wages by age

(b) Average auto-covariance function

(c) Estimated experience effects in transitory wage shocks

(d) Decomposition of wage variance

Figure 15: Model fit for estimated wage process (year effects)

Source: Kaplan (2010), AR(1) dynamics in the transitory component

Misspecification (?)

- 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.[13]

- 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.

[13] 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. That, the implications of 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. [14]
- 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.

[14] If, for example, the fixed-effects wage profiles overstate average wages at age 60 by 10% in the simulated sample, then it is likely that wages have been overestimated at age 60 by 10% in the PSID data. Therefore, the candidate for the unobserved average wage at age 60 is the fixedeffects estimate from the PSID data, less 10%. This new candidate wage profile is fed into the model and the procedure is repeated. If, for example, the fixed-effects profile using simulated data still indicates a 1% upward bias, the true candidate wage profile is reduced by an additional 1%.

Acemoglu, D. (2002). Technical change, inequality, and the labor market. *Journal of Economic Literature*, 40:7–72.

Acemoglu, D., Aghion, P., and Violante, G. (2001). Deunionization, technical change and inequality. *Carnegie-Rochester Conference Series on Public Policy, 55(1) , pp. 229-264.*, 55(1):229=264.

Acemoglu, D. and Autor, D. H. (2011). Skills, tasks and technologies: Implications for employment and earnings. In Ashenfelter, O. and Card, D., editors, *Handbook of Labor Economics*, volume 4, chapter 12, pages 1043–1171. Elsevier.

Aiyagari, S. R. (1994). Uninsured idiosyncratic risk, and aggregate saving. *Quarterly Journal of Economics*, 109:659–684.

Ameriks, J. and Zeldes, S. P. (2004). How do household portfolio shares vary with age. Mimeo, Columbia University.

Bloom, N., Romer, P., and Reenen, J. V. (2010). A trapped factors model of innovation. Mimeo, Stanford University and London School of Economics.

Blundell, R., Pistaferri, L., and Preston, I. (2008). Consumption inequality and partial insurance. *American Economic Review*, 98(5):1887–1927.

Budria, S., Díaz-Giménez, J., Quadrini, V., and Ríos-Rull, J.-V. (2002). Updated facts on the u.s. distributions of earnings, income and wealth. *Federal Reserve Bank of Minneapolis Quarterly Review*, Summer:2–35.

Castañeda, A., Díaz-Giménez, J., and Ríos-Rull, J. V. (2003). Accounting for earnings and wealth inequality. *Journal of Political Economy*, 111(4):818–857.

De Nardi, M. (2004). Wealth inequality and intergenerational links. *res*, 71(3):743–768.

Díaz-Giménez, J., Glover, A., and Ríos-Rull, J. V. (2011). Facts on the distributions of earnings, income, and wealth in the united states: 2007 update. *Federal Reserve Bank of Minnepolis Quarterly Review*, 34(1):2–31.

French, E. (2005). The effects of health, wealth, and wages on labor supply and retirement behavior. *Review of Economic Studies*, 72:395–427.

Guvenen, F. (2009). An empirical investigation of labor income processes. *Review of Economic Dynamics*, 12(1):58–79.

Hall, R. E. and Mishkin, F. S. (1982). The sensitivity of consumption to transitory income: Estimates from panel data on households. *Econometrica*, 50(2):461–481.

Heathcote, J., Perri, F., and Violante, G. (2010). Unequal we stand: An empirical analsys of economic inequality in the united states, 1967-2006. *Review of Economic Dynamics*, 13(1):15–51. Special Issue, January 2010.

Heathcote, J., Storesletten, K., and Violante, G. (2005). Two views of inequality over the life cycle. *Journal European Economic Association*, 3(2-3):765–775.

Hornstein, A., Krusell, P., and Violante, G. L. (2005). The effects of technical change on labor market inequalities. In Aghion, P. and Durlauf, S., editors, *Handbook of Economic Growth*, volume 1, chapter 20, pages 1275–1370. Elsevier.

Hryshko, D. (2010). Rip to hip: The data reject heterogeneous labor income profiles. Mimeo, University of Alberta.

Hugget, M. (1993). The risk-free rate in heterogenous-agent incomplete-insurance economies. *Journal of Economic Dynamics and Control*, 17(5/6):953–970.

Huggett, M. (1996). Wealth distribution in life-cycle economies. *Journal of Monetary Economics*, 38(3):469–494.

Imrohoroglu, A. (1989). The costs of business cycles with indivisibilities and liquidity. *Journal of Political Economy*, 97:1364–83.

Jappelli, T. and Pistaferri, L. (2010). Does consumption inequality track income inequality in italy? *Review of Economic Dynamics*, 13(1):133–153. Special Issue, January 2010.

Katz, L. F. and Autor, D. H. (1999). Changes in the wage structure and earnings inequality. In Ashenfelter, O. and Card, D., editors, *Handbook of Labor Economics*, volume 3, chapter 26, pages 1463–1555. Elsevier.

Katz, L. F. and Murphy, K. M. (1992). Changes in relative wages, 1963–1987: Supply and demand factors. 107(1):35–78.

Krueger, D., Perri, F., Pistaferri, L., and Violante, G. L. (2010). Cross sectional facts for macroeconomists. *Review of Economic Dynamics*, 13(1). Special Issue, January 2010.

Krüger, D. and Perri, F. (2006). Does income inequality lead to consumption inequality? *Review of Economic Studies*, 73(1):163–193.

Krusell, P., Ohanian, L. E., Ríos-Rull, J.-V., and Violante, G. L. (2000). Capital-skill complementarity and inequality: A macroeconomic analysis. *Econometrica*, 68(5):1029–1054.

Meghir, C. and Pistaferri, L. (2010). Earnings, consumption and lifecycle choices. NBER Working Paper No. 15914.

Olivetti, C. and Petrongolo, B. (2009). Unequal pay or unequal employment? a cross-country analysis of gender gaps. *Journal of Labor Economics*, 26(4):621–654.

Quadrini, V. (1997). Entrepreneurship, saving and social mobility. 3(1):1–40.

Ríos-Rull, J.-V. (1995). Models with heterogenous agents. In Cooley, T. F., editor, *Frontiers of Business Cycle Research*, chapter 4. Princeton University Press, Princeton.

Storesletten, K., Telmer, C., and Yaron, A. (2004a). Consumption and risk sharing over the life cycle. *Journal Monetary Economics*, 51(3):609–633.

Storesletten, K., Telmer, C., and Yaron, A. (2004b). Cyclical dynamics in idiosyncratic labor market risk. *Journal of Political Economy*, 112(3):695–717.