Recent Trends in Top Income Shares in the USA:
Reconciling Estimates from March CPS and IRS Tax Return Data

Richard V. Burkhauser
Cornell University

Shuaizhang Feng
Princeton University and Shanghai University of Finance and Economics

Stephen P. Jenkins
University of Essex

Jeff Larrimore
Cornell University

August 27, 2009

Abstract

Although the vast majority of US research on trends in the inequality of family income is based on public-use March Current Population Survey (CPS) data, a new wave of research based on Internal Revenue Service (IRS) tax return data reports substantially higher levels of inequality and faster growing trends. We show that these apparently inconsistent estimates can largely be reconciled once one uses internal CPS data (which better captures the top of the income distribution than public-use CPS data) and defines the income distribution in the same way. Using internal CPS data for 1967–2006, we closely match the IRS data-based estimates of top income shares reported by Piketty and Saez (2003), with the exception of the share of the top 1 percent of the distribution during 1993–2000. Our results imply that, if inequality has increased substantially since 1993, the increase is confined to income changes for those in the top 1 percent of the distribution.

Introduction

The March Current Population Survey (CPS) public-use files have been the primary data source used to study income inequality trends in the USA.[1] The consensus finding of research based on these data is that household income inequality increased substantially in the 1970s and 1980s, and continued to increase but at a much slower pace starting in the 1990s (Gottschalk and Danziger 2005, Daly and Valetta 2006, and Burkhauser, Feng and Jenkins 2009).

The most notable alternative source for studying income inequality trends derives from tax return data. In their seminal paper, Piketty and Saez (2003) use data from Internal Revenue Service (IRS) Statistics of Income tax returns to analyze income inequality trends in the USA. Their paper was one of the first in a rapidly expanding literature that has used tax return data to examine income inequality trends around the world. See Piketty (2003) for France, Atkinson (2005) for the UK, Saez and Vaell (2005) for Canada, Bach, Corneo and Steiner (2009) for Germany, Dell (2005) for Germany and Switzerland, and Atkinson and Leigh (2007) for Australia. Atkinson and Piketty (2007) and Leigh (2009) provide comprehensive reviews of this literature.

One of Piketty and Saez's major contributions derives from being able to observe income inequality trends over a much longer period than previous researchers: tax return data are available for years well before any survey data on income was collected. However, their findings have also sparked debate about inequality trends over relatively short periods, and recent years in particular. For a flavor of the debate on this topic, see the blog postings by leading economists and others on the Economists View website (2007). Reynolds (2007) provides an illustration of

---

[1] See Atkinson, Rainwater, and Smeeding (1995), Atkinson and Brandolini (2001) and Gottschalk and Smeeding (1997) for reviews of the income distribution literature. For more recent examples of the use of the public-use CPS in measuring inequality trends in the USA, see Gottschalk and Danziger (2005) and Burkhauser, Feng and Jenkins (2009).

how the work by Piketty and Saez has altered the popular view of recent trends in income inequality and a critique of their results.

In contrast to research based on CPS data that finds income inequality slowing in the 1990s, Piketty and Saez (2003, 2008) find that the share of total income held by the very richest groups grew during the 1990s and, with the exception of the period from 2000–2002, continued to rise rapidly through the beginning of the 21st century as well. What explains the differences in inequality trends found by researchers using these two types of data?

One explanation is that there are deficiencies in one or both of these data sets that limit researchers' abilities to observe the true trends in inequality. Critics of those using the public-use CPS to measure income inequality argue that topcoding and underreporting of top incomes restricts the survey's ability to observe income changes for those at the top of the distribution. See inter alia Levy and Murnane (1992), Slemrod (1996), Burkhauser, Couch, Houtenville and Rovba (2003-2004), Piketty and Saez (2006*b*), and Burkhauser, Feng, and Jenkins (2009). Thus, to the extent that income inequality changes are due to changes in the topcoded portion of the CPS, researchers using this data may mismeasure trends in income inequality.

Using IRS data to measure income inequality also has potential limitations, however. Critics point out that tax filers have a financial incentive to report their income in ways that limit their tax liabilities and, as a result, filing behavior is sensitive to changes in the personal income tax rate. There are several fiscal manipulation strategies that are sensitive to changes in marginal tax rates and income reporting rules. These include reclassifying income as either wage earnings or business profits depending on which is taxed less (Sivadasan and Slemrod 2006), receiving untaxed fringe benefits in lieu of wage compensation (Woodbury and Hammermesh 1992), or deferring compensation through stock options or deferred compensation packages (Scholes and

Wolfson 1992, Goolsbee 2000). Since high income earners are the individuals most able to adjust the way that they receive and report income, tax return data may especially not be able to capture income at the top of the distribution accurately. For example, Slemrod (1995) and Reynolds (2006) suggest that tax-law changes since the 1970s have provided incentives for the very rich to switch their reported income from Subchapter-C corporation profits, which are not reported on personal income tax forms, to S-corporation profits and personal wage income, which are reported. They suggest that this, in turn, has led researchers using tax return data to overstate the actual rise in income among the very rich. See Feenberg and Poterba (1993) for an earlier discussion of this problem and a summary of the difficulties measuring top incomes with tax records data.

Piketty and Saez (2003) acknowledge that this type of fiscal manipulation may impact measures of top income shares, but argue that such effects are only problematic for short-term trends rather than the long run trends in income inequality which are their primary concern. However, for researchers interested in the relatively short-term trends in income inequality of recent years, time-shifting of income may still pose a problem depending on the time-frame of the deferred compensation plan. Additionally, while time-shifting of income may only impact income inequality in the short-term, income that is received in ways other than through labor earnings – such as through higher non-taxable fringe benefits or the reporting of what had been wage earnings in previous years as business profits – will never be reported on personal income tax forms and thus could have implications on long-term income inequality trends. Thus, to the extent that changes in reporting rules alter the way income is reported at the top of the distribution, researchers using IRS tax return data may mismeasure actual changes in income inequality.

Yet another potential explanation for the differences in estimated inequality trends is that they result from differences in the definition of income and how its distribution is summarized rather than differences in the data sources themselves. Although all the researchers using public-use CPS data and IRS tax data examine "inequality" in the broad sense, there are substantial differences in their definitions of "income" (the sources included – most especially the inclusion of government transfers and non-taxable income in the former and its exclusion in the latter – and whether there is adjustment for differences in "needs"), the income recipient unit (tax units versus households and individuals within them), and how best to measure inequality (in terms of top income shares versus a more comprehensive measure such as the Gini coefficient).

To some extent, these differences in practice have evolved because of the nature of the data examined. For example, researchers who use public-use CPS data, which has a high prevalence of topcoded values at the top of the income distribution, often measure inequality using the ratio of the 90th percentile to the 10th percentile ("p90/p10") to mitigate problems arising from topcoding. (See, Burkhauser, Feng, and Jenkins 2009 for a discussion of the limitations of this measure.) Researchers using tax return data focus on top income shares since many low income individuals do not file a tax return and so it is not possible to directly derive measures of income inequality that take account of the income shares of poorer groups (Piketty and Saez 2006*a*).

To date, no researchers have attempted to bridge the gap between the CPS- and IRS-based literatures to determine the extent to which the differences in inequality estimates emanating from these two literatures arise from differences in the ability of these two data sources to capture top incomes or from the application of different income constructs based on these data sources. In this paper, we do just that.

Using internal CPS data, we examine the trends in income inequality since 1967 using the inequality measures and income distribution definitions developed by Piketty and Saez (2003) and others using tax return data. Doing so, we are able to closely match their results. Our estimates of top income shares are nearly identical for groups in the richest tenth with the exception of the richest 1 percent, and our estimates of trends differ only slightly. Even for estimates of the share held by the top 1 percent, the two data sources are broadly in agreement about trends over much of the past 40 years. It is only during a six year period in the late 1990s that the trends diverge for reasons that are not easily explained by changes in the nature of the two data sources.

In the next section we describe the internal CPS data that we use and our methods for overcoming topcoding problems. We then compare the definitions of "income" and income-receiving unit used by Piketty and Saez with those used by researchers using CPS data. Using internal CPS data we derive two series of estimates of top income shares, each one corresponding to the two research traditions. We go on to show that the CPS-based series of estimates based on Piketty and Saez-type definitions corresponds closely with Piketty and Saez's (2003) IRS-based series of estimates, though with some exceptions that we discuss. We also consider the reasons for the divergences between the series.

Data

Our analysis derives from access to internal CPS data which are identical to the data used by Census Bureau researchers in their official work (see e.g. U.S. Census Bureau, various years). These data measure top incomes much better than the data released in public-use CPS files. To protect the confidentiality of its respondents, the Census Bureau censors ("top codes") each of

the income sources received by individuals. This practice must be addressed in order to derive

sensible estimates of top income shares using CPS data. The advantage of internal data over

public-use data is that the prevalence of topcoding is very much lower.[2] For example, in 2004,

0.5 percent of individuals lived in a household in which some source of income was topcoded in

the internal data compared to 4.6 percent in the public-use data.

Even the small extent of censoring in the internal CPS data produces biased estimates of

top income shares. To address this issue, we use a multiple imputation approach in which values

for censored observations in the internal data are multiply imputed using draws from a

parametric model of the income distribution fitted to the internal data.

Our multiple imputation approach is the same as that used by Burkhauser et al. (2008)

and described in detail by Jenkins et al. (2009). The approach involves five steps. First, for each

year's data, we fit a Generalized Beta of the Second Kind (GB2) distribution by maximum

likelihood, accounting for individual-level right-censoring.[3] To ensure that model fit is

maximized at the top of the distribution, the GB2 is fitted using observations in the richest 70

percent of the distribution only (with appropriate corrections for left truncation in the ML

procedure). Second, for each observation with a censored income, we draw a value from the

income distribution that is implied by the fitted GB2 distribution, using an appropriate stochastic

procedure. Third, using the distribution comprising imputations for censored observations and

---

[2] For many measures of income inequality such as the Gini coefficient and General Entropy inequality measures researchers can closely replicate the results from the internal CPS data by using cell-means of topcoded incomes that are provided back to 1975 in Larrimore, et al. (2008). However, by design cell-means assume that all topcoded individuals have the same income. As a result, using cell-means to approximate top income shares with the public-use CPS data will lead to an overestimation of the income held by the 90th to 99th percentile groups and an underestimation of the income held by the top 1 percent of the distribution.

[3] The GB2 model is widely used in the income distribution literature, and shown to fit income distributions extremely well across different periods and countries: see e.g. Bordley, McDonald and Mantrala (1996), Brachmann, Stich and Trede (1996), Bandourian, McDonald, and Turley (2003), and Jenkins (2009). Since the GB2 is a four-parameter distribution, its shape is more much flexible than that of the commonly-used Pareto distribution, and hence fits the data better.

observed incomes for non-censored observations, we estimate our various inequality indices. Fourth, we repeat steps 2 and 3 one hundred times, and finally, we derive inequality estimates by combining the one hundred sets of estimates from each of the one hundred data sets for each year using the 'averaging' rules proposed by Rubin (1987) and modified by Reiter (2003) to account for imputation variability.

Internal CPS data augmented with multiply imputed values for censored incomes provide the best available estimates of the income distribution using the CPS. These distributions are the source for all the CPS-based estimates of top income shares reported below that we compare with the top income share estimates of Piketty and Saez (2003).[4] We have also undertaken all our calculations of top income shares using CPS internal data used "as is", without imputations for censored values. All the conclusions we draw later regarding income shares for income groups outside of the top 1 percent are unchanged. For the top 1 percent, using the unaltered internal data rather than multiply imputed internal data reduces estimates of income shares, but conclusions about trends are similar. See Appendix A for further details.

Methods: Three Definitions of the Income Distribution

There are three substantial methodological differences between research based on the CPS and research based on the IRS tax return data. The first is the inequality measures used. Most CPS research uses either inequality indices such as the Gini or Theil coefficients that use data on all incomes, or indices like p90/p10 that ignore incomes at the very top of the income distribution. In contrast, tax data researchers focus on the top of the income distribution, defining

---

[4] Imputation of censored incomes has been used previously in inequality research based on public-use March CPS data: see e.g. Fichtenbaum and Shahidi (1988) and Bishop, Chiou, and Formby (1994) who apply single imputation methods using fitted Pareto distributions. Imputation is also nearly universally used in the inequality literature based on tax return data since income is only reported in income bands that do not necessarily coincide with percentile cut-points. See e.g. Piketty and Saez (2003), Piketty (2003), Dell (2005), or Saez and Vaell (2005).

inequality in terms of top income shares – the share of total income held by the richest 10 percent, the richest 5 percent, or the richest 1 percent, and so on – with larger income shares indicating greater inequality.

The other two differences in method concern the definition of the income distribution, specifically: what is counted as "income" and what is the income-receiving unit. CPS-based researchers have typically defined income as pre-tax post-transfer income excluding capital gains: see e.g. Gottschalk and Danziger (2005) and Burkhauser et al. (2008).[5] This income is aggregated to the household level, and deflated using an equivalence scale to account for differences in economies of scale and "needs" (the square root of household size is a commonly-used scale). Attributing the same size-adjusted household income to each individual within the same household, researchers examine the distribution of income among individuals.

Piketty and Saez (2003) and other researchers using tax data use different definitions. Piketty and Saez define income to include any income reported on IRS tax returns before deductions and excluding capital gains. This encompasses "salaries and wages, small business and farm income, partnerships and fiduciary income, dividends, interest, rents, royalties, and other small income reported as income" (Piketty and Saez 2003, pp. 5–6). The most notable difference between this income definition and the CPS one is that it excludes most transfer income, which is generally not taxable and not included in the adjusted gross income reported on tax returns. Hence it is close to the individual's market income, which is also known as pre-tax pre-transfer income in the broader income inequality literature.[6] See Scholz and Levine (2002),

---

[5] In international comparisons of income inequality, it is most common to include the effect of both government transfer programs and tax policies by measuring post-tax, post-transfer income. See Atkinson and Brandolini (2001) and Gottschalk and Smeeding (1997) for reviews of this literature.

[6] In the wage inequality literature researchers tend to primarily be interested in how different types of workers – e.g. low vs. high skilled, women vs. men, etc. – are rewarded in the labor market. Hence in this literature it is common to measure pre-tax wage rates or labor earnings. Pre-tax pre-transfer market income is an extension of this concept to all factors of production. Traditionally, researchers interested in income inequality have focused on how it relates to

Corneo and Fong (2008), and Bach, Corneo, and Steiner (2009) for examples of this type of measure.

Piketty and Saez (2003) aggregate income to the level of the tax unit rather than to the level of the household, do not adjust for differences in tax unit size, and they examine the distribution among tax units rather than among individuals. An important issue in this literature is that not all individuals in the USA file a tax return, with non-filers generally having lower incomes. Therefore, estimates of the income share of the top 10 percent of tax filers understate the number of tax filers relative to the situation in which non-tax filers are included in the base. That is, when the number of "potential tax filing units" (filers plus non-filers) is the base, a higher share of actual tax filers and hence a larger share of reported pre-tax pre-transfer income must be included in order to correctly measure overall income inequality. To address this issue, Piketty and Saez (2003) estimate the total number of potential tax units and calculate the number of returns that make up the top income groups using this number. They define a potential tax unit as a married couple of any age, divorced or widowed individual of any age, or single individual over the age of 20. See the Data Appendix of Piketty and Saez (2007) for further details.

Definitions of income and the unit of analysis are important because variations in each can be expected to lead to different estimates of the degree of inequality. For example, we expect the inclusion of transfer income in income (as is done by CPS researchers) to reduce inequality because transfer income is targeted at poorer families. Additionally, low income individuals who

---

one's ability to consume and hence include government transfers in the US literature and both taxes and transfers in the international comparative literature. Thus in those literatures pre-tax pre-transfer income is rarely used by itself but rather to distinguish between incomes generated in the absence of government and a fuller measure of income which includes government taxes and transfers. In the CPS-based literature this has generally meant including cash-transfers for inequality calculations – thus using a pre-tax, post-transfer income definition. Some researchers, including the National Research Council Panel on Poverty and Family Assistance, have advocated moving even farther from the pre-tax, pre-transfer market income definition when analyzing poverty by including taxes and non-cash transfers in US income inequality calculations. For a further discussion of the effect of such proposals on poverty rates and income inequality, see Burtless and Smeeding (2002).

need to share costs and lower living expenses are more likely to live in larger households with individuals outside of their tax unit. Therefore, aggregating income to the household level rather than the tax unit, and adjusting for economies of scale using an equivalence scale, is expected to yield an inequality estimate that is lower than for the distribution of pre-tax pre-transfer income among tax units.

The two CPS series that we use are defined as follows. First, our "traditional" CPS series, labeled "CPS-Post-HH", refers to the estimates based on the distribution of size-adjusted pre-tax post-cash transfer household income among individuals. Size adjustment uses the square root of household size.

The second CPS-based series, "CPS-Pre-TU", uses Piketty-Saez-type definitions of the income distribution. That is, we consider distributions of non-size-adjusted pre-tax pre-transfer tax unit income among tax units. Since tax unit identifiers are not provided in the CPS, we follow Piketty and Saez's procedures to determine potential tax units. All single individuals over the age of 20, married couples, and divorced or widowed individuals are considered to head a tax unit. Never-married children under the age of 20 are considered dependents and are assigned to the tax unit of their parent or guardian.[7] Our measure of pre-tax pre-transfer income includes income from wages and salaries, self-employment, farm income, interest, dividends, rents, trusts, and retirement pension income – which closely matches the taxable income sources included in the IRS tax return data analyzed by Piketty and Saez. In both cases, capital gains are excluded. Although a small number of taxable transfers are excluded by this definition, the broad income

---

[7] In the small number of cases where never-married individuals under age 20 live in a household without a parent or guardian, we assigned them to the tax-unit of the primary family in the household or the oldest adult in the household when there is no primary family. Only if there are no adults over the age of 20 in the household are they considered their own tax-unit. Different procedures for classifying these individuals were tested, including removing them from the sample, and assigning them their own tax units. These procedures produced substantively similar results.

categories used by the CPS prior to 1987 make it difficult to separate these taxable transfers from non-taxable transfers consistently across the entire period. Since the vast majority of transfer income is non-taxable, our best approximation to Piketty and Saez's income definition necessarily excludes this income source.

Comparisons between the CPS-Post-HH and CPS-Pre-TU series are informative about how much of the difference in top share estimates can be attributed to differences in definitions, whereas comparisons between the CPS-Pre-TU series and the "Piketty-Saez" estimates reported by Piketty and Saez (2003, 2008) are informative about how much of the difference in estimates can be attributed to differences in the underlying data source.

In order to contrast the three series at several points in the income distribution, we examine income shares for three groups within the top 10% of the distribution each year. We consider the fortunes of those with incomes between the 90th and 95th percentiles of the distribution (the "p90–p95 group"), those with incomes between the 95th and 99th percentiles of the distribution (the "p95–p99 group"), and those in top 1 percent.

Top Income Shares: IRS- and CPS-based Series Compared

In Figures 1 through 3 we provide our estimates of top income shares for three series defined earlier. The income shares for the p90–p95 group are presented in Figure 1, the shares for the p95–p99 group are presented in Figure 2, and the shares for the top 1 percent are presented in Figure 3.

For all three groups, the estimates of income shares according to the CPS-Post-HH series are smaller than the corresponding ones from the Piketty-Saez series. This is unsurprising given the two very different income definitions used. Because a much greater share of non-taxable

government in-cash transfers – AFDC/TANF, Social Security benefits, etc. – are held by the poorest 90 percent of the pre-tax post-transfer (CPS-Post-HH definition) distribution, we would expect the income share of the top 10 percent of the pre-tax post-transfer income distribution to be smaller than the income share for the top 10 percent of the Piketty-Saez gross income distribution in all years. This is the case.

But, once we control for differences in definitions, the differences in estimates of income share held by these high income groups based on CPS and IRS data are much smaller in both level and trend. This can be seen by comparing corresponding estimates in the CPS-Pre-TU and Piketty-Saez series. For the p90–p95 group (Figure 1), the CPS-Pre-TU series and Piketty-Saez share estimates are almost identical in the beginning of the period. The increase in the CPS-Pre-TU series p90–p95 group's income share over the 40 year period is somewhat greater than the Piketty-Saez estimates: a rise from 10.9 percent to 12.5 percent, compared to a rise from 11.0 percent to 11.9 percent. But, even with the slight trend differences, the income shares in each year are always close to each other. For the p95–p99 group (Figure 2), levels and trends using the CPS-Pre-TU and Piketty-Saez series are even closer, although the CPS-Pre-TU series again shows a slightly greater upward trend than the IRS data.

In addition to comparing the income share of the p90–p95 and p95–p99 groups, we also considered the sources from which individuals in these groups received their income. However, the GB2-based multiple imputation procedure must be performed on total household income and thus cannot distinguish source-level incomes for this analysis. While this prevents us from comparing income sources for the top 1 percent of the distribution, since most individuals in the p90–p95 and p95–p99 groups are not censored we can use the unadjusted internal data to compare the sources of income for members of these groups. As discussed in Appendix B, for

the p90–p95 and p95–p99 income groups, the sources of income for members of these groups are also quite similar between the CPS Pre-TU series and the Piketty-Saez Series. For example, in any given year between 85.1 to 89.3 percent of income received by members of the p90–p95 group comes from wages in the CPS Pre-TU data. This compared to a range of 86.9 to 91.6 percent of income received from wages in this group when using the IRS tax records data. Among the p95–p99 income group, the income shares are equally as similar, with the 74.8 to 85.7 percent of income coming from wages in the CPS data and 73.3 to 84.4 percent of income coming from wages in the IRS tax records data.

Thus far, we have restricted our attention to groups with incomes lying between the 90th and the 99th percentiles. What about the top 1 percent? It is only within this group that we see larger differences in results across the datasets. Figure 3 shows that the CPS-Post-HH series leads to a smaller share estimate than the other two series. This is similar to our findings for the p90–p95 and p95–p99 income groups and is expected given the different income definitions. However, in contrast to the earlier findings for the other two income groups, while controlling for differences in definitions reduces this gap, a more sizable unexplained gap remains. It is worth emphasizing, however, that while the remaining difference is greater than for the other two income groups analyzed, the differences in absolute terms between the CPS Pre-TU series and the IRS series are relatively small, at least in earlier years. Before 1986 the income share for the top 1 percent is between 1 and 2 percentage points greater for the Piketty-Saez estimates relative to the CPS-Pre-TU series, although this difference expands in later years.

Arguably, inequality trends over time are more important to researchers than inequality levels. In both the CPS Pre-TU series and the Piketty-Saez series we find slower growth in the share of income held by the p90–p95 and p95–p99 groups starting in the early 1990s than was

the case in the 1980s. Thus, both the CPS and IRS data sources seem to yield the result that whatever inequality growth occurred in the 1990s was largely confined to increases in the share of income held by the top 1 percent of the income distribution.

So what precisely has been happening to the top 1 percent's share? Prior to 1986, the trends in the income share for this group are remarkably similar according to all three series. Table 1 shows the average annual percent increases in the top 1 percent's income share for seven subperiods. The two pre-1986 periods are the relatively low inequality growth period of the 1970s and the higher inequality growth period from 1980–1986. Each of the three series shows similarly small inequality growth in the 1970s, with the Piketty-Saez series and the CPS Post-HH series each showing a very slight decrease in the top 1 percent income shares and the CPS Pre-TU series showing a very slight increase. The 1980–1986 period is even more similar, as the Piketty-Saez series shows almost identical average growth in the share held by the top 1 percent as found using the two CPS series. It is only after 1986 that more substantial differences between the series begin to appear. The first of these differences occurs from 1986–1988, when the Piketty-Saez series shows a dramatic 22.1 percent annual increase in the share of income going to the top one percent. The increase according to the CPS-Pre-TU series is a more moderate 2.0 percent.

This divergence between series subsides in the period immediately after 1988. Compared to the CPS-Post-HH series, the Piketty-Saez series shows moderately higher growth in the income share of the top 1 percent – growth of 0.6 percent per year compared to no growth in the CPS-Post-HH series from 1988 to 1992. Much of this difference, however, is simply due to the different income distribution definitions. When the CPS-Pre-TU series is used instead, the difference in the top 1 percent's income share between this series and the Piketty-Saez one is a

much smaller 0.2 percent per year. Thus, for the entire period between 1967 and 1992 with the exception of 1986–1988, the trends in the income share of the top 1 percent are similar according to both data sources – if similar income definitions are used.

From 1992–1993, the trends diverge again across series. In this year, both CPS series increase by over 40 percent while the IRS series falls by 4.9 percent. But it is only from 1993–2000 that the IRS series shows a sustained increase in the share of income held by the top 1 percent relative to CPS-Pre-TU series. Over this period, the Piketty-Saez series estimates that the top one percent's share was rising at an accelerated pace. The 4.1 percent annual increase is more than twice the rate of increase in the early 1980s. By contrast, the CPS-Pre-TU series yields an annual increase of only 1.5 percent in the income share of the top 1 percent – which is a slower rate of increase than seen in the 1980s.

After the divergence for the 1990s, trends across series converged again from 2000–2006 if similar income distribution definitions are used. During this period, all three series show similar increases of between 1.3 and 1.5 percent average annual increases in the income share for the top 1 percent.

So, for most of the past 40 years, the trends in top income shares are similar – once similar income definitions are used. There are no major differences in the trends implied by the difference sources for the income shares of those with incomes between the 90th and 99th percentiles. It is only during the periods 1986–1988, 1992–1993, and 1993–2000, that the two sources show markedly different trends and only for the top 1 percent of the population.

Explaining the differences in trends in the share of the top 1 percent

What explains the divergences between series in estimates of the share of the top 1 percent for the periods 1986–1988, 1992–1993, and 1993–2000. We believe that the results for the first two periods arise from well-known limitations of the IRS tax return data and of the CPS, respectively.

For 1986–1988, we believe that the increased share of the top 1 percent shown by the Piketty-Saez series primarily reflects a change in tax policy rather than any genuine change in the incomes controlled by the richest 1 percent. The Tax Reform Act of 1986 provided substantial incentives for the very richest tax units to switch reported income from Subchapter-C corporations to Subchapter-S income and wage income. The tax law changes likely created a behavioral effect in how income is reported, which led to the very large observed increase in top income shares in IRS personal tax return data over the course of these two years. See Slemrod (1996) and Reynolds (2006) for a fuller discussion of this type of issue in the Piketty and Saez (2003) data, and see Feenberg and Poterba (1993) for a more general discussion of the problems of measuring income inequality using tax return data. Of course, Piketty and Saez (2003) recognize the potential impact of such "fiscal manipulation" (2003, p.3), but they do not address the issue in detail because of their focus on long-run trends in top income shares.

The divergence in series for 1992–1993 reflects fundamental changes in the design of the CPS, rather than a real change in income inequality. Over these years, the Census Bureau implemented a major redesign of the survey instrument, including a change to computerized rather than paper-based data collection methods. (See Ryscavage 1995 and Jones and Weinberg 2000 for details.) These changes improved the ability of the CPS to record all incomes but especially top incomes. We believe that this explains the increase of more than 40 per cent in the

top 1 percent's share according to the two CPS-based series (Table 1). Notice the much more modest change in the Piketty-Saez series over these years.

What explains the divergences for 1993–2000? It may be the case, as Reynolds (2006) suggests, that changes in tax rules, requiring executive stock options to be reported as taxable income, led to the estimated rise in income share of the top 1 percent according to the Piketty-Saez series. According to this hypothesis, this group's income share has always been higher than observed (implying a greater difference between the Piketty-Saez and CPS-Pre-TU series). And importantly, trends according to the two series are more similar on the grounds that the more rapid increase in the Piketty-Saez series in the 1990s was an artifact of the changes in tax accounting rules.

Another possible explanation, also suggested by Reynolds (2006), is that a greater increase in the use of tax-deferred savings accounts (401k plans, Keogh plans and IRA tax shelters) by individual in top income groups outside the top 1 percent may explain part of the rise in the income share of the top 1 percent in the Piketty-Saez series for the late 1990s. This hypothesis would be consistent with our results for the p90–p95 and p95–p99 groups, for which the CPS-based series showed very slightly higher increases in income shares than the Piketty-Saez series.

Either of these explanations for the diverging trends is plausible. So too may be the view that the CPS did an increasingly poorer job of capturing top incomes in the late 1990s. But, if this is the explanation, the timing of the differences is curious. After the CPS redesign in 1993, it was better able to capture top incomes, as evidenced by the artificial jump in inequality in both of our CPS series between 1992 and 1993. Moreover, the prevalence of censoring during this period – after the internal data's topcodes were increased – was lower than it was in the mid-

1980s or in the early 21st century.[8] So the CPS design changes should have increased the survey's ability to accurately observe top incomes during this period.

Additional work is necessary to determine what precisely happened to the very highest income shares over this period and thereby provide a comprehensive reconciliation of the differences between the CPS-based and IRS tax return-based series.

Income inequality trends using Gini coefficients

Thus far we have explored the ability of CPS data to capture trends in the share of pre-tax pre-transfer income going to top tax units in the IRS tax record data as measured by Piketty and Saez. In this section we explore the sensitivity of inequality levels and trends to one's choice of inequality index as well as sources of income and income receiving unit. Since a top income share is the only inequality measure that can be readily derived from IRS tax record data we focus this part of the analysis on the CPS data.[9] Figure 4 compares Gini coefficients based on the CPS Post-HH income series to those from the CPS Pre-TU income series. If the choice of income definition and income sharing unit did not matter, we would expect to find similar levels and trends in Gini coefficients using each of these two series. Instead we find that using pre-transfer, tax-unit data rather than post-transfer household income yields substantially higher observed levels of income inequality. Using post-transfer household income, estimates of the Gini coefficient range from a low of 0.35 in 1968 to a high of 0.46 in 2006. Switching to pre-

---

[8] See Larrimore et al. (2008) for detailed information about the prevalence of censoring in the internal CPS data year by year.
[9] Leigh (2007) uses unit record data to compare top income shares with other inequality estimates, but using panel data regressions applied to cross-national comparative data. Ours is the first study to investigate this issue using a long run of comparable microdata from the same country.

transfer tax-unit income increases observed inequality 30 to 40 percent over these levels, with a minimum Gini coefficient of 0.47 in 1968 and a maximum of 0.59 in 2006.[10]

This dramatic difference in income inequality levels occurs for two reasons. First, defining the sharing unit as the tax unit causes adult children living with their parents or other individuals who have little or no independent income but are supported by other household members to be counted independently. As a result, the fraction of the population who appear to have no income is much higher when considering tax-units than when considering households resulting in increases in measured income inequality.

Second, transfer income is predominantly received by individuals in the lower tail of the income distribution. Ignoring transfer income therefore reduces the income of individuals at the bottom of the distribution resulting in increases in measured income inequality. Of course, these factors also affect calculations of income inequality when using the top income shares measure as well. However, the difference in observed inequality between the household and tax-unit income series is much larger when using the Gini coefficient than was seen for the top income shares. This is because the Gini coefficient incorporates information about inequality differences throughout the income distribution, and the top income share measures do not. Since transfers are most relevant for individuals lower in the income distribution, they have a larger impact on comprehensive inequality measures such as the Gini coefficient. The substantial differences in levels of observed inequality from what appears to be a relatively innocuous change in income and income receiving definitions, however, illustrates the importance of careful attention to detail when comparing income inequality calculations. But, as can also be seen in Figure 4, these

---

[10] In addition to the difference in sharing unit and the inclusion of transfer income, the series also differ in that the post-transfer household income (CPS Post-HH) is size-adjusted and evaluated at the individual level the pre-transfer tax-unit income (CPS Pre-TU) is not. Not size-adjusting household income would slightly reduce the level difference between the Gini coefficients in the two series, but most of the difference remains. The size-adjustment does not affect estimates of the trends in inequality.

choices appear to be less critical with respect to trends. As will be seen below, this is not the case with respect to the choice of income inequality index.

Comparing income inequality trends using Gini coefficients and top income shares

Finally, we also consider how the choice of income inequality index affects measured household income inequality. Using the two CPS-based series, we compare the observed growth in income inequality using the Gini coefficient to the trend in the income share of the top 1 percent and the top 10 percent of the population. By using the same sample to compare results for these three inequality measures, we can determine the extent to which the choice of inequality measures influences the observed trends in income inequality.

Table 2 shows the average annual percent increases using these three income inequality measures for seven subperiods since 1967 and for the entire 40 year period. This is done using the CPS Post-HH series and the CPS Pre-TU series – our two series for which all three inequality metrics can be calculated. Using either income series, the two top income share series exhibit faster inequality growth than the Gini series when considering the entire 40 year period. However, much of this difference comes from the substantial, artificial jump in top income shares between 1992 and 1993.

When considering the subperiods, the pattern is mixed with the top 1 percent's income share exhibiting higher growth than the Gini coefficient in some periods (1980–1988, 1993–2000, and 2000-2006) and slower growth in others (1967-1980 and 1988–1992). During the period of greatest disagreement between the two literatures – the late 1990s where the IRS-based literature has observed much larger increases in income inequality – this difference is quite large. Using the CPS Post-HH series, the growth in inequality as measured by the top 1 percent's income

share grew an average of 1.6 percent per year. This compares to an average annual growth of just 0.2 percent per year in the Gini coefficient. The difference is similarly large when using the CPS Pre-TU series, with the top 1 percent's income share growing an average of 1.5 percent per year and the Gini showing no growth over the period. (The growth in the top 10 percent's income share is much closer to that of the Gini.)

These results can partially explain why researchers examining top income shares using IRS tax records have found continued inequality growth through the 1990s while researchers examining Gini coefficients using CPS data have not. We previously observed some differences in inequality trends between the two datasets during this period even using the same inequality measure. However, Table 2 shows that differences in the inequality trends observed in these two literatures also stem from differences in the inequality index used. For researchers interested in inequality across the entire distribution, which the Gini coefficient is superior for measuring, inequality growth in the 1990s was dramatically slower than that in the 1980s. But for researchers interested in comparing the income differentials between the very top income holders and the rest of society, then the slowdown in inequality growth in the 1990s was far less substantial.

Summary and Conclusions

We analyze trends in top income shares in the USA over four decades (1967–2006), with the goal of reconciling estimates derived from the CPS with those reported by Piketty and Saez (2003) and derived from IRS tax return data. Our CPS-based estimates draw on the internal data used by the Census Bureau to produce their official income statistics, which is a much better

source for examining income distribution trends than CPS public-use data because the prevalence of topcoding is substantially smaller.

When applying a Piketty-Saez-type definition of the income distribution to CPS data, we derive estimates of top income shares that are remarkably similar in terms of both levels and trends to those reported by Piketty and Saez (2003, 2008) for both the p90–p95 and p95–p99 groups. The shares grew in the 1980s and then slowed starting in the early 1990s. For the top 1 percent, our CPS-Pre-TU series provides a slightly lower share estimates than the Piketty-Saez series does but, with the exception of the period 1993–2000, the trends in the series are remarkably similar. Thus, we conclude that the differences in inequality trends observed by researchers using these two data sources are not primarily due to deficiencies in either data source but rather to the traditions of income inequality measurement used in the two literatures.

To explore this possibility further we also measure income inequality using Gini coefficient in the March CPS data, and compare results using the Piketty and Saez–style source of income and income receiving unit definitions (CPS Pre-TU) and those using standard source of income and income receiving unit definitions (CPS Post-HH). Using Piketty-Saez pre-tax pre-transfer, tax-units substantially increases observed levels of income inequality but does not greatly impact trends (Figure 4). In contrast, when using identical data, source of income, and income receiving units but different inequality measures, we found that the growth in the income share of the top 1 percent of the population substantially outpaced measured inequality using the Gini coefficient (Table 2).

Thus, we conclude that at least part of the differing views in the two literatures about recent trends in income inequality can be attributed to differences in the literatures' measures of income inequality. Specifically, while the income divergence between the very top income

22

holders and the rest of society was growing in the 1990s, the growth in income inequality across the entire distribution occurred at a more moderate pace.

When we use the same measure of income inequality – the income share of the top 1 percent – and similar income definitions – pre-transfer, tax-unit income – with the CPS data we are for the most part able to very closely capture the same levels and trends Piketty and Saez find using the IRS tax record data. The only unexplained divergence in the observed income inequality between the two datasets occurs over the period 1993–2000. It is possible that in this period of rapid economic growth, the CPS was unable to capture the rise in pre-tax pre-transfer income of the very richest people. It is also possible that behavioral effects caused by changes in the tax laws made it more likely for an increase in the sheltering of income by those at the top of the distribution but outside the top 1 percent, which then exaggerated the change in incomes recorded by IRS tax return data. Hence the difficulty of disentangling real changes in the share of income controlled by the very richest income tax units from changes in the way they report their income as the source of these yearly changes in inequality. But despite this limitation, users of both CPS and of IRS tax return data should be comforted by our finding that, for most groups at the top and for most of the past four decades, the differences in estimates from the two data sources are relatively minor.

References

Atkinson, Anthony B. 2005. "Top Incomes in the UK Over the Twentieth Century." *Journal of the Royal Statistical Society, Series A* 168: 325–343.

Atkinson, Anthony B. and Andrea Brandolini. 2001. "Promises and Pitfalls in the Use of Secondary Data Sets: Income Inequality in OECD Countries as a Case Study." *Journal of Economic Literature*, 39 (3): 771–799.

Atkinson, Anthony B. and Andrew Leigh. 2007. "The Distribution of Top Incomes in Australia", *Economic Record*, 83(262): 247–61.

Atkinson, Anthony B. and Piketty, Thomas 2007. *Top Incomes over the Twentieth Century: A Contrast Between Continental European and English Speaking Countries*, Oxford: Oxford University Press.

Atkinson, Anthony B., Lee Rainwater, and Timothy Smeeding. 1995. *Income Distribution in OECD Countries. Evidence from the Luxembourg Income Study*. Social Policy Studies No. 18. Paris: Organization for Economic Cooperation and Development.

Bach, Stefan, Giacomo Corneo, and Viktor Steiner. 2009. "From Bottom to Top: The Entire Income Distribution in Germany, 1992-2003." *Review of Income and Wealth*, 55(2): 303-330.

Bandourian, Ripsy, James B. McDonald, and Robert S. Turley. 2003. "A Comparison of Parametric Models of Income Distribution Across Countries and Over Time." *Estadistica*, 55: 135–152.

Bishop, John A., Jong-Rong Chiou, and John P. Formby. 1994. "Truncation Bias and the Ordinal Evaluation of Income Inequality." *Journal of Business and Economic Statistics*, 12: 123–127.

Bordley, Robert F., James B. McDonald, and Anand Mantrala. 1996. "Something New, Something Old: Parametric Models for the Size Distribution of Income." *Journal of Income Distribution*, 6: 91–103.

Brachmann, Klaus, Andreas Stich, and Mark Trede. 1996. "Evaluating Parametric Income Distribution Models." *Allgemeines Statistisches Archiv*, 80: 285–298.

Burkhauser, Richard V., Kenneth A. Couch, Andrew J. Houtenville, and Ludmila Rovba. 2003-2004. "Income Inequality in the 1990s: Re-Forging a Lost Relationship?" *Journal of Income Distribution,* 12 (3–4): 8–35.

Burkhauser, Richard V., Shuaizhang Feng, Stephen Jenkins, and Jeff Larrimore. 2008. "Estimating Trends in US Income Inequality Using the Current Population Survey: The

Importance of Controlling for Censoring," NBER Working Paper number w14247, Cambridge MA: National Bureau of Economic Research.

Burkhauser, Richard V., Shuaizhang Feng, and Stephen P. Jenkins. 2009. "Using the P90/P10 Ratio to Measure US Inequality Trends with Current Population Survey Data: A View from Inside the Census Bureau Vaults." *Review of Income and Wealth*, 55(1): 166–185.

Burtless, Gary and Timothy M. Smeeding. 2002. "The Level, Trends, and Composition of Poverty." in *Understanding Poverty*, ed. Sheldon Danziger and Robert H. Haveman, Cambridge, MA: Harvard University Press, 27–68.

Corneo, Giacomo and Christina M. Fong. 2008. "What's the Monetary Value of Distributive Justice?" *Journal of Public Economics*, 92(1-2): 289–308.

Daly, Mary C. and Robert G. Valletta. 2006. "Inequality and Poverty in the United States: The Effects of Rising Dispersion of Men's Earnings and Changing Family Behavior." *Economica,* 73, 75-98.

Dell, Fabien. 2005. "Top Incomes in Germany and Switzerland Over the Twentieth Century," *Journal of the European Economic Association*, 3(2–3): 412–421.

Economists View website. 2007. "Increasing Inequality is Not a Statistical Illusion." http://economistsview.typepad.com/economistsview/2007/01/increasing_ineq.html Last accessed August 28, 2009.

Feenberg, Daniel and James M. Poterba. 1993. "Income Inequality and the Incomes of Very High-Income Taxpayers: Evidence from Tax Returns," in *Tax Policy and the Economy*, ed. James M. Poterba, Cambridge, MA: NBER/MIT Press, 145–177.

Fichtenbaum, Rudy and Hushang Shahidi. 1988. "Truncation Bias and the Measurement of Income Inequality." *Journal of Business and Economic Statistics*, 6: 335–337.

Goolsbee, Austan. 2000. "What Happens When You Tax the Rich? Evidence from Executive Compensation." *Journal of Political Economy*, 108 (2): 352–378.

Gottschalk, Peter and Timothy M. Smeeding. 1997. "Cross-National Comparisons of Earnings and Income Inequality." *Journal of Economic Literature,* 35 (2): 633–687.

Gottschalk, Peter and Sheldon Danziger. 2005. "Inequality of Wage Rates, Earnings and Family Income in the United States, 1975–2002." *Review of Income and Wealth*, 51 (2): 231–254.

Jenkins, Stephen P. 2009. "Distributionally-Sensitive Inequality Indices and the GB2 Income Distribution." *Review of Income and Wealth*, 55 (2): 392–398.

Jenkins, Stephen P., Richard V. Burkhauser, Shuaizhang Feng, and Jeff Larrimore. 2009. "Measuring Inequality with Censored Data: A Multiple Imputation Approach." US Census Bureau Center for Economic Studies Working Paper CES-WP-09-05. http://www.ces.census.gov/index.php/ces/discussion_papers?detail_key=101854

Jones, Arthur F., Jr., and Weinberg, Daniel H. 2000. *The Changing Shape of the Nation's Income Distribution*. U.S. Census Bureau, Current Population Reports, P60-204. Washington, DC: GPO

Larrimore, Jeff, Richard V. Burkhauser, Shuaizhang Feng and Laura Zayatz. 2008. "Consistent Cell Means for Topcoded Incomes in the Public Use March CPS (1976–2007)." *Journal of Economic and Social Measurement*, 33(2–3): 89–128.

Leigh, Andrew. 2007. "How Closely do Top Income Shares Track Other Measures of Inequality?" *Economic Journal*, 117(524): F619–F633.

Leigh, Andrew. 2009. "Top Incomes" in *The Oxford Handbook of Economic Inequality,* eds. Wiemer Salverda, Brian Nolan, and Timothy M. Smeeding. New York: Oxford University Press, 150-176.

Levy, Frank and Murnane, Richard J. 1992. "U.S. Earnings Levels and Earnings Inequality: A Review of Recent Trends and Proposed Explanations." *Journal of Economic Literature*, 30(3): 1333–1381.

Piketty, Thomas. 2003. "Income Inequality in France, 1901–1998," *Journal of Political Economy*, 111(5): 1004–1042.

Piketty, Thomas and Emmanuel Saez. 2003. "Income Inequality in the United States, 1913–1998." *Quarterly Journal of Economics*, 118 (1): 1–39.

Piketty, Thomas and Emmanuel Saez, 2006*a*. "The Evolution of Top Incomes: A Historical and International Perspective," *American Economic Review, Papers and Proceedings*, 96(2): 200–205.

Piketty, Thomas and Emmanuel Saez. 2006*b*. "Response by Thomas Piketty and Emmanuel Saez to: The top 1% … of What? By Alan Reynolds," Available online at: http://www.econ.berkeley.edu/~saez/answer-WSJreynolds.pdf

Piketty, Thomas and Emmanuel Saez, 2007. "Income and Wage Inequality in the United States, 1913-2002" in *Top Incomes over the 20th Century: A Contrast Between Continental European and English Speaking Countries*, eds. Anthony B. Atkinson and Thomas Piketty, Cambridge, MA: NBER/MIT Press, 141–225.

Piketty, Thomas and Emmanuel Saez. 2008. "Income Inequality in the United States, 1913-1998 (Tables and Figures Updated to 2006)," available online at: http://www.econ.berkeley.edu/~saez/TabFig2006.xls

Reiter, Jerome P. 2003. "Inference for Partially Synthetic, Public Use Microdata Sets," *Survey Methodology*, 29: 181–188.

Reynolds, Alan. 2006. *Income and Wealth*. Westport, CT: Greenwood Press.

Reynolds, Alan. 2007. "Has U.S. Income Inequality Really Increased?" *Policy Analysis*, 586: 1–24.

Rubin, Donald B. 1987. *Multiple Imputation for Nonresponse in Surveys*. New York: Wiley.

Ryscavage Paul 1995. "A Surge in Growing Income Inequality?" *Monthly Labor Review*, 188 (8): 51–61.

Saez, Emmanuel and Michael R. Veall. 2005. "The Evolution of High Incomes in Northern America: Lessons from Canadian Evidence," *American Economic Review*, 95(3): 831–49.

Scholes, Myron and Mark A. Wolfson. 1992. *Taxes and Business Strategy: A Planning Approach*. Englewood Cliffs: NJ: Prentice Hall.

Scholz, John K. and Kara Levine. 2002. "The Evolution of Income Support Policy in Recent Decades" in *Understanding Poverty*, ed. Sheldon Danziger and Robert H. Haveman, Cambridge, MA: Harvard University Press, 193–228.

Sivadasan, Jagadeesh and Michael Slemrod. Forthcoming. "Tax Law Changes, Income Shifting, and Measured Wage Inequality: Evidence from India," *Journal of Public Economics*.

Slemrod, Joel. 1995. "Income Creation or Income Shifting? Behavioral Responses to the Tax Reform Act of 1986," *American Economic Review, Papers and Proceedings*, 85(2): 175–180.

Slemrod, Joel. 1996. "High-income Families and the Tax Changes of the 1980s: The Anatomy of Behavioral Response," in *Empirical Foundations of Household Taxation*, eds. Martin Feldstein and James M. Poterba. Chicago: University of Chicago Press (for NBER).

U.S. Census Bureau. Various years. *Income, Poverty, and Health Insurance Coverage in the United States: various years*. Current Population Reports, P-60 series, Washington DC: GPO.

Woodbury, Stephen A. and Daniel S. Hamermesh. 1992. "Taxes, Fringe Benefits, and Faculty," *Review of Economics and Statistics*, 74(2): 287–296.

Figure 1: Estimates from CPS and IRS tax return data of the share of total income held by units with incomes between the 90th and 95th percentiles, 1967–2006
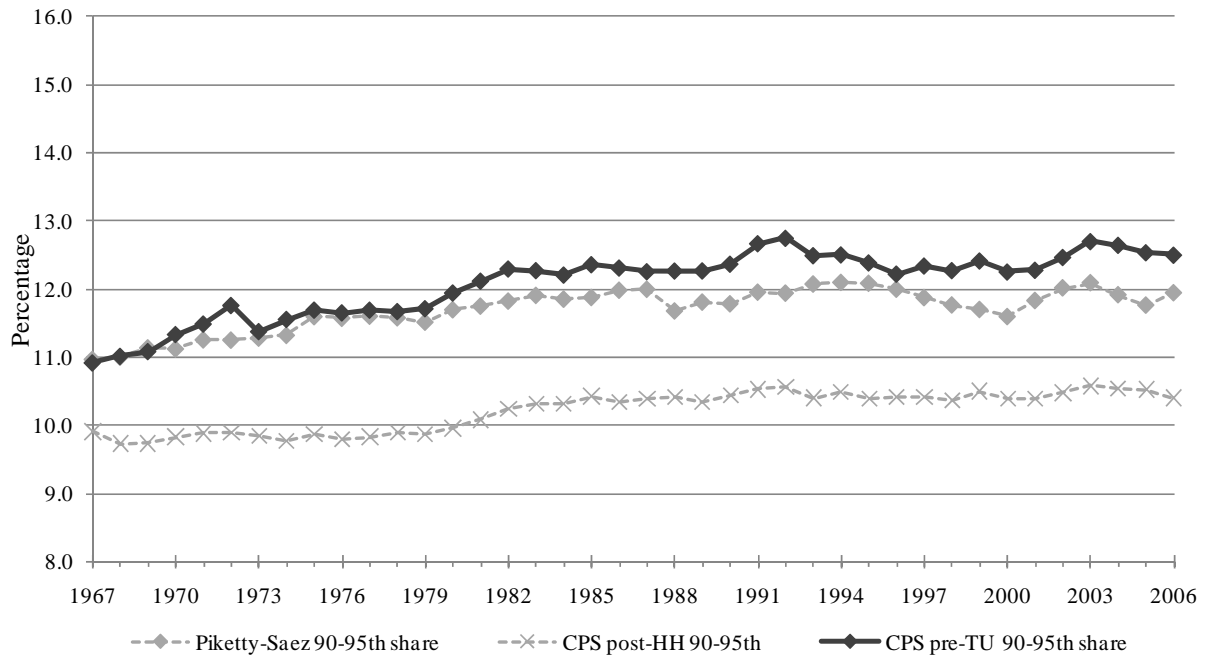
Sources. The Piketty-Saez series is taken from Piketty and Saez (2003, 2008). It refers to the distribution of pre-tax pre-transfer income among tax units. The CPS-based series were derived by the authors from CPS internal data. The CPS-Pre-TU series was derived using the Piketty-Saez definition; the CPS-Post-HH series refers to the distribution of size-adjusted pre-tax post-transfer household income among individuals. See main text for further details.

Figure 2: Estimates from CPS and IRS tax return data of the share of total income held by units with incomes between the 95th and 99th percentiles, 1967–2006
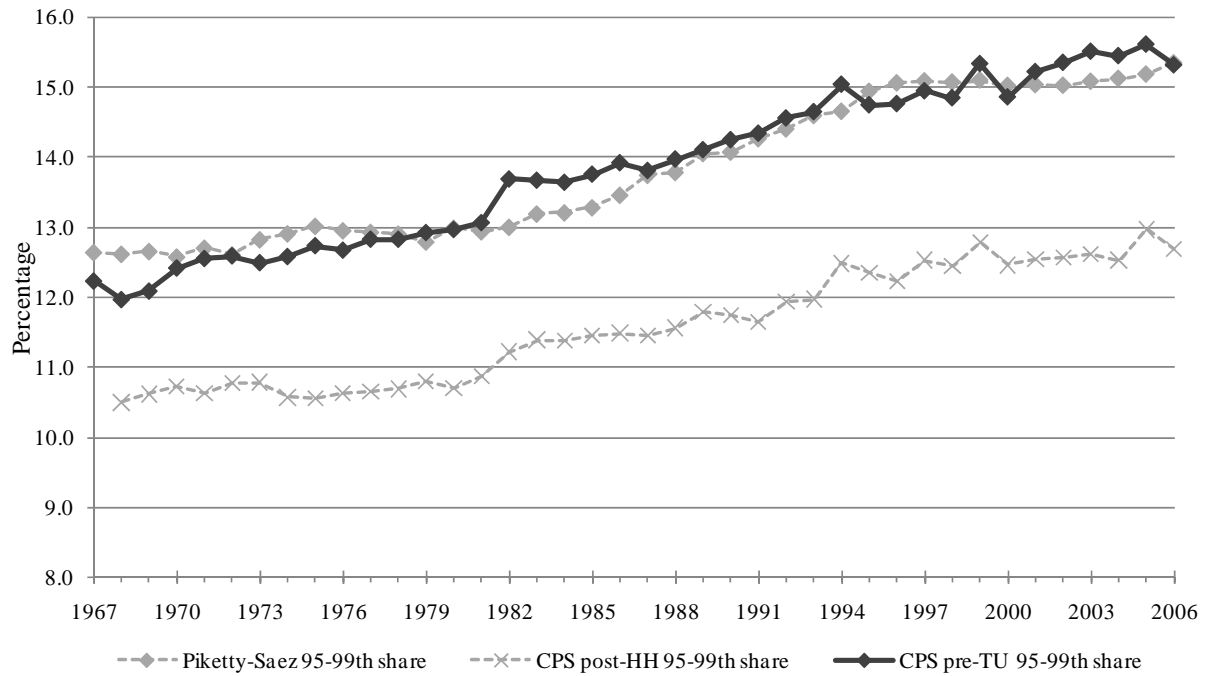
Sources: see note to Figure 1.

Figure 3: Estimates from CPS and IRS tax return data of the share of total income held by the top 1 percent, 1967–2006
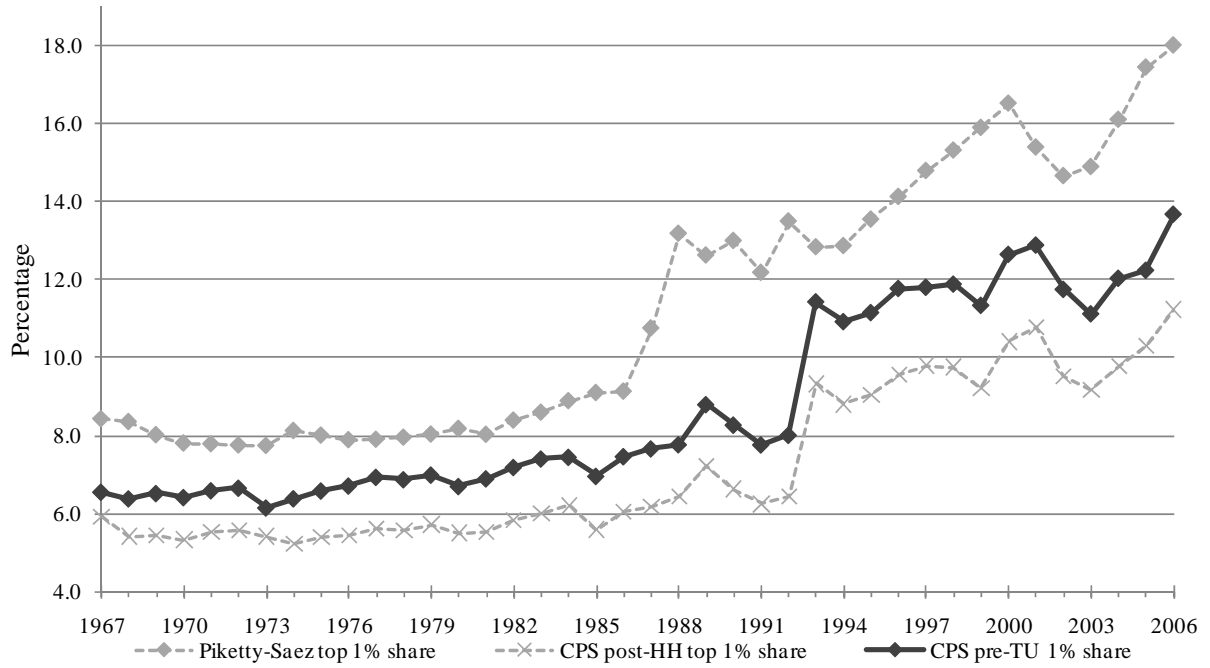
Sources: see note to Figure 1.

Figure 4: Estimates from CPS data of income inequality measured by the Gini coefficient, 1967–2006
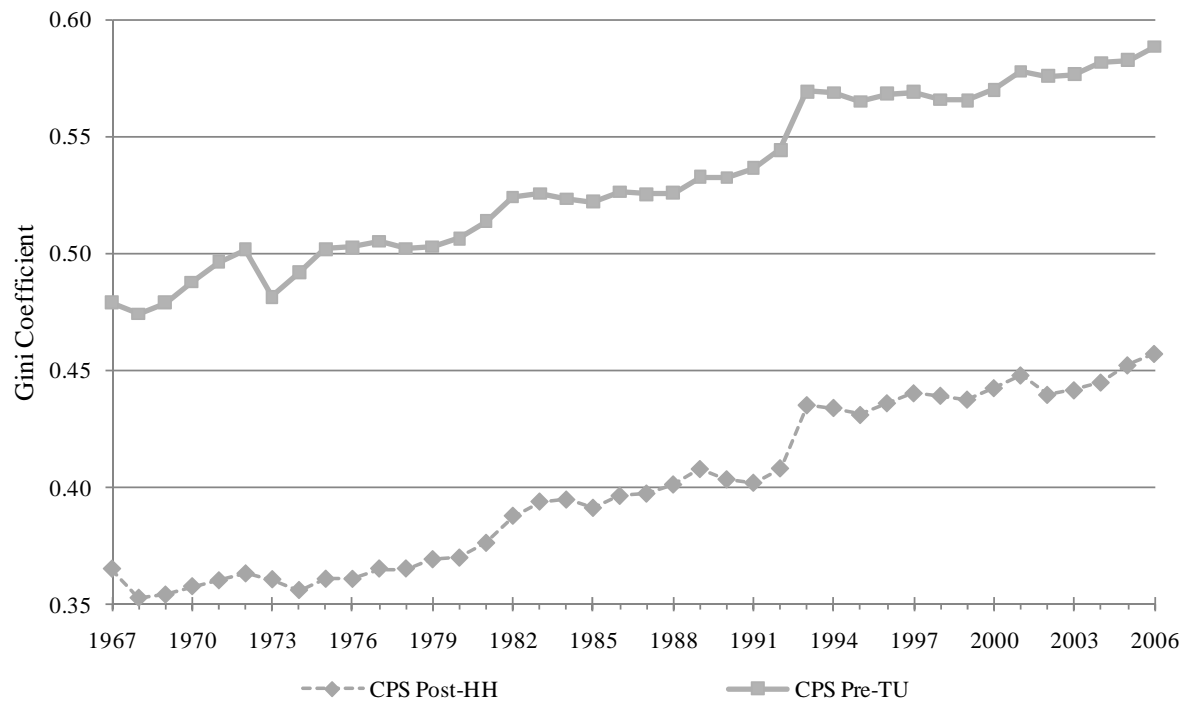
Sources: see note to Figure 1.

Table 1: Average annual percentage change in income share of the top 1 percent, by subperiod between 1967 and 2006

| Subperiod | March CPS | | IRS tax return data |
	Size-adjusted pre-tax post-transfer household income among individuals ("CPU-Post-HH")	Pre-tax pre-transfer tax unit income among tax units ("CPU-Pre-TU")	Pre-tax pre-transfer tax unit income among tax units ("Piketty-Saez")
1967–1980	–0.5	0.4	–0.2
1980–1986	1.7	1.9	1.9
1986–1988	3.2	2.0	22.1
1988–1992	0.0	0.8	0.6
1992–1993	45.0	42.5	–4.9
1993–2000	1.6	1.5	4.1
2000–2006	1.3	1.4	1.5

Note: sources for calculations as described in the note to Figure 1.

Table 2: Average annual percentage change in income inequality using three inequality measures, by subperiod between 1967 and 2006

| | CPS Post-HH | | | CPS Pre-TU | | |
	Gini	Share of Top 1%	Share of Top 10%	Gini	Share of Top 1%	Share of Top 10%
1967–1980	0.1	–0.5	–0.2	0.4	0.2	0.5
1980–1986	1.2	1.7	1.1	0.7	1.9	1.1
1986–1988	0.6	3.2	0.9	–0.1	2.0	0.5
1988–1992	0.4	0.0	0.5	0.9	0.8	1.0
1992–1993	6.6	45.0	9.6	4.6	42.5	9.2
1993–2000	0.2	1.6	0.7	0.0	1.5	0.4
2000–2006	0.6	1.3	0.5	0.5	1.4	0.7
1967–2006	0.6	2.3	0.7	0.6	2.8	1.0

Sources: see note to Figure 1.

Appendix A. Findings based on unadjusted internal CPS data

To account for censoring (albeit limited) in the internal CPS data, we used a multiple imputation (MI) approach, as described in the main text. Imputation of some kind is necessary when one wishes to calculate income inequality for the entire income distribution including topcoded observations, and have been commonly-used in both the CPS-based inequality literature and the IRS-based inequality literature.

To investigate the potential sensitivity of our findings to the use of imputation methods, we have also undertaken all our calculations of top income shares using unadjusted internal CPS data used "as is", without imputations for censored values. Appendix Figures A1 through A3 provide the top income shares using the pre-tax, pre-transfer tax-unit income definition using both the unadjusted internal CPS data series and our series derived using the MI procedure.

For the p90–p95 income group and the p95–p99 income group, the levels and trends in income shares derived from the unadjusted CPS data closely match those from the MI series, and both are close to the levels and trends seen by Piketty and Saez (2003) using IRS tax records. This is not unexpected as less than 1 percent of individuals lived in a household in which some source of income was censored in the internal data so the censoring primarily impacts those in the very top income group.

For the top 1 percent income share, using the unadjusted CPS data results in a lower level of measured income inequality and a slightly lower income inequality growth than when using the MI series. The general patterns of the inequality increases are similar, however, with the top 1 percent income share increasing at a pace similar to that seen by Piketty and Saez in the 1980s (although the unadjusted internal data observes the increase later in the 1980s than the other two series). As with the MI series, the rate of increase in the top 1 percent income share then slows in

the 1990s compared to that seen by Piketty and Saez before showing similar patterns again in the early 21st century.

Thus, even for researchers using the unadjusted internal data without our GB2-based multiple imputations for the small number of internally censored observations, our main findings hold. Controlling for differences in income definitions and inequality measurements, the results using CPS and IRS data are extremely consistent for almost all of the past 30 years with the exception of the mid- to late-1990s.

Appendix Figure A1: Estimates from CPS data with and without the GB2 imputation procedure for censored observations of total income held by units with incomes between the 90th and 95th percentiles, 1967–2006
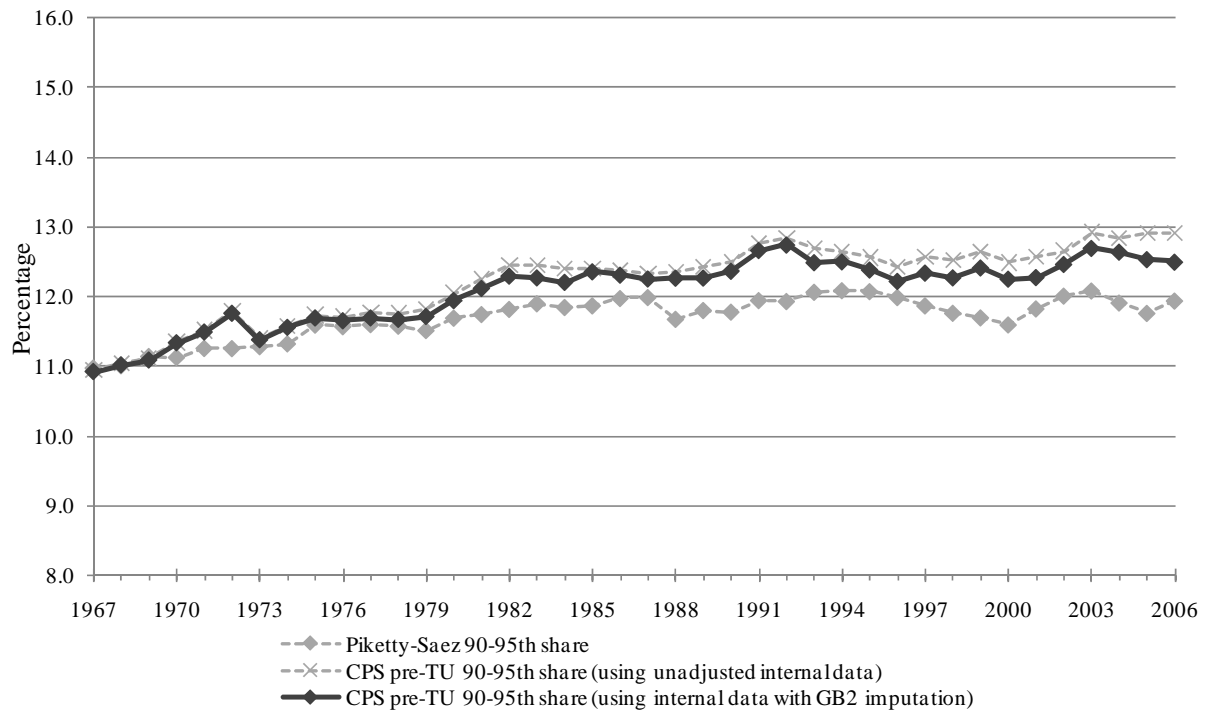
Sources. The Piketty-Saez series is taken from Piketty and Saez (2003, 2008). It refers to the distribution of pre-tax pre-transfer income among tax units. The CPS-based series were derived by the authors from CPS internal data. The CPS-Pre-TU series with the GB2 imputation matches the CPS-Pre-TU series from the main text, using our GB2 imputation to derive censored incomes in the internal data. The CPS-Pre-TU series using unadjusted internal data uses the unadjusted internal data "as is". See Appendix A and the main text for further details.

Appendix Figure A2: Estimates from CPS data with and without the GB2 imputation procedure for censored observations of total income held by units with incomes between the 95th and 99th percentiles, 1967–2006
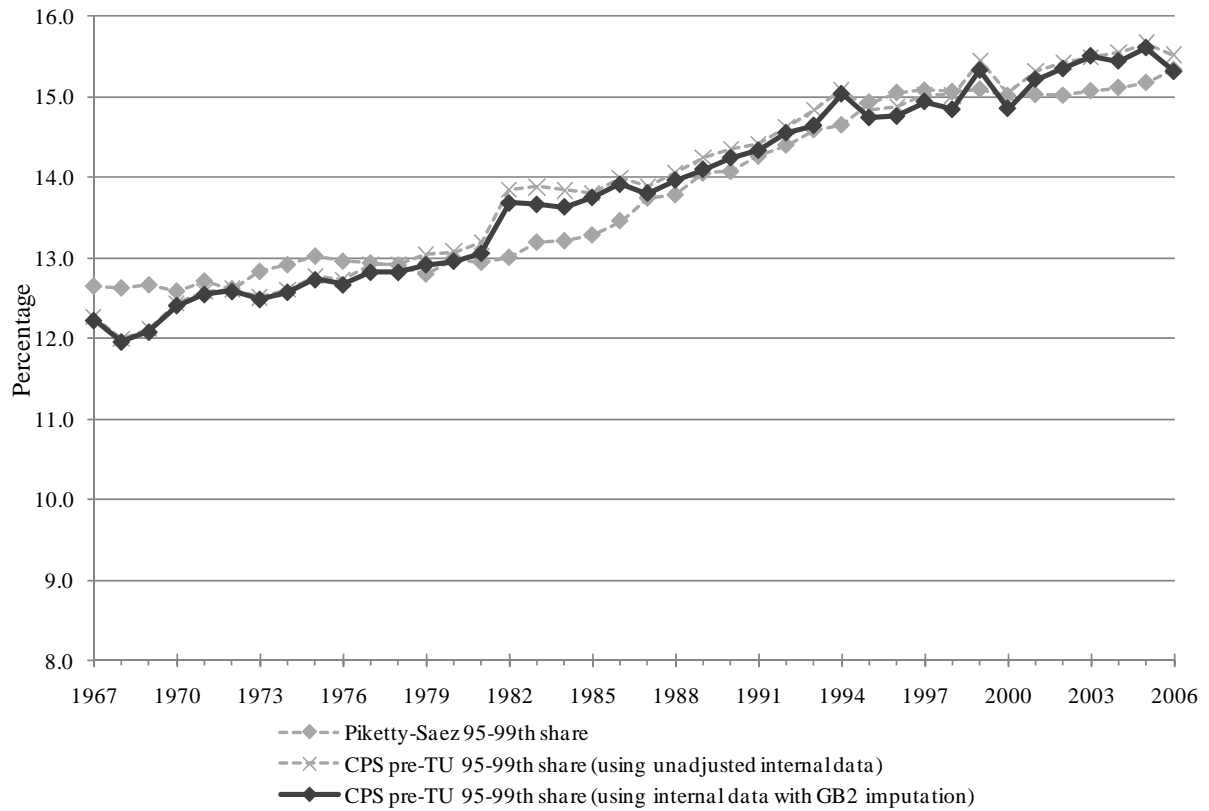
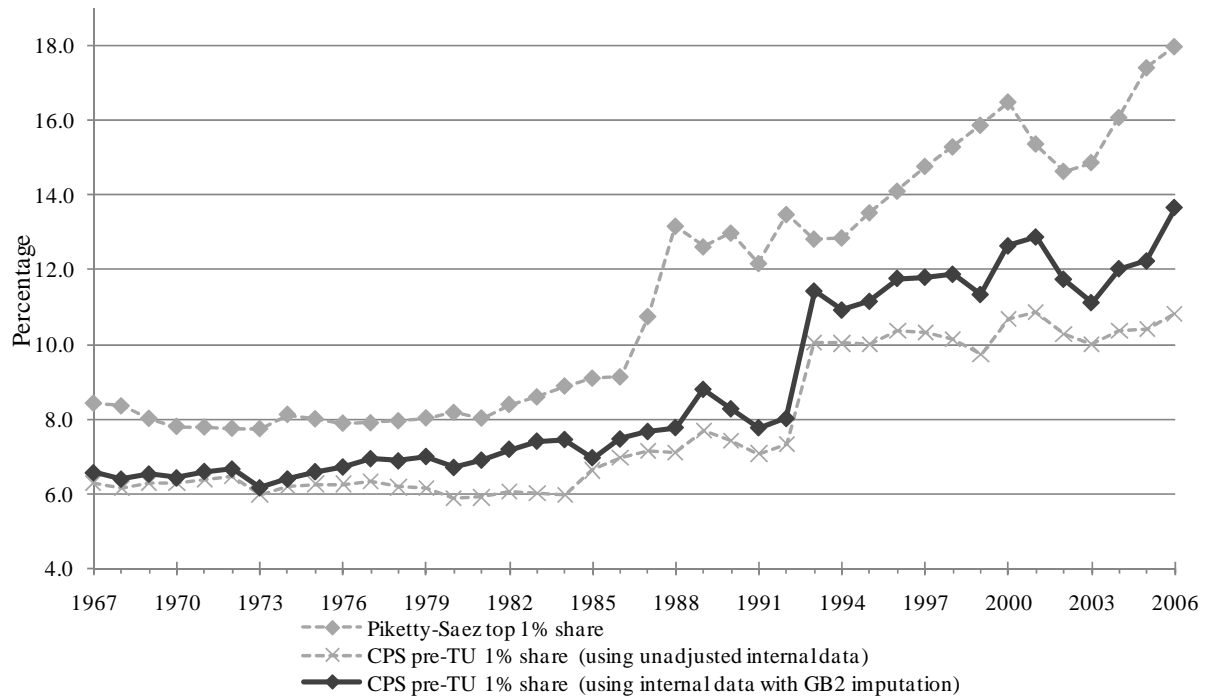Sources: see note to Appendix Figure A1.

Appendix Figure A3: Estimates from CPS data with and without the GB2 imputation procedure for censored observations of total income held by units with incomes in the top 1 percentile, 1967–2006

Sources: see note to Appendix Figure A1.

Appendix B. Sources of income: unadjusted internal CPS data versus Piketty-Saez data

Having established that in general the share of income in March CPS data going to the top 10 percent of the distribution closely matches that found in IRS tax record data by Piketty and Saez (2007), we also seek to understand how the sources of income compare for these individuals. When doing so, it is necessary to use the unadjusted internal data rather than the data based on our GB2-based multiple imputation (MI) procedure. This is because the MI procedure used to overcome the censoring problem must be performed on total income, rather than on each individual income source. As such, it is not possible to distinguish information about income sources in the imputed data.

Of course, by using the unadjusted internal CPS data rather than the MI data, we are unable to observe the actual incomes, or the sources of those incomes, for censored individuals. Therefore, since many individuals in the top 1 percent of the income distribution are censored we are only able to provide meaningful income source information for the p90–p95 and p95–p99 income groups where censoring is less restrictive.

As illustrated in Appendix Table B1, the sources of income are quite similar for the p90–p95 income groups in the CPS and IRS data. The percent of income among members of this group received from wages ranges from 85.1 to 89.3 percent of income when using the CPS Pre-TU data, compared to a range of 86.9 to 91.6 percent of income when using the IRS tax records data. While there are some year-to-year fluctuations in the income received from wages, the level is remarkably stable in both the IRS and CPS data over the 40 year period.

Among the p95–p99 income group, the income shares are also as similar, with the share of income received from wages ranging from 74.8 to 85.7 percent of income in the CPS data and from 73.3 to 84.4 percent of income in the IRS tax records data (Appendix Table B2). There are

only 2 years where the difference in the share of income received from wages is greater than 4 percent in the two datasets. Additionally, both datasets show increases in the portion of wages from income of approximately 7.5 percent over the 40 year period while the income from entrepreneurial activities declined. The only substantial difference between the series is that the IRS tax records data indicate that the portion of income from assets declined since 1967, whereas the CPS data suggest that asset income increased in importance to these high-income individuals. In general, however, not only do the IRS and CPS data closely match the share of income received by top earners in the income distribution, but they also provide similar results for the sources of that income.

Appendix Table B1: Income composition by source for tax-units with incomes between the 90th and 95th percentiles of the income distribution, 1967–2006

Year	CPS Pre-TU (Unadjusted)			Piketty-Saez		
	Wage	Entrepreneurial	Asset	Wage	Entrepreneurial	Asset
1967	86.3	10.3	3.3	88.2	7.3	4.6
1968	87.3	9.5	3.2	88.6	7.0	4.3
1969	88.8	7.7	3.4	88.6	6.8	4.6
1970	89.3	7.5	3.2	89.2	6.0	4.7
1971	88.1	8.8	3.2	90.1	5.6	4.3
1972	87.9	8.9	3.2	89.6	5.9	4.6
1973	87.3	9.1	3.6	88.8	6.4	4.9
1974	86.9	9.2	3.9	86.9	6.6	6.5
1975	87.5	8.6	3.9	88.7	5.6	5.7
1976	87.9	7.7	4.3	88.4	5.8	5.8
1977	88.2	7.7	4.1	88.7	5.4	5.9
1978	88.7	7.0	4.3	88.4	5.8	5.7
1979	87.8	7.6	4.6	89.1	5.2	5.7
1980	89.1	6.0	5.0	88.6	4.5	6.9
1981	87.3	6.6	6.1	88.1	3.7	8.2
1982	87.2	6.0	6.8	89.2	2.5	8.3
1983	85.1	7.5	7.3	89.5	3.4	7.1
1984	86.2	6.1	7.7	89.9	3.2	6.8
1985	86.9	5.8	7.4	89.9	3.2	6.8
1986	86.9	6.4	6.6	90.1	3.8	6.0
1987	86.5	8.0	5.4	90.1	4.3	5.6
1988	86.7	6.9	6.4	89.4	4.9	5.8
1989	85.7	7.8	6.5	88.6	4.9	6.5
1990	85.7	7.5	6.8	88.7	4.7	6.6
1991	86.3	7.8	5.9	89.4	4.7	5.9
1992	87.0	7.4	5.7	90.9	4.3	4.8
1993	88.2	6.3	5.6	90.9	5.0	4.2
1994	89.2	5.6	5.2	91.1	5.0	3.9
1995	88.4	5.6	6.0	91.6	4.5	3.9
1996	86.4	6.4	7.2	90.8	4.7	4.6
1997	85.7	6.2	8.2	91.0	4.8	4.2
1998	86.1	6.1	7.8	91.1	4.9	4.0
1999	85.4	6.4	8.2	90.6	5.5	3.9
2000	87.4	5.7	6.9	89.7	5.6	4.7
2001	87.8	5.8	6.5	91.2	5.0	3.8
2002	89.2	5.5	5.3	89.9	6.2	4.0
2003	88.5	5.4	6.0	90.1	6.3	3.6
2004	88.9	5.3	5.8	89.4	6.8	3.8
2005	88.5	5.0	6.5	88.1	7.5	4.4
2006	86.9	6.1	7.1	88.2	6.8	5.1

Sources: The Piketty-Saez series is calculated from Piketty and Saez (2007, 2008). The CPS-Pre-TU series using unadjusted internal data uses the unadjusted internal data "as is". See Appendix A and the main text for further details.

Entrepreneurial income includes self-employment and farm income. Asset income includes interest from interest, dividends, and rents. For comparability with the source-decomposition results presented in Piketty and Saez (2007), income from other sources are excluded and the sum of incomes from wages, entrepreneurial activities, and asset income is scaled to sum to 100 percent. Other forms of income represent less than 4 percent of income in all years.

Appendix Table B2: Income composition by source for tax-units with incomes between the 95th and 99th percentiles of the income distribution, 1967–2006

Year	CPS-Pre TU (Unadjusted)				Piketty-Saez		
	Wage	Entrepreneurial	Asset		Wage	Entrepreneurial	Asset
1967	74.8	18.8	6.4		73.3	17.4	9.3
1968	76.8	16.8	6.3		73.7	17.2	9.1
1969	77.9	15.3	6.8		75.3	16.1	8.6
1970	78.9	14.7	6.4		77.1	14.1	8.7
1971	79.1	14.6	6.3		77.6	13.4	9.0
1972	77.8	16.1	6.1		76.4	14.6	9.0
1973	75.9	17.3	6.8		74.2	16.0	9.8
1974	78.1	15.3	6.6		74.3	15.5	10.3
1975	78.9	14.7	6.4		77.4	13.3	9.3
1976	79.4	13.7	6.9		77.9	12.7	9.4
1977	79.0	13.9	7.0		78.1	12.4	9.5
1978	77.8	15.4	6.8		78.0	12.6	9.4
1979	78.3	13.1	8.6		78.4	11.5	10.1
1980	80.6	11.2	8.2		79.7	8.5	11.9
1981	79.2	11.0	9.8		80.6	6.1	13.2
1982	79.5	10.6	10.0		81.2	5.4	13.5
1983	78.9	10.8	10.3		83.4	5.7	10.9
1984	78.9	10.3	10.8		81.8	6.3	11.9
1985	81.2	8.7	10.1		82.9	6.6	10.5
1986	81.1	9.9	9.0		83.3	7.3	9.4
1987	80.1	10.5	9.4		81.8	8.9	9.3
1988	80.5	10.6	8.9		80.3	10.4	9.3
1989	78.1	11.3	10.6		79.3	10.3	10.3
1990	78.9	10.3	10.8		80.5	9.8	9.7
1991	79.1	11.0	9.9		80.8	10.2	9.0
1992	82.3	8.7	9.0		82.6	10.5	6.9
1993	80.8	10.3	9.0		83.2	10.7	6.1
1994	81.5	9.4	9.1		82.9	10.8	6.3
1995	82.8	7.2	10.0		82.9	10.8	6.3
1996	81.9	7.8	10.3		82.4	11.1	6.5
1997	78.1	8.7	13.2		82.0	11.0	7.1
1998	78.7	8.2	13.1		82.2	11.4	6.5
1999	76.8	10.5	12.7		82.2	11.3	6.5
2000	81.5	8.9	9.7		82.3	11.0	6.7
2001	83.5	7.5	9.0		83.2	10.9	5.9
2002	85.7	7.6	6.7		84.1	10.6	5.3
2003	84.1	7.8	8.2		84.4	10.6	5.0
2004	83.5	7.7	8.8		83.2	11.5	5.3
2005	83.5	7.3	9.3		81.5	12.6	5.9
2006	82.3	7.4	10.3		80.9	12.2	6.9

Sources: See note to Appendix Table B1.