Income Inequality, Mobility and Turnover at the Top in the U.S., 1987 – 2010

Gerald Auten
Geoffrey Gee
And
Nicholas Turner

Cross-sectional Census data, survey data or income tax returns (Saez 2003) generally show increasing income inequality in the U.S. for the past several decades. Other studies have found increasing variability of earnings over this period, which is potentially related and could have contributed to the increases in measured inequality. Considerations of fairness, policy responses and a fuller understanding of inequality suggest that it is important to look at how incomes change over time as well as cross-section distributions.

A previous paper (Auten and Gee, 2009) examined 10-year income mobility over the periods 1987-1996 and 1996-2005 using large panels of tax returns. That paper found that in both periods more than half of individuals age 25 and above changed income quintiles and that about half of those initially in the bottom quintile moved up one or more quintiles. Relative income mobility was virtually identical over the two periods. While the wider income gaps might have been expected to reduce relative income mobility, this was offset by larger changes in absolute income, i.e., greater absolute income mobility, in the more recent period. Median real incomes, adjusted for household size and following individuals when they changed marital status or were in different households, increased by 7.8 percent from 1987 to 1996 and by 22.7 percent from 1996 to 2005. Real incomes increased for 58 percent of individuals from 1987 to 1996 and for 68 percent of individuals in the more recent period.

In our new project, we examine income mobility more broadly by linking a series of tax return files from 1987 onward to population files of tax return and administrative record data to create multiple tax-filer-based panels. Consequently we can examine questions about income mobility along additional dimensions including lifecycle, within-cohort, and intergenerational mobility. We present new results on within-cohort income mobility, persistence and turnover at the top of the income distribution, and the income movements of the generations from 1987 forward.

We begin with the large Statistics of Income (SOI) individual income tax cross-section samples chosen to represent the filing population of a particular year. Information on individual taxpayers and dependents in later years are found in population level tax return data available on the IRS Compliance Data Warehouse (CDW). We supplement the tax return data with Social Security Administration records on birth and death dates and with data from W2, 1099-SSA, 1099-INT, and 1099-DIV tax forms.[1] If no tax return is found, we use the administrative records to estimate the individual's income.

The analysis uses a comprehensive measure of cash income that includes tax-exempt interest and taxable and non-taxable Social Security benefits in addition to wages, investment income, net business income, capital gains and other types of income subject to tax (See Auten and Gee,

[1] Since these files are largely unedited outside of automatic processing checks, we used the available data and applicable tax code to further clean the data.

2009). We track primary and secondary taxpayers separately and follow them even when they change marital status or household and adjust incomes for family size when ranking households by the commonly used approach of dividing income by the square root of family size.[2]

The use of tax return data has both advantages and disadvantages measuring income inequality and dynamics. Tax return data are best for measuring incomes of middle and upper income individuals and households. They are weakest for the low income population because of the filing threshold and because incomes from the informal sector are generally unreported. A small number of taxpayers with negative incomes from business losses are actually quite wealthy, and clearly different than others in the lowest income quintile. Some taxpayers with apparently high incomes have relatively low net economic incomes because nearly offsetting expenses for employee business expenses and gambling losses can only be deducted as itemized deductions. Further, since 1987, generally 11 or 12 million tax returns and as much as 11.5 percent of all returns, have been filed by dependents and individuals under age 20. Failing to adjust for these dependent returns can distort income distribution computations based on published tax data. Nevertheless, tax return information is valuable in examining income mobility and inequality issues since other data sets suffer from comparable issues. Use of the population-level CDW data allows us to obtain partial income estimates for taxpayers and dependents from earlier years even when they are non-filers in later years.

Long Term Within-Cohort Mobility

The large SOI cross-section samples allow us to focus on narrow age ranges and compare similar reference groups, thereby abstracting from mobility that arises from life cycle effects. Table 1 shows relative income mobility over the 20-year period from 1987 to 2007 of approximately 30,000 nondependent primary and secondary taxpayers ages 35 through 40 who filed a 1987 tax return and represent approximately 20.5 million individuals. The 1987 quintiles on the vertical axis are based on nondependent taxpayers ages 35 through 40 in the 1987 SOI sample. Similarly, 2007 quintiles are calculated using the 2007 SOI sample of nondependent taxpayers ages 55 through 60. Taxpayers in these age ranges typically have relatively high filing rates. Taxpayers with negative cash income are segregated from first quintile taxpayers into a separate cell such that the sum of the "negative quintile" and the first quintile represent 20-percent of the universe. Each cell represents the percentage of 1987 quintile for a particular outcome. For instance, roughly 26 percent of the 35-to-40 year-olds in the second quintile in 1987 were in the second quintile of 55-to-60 year-old taxpayers 20 years later.

Considering only those whose information was found in both years, approximately half of taxpayers in the first and fifth quintile remained in the same quintile 20 years later. About one-fourth of those in the bottom moved up one quintile, while 4.6 percent moved to the top quintile. The overall set of estimates suggests that while there is some persistence among observed

[2] This family size adjustment is used by CBO(2011) and for official Treasury Department distribution tables and produces results similar to other commonly used adjustments. It especially improves income measurement in the cases of married individuals who later file separately and single individuals who are married in the ending year. Since primary and secondary taxpayers are followed separately, they are counted separately in determining the income quintiles of the taxpayer population. Thus, a married couple filing jointly is counted as two observations. CBO (2011) counts all members of a household (including children) in determining population quintiles.

taxpayers there is also meaningful movement even within this narrow cohort. Movements to and from the Top 10-, 5-, and 1-percent are a logical extension of the fifth quintile. While those in the top groups have much higher probabilities of being there in the later year, there are taxpayers that start from the bottom and move to the top and vice versa.

Analyzing movement across income quintiles is complicated because the likelihood of appearing in Q1 through Q5 in 2007 is correlated with income in 1987. Total attrition was 16 percent in Q1, 4.4 percent for Q5 and 3.3 percent for those in the top 1 percent in 1987.[3] Taxpayers in lower income groups in 1987 are more likely to have died and less likely to be tax filers in 2007. In our sample, taxpayers in the first quintile were more than twice as likely to die relative to taxpayers in the fifth quintile (9 percent versus 3.5 percent). The 7 percent of SSNs appearing in 1987 but having no tax record in 2007 might be assumed to have had little or no income that year, but there are other possible explanations. Such individuals could have died without the death being recorded by SSA, have filed under an incorrect SSN in 1987, have left the country, or be a nonfiler with no income reported to the IRS.[4] The use of tax return information means that we are unable to observe the person's household situation or participation in the informal economy. Among this 7 percent, CDW data back to 1997 show that 1 percent appeared on at least one more recent tax return, an additional 5 percent had some type of IRS administrative record , and only 1 percent had no tax record in any year. It may be possible to learn more about the situation of these individuals from additional records in 2007 or other tax years.

[3] This attrition rate compares favorably to the 25 percent attrition for some studies using PSID data. While low-income individuals are not required to file, the filing thresholds are generally lower than Census poverty standards and additional low-income individuals have an incentive to file to claim tax refunds and refundable credits such as the Earned Income Tax Credit.

[4] Only a small number of the "not found" taxpayers were recorded as dying within the subsequent two years or had a high number taxpayer identification number given to certain resident aliens that may have been temporary residents. It is well-known that there were many SSNs problems in the 1980s. For example, the number of dependents claimed fell by more than 5 million in 1987 when taxpayers were first required to report SSNs for claimed dependents. Secondary SSNs were not checked as carefully until the 1990s. Some fraudulent returns were known to have been filed in order to claim tax refunds and some primary or secondary SSNs appeared in more than one household and this was more difficult to detect in the 1980s. On the other hand, the fact of having a valid age suggests that an actual person was being reported in 1987.

TABLE 1: STATUS IN 2007 OF TAXPAYERS AGE 35-38 IN 1987

1987 INCOME CLASS	2007										
	Missing		Income Quintiles						Top Income Classes		
	Died	No tax record	Negative	Lowest	Second	Middle	Fourth	Highest	Top 10 percent	Top 5 percent	Top 1 percent
Negative	4.13	8.96	3.53	28.11	23.57	4.38	8.99	18.34	9.17	5.54	3.06
Lowest	8.99	7.07	0.99	42.45	19.45	12.09	5.16	3.79	1.93	1.00	0.23
Second	6.71	2.71	0.57	21.87	25.85	21.18	14.04	7.07	2.39	1.11	0.11
Middle	5.14	1.75	0.42	13.53	18.47	24.18	23.46	13.06	4.06	1.66	0.24
Fourth	4.78	0.98	0.41	9.73	14.14	19.51	27.45	22.99	7.64	3.05	0.42
Highest	3.46	0.91	0.87	6.16	7.14	12.88	22.27	46.30	27.67	15.98	3.54
Total	5.78	2.71	0.68	18.70	17.04	17.88	18.48	18.73	8.78	4.59	0.93
Top 10%	3.31	0.61	1.14	5.65	5.72	9.34	17.72	56.51	38.39	24.41	6.05
Top 5%	3.25	0.62	1.20	5.43	4.28	7.33	11.16	66.74	50.50	36.93	9.89
Top 1%	2.71	0.59	1.41	4.68	3.34	3.72	6.94	76.60	67.75	58.15	23.31

Note: 1987 centiles are based on 1987 taxpayers age 25 to 65. 2007 centiles are based on non-dependent taxpayers age 55 to 60. Returns with negative incomes and top centile groups are not included in the income quintiles. No tax record indicates no tax return and no W-2, 1099-SSA or 1099-INT information return was found. Rows sum to 100 percent.

Intergenerational Mobility

A common question regarding income mobility is whether children have greater economic success than their parents. We shed some light on the issue in Table 2 where we compare the income positions of the families of 11.2 million dependents (represented by 19,500 sample observations) ages 15-18 on 1987 tax returns to the income positions of those dependents relative to their peers 20 years later in 2007. The 1987 quintiles on the vertical axis are calculated with nondependent 25 through 65 year-old taxpayers on the 1987 SOI file. Dependents on returns are placed into those quintiles based on the cash income of the parent/guardian. The 2007 quintiles are calculated with nondependent taxpayers ages 35 through 38 found on 2007 CDW tax returns. Since the cash income profile is still somewhat steep at these ages, we calculate quintiles for 35-, 36-, 37-, and 38-year-olds, place the 1987 dependents into the appropriate age-based quintile, and aggregate those results to control for both a year and age-specific effect.[5]

Relative to the 1987 taxpayer mobility discussed above, we observe less persistence and more upwards and downwards mobility from the first and fifth quintiles, respectively. Out of dependents in the first quintile, roughly 27- and 10-percent are in the bottom and top quintiles relative to their peers in 2007. Similarly, out of dependents in the fifth quintile, roughly 39- and 8-percent are in the top and bottom quintiles relative to their peers. Mirroring the patterns in

[5] While this approach controls for the rising income-age profile, it misses any mobility of those who, for example, make an upward income movement one year later than their contemporaries, such as might result from being held back a year before starting kindergarten or being slow to finish a PhD program.

Table 1, dependents from the bottom quintiles are more likely to either be confirmed dead by SSA or not found on a tax return nor administrative record. Not surprisingly, mortality is far less of an issue with the 15- to 18-year-old group in Table 2 compared to the sample in Table 1. Failing to be located remains a consistent problem across the two samples, with a differential likelihood of having a tax record in 2007 that is correlated with income group in 1987.

TABLE 2: STATUS IN 2007 OF DEPENDENTS AGE 15-18 IN 1987

	2007										
1987	Missing		Income Quintiles						Top Income Classes		
INCOME CLASS	Died	Not Found	Negative	Lowest	Second	Middle	Fourth	Highest	Top 10 percent	Top 5 percent	Top 1 percent
Negative	0.10	3.00	0.54	10.36	15.89	26.01	21.14	22.96	16.39	7.33	1.65
Lowest	2.27	8.07	0.53	26.57	17.35	18.45	17.08	9.67	4.22	1.67	0.36
Second	1.48	4.90	0.68	16.77	19.13	20.97	21.61	14.46	5.92	2.28	0.44
Middle	1.37	3.09	0.38	13.54	15.11	21.32	24.22	20.97	10.19	5.31	0.78
Fourth	1.00	2.12	0.79	10.41	13.30	19.30	24.75	28.32	13.92	5.77	1.29
Highest	1.17	1.90	0.55	8.17	9.08	15.72	24.10	39.31	25.19	15.26	4.23
Top 10%	1.17	1.59	0.76	8.64	7.55	12.33	21.77	46.19	30.16	18.89	5.53
Top 5%	1.05	1.50	0.83	7.50	6.84	11.68	20.07	50.54	34.28	22.93	7.65
Top 1%	1.09	1.87	1.38	6.37	7.57	10.64	15.53	55.54	43.22	32.49	13.65

Note: 1987 dependents are classified using the income from the return on which they were claimed as a dependent. 1987 centiles are based on 1987 taxpayers age 25 to 65. 2007 centiles are based on age 35 to 38 non-dependent taxpayers at each age. Returns with negative incomes are not included in the lowest income quintile. Not found indicates no tax return, W-2, 1099-SSA or 1099-INT was found. Rows sum to 100 percent.

Persistence at the Top

In recent years there has been renewed interest in the top of the income distribution from both policymakers and the general public. Despite this attention, there is relatively little known about the persistence of individuals at the very top of the income distribution in the U.S. Do most high income individuals receive large amounts of income every year, or do they have a few years of very high income that places them at the top of the distribution for a short period? Our earlier paper (Auten and Gee, 2009) found that nearly 60 percent of those in the top 1 percent in the beginning year of each period had dropped to a lower centile by the 10th year. Less than one-fourth of the individuals in the top 1/100th percent in 1996 remained in that group in 2005.[6]

Table 3 sheds additional light on the persistence of high-income tax returns in the top one percent of tax returns over 5-year periods. The top percentile of cash income is defined in each year by tax returns with a primary filer who is age 25 or older. The columns show the persistence in the top 1 percent for up to five consecutive years. For example, column T+2 for the 1991 row

[6] Data generally show that the higher the income, the greater the turnover, reflecting regression to the mean and one-time realizations of large amounts of income. For example, IRS Statistics of Income tables on the top 400 taxpayers over the 1992 to 2008 period show that 3,672 individuals appeared as either a primary or secondary taxpayer in one or more years. 2,676 individuals had appeared only once during this period (but may have appeared before or after it), while only 4 appeared in all 17 years.

shows that 39 percent of taxpayers who were in the top one percent in 1991 were also in the top one percent in 1992 and 1993. To abstract from retirement decisions, the sample used for Table 3 includes individuals who are 25-60 in the base year (note that the figures may include some effects from early retirement.)

One year persistence (T+1) rates range from 68 to 70 percent during the years 1992-1999 and are roughly 66 percent between 2002 and 2006. The lower one year persistence in 1991, 1999 through 2001 and 2007 is likely the result of macro economic factors. During periods of recession, financial income including capital gains and dividends, as well as business income contract and alter the composition of top earners.[7] Five year persistence rates suggest that as much as one-third of all taxpayers in the top one percent in a given year remain there for five consecutive years during this period. Given that the sample focuses on an age group that is likely to be still working, this implies that many high income earners remain high earning in later years.

TABLE 3: PERSISTENCE IN THE TOP 1 PERCENT OF TAXPAYERS

	T+1	T+2	T+3	T+4	T+5
1991	0.61	0.46	0.39	0.33	0.29
1992	0.68	0.54	0.45	0.39	0.34
1993	0.69	0.56	0.47	0.41	0.35
1994	0.70	0.56	0.47	0.40	0.36
1995	0.68	0.54	0.44	0.39	0.33
1996	0.68	0.53	0.46	0.38	0.32
1997	0.66	0.55	0.43	0.36	0.30
1998	0.70	0.51	0.40	0.32	0.28
1999	0.52	0.37	0.29	0.24	0.21
2000	0.60	0.42	0.34	0.29	0.26
2001	0.62	0.47	0.39	0.33	0.30
2002	0.66	0.51	0.42	0.37	0.32
2003	0.66	0.51	0.43	0.37	0.31
2004	0.65	0.51	0.43	0.34	0.28
2005	0.65	0.51	0.39	0.31	0.27
2006	0.65	0.45	0.34	0.30	Na
2007	0.59	0.41	0.34	na	Na
2008	0.62	0.49	na	na	Na
2009	0.68	na	na	na	Na

Notes: Sample includes primary and secondary taxpayers age 25 to 60 in the base year who appear in each year between the base year and year T+X. For example, in base year 1991 row, the share in the T+3 column includes taxpayers in the top 1% in 1991 (between ages 25 and 60 in 1991) who also appear in the top 1 percent in 1992, 1993 and 1994. na indicates periods out of the sample window.

[7] In related work we find that the composition of income among the top one percent is related to persistence in way that is consistent with this interpretation.

Taxpayers by Birth Cohort and Age

Another way to look at the changing composition of the top 1 percent is to track taxpayers in cross section files according to their birth cohorts. Figure 1 shows the percentages of the top 1 percent born prior to 1946, from 1946 to 1955 (the early boomers), 1956 to 1965 (later boomers), 1966 to 1976 (Gen X), and from 1977 to 1994 (Gen Y). Over the period 1987 to 2010, the share of pre-1946 group declined steadily from 77 percent to about 20 percent. By 1987, the early boomers, then age 32 to 41, already accounted for 19 percent of the top 1 percent. By 2003 (ages 48 to 57) this group reached its peak share of 33 percent before declining to 29 percent in 2010. The later boomers reached a 19 percent share at ages 34 to 43 in 1999. At ages 34 to 43, Gen X accounted for 18 percent of those in the top 1 percent. The differences between these "generations" are small, the later ones being perhaps a year or two behind the early boomers. Gen Y was only beginning to be seen in the top 1 percent with 3 percent in 2010. Looked at slightly differently, over this period the shares of the top 1 percent by age group averaged 5 percent for those aged 25 to 34, 19 percent for those age 35 to 44, 28 percent for those age 45 to 54, 25 percent for those age 55 to 64 and 23 percent for those age 65 and over, with the shares varying modestly over time with the size of the cohort and other factors. While these cohort and age shares do not track the movement of specific individuals in and out of the top 1 percent, they show another dimension of the turnover in the top 1 percent.

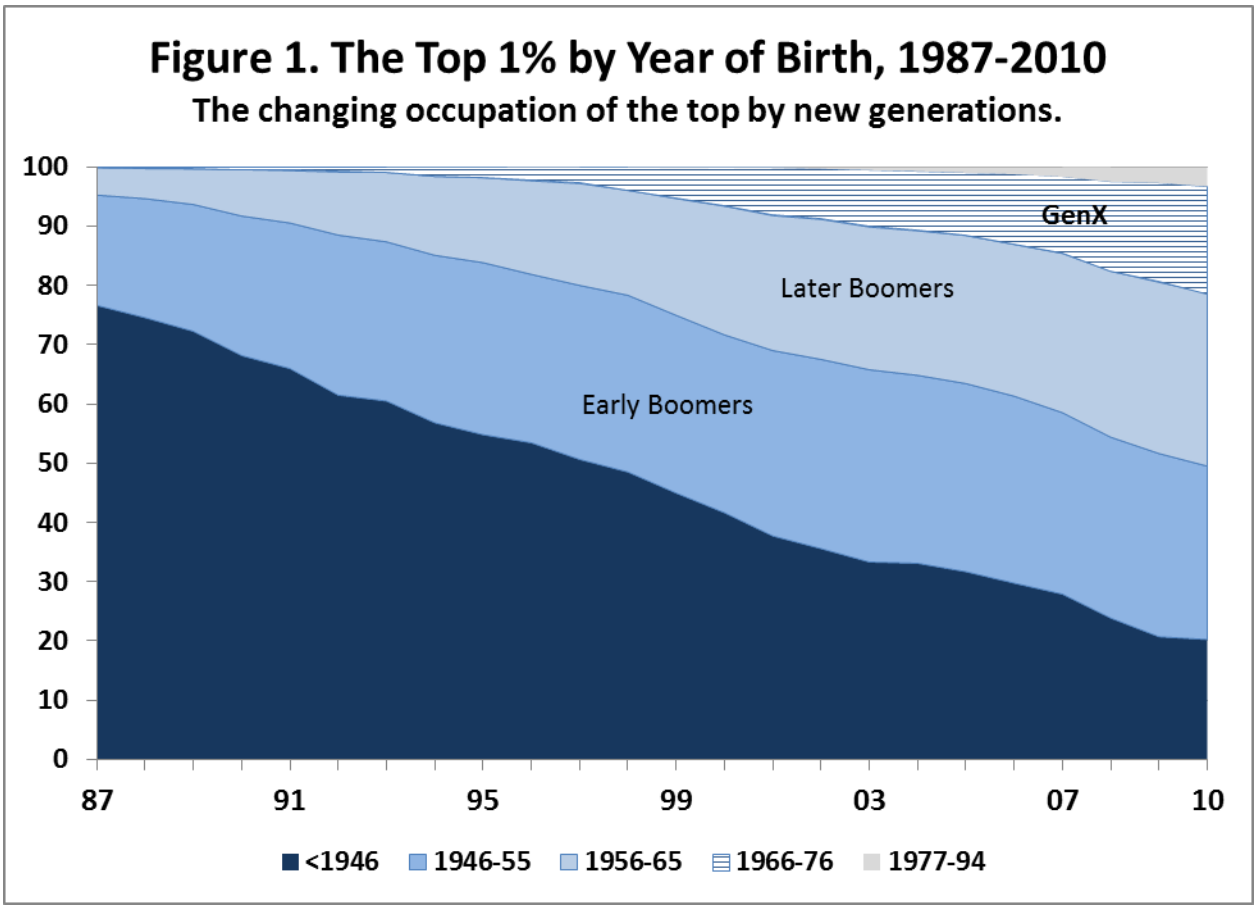
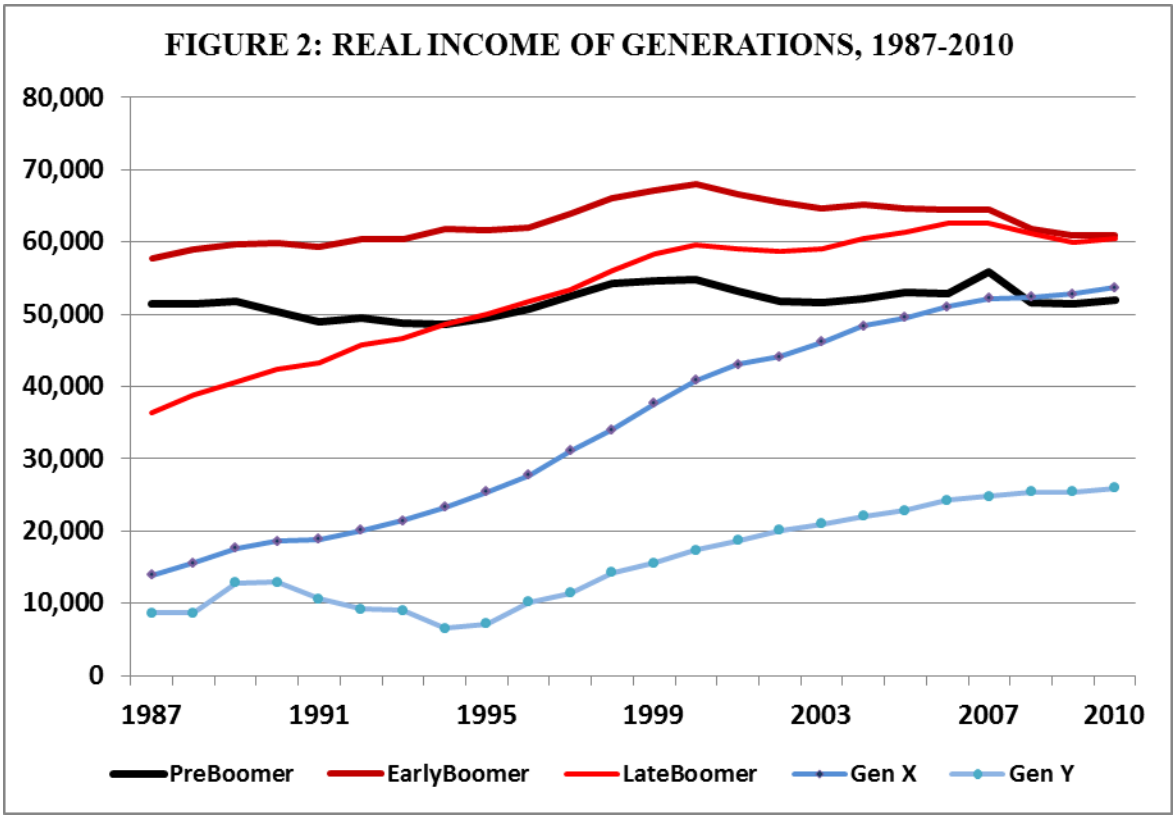
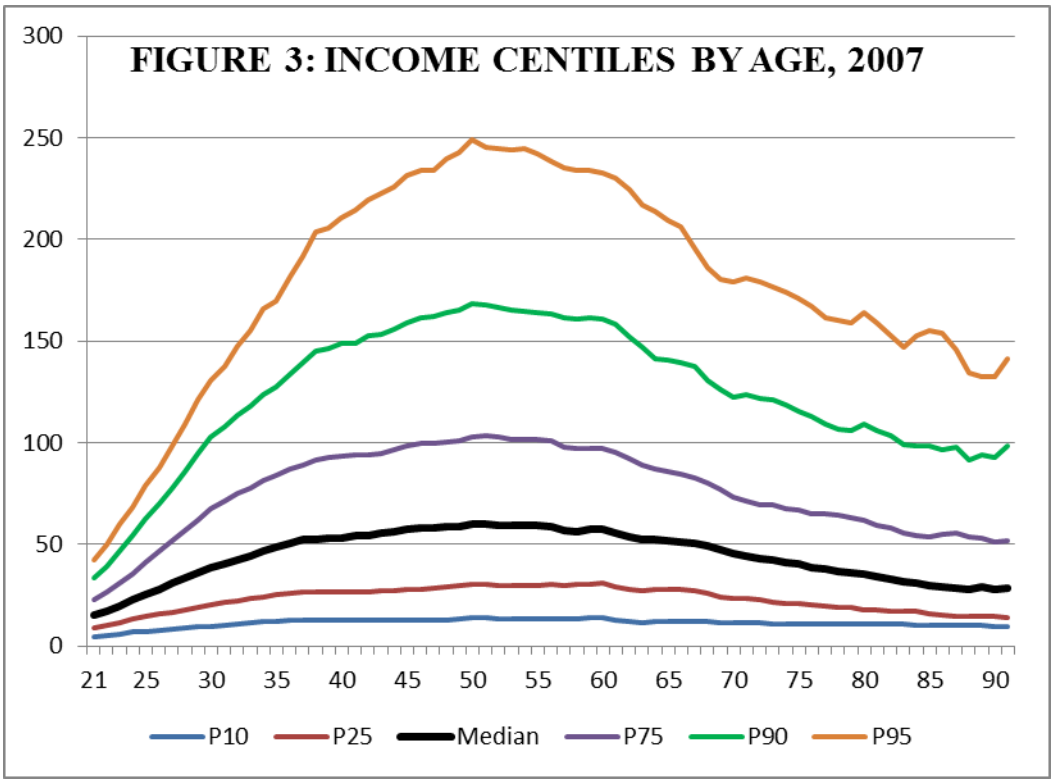


Figure 2 provides a different perspective on the generations by tracking the median real income of each group from 1987 through 2010. Median real income of the early boomers increased steadily until 2000 and then declined through 2010. Income of the later boomers increased through 2006 before declining in the recession period. In contrast, the median real income of GenX continued to increase throughout the recession period. In general, the story is that the income of each generation eventually catches up and surpasses that of the prior ones.



Our work in this area highlights the importance of the life cycle in income mobility and in what one observes about inequality at points in time. The steep income profiles for individuals in their 30s is one reason why we focused on a relatively narrow age group initially between ages 35 and 40. The following figure shows key centile breakpoints for the incomes of taxpayers age 21 through 90 (with those above age 90 lumped at age 91) in 2007. The cutoff for the top 5 percent at each age rises dramatically from about $45,000 at age 21 to $250,000 at age 50 before declining more gradually to under $150,000 for those in their late 80s. The cutoffs for the top 10 and 25 percent peak in the early 50s before declining more gradually. The age-income profiles for median and the 10th and 25th percentiles are lower and somewhat flatter. While this cross-section includes only those who have filed tax returns, a much large share of the population is represented this year because the number of non-dependent tax returns filed increased by over 14 million (11.5 percent) in 2007 because of tax rebates before declining by 10 million returns to a more normal level in 2008. The age-income profile also illustrates the gradual movement of individuals through the income distribution over time.



Conclusions

This paper extends the literature on income mobility and inequality by providing new evidence on additional dimensions of the dynamics of income over time: long-term (20-year) income changes over the primary working ages, intergenerational mobility of dependents, short-term persistence of taxpayers in the top 1 percent and the movement of successive generations through the top 1 percent since 1987. The use of CDW data allows us to greatly reduce the attrition inherent in longitudinal studies and obtain at least partial information on individuals who do not file tax returns. We leave it to the reader to look at the results and decide whether they think the observed mobility and turnover at the top are sufficient or good.

REFERENCES

Auten, Gerald, and Robert Carroll. 1999. "The Effect of Income Taxes on Household Income," *Review of Economics and Statistics*. 81(4): 681-93.

Auten, Gerald, and Geoffrey Gee. 2009. "Income Mobility in the United States: New Evidence from Income Tax Data." *National Tax Journal*, 57(2): 301-328.

Congressional Budget Office. 2011. Trends in the Distribution of Household Income Between 1979 and 2007. Washington, DC.

Piketty, Thomas and Emmanual Saez. 2003. "Income Inequality in the United States, 1913-1998." *Quarterly Journal of Economics*, 118(1): 1-39. (Updated through 2010 at http://elsa.berkeley.edu/~saez/)

U.S. Internal Revenue Service, Statistics of Income Division. (n.d.) "The 400 Individual Income Tax Returns Reporting the Highest Adjusted Gross Incomes Each Year, 1992-2008." Washington, DC. http://www.irs.gov/taxstats.

APPENDIX TABLE 1: INDIVIDUAL INCOME TAX RETURNS BY DEPENDENT STATUS AND AGE, 2007-2010

	Returns in 1000s				Percent of All Returns		
Year	All returns	Dependent plus under age 20	Non-dependent under age 20	Dependent	Dependent plus under age 20	Non-dependent under age 20	Dependent
1985	101,660	7,739	6,955	784	7.6	6.8	0.8
1986	103,045	7,733	7,058	675	7.5	6.8	0.7
1987	106,996	12,150	2,809	9,341	11.4	2.6	8.7
1988	109,708	12,601	2,613	9,988	11.5	2.4	9.1
1989	112,136	12,960	2,579	10,381	11.6	2.3	9.3
1990	113,717	12,572	2,226	10,347	11.1	2.0	9.1
1991	114,730	11,963	2,187	9,776	10.4	1.9	8.5
1992	113,605	11,361	2,060	9,301	10.0	1.8	8.2
1993	114,602	11,169	1,895	9,274	9.7	1.7	8.1
1994	115,943	11,490	1,925	9,564	9.9	1.7	8.2
1995	118,218	12,046	1,884	10,162	10.2	1.6	8.6
1996	120,351	12,362	1,711	10,651	10.3	1.4	8.8
1997	122,422	13,072	1,793	11,280	10.7	1.5	9.2
1998	124,771	12,899	1,805	11,094	10.3	1.4	8.9
1999	127,075	13,556	2,030	11,527	10.7	1.6	9.1
2000	129,374	13,678	1,944	11,735	10.6	1.5	9.1
2001	130,255	12,755	1,871	10,884	9.8	1.4	8.4
2002	130,076	11,917	1,691	10,226	9.2	1.3	7.9
2003	130,424	11,374	1,709	9,665	8.7	1.3	7.4
2004	132,226	11,353	1,652	9,701	8.6	1.2	7.3
2005	134,373	11,541	1,844	9,698	8.6	1.4	7.2
2006	138,395	11,764	1,823	9,941	8.5	1.3	7.2
2007	153,560	12,310	1,924	10,386	8.0	1.3	6.8
2008	142,451	11,456	1,924	9,532	8.0	1.4	6.7
2009	140,494	9,476	1,629	7,847	6.7	1.2	5.6
2010	142,892	9,341	1,541	7,800	6.5	1.1	5.5

Notes: Data are IRS Statistics of Income Individual Income Tax Returns. Each year's data includes some late filed tax returns from prior years. 2007 data include returns apparently filed only to receive a tax rebate and not included in published IRS statistics. The definition of a dependent filer changed in 1987.