Comparing Estimates of Family Income in the Panel Study of Income Dynamics and the March Current Population Survey, 1968-2007

Elena Gouskova, Patricia Andreski, and Robert F. Schoeni
Survey Research Center, Institute for Social Research
University of Michigan

March, 2010

Comparing Estimates of Family Income in the Panel Study of Income Dynamics and the March Current Population Survey, 1968-2007

Elena Gouskova, Patricia Andreski, and Robert F. Schoeni
Institute for Social Research
University of Michigan

March 2010

I. INTRODUCTION

The Panel Study of Income Dynamics (PSID) is a nationally representative longitudinal study of families and individuals that began in 1968. The initial focus of the PSID was to examine dynamics of employment, earnings, and income over the life cycle through interviews with roughly 5,000 families. The PSID continues to interview many of these same families today, as well as their descendents.

Although the PSID has always had a high response rate of 94-98 percent between each wave, cumulative non-response over the 39-year period is substantial. Moreover, several changes have been implemented to the PSID since the mid-1990s, including the following:

- Change from a Pencil and Paper Telephone Interview to a Computer-Assisted Telephone Interview in 1993;
- Suspension of roughly one-half of the low-income sample in 1997;
- Addition in 1997 of a sample of families who immigrated to the US since 1968;
- Switch to biannual interviewing in 1999; and
- A doubling of the length of the interview between 1995 and 1999.

As a result, it is important to continually reassess the quality of the PSID data. In this report we investigate the quality of one of the most important data elements – total family income – updating through 2007 a similar report that examined these data through the 2005 wave (http://psidonline.isr.umich.edu/Publications/Papers/tsp/2007-01_Comparing_Estimates_PSID.pdf).

One way to examine the quality of the data is to compare it with a gold standard, or a set of estimates that are widely believed to be highly accurate. For family income, such a gold standard does not exist. However, perhaps the most widely used data source for

cross-sectional estimates of family income in the United States is the March Current Population Survey (CPS), which is the basis for the government's official estimates of income and poverty. The objective of this study is to compare estimates of family income between the PSID and the CPS for the entire history of the PSID, survey years 1968 through 2007. Our approach is to use visualization techniques to assess qualitatively the disparities in the empirical distributions of income in the PSID and CPS.

Our results show that the distributions match fairly closely in the range between the 5th and 95th percentiles throughout the 39-year history of the PSID; historically the PSID estimates have been somewhat higher than the CPS estimates, but the trends are quite similar. The two data sets show less agreement at the upper and lower five percentiles of the distribution.

In the next section we briefly describe the data and discuss methodological difficulties related to the comparison of PSID and CPS data. The results are reported in Section III, while the final section summarizes and discusses next steps.

II. DATA

In the analysis we use CPS total household income data and PSID total family income data beginning in the first year that the PSID data were collected – 1968 – through the latest year of data – 2007. All PSID income data are publicly available in the PSID Data Center (at www.psidonline.org). For the CPS we used the version of the data distributed by Unicon Research Corporation (www.unicon.com).

The annual PSID sample size ranges from about 5,000 to 8,000 families. The corresponding numbers for CPS samples of households are roughly 46,000 to 80,000. Because of its large sample, the CPS is able to capture distributional characteristics of the whole population relatively accurately. To correct for the non-randomness in both data sets, weights are used to calculate all estimates. For the PSID, the core family weights are used. In the PSID analysis we analyzed the sample of core families, i.e. families directly related to the original sample of 1968 plus the immigrant sample added in 1997; the Latino sample that was interviewed from 1990 to 1995 was not included.

Both surveys collect income during the calendar year prior to the year in which the data were collected. For example, the (survey year) 2007 March CPS collects data on

income received during (income year) 2006. All estimates are expressed in constant 2006 dollars using the CPI-U. (ftp://ftp.bls.gov/pub/special.requests/cpi/cpiai.txt).

The comparison of PSID and CPS data is not straightforward. The major difficulty is that the surveys use different definitions of "family." This difference comes from the fact that the PSID is a longitudinal study following the same set of families over nearly 40 years. In order to do this successfully the PSID adopted a definition of family that fits the study design. As a result, the PSID definition of family is broader, encompassing unmarried couples living together and sharing resources as well as single-person "families."

In order to obtain the most comparable estimates of income, we base our analysis on the CPS <u>household</u> unit rather then on the <u>family</u> unit. The definition of CPS household comes closer to matching PSID family than does CPS family. However, while close, the PSID family and the CPS household are still not the same concepts. First, not all people living within a household that contains a PSID family are members of that family. Furthermore, the PSID does not collect information on income of household members who are not members of the PSID family. For this reason we would expect the PSID estimate of family income to be lower than the CPS estimate of household income.

One of the other major differences between the CPS definition of household and the PSID concept of a family is represented by cases where one or more PSID families reside in the same household. This happens, for example, when a grown child marries and leaves the parental home to live independently, but then eventually comes back to live with their parents. It is PSID practice to treat the parent's family and the new adult child's family (even if it consists of a single person) as separate "families" and obtain full, independent interviews from both of them.

When the PSID began in 1968 each PSID household had only one PSID family. But over time as family members split up and then joined back together again, the share of PSID family units living in a household with another PSID family unit began to rise. Since about the late 1970s, roughly 4 to 9 percent of PSID families lived in the same household as another PSID family (Table 1). The drop in 1997 is explained by the fact that after 1996 approximately 2,000 low-income families were dropped from the study. This drop also suggests that the practice of living in the same dwelling is more likely

3

among low-income families. The drop in 1994 is due to the fact that a large number of families that had attrited prior to 1993 were brought back into the PSID in this year. During this effort to bring back attriting families, it was also decided that if there were more than one PSID attriting family living in the same household, only one interview would be conducted, thereby merging all family units into one.

In order to account for the fact that more than one PSID family unit may live in the same household, we aggregated income for all family units living together; we refer to this new measure of income as "PSID aggregated family income." We use the term "aggregated family" rather than "household" to underscore the fact that this unit still does not represent a household as defined by the CPS. For the aggregated families, we used a simple average of the weights of these families to serve as their weights in all analyses reported in this study.

III. RESULTS

PSID Aggregate Income Versus PSID Family Income

The first step is to aggregate income for PSID households with more than one PSID family. A comparison of the 50th percentile of the distribution of total family income and total aggregated family income is shown in Figure 1. After the late-1970s, on average, aggregated family income is about 5 percent higher.

In a prior study using data through the 2005 wave ((http://psidonline.isr.umich.edu/Publications/Papers/tsp/2007-01_Comparing_Estimates_PSID.pdf)), we investigated the extent to which the increase due to aggregating family income is larger for low versus high income families. We restate the findings here which are summarized in Figure 2, panels A-C. In panel A we plot the ratio of aggregated family income to "unaggregated" family income for the 5th percentile versus the 90th percentile; each income year 1967-2004 generates a combination of estimates for the 5th and 90th percentiles, so there are 34 data points on the chart (because PSID began interviewing biannually after 1997). We see that at the bottom 5th percentile, the ratio ranges from 1.0 (typically in the first 5-10 years of PSID interviewing before PSID families began co-residing) to 1.15, with a substantial number of years between 1.05 and 1.10. Aggregating is much less important (in relative terms)

4

for the higher income families, with aggregated income never more than 5 percent higher than unaggregated income. As a result, almost all of the data points fall below the 45-degree line in Panel A; that is, aggregating income among households with more than one PSID family is more important for the bottom of the income distribution than the top.

Panels B and C plot the 5th percentile versus the 50th and 30th percentiles, respectively, to determine whether aggregating income is only important for the very poorest families. We find that for many years aggregation is actually more important for families at the median of the distribution than for families at the 5th percentile. The evidence is even stronger at the 30th percentile. In sum, aggregation of families within the same household is quite important for the poorest families, of relatively little importance for the highest income families, but also of importance for middle-income families.

PSID Versus CPS

The central objective of the study is to compare different percentiles of the distribution of total aggregate family income in the PSID with total household income in the CPS. The official tabulations of the 20th, 40th, 60th, 80th and 95th household income percentiles based on the CPS are provided by US Census Bureau.[1] For our analysis we want to compare the full distribution, not just these five percentiles. Therefore, we have calculated our own estimates based on the CPS. To make sure that we are using the CPS data appropriately, we first compared our calculations with the published tabulations, and this comparison is displayed in Table 2. For almost every year and percentile, our estimates are within 1 percent of the published tabulations. Therefore, we proceed to examine additional points in the income distribution.

The results of the analysis are summarized in Figures 3A-3D for various points in the income distribution. The PSID and CPS track each other fairly closely throughout the 39 years of the panel study. Moreover, this result is true for all points in the distribution between, roughly, the 5th and 95th percentile. It is only the tails of the distribution where the estimates diverge substantially.

Two exceptions should be noted. First, the PSID estimate for income (survey) year 1992 (1993) is unusually high relative to both the CPS for 1992 and the PSID in 1991

---

[1]These estimates are available at: http://www.census.gov/hhes/www/income/histinc/h01ar.html.

and 1993. This divergence is especially large for the 70th and 80th percentiles. Second, the peak of the boom in the late 1980s was 1989 according to the CPS, while for the PSID the peak was 1-2 years earlier at most percentiles.

IV. CONCLUSIONS

This report examines the comparability of the estimates of total family/household income reported in the PSID and the CPS. At almost all points in the distribution, we find that the estimates based on the PSID are higher than the estimates based on the CPS. Moreover, the magnitude of the gap is fairly constant through the 39-year history of the PSID. While there are some unexplained differences that need to be investigated, the close agreement in the trends is remarkable given the substantial differences between the two surveys and the amount of change that both surveys have undergone during the past four decades.

Table 1. Share of PSID Families Living in the Same Household
as Another PSID Family

Survey year	Percent	Year	Percent
1968	0.0%	1985	8.8%
1969	0.0%	1986	8.5%
1970	0.0%	1987	8.4%
1971	0.3%	1988	8.8%
1972	1.0%	1989	9.1%
1973	2.5%	1990	8.3%
1974	2.0%	1991	8.6%
1975	2.9%	1992	9.3%
1976	3.1%	1993	9.1%
1977	3.6%	1994	6.7%
1978	4.3%	1995	7.1%
1979	4.0%	1996	7.3%
1980	6.5%	1997	4.7%
1981	7.6%	1999	4.5%
1982	8.8%	2001	4.8%
1983	8.9%	2003	5.4%
1984	8.7%	2005	6.1%
		2007	6.3%

	20th Percentile		40th Percentile		60th Percentile		80th Percentile		95th Percentile	
Income year	Our Estimate	Ratio*	Our Estimate	Ratio*	Our Estimate	Ratio*	Our Estimate	Ratio*	Our Estimate	Ratio*
1967	3,000	1.00	5,800	1.01	8,254	1.01	11,700	1.01	18,200	1.04
1968	3,323	1.00	6,300	1.00	9,030	1.00	12,688	1.00	19,850	1.00
1969	3,600	0.99	6,884	1.00	9,937	1.00	13,900	1.00	21,769	1.00
1970	3,687	1.00	7,065	1.00	10,276	1.00	14,661	1.00	23,175	1.00
1971	3,800	1.00	7,244	1.00	10,660	1.00	15,200	1.00	24,138	1.00
1972	4,050	1.00	7,800	1.00	11,530	1.00	16,500	1.00	26,560	1.00
1973	4,418	1.00	8,393	1.01	12,450	1.00	17,985	1.00	28,509	1.02
1974	4,758	1.02	8,943	1.01	13,143	1.01	19,048	1.01	30,280	1.01
1975	5,000	1.00	9,384	1.00	14,180	1.00	20,360	1.00	32,129	1.00
1976	5,405	1.00	10,070	1.00	15,340	1.00	22,070	1.00	35,000	1.00
1977	5,734	1.00	10,800	1.00	16,462	1.00	24,000	1.00	38,000	1.00
1978	6,318	1.00	11,946	1.00	18,075	1.00	26,288	1.00	42,050	1.00
1979	7,000	1.00	13,024	1.00	20,030	1.00	29,067	1.00	47,000	1.00
1980	7,478	1.00	14,020	1.00	21,500	1.00	31,474	1.00	50,300	1.01
1981	8,024	1.00	15,000	1.00	23,200	1.00	34,300	1.00	55,200	1.00
1982	8,400	1.00	15,976	1.00	24,410	1.00	36,398	1.00	60,040	1.00
1983	8,819	1.01	16,500	1.01	25,379	1.01	38,325	1.01	62,835	1.01
1984	9,500	1.00	17,780	1.00	27,393	1.00	41,380	1.00	68,500	1.00
1985	9,941	1.00	18,704	1.00	28,975	1.00	43,550	1.00	72,004	1.00
1986	10,247	1.00	19,600	1.00	30,419	1.00	45,950	1.00	77,091	1.00
1987	10,800	1.00	20,500	1.00	32,000	1.00	48,363	1.00	80,928	1.00
1988	11,382	1.00	21,500	1.00	33,506	1.00	50,593	1.00	85,640	1.00
1989	12,096	1.00	23,000	1.00	35,350	1.00	53,710	1.00	91,733	1.00
1990	12,500	1.00	23,662	1.00	36,200	1.00	55,205	1.00	94,700	1.00
1991	12,588	1.00	24,000	1.00	37,070	1.00	56,760	1.00	96,399	1.00
1992	12,664	0.99	24,300	0.99	38,000	1.00	58,200	1.00	99,270	1.00
1993	12,967	1.00	24,679	1.00	38,793	1.00	60,300	1.00	102,800	1.02
1994	13,426	1.00	25,200	1.00	40,100	1.00	62,841	1.00	106,546	1.03
1995	14,400	1.00	26,914	1.00	42,004	1.00	65,276	1.00	113,942	0.99
1996	14,768	1.00	27,760	1.00	44,012	1.00	68,150	1.00	119,834	1.00
1997	15,400	1.00	29,200	1.00	46,000	1.00	71,700	1.00	127,590	0.99
1998	16,116	1.00	30,412	1.00	48,509	1.00	75,400	0.99	132,876	0.99
1999	17,196	1.00	32,000	1.00	50,685	0.99	80,000	0.99	142,894	0.99
2000	17,920	1.00	33,004	1.00	52,372	1.00	82,250	0.99	146,825	0.99
2001	17,974	1.00	33,350	1.00	53,150	1.00	84,135	0.99	151,370	0.99
2002	17,916	1.00	33,396	1.00	53,300	1.00	84,322	1.00	150,450	1.00
2003	17,984	1.00	34,000	1.00	54,608	1.00	87,140	1.00	154,953	0.99
2004	18,500	1.00	34,761	1.00	55,600	0.99	88,334	1.00	157,350	1.00
2005	19,100	1.00	36,000	1.00	57,838	1.00	92,048	1.00	166,000	1.00
2006	20,010	1.00	37,812	1.00	60,014	1.00	97,413	1.00	174,020	1.00

Table 2. Income Limites for Each Fifth and Top 5 Percent of households Calculated with CPS Data

"Our estimate" is the estimate we calculated using the CPS microfiles. "Ratio" is the ratio of the published census estimate to "our estimate."

Figure 1. PSID Family Income and PSID Aggregated Family Income, 50th Percentile: 1967-2006
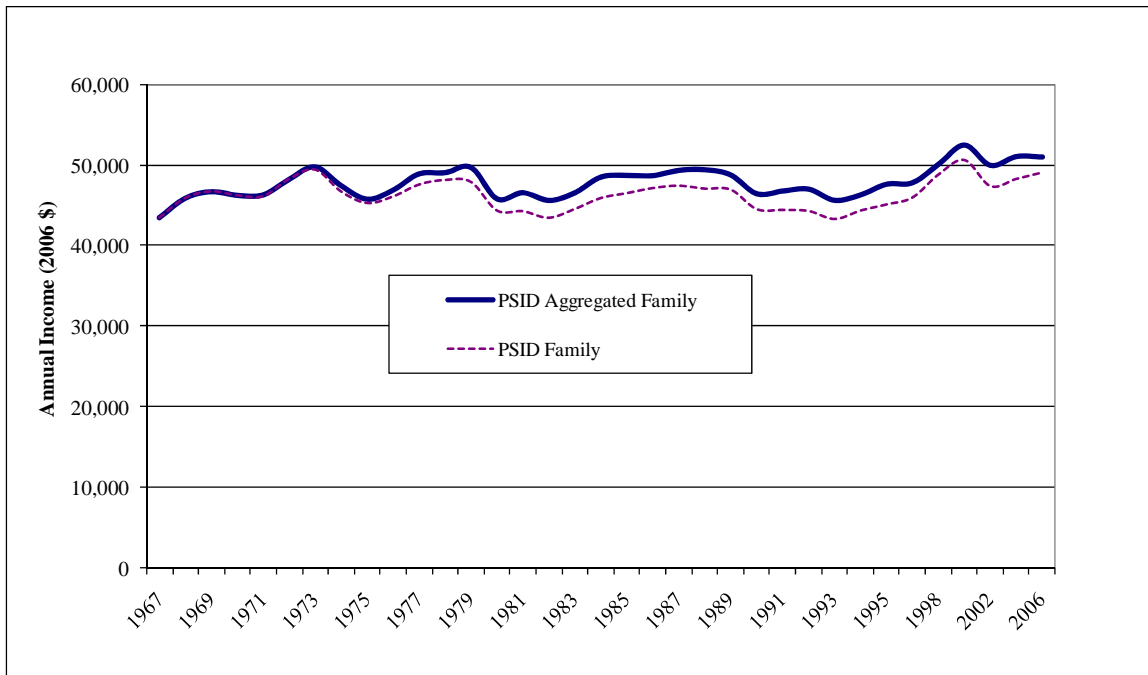
Figure 2. Plots of Relative Values (PSID Aggregate Family Income/PSID Family Income) for Different Percentiles
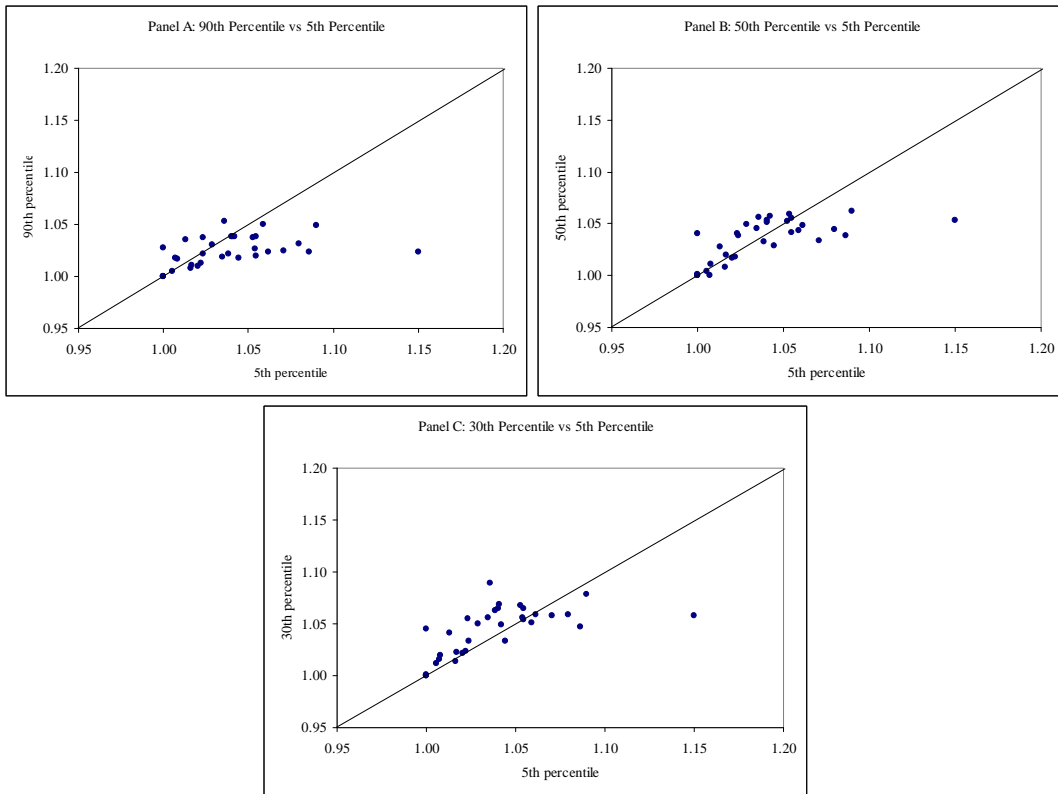[Each point is a given year, 1967-2004]

Panel A: 90th Percentile vs 5th Percentile

Panel B: 50th Percentile vs 5th Percentile

Panel C: 30th Percentile vs 5th Percentile

Figure 3A. 10th-40th Percentiles of PSID Aggregated Family Income and CPS Household Income, 1967-2006
[Top (blue/dotted) line=PSID; Bottom (purple/solid) line=CPS]
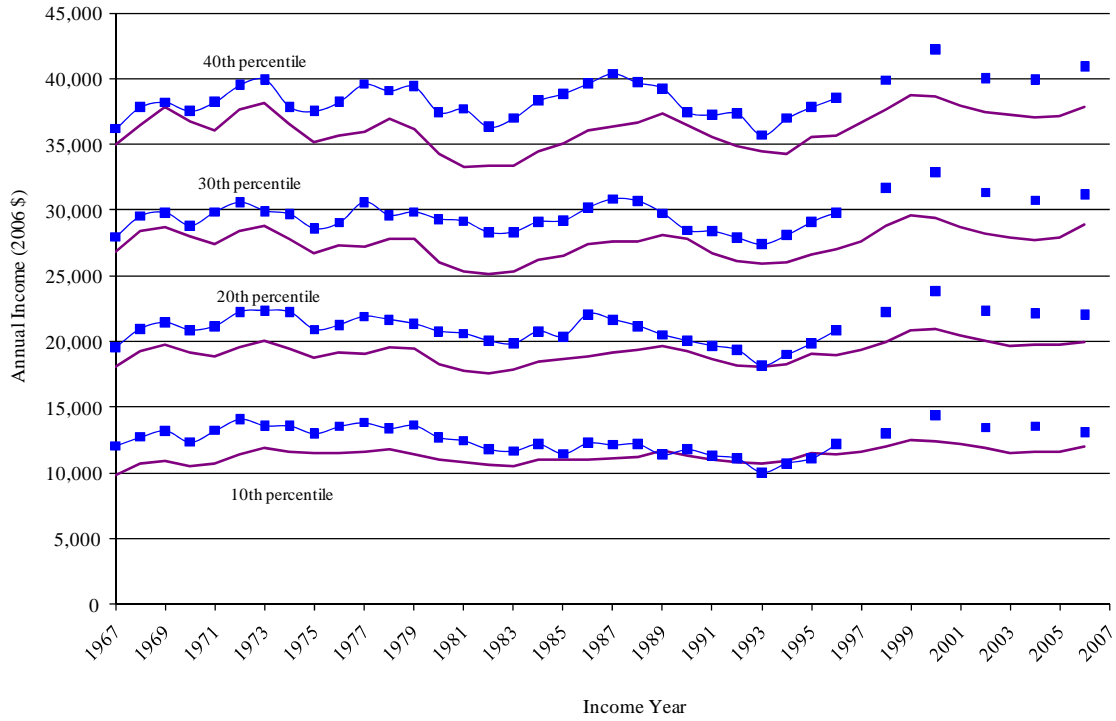
Figure 3B. 50th-80th Percentiles of PSID Aggregated Family Income and CPS Household Income, 1967-2006
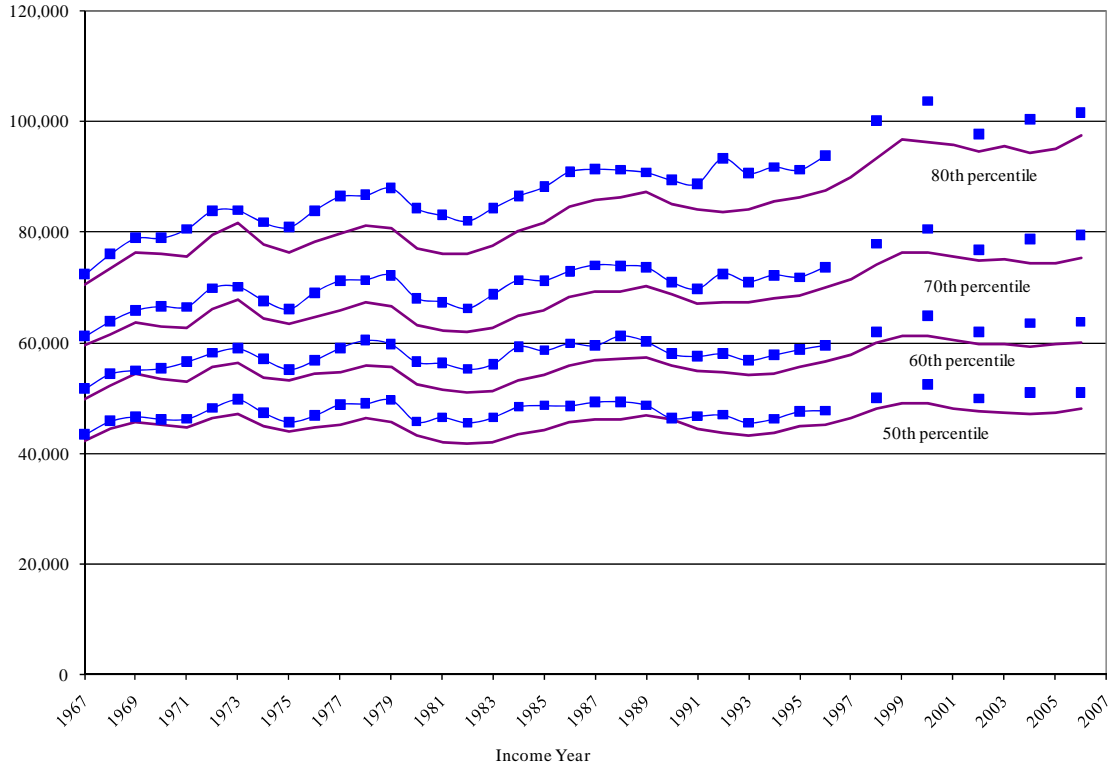[Top (blue/dotted) line=PSID; Bottom (purple/solid) line=CPS]

80th percentile

70th percentile

60th percentile

50th percentile

Income Year

Figure 3C. 1st-5th Percentiles of PSID Aggregated Family Income and CPS Household
Income, 1967-2006
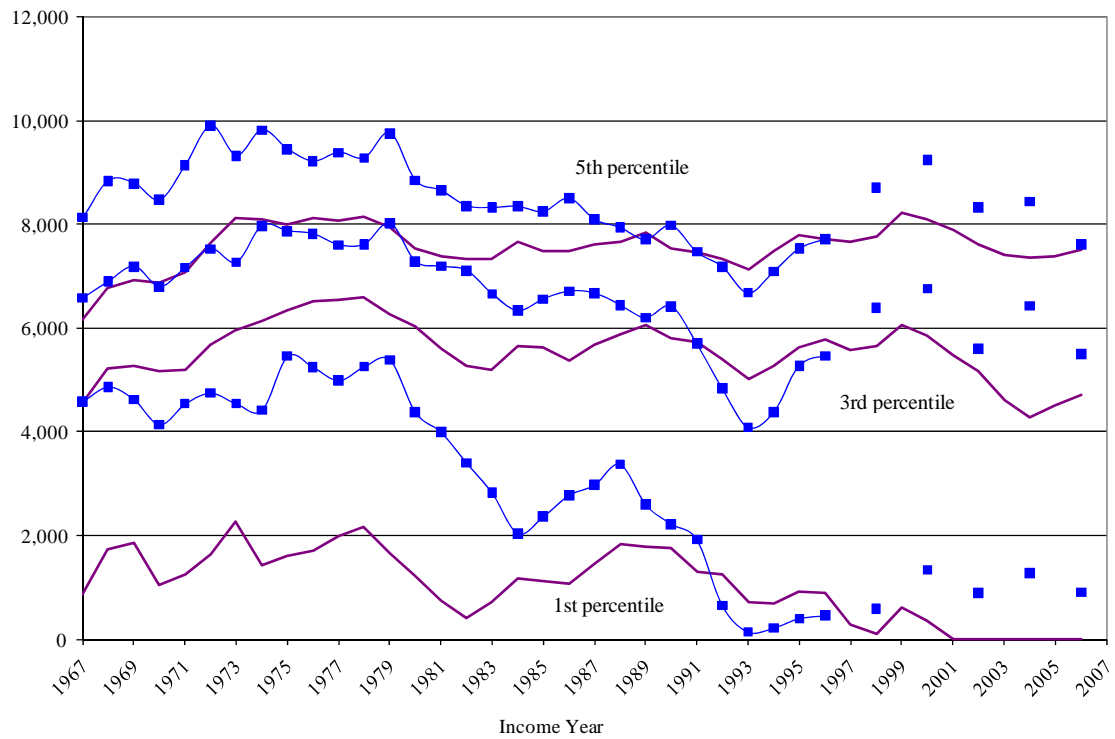[Top (blue/dotted) line=PSID; Bottom (purple/solid) line=CPS]

Income Year

Figure 3D. 90th-99th Percentiles of PSID Aggregated Family Income and CPS Household Income, 1967-2006
[Top (blue/dotted) line=PSID; Bottom (purple/solid) line=CPS]

Income Year