# Cross-Sectional Facts for Macroeconomists: United States (1967-2006)*

**Jonathan Heathcote**
Federal Reserve Board of Governors, Georgetown University, and CEPR
jonathan.h.heathcote@frb.gov

**Fabrizio Perri**
University of Minnesota, Minneapolis Fed, CEPR, and NBER
fperri@umn.edu

**Giovanni L. Violante**
New York University, CEPR, and NBER
glv2@nyu.edu

This draft: June 30, 2008
very preliminary and incomplete! Please do not quote

**Abstract**

This paper documents time trends for U.S. cross-sectional inequality in individual wages and hours, and household earnings and consumption. The empirical analysis covers the period 1967-2006 and is based on three data sources: CPS, PSID and CEX. An effort is made to create comparable samples across the three data sets.

*We are grateful to Greg Kaplan and Kai Steverson for outstanding research assistance. We thank Dean Lillard for providing data on transfers for the 1994-2003 waves of PSID. Heathcote and Violante thank the National Science Foundation (Grant SES-0418029). The opinions expressed here are those of the authors and not necessarily those of the the Federal Reserve Bank of Minneapolis or the Board of Governors of the Federal Reserve System

# 1 Introduction

The rise in economic inequality in the United States has been extensively studied. Relative to the large existing literature, the contribution of this paper, following the RED guidelines, is three fold. First, rather than focusing on a specific measure of economic wellbeing, we begin with changes in the structure of relative individual wages, as our most primitive measure of inequality, and from there take a series of steps to contrast inequality in individual wages to that in individual earnings, household earnings, pre-government income, disposable income, and, ultimately, consumption. Along the way we consider the impact on measured inequality of individual labor supply, household income pooling, private transfers and asset income, government redistribution, and household net saving. Second, we systematically compare and contrast the pictures for evolving inequality as presented by three widely-used datasets: the March supplement to the Current Population Survey (CPS), the Panel Study of Income Dynamics (PSID), and the Consumer Expenditure Survey (CEX). Third, we estimate the evolution of individual income dynamics and the evolution of within-cohort inequality over the life-cycle, in both cases carefully addressing the fact that our surveys follow households during a period of rising risk and inequality.

While many of our findings are familiar, we also document some interesting new results.

- Generally speaking, the CPS, CEX and PSID align very closely in terms of how the distribution of income has evolved over time. This is reassuring, since it means that researchers can estimate individual income dynamics from the PSID, or measure consumption inequality in the CEX, and safely make comparisons to cross-sectional moments from the much larger CPS sample.

- Cyclical fluctuations in CPS per capita income are much larger than in NIPA personal income

- While measures of average consumption in the CEX exhibit large differences in levels and trends relative to their NIPA counterparts, per capita food consumption in the CEX and the PSID align closely.

- The dynamics of relative wages, both between and within groups, align remarkably closely across the CPS, the CEX and the PSID.

- Relative to the CPS and the CEX, the PSID misses the widening in inequality at the bottom of the income distribution in the 1970s and the early 1980s

- Residual (within-group) male wage inequality rises steadily between 1967 and 2005, while residual household earnings inequality is relatively stable after the early 1980s

- In cross-section, government and transfers compress household income inequality dramatically. However, the inequality-reducing effect of public transfers diminished considerably in the late 1990s.

- The incomes of those at the bottom of the income distribution are much more cyclical than those at the top

- The increase in within-cohort inequality over the life-cycle appears much smaller when the rise in overall US inequality is modelled through time effects rather than cohort effects. Moreover, the age-profile for wage inequality is concave, while that for household earnings inequality is convex.

- Models for individual wage and income dynamics estimated on individual growth rates deliver grossly counter-factual implications for age profile of within-cohort dispersion.

The rest of the paper is organized as follows. Section 2 describes the three data sources: the CPS, the PSID, and the CEX. Section 3 presents the results of our empirical analysis, and Section 4 concludes. Many details of the empirical analysis are omitted from the main text and collected in the Appendix, to which we will refer throughout the paper.

# 2 The three data sets

In this section, we briefly describe CPS, PSID and CEX. The Appendix contains more detail on each source of data, the definition of the key variables in each data set, and a discussion of how we overcame certain shortcomings of each data set in constructing the baseline samples.

## 2.1 CPS

The CPS is the source of official government statistics on employment and unemployment and is designed to be representative of the civilian non-institutional population. Currently around

55,000 households are interviewed monthly out of around 60,000 eligible housing units sampled. The Annual Social and Economic Supplement (ASEC) formerly known as the Annual Demographic Survey applies to the sample surveyed in March, and extends the set of demographic and labor force questions asked in all months to include detailed questions on income. For the ASEC supplement, the basic CPS monthly sample of around 60,000 households is enlarged by adding an additional 4,500 hispanic households (since 1976) and, more importantly, by the addition of 34,500 extra households in 2002 as part of an effort to improve state estimates of children's health insurance coverage: this sample expansion is known as the SCHIP sample. We use the Unicon CPS Utilities version of the March CPS. The number of households in the Unicon sample rises from around 46,700 in income year 1967 to 75,900 in 2005

The basic unit of observation is a housing unit, and for the CPS we report statistics on inequality at the level of the household (rather than at the level of the family). The CPS householder refers to the person, or one of the persons (the first one listed by the respondent answering the question), in whose name the housing unit is owned or rented, and is the "reference person" to whom the relationship, if any, of other household members is recorded. The March CPS contains detailed demographic data for each household member and labor force and income information for each household member aged 15 or older. Labor force and income information correspond to the previous year.

We use the March supplement weights to produce our estimates. Weights are chosen to make the CPS sample representative of the US population, and apply at the individual level. For household level variables, we use the household weight. The supplement weights differ from the usual monthly CPS weights, reflecting differences in the sample, particularly the inclusion of the SCHIP subsample.

The Appendix contains more details on the March CPS, and the exact definition of each variable used in this study.

## 2.2 PSID

The Panel Study of Income Dynamics (PSID) is a longitudinal study of a representative sample of U.S. individuals (men, women, and children) and the family units in which they reside. From 1968, the PSID interviewed individuals from families in the initial sample, whether or not they were living in the same dwelling or with the same people. Adults have been followed as they have grown older, and children have been observed as they have advanced into adulthood, forming family units of their own (the "split-offs"). New members of the initial families (e.g., newborn children) and of split-offs families (e.g., new spouses of grown-up members in the initial sample) were systematically added to the sample. The survey is thus structured as an unbalanced panel. Survey waves are annual from 1968 to 1997. Since 1997, the PSID became biennial. In what follows we use all the yearly surveys (1967-1996) and biennial surveys from 1999, 2001 and 2003.

The PSID data was collected in face-to-face interviews based on paper and pencil questionnaires between 1968 and 1972. Thereafter, the majority of interviews were conducted over the telephone. In 1993, the PSID introduced the use of computer-assisted telephone interviewing (CATI) and administered its questionnaire this way since then.[1] The annual response rate is very high, typically around 97%.

The PSID data files provide a wide variety of information about both families and individuals. The focus of the data is economic and demographic, with substantial detail on income sources and amounts, employment status and history, family composition changes, and residential location. Information gathered in the survey applies either to the circumstances of the family unit as a whole (e.g., type of housing, food consumption expenditures) or to particular persons in the family unit (e.g., wages, earnings, education level). While some information is collected about all individuals in the family unit, the greatest level of detail is ascertained for the primary adults in the family unit, i.e. the head and the spouse, when present.[2]

We base our empirical analysis on the so called "SRC sample". It is a cross-sectional national sample which was originally drawn in 1968 by the Survey Research Center (SRC). It is an equal probability sample of households from the 48 contiguous US states and was designated to yield just below 3,000 completed interviews to families. Using the SRC sample alone does not require sample weights.[3]

The Appendix contains more details on PSID, and lists the exact definition of each variable used in this study.

[1]In the 1999 wave, 97.5% of the interviews were conducted over the phone, and all interviews were conducted using computer-based instruments.

[2]In PSID, the head is usually the male above 16 years-old with the highest financial responsibility. In households without males, it is the female older than 16 with the highest financial responsibility. See the Appendix for more details on the definition of head.

[3]See the Appendix for a description of the other samples in PSID.

## 2.3 CEX

The Consumer Expenditure Survey (CEX) consists of two surveys, the quarterly Interview Survey and the Diary Survey both collected for the Bureau of Labor Statistics by the U.S. Census Bureau. It is the only US dataset that provides detailed information about households consumption expenditures. As the Diary survey focuses only on expenditures on small, frequently purchased items (such as food, beverages and personal care items) while the interview survey aims to provide information for up to 95% of the typical household consumption expenditures in this study we will focus only on the Interview survey (see Attanasio, Battistin and Ichimura for a study that jointly use the Diary and the Interview survey). The CEX Interview Survey is a rotating panel of households that are selected to be representative of the U.S. population. It started in 1960, but continuous data are available only from the first quarter of 1980 until the first quarter of 2007. Each quarter the survey contains detailed information on quarterly consumption expenditures and on demographic characteristics for all households interviewed during that quarter. After a first preliminary interview (not available in the public access files), each household is interviewed for a maximum of four consecutive quarters. In the second and fifth interviews, household members are asked questions about earnings, other sources of income, hours worked, and taxes paid for the past year. Records of households in their third and fourth interview also contain income information but it is the one collected either in the second or the fifth. For the cross sectional consumption statistics computed in this paper the universe of records for analysis in the CEX is constituted by all household/quarter observations that satisfy the sample restriction we discuss below. For the cross sectional wages, income, taxes and hours the universe of records of analysis in the CEX is all household/quarter observation who satisfy the sample restrictions below, who are in the second or fifth interview (in which income questions are asked) or which are in the sample for their third or fourth interview and the income figures in their record is not reported in any other interview. [4]. For details about the exact definition of variables, about treatment of data, about time aggregation and deflation and how we deal with top-coding, bracketing and imputation please see the appendix at the end.

[4] We include these households as there is non negligible number of households who only appear in the CEX files only in the third of four interview

## 2.4 Comparison across data sets

he similarities across surveys, in general, are large and make the comparison across the three data sets meaningful and appropriate. However, definitions of some key variables are different, which often explains divergence in sample means or trends.

Individual labor income is defined in all three surveys as the sum of all income from wages, salaries, commissions, bonuses, and overtime, and the labor part of self-employment income. The PSID uses a 50-50 rule in splitting self-employment income between capital and labor components, while in CPS and CEX data we do a separate imputation based on 2/3 to labor income and 1/3 to asset income. One important difference in the definition of asset income is that only in the CEX is it possible to impute rents from owner occupied housing across the entire sample period. Also the calculation of taxes differs across data sets. The PSID includes a variable for household income taxes only up until 1991. Rather than using this variable, we use the NBER's TAXSIM program to calculate an estimate of household federal income taxes, and state income taxes that is comparable across all years in the sample. The CPS contains imputed values for federal and state income taxes, social security payroll taxes, and the earned-income tax credit for the 1979-2004 income years. The CEX finally asks each household member in the second and fifth interview a question about taxes (including federal state and local) paid in the previous year.

Hours worked are defined as the sum of annual hours worked on the main job, on extra jobs, plus annual hours of overtime in all three surveys. In the PSID, they are computed by using information on "usual hours worked per week" and the "number of actual weeks worked" in the last year. In CPS, they are computed in exactly the same way only from 1975 onwards. Before 1975, there was no question on "usual hours worked per week" so we are forced to measure hours as weeks worked last time times hours worked last week (see the CPS Appendix for more details).

In all three data sets, the hourly wage is defined as labor income divided annual hours of work.

An important discrepancy between PSID on one side and CPS and CEX is that in the former the unit of analysis is not the household, but the "family unit" (FU). The difference comes from the fact that PSID is a longitudinal survey following the same set of families, and their split-offs, since 1968. In CPS and CEX a household is composed by all persons, related or unrelated, living together in a dwelling unit. In PSID, the family unit is defined as a group of people living together who are usually related by blood, marriage, or adoption. These definitions are close, but not perfectly overlapping. For example, not all people living within a household containing a PSID member are members of that family (e.g., if they are friends). This difference may affect statistics on household size and equivalence scales. Moreover, the PSID does not collect information on household members who do not belong to the family, so discrepancies can arise in calculations involving household income (e.g., average household income in the PSID should be lower).

There are three additional important issues in PSID that make the sample not perfectly comparable to CPS and CEX. First, prior to 1975 and post 1994, labor income and hours worked are not reported in the PSID for household members who are not heads or spouses. All our calculations in PSID only refer to heads and spouses, whereas in CPS and CEX they include also other individuals who are non-heads and non-spouses. Second, the SRC sample may have somewhat lost representativeness of the U.S. population over time because it is based on a set of families chosen in 1968, and as such, it does not account for the new inflows of immigrants in the US. Third, the CPS imputes values for missing income data, while the PSID and the CEX do not.

A final remark on top coding. While top-coding affects very few observations in PSID, it has a non-sizeable impact on CPS statistics. In both datasets, we forecast mean values for top-coded observations by extrapolating a Pareto density fitted to the non-top-coded upper end of the observed distribution. We apply this procedure separately to each component of income in each year: thus estimated distribution parameters that defines inequality are allowed to vary both across different components of income and across time. Towards the end of our sample period, the CPS reports cell means for top-coded observations, which we use to evaluate the regression procedure.[5] In addition to top-coding for public use, the March CPS also suffers from internal censoring problems, meaning that interviewers were unable to record income levels above particular thresholds ($99,989 for wage income between 1975 and 1984, for example). Our extrapolation-based approach to top-coding automatically takes care of this problem.

Table 1 reports some key demographic characteristics of the population aged 25-60 in the

[5]In the interests of maintaining as consistent as possible an approach to measurement, we do not use these cell means directly to replace top-coded observations for primary earned income. However, at the same time that the CPS started to report cell means, they also dramatically reduced public-use top code values for a range of income variables, which in practice forces us to assign cell mean values to top-coded observations for categories of unearned income, and secondary sources of earned income (see the Appendix for more details).

three data sets. Both the levels and trends align quite well. Note that the fraction of white males declines much faster in CPS, confirming that the CPS may capture better the –largely non-white– immigrant population. In addition, a significantly larger fraction of households (families) in the PSID contain married couples, indicating that the PSID under-samples non-traditional households.

## 2.5 Sample selection

In each of our three datasets, we construct three different samples, that we label samples A, B, and C.

Sample A is the most inclusive, and is essentially a cleaned version of the raw data. We only drop records if 1) there is no information on age for either the head or the spouse, 2) if either the head or wife has positive labor income but zero annual hours, and 3) if either the head or wife has an hourly wage less than half the corresponding Federal minimum wage in that year. In PSID, we set to missing consumption observations where equivalized household food expenditures are less than 1,000 dollars. In CEX, in order to reduce measurement error in income (see Nelson, 1996), we exclude households that are flagged as incomplete income reporters. Sample A captures the entire population and is used for Figures 1-3, where we compare per-capita means from micro-data to NIPA aggregates.

Sample B is further restricted by dropping a household from sample A if no household member is between the ages of 25 and 60 (in the PSID we drop households if neither the head nor the spouse falls in this age range). The household head is the oldest working age male, as long as there is at least one working-age male in the household - otherwise the head is the oldest working age female. Sample B is our household-level sample and is used for Figures 7-14, where we study household-level inequality.

To obtain sample C, an individual is retained from sample B only if she/he has annual work hours of at least 260 hours. Sample C is our individual-level sample and is used for Figures 4, 5, and 6.

Table 2 shows the number of household/year observation that were lost at each stage of the selection process.

# 3 Results

We begin by comparing the evolution of average household earnings, income and consumption in our micro data to the official NIPA, over the period of interest, 1967-2005. Next, we compare trends in average wages, and hours worked in CPS and PSID. We then move to analyzing time trends in individual-level inequality statistics, and in household-level dispersion measures, for various definitions of income and consumption.

Next, we study the evolution of inequality over the life cycle. Then, we perform log-Normality tests for various measures of income and consumption. Finally, we estimate stochastic processes for wages, household earnings and disposable income, allowing for time variation in some of the parameters.

In what follows, we express all income and expenditure variables in year 2000 dollars. Our equivalence scale is based on the OECD scale, and assigns a value of 1.0 to the first adult, a weight of 0.7 to each additional adult, and a weight of 0.5 to each child.[6]

## 3.1 Trends in means and comparison with NIPA

Before exploring income inequality, we begin by comparing measures of average survey income to their analogues in the national income accounts.

The income definition that is conceptually most similar across the CPS and the National Income and Product Accounts is wage and salary income. Two minor differences, however, are worth noting (Ruser, Pilot and Nelson, 2004). The first involves the treatment of S corporation profits. The BEA classifies as dividends all S corporation profits distributed to shareholders, while the Census treats these profits as wage and salary income if the recipients are shareholder-employees. The second is that the BEA (but not the CPS) makes an upwards adjustment for wage and salary income earned in the underground economy from legal but "off the books" activities.

The top panel of Figure 1 compares average wage and salary income in the CPS and the CEX to the corresponding NIPA variable (Table 2.1, line 3, "wage and salary disbursements").[7] All series are in per capita, real and logged. The US population estimate is from NIPA Table 7.1, line 16. The price deflator used is the Bureau of Labor Statistics CPI-U series, all items.

[6] In PSID, a child is a family member younger than 18 years-old. In CPS,...

[7] We do not plot a PSID series here, because the PSID data does not always allow for a clean separation between wage and salary income and self-employment income.

Wage and salary income aligns remarkably well across CPS and NIPA, in terms of levels, trends, and business cycle fluctuations. On average across the 1967-2005 period, the CPS statistic exceeds its NIPA counterpart by 0.27 percent. The average absolute discrepancy is 1.1 percent. The reliability of CPS wage and salary reporting is confirmed by Roemer (2002), who matches individuals in the March CPS to detailed earnings records from the Social Security Administration. He finds that workers with part-year, part-time workers have underestimated March CPS wages (CPS/DER ratio around 90 percent), but that for all other groups wages align very closely. In the early 1990s CPS wages rise somewhat more rapidly than in the NIPA, a finding previously noted by Roemer. Conversely there is less evidence of a decline in CPS wages in the early 2000s.

We next consider total income. The CPS measure used here is total pre-tax cash income, our construction of the concept labeled "money income" by the CPS. This is wages and salaries, self-employment income, net financial income, and public and private transfers. The NIPA measure is "personal income", Table 2.1, line 1.

There are much larger conceptual and practical differences between these two measures of income. The main conceptual differences are that the survey-based measures record cash income received directly by individuals, while the PSID records cash and in-kind income collected on behalf of individuals.

The by versus on behalf of distinction means that dividends, interest and rents received on behalf of individuals by pension plans, nonprofits and fiduciaries is in PSID income but not survey income. The cash versus cash plus in-kind distinction means that employer contributions for employee pension and health insurance funds are in NIPA income, but not survey income. Similarly the PSID includes (but the surveys exclude) the imputed rental value of owner-occupied housing and in-kind transfers such as Medicare, Medicaid and food stamps. In the other direction, the surveys include but the PSID excludes personal contributions for social insurance, income from private pension and annuities plans, income from government employee retirement plans, and income from interpersonal transfers, such as child support. Table 1 in Ruser et. al. (2004) provides a careful and detailed account of these differences. They find that in 2001, 64 percent of the $2.23 trillion gap between aggregate NIPA personal income and aggregate CPS money income can be accounted for by differences in income types across the two measures (see also Roemer, 2000).

In addition to these conceptual differences, one should expect an additional gap between

PSID and survey-based estimates because of the practicalities of how the BEA constructs its estimates. In addition to the under-ground economy adjustment for wages and salary discussed above, the BEA also adjusts for an estimate of proprietors income earned but not reported on tax returns (the basic source for NIPA estimates). This adjustment is based on evidence that proprietors' actual income in 1999 was more than double levels reported on tax returns. Ruser et al. note that it is likely that respondents who underreport to the IRS also underreport in voluntary surveys, though the CPS, PSID and CEX do not attempt to adjust for underreporting. Comparing various other components of income across the CPS and other independent estimates, Ruser et al. note that under-reporting in the CPS seems to be important for private and government retirement income, interest and dividend income, and social security income. One implication of this under-reporting for most types of unearned income is that in our sample, wages and salary income (not counting any part of self-employment income) accounts for fully three quarters of total per-capita income, averaged over the sample period.

On average across the sample period, CPS income falls 21 percent short of NIPA income. The gap widens over time, by around 10 percentage points of NIPA income. The episodes in which the NIPA-CPS income gap widens most rapidly do not correspond to changes in the set of income questions, or to changes in imputation procedures. A large part of this widening gap can be attributed to rising employed contributions for employee pension and social insurance funds, which do not appear in CPS income: these were 4.3 percent of NIPA personal income in 1967, and 9.0 percent in 2005.

The levels of per-capita income in the CPS, PSID and CEX are very similar. The PSID exhibits somewhat faster growth over the 1969-2002 period, such that the NIPA-PSID differential is broadly stable. Compared to the corresponding NIPA and CPS measures, CEX per capita income grows somewhat more slowly in the 1980s, at a very similar rate in the 1990s, and more rapidly in the 2000-2005 period.

Our CPS estimate for per capita income exceeds that in the historical series published by the Census. In part, this is because our approach to dealing with top-coded observations addresses the problem associated with internal censoring. The gap between our CPS-derived estimate for per capita income and the official Census series narrows from over 7 percent at the start of the sample period, to less than 1 percent at the end. Between 1992 and 1993, when the Census internal censoring point for primary job rose from $300,000 to $1m, the gap narrows from 5.3 percent to 2.5 percent.

The CPS, PSID and CEX all capture the business cycle fluctuations evident in the NIPA income series. However, cyclical fluctuations appear larger in the CPS than in the NIPA. From peak to trough, percentage real income declines in the CPS (NIPA) for the recessions in the mid 70s, early 80s, early 90s and early 00s are 3.9 (2.2), 6.6 (2.9), 5.1 (2.3) and 2.2 (1.3). While recession declines in per-capita pre-tax income are roughly twice as large in the CPS, declines in wages and salary are very similar in magnitude. Thus the difference in business cycle dynamics must be attributed either to definitional differences or to differential measurement error in unearned components of income. Either way these differences raise important questions about the welfare costs of business cycles, since aggregate income is likely a better proxy than aggregate earnings for aggregate consumption.

Figure 2 reports measures of aggregate consumption for the CEX and PSID and contrasts them with comparable aggregates from NIPA. The first panel reports aggregate expenditure on food (including alcoholic beverages and food away from home) for PSID, CEX and NIPA. The figure confirms that food expenditures in the CEX and PSID track each other rather well (see Blundell Pistaferri and Preston, 2004 for a similar finding in the earlier part of our sample) but also confirms that they are both much lower than food expenditures from NIPA and that the gap between the two series is increasing over time at a rapid pace. This growing discrepancy is evident, even more markedly, for a broader definition of non durable consumption (the northeast panel), and for expenditures on durable goods and for housing services (the bottom panels), although to a lesser extent. Few researchers have investigate the causes of this growing discrepancy (see for example Slesnick, 2002 or Garner et. al 2006) but a conclusive answer has not yet been provided. What makes this discrepancy particular puzzling is that it seems to be very specific to the US consumption case. For example in UK or Italy (see the articles in this volume) there is a remarkable degree of co-movement between the aggregate consumption from survey and the one from NIPA, while figure 2 above has shown that the income data from CEX tracks NIPA income quite well.

Figure 3 plots average wages and hours over the sample period. The estimates of average hours in Figure 3 use all 25-60 year-old individuals in Sample B, including those working zero hours. Average wages apply to Sample C, which excludes all individuals working less than 260 hours in the year. Wages are computed as annual pre-tax earnings divided by annual hours, where earnings includes a fraction (2/3 in CPS and CEX, 1/2 in PSID) of self-employment income. Prior to income year 1975, CPS information on hours - and thus wages - is not ideal for two reasons: the question about weekly hours refers to hours worked last week (rather than usual weekly hours), and information about weeks worked in the previous year is available only in intervals. We have used information for years in which both measures of hours are available to splice together estimates for the 1967-1974 period and those for the later period.

All three datasets show substantial and similar real wage gains for women. Average male wages grow somewhat in the PSID, but are broadly stable in the CPS, with losses in the late 1970s recouped in the late 1990s. Business cycle fluctuations are evident in all average wage series, while the PSID exhibits an upward blip for both men and women in 1992-1993.

Average male hours decline somewhat over the sample period, especially in the CPS. In contrast, the bottom right panel of Figure 3 shows a dramatic secular rise in female labor force participation that stalls towards the end of the sample period. Average male hours in the CEX are consistently lower than in the CPS, while male hours in the PSID are consistently higher. The PSID-CPS discrepancy presumably reflects differences in survey design: in particular, the PSID collects information only on male heads-of-household, while the CPS includes all men in the household. Our CPS estimates align very closely by year and age group with the decennial Census-based estimates of McGrattan and Rogerson (2004, Table 3).

## 3.2 Individual-level inequality

**Wages** We begin our discussion of individual-level inequality from wages. Figure 4 displays four measures of dispersion in hourly wages for the entire working-age population, i.e. men and women aged 25-60. Recall that we condition on individuals who work at least 260 hours per year, with wage at least half the legal Federal minimum wage. The variance of log hourly wages and the Gini coefficient of hourly wages show a similar steady upward trend throughout the period. The only deceleration in the rise of wage dispersion seem to occur in the 1970s when, as we will document below, the education wage premium shrinks.

PSID, CPS and CEX are in broad agreement, even cyclical movements around the trend are somewhat synchronized. The trend in the PSID data seem slightly steeper, and the level of inequality is higher throughout, but only by just about one point in both measures. Quantitatively, the rise in wage inequality is substantial. The variance of log wages rises by 15-16 log points, and the Gini by around 8 log points. A unique feature of CEX data is the fast increase in inequality after 2003.

Turning to the percentile ratios, we uncover different trends in the top half and bottom half of the wage distribution. The 90th-50th percentile ratio (P90-P50) confirms the steady upward trend in inequality throughout the sample period: in both CPS and PSID, the ratio increases from 1.9 to 2.3 over the period 1975-2005. CEX data line up well with the other two surveys.

The dynamics of inequality in the lower half of the wage distribution are somewhat different. In CPS, the rise in the 50th-10th percentile is concentrated in one decade, the 1980s, whereas PSID data show also an increase in the 1970s. However, both data sets show no change, or even a small decline, since the early 1990s. CEX data show the same qualitative pattern, with a less severe trend break in the 1990s.[8] Thus, while the rise in US wage dispersion in the 1980s is visible throughout the distribution, in the 1990s it is concentrated in the top half.

There is a large empirical literature documenting the evolution of cross-sectional wage inequality in the United States since the mid 1960s. The two most recent and comprehensive surveys are Katz and Autor (1999), and Eckstein and Nagypal (2002). A more up to date account is provided by Autor, Katz and Kearney (2008).[9] All these papers are based on CPS data, and focus only on full-time, full-year workers, i.e. individuals who work at least 35 hours per week and fourty-plus weeks per year. Our analysis is based on a much broader sample, given the more inclusive criterion on hours worked. Nevertheless, the qualitative trends we uncover are very similar to these previous studies. A unique contribution of our study is to document that measured changes of the wage structure in CEX and PSID line up very well with those in the larger CPS sample.

A debate developed recently on whether the rise in US inequality was mostly an episodic event of the 1980s which plateaud by the end of the decade and never recurred (Card and DiNardo, 2002; Lemieux, 2006) or, rather, a long-term trend towards a wider wage structure (Autor, Katz, and Kearney, 2008) that started in the 1970s and is still ongoing. This different view is grounded in the different data used. The "episodic" interpretation of widening wage dispersion is based on the May Outgoing Rotation Group (ORG) samples of CPS which has point-in-time measures of usual hourly wages. The "long-run" interpretation is based on March CPS, the data we use as well, where hourly wages are constructed as annual earnings divided by annual hours worked.

[8] Lee (1999) and Card and DiNardo (2002) claim that the US federal minimum wage has a large impact in shaping the bottom of the wage distribution. The minimum wage was stable at around $6.50 (in 2000 dollars) until 1979, then it declined steadily for about a decade until $4.50, and then slowly increases again since then. If plotted together, the P50-P10 ratio and the (inverse) of the minimum wage comove very closely.

[9] Historically, the widening of the US wage structure during the 1980s was first documented by Bound and Johnson (1992), Katz and Murphy (1992), Murphy and Welch (1992), and Juhn, Murphy and Pierce (1993).

Interestingly, we find that both PSID and CEX give support to the "long-run" view.[10] In particular, we uncover that, even in the bottom half of the distribution, where the rise in dispersion was concentrated in the 1980s, after 2000 there seems to be a renewed growth in inequality in all three data sets.

**Observables and residuals** In order to understand the sources of the rise in US wage inequality, it is important to distinguish the role of some key observable demographics such as education, age and gender. We perform this decomposition in Figure 5. We define the male education premium as the ratio between the average hourly wage of male workers with at least 16 years of schooling to the average wage of male workers with less than 16 years of schooling. The pattern emerging from PSID and CPS is the well documented U shape: following a decline until the late 1970s, the college wage premium starts rising steadily. In 1975, US college graduates earned 40% more than high-school graduates, while in 2005 they earned 90% more.

In the US, the fraction of 25+ years old men who completed college rose steadily from 13% in 1967 to 29% in 2005 (US Census Bureau). A vast literature argues that the trends in relative quantities and relative wages of college educated workers since 1980s can only be reconciled through a skill-biased demand shift which economists have associated to the technological shift towards ICT and to globalization (e.g., Acemoglu, 1999; Katz and Murphy, 1992; Krusell et al., 2000; Hornstein et al., 2005).[11]

The experience wage premia plotted in Figure 5 is defined as the ratios between the hourly wage of 45-55 years old to the hourly wage of 25-35 years old. PSID and CPS show an increase for both men and women. The increase for males is substantial: the experience premium more than doubles (from 20% to 40%) since 1975. The increase for women is smaller and occurs somewhat later. An interpretation of these trends based on relative quantities alone goes a long way, at least qualitatively. Due to the "baby-boom generation" which entered the labor market in the early 1970s, the relative labor supply of the 25-35 years old starts increasing, and their relative price decline. After roughly 20 years, i.e., around 1990, this large cohort starts entering the 45-55 group and depresses the experience premium, as evident from the data. Eckstein and Nagypal (2002, Figure 15) plot the coefficient on experience from a standard

[10] Also in CEX and PSID our measure of hourly wage is constructed as annual earnings divided by hours worked last year.

[11] Eckstein and Nagypal (2002) and, more recently, Lemieux (2006) document that the premium for post-graduate education increased even faster.

Mincerian wage regression and find a pattern very close to ours: the experience premium for women is much lower than for men, and for both sexes it rises in the 1970s and 1980s and stabilizes in the 1990s.

The plot of the gender wage premium in Figure 5 shows a stable gap throughout the 1970s, and then a sharp narrowing of the gap until the early 1990s. CPS data reveal that men were paid on average 65% more than women in 1975, and they only earned 30% more in 2005. Overall, PSID and CPS are in agreement, except for the last decade. While the gender gap in CPS keeps declining, albeit at a slower rate, the PSID data reveal a rebound.[12]

The last panel of Figure 5 contains residual male wage inequality, measured as the variance of the log wage residuals from a regression on standard demographics.[13] Residual wage dispersion rises throughout the period. A comparison with the variance of "raw" wage inequality for men (see the first panel of Figure 6 below) reveals that residual inequality explains the bulk of the increase in cross-sectional wage inequality in the 1970s, but only about 60% of the rise since 1980 – the rest being explained by observable characteristics, specifically education and experience.

**Labor supply** The first panel in Figure 6 plots male and female wage dispersion, measured by the variance of log hourly wages, conditional on working at least 260 hours, and earning hourly above half the minimum wage. Both sources of data show a strong rise in female wage inequality as well. However, the rise for women starts later –in the early 1980s– and is somewhat less pronounced than the one for males, overall. In CPS, female wage dispersion grows by 0.16 log points, compared to 21 log points for males, over 1967-2005. Eckstein and Nagypal (2002, Figure 3) report similar findings.

The second panel of Figure 6 plots the variance of log hours, conditional on working at least 260 hours. Here, CPS and PSID tell a virtually identical story. The variance in log hours worked has been stable for men, around 0.12. Interestingly, both series show that inequality in male hours is sharply counter-cyclical.

Hours dispersion, instead, has declined sharply for women: the variance falls from 0.28 to 0.20. As women become gradually more attached to the labor force, increase their hours worked

[12] The PSID sample is more than 10 times smaller than CPS, and the series are more volatile. Only a few more years of data can confirm whether this discrepancy is due to short-term volatility, or to differential long-term trends.

[13] For regressions based on PSID data, the regressors are dummies for family, education, and race composition of the household, and a quartic in heads' age. The exact specification of the CPS regression is reported in the Appendix.

(recall Figure 3), and look more and more like men in that many of them hold full time jobs –which explains the convergence in hours inequality.

The last two panels plot the correlation between hours and wages in CPS and PSID for men and women. Both CPS and PSID show a sharp rise in the male correlation which peaks in the mid 1980s and then flattens out. The rise is more pronounced in CPS, and its level is higher. In 2000, this correlation is roughly 0.10 in CPS and -0.03 in PSID. Recall that hourly wages are computed as earnings divided by hours, so measurement error in hours and earnings, pervasive in survey data, tends to bias this correlation. Reporting error in earnings biases the correlation towards zero, whereas error in hours biases it downward. One interpretation of the plot is that there is more measurement error in PSID.

Except for the first few observations in the late 1960s, CPS and PSID line up also in the dynamics of female wage-hours correlation. Similarly to men, the female correlation rises during 1975-1985, but then it declines significantly at a time when the male wage-hours correlation levels off. On average, over the period, wages and hours are more correlated for women than for men.

## 3.3 Household-level inequality

**Evidence from CPS** The top panel of Figure 7 plots three versions of the variance of log household earnings in CPS: non-equivalized, equivalized, and residual, where the residuals are computed from the same regression used in Figure 5.[14] The equivalization reduces slightly the level, but it has no impact on the trend of the variance which grows until the early 1990s by roughly 30 log points, and then levels off thereafter. The figure shows that key demographic characteristics of the household explain about 40% of the variance. Consistently with what we observed for wages, the growth in residual inequality accounts for most of the early increase in the raw variance and little thereafter.

Figure 7a provides more detail on the evolution of the distribution of equivalized log household earnings in CPS, during 1967-2005. The top panel plots densities for decades 1970, 1980, 1990 and 2000, between the normalized values of 2 and -2.[15] The rise in household earnings dispersion is very evident from how the shape of the distribution changes across decades. Note

[14]In computing household-level inequality statistics, we trimmed the bottom 0.5% of the distribution because we noticed that the variance of the logs is very sensitive to low earnings outliers. For consistency, we compute all the other measures of inequality (Gini, and percentile ratios) on the same trimmed sample.

[15]The densities are normalized so that mean earnings equal to zero in each decade.

also that the distribution is notably left-skewed, hence far from Normal. We return on this point in Section 3.4.

The bottom panel of Figure 7a plots the trends in percentiles at different points in the distribution (all normalized to zero in 1967), together with the shaded areas denoting NBER recessionary periods. The panel shows clearly the fanning out of the distribution, particularly strong until 1990. Moreover, it appears that household earnings of the income-poor decreased in real terms for long periods of time: according to CPS, in 2000 households below the 10th percentile earned as much as they earned in 1970, in real terms. Finally, the plot shows that household earnings are procyclical at each percentile, but business cycle fluctuations are much more severe at the bottom of the distribution, where fluctuations in household earnings are grater in percentage terms.

**CPS-PSID comparison** We now turn to the analysis of the distribution of household earnings in PSID. The bottom panel of Figure 7, based on PSID data, paints a picture which is at odds with CPS. The non-equivalized variance has a similar trend to the CPS data, but the overall size of the increase is smaller –about 2/3 of that in CPS. The equivalization does not just shift the series, as in CPS, but it makes the series even flatter, amplifying the discrepancy with CPS. Isolating the effect of observable demographics, instead, has a similar impact in the two data sets.

Figure 8 sheds further light on this important discrepancy between data sets.[16] The bottom two panels of Figure 8 make clear that CPS and PSID are in disagreement on the dynamics of inequality at the bottom of the earnings distribution, in particular below the 25th percentile. The trend in the P90-P50 ratio, instead, is remarkably similar across data sets. A comparison of Gini and variance of the log reveals an interesting finding: the evolution of the Gini is much closer to that of the P90-P50 ratio, whereas the trend in the variance of log household earnings follows closely the P50-P10 ratio. Qualitatively, this is not surprising, since that the logarithmic function exacerbates the role of small earnings values, but the quantitative impact is strikingly large.[17]

In Section 2, we have discussed some important differences between PSID and CPS. First of all, a household in PSID is really a family unit, while a household in CPS can include two

[16] To retain the same sample across all these different measures of inequality, we always drop those observations with zero earnings which are missing, by construction, for the variance of logs. After dropping these observations, we trim at the bottom 0.5% of the distribution.

[17] Recall that we do trim the top and bottom 0.5% of the distribution of equivalized earnings.

or more families. This suggest that the trends in "household equivalence scale" could diverge, and this can account for part of the discrepancy in equivalized earnings inequality. Second, the empirical literature on poverty, finds consistently that poverty rates are higher in CPS (see, for example, Duncan and Rodgers, 1991, Figure 1). And thus, plausible sources for the differences we have uncovered are both the dynamics of the equivalence scale, and the accuracy in representing the bottom of the (unequivalized) household earnings distribution.

Figure 8a plots some useful statistics on family composition in the two data sets. While, as expected, the average equivalence scale (hence household size) is somewhat larger in CPS, the trends in the two data sets are closer and show a dramatic decline in household size over the period (see also Table 2). However, the variance of the equivalence scale declines much faster in PSID in the 1970s, exactly over the period where the trends in the variance of equivalized earnings disagree the most. From Figure 8a, it is clear that the discrepancy arises because of the variance of the number of adults, rather than children. This finding is related to the discrepancy in the number of households with spouse present documented in Table 1.

Figure 8b plots the 5th, 10th, and 25th percentiles for unequivalized household earnings in the two data sets, normalized to CPS values in 1967. The differences between trends in the percentiles gets more sizeable as one moves towards the bottom of the distribution. Overall, it seems that even before equivalization, PSID and CPS reveal a contrasting picture of the very bottom of the household earnings distribution which, in turn, has a large impact on the trend of the variance of log household earnings.

Figure 9 displays the evolution of various measures of cross-sectional household income dispersion, beginning with labor earnings for the main earner and moving to increasingly broad measures of household income. For each type of income, moments are computed for the same set of households: households in Sample B that also have positive household earnings.[18] The differences in the dynamics for earnings inequality between the CPS and the PSID noted in Figure 7 extend to broader measures of income, with the CPS indicating a larger rise in inequality overall, and particularly in the 1970s. However, the CPS and PSID pictures are remarkably similar in terms of what they say about how broadening the measure of income affects overall inequality.

Moving from earnings for the main earner to total household earnings increases the variance

[18]For each measure of equivalized income, we trim the bottom 0.5 percent of observations according to that particular definition of income. .

of log earnings. Thus low main-earner earnings does not signal the existence of compensating earnings from other household members. Adding private transfers reduces income inequality in both datasets. This reflects the fact that households containing retirees tend to have lower earnings, but higher private retirement income. Adding asset income, by contrast, has little impact on overall income dispersion, and actually increases inequality towards the end of the sample period. Public transfers play a very important role in compressing inequality, especially in the CPS. Interestingly, however, the income compression associated with public transfers declines in the 1990s in both datasets. Finally, the tax code also appears to be quite progressive, reducing the variance of income by around 10 log points.

Between 1967 and 1980, the CPS data indicate that private and public transfers worked to substantially offset the impact on pre-tax income inequality associated with rising inequality in household earnings: the variance of log equivalized household earnings increased by almost 20 points, while the variance of pre-tax income rose by less than 5 points. From 1980 to 2005, however, the variances of household earnings, pre-government income, and disposable income all increased by similar magnitudes. We plan to explore whether the correct interpretation of these results is that government tax and transfer policy works to compress income inequality in the short run but not in the long run, or alternatively whether the amount of redistribution implicit in the tax and transfer system has declined over time.

## 3.4 Life-cycle profiles of inequality

The age profiles of inequality in wages, hours, earnings and consumption are key inputs to calibrate and analyze life-cycle heterogeneous agents models with incomplete markets (see Storesletten, Telmer and Yaron, 2004; Guvenen, 2007; Huggett, Ventura and Yaron, 2008).

The key statistical problem in isolating the age profile from repeated cross-sections is that age, time and cohort are linearly dependent, and hence one can either fully control for time or cohort effects. For a long time, the standard approach in the literature has been the one advocated by Deaton and Paxson (1994) in their seminal article on life-cycle consumption inequality, where the regression includes a full set of cohort dummies, whereas time dummies are constrained to capture only business-cycle effects, but no long-run trends, i.e. they are forced to sum to zero. However, recently, Heathcote, Storesletten and Violante (2005) argued that the choice whether fully controlling for time or cohort effects can significantly change the estimated age profiles. Here, we use both approaches and report both set of findings.

Let $m_{a,c,t}$ be a cross-sectional moment of interest (e.g., the variance of log earnings) for the group of households, or individuals, of age $a$ belonging to cohort $c$ (hence, observed at date $t = c + a$). To isolate the age profile, we run the two alternative regressions

$$\begin{aligned} m_{a,c,t} &= \boldsymbol{\beta}_a'\mathbf{D}_a + \boldsymbol{\beta}_t'\mathbf{D}_t + \varepsilon_{a,c,t} \\ m_{a,c,t} &= \boldsymbol{\beta}_a'\mathbf{D}_a + \boldsymbol{\beta}_c'\mathbf{D}_c + \varepsilon_{a,c,t}, \end{aligned} \tag{1}$$

where $\mathbf{D}_t, \mathbf{D}_c$ and $\mathbf{D}_a$ are vectors with entries corresponding to a full set of time (year), cohort and age dummies, respectively. The vectors $\boldsymbol{\beta}_t$, $\boldsymbol{\beta}_c$ and $\boldsymbol{\beta}_a$ are the corresponding vector of coefficients. The lines labelled "year effects" plot the estimated values of $\boldsymbol{\beta}_a$ in the first regression where we control for year effects, and the lines labelled "cohort effects" plot the estimated values of $\boldsymbol{\beta}_a$ in the second regression, where we control for cohort effects only.[19]

Figure 13 is based on CPS data. The key finding is that for both wages and earnings, when controlling for time effects, the life-cycle profiles of inequality are remarkably less steep than when controlling for cohort effects. For example, the variance of log wages grows by 20 log points from age 25 to age 60 with year effects, and by 35 log points when controlling for cohort effects. We find differences of the same order of magnitude for household earnings. Whether one controls for time or cohort effects in the regressions does not seem to matter for hours dispersion, whose age profile is U shaped, but much flatter than for income (note the different scale).

The curvature of the life-cycle profile is also of interest, since it can shed light on the nature of the underlying stochastic process generating inequality. Wage inequality grows roughly linearly, perhaps showing a slightly concave profile, whereas non-equivalized household earnings inequality features a convexity which is clearly inherited from the life-cycle path in the dispersion of hours worked.

A linear or concave profile is consistent with an income process with unit root, or with a very persistent autoregressive component. Convexity has been often indicated as evidence of "heterogeneous income profiles" (Lillard and Weiss, 1979; Baker, 1997; Guvenen, 2007).[20] The fact that the convexity is not present in wages, arguably a more exogenous income concept, but only inherited by earnings through hours worked, arguably an endogenous choice of individuals, suggests that it can be misleading to use this aspect of the data to make a case for profile

[19] In Figures 13 and 14, the values of $\beta_a$ for the first age group are always normalized to zero.

[20] See Guvenen (2007) for a formal explanation of why the model with heterogeneous income profiles can generate a convexity in the variance of earnings over the life cycle.

heterogeneity. A more promising approach would be to estimate an income process (with or without profile heterogeneity) on wage data, and then through a labor supply model generate endogenously the convexity in earnings inequality.

Finally, we note that the equivalization has also a remarkable impact on the life-cycle profile of earnings inequality. Equivalized earnings dispersion grows 50% less than its non-equivalized counterpart, over the age range 25-50. Equivalization matters much less after age 50.

Figure 14, based on PSID data, confirms virtually all the findings from CPS. Both the magnitude of the increase in life-cycle inequality, and the shape of the profiles are very close to those estimated on CPS data. Overall, we confirm the findings in Heathcote, Storesletten and Violante (2005) that, when controlling for time effects, inequality grows more slowly over the life-cycle.

## 3.5 Tests for log-normality of income and consumption

**TBC**

## 3.6 Estimation of income dynamics

There is a long tradition in labor economics in estimating structural models of income dynamics from panel data (starting, perhaps, from Lillard and Willis, 1978; Lillard and Weiss, 1979; MaCurdy, 1982). Recently, this literature spilled over into quantitative macroeconomics. Individual-level, or household-level, income dynamics are a key ingredient of the calibration and estimation of heterogeneous-agent incomplete-market models (e.g., Imrohoroglu, 1989; Huggett, 1993; Aiyagari, 1994; Rios-Rull, 1996).

In this section, we use PSID data from 1967-1996, 1998, 2000 and 2002 to estimate the dynamics of individual wages and equivalized household earnings. Wage dynamics are useful in the parameterization of models with endogenous labor supply decisions, and household earnings in models with inelastic labor supply.

As common in the literature, we focus on "residual" inequality, i.e., log wages and earnings residuals from a standard Mincerian regression. We use the same specification chosen to produce Figures 5 and 7. These figures show clearly that the variance of residual inequality grew over this period, hence the model to be estimated must allow for non-stationarity of the parameters, a well-known fact since Gottschalk and Moffitt (1994).

**Statistical models** Let $y_{i,a,t}$ be the log hourly wage (equivalized annual earnings) for individual (household) $i$ of age $a$ at date $t$. The most general model we estimate is the composite of an AR1 and an MA1 component,

$$\begin{aligned} y_{i,a,t} &= z_{i,a,t} + \varepsilon_{i,a,t} + \theta\varepsilon_{i,a-1,t-1} \\ z_{i,a,t} &= \rho z_{i,a-1,t-1} + \eta_{i,a,t} \end{aligned} \tag{2}$$

where $\varepsilon_{i,a,t}$ and $\eta_{i,a,t}$ are uncorrelated over time, i.i.d. across individuals, and orthogonal to each other. We denote $var_t(\varepsilon_{i,a,t})$ as $\sigma_{\varepsilon t}$, and $var_t(\eta_{i,a,t})$ as $\sigma_{\eta t}$. Moreover, we assume that at age $a = 1$, $var(z_{i,1,t}) = \sigma_{z0}$. In other words, we allow time variation in the variances of the AR1 and MA1 shocks, but we fix the coefficients $\rho$ and $\theta$, as well as the variance of the initial persistent component (this approach is standard in the literature). We denote this model as AR1MA1.

Other statistical models that we estimate are all particular cases of (2). With $\rho = 1$, the model in (2) becomes a permanent-MA1 (PMA1) model. With $\theta = 0$, it becomes an AR1-transitory (AR1T) model. When $\rho = 1$ and $\theta = 0$, it becomes a simple permanent-transitory (PT) model.

**Methodology** In the literature, there are two approaches to the estimation of income processes. The first, diffused in labor economics (e.g., Abowd and Card, 1989; Meghir and Pistaferri, 2004), uses moments based on income growth rates –or first-differences in log income. The second, more common in macroeconomics (e.g., Storesletten, Telmer, and Yaron, 2004; Guvenen, 2007; Heathcote, Storesletten and Violante, 2008) uses moments in log income levels.

The two approaches diverge in the set of moments that yield identification. For example, consider a stationary version of the simple PT model. In first-differences, the model is estimated based on the following restrictions:

$$cov(\Delta y_t, \Delta y_{t-j}) = \begin{cases} \sigma_\eta + 2\sigma_\varepsilon & j = 0 \\ -\sigma_\varepsilon & j = 1 \\ 0 & j > 1 \end{cases} \tag{3}$$

Therefore, the covariance in growth rates at one lag ($j = 1$) identifies the variance of the transitory shock, and the variance of growth rates identifies, residually, the variance of permanent shocks. The model is just identified.[21]

[21]It is easy to see that with the MA1 component, we add one additional parameter $\theta$ and one moment, since now the covariance term is non zero for $j \leq 2$.

In levels, the model is estimated based on moment restrictions of the type:

$$cov\left(y_{a,t}, y_{a-j,t-j}\right) = \begin{cases} \sigma_{z0} + (a-1)\,\sigma_\eta + \sigma_\varepsilon & j = 0 \\ \sigma_{z0} + (a-j-1)\,\sigma_\eta & 0 < j < a \end{cases}$$

and therefore, the covariance at lag $j = 1$ for the age group $a = 2$ identifies $\sigma_{z0}$. The difference between the covariance at lag $j > 1$ (age $a > 1$) and lag $j - 1$ (age $a - 1$)–i.e., the slope of the autocovariance function– identifies $\sigma_\eta$. Finally, the drop in the autocovariance function between lag $j = 0$ and lag $j = 1$ identifies $\sigma_\varepsilon$. Note that the model is overidentified, since moments can be constructed for different cohorts/age groups.[22]

Both approaches have strengths and weaknesses. On the one hand, the estimation in levels allows also identification of the variance of the initial conditions $\sigma_{z0}$ which, by construction, cannot be identified with growth rates alone. Moreover, moments in first-differences are obtained as linear combinations of moments in levels –whereas the opposite is not true– which means that moments in first-differences are implicitly used in the level estimation, even though the importance given to them depends on the weighting matrix used. On the other hand, the estimation in first-differences has the advantage of being robust to cohort effects, whereas the identification strategy outlined above for the levels estimation assumes away, but it does not hinge upon, cohort effects.

Below, we perform the estimation using both approaches and compare findings. Throughout the analysis, we use a minimum distance estimator (Chamberlain, 1984), with moments weighted by the identity matrix –as suggested by the Monte-Carlo study of Altonji and Segal (1994)– the standard methodology in the literature.

**Results for hourly wages** The results on hourly wages are presented in Table 3. The first three columns report results for the estimation in first-differences, and the last four report results for the estimation in levels.

The PT model estimated in first-differences yields average estimates of $\sigma_\varepsilon$ around 0.051 and of $\sigma_\eta$ around 0.036. Moments based on first-differences are particularly sensitive to outliers and measurement error. This problem especially affects estimates of $\sigma_\varepsilon$ which captures transitory changes. When we excluded all observations with log-change greater than one in absolute value,

[22] The identification of the non-stationary model is somewhat more complicated, but it follows the same logic. See Blundell, Pistaferri and Preston (2008) for a discussion of identification of the non-stationary version of the model based on moments in first-differences. Heathcote, Storesletten and Violante (2008) list the necessary assumptions for identification of the non-stationary version of the model with moments in levels, and missing years (like PSID after 1997). Here we follow their approach.

we ended up with a variance of transitory shocks about half as large, while the estimate of $\sigma_\eta$ decreased only slightly. In what follows, we present results for this "trimmed" sample.

Adding an MA1 component –with $\theta$ estimated at 0.11– increased somewhat the transitory variance and further reduced the permanent variance to 0.026.

The estimation of the PT model in levels yields sharply different conclusions from the estimates in growth rates. The permanent variance is estimated to be three times smaller ($\sigma_\eta = 0.008$), and the transitory variance three times larger ($\sigma_\varepsilon = 0.09$) . Adding an MA1 component reduces the transitory variance, i.e., it has the opposite effect relative to the estimates in growth rates, since $\theta$ is now estimated to be large and negative ($\theta = -0.64$) rather than small and positive. Allowing for autocorrelation lower than one leads to a point estimate of $\rho$ around 0.98 and, naturally, to a slightly larger persistent variance $\sigma_\eta$.

Figure 15 (top two panels) visualizes the time trends in the estimated variances of shocks to log wages. To make the trends sharper we plot 5-year moving averages. Both types of estimations agree on the rise of the permanent/persistent variance in the 1970s and the rise in the transitory variance in the late 1990s. However, the estimation in levels also detects a rise in the permanent variance in the late 1990s, and a sharp rise in the transitory component until the mid 1980s which are not present in the growth-rate estimation.

The divergence in the estimates of $(\sigma_\eta, \sigma_\varepsilon, \theta)$ obtained through the two approaches are striking. Which model provides a better fit? To answer this question we have isolated four key empirical moments: the age profile of the variance of log wages, the autocorrelation function of log wages, the variance of wage growth rates, and the covariance of wage growth rates at one lag. Figure 16 plots the fit of the PT model in first-differences and Figure 17 the fit of the AR1MA1 model in levels.

Because of the exact identification, the estimation in first-differences perfectly fits the last two moments. However, the model predicts a rise in wage inequality over the life-cycle which is completely implausible – roughly five times as large as in the data. Also the fit of the autocorrelation function in levels is modest at best. The low estimated transitory variance is reflected in the small initial drop, and the large permanent variance in the excessively steep downward trend.

Turning to Figure 17 (where the same scale as in Figure 18 is used), we note that the model can reproduce very well the growth of wage inequality over the life-cycle and the entire autocorrelation function for log wages. As expected, the fit is less satisfactory for moments in first differences, although the gaps between model and data are not as spectacular as for those documented in Figure 16.[23]

If the PT or PMA1 models are the true data generating process, then the only explanation for the discrepancy in the estimated coefficients in Table 3 would be that the small-sample bias operates differently across the two set of moments. Alternatively, it is possible that the PT or PMA1 models are misspecified (and, for example, the true model has $\rho < 1$ and $\theta < 0$) and this misspecification bias is much more severe for the growth-rates estimation. Further work should investigate these two hypotheses.

**Results for household earnings** The results on household earnings are presented in Table 4, with the trends in the variances plotted in the bottom two panels of Figure 15. All the findings documented above remain true qualitatively also for equivalized household earnings.

# 4 Conclusions

`TO BE ADDED`

[23]It should be noted that the line denoted as "data" in the two bottom panels of Figures 16 and 17 is not exactly the same. The reason is that in the estimation in levels, the objective function is unweighted, hence the model tries to match an unweighted mean of cohort-specific covariances of growth rates at zero and one lags. In the estimation in first-differences, the covariances are constructed as weighted sample means.

# A CPS

**Survey description.** Currently around 55,000 households are interviewed monthly out of around 60,000 eligible housing units sampled. Each household is interviewed once a month for four consecutive months one year, and again for the corresponding time period a year later. The Annual Social and Economic Supplement (ASEC) formerly known as the Annual Demographic Survey applies to the sample surveyed in March, and extends the set of demographic and labor force questions asked in all months to include detailed questions on income. For the ASEC supplement, the basic CPS monthly sample of around 60,000 households is enlarged by adding an additional 4,500 hispanic households (since 1976) and, more importantly, by the addition of 34,500 extra households in 2002 as part of an effort to improve state estimates of children's health insurance coverage: this sample expansion is known as the SCHIP sample. Some of these extra households are interviewed in February or April rather than March. The ASEC sample universe also differs from the standard CPS in that it includes military personnel living in households with at least one civilian adult. Including the basic CPS sample, approximately 98,000 housing units were in sample for the 2007 ASEC, of which 83,200 were determined to be eligible for interview, leading to about 76,100 interviews obtained.

There have been a succession of changes over time in the March CPS involving the sample construction, interview methods, data processing and imputation methods, weighting (reflecting new decennial Census population counts), and the structure and content of the questions themselves. More detailed questions about income were asked beginning with the 1976 survey, and the set of questions was expanded again in 1988.

For March 1988 two files are available: the regular and the rewrite file, which includes revised procedures for weighting and imputations (a previous change to the imputation procedure occurred in 1976). We use the rewrite file, which is recommended for comparison with future years. Two files are also available for 2001: including or excluding the SCHIP sample expansion. We use the smaller sample. The largest changes in the basic CPS survey methodology came in 1994, with the introduction of computer-assisted interviewing, and associated redesign of the questionnaire.

Notwithstanding these and other changes, the basic structure of the March CPS has remained remarkably intact over time.

We use the March supplement weights to produce our estimates. Weights are chosen to make the CPS sample representative of the US population, and apply at the individual level. For household level variables, we use the household weight, which is equal to the family weight of the household reference person, which is the reference person's weight, unless the reference person is a married man in which case it is the weight of his wife. The supplement weights differ from the usual monthly CPS weights, reflecting differences in the sample, particularly the inclusion of the SCHIP subsample. For individual level variables we use individual weights, which can differ across individuals within a household because different household members have different demographic characteristics (age, sex, race, ethnicity) which are inputs to the CPS weighting procedure.

**Sample selection**. Our basic sample selection strategy is outlined in the text: here we describe the details of how this applies to the CPS. To generate our Sample A, the cleaned version of the entire dataset, we start by dropping households that do not have a reference person, or that have more than one reference person (there are no such households from income year 1993 onwards). We then drop households in which there are household members with negative or zero weights (there are only a handful of such households from 1975 onwards). Next we drop households in which there are members with positive earnings but zero weeks worked (there are no such households from 1989 onwards). Next we drop households in which there is an individual whose hourly wage is less than half the legal minimum in that year. To apply a consistent sample selection rule across the whole sample period, we define the hourly wage here using the hours worked last week variable, which is available throughout the sample period (see below). There are no missing values for variables in the CPS, since missing values are imputed (see below). We do not exclude observations with imputed values, even if all income variables are imputed. This defines the basic "NIPA" sample used for comparison with BEA estimates of income in Figure 1.

Sample B, the starting point for measuring inequality among the population of working age households, is Sample A less all households in which there are no individuals aged between 25 and 60, inclusive. A minor difference relative to the PSID is that, since we have income data for all household members, the CPS version of Sample B retains households as long as any household member falls in the 25-60 age range, even if both the CPS reference person and their spouse fall outside the range. The CPS estimates of average hours in Figure 3 uses all individuals in Sample B.

The estimates for measures of income inequality in Figures 7-9 and 13-14 are for a subset of Sample B. In each year, we drop households with zero household earnings. Then, for each different variable of interest (eg unequivalized household earnings or equivalized pre-tax household income) we trim the lowest 0.5% of observations. Thus, when we apply different measures of dispersion to equivalized household earnings in Figure 8, we apply them to exactly the same set of households. In Figure 9, when we compare inequality across different measures of income, using the variance of log metric, the sample of households is the same for each measure of income, except that there is some variation in the identities of the 0.5% of households that are trimmed.

Sample C, used for statistics involving wages, is a sample of individuals from households in Sample B, aged 25-60 and with annual hours greater than 260, where annual hours are computed using hours last week prior to the 1975 income year, and using usual hours after it becomes available in 1975. Then, for 1975 onwards, we drop individuals with wages (computed using usual hours) below half the minimum (recall that Sample A applies a similar screen, but using a different measure of hours). The plots for wage dispersion over the life-cycle in Figures 13 and 14 use Sample C for the period 1975 onwards.

**Hours.** Recall that we compute an individual's wage as annual earnings divided by annual hours worked. To compute hours worked last year we multiply weeks worked last year (wkslyr) by a measure of hours worked per week. Up to and including income year 1974 we are forced to use hours worked last week (hours), while from 1975 onwards a new variable (hrslyr) becomes available which measures usual hours per week last year. One would expect this latter measure to produce a much more accurate estimate for an individual's annual hours, and thus for his annual wage. We compute hours and wages both ways for the 1975-2005 period. Reassuringly, we find that trends in the variances of hours and wages are very similar over this period, while there is some difference, unsurprisingly, in levels of inequality - there is less variance in wages using the better measure. We also find a very similar increase in the correlation between individual hours and individual wages using the two different approaches, though the level of the correlation is much lower using the hours-last-week question. This reflects the well-known division bias: mis-measurement in hours translates automatically to mis-measurement in the inverse direction in wages, and thus drives down the observed wage-hour correlation.

Prior to 1975 income year, in addition to having to use hours last week (rather than usual weekly hours) there is a second reason why our measure of hours is of lessor quality, which is that the March CPS data files record weeks worked in intervals rather than as specific integers (even though the original questionnaires for the 1970-1975 survey years asked for integer responses). Based on the weeks worked distributions in income years 1975 forwards, Unicon converts interval codes into estimates of cell means. We compute an individual's wage as individual earnings divided by hours last week times estimated-cell-mean weeks worked.

Taken together, measures of hours and wages prior to income year 1975 are more uncertain than in later years, and estimates of first and second moments for this period should be viewed accordingly.

**Imputation**. The CPS is subject to two sources of nonresponse: noninterview households and item nonresponse. To compensate for this data loss, the weights on noninterviewed households are distributed among interviewed households. The second source is item nonresponse, meaning a respondent either does not know or refuses to provide the answer to a question. The Census Bureau imputes missing income data using a "hot deck" procedure which matches individuals with missing observations to others with similar demographic and economic information who did answer the questions. For example, the weekly earnings hot deck is defined by age, race, sex, usual hours, occupation and educational attainment. Before any edits are applied, the data is sorted geographically so that missing values are allocated from geographically close records.

We do not exclude households with imputed income because imputation is widely-used, especially for asset income categories. Thus dropping households with imputed values would drastically reduce the sample size, and call into question the appropriateness of the CPS-provided weights. Response rates for the CPS are high relative to other large household surveys, but have been declining over time. Moreover, for households nonresponse rates are higher for income than for other kinds of questions. Atrostic and Kalenkoski (2002) report response rates, defined as percent of all recipients (reported and imputed) who also reported an amount for the 1990 March CPS and the 2000 March CPS. Response rates for earnings from longest job (incer1) fell from 81.2 percent to 72.4 percent. Response rates for interest and dividend income fell from over 70 percent to below 50%. In terms of the share of income imputed, 26.8 percent of total wage and salary earnings, 43.8 percent of non-farm self-employment income, and 64.1 percent of interest and dividend income was imputed in 2000. For a significant fraction of households all income items are imputed.

**Topcoding.** Topcoding is an important issue to address in the CPS, both for computing means, and for measuring the evolution of inequality at the top of the income distribution.

Public top-code thresholds vary widely across income categories, and across time. An additional problem is that the Census Bureau's internal data is also subject to censoring (to economize on computer tape, and to protect against gross errors). For example, the public use censoring point for the variable incwag (income from wages and salaries) was $50,000 for the income years 1975-1980, $75,000 for 1981-1983 and $99,999 for 1984-1986. For the same variable, the internal CPS censoring points were $99,999 for the period 1975-1984, and $250,000 for 1985-1986.

We deal with top-coded observations by assuming the underlying distribution for each component of income is Pareto, and we follow a suggestion of David Domeij by forecasting the mean value for top-coded observations by extrapolating a Pareto density fitted to the non-top-coded upper end of the observed distribution. This procedure automatically takes care of the internal censoring problem, since the internal threshold always exceeds the public use limit. It also has the advantage that in principle it adjusts appropriately to changes in top code thresholds.

We apply this procedure at the most disaggregated decomposition of income possible. Thus, for example, for each year we divide the set of observations for the variable incer1 (income from primary source) according to whether or not they are flagged as wage and salary or self employment, and run separate regressions on the two sets of observations. This is important for two reasons. First, for any given individual, while one type of income may be top-coded others will not be. Second, there is more upper tail concentration in some types of income than others.

Beginning in income year 1995 the CPS started reporting cell means for top-coded observations, with cells identified by gender, race and work experience. This allows us to assess the performance of the regression procedure. We find that the regression approach generally performs very well for most income categories. It leads us to slightly over-predict income from primary source flagged as wages and salary over the 1995-2005 income year period, and to slightly under-predict interest income.

Since our primary goal is to measure changes in inequality consistently over time, we use the regression approach for the primary income variable through the sample period, even when cell means are available. However, at the same time that the Census began reporting cell means, they drastically reduced public use censoring points for many income categories: the threshold for interest income declined from $99,999 to $35,000 between income years 1997 and 1998 and to $25,000 in 2002, while the threshold for dividend income declined from $99,999 to $15,000. We found that when the distribution is truncated too far to the left, the Pareto-extrapolation procedure does not always perform well. Thus for income years 1998 to 2005 we use cell means for all income categories, except income from primary source.

Comparing our per-capita salary estimates, derived using the cell means, to figures made publicly available by the CPS (http://www.census.gov/hhes/www/income/dinctabs.html), the differences are tiny: less than $50 in all income years between 1995 and 2005, except 1999, where our estimate is $383 below the public number. However, there are errors in the reported cell means for earnings for the 2000 survey year (1999 income year): for example, the replacement value for earnings (topcode value $150,000) for male, non-black non-hispanic full-year full-time workers falls from $306,731 in 1999 to $229,340 in 2000, and then rises to $335,115 in 2001. Larrimore et. al. (2008) were granted access to internal CPS data, and report a 2000 cell mean for this group of $300,974.

The precise procedure we follow to compute top-coding adjustments is as follows. First, for a particular income variable, we identify the existence of top-coded observations. Then we sort observations in ascending order by income. The sample for our least-squares regression is the top (weighted) decile of non-zero, non-top-coded observations. For each individual $i$ with income $y_i$ we compute the fraction of households in our sample (including top coded households) with income greater than $y_i$, which we denote $v_i$. We then regress $\ln(v)$ on a constant and $\ln(y)$, and set the adjustment factor to $\beta/(1+\beta)$, where $\beta$ is the estimated coefficient on income. For a given income type in a given year, all top coded observations are assigned an income value equal to the top-code threshold times this adjustment factor.

**Demographic variables.** First we note that demographic variables (age, years of education, etc) refer to the survey year, while questions about income refer to the previous year. We do not attempt to adjust for this timing discrepancy. Thus, for example, the CPS version of Sample B for income year 1980 corresponds to households who in March 1981 reported at least one households member aged 25-60.

*Head* If there are any 25-60 year-old males in the household, the oldest male is the head. If there are no such males, the oldest 25-60 year-old female is the head. Note that this definition of head makes no connection to the identity of the CPS reference person. *Education* We define an individual to be college educated if they have 16 years of schooling or more. *Race* We divide individuals into those identifying as "white" and those that do not, who we label "non-white". Until 1988 the only non-white options were "black" or "other". In 1988, American Indian and Asian were added as additional options. In 1996 the "other" option was dropped. In 2003 many new options were added.

**Dispersion related to observables, and residual inequality.** For the plots of residual dispersion in Figures 5 and 7 we proceed as follows.

The sample for Figure 7 is in Sample B in which there are either one or two adults (a head and non-head) aged 25-60. These households constitute around 96% of all households in Sample B. The sample for Figure 5 is the subset of these households with a male head (where head is defined above).

Both sets of regressions use exactly the same set of regressors. The independent variables for the two-adult households are: 3 race dummies (white-white, non-white-non-white, mixed-race), 2 sex dummies (male-female, same-sex), 4 education dummies (college-college, college-non-coll, non-coll-college, non-coll-non-coll), average years of education for all adults, a quadratic in age (actual age minus 25) for the head, a quadratic in age for the non-head, number of household members below age 25, number of members above age 60. Note that this specification admits the possibility that earnings in households in which only the head has a college degree (college-non-coll) might differ from those in which only the non-head has a degree (non-coll-college). Note also that by construction there are no female-male households. The independent variables for the one-adult households are analogous: 2 race dummies, 2 sex dummies, 2 education dummies, years of education, a quadratic in age, number below age 25, number above age 60.

**Income measures.** Over our sample period there have been two important changes in the set of income questions asked in the March CPS, one beginning in the 1975 income year, and a second in the 1987 income year. However, these changes appear to have a negligible impact on either total income, or its division between different classes of income. The exception to this is for private transfers, which increases from 1.9 percent of pre-tax income in 1974 to 3.5 percent in 1975 (where these figures apply to Sample A).

*Wages and salary*

1967-1986 incwag

1987-2005 incer1 (if ernsrc=1 (wage and salary)) + incwg1

incwag = income from wage and salary; incer1 = earnings from longest job before deductions; incwg1 = income from other wage and salary

*Self employment income*

1967-1986 incse + incfrm

1987-2005 incer1 (if ernsrc=2 or 3 (farm or non-farm self-employment)) + incse1 + incfr1

incse = income from non-farm self-employment; incfrm = income from farm or nonincorporated self-employment; inse1 = income from other work – own business self-employment; incfr1 = income from other work – farm self-employment

*Private transfers*

1967-1974 incoth

1975-1986 incret + incalc + incoth

1987-2005 incoth + incalm + inchld + incds1 + incds2 + incont + incrt1 + incrt2 + incsi1 + incsi2

incoth = income from other sources; incret = income from retirement funds; incalc = income from alimony and child support; incalm = income from alimony; inchld = income from child support; incds1 = income from disability income – primary source; incds2 = income from disability income – secondary source; incont = income from contributions, assistance from friends; incrt1 = income from retirement income – primary source; incrt2 = income from retirement income – secondary source; incsi1 = income from survivors income – primary source; incsi2 = income from survivors income - secondary source

*Net financial income*

1967-1974 incint

1975-1986 incint + incdiv

1987-2005 incint + incdv2 + incrnt

incint = income from interest, dividends and net rentals; incdiv = income from dividends, rents and trusts; incdv2 = income from dividends; incrnt = income from rent

*Public transfers*

1967-1974 incpa + incomp + incss

1975-1986 incpa + incomp + incss + incsec

1987-2005 incpa + incss + incsec + inced + incvet + incwcp + incuc

incpa = income from public assistance or welfare; incomp = income from unemplymt/workers comp/veterans payments/govt pensions; incss = income from social security or railroad retirement – from US govt; incsec = income from supplemental security; inced = income from educational assistance; incvet = income from veterans payments; incwcp = income from worker's compensation; incuc = income from unemployment compensation

*Taxes (imputed)*

1979-2005 fedtaxbc + statetaxbc + fica - eitcrd

fedtaxbc = federal income tax liability, before credits; statetaxbc = state income tax liability, before credits; fica = social security retirement payroll deduction; eitcrd = earned income tax credit

Given the various income components described above, the different measures of income used in the paper are constructed successively as follows, following the project guidelines:

*Labor income* wages and salary + 2/3 self-employment income

*Labor income plus* labor income + private transfers

*Pre-government income* labor income plus + net financial income

*Pre-tax income* pre-government income + public transfers

*Disposable income* pre-tax income - taxes

Household-level measures of income are constructed by adding up the income of all household members

# B PSID

**Definition of "head".** The head of the family unit (FU) must be at least 16 years old, and the person with the most financial responsibility in the FU. If this person is female and she has a husband in the FU, then he is designated as head. If she has a boyfriend with whom she has been living for at least one year, then he is head. However, if she has 1) a husband or a boyfriend who is incapacitated and unable to fulfill the functions of head, 2) a boyfriend who has been living in the FU for less than a year, 3) no husband/boyfriend, then the FU will have a female head. A new head is selected if last year's head moved out of the household unit, died or became incapacitated, or if a single female head has gotten married. Also, if the family is a split-off family (hence a new family unit in the sample), then a new head is chosen.

**Samples.** In addition to the SRC sample, described in the main text, the second sample which belonged to the original 1968 survey is part of the Survey of Economic Opportunity (SEO) which was conducted by the Bureau of the Census for the Office of Economic Opportunity. The PSID selected about 2,000 low-income families with heads under the age of sixty from SEO respondents. In 1997, the SEO sample was reduced by two thirds.

In 1990, PSID added 2,000 Latino households, including families originally from Mexico, Puerto Rico, and Cuba. While this sample (the so called "Latino sample") did represent three major groups of immigrants, it missed out on the full range of post-1968 immigrants, Asians in particular. Because of this crucial shortcoming, and a lack of sufficient funding, the Latino sample was dropped after 1995. A sample of 441 immigrant families, including Asians, was added in 1997 (the so called "Immigrant sample").[24]

**File structure of the PSID data.** Information on *family-level* variables and on *individual-level* variables (for individuals in families belonging to the PSID sample) are split in two different sets of files. There are several family-level files, one for each year (*Single-year Family Files*), which contain one record for each family interviewed in the specified year. Note that individual income measures, and a large set of other individual-level variables (e.g., race, marital status) are contained in the *family files.* There is only one cross-year individual file with some individual-level data (e.g. education) collected from 1968 to the most recent interviewing wave (*Cross-year Individual File*). The file also contains the ID of the family with whom the person is associated in each year, which can be used to match individual-level data and family-level data.

The PSID contains many useful data supplements. The *Family Income-Plus Files, 1994-2001* contain various constructed income variables for household income and its components. The *Hours of Work and Wage Files, 1994-2001* contain constructed variables for total annual hours worked of heads and wives. The *Wealth Supplement File* includes detailed wealth information for 1984, 1989, 1994, 1999, 2001, 2003, and 2005. It can be linked to the rest of PSID data. Finally, a *Validation Study* was designed to assess the quality of economic data obtained in the PSID. The first wave of the Validation Study was conducted in 1983 and a second wave was conducted in 1987. For the Validation Study, the standard PSID questionnaire was administered to a sample drawn from a single large manufacturing firm. Questionnaire results were compared to company records to verify respondents' answers to questions such as earnings and hours worked. This source of data has been frequently used in the past to assess the size of measurement error in earnings and hours.

**Data quality.** Traditionally the PSID data has been released in two stages –an *early release* file with variables named ERxxxxx, and a *final release* file with variables named Vxxxx. The

[24]The so called "PSID core sample" combines the SRC, SEO and Immigrant samples. If one plans to use also these three samples together, weights should be used.

final release file contains data that has been subject to more stringent cleaning and checking processes and contains a number of constructed variables (e.g., total annual labor income of the head and wife). From 1994 on the final release files have not been made available. Instead, clean variables for labor income, annual hours and several other variables, are available in a number of supplementary data sets. These include the *Family Income-Plus Files* which contain various constructed income variables, the *Hours of Work and Wage Files*, which are used for data on annual hours worked, and the *Retrospective Occupation Files.*

**Top coding and bracketed variables.** We deal with top-coded observations by assuming the underlying distribution for each component of income is Pareto, and by forecasting the mean value for top-coded observations by extrapolating a Pareto density fitted to the non-top-coded upper end of the observed distribution. Variables with top-coded observations for which this imputation procedure was used are marked in Table A.

In some of the early waves, a number of income measures were bracketed. For these variables, we use the midpoint of each bracket, and $1.5\times$ the top-coded thresholds for observations in the top bracket. Bracketed variables are marked in Table A.

**Variable Definitions.** In the PSID all the questions are retrospective, i.e. variables in survey-year $t$ refer to calendar year $t-1$. The interview is usually conducted around March. A complete listing of the original PSID variables used in the construction of the variables in the final data set, year by year, can be found in Table A. When variables were not defined consistently across years (for example race was categorized differently in different years), the variables were recoded based on their original (and less detailed) coding, so as to be consistent across years.

A detailed definition of the key variables used in the study follows below:

*Labor Income.* For heads and wives, annual labor income includes all income from wages, salaries, commissions, bonuses, overtime and the labor part of self-employment income. The PSID splits self-employment income into asset and labor components using a 50-50 rule.

*Annual Hours of Work.* For heads and wives, it is defined as the sum of annual hours worked on the main job, on extra jobs, plus annual hours of overtime. It is computed by the PSID using information on usual hours worked per week and the number of actual weeks worked in the last year.

*Hourly Wage.* It is defined as Labor Income divided Annual Hours of Work.

*Household Labor Income.* It is defined as the sum of head and wife Labor Income.

*Labor Income Plus.* It is defined as Household Labor Income plus private transfers. Private transfers include alimony, child support, help from relatives, miscellaneous transfers, private retirement income, annuities and other retirement income.

*Financial Asset Income.* It includes income from interests, dividends, trust funds, and the asset part of self-employment income.

*Total Asset Income.* It includes Financial Asset Income plus rental income. We do not include an imputed rental value for owner-occupied housing in the definition of rental income.

*Household Pre-Government Income.* It is the sum of Labor Income Plus and Total Asset Income

*Household Income Plus Benefits.* It is the sum of Household Pre-Government Income plus public transfers. Public transfers include payments from the Aid to Families with Dependent Children (AFDC) program, Supplemental Security Income payments, other welfare receipts, plus social security benefits, unemployment benefits, worker's compensation and veterans' pensions. In the 1968 and 1969 interview years, many items are missing, so we start computing this measure from the 1970 survey (actual year 1969).

*Taxes.* An estimate of household federal income taxes, and state income taxes is computed based on the NBER's TAXSIM program.

*Household Disposable Income.* It is constructed as the sum of Household Pre-Government Income plus public transfers less federal and state taxes.

*Food Consumption.* It is defined as total expenditures on food eaten at home, on food eaten out of home, on food delivered, and on food purchased using food stamps. There is no food data available in the 1973, 1988 and 1989 interview years, except for food purchased using food stamps, so we omit those years in all calculations using this variable.

## C CEX

# References


[1] Acemoglu, Daron 2002. "Technical Change, Inequality and the Labor Market." *J. Economic Literature* 40 (March): 7-72.

[2] Aiyagari, Rao S. 1994. "Uninsured Idiosyncratic Risk and Aggregate Saving." *Q.J.E.* 109 (August): 659-84.

[3] Altonji, Joseph G., and Lewis M. Segal. 1996. "Small-Sample Bias in GMM Estimation of Covariance Structures." *J. Bus. and Econ. Statis.* 14 (July): 353-66.

[4] Atrostic, B.K., and C. Kalenkoski. 2002. "Item Response Rates: One Indicator of How Well We Measure Income". Proceedings of the Annual Meetings of the American Statistical Association.

[5] Autor, David H. , Lawrence F. Katz, and Melissa S. Kearney. 2008. "Trends in U.S. Wage Inequality: Revising the Revisionists." *Review of Economics and Statistics.* 90 (May): 300-23.

[6] Attanasio Orazio, Eric Battistin, and Hide Ichimura. 2005, "What Really Happened to Consumption Inequality in the US?" in E. Berndt and C. Hulten (eds.), *Measurement Issues in Economics—The Paths Ahead: Essays in Honor of Zvi Griliches,* University of Chicago Press, Chicago

[7] Baker, Michael. 1997. "Growth-Rate Heterogeneity and the Covariance Structure of Life-Cycle Earnings." *Journal of Labor Economics.* 15(2): 338-75.

[8] Blundell, Richard, Luigi Pistaferri, and Ian Preston. 2004. "Imputing consumption in the PSID using food demand estimates from the CEX " Institute for Fiscal Studies Working paper 04/27

[9] Blundell, Richard, Luigi Pistaferri, and Ian Preston. 2008. "Consumption Inequality and Partial Insurance." *A.E.R.* forthcoming.

[10] Bound, John, and George Johnson. 1992. "Changes in the Structure of Wages in the 1980s: An Evaluation of Alternative Explanations". *American Economic Review*, 82(3): 371-382.

[11] Burkhauser, R.V., S. Feng and S.P. Jenkins. 2007. "Using the P90/P10 Index to Measure US Inequality Trends with Current Population Survey Data: A View from Inside the Census Bureau Vaults". ISER Working Paper 2007-14

[12] Card, David and John E. DiNardo. 2002. "Skill-Biased Technological Change and Rising Wage Inequality: Some Problems and Puzzles." *Journal of Labor Economics* 20(4): 733-783.

[13] Chamberlain, Gary. 1984. "Panel Data." In *Handbook of Econometrics*, vol. 2, edited by Zvi Griliches and Michael D. Intrilligator. Amsterdam: North Holland.

[14] Cutler, David M., and Lawrence F. Katz. 1991. "Macroeconomic Performance and the Disadvantaged." *Brookings Papers Econ. Activity* 0: 1-74.

[15] Deaton, Angus, and Christina Paxson. 1994. "Intertemporal Choice and Inequality." *Journal of Political Economy* 102: 437-467.

[16] Duncan, Greg, and Willard Rodgers. 1991. "Has Children Poverty Become More Persistent?" *American Sociological Review.* 56(4): 538-550.

[17] Eckstein, Zvi, and Eva Nagypal. 2004. "The Evolution of U.S. Earnings Inequality: 1961-2002." *Fed. Reserve Bank Minneapolis Quart. Rev.* 28 (December): 10-29.

[18] Garner, Thesia, George Janini, William Passero, Laura Paszkiewicz and Mark Vendemia. 2006 "The CE and the PCE: a comparison." Monthly Labor Review, September: 20-46

[19] Gottschalk, Peter, and Robert A. Moffitt. 1994. "The Growth of Earnings Instability in the U.S. Labor Market." *Brookings Papers Econ. Activity* 0: 217-72.

[20] Guvenen, Fatih. 2007. "An Empirical Investigation of Labor Income Processes." Mimeo, University of Minnesota.

[21] Heathcote, Jonathan, Kjetil Storesletten, and Giovanni L. Violante. 2008. "The Macroeconomic Implications of Rising Wage Inequality in the United States." Mimeo, NYU.

[22] Heathcote, Jonathan, Kjetil Storesletten, and Giovanni L. Violante. 2005. “Two Views of Inequality Over the Life-Cycle.” *J. European Econ. Assoc.* 3 (April-May): 543-52.

[23] Hornstein, Andreas, Per Krusell and Giovanni L. Violante. 2005. “The Effects of Technical Change on Labor Market Inequalities.” In *Handbook of Economic Growth*, vol. 1B, edited by Philippe Aghion and Steven Durlauf. Amsterdam: North-Holland.

[24] Huggett, Mark. 1993. “The Risk Free Rate in Heterogeneous-Agents Incomplete-Insurance Economies.” *J. Econ. Dynamics and Control* 17 (September-November): 953-69.

[25] Huggett, Mark, Gustavo Ventura, and Amir Yaron. 2008. “Sources of Lifetime Inequality.” Mimeo, Georgetown University.

[26] Imrohoruglu, Ayse. 1989. “Cost of Business Cycles with Indivisibilities and Liquidity Constraints.” *J.P.E.* 97 (December): 1364-83.

[27] Johnson, David, and Stephanie Shipp. 1997. “Trends in Inequality Using Consumption-Expenditures: The U.S. from 1960 to 1993.” *Rev. Income and Wealth* 43 (June): 133-52.

[28] Juhn, Chinhui, Kevin M. Murphy, and Brooks Pierce. 1993. “Wage Inequality and the Rise in Returns to Skill.” *J.P.E.* 101 (June): 410-42.

[29] Katz, Lawrence F., and David H. Autor. 1999. “Changes in the Wage Structure and Earnings Inequality.” In *Handbook of Labor Economics*, vol. 3A, edited by Orley Ashenfelter and David Card. Amsterdam: North Holland.

[30] Katz, Lawrence F., and Kevin M. Murphy. 1992. “Changes in Relative Wages, 1963-1987: Supply and Demand Factors.” *Q.J.E* 107 (February): 35-78.

[31] Krueger, Dirk, and Fabrizio Perri. 2006. “Does Income Inequality Lead to Consumption Inequality? Evidence and Theory.” *Rev. Econ. Studies* 73 (January): 163-93.

[32] Lee, D. 1999. “Wage inequality in the United States during the 1980s: Rising dispersion or falling minimum wage?” *Quarterly Journal of Economics* 114(3): 9771023.

[33] Lemieux, Thomas. 2006. “Post-Secondary Education and Increasing Wage Inequality.” NBER Working Paper 12077.

[34] Lillard, Lee, and Yoram Weiss. 1979. “Components of Variation in Panel Earnings Data: American Scientists, 1960-1970.” *Econometrica* 47: 437-454.

[35] Lillard, Lee, and Robert J. Willis. 1978. “Dynamic Aspects of Earning Mobility. *Econometrica* 46(5): 985-1012.

[36] Larrimore, J., R.V. Burkhauser, S. Feng and L. Zayatz. 2008. ”Consistent Cell Means for Topcoded Incomes in the Public Use March CPS (1967-2007)”. NBER Working Paper 13941.

[37] McGrattan, E., and R. Rogerson. 2004. ”Changes in Hours Worked, 1950-2000”. Federal Reserve Bank of Minneapolis Quarterly Review 28(1), 14-33.

[38] Meghir, Costas, and Luigi Pistaferri. 2004. “Income Variance Dynamics and Heterogeneity.” *Econometrica* 72 (January): 1-32.

[39] Murphy, Kevin, and Finis Welch. 1993. “The Structure of Wages,” *Quarterly Journal of Economics*: 107(1): 285-326.

[40] Nelson, Julie 1994, “On Testing for Full Insurance Using Consumer Expenditure Survey Data: Comment,” *Journal of Political Economy,* 102, 384–394.

[41] Ríos-Rull, José-Víctor. 1995. “Models with Heterogeneous Agents.” In *Frontiers of Business Cycle Research*, edited by Thomas F. Cooley. Princeton: Princeton University Press.

[42] Roemer, M. 2000. ”Assessing the Quality of the March Current Population Survey and the Survey of Income and Program Participation Income Estimates”. Census Bureau Working Paper

[43] Roemer, M. 2002. "Using Administrative Earnings Records to Assess Wage Data Quality in the March Current Population Survey and the Survey of Income and Program Participation". Census Bureau Working Paper

[44] Ruser, J., A. Pilot and C. Nelson 2004. "Alternative Measures of Household Income: BEA Personal Income, CPS Money Income, and Beyond". Federal Economic Statistics Advisory Committee Working Paper

[45] Slesnick, Daniel T. 2001. *Consumption and Social Welfare: Living Standards and Their Distribution in the United States*. New York: Cambridge University Press.

[46] Storesletten, Kjetil, Chris I. Telmer, and Amir Yaron. 2004. "Consumption and Risk Sharing over the Life Cycle." *J. Monetary Econ.* 51 (April): 609-33.

| | 1967-1979 | | | 1980-1989 | | | 1990-1999 | | | $\geq$ 2000 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | PSID | CPS | CEX | PSID | CPS | CEX | PSID | CPS | CEX | PSID | CPS | CEX |
| Avg. male age | 41.23 | 41.01 | | 39.75 | 39.70 | | 40.93 | 40.37 | | 42.14 | 41.74 | |
| Avg. female age | 39.06 | 41.27 | | 37.90 | 40.01 | | 39.27 | 40.58 | | 40.69 | 41.99 | |
| % white male | 0.898 | 0.893 | | 0.919 | 0.868 | | 0.917 | 0.845 | | 0.901 | 0.825 | |
| % male $\geq$16 years edu | 0.230 | 0.194 | | 0.302 | 0.260 | | 0.329 | 0.280 | | 0.319 | 0.302 | |
| % female $\geq$16 years edu | 0.138 | 0.119 | | 0.222 | 0.191 | | 0.268 | 0.246 | | 0.282 | 0.298 | |
| Avg. household size | 3.327 | 3.336 | | 2.978 | 2.978 | | 2.891 | 2.891 | | 2.772 | 2.825 | |
| % households with spouse | 0.779 | 0.720 | | 0.731 | 0.622 | | 0.690 | 0.574 | | 0.673 | 0.549 | |

Table 1: Some key demographic characteristics of the 25-60 years-old in PSID, CPS and CEX. In PSID, a household is defined as “family unit”, see the main text.

| | PSID (1967-96, 98, 00, 02) | | CPS (1967-2005) | | CEX (1980-2006) | |
|---|---|---|---|---|---|---|
| | dropped | remaining | dropped | remaining | dropped | remaining |
| **Initial Sample** | | **123,786** | | **2,217,997** | | |
| Missing/Miscoded Household Info | 7 | 123,779 | 8,993 | 2,209,004 | | |
| Pos. Labor Inc. & Zero Hours | 309 | 123,470 | 17,599 | 2,191,405 | | |
| Wage <0.5*Minimum Wage | 5,826 | 117,644 | 121,367 | | | |
| **Sample A** | | **117,644** | | **2,070,038** | | |
| Age Selection, 25-60 | 32,317 | 85,327 | 524,609 | 1,545,429 | | |
| **Sample B** | | **85,327** | | **1,545,429** | | |
| Individual with Hours<260 | 4,736 | 80,591 | 52,500 | 1,429,929 | | |
| **Sample C** | | **80,591** | | **1,492,929** | | |

Table 2: Sample selection in PSID, CPS and CEX.

| | Estimation in first differences | | | | | | Estimation in levels | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Perm+Trans (Untrimmed) | | Perm+Trans | | Perm+MA1 | | Perm+Trans | | Perm+MA1 | | AR1+Trans | | AR1+MA1 | |
| $\sigma_{z0}$ | – | | – | | – | | 0.0933 | | 0.0922 | | 0.1128 | | 0.1092 | |
| $\theta$ | 0.0000 | | 0.0000 | | 0.1075 | | 0.0000 | | -0.6378 | | 0.0000 | | -0.3833 | |
| $\rho$ | 1.0000 | | 1.0000 | | 1.0000 | | 1.0000 | | 1.0000 | | 0.9769 | | 0.9803 | |
| | $\sigma_\varepsilon$ | $\sigma_\eta$ | $\sigma_\varepsilon$ | $\sigma_\eta$ | $\sigma_\varepsilon$ | $\sigma_\eta$ | $\sigma_\varepsilon$ | $\sigma_\eta$ | $\sigma_\varepsilon$ | $\sigma_\eta$ | $\sigma_\varepsilon$ | $\sigma_\eta$ | $\sigma_\varepsilon$ | $\sigma_\eta$ |
| 1967 | – | – | – | – | – | – | 0.0439 | 0.0046 | 0.0311 | 0.0044 | 0.0198 | 0.0125 | 0.0173 | 0.0111 |
| 1968 | 0.0342 | 0.0324 | 0.0220 | 0.0270 | 0.0274 | 0.0215 | 0.0357 | 0.0163 | 0.0216 | 0.0170 | 0.0140 | 0.0217 | 0.0092 | 0.0216 |
| 1969 | 0.0253 | 0.0237 | 0.0210 | 0.0251 | 0.0264 | 0.0196 | 0.0401 | 0.0061 | 0.0258 | 0.0058 | 0.0166 | 0.0161 | 0.0109 | 0.0148 |
| 1970 | 0.0321 | 0.0364 | 0.0193 | 0.0237 | 0.0246 | 0.0180 | 0.0516 | 0.0000 | 0.0394 | 0.0000 | 0.0296 | 0.0069 | 0.0284 | 0.0046 |
| 1971 | 0.0311 | 0.0293 | 0.0211 | 0.0220 | 0.0264 | 0.0161 | 0.0686 | 0.0074 | 0.0509 | 0.0065 | 0.0389 | 0.0233 | 0.0353 | 0.0211 |
| 1972 | 0.0384 | 0.0283 | 0.0235 | 0.0277 | 0.0294 | 0.0216 | 0.0651 | 0.0000 | 0.0462 | 0.0000 | 0.0399 | 0.0043 | 0.0366 | 0.0030 |
| 1973 | 0.0473 | 0.0302 | 0.0247 | 0.0282 | 0.0304 | 0.0223 | 0.0845 | 0.0000 | 0.0648 | 0.0000 | 0.0606 | 0.0069 | 0.0590 | 0.0046 |
| 1974 | 0.0443 | 0.0290 | 0.0256 | 0.0332 | 0.0316 | 0.0272 | 0.0819 | 0.0017 | 0.0573 | 0.0014 | 0.0560 | 0.0118 | 0.0510 | 0.0105 |
| 1975 | 0.0419 | 0.0283 | 0.0226 | 0.0275 | 0.0289 | 0.0212 | 0.0827 | 0.0073 | 0.0625 | 0.0070 | 0.0550 | 0.0171 | 0.0523 | 0.0153 |
| 1976 | 0.0488 | 0.0253 | 0.0222 | 0.0283 | 0.0283 | 0.0215 | 0.0863 | 0.0107 | 0.0600 | 0.0109 | 0.0562 | 0.0215 | 0.0503 | 0.0200 |
| 1977 | 0.0377 | 0.0345 | 0.0163 | 0.0320 | 0.0218 | 0.0259 | 0.0738 | 0.0007 | 0.0500 | 0.0007 | 0.0453 | 0.0076 | 0.0402 | 0.0065 |
| 1978 | 0.0372 | 0.0406 | 0.0236 | 0.0369 | 0.0288 | 0.0304 | 0.0716 | 0.0075 | 0.0546 | 0.0072 | 0.0437 | 0.0152 | 0.0425 | 0.0136 |
| 1979 | 0.0504 | 0.0375 | 0.0200 | 0.0347 | 0.0253 | 0.0298 | 0.0789 | 0.0110 | 0.0591 | 0.0108 | 0.0495 | 0.0211 | 0.0481 | 0.0192 |
| 1980 | 0.0450 | 0.0360 | 0.0216 | 0.0374 | 0.0269 | 0.0317 | 0.0841 | 0.0178 | 0.0659 | 0.0172 | 0.0510 | 0.0304 | 0.0509 | 0.0281 |
| 1981 | 0.0394 | 0.0316 | 0.0222 | 0.0316 | 0.0279 | 0.0257 | 0.1050 | 0.0080 | 0.0837 | 0.0072 | 0.0722 | 0.0174 | 0.0735 | 0.0148 |
| 1982 | 0.0387 | 0.0310 | 0.0226 | 0.0344 | 0.0287 | 0.0280 | 0.1018 | 0.0130 | 0.0745 | 0.0130 | 0.0686 | 0.0232 | 0.0654 | 0.0217 |
| 1983 | 0.0410 | 0.0501 | 0.0189 | 0.0353 | 0.0248 | 0.0288 | 0.1125 | 0.0114 | 0.0870 | 0.0107 | 0.0784 | 0.0224 | 0.0775 | 0.0199 |
| 1984 | 0.0469 | 0.0435 | 0.0210 | 0.0374 | 0.0263 | 0.0309 | 0.1119 | 0.0090 | 0.0870 | 0.0083 | 0.0786 | 0.0186 | 0.0797 | 0.0162 |
| 1985 | 0.0432 | 0.0347 | 0.0232 | 0.0303 | 0.0291 | 0.0241 | 0.1142 | 0.0124 | 0.0920 | 0.0115 | 0.0818 | 0.0220 | 0.0842 | 0.0197 |
| 1986 | 0.0313 | 0.0488 | 0.0183 | 0.0373 | 0.0238 | 0.0315 | 0.1284 | 0.0017 | 0.1020 | 0.0009 | 0.0983 | 0.0100 | 0.1000 | 0.0075 |
| 1987 | 0.0445 | 0.0452 | 0.0226 | 0.0344 | 0.0275 | 0.0284 | 0.1221 | 0.0028 | 0.0915 | 0.0026 | 0.0945 | 0.0105 | 0.0918 | 0.0091 |
| 1988 | 0.0433 | 0.0395 | 0.0220 | 0.0302 | 0.0277 | 0.0249 | 0.1093 | 0.0122 | 0.0776 | 0.0128 | 0.0829 | 0.0209 | 0.0782 | 0.0198 |
| 1989 | 0.0416 | 0.0351 | 0.0222 | 0.0338 | 0.0277 | 0.0279 | 0.0982 | 0.0024 | 0.0810 | 0.0007 | 0.0742 | 0.0100 | 0.0761 | 0.0076 |
| 1990 | 0.0463 | 0.0405 | 0.0192 | 0.0352 | 0.0244 | 0.0298 | 0.1095 | 0.0137 | 0.0788 | 0.0136 | 0.0858 | 0.0231 | 0.0800 | 0.0218 |
| 1991 | 0.0436 | 0.0407 | 0.0205 | 0.0341 | 0.0255 | 0.0285 | 0.1009 | 0.0000 | 0.0755 | 0.0000 | 0.0818 | 0.0053 | 0.0787 | 0.0029 |
| 1992 | 0.0917 | 0.0493 | 0.0336 | 0.0343 | 0.0400 | 0.0278 | 0.1096 | 0.0162 | 0.0800 | 0.0147 | 0.0899 | 0.0260 | 0.0842 | 0.0237 |
| 1993 | 0.0842 | 0.0441 | 0.0282 | 0.0285 | 0.0357 | 0.0224 | 0.0867 | 0.0038 | 0.0592 | 0.0039 | 0.0691 | 0.0114 | 0.0621 | 0.0105 |
| 1994 | 0.0764 | 0.0455 | 0.0227 | 0.0316 | 0.0292 | 0.0244 | 0.0871 | 0.0084 | 0.0682 | 0.0055 | 0.0698 | 0.0174 | 0.0688 | 0.0134 |
| 1995 | 0.0702 | 0.0376 | 0.0209 | 0.0360 | 0.0266 | 0.0294 | 0.0969 | 0.0045 | 0.0698 | 0.0136 | 0.0792 | 0.0142 | 0.0721 | 0.0192 |
| 1996 | 0.0722 | 0.0417 | 0.0328 | 0.0276 | 0.0354 | 0.0244 | 0.1664 | 0.0057 | 0.0938 | 0.0056 | 0.1595 | 0.0125 | 0.1194 | 0.0113 |
| 1997 | 0.0805 | 0.0374 | 0.0529 | 0.0330 | 0.0539 | 0.0311 | 0.1207 | 0.0150 | 0.0793 | 0.0149 | 0.1076 | 0.0235 | 0.0861 | 0.0207 |
| 1998 | 0.0887 | 0.0331 | 0.0730 | 0.0385 | 0.0724 | 0.0377 | 0.0750 | 0.0243 | 0.0648 | 0.0242 | 0.0558 | 0.0346 | 0.0582 | 0.0330 |
| 1999 | 0.0959 | 0.0224 | 0.0718 | 0.0301 | 0.0711 | 0.0298 | 0.1171 | 0.0142 | 0.0890 | 0.0142 | 0.1050 | 0.0219 | 0.0967 | 0.0207 |
| 2000 | 0.1031 | 0.0118 | 0.0706 | 0.0217 | 0.0697 | 0.0219 | 0.1592 | 0.0042 | 0.1133 | 0.0042 | 0.1542 | 0.0092 | 0.1352 | 0.0085 |
| Average | 0.0510 | 0.0363 | 0.0266 | 0.0318 | 0.0317 | 0.0263 | 0.0916 | 0.0082 | 0.0674 | 0.0081 | 0.0669 | 0.0170 | 0.0626 | 0.0154 |

Table 3: Results of the income process estimation: individual hourly wages.

| | Estimation in first differences | | | | | | Estimation in levels | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Perm+Trans (Untrimmed) | | Perm+Trans | | Perm+MA1 | | Perm+Trans | | Perm+MA1 | | AR1+Trans | | AR1+MA1 | |
| $\sigma_{z0}$ | – | | – | | – | | 0.0904 | | 0.0884 | | 0.1181 | | 0.1090 | |
| $\theta$ | 0.0000 | | 0.0000 | | 0.1959 | | 0.0000 | | -0.4576 | | 0.0000 | | -0.3148 | |
| $\rho$ | 1.0000 | | 1.0000 | | 1.0000 | | 1.0000 | | 1.0000 | | 0.9699 | | 0.9773 | |
| | $\sigma_\varepsilon$ | $\sigma_\eta$ | $\sigma_\varepsilon$ | $\sigma_\eta$ | $\sigma_\varepsilon$ | $\sigma_\eta$ | $\sigma_\varepsilon$ | $\sigma_\eta$ | $\sigma_\varepsilon$ | $\sigma_\eta$ | $\sigma_\varepsilon$ | $\sigma_\eta$ | $\sigma_\varepsilon$ | $\sigma_\eta$ |
| 1967 | – | – | – | – | – | – | 0.1441 | 0.0057 | 0.1139 | 0.0054 | 0.1159 | 0.0166 | 0.1062 | 0.0132 |
| 1968 | 0.0256 | 0.0654 | 0.0161 | 0.0286 | 0.0246 | 0.0192 | 0.1647 | 0.0193 | 0.1377 | 0.0199 | 0.1359 | 0.0308 | 0.1307 | 0.0277 |
| 1969 | 0.0358 | 0.0691 | 0.0158 | 0.0308 | 0.0254 | 0.0204 | 0.1358 | 0.0089 | 0.1028 | 0.0093 | 0.1053 | 0.0225 | 0.0954 | 0.0195 |
| 1970 | 0.0407 | 0.0801 | 0.0122 | 0.0395 | 0.0211 | 0.0291 | 0.1081 | 0.0045 | 0.0892 | 0.0041 | 0.0752 | 0.0189 | 0.0740 | 0.0146 |
| 1971 | 0.0294 | 0.0747 | 0.0126 | 0.0440 | 0.0206 | 0.0336 | 0.1171 | 0.0080 | 0.1047 | 0.0074 | 0.0800 | 0.0245 | 0.0845 | 0.0196 |
| 1972 | 0.0479 | 0.0846 | 0.0161 | 0.0299 | 0.0233 | 0.0212 | 0.1412 | 0.0006 | 0.1246 | 0.0001 | 0.1021 | 0.0153 | 0.1064 | 0.0105 |
| 1973 | 0.0481 | 0.0826 | 0.0149 | 0.0395 | 0.0221 | 0.0325 | 0.1364 | 0.0000 | 0.1213 | 0.0000 | 0.0989 | 0.0109 | 0.1055 | 0.0066 |
| 1974 | 0.0304 | 0.0791 | 0.0166 | 0.0408 | 0.0244 | 0.0327 | 0.1281 | 0.0022 | 0.1075 | 0.0023 | 0.0901 | 0.0148 | 0.0911 | 0.0118 |
| 1975 | 0.0427 | 0.0987 | 0.0138 | 0.0397 | 0.0219 | 0.0315 | 0.1790 | 0.0116 | 0.1644 | 0.0102 | 0.1366 | 0.0280 | 0.1450 | 0.0217 |
| 1976 | 0.0538 | 0.0867 | 0.0160 | 0.0425 | 0.0230 | 0.0342 | 0.1618 | 0.0046 | 0.1338 | 0.0051 | 0.1207 | 0.0160 | 0.1204 | 0.0132 |
| 1977 | 0.0302 | 0.0981 | 0.0132 | 0.0436 | 0.0218 | 0.0354 | 0.1721 | 0.0045 | 0.1540 | 0.0037 | 0.1301 | 0.0179 | 0.1370 | 0.0129 |
| 1978 | 0.0278 | 0.1247 | 0.0137 | 0.0418 | 0.0209 | 0.0329 | 0.1528 | 0.0071 | 0.1362 | 0.0070 | 0.1140 | 0.0162 | 0.1204 | 0.0134 |
| 1979 | 0.0372 | 0.1165 | 0.0094 | 0.0507 | 0.0171 | 0.0424 | 0.2080 | 0.0133 | 0.1781 | 0.0129 | 0.1658 | 0.0291 | 0.1657 | 0.0240 |
| 1980 | 0.0299 | 0.1092 | 0.0138 | 0.0450 | 0.0215 | 0.0348 | 0.1856 | 0.0198 | 0.1667 | 0.0187 | 0.1380 | 0.0383 | 0.1469 | 0.0323 |
| 1981 | 0.0474 | 0.0814 | 0.0128 | 0.0474 | 0.0218 | 0.0377 | 0.1922 | 0.0086 | 0.1674 | 0.0086 | 0.1467 | 0.0207 | 0.1504 | 0.0172 |
| 1982 | 0.0551 | 0.1043 | 0.0182 | 0.0491 | 0.0279 | 0.0373 | 0.1872 | 0.0014 | 0.1517 | 0.0020 | 0.1454 | 0.0117 | 0.1407 | 0.0091 |
| 1983 | 0.0514 | 0.0820 | 0.0140 | 0.0390 | 0.0222 | 0.0301 | 0.1832 | 0.0115 | 0.1606 | 0.0102 | 0.1411 | 0.0253 | 0.1441 | 0.0203 |
| 1984 | 0.0370 | 0.1327 | 0.0150 | 0.0461 | 0.0235 | 0.0362 | 0.2295 | 0.0134 | 0.1961 | 0.0129 | 0.1833 | 0.0315 | 0.1823 | 0.0259 |
| 1985 | 0.0319 | 0.0771 | 0.0167 | 0.0355 | 0.0255 | 0.0256 | 0.1542 | 0.0000 | 0.1252 | 0.0000 | 0.1197 | 0.0024 | 0.1187 | 0.0007 |
| 1986 | 0.0686 | 0.1011 | 0.0140 | 0.0445 | 0.0223 | 0.0355 | 0.1986 | 0.0108 | 0.1805 | 0.0091 | 0.1600 | 0.0281 | 0.1671 | 0.0218 |
| 1987 | 0.0331 | 0.0811 | 0.0125 | 0.0458 | 0.0200 | 0.0369 | 0.1587 | 0.0087 | 0.1408 | 0.0084 | 0.1220 | 0.0205 | 0.1282 | 0.0171 |
| 1988 | 0.0385 | 0.0729 | 0.0150 | 0.0383 | 0.0221 | 0.0299 | 0.2021 | 0.0113 | 0.1753 | 0.0106 | 0.1677 | 0.0224 | 0.1692 | 0.0180 |
| 1989 | 0.0362 | 0.0881 | 0.0159 | 0.0337 | 0.0235 | 0.0261 | 0.1422 | 0.0024 | 0.1281 | 0.0015 | 0.1098 | 0.0139 | 0.1174 | 0.0102 |
| 1990 | 0.0206 | 0.1009 | 0.0121 | 0.0420 | 0.0197 | 0.0343 | 0.1716 | 0.0062 | 0.1527 | 0.0055 | 0.1420 | 0.0165 | 0.1460 | 0.0124 |
| 1991 | 0.0465 | 0.1243 | 0.0163 | 0.0392 | 0.0237 | 0.0303 | 0.1465 | 0.0123 | 0.1272 | 0.0118 | 0.1185 | 0.0234 | 0.1205 | 0.0200 |
| 1992 | 0.0669 | 0.1059 | 0.0210 | 0.0405 | 0.0303 | 0.0315 | 0.1758 | 0.0170 | 0.1558 | 0.0144 | 0.1489 | 0.0286 | 0.1518 | 0.0227 |
| 1993 | 0.0653 | 0.0856 | 0.0197 | 0.0394 | 0.0295 | 0.0301 | 0.1210 | 0.0123 | 0.1008 | 0.0132 | 0.0957 | 0.0240 | 0.0958 | 0.0219 |
| 1994 | 0.0511 | 0.0942 | 0.0158 | 0.0396 | 0.0250 | 0.0298 | 0.1307 | 0.0068 | 0.1242 | 0.0023 | 0.1058 | 0.0197 | 0.1151 | 0.0121 |
| 1995 | 0.0594 | 0.0669 | 0.0147 | 0.0383 | 0.0232 | 0.0283 | 0.1369 | 0.0000 | 0.1127 | 0.0122 | 0.1145 | 0.0105 | 0.1098 | 0.0173 |
| 1996 | 0.0460 | 0.0871 | 0.0181 | 0.0373 | 0.0217 | 0.0324 | 0.1826 | 0.0187 | 0.1501 | 0.0191 | 0.1653 | 0.0324 | 0.1524 | 0.0289 |
| 1997 | 0.0577 | 0.0691 | 0.0317 | 0.0434 | 0.0329 | 0.0403 | 0.1702 | 0.0137 | 0.1396 | 0.0139 | 0.1479 | 0.0257 | 0.1390 | 0.0220 |
| 1998 | 0.0694 | 0.0510 | 0.0453 | 0.0494 | 0.0440 | 0.0482 | 0.1577 | 0.0087 | 0.1292 | 0.0087 | 0.1304 | 0.0190 | 0.1255 | 0.0161 |
| 1999 | 0.0775 | 0.0534 | 0.0499 | 0.0442 | 0.0484 | 0.0437 | 0.1495 | 0.0250 | 0.1232 | 0.0251 | 0.1315 | 0.0351 | 0.1241 | 0.0324 |
| 2000 | 0.0856 | 0.0559 | 0.0545 | 0.0389 | 0.0528 | 0.0392 | 0.1413 | 0.0413 | 0.1171 | 0.0414 | 0.1325 | 0.0512 | 0.1228 | 0.0488 |
| Average | 0.0450 | 0.0884 | 0.0176 | 0.0409 | 0.0248 | 0.0326 | 0.1614 | 0.0091 | 0.1387 | 0.0090 | 0.1274 | 0.0215 | 0.1281 | 0.0178 |

Table 4: Results of the income process estimation: equivalized household earnings.

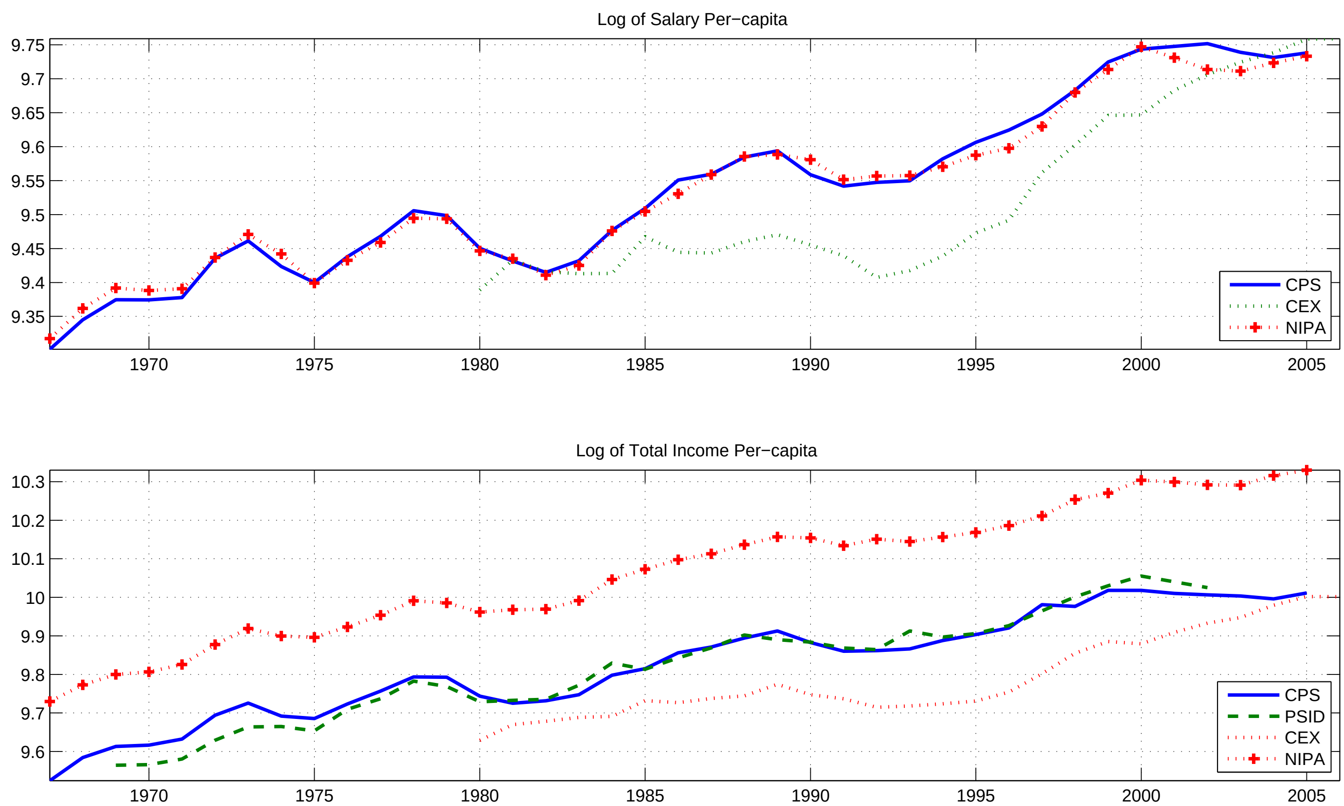


Figure 1: Comparison between averages in micro data and averages in NIPA: salary and total income

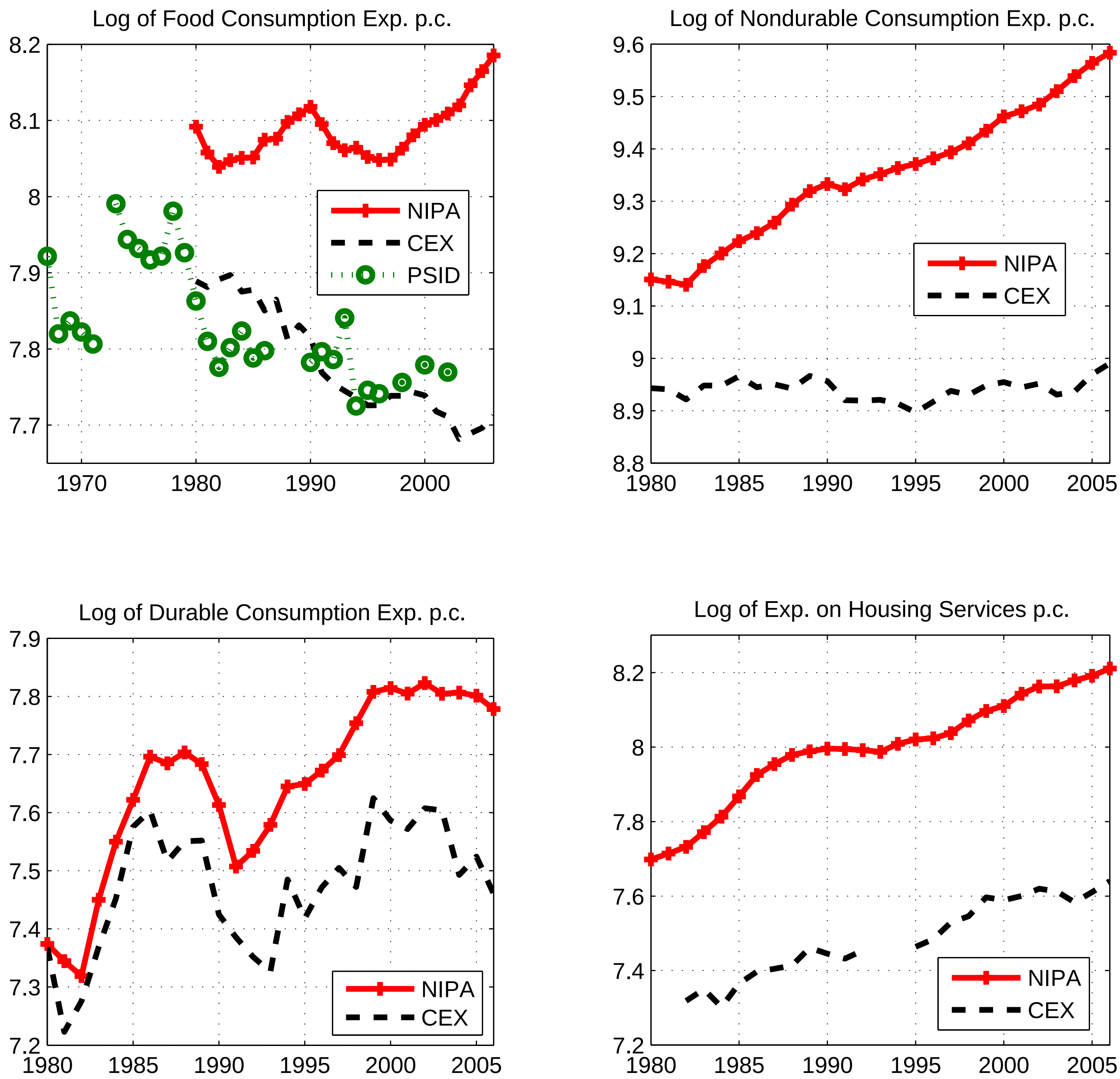


Figure 2: Comparison between averages in micro data and averages in NIPA: consumption

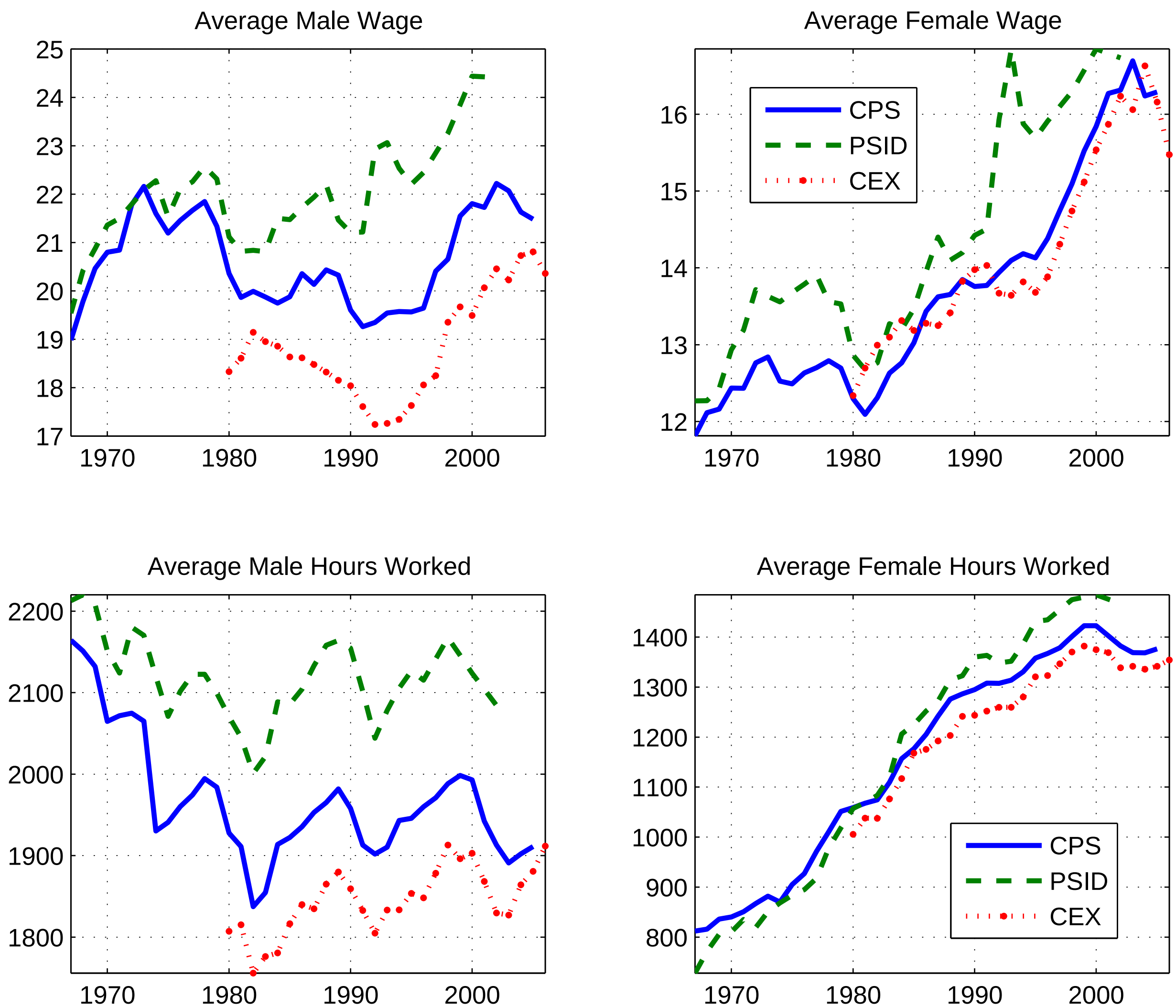


Figure 3: Comparison between CPS, PSID and CEX averages: wages and hours worked

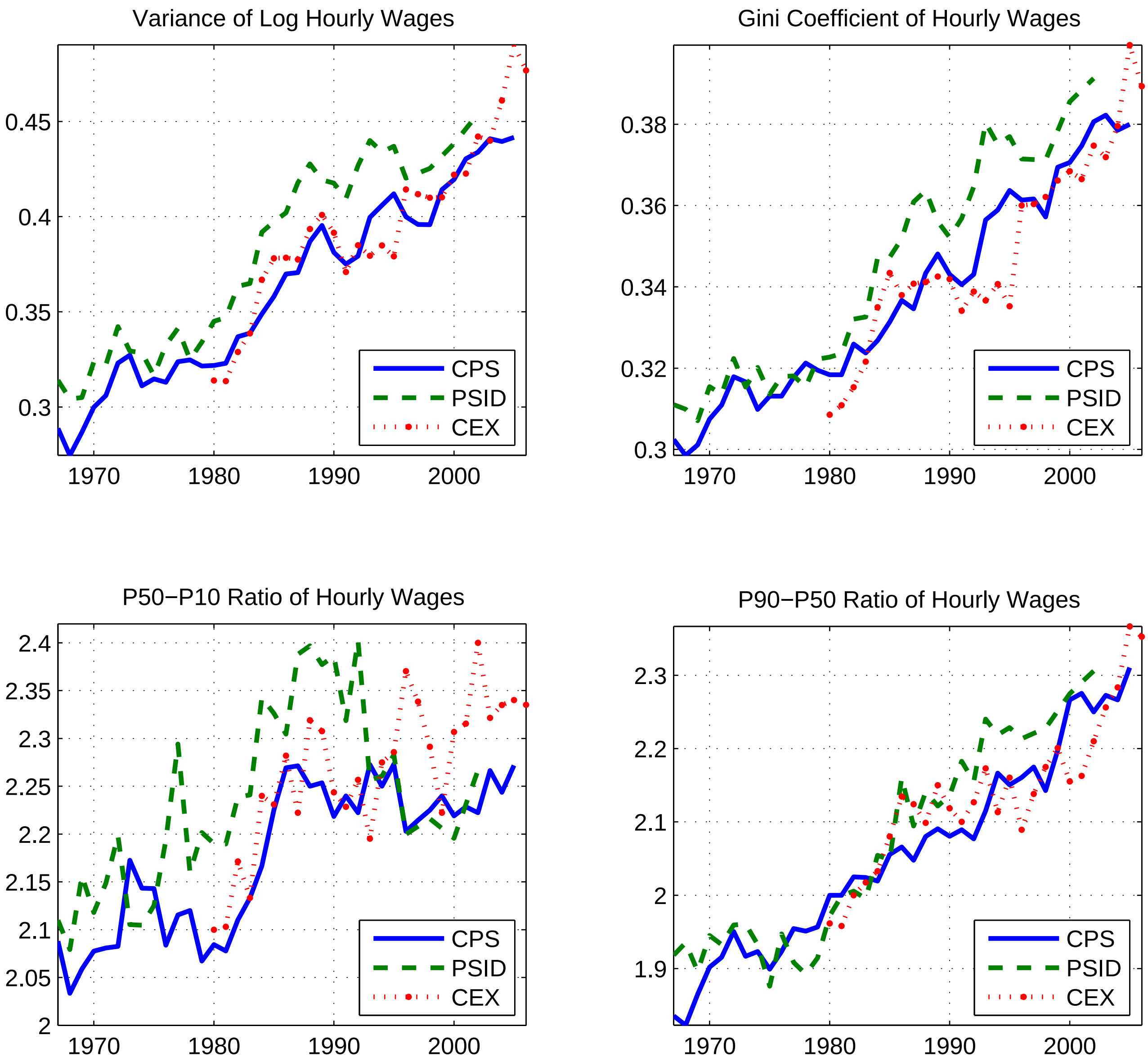


Figure 4: Wage Inequality in CPS, PSID, and CEX

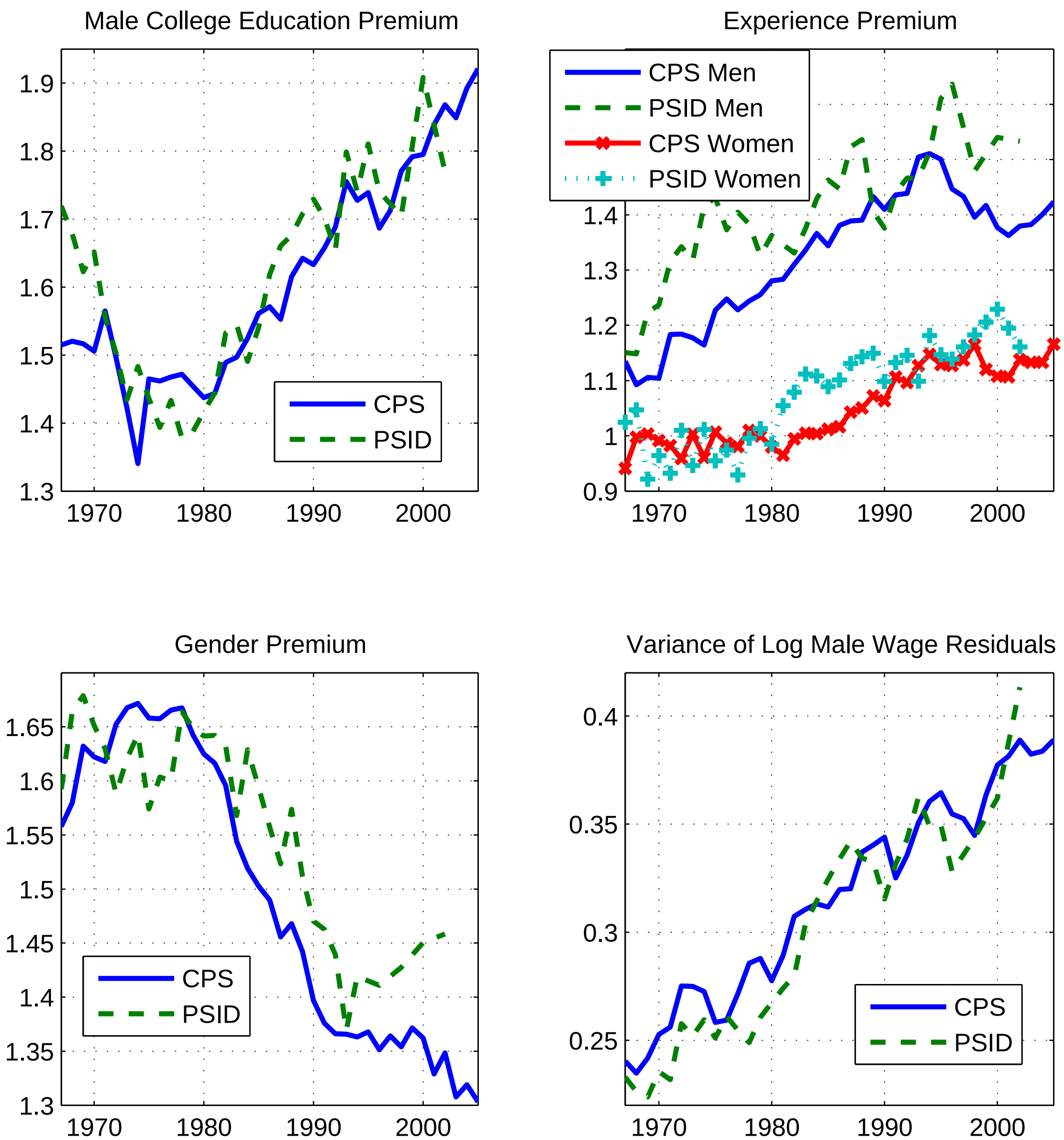


Figure 5: Education, experience and gender wage premia and residual wage inequality in CPS and PSID

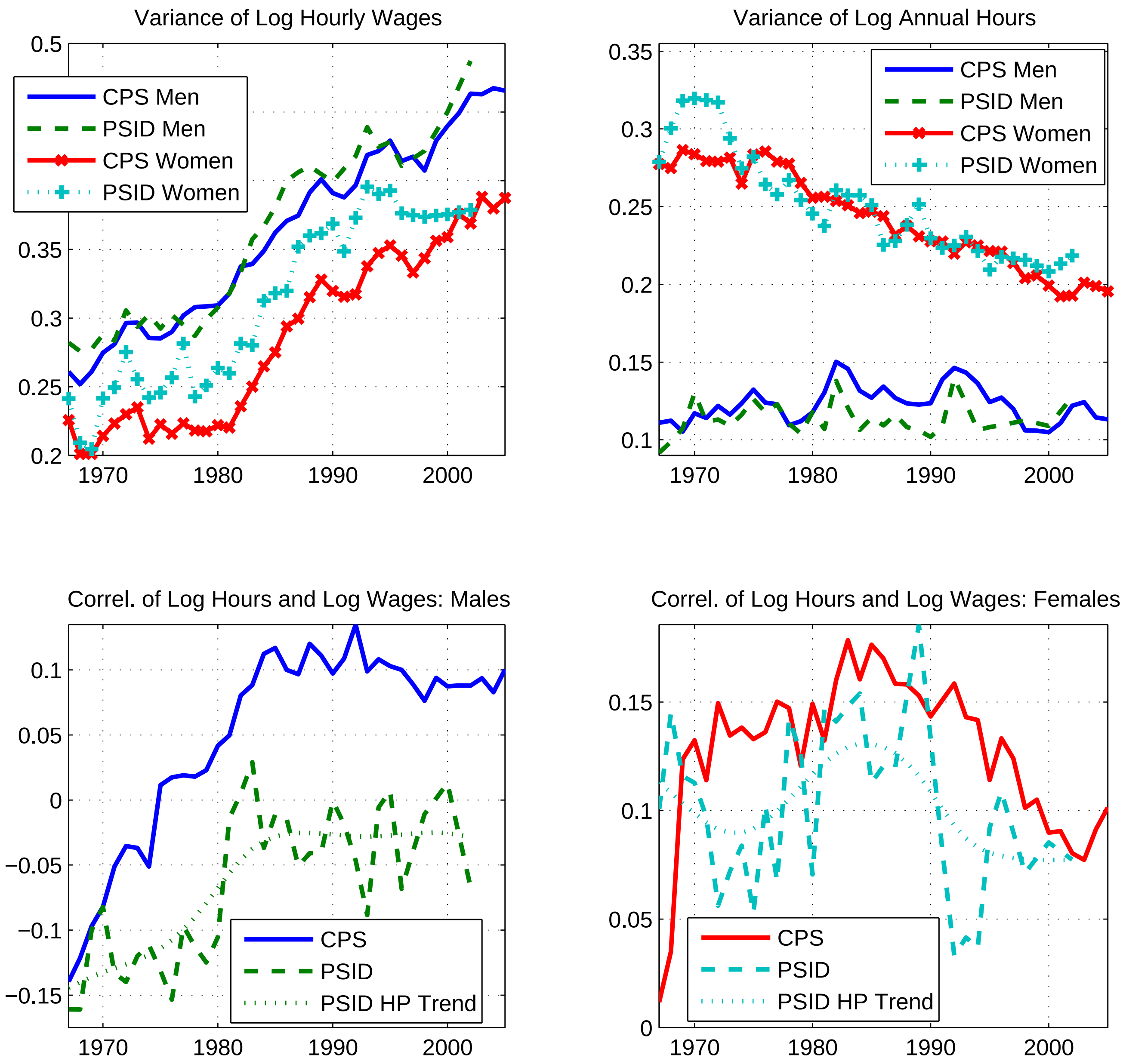


Figure 6: Inequality in labor supply of males and females in CPS and PSID

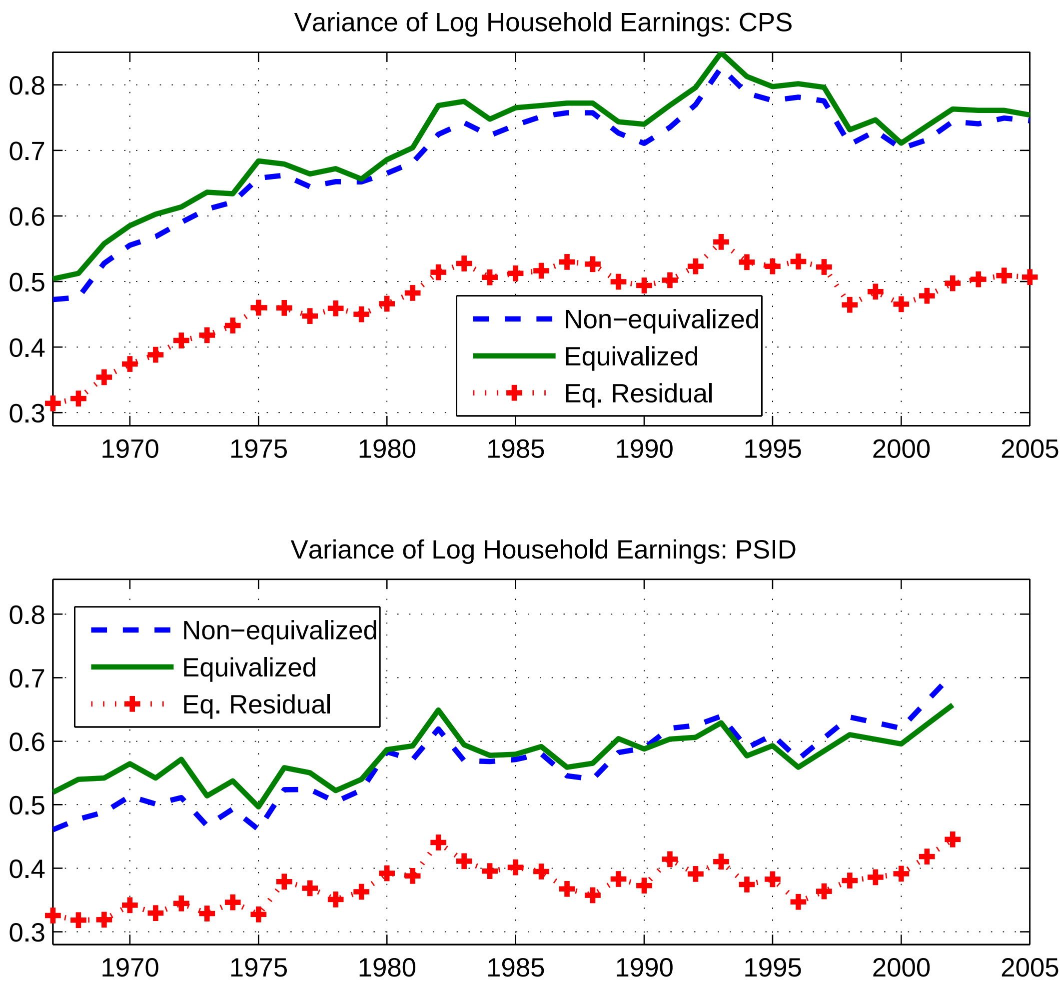


Figure 7: Household earnings inequality in CPS and PSID

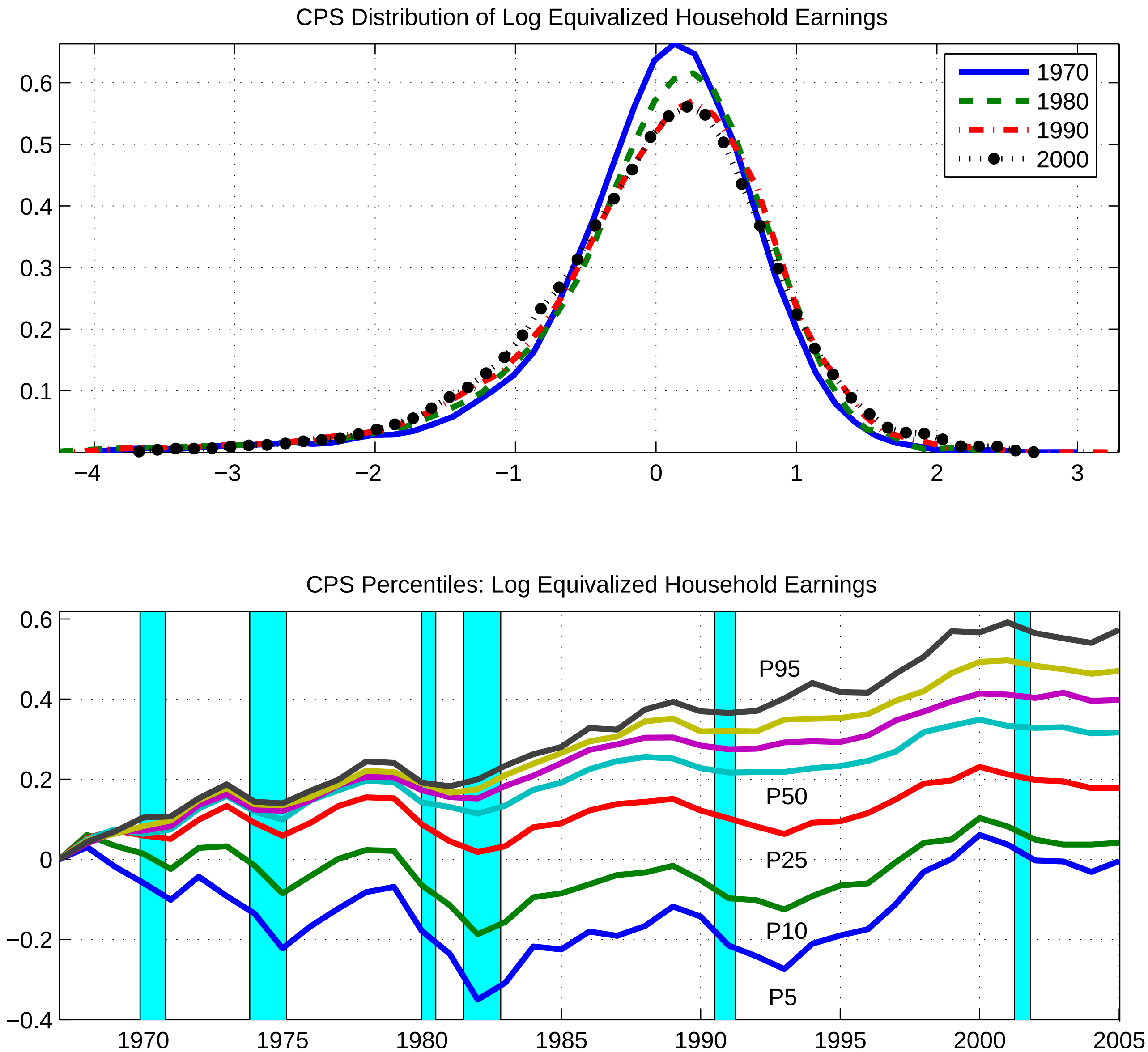


Figure 7-a: Top panel: Earnings densities for decades 1970 to 2000. Bottom panel: trends in percentiles (shaded areas are NBER recessions)

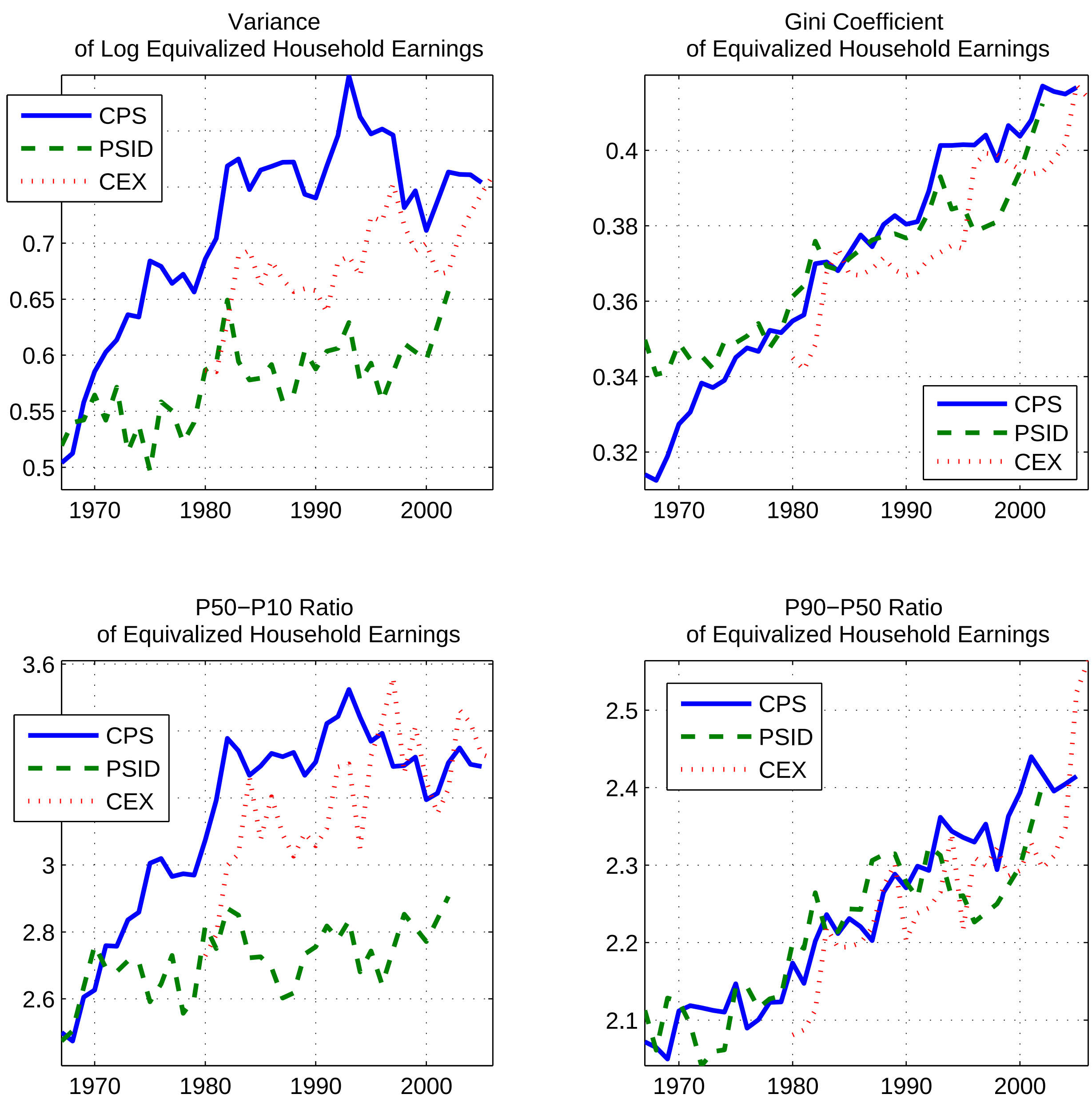


Figure 8: Household earnings inequality in CPS, PSID, and CEX

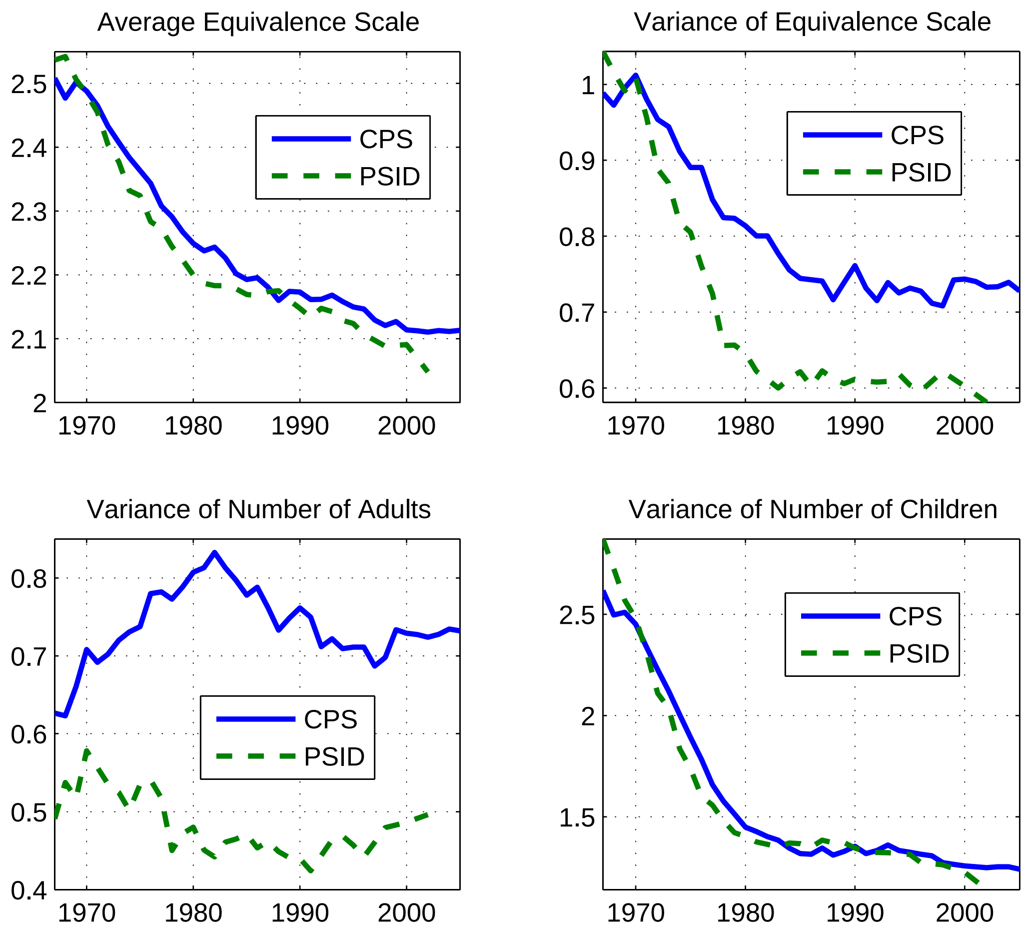


Figure 8-a: Some statistics on the distribution of household size in CPS and PSID

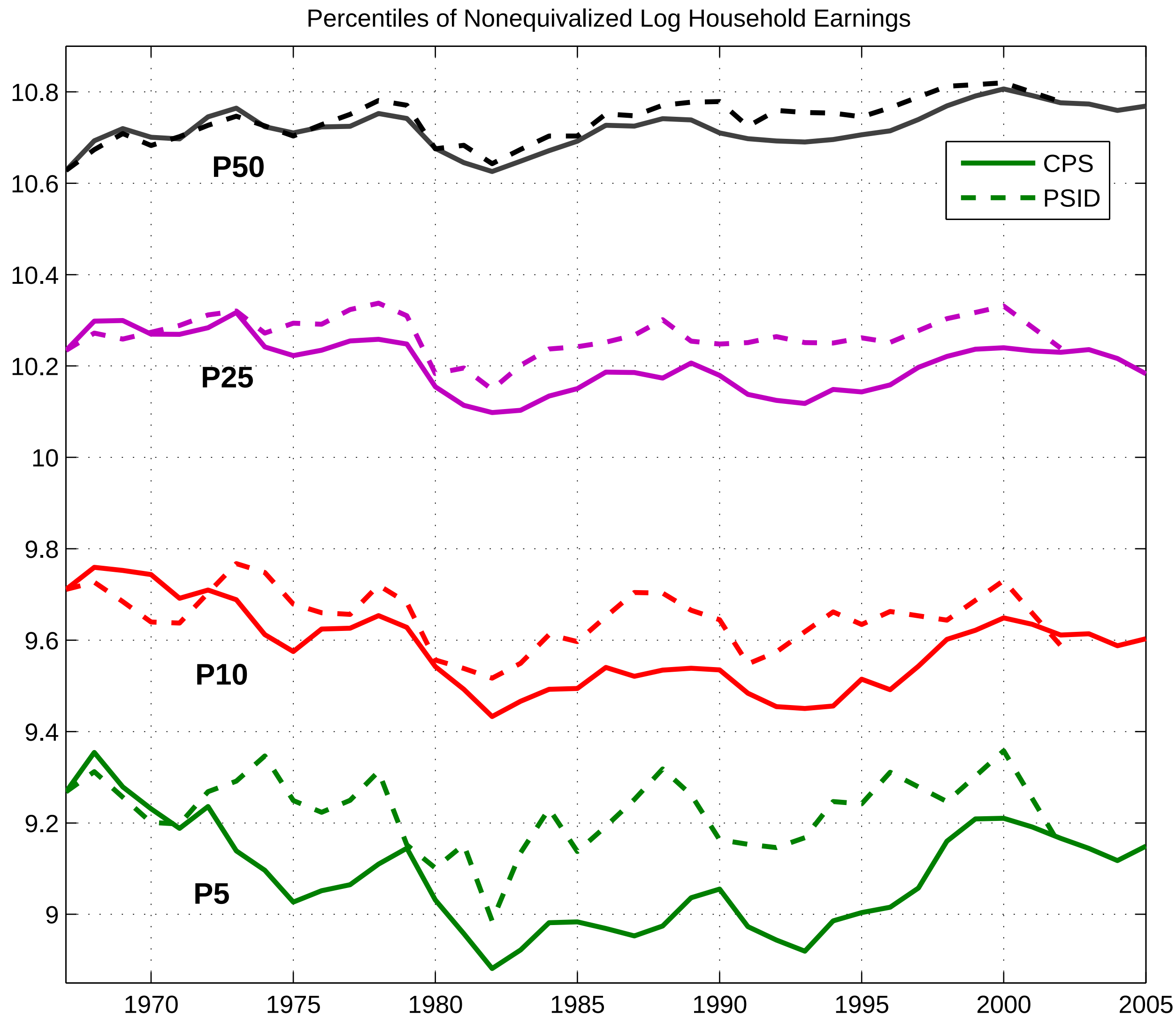


Figure 8-b: Percentiles of the unequivalized household earnings distribution in CPS and PSID

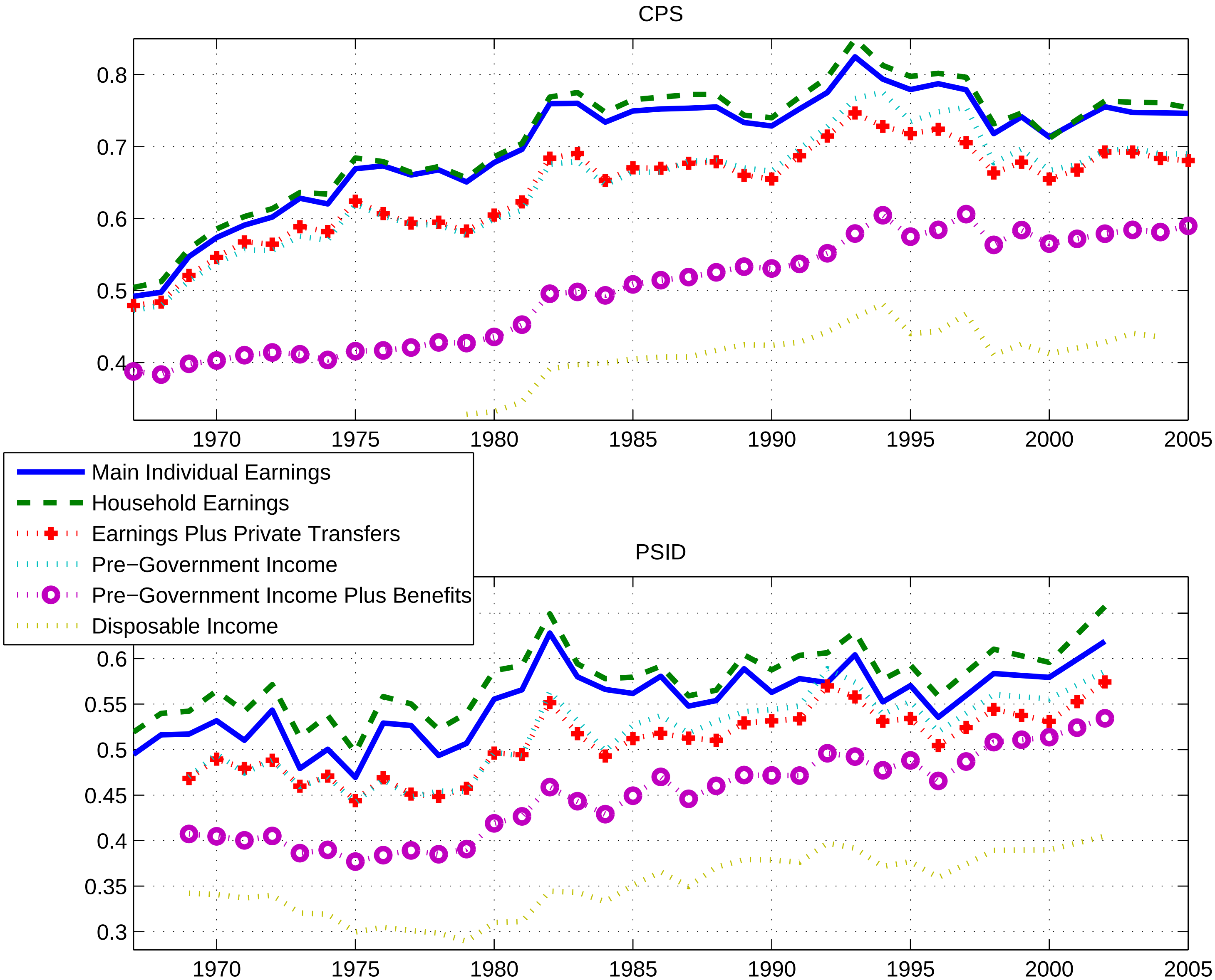


Figure 9: From equivalized individual earnings to household disposable income: CPS and PSID

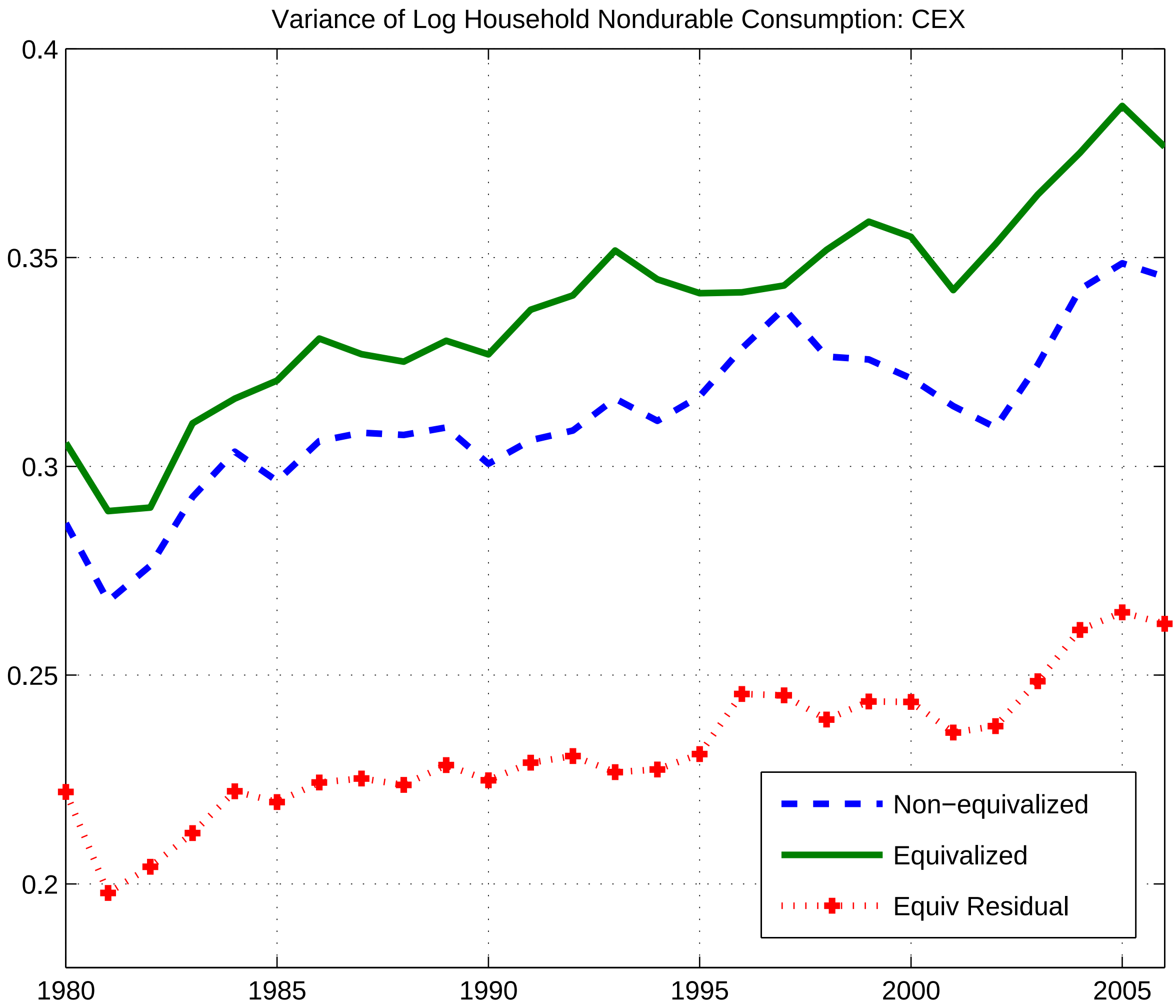


Figure 10: Nondurable consumption inequality from CEX

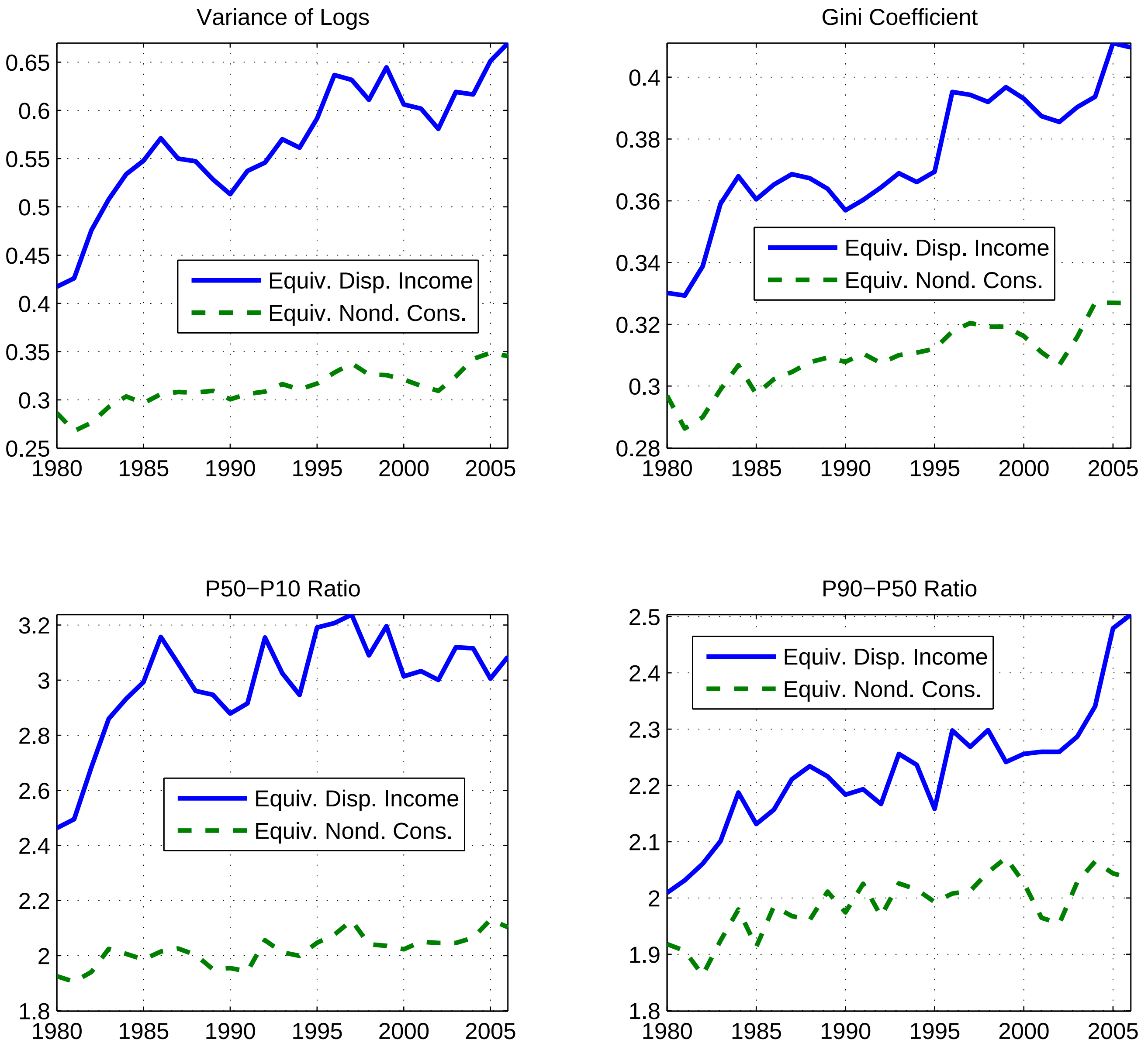


Figure 12: Inequality in equivalized disposable income and nondurable consumption from CEX

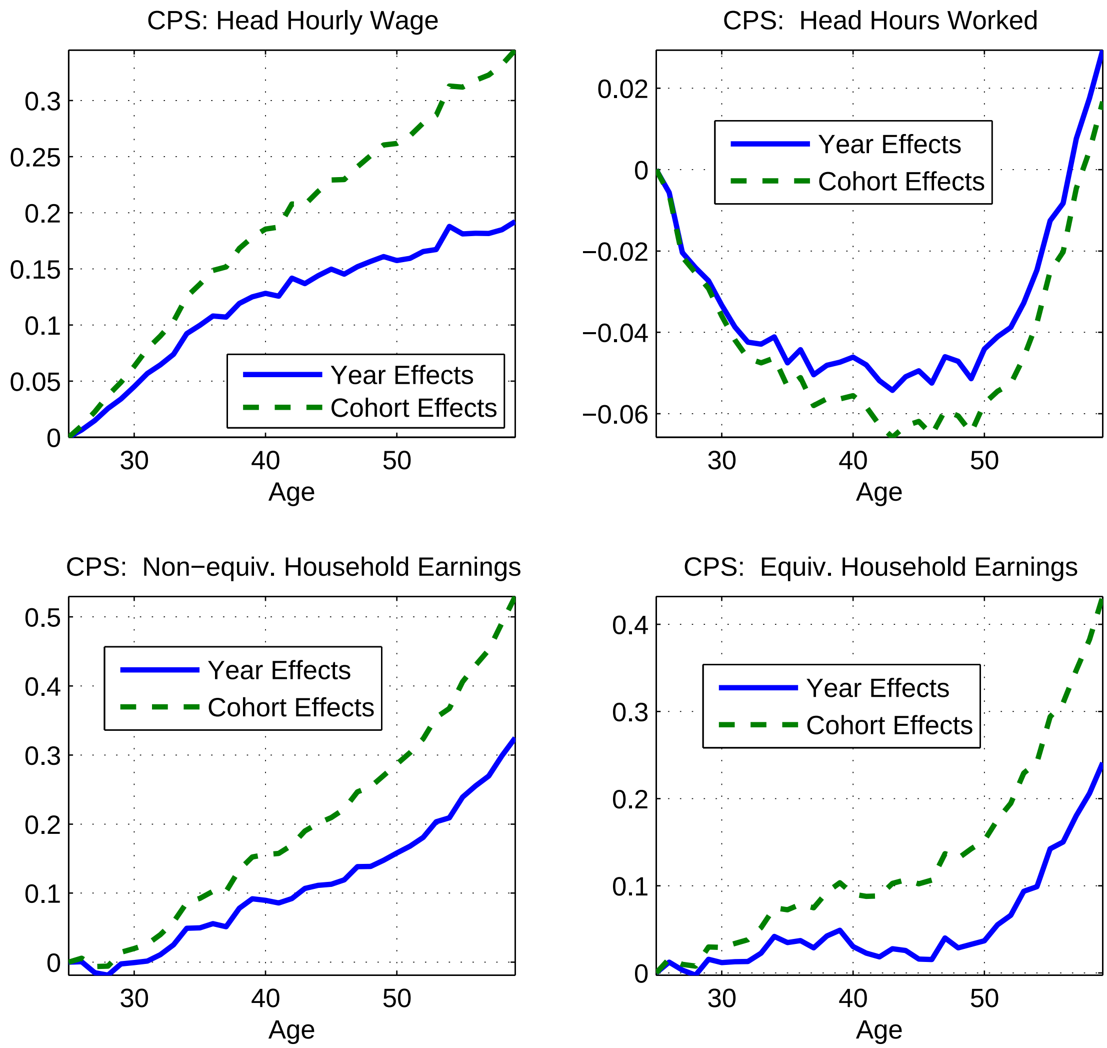


Figure 13: Life-cycle inequality from CPS: controlling for time and cohort effects

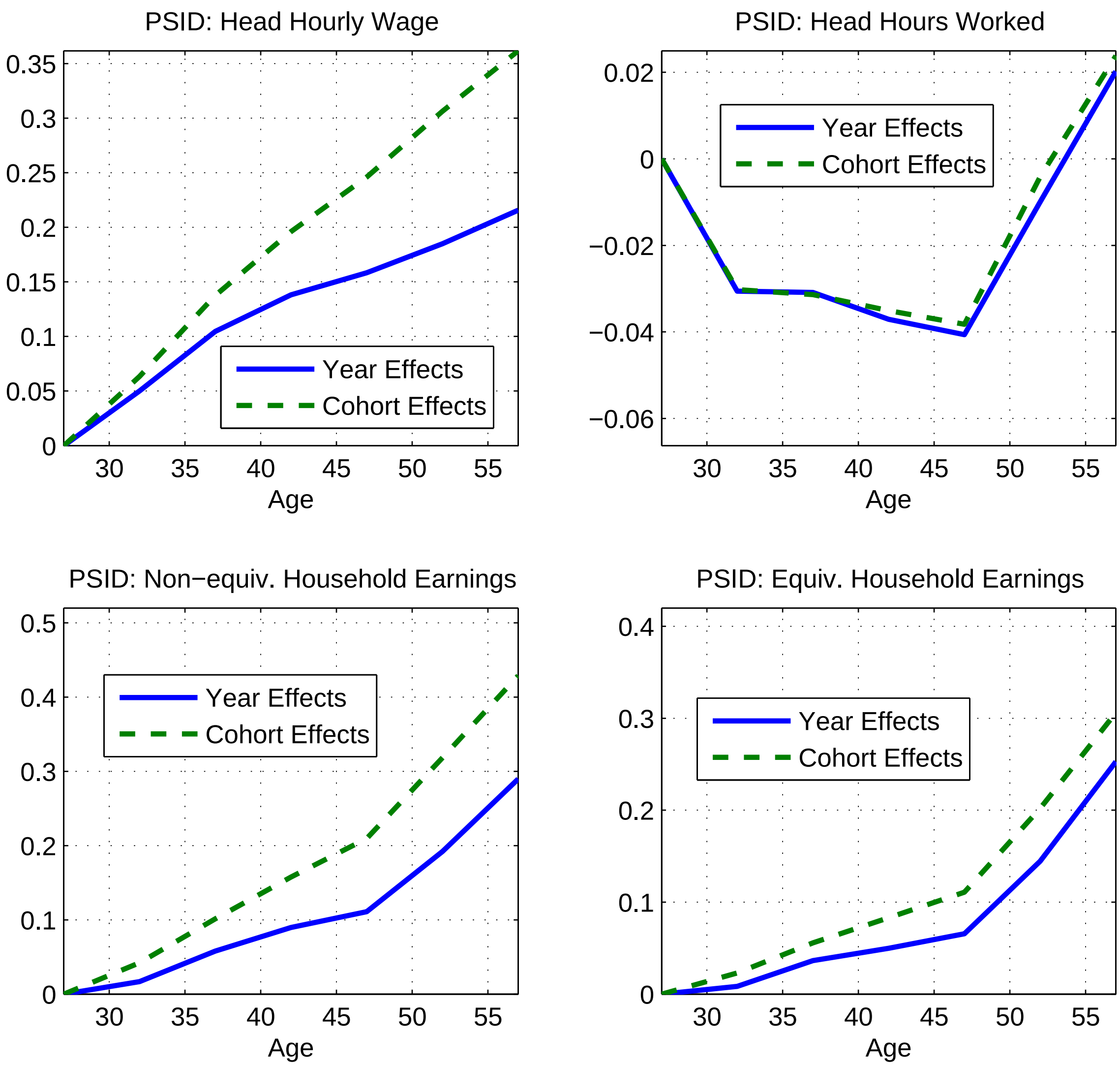


Figure 14: Life-cycle inequality from PSID: controlling for time and cohort effects

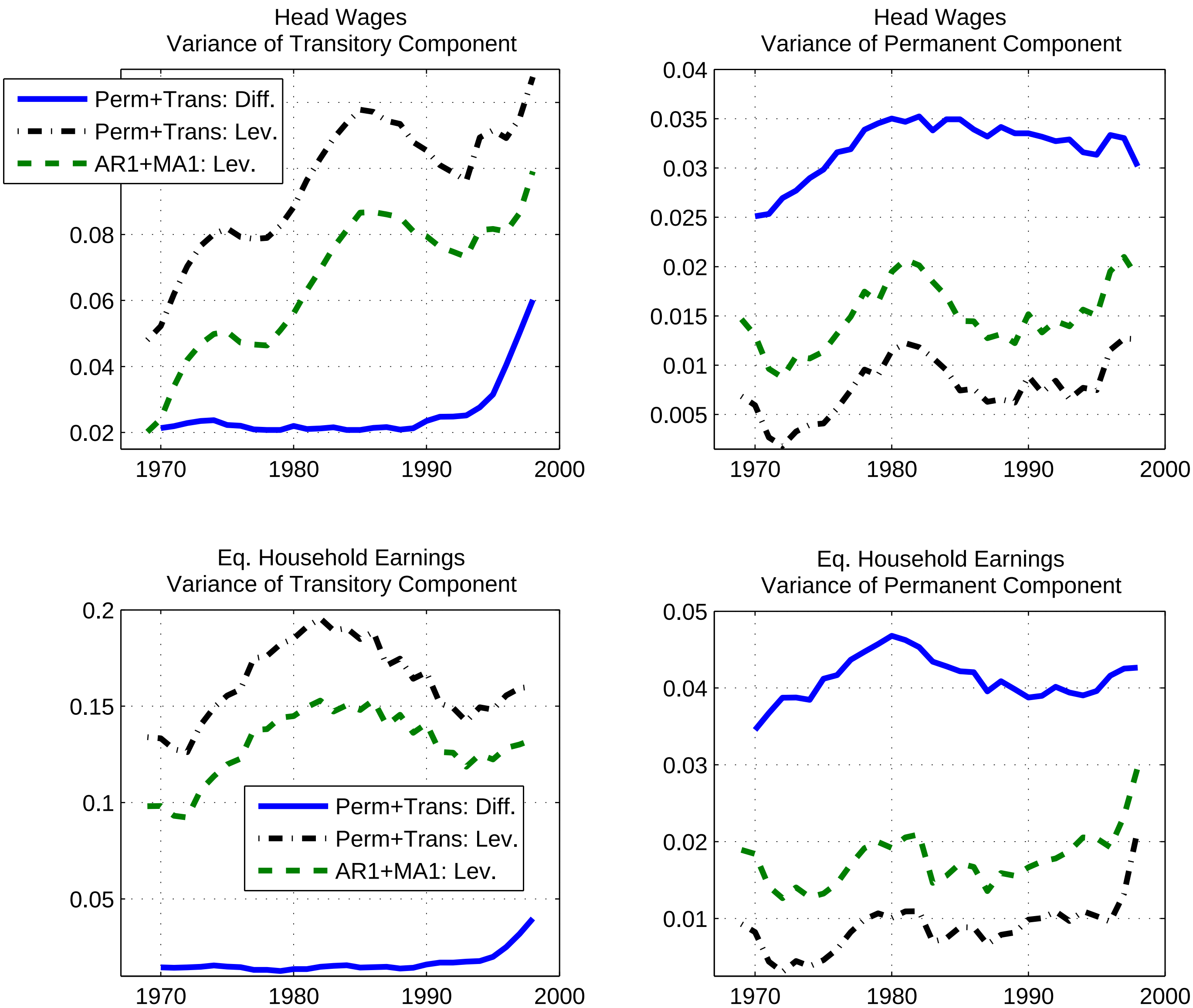


Figure 15: Estimates of the variance od transitory and permanent/persistent components (5-year moving averages)

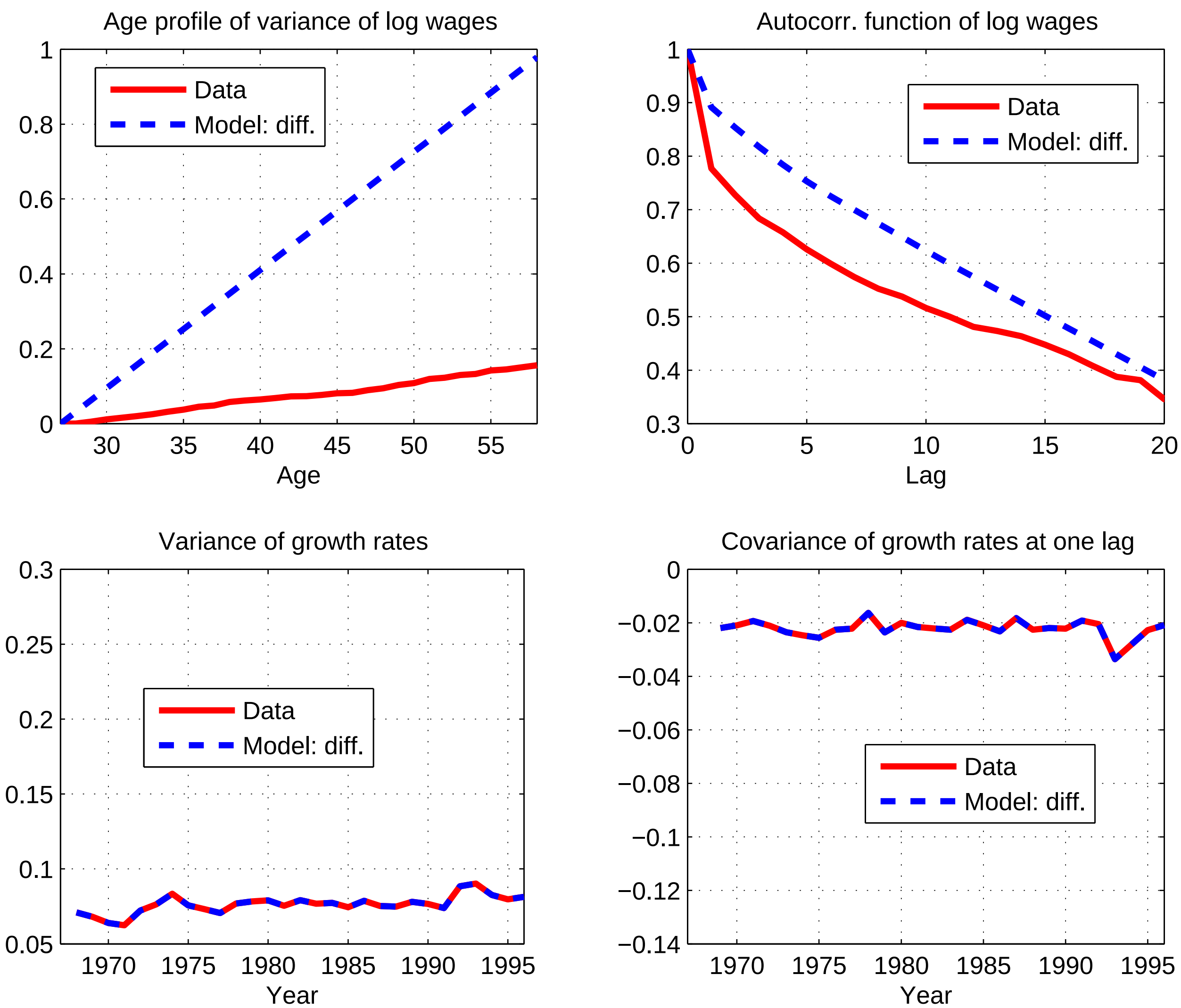


Figure 16: Fit of the permanent-transitory model estimated in first differences

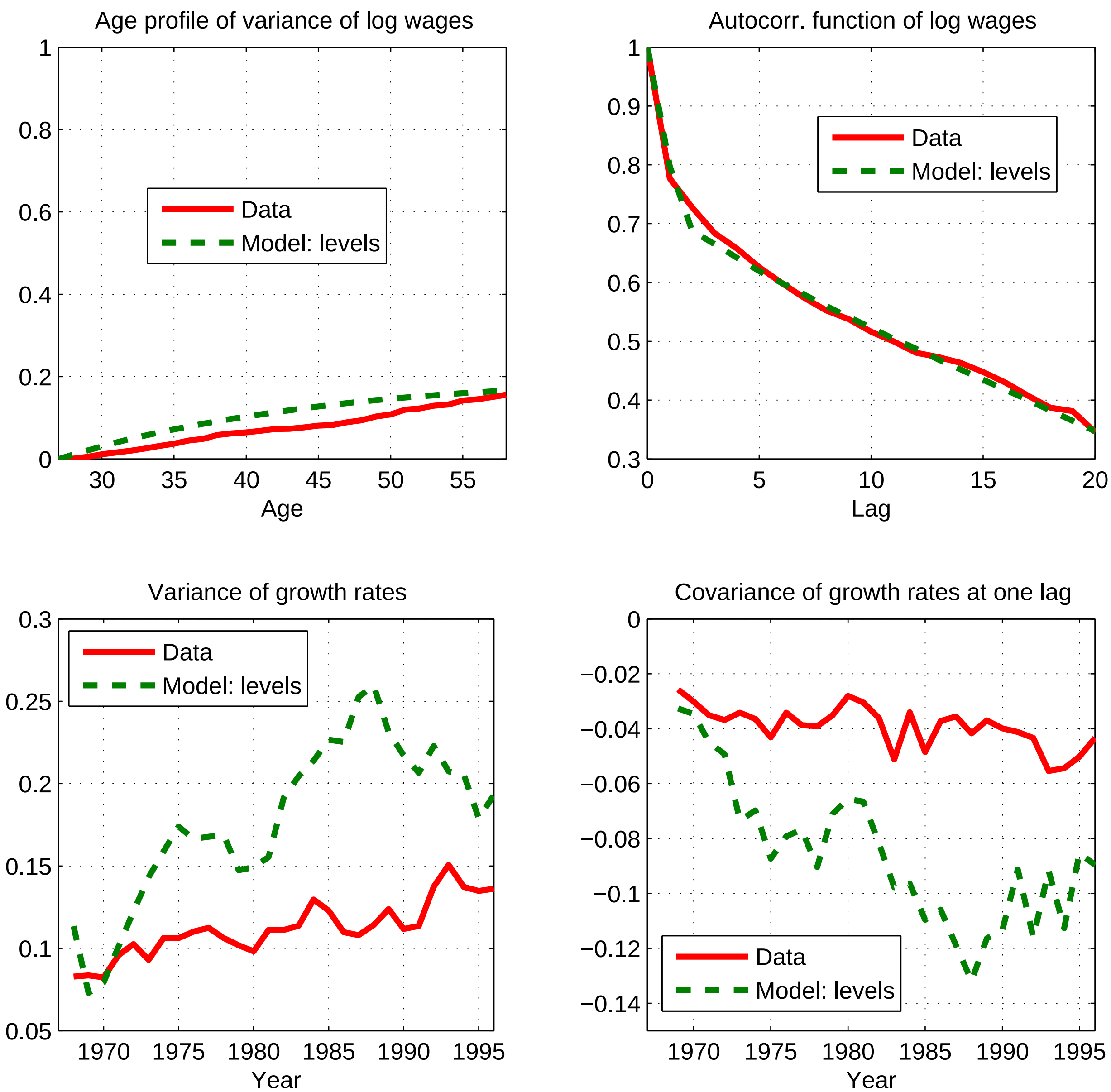


Figure 17: Fit of the AR1-MA1 model estimated in levels