

WHAT'S BEEN HAPPENING TO UNITED STATES INCOME INEQUALITY?

*Properly adjusted, for at least the bottom 99 percent of the income distribution,
the rise in income inequality since 1993 has been small*

BY RICHARD V. BURKHAUSER

A condensed and simplified version of the Downing Lecture delivered at the University of Melbourne on 23 October 2008. The full paper is published in Burkhauser, Richard V., Shuaizhang Feng, Jeff Larrimore and Stephen P. Jenkins , 2008, 'Trends in United States Income Inequality Using the Internal March Current Population Survey: The Importance of Controlling for Censoring.' *NBER Working Paper* w14247, August 2008.*

Introduction

The public-use version of the March Current Population Survey (CPS), an annual cross-sectional survey of over 50,000 American households, is the primary data source used by public policy researchers and administrators to investigate trends in average U.S. income and its distribution. Despite the widely held view that United States income inequality has increased substantially since the 1980s, our research, which derives from unprecedented access to internal CPS data, tells a very different story. Most of the evidence of a large increase in income inequality since 1993 has come either from those who do not adjust for topcoding in the public-use CPS or from Internal Revenue Service (IRS) administrative record files that have their own consistency

*Based on research conducted while Burkhauser, Feng and Larrimore were Special Sworn Status researchers of the U.S. Census Bureau at the New York Census Research Data Center at Cornell University. Conclusions expressed are those of the authors and do not necessarily reflect the views of the U.S. Census Bureau. This research has been screened to ensure that no confidential data are disclosed.

problems. In a practice called topcoding, for incomes above some value in the public-use CPS—the topcode threshold—the Census Bureau reports income as equal to this topcode threshold rather than providing the exact recorded value from the internal CPS.

Using various layers of CPS data (see Table 1 for a more precise definition of the various layers of CPS data we use) we show why not adjusting for topcoding in the public-use CPS data will falsely show that American income inequality has been rapidly growing. Once properly adjusted, we find that for at least the bottom 99 percent of the income distribution, the rise in income inequality since 1993 has been small and its yearly growth much slower than in the previous two decades. Our results hold even when we estimate income values for the very top part of the income distribution missing in the internal CPS data. Our findings are consistent with those found using IRS data on the 90th–99th percentile groups, only differing with respect to the top 1 percent of the income distribution. It is uncertain to what degree this difference is the result of our decreasing ability to capture income at the very highest income levels, even using internal CPS data, or of behavioral changes in the way that individual tax units report their adjusted gross income on their tax returns captured in the IRS data.

(Place Table 1 here)

Public-use vs. Internal CPS data.

For confidentiality reasons, the Census Bureau does not provide full information in the public-use CPS, which it makes available to the general public, on the amount of income from each income source – for example wages and salary, interest, dividends, etc. - found in the internal CPS. Beginning in 1995, the Census Bureau has also provided the mean of all topcoded values for these income sources. Using the mean value of all these topcoded values rather than the topcode threshold for that source of income captures more accurately, income values above the threshold from that source. But, as we will see, this leads the unwary to confuse an increase in our ability to measure higher income values with a real change in the income of richer people.

For their official work, Census Bureau researchers use the internal March CPS that is less severely censored. This is the data set we use. Doing so, we analyse levels and trends in US

inequality using Gini coefficients between 1975 and 2004 derived from the internal CPS and compare them with estimates from several series derived from the public-use CPS. (See Table 1). The Gini coefficient is the most common way to measure inequality in the distribution of income across a country's population. The Gini value must be between zero and one. Income that is perfectly distributed across the population, has a value of zero, while a perfectly unequal distribution—one person holds the country's entire income—has a value of one.

Measuring trends in income inequality.

Figure 1 shows that those who simply use the unadjusted public-use CPS (Public-Unadjusted) will find that income inequality jumps dramatically between 1994-1995—the Gini value increases from 0.395 to 0.422, a single year change far greater than in any prior or subsequent year. This is caused solely by an increase in topcoding and the use of a Census Bureau-derived mean value rather than the topcoded value for all values above the topcoded value. Using the unadjusted internal CPS (Internal-Unadjusted), we find no such increase between 1994-1995. Rather, what is happening is that prior to 1995 the Public-Unadjusted CPS Gini values substantially understate income inequality because they fail to fully account for income values above the topcodes. Once the Census Bureau provided the mean value of all these topcoded values, the now more precise Public-Unadjusted CPS Gini values match the higher Internal-Unadjusted Gini values. Failure to account for this change in methodology will grossly overstate US inequality increases before and after 1994-1995.

(Place Figure 1 here.)

This problem is not solved by simply ignoring Census Bureau mean values after 1994 as can be seen in the Public-NoMean Gini series. This series still inconsistently topcodes high values and underestimates inequality after 1994, as can be seen by the way its Gini values fall further and further below the Internal-Unadjusted Gini values. We solve this problem by deriving a mean value for all topcoded incomes in the public-use CPS, for each year back to 1975. When we use the public-use CPS together with our extended mean series (Public-Mean) in Figure 1, we match the Internal-Unadjusted Gini values in every year.

However, the internal CPS data is itself censored albeit to a substantially smaller extent than the public-use CPS. Hence, it too has time-inconsistencies, especially in 1992-1993 as can be seen by the jump in the Internal-Unadjusted Gini values between these years. To both control for inconsistent censoring and capture the missing part of the internal CPS data, we use a multiple imputation approach in which, for each year, out-of-sample values, that is values that are top-coded in the internal CPS are imputed based on assumptions about what their distribution would look like and with data from the lower in-sample values that we do have in the internal CPS data. Unsurprisingly, as can also be seen in Figure 1, we find that compared to estimates derived from our multiple imputation approach (Internal-MI), that contains these higher imputed values, all the other series understate the level of inequality in all years.

However, just as was the case for our Public-Mean series and the Internal-Unadjusted series it replicated, the Internal-MI series reveal the same trends: an increase in inequality over the entire period 1975–2004, but with a **rate** of increase noticeably lower after 1993 compared to before 1993. In each series, average inequality is found to increase much more prior to 1992 than after 1993. And in each series the jump in 1992-1993 is far higher than in any other period and consistent with the argument that a change in the measurement of inequality rather than a real change in inequality is its cause.

CPS vs. IRS results for the share of income held by high income groups.

Finally, we compare our CPS-based estimates of trends in top income shares using our Internal-MI adjusted CPS data to the estimates of top income shares derived by Piketty and Saez (2003) from IRS administrative files. As can be seen in Figure 2 both our (the red lines) and their (the blue lines) estimates of the income share held by the top 90th–95th percentile group have a relatively flat trend during the period 1975–2004, although the Piketty and Saez values are slightly higher in level. Similarly, for the shares held by the top 95th–99th percentile group, despite slight differences in levels, the two series exhibit remarkably similar trends over the 30 year period. In contrast, while the share of income held by the richest 1 percent (which can be found by taking the difference between the two top blue line for them and the top two red lines for us) increased substantially over this period according to both our CPS-based and the IRS-

based Piketty and Saez (2003) estimates), their estimates are much larger—from 8.0 to 16.1 percentage points compared to an increase from 5.4 to 9.8 percentage points in our series.

(Place Figure 2 here.)

But this difference is even greater when it is observed that one-third of the increase in the income share of the top 1 percent in our series from 1975–2004 occurred in 1993, and is primarily attributable to the CPS redesign. Hence a significant minority of this increase in the share of income held by the top 1 percent in our series is likely due to better measurement of their income by the CPS rather than by an actual increase in the share of income they held.

Effect of change in legislation

This same problem of changes in measurement versus changes in real income held by the richest 1 percent of the distribution is also likely to explain at least some part of the rise in income shares at the top of the income distribution in the Piketty and Saez (2003) series. As can also be seen in Figure 2, there is a dramatic 4 percentage point jump in the share of gross taxable income held by the highest 1 percent of tax units reported by Piketty and Saez (2003) between the years 1986 and 1988. (Compare the change in the difference between the top two blue lines over these years to confirm this). This finding must, to some degree, be the result of changes in the way the very richest tax units chose to report their income as a result of the Tax Reform Act of 1986 rather than the result of genuine changes in income inequality. This legislation provided substantial incentives for the very richest tax units to switch income from Subchapter-S corporations to wage income. Subtracting the 1986–1988 jumps in their series would cut in half the increase in the share of income held by the richest 1 percent over the whole period. Their data for other years may also be subject to this same type of change in tax reporting behavior, albeit to a lesser extent.

A further puzzle

Keeping this in mind, it is difficult to determine what is going on at the very top of the income distribution using either the CPS data or the IRS data. Piketty and Saez (2003) find greater increases in inequality after 1993 than we do primarily because of greater growth in the share of income held by the richest 1 percent in their IRS data. It is uncertain to what degree this difference is the result of our decreasing ability to capture income at the very highest income levels, even using internal CPS data, or of behavioral changes in the way that individual tax units report their adjusted gross income on their tax returns. The CPS may be less able to capture this income going to the top of the income distribution. But it also may be the case, as Reynolds (2006) argues, that a greater increase in the use of tax-deferred savings accounts - 401k plans, Keogh plans and IRA tax shelters - by those in richer percentile groups but not in the very richest 1 percent of the adjusted gross income distribution of tax units may also explain part of the rise in the top income share reported by Piketty and Saez (2003).

Concluding observations

Our results suggest that, for at least the poorest 99 percent of the income distribution, the increase in US income inequality since 1993 is significantly slower than in the previous two decades. Based on our Internal-MI series we find the level of income inequality rises when we include an estimate that includes the very top part of the income distribution censored in the unadjusted internal CPS data. But even in these estimates, the rise in income inequality slowed after 1993. Our findings are consistent with those found by Piketty and Saez (2003) for the 90th – 99th percentile groups. It is only with respect to the richest 1 percent that we differ. And it is here that we are at the limits of current knowledge, both with respect to the CPS because of its difficulty in obtaining information on the highest income households, and with respect to the IRS data because of behavioral effects caused by changes in the tax laws. It is difficult to fully understand how much of the yearly changes in inequality are the result of real changes in the incomes of the very richest income tax units and how much is due simply to changes in the way they report that income.

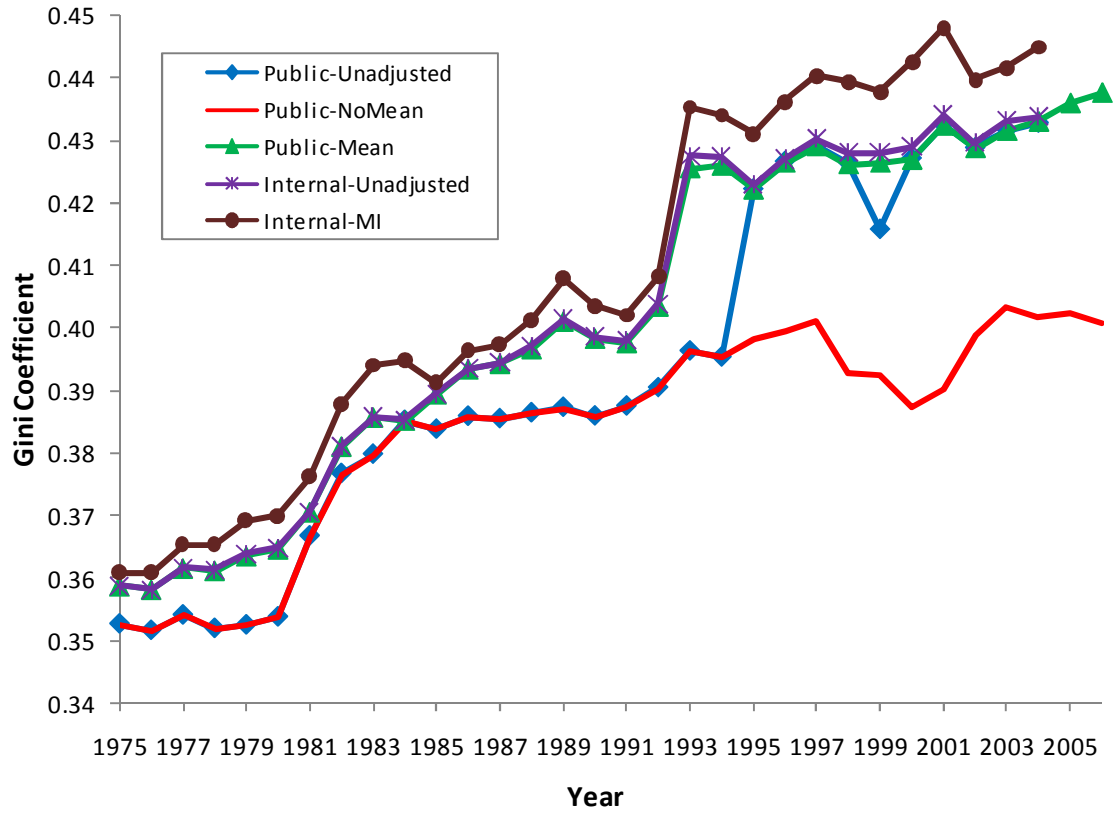
Professor Burkhauser is Sarah Gibson Blanding Professor of Public Policy in the Department of Policy Analysis and Management, Professor of Economics in the Department of Economics, Cornell University, and 2008 R I Downing Fellow at the University of Melbourne.

References

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

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

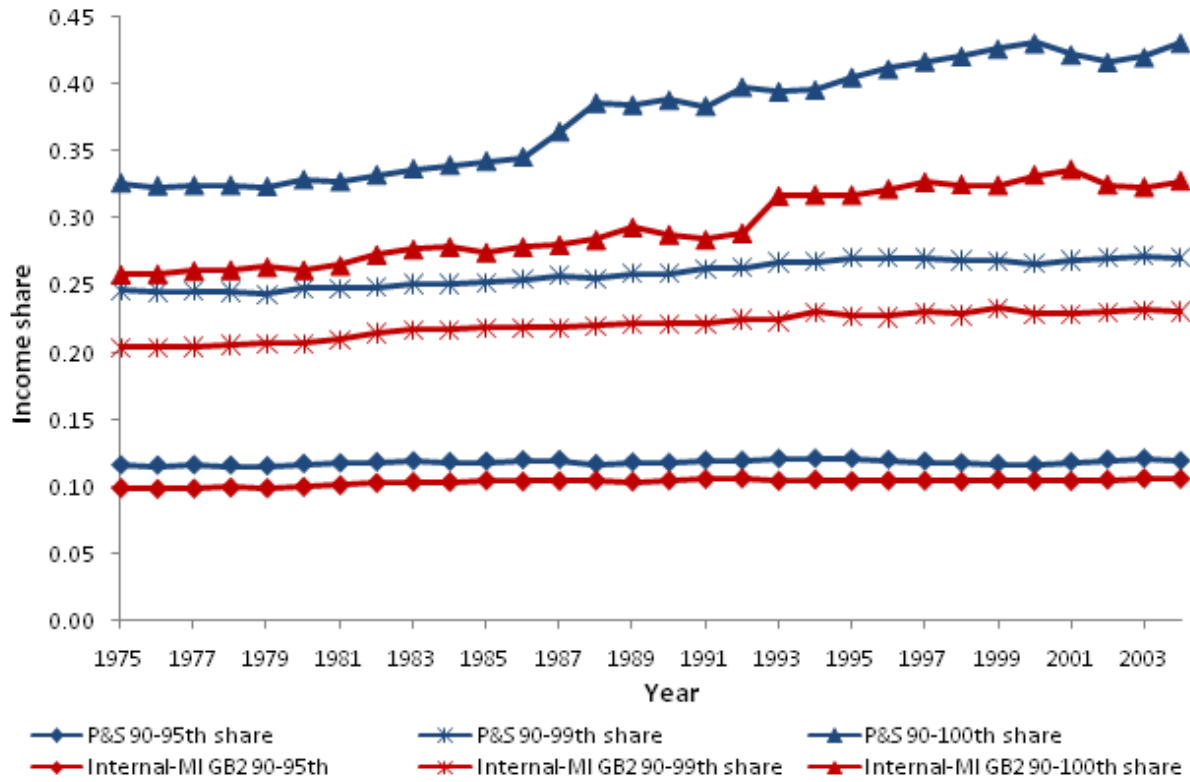
Figure 1. Inequality Estimates using Alternative Layers of CPS data, 1975–2006



Internal data were not available for years after 2005.

Source: Burkhauser, Feng, Jenkins and Larrimore (2008).

Figure 2. Share of Income Held by Top 10 Percent: IRS vs. CPS, 1975–2004



Source: <http://elsa.berkeley.edu/~saez/> and Burkhauser, Feng, Jenkins and Larrimore (2008).

Table 1: Definitions for Income Distribution Series by Source and Censoring Method

Acronym	Source	Method for Addressing Censoring Issues
Internal-Unadjusted	Internal	Uses internal data as provided in Census Bureau files, without any adjustments
Internal-MI	Internal	Topcoded observations replaced by imputations derived from GB2 imputation model fitted to internal data; inequality estimates derived using multiple imputation combination methods.
Public-Unadjusted	Public Use	Uses public use data as provided in Census Bureau files; includes Census Bureau cell mean imputations for topcoded observations from 1995 onwards
Public-Mean	Public Use	Uses public use data as provided in Census Bureau files; includes cell mean imputations for topcoded observations for all years
Public-NoMean	Public Use	Uses public use data as provided in Census Bureau files, except that no cell mean imputations used for any year (topcoded values used 'as is').

Source: Burkhauser, Feng, Jenkins and Larrimore (2008).

