Richard V. Burkhauser, Shuaizhang Feng, Stephen P. Jenkins and Jeff Larrimore

Estimating trends in US income inequality using the Current Population Survey: the importance of controlling for censoring

Article (Accepted version) (Refereed)

Original citation:
Burkhauser, Richard V., Feng, Shuaizhang, Jenkins, Stephen P. and Larrimore, Jeff (2010) *Estimating trends in US income inequality using the Current Population Survey: the importance of controlling for censoring.* The Journal of Economic Inequality, 9 (3). pp. 393-415 ISSN 1569-1721 DOI: 10.1007/s10888-010-9131-6

Estimating Trends in US Income Inequality Using the Current Population Survey: The Importance of Controlling for Censoring

Richard V. Burkhauser[1]
(Cornell University)

Shuaizhang Feng[2]
(Princeton University and Shanghai University of Finance and Economics)

Stephen Jenkins[3]
(University of Essex)

Jeff Larrimore[4]
(Cornell University)

[1] (*For correspondence and/or offprints*) 125 MVR Hall, Department of Policy Analysis & Management, Cornell University, Ithaca, NY 14853, USA; rvb1@cornell.edu
[2] 411A Robertson Hall, Princeton University, Princeton, NJ 08540, USA; shuaizhang.feng@gmail.com
[3] Institute for Social and Economic Research, University of Essex, Winhoe Park, Colchester CO4 3SQ, United Kingdom; stephenj@essex.ac.uk
[4] 404 Uris Hall, Department of Economics, Cornell University, Ithaca, NY 14853, USA; jhl42@cornell.edu

Estimating Trends in US Income Inequality Using the Current Population Survey: The Importance of Controlling for Censoring

Abstract

We analyze trends in US size-adjusted household income inequality between 1975 and 2004 using the most commonly used data source – the public use version of the March Current Population Survey. But, unlike most researchers, we also give substantial attention to the problems caused by the topcoding of each income source in the CPS data. Exploiting our access to Census Bureau internal CPS data, we examine estimates from data incorporating imputations for topcoded incomes derived from cell means and estimates from data multiply-imputed from parametric distribution models. Our analysis yields robust conclusions about inequality trends. The upward trend in US income inequality that began in the mid-1970s and increased in the 1980s slowed markedly after 1993.

Key Words: Censoring, CPS, Time Trend, Topcoding, US Income Inequality

JEL Classifications: D31, C81

1. Introduction

The public use files of the March Current Population Survey (CPS) are the primary data sources used for investigating trends in the distributions of individual earnings and of household income in the USA, as well as for undertaking cross-national comparisons involving the USA. The consensus of this research is that earnings and household income inequality increased substantially during the 1970s and 1980s. This growth in household income inequality outpaced that seen in most other OECD countries and by the early 1990s the USA had the highest level of inequality among 19 OECD countries in the Luxemburg Income Study (Gottschalk and Smeeding [32]).[5]

More recently, researchers using public use CPS data have found that the growth in wage inequality slowed in the 1990s (Card and Dinardo [21], Lemieux [44], and Autor, Katz, and Kearney [8]). However, in contrast to the findings for wage inequality, there are conflicting views about whether inequality in income continues to increase in the USA. Gottschalk and Danziger [33], using public use CPS data and measuring inequality by the ratio of the 90th percentile to the 10th percentile (the P90/P10 ratio), find that the rise in income inequality among their sample of working age adults slowed in the 1990s, which is similar to the pattern they find for wage inequality. However, others argue that the growth in US income inequality has continued unabated. Most prominent among them, Piketty and Saez, using data from US Internal Revenue Service (IRS) administrative records, report sustained growth in the income share of the richest 10% of tax filers throughout the 1980s to

[5] For systematic reviews of the cross-national inequality literature, see Atkinson, Rainwater and Smeeding [7], Gottschalk and Smeeding [32], and Atkinson and Brandolini [3] regarding household income inequality and Gottschalk and Smeeding [32] and Acemoglu [1] regarding earnings inequality. For examples of the use of the public use CPS in measuring inequality trends in the USA, see Burkhauser, Couch, Houtenville and Rovba [15], and Gottschalk and Danziger [33] regarding household income inequality, and Katz and Murphy [40]; Juhn, Murphy, and Pierce [38], and Card and Dinardo [21] regarding earnings inequality.

1998 ([49], Figure 1).[6] Since the P90/P10 ratio by definition does not capture changes in the upper tail of an income distribution, the results of Piketty and Saez are not necessarily inconsistent with the findings of Gottschalk and Danziger if, for instance, growth in inequality slowed in the middle of the distribution but continued to rise in the upper tail. But if it is the case that most of the changes in US incomes have been within the upper tail, then it suggests that public use CPS data may not be of much value in capturing inequality change in the distribution, especially if inequality is measured using the P90/P10 ratio.

One reason Gottschalk and Danziger and other researchers use P90/P10 ratios to measure inequality rather than measures that capture dispersion throughout the entire distribution (like the Gini coefficient) is that topcoding in public use CPS data has made it difficult to consistently observe changes at the top of the distribution.[7]

Piketty and Saez's approach avoids this limitation of public use CPS data by using IRS data on 'adjusted gross income' made available to the research community in grouped form. But these data also have limitations. Because the share of Americans paying personal income taxes has varied greatly over time, researchers like Piketty and Saez usually only look at the share of taxable income accruing to the top income earning groups and do not consider trends in inequality measured over all income groups.[8] In addition, critics of the use of these data for measuring trends in income inequality point out that tax filers are sensitive to changes in personal income tax rate. So tax-rate changes since the 1970s which have

[6] Updated tables from this paper are available on Saez's website [55]. The updated data show that with the exception of a decline in inequality from 2000–2002, this inequality growth has continued through 2006, their last available year of data.

[7] Burkhauser, Feng, and Jenkins [16] show that even P90/P10 ratios do not completely overcome the problem of topcoding in the CPS data because CPS topcoding is done for each of its sources of income. They find a significant minority of persons below the 90th percentile of the household income distribution have some source of their own income or of another household member's income that is topcoded.

[8] Since the overrepresentation of top income earners in the tax-filing population is not unique to the USA, top income shares are the standard approach used in international tax-records based analyses of inequality as well. See, for example, Piketty [48] for France, Atkinson [2] for the UK, Saez and Vaell [56] for Canada, Bach, Corneo and Steiner [9] for Germany, Dell [26] for Germany and Switzerland, and Atkinson and Leigh [4] for Australia. Atkinson and Piketty [5], Leigh [43], and Atkinson, Piketty, and Saez [6] provide comprehensive reviews of the literature using tax records to measure long-term inequality trends.

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. This shift would then overstate the actual rise in inequality among the very rich.[9] For a flavour of the heated debates on this topic, see the blog postings by leading economists and others on the Economists View website [28].

In this paper, we address the topic of US income inequality trends in a different and distinctive way. While we use public use CPS data as many previous researchers have, we also employ the more complete internal Census Bureau versions of the CPS data that provide considerably more information about the incomes of those at the top of the income distribution. With these better data, we are able to confront the problems of topcoding in the public use CPS data more directly.[10] Access to internal data also means that we are able to use broad-based measures of inequality that are sensitive to changes at the top of the income distribution (unlike percentile ratios such as P90/P10). So, for the first time, we are able to measure long term inequality trends for the complete distribution of income in the USA, accounting for the experience of both individuals in the upper tail of the distribution and those farther down.

Topcoding arises because the US Census Bureau is required to protect the confidentiality of CPS respondents. For each source of income received by an individual above a source-specific amount (the 'topcode'), the public use data made available to the research community contain the topcode rather than the reported income value. However, official statistics reported by the US Census Bureau [59] on mean income and income

[9] See Slemrod [58] and Reynolds [53] for a fuller discussion of the 'tax elasticity' issue in the Piketty and Saez [49] data, and Feenberg and Poterba [29] for a more general discussion of the problems of measuring income inequality using income tax records.

[10] Another issue bedeviling analysis of trends is the change in the CPS data collection design, especially the change to computer-assisted personal interviewing between calendar years 1993 and 1994. We refer to these later.

inequality are based on *internal* data from the March CPS that are much less severely censored. We use these internal data too.

Given that it is the predominant dataset for United States income statistics, this paper and much of the earlier research on topcoding focuses on the March CPS. However, issues of confidentiality make topcoding of income a necessity for most national datasets both in the US and internationally and researchers using these data employ similar techniques to those used by those using March CPS data.[11] For instance, in the USA, the National Longitudinal Survey of Youth (NLSY97) topcodes some of its income sources as does the Panel Study of Income Dynamics (PSID), the Survey of Income and Program Participation (SIPP), and the American Community Survey (ACS). In Great Britain, in order to comply with the 2007 Statistics and Registration Services Act, the Annual Population Survey and the Quarterly Labour Force Survey have introduced topcodes on earnings data in their main public release files. In numerous countries including Germany, Austria, and the United States, the wage data that are available from social insurance or social security administrative registers are right censored at the earnings level corresponding to the upper limit to social insurance contributions.[12]

We analyze trends in the inequality of size-adjusted pre-tax post-transfer household income in the USA using the March CPS between 1975 and 2004, comparing the estimates derived from three data series and both public use and internal CPS data. Our first series, which we label *Internal-Unadjusted*, is based on the internal data used 'as is', i.e. using income data among which the prevalence of topcoding is substantially lower than for public

[11] For example Gottschalk and Smeeding [32] consistently topcode cross-national data in the Luxemburg Income Study at 10 times the median reported income for each country, Pischke [50] sets topcoded incomes in the SIPP to the topcode threshold, and Lubotsky [46] sets topcoded incomes in US Social Security Earnings Records to 1.38 times the topcode threshold.

[12] Other datasets, such as the International Society Survey Program collect income information in income bands rather than as a continuous variable. This poses a similar challenge to topcoding of a continuous income variable when assigning income to the open-ended top income band. Similar techniques to those used for correcting for topcoding, such as assigning a multiple of the top-band's threshold, are often used when handling data of this structure. See e.g. Kahn [39].

use data. The second series, labeled *Public-CellMean*, uses cell means from the internal data to impute values for topcoded incomes in the public use data. That is, using internal data, we have derived a cell mean series in which topcoded incomes in the public use data, for each year back to 1975, are replaced by the mean of all income observed above the topcode for any income source of any individual in the public use data that has been topcoded.[13] This cell mean series can be used in conjunction with cell means provided by the Census Bureau for 1995 onwards to create a set of cell means for topcoded observations that covers the whole period. See Larrimore, Burkhauser, Feng, and Zayatz [41] for these cell mean values and for a detailed discussion of the procedures used to create them.

The estimates of mean income and of income inequality for the US private household population that we derive using our Public-CellMean series closely match the corresponding estimates that we derive from internal data.[14] If the internal data contained complete information about the topcoded incomes, then the cell mean approach would almost completely mitigate the biases in estimates that are due to topcoding. However, the internal data are also censored both out of concern about the accuracy of some high income responses and to reduce volatility in income statistics resulting from random sampling error among these top earners. While censoring in the internal data reduces the volatility of income statistics, this censoring is problematic since it causes the internal CPS data to understate the true level of income at the top of the US income distribution. Additionally, since increases in the internal censoring thresholds have not been performed systematically, the variation in the fraction of the population censored over time may also bias the trends in income held by

[13] Each CPS survey measures income from the previous calendar year. In this paper, all references are to the income year, so when we discuss the year 1975, this refers to income from various sources that members of the household received in 1975 reported at the March 1976 Current Population Survey interview.

[14] In 2006, we were granted permission to use the internal March CPS to test the sensitivity of measured income inequality in the public use CPS. We were given access to internal March CPS records from 1975–2004. These data include information on income above the public use topcode thresholds for respondents in the March CPS survey up to the *processing limit* in the March CPS data. The internal March CPS data the Census Bureau uses to produce the statistics in its official Census publications are subject to these same processing limits. However these processing limits are lower than the *data collection limits* for income in the March CPS, which are the limits on the actual values collected in the March CPS.

these individuals. Consequently, the internal censoring that exists in the internal March CPS data used for official Census Bureau estimates and for our *Internal-Unadjusted* series is expected to understate the level and misstate the trends in US income inequality.

Our third data series, labeled *Internal-MI*, accounts for this issue of censoring in the internal data. We use a multiple imputation approach in which, for each year, values for censored observations in the internal data are multiply imputed using draws from a Generalized Beta distribution of the Second Kind (GB2) fitted to internal data. Using this Internal-MI series, we investigate the extent to which the inconsistent exclusion of the top part of the distribution affects estimates of the level and trends in inequality. We find that, for every year, inequality estimates from the Internal-Unadjusted series are lower than the corresponding estimates derived from the Internal-MI series. However, the two series reveal the same trends: the upward trend in US income inequality that began in the mid-1970s slowed down markedly after 1993.

In the next section, we explain the nature of censoring in the CPS public use and internal data in greater detail, and explain the derivation of the Public-CellMean, Internal-Unadjusted and Internal-MI series. At the heart of the paper, Section 3, are comparisons of estimates of income inequality derived from these series. Section 4 presents a summary and conclusions.

Throughout the paper, income is defined in a conventional way. Income is pre-tax post-cash-transfer household income excluding capital gains, adjusted for differences in household size using the square root of household size.[15] Each individual is attributed with the size-adjusted income of the household to which he or she belongs. Income refers to

[15] We follow the same procedure for generating size-adjusted household income as discussed in Burkhauser and Larrimore [19]. These procedures, particularly the use of the square root of household members to adjust for household size are standard in the international comparative income distribution literature: see for instance Atkinson, Rainwater and Smeeding [7]. Its use is increasing in studies of US income distribution trends. See Burkhauser, Couch, Houtenville and Rovba [15], Gottschalk and Danziger [33], and Burkhauser, Osaki, and Rovba [20].

income for the calendar year preceding the March interview. We convert the small number of negative and zero household income values each year to one dollar prior to our calculations because a number of inequality indices are defined only for positive income values. Our samples for each year are all individuals in CPS respondent households, excluding individuals in group quarters or in households containing a member of the military. All statistics are calculated using the relevant CPS sampling weights.

2. Censoring in the March CPS

2.1 Topcoding in CPS public use data

In the March CPS, a respondent in each household is asked a series of questions on the sources of income for the household. Starting in 1975, respondents reported income from 11 sources and since 1987 they have done so for income from 24 sources. These income sources go well beyond labor market earnings and include numerous non-labor earnings sources such as interest, dividends, social security income, and unemployment payments. See Appendix Table 1 for a list of these income sources. Rather than simply topcoding high total household income values in the public use data, the US Census Bureau topcodes the high values of each source of household income. See Larrimore, Burkhauser, Feng and Zyatz [41] for a full list of topcode thresholds used over time in the public use data. Prior to 1995, the Census Bureau assigned the topcode threshold from that source of income to all topcoded income values in the public use data. Since 1995, the Census Bureau has substituted a cell mean value derived from the internal data to each topcoded value in the public use data. An additional complication is that household income is the aggregation of multiple income sources (income types and household members), each of which may be topcoded. As a result, the prevalence of topcoding in household income is significantly greater than for any particular income source, and topcoded household income values are not necessarily the highest incomes – they may occur throughout the income distribution.

Because the Census Bureau cell mean series starts in 1995, using the public use data without correcting for this major change in the values imputed to the highest incomes, source by source, results in a significant increase in measured income and income inequality in 1995 and subsequently simply because of the closer correspondence between the true reported income and the value included in the public use file. Hence, not taking this improvement in measurement into account overestimates how much actual income increased in 1995 among those at the highest income levels and also overestimates the level of income inequality.

The major change in the public use data in 1995 is a specific example of the more general problem that income topcoding presents. Topcodes in the public use data have increased in a non-systematic manner over time, and part of the apparent trend in average income and income dispersion in uncorrected data is caused by topcoding capturing a larger portion of the income distribution.[16]

Despite the Census Bureau's attempt to alleviate the problem of topcoding, their cell means have generally been ignored by researchers studying US income distribution trends for periods spanning the 1990s since to do otherwise would exacerbate time-inconsistencies that arise from using unadjusted public use data for the pre-1995 period and CPS data with cell mean imputations thereafter.

Researchers have typically followed two strategies to deal with the CPS topcoding issues. First, they have focused on the P90/P10 ratio as the summary measure of inequality, in the belief that this largely insulates them from topcoding problems.[17] Burkhauser, Feng, and

[16] See Levy and Murnane [45] for an early review of the income distribution literature and a more formal statement of this problem.

[17] Gottschalk and Danziger [33], Gottschalk and Smeeding [32], and Daly and Valletta [24], for example, use P90/P10 ratios to measure US income inequality. Juhn, Murphy, and Pierce [38], Dinardo, Fortin, and Lemieux [27], Lemieux [44], and Autor, Katz, and Kearney [8] use P90/P10 ratios to examine US labor earnings inequality.

Jenkins [16] show that using the P90/P10 ratio does not completely alleviate CPS topcoding problems, since incomes are topcoded by income source (not by total income) and so there are topcoded incomes below the 90th percentile of size-adjusted household income. Additionally, the P90/P10 ratio is based on just two income values, and so captures different aspects of the income distribution than widely-used measures of dispersion such as the Gini coefficient or other indices summarizing income differences throughout the income distribution.

Second, instead or as well, researchers have used 'consistent topcoding' methods, i.e. they have ascertained the quantile of the income distribution at which topcoding bites in the data for each of the years being studied, and then applied the minimum of these topcodes consistently across all the years.[18] Consistent top coding is useful because it is consistent by construction, but it is limited because it also implies that a substantial range of the income distribution is simply ignored, and arguably it is at the top of the distribution where many of the recent changes have been occurring (cf. Piketty and Saez [49]).[19] In this paper, we eschew use of consistent topcoding and the P90/P10 ratio in favour of methods that take greater account of top incomes (exploiting access to CPS internal data) and use a range of distributionally-sensitive inequality indices (the Gini coefficient and three members of the Generalized Entropy class of indices).

[18] Gottschalk and Danziger [33] apply consistent topcoding methods. See also Burkhauser, Butler, Feng, and Houtenville [14], Feng, Burkhauser, and Butler [30], and Burkhauser, Oshio, and Rovba [20].

[19] Until recently, consistent topcoding was the best method available for dealing with topcoding, as it allowed researchers to analyze a consistent fraction of the population. However, two recent events have led consistent topcoding to become a less useful correction method (which is why we do not report our findings based on its application in this paper). The first event is the rise in the fraction of the population that is topcoded on any source of household income, which reached almost 6% in 2006. As a result of these increases in topcoding, consistent topcoding means that researchers consider a decreasing fraction of the population. Second, and more importantly, is the recent availability of a superior alternative – cell means for topcoded incomes back to 1975 provided in Larrimore, Burkhauser, Feng and Zayatz [41]. However, while the cell-mean series allows researchers to successfully mimic results from the internal data, for those interested in what is happening for the entire distribution, including those with incomes greater than the internal data censoring points, it is necessary to use more sophisticated techniques such as those described in this paper.

2.2 Censoring in the internal CPS data

Using public use data with cell mean imputations (the Public-CellMean series) cannot solve the problem of censoring entirely, because the internal data from which the cell means are derived are themselves censored, albeit to a lesser degree than the public use data. As was the case in public use data prior to 1995, any censored observations in the internal data are reported as having source income equal to the censoring threshold.

This censoring in the internal data originated due to limits on the number of digits provided for responses on the survey questionnaire and early electronic data records. Until 1985, responses for each income source were limited to just 4 or 5 digits. However, after the Census Bureau processing systems were upgraded in 1985 to allow for responses with more digits, this censoring persisted with two justifications. The first is a concern about the accuracy of these high responses and the possibility for a recording error to have a substantial impact on inequality statistics. The second justification is that the small number of very high incomes may lead to wide swings in income statistics from random sampling error depending on which high earners are selected for the survey in any given year. Thus, to limit the effect of high income outliers on inequality statistics, the Census Bureau has maintained some degree of internal censoring in all years of the survey.[20] See Larrimore, Burkhauser, Feng and Zayatz [41] for a full list of internal censoring thresholds over time and see Ryscavage [54], Jones and Weinberg [37], and Burkhauser, Butler, Feng and Houtenville [14], for earlier discussions of the problems of censoring in the public use and internal data.

The trend in the percentage of individuals with censored size-adjusted household income in the public use and internal data is provided in Figure 1. Although the percentage of individuals with topcoded income in any given year is generally substantially lower in the

[20] Since internal censoring impacts not just research based on the internal CPS data but also that based on the public-use data using cell-means, the Census Bureau could alleviate many of the censoring concerns without compromising respondent confidentiality by providing a cell-mean series using the uncensored *collection limits data* to supplement the currently produced cell-means based off of the *processing limits data*.

internal data than in the public use data, the number of people in the internal data affected by censoring is non-negligible. The fraction is typically around 0.5% of the population, and has been rising slowly since 1985.

<Figure 1 near here>

Because we had access to the internal CPS data, we are able for the first time to correct for censoring in these data back to 1975, and hence provide a more consistent measure of trends in the income distribution and compare these trends with those found in either the public use or the unadjusted internal data.

2.3 A Multiple Imputation Approach

To correct for censoring and also capture income dispersion across the whole distribution, we develop a procedure to impute values to incomes that are censored in the internal CPS data. This yields the Internal-MI series.

We use a parametric model of the income distribution since, by definition, the 'true' uncensored underlying income in the internal data are not available to us. Unlike most previous methods used to estimate topcoded incomes which led to a single imputation for each topcoded value, we derive multiple imputed values using a suitable randomization procedure which leads to multiple distributions for each year, and account for potential stochastic imputation error by averaging estimates (hence the label 'multiple imputation').

The parametric model of the income distribution used for the imputation model is the Generalized Beta of the Second Kind (GB2). The GB2 distribution has a probability density function

$$f(y) = \frac{ay^{ap-1}}{b^{ap}B(p,q)\left[1+(y/b)^{a}\right]^{p+q}}, y > 0 \tag{1}$$

and cumulative density function (CDF)

$$F(y) = I(\, p, q, (y/b)^{a} \,/\, [1 + (y/b)^{a}]\,), y > 0, \tag{2}$$

where parameters *a*, *b*, *p*, *q*, are each positive. $B(p, q) = \Gamma(p)\Gamma(q)/\Gamma(p + q)$ is the Beta function, $\Gamma(.)$ is the Gamma function, and $I(p, q, x)$ is the regularized incomplete beta function also known as the incomplete beta ratio. Parameter *b* is a scale parameter, and *a*, *p*, and *q* are each shape parameters. The GB2 distribution is a flexible form widely used in the income distribution literature, and a number of studies have shown that it fits income distributions extremely well across different times and countries: see e.g. Bordley, McDonald and Mantrala [12], Brachmann, Stich and Trede [13], Bandourian, McDonald, and Turley [10], and Jenkins [34]. Feng, Burkhauser, and Butler [30] fitted GB2 distributions to labor earnings using public use CPS data and argue that their Gini coefficients calculated from the fitted GB2 distributions provide more plausible estimates of inequality trends than do the Gini coefficients derived from unadjusted internal data reported by the Census Bureau. The flexible nature of the GB2 distribution also incorporates many other distributions as special cases. For example, the Singh-Maddala (Burr type 12) distribution is the special case of the GB2 distribution when $p = 1$; the Dagum (Burr type 3) distribution is the special case when $q = 1$; and the lognormal distribution is a limiting case. For details, see McDonald [47] and Kleiber and Kotz [41]. In addition, the GB2 becomes Pareto-like as income becomes increasingly large (Shluter and and Trede [57] 164, fn. 11).

Our multiple imputation approach differs from earlier studies that used parametric models to derive imputed values for topcoded incomes: see Fichtenbaum and Shahidi [31] and Bishop, Chiou, and Formby [11].[21] First, those authors fitted a one-parameter Pareto distribution to the upper ranges of the income distribution, rather than the more flexible four-parameter GB2 model. Second, our approach takes account of the fact that, rather than being common to all individuals, censoring levels vary across individuals (because censoring is

[21] Piketty and Saez [49] also use Pareto distributions to perform parametric interpolation in the upper tail of the distribution, although they do so to estimate the average income level of each of their income groups rather than estimating the share of income allocated to topcoded individuals.

done at the level of each income source rather than at the level of aggregate household income). Third, by using multiple imputation methods, rather than a single imputation, we account for the variability intrinsic in the imputation process.

Our multiple imputation approach involves five steps. First, for each year's data, we fit a GB2 distribution by maximum likelihood, accounting for individual level right-censoring. In fitting the GB2 distribution, we model individual household size-adjusted income at the aggregate level, not for each income source separately. Doing so the other way would require modeling the joint distribution of 24 income sources (11 before 1987) for various members of each household and, as a result, our approach is more parsimonious and tractable. 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 (using appropriate corrections for left truncation in the ML procedure).[22] We specify the sample log-likelihood for each year's data as

$$\ln L = \sum_{i=1}^{N} w_i \left\{ \frac{c_i \ln[1 - F(y_i)] + (1 - c_i) \ln[f(y_i)]}{1 - F(z)} \right\} \tag{3}$$

where $i = 1,\ldots, N$, indexes each individual sample observation, w_i is the sample weight for i, and $c_i = 1$ if i is an observation with a right censored household income value, and $c_i = 0$ otherwise. The denominator of the expression adjusts for left truncation: z is the income level corresponding to the left truncation point.

Second, for each observation with a censored income, we draw a value from the income distribution that is implied by the fitted GB2 distribution. Given the fitted GB2 CDF, $\hat{F}(y)$, the corresponding CDF for each topcoded observation i is, using standard formulae for truncated distributions:

$$\hat{G}(y_i) = [\hat{F}(y_i) - \hat{F}(t_i)]/[1 - \hat{F}(t_i)] \tag{4}$$

[22] The 30th percentile was chosen as the left truncation point after experiments balancing goodness of fit with ease of maximization.

where t_i is the topcode for i, and y_i is the 'true' value for that observation (which we are unable to observe). Letting $u_i = \hat{G}(y_i)$, and inverting the expression for the income distribution among topcoded observations, we have

$$y_i = \hat{F}^{-1}\left(u_i[1 - \hat{F}(t_i)] + \hat{F}(t_i)\right). \tag{5}$$

A value of y_i for each topcoded observation is generated by substituting into this expression a value of u_i equal to a random draw from a standard uniform distribution. Since uncensored observations are observed, this imputation is only used to estimate censored observations and the actual values of uncensored observations are used directly in all calculations.

Third, using the distribution comprising imputations for censored observations and observed incomes for non-censored observations, we estimate our various inequality indices. Fourth, we repeated steps 2 and 3 one hundred times, and finally, we combine the one hundred sets of estimates from each of the one hundred data sets for each year using the 'averaging' rules proposed by Rubin [52] and modified by Reiter [51] for the case of partially synthetic data. See Jenkins, Burkhauser, Feng and Larrimore [36] for further details of our multiple imputation approach to estimation and inference.[23]

This paper performs the multiple imputation procedure only on internal March CPS data, but a similar approach could be used to impute all topcoded incomes in the public use data. While, as expected, the GB2 imputation on public use data generally yields higher levels of inequality than that seen using unadjusted internal data, it finds somewhat lower levels of inequality than that seen from using the GB2 imputation on internal data. This difference occurs because the substantially greater information on top incomes provided by the internal data improves the fit of the GB2 imputation and thus yields improved estimates of inequality including the top range of the income distribution. See Jenkins, Burkhauser,

[23] A referee has suggested that an alternative approach to inference could be based on non-standard bootstrap resampling, in which bootstrap samples are drawn from the semiparametric estimate of each income distribution (one combining a parametric estimate of the upper tail of topcoded observations with a non-parametric estimate for the rest of the distribution). A related approach is discussed by Davidson and Flachaire [24].

Feng, and Larrimore [36] for a comparison of the Internal-MI estimates presented here with estimates derived from the same multiple imputation approach applied to public-use data from 1995–2004.

3. Estimates of inequality levels and trends

3.1 The revised cell mean series (Public-CellMean)

Figure 2 reports our estimates of income inequality levels and trends derived from the Public-CellMean and Internal-Unadjusted series that we have described so far. To highlight the consequences of ignoring the additional information about top incomes that is incorporated in these series, we compare them with two other series. The *Public-Unadjusted* series is based on unadjusted public use data which include Census Bureau cell mean imputations for topcoded observations from 1995 onwards (it is derived from the data that is available to researchers in the public use files nowadays). The *Public-NoCellMean* series is the same as the Public-Unadjusted one, except that all cell mean information is discarded. Inequality is summarized using the Gini coefficient. Because public use data are topcoded, mean household income and Gini coefficients derived from them are understated relative to those estimated from internal data unless cell mean imputations are used, since cell means provide more information about top incomes.

<Figure 2 near here>

In every year, the Gini estimates based on the Public-NoCellMean series are below those from the Internal-Unadjusted series. To correct for this difference (and for several other reasons), the Census Bureau revised its topcoding procedures in income year 1995. The result is reflected in the difference between the Gini estimates derived from the Public-NoCellMean and Public-Unadjusted series. Prior to 1995, the series are identical. After 1995, the Public-

Unadjusted Gini estimates are very similar to the Internal-Unadjusted estimates (with the exception of 1999), and are substantially greater than those derived from the Public-NoCellMean series. (We believe that the 1999 result reflects an error in the Census Bureau cell mean series.) Although the Gini estimates from the Public-Unadjusted series now better represent the Gini estimates derived from the Internal-Unadjusted data, anyone uncritically using Gini values from the Public-Unadjusted data to describe changes in income inequality in the United States would greatly exaggerate its increase in 1995.

Our Public-CellMean series corrects this problem. Figure 2 shows that the Gini estimates derived from these data match those derived from the Internal-Unadjusted data in all years including 1999. Thus the addition of our cell mean imputations to the CPS public use data allows researchers without access to the Internal-Unadjusted data to produce virtually the same inequality levels and trends found in the Internal-Unadjusted data.

We would also draw attention to the substantial jump in inequality between 1992 and 1993 according to the Public Cell-Mean and Internal-Unadjusted series. However, much of this jump is simply due to changes in CPS data collection methods that were implemented by the Census Bureau at that time. (See Ryscavage [54] and Jones and Weinberg [37].) The vertical dotted line between 1992 and 1993 in Figure 2 (and Figure 3 below) is included in order to draw attention to this 'structural break' in all our series.

3.2 Long term trends in US income inequality

Given how closely our Public-CellMean series matches the levels and trends in inequality in the internal CPS data, using cell means would almost completely mitigate the problem of topcoding were it not for internal censoring. To address the extent of internal censoring on measures of inequality, we now extend our analysis of long term trends by comparing estimates of inequality derived from the Internal-Unadjusted and the Internal-MI series over the period 1975–2004. We do not include the other public use series in our analysis going forward since each of these series are inferior to the Public-CellMean series in their ability to track inequality for the entire distribution including those with topcoded incomes. We stop at 2004 because that is the last year of internal CPS data to which we had access. Our discussion is in terms of the Gini coefficient initially, but then we discuss trends in Generalized Entropy inequality indices.[24]

Figure 3, panel (a), depicts the trends over the period 1975–2004 in the Gini coefficient derived from the Internal-Unadjusted and the Internal-MI series. According to the former, there was a dramatic increase in inequality between 1975 and 1992, with an acceleration of this increase starting in 1980. There was then a substantial jump in inequality between 1992 and 1993 due to changes in CPS data collection methods (see above). From 1993 onwards, while inequality continued to rise, the rate of increase is considerably slower than for the pre-1993 period.

Since the Internal-MI series captures the very highest incomes, it leads to higher measured inequality levels than the Internal-Unadjusted counterpart in all years. But the trends in inequality represented by the series are similar. So, even though the Internal-Unadjusted series underestimates the level of inequality in any given year, it is accurately capturing the inequality trends.

The Gini coefficient rose between 1992 and 1993 more sharply for the Internal-MI series than for the Internal-Unadjusted ones. This is additional evidence that the 1993 jump in inequality is primarily due to other survey design changes rather than to problems related to censoring. That the jump is larger for the Internal-MI series also indicates that the survey-design changes may be disproportionately impacting the reporting of incomes at the top of

[24] For a discussion of robustness of inequality results to alternate inequality indices, see Jenkins et al. [36] which tracks recent trends in Lorenz curves for the US income distribution. For a comparison of inequality trends using Gini coefficients to those based on the income share of the top percentiles of the distribution, see Burkhauser et al. [18].

the distribution. Such an observation is consistent with the fact that the increase in inequality between 1992 and 1993 using public use CPS data – in which the topcodes are lower – do not show similar increases between those two years (see Figure 2).

3.3 Do conclusions differ if different measures of inequality are used?

The Gini coefficient is just one of many commonly-used indices of inequality. It incorporates particular assumptions about how income differences are aggregated at different parts of the income distribution: it is relatively sensitive to income differences in the middle of the distribution ('middle sensitive'). This raises the question of whether alternative inequality measures, ones incorporating different aggregation assumptions, might lead to different conclusions about inequality trends, depending, for example, on whether changes in dispersion occur mainly at the top or the bottom of the distribution. We explore the sensitivity of our results to the choice of inequality measure by repeating our analysis using three indices from the Generalized Entropy parametric family.[25] We employ the three most commonly used Generalized Entropy (GE) measures: $I(0)$, the mean logarithmic deviation (MLD), which is a relatively bottom-sensitive index; $I(1)$, the Theil index, which is relatively middle-sensitive; and $I(2)$, half the squared coefficient of variation, which is an index that is top-sensitive.

We calculated each of these GE indices using the Internal-Unadjusted and Internal-MI series, and graph the estimates in Figures 3, panels (b), (c) and (d) for $I(0)$, $I(1)$ and $I(2)$ respectively.[26] In Table 1 we provide the average annual percentage change in inequality, by subperiod, for each of these inequality indices as well as the Gini coefficient, using both our data series. Since the trends in inequality are markedly different in the periods before and

[25] See e.g. Cowell [22] or Jenkins and Van Kerm [35] for detailed discussion of these indices.
[26] Trends in the GE indices using the *Public-CellMean* and *Public-Unadjusted* data are available in the extended working-paper version of this article (Burkhauser et al. [17])

after 1980 and 1993, we divide the table into four sub-periods: 1975–1980, 1980–1992, 1992–1993, and 1993–2004.

<Table 1 near here>

In all cases, measured inequality in a given year is greatest using the Internal-MI series, indicating that the Internal-Unadjusted series understate the true level of inequality. Additionally, for each index (with the exception of $I(2)$ perhaps), the trends in inequality using both our data series are quite similar. However, there are differences in the observed trends in inequality depending on which inequality index is chosen.

The results for $I(0)$ clearly differ from those based on the Gini coefficient: compare panels (a) and (b) of Figure 3. First, with $I(0)$, the difference in inequality levels between the two data series is smaller than the corresponding difference for the Gini (panel a). This is unsurprising because $I(0)$ is a relatively bottom-sensitive index, and the differences between these two different data series arise at the top of the distribution. Additionally, there was a marked inequality growth pause in the middle of the 1980s. Since $I(0)$ is a bottom-sensitive inequality measure, this suggests that the most substantial changes in the income distribution were occurring towards the top of the distribution rather than towards the bottom during this period in the 1980s. Additionally, $I(0)$ showed less signs of slowing down in the 1990s than our other series, suggesting that the slowdown in inequality in the 1990s was driven by changes near the top of the distribution.

Inequality trends according to the $I(1)$ index are similar to those using the Gini coefficient: compare panels (c) and (a) in Figure 3. For both series, $I(1)$ shows a rapid increase in inequality from 1980 through 1992, with a slower increase in inequality starting in 1993.

According to $I(2)$, the difference between the Internal-MI and Internal-Unadjusted series is greater during the post-1993 period than for the other three indices: compare panel (d) with the other panels in Figure 3. Since the $I(2)$ is a top-sensitive inequality index, these patterns simply reflect the fact that topcoding mainly affects incomes near the top of the distribution. However, the $I(2)$ estimates also exhibit greater year-on-year variability in the post-1993 period than do the other three indices. These results may arise because $I(2)$ estimates tend to have larger sampling variability relative to these other indices, other things being equal.[27] Thus, more caution is warranted when discussing inequality trends using this index.

4. Conclusions

We analyze trends in income inequality in the USA over three decades (1975–2004), drawing on internal CPS data that are not generally available to researchers outside the Census Bureau. Although the CPS is the most authoritative survey data source for studying US income inequality, the topcoding of each of its many sources of income artificially lowers the level of inequality and potentially affects estimates of inequality trends, whether one uses public or internal CPS data. This has necessitated that researchers interested in trends in income inequality either use CPS data and focus only on inequality in lower portions of the distribution, or use other data sources such as IRS administrative records that allow one to capture inequality trends near the top of the distribution but not to observe trends at lower incomes.

We consider several methods for addressing censoring and, as a byproduct of this work, derive a revised cell mean imputation series that can be applied to public use data. This series allows researchers to more accurately capture levels and trends in the internal data. In addition, we take account of censoring in the internal data using a multiple imputation

[27] $I(2)$ is also relatively prone to non-robustness to the effects of outliers in the sense discussed by Cowell and Victoria-Feser [23]. However, we would stress the role of sampling variability, since the 'averaging' process used to combine the estimates from our 100 multiply imputed data sets is likely to smooth out the effects of any outliers being added by the imputation process.

approach that uses parametric methods to derive imputations for topcoded incomes and allows us to observe inequality trends for the entire US income distribution.

When the Gini coefficient is used to measure inequality, we find the level of income inequality according to our Internal-MI series is slightly higher in corresponding years than inequality according to the Internal-Unadjusted series before 1993 and significantly higher after 1993. But despite this difference in levels, the *trends* in inequality we find using the two series are not significantly different. Inequality rose over the whole period 1975–2004, but the rate of increase slowed after 1993.

The story about trends differs when we use the three Generalized Entropy inequality indices, but the differences in trends are readily explicable in terms of the properties of the indices, specifically their differences in sensitivity to income differences in different ranges of the income distribution. $I(0)$ rose steadily during 1975–2004 (apart from a pause during the mid-1980s), showing no sign of slowing down after 1993. But our $I(1)$ and $I(2)$ series are similar to our Gini series in their trends, subject to the caveat that $I(2)$ indices are less precisely estimated and show larger year-on-year variation. Because $I(0)$ is more bottom-sensitive than the other inequality measures, our results suggest that what is happening in the upper part of the income distribution is responsible for the 1990s slowdown in rate of inequality increase.

Our results for household income are largely consistent with other research that has used public use CPS data to examine levels and trends in wage inequality (e.g. Autor, Katz and Kearney [8]) or income inequality for lower portions of the income distribution (e.g. Gottschalk and Danziger [33]) but differ from the results for income inequality found by Piketty and Saez [49] using IRS administrative record data. The differences between our results and those of Piketty and Saez may be partially due to differences in the income units (since Piketty and Saez consider distributions of income among tax units while we explore household income among individuals), as well as differences in the types of income reported in the two datasets. A valuable avenue for future research would be to explore the differences between the IRS and CPS datasets to more fully understand the differences in these results, as we have begun to do in our current research (see Burkhauser et al. [18]).

Acknowledgements:

The research in this paper was 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. Any opinions and conclusions expressed herein are those of the author and do not necessarily represent the views of the U.S. Census Bureau. All results have been reviewed to ensure that no confidential information is disclosed. Supports for this research from the National Science Foundation (award nos. SES-0427889 SES-0322902, and SES-0339191) and the National Institute for Disability and Rehabilitation Research (H133B040013 and H133B031111) are cordially acknowledged. Jenkins's research was supported by core funding from the University of Essex and the UK Economic and Social Research Council for the Research Centre on Micro-Social Change and the United Kingdom Longitudinal Studies Centre. We thank Ian Schmutte, Arnie Reznek, Laura Zayatz, the Cornell Census RDC Administrators, and all their U.S. Census Bureau colleagues who have helped with this project. This paper has also benefited from comments made by Peter Gottschalk, Bruce Meyer, and other participants at the Empirical Economics Workshop of Shanghai University of Finance and Economics, Shanghai, May 2008.

References

1. Acemoglu, D.: Cross-country Inequality Trends. Econ. J. **113**, F121–F149 (2003)
2. Atkinson, A.B.: Top Incomes in the UK Over the Twentieth Century. J. Roy. Statistical Society, **168,** 325–343 (2005).
3. Atkinson, A.B., Brandolini, A.: Promises and Pitfalls in the Use of Secondary Data Sets: Income Inequality in OECD Countries as a Case Study. J. Econ. Lit. **39(3)**, 771–799 (2001)
4. Atkinson, A.B., Leigh, A.: The Distribution of Top Incomes in Australia, Econ. Rec. **83**, 262 (2007), 247–61.
5. Atkinson, A.B., Piketty, T.: Top Incomes over the Twentieth Century: A Contrast Between Continental European and English Speaking Countries (Oxford: Oxford University Press, 2007).
6. Atkinson, A.B., Piketty, T., Saez, E.: Top Incomes in the Long Run of History. NBER Working Paper w15408, http://www.nber.org/papers/w15408 Cited 13 Dec 2009 (2009)
7. Atkinson, A.B., Rainwater, L., Smeeding, T.: Income Distribution in OECD Countries. Evidence from the Luxembourg Income Study. Social Policy Studies No. 18. Organization for Economic Cooperation and Development, Paris (1995)
8. Autor, D., Katz, L., Kearney, M.: Trends in U.S. Wage Inequality: Revising the Revisionists. Rev. Econ. Statist. **90(2),** 300–323 (2008)
9. Bach, S., Corneo, G., Steiner, V.: From Bottom to Top: The Entire Income Distribution In Germany, 1992-2003. Rev. Income Wealth **55(2), 303-330** (2009),
10. Bandourian, R., McDonald J.B., Turley, R.S.: A Comparison of Parametric Models of Income Distribution across Countries and Over Time. Estadistica **55,** 135–152 (2003)

11. Bishop, J.A., Chiou, J., and Formby, J.P.: Truncation Bias and the Ordinal Evaluation of Income Inequality. J. Bus. Econ. Statist. **12**, 123–127 (1994)

12. Bordley, R.F., McDonald, J.B., Mantrala, A.: Something New, Something Old: Parametric Models for the Size Distribution of Income. J. Income Distribution **6**, 91–103 (1996)

13. Brachmann, K., Stich, A., Trede, M.. Evaluating Parametric Income Distribution Models. Allgemeines Statistisches Arch. **80**, 285–298 (1996)

14. Burkhauser, R.V., Butler, J.S., Feng, S., and Houtenville, A.: Long-Term Trends in Earnings Inequality: What the CPS Can Tell Us. Econ. Letters **82**, 295–299 (2004)

15. Burkhauser, R.V., Couch, K.A., Houtenville, A.J., Rovba, L.: Income Inequality in the 1990s: Re-Forging a Lost Relationship? J. Income Distribution **12(3–4)**, 8–35 (2003-2004)

16. Burkhauser, R.V., Feng, S., Jenkins, S.P.: Using the P90/P10 Ratio to Measure US Inequality Trends with Current Population Survey Data: A View from Inside the Census Bureau Vaults. Rev. Income Wealth **55(1)**, 166–185 (2009)

17. Burkhauser, R.V., Feng, S., Jenkins, S.P., Larrimore, J.: Estimating Trends in US Income Inequality Using the Current Population Survey: The Importance of Controlling for Censoring. NBER Working paper 14247, http://www.nber.org/papers/w14247 Cited 13 Dec 2009 (2008)

18. Burkhauser, R.V., Feng, S., Jenkins, S.P., Larrimore, J.: Recent Trends in Top Income Shares in the USA: Reconciling Estimates from March CPS and IRS Tax Return Data. NBER Working Paper w15320, http://www.nber.org/papers/w15320 Cited 13 Dec 2009 (2009)

19. Burkhauser, R.V., Larrimore, J.: Trends in the Relative Economic Well Being of Working-Age Men with Disabilities: Correcting the Record using Internal Current Population Survey Data. J. Disabil. Pol. Stud. **20(3)**, 162-169 (2009)

20. Burkhauser, R.V., Oshio, T., Rovba, L.: How the Distribution of After-Tax Income Changed Over the 1990s Business Cycle: A Comparison of the United States, Great Britain, Germany, and Japan. J. Income Distribution. **17(1)**, 87–109 (2008)

21. Card, D., DiNardo, J.E.: Skill-Biased Technological Change and Rising Wage Inequality: Some Problems and Puzzles. J. Lab. Econ. **20(4)**, 733–783 (2002)

22. Cowell, F.A.: Measurement of Inequality. In: Anthony B. Atkinson and François Bourguignon (eds), Handbook of Income Distribution, pp. 87–166. Elsevier North Holland, Amsterdam (2000)

23. Cowell, F.A., Victoria-Feser, M.: Robustness Properties of Inequality measures. Econometrica. **64**, 77–101 (1996)

24. Daly, M.C., Valletta, R.G.: Inequality and Poverty in the United States: The Effect of Rising Dispersion of Men's Earnings and Changing Family Behavior, Economica. **73(289)**, 75–98 (2006)

25. Davidson, R., Flachaire, E.: Asymptotic and Bootstrap Inference for Inequality and Poverty Measures, Journal of Econometrics **141**, 141-166 (2007)

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

27. DiNardo, J., Fortin, N., and Lemieux, T.: Labor Market Institutions and the Distribution of Wages, 1973–1992: A Semi-Parametric Approach, Econometrica. **64(5)**, 1001–1044 (1996)

28. Economists View:

http://economistsview.typepad.com/economistsview/2007/01/increasing_ineq.html. Cited 24 Feb 2009 (2007)

29. Feenberg, D., Poterba, J.: Income Inequality and the Incomes of Very High-Income Taxpayers: Evidence from Tax Returns. In: Poterba, J.M. (ed.) Tax Policy and the Economy, 7, pp. 145-177. NBER/MIT Press, Cambridge, MA (1993)

30. Feng, S., Burkhauser, R. V., Butler, J.S.: Levels and Long-Term Trends in Earnings Inequality: Overcoming Current Population Survey Censoring Problems Using the GB2 Distribution. J. Bus. Econ. Statist. **24(1)**, 57–62 (2006)

31. Fichtenbaum, R., Shahidi, H.: Truncation Bias and the Measurement of Income Inequality. J. Bus. Econ. Statist. **6**, 335–337 (1988)

32. Gottschalk, P., Smeeding, T.M.: Cross-National Comparisons of Earnings and Income Inequality. J. Econ. Lit. **35(2)**, 633–687 (1997)

33. Gottschalk, P., Danziger,S.: Inequality of Wage Rates, Earnings and Family Income in the United States, 1975–2002. Rev. Income Wealth. **51(2)**, 231–254 (2005)

34. Jenkins, S.P.: Distributionally-Sensitive Inequality Indices and the GB2 Income Distribution. Rev. Income Wealth **55(2),** 392–398 (2009)

35. Jenkins, S.P., and Van Kerm, P.: The Measurement of Economic Inequality. In: W. Salverda, B. Nolan and Timothy M. Smeeding (eds), The Oxford Handbook on Economic Inequality. Oxford University Press, Oxford (2009)

36. Jenkins, S.P., Burkhauser, R.V., Feng, S., Larrimore, J.: Measuring Inequality with Censored Data: A Multiple Imputation Approach. ISER Working Paper 2009-04. Colchester, UK: Institute for Social and Economic Research, University of Essex. http://www.iser.essex.ac.uk/publications/working-papers/iser/2009-04. Cited 24 Feb 2009 (2009)

37. Jones, A.F., Weinberg, D.H.: The Changing Shape of the Nation's Income Distribution. Current Population Reports, U.S. Census Bureau, June 2000. http://www.census.gov/prod/2000pubs/p60-204.pdf. Cited 24 Feb 2009 (2000)

38. Juhn, C., Murphy, K.M., Pierce, B.: Wage Inequality and the Rise in Returns to Skill, J. Polit. Economy **101(3)**. 410–442. (1993)

39. Kahn, L.M.: Collective Bargaining and the Interindustry Wage Structure: International Evidence. Economica **65(260)**, 507-534

40. Katz, L.F., and Murphy, K.M.: Changes In Relative Wages, 1963-1987: Supply and Demand Factors. Quart. J. Econ. **107(1)**, 35–78 (1992)

41. Kleiber, C., Kotz, S.: Statistical Size Distributions in Economics and Actuarial Sciences. Hoboken, NJ: John Wiley (2003)

42. Larrimore, J., Burkhauser, R.V., Feng, S., Zayatz, L.: Consistent Cell Means for Topcoded Incomes in the Public Use March CPS (1976–2007). J. Econ. Soc. Meas. **33(2-3)**, 89–128 (2008)

43. Leigh, A.: How Closely do Top Income Shares Track Other Measures of Inequality? Econ. J. **117(524)**, F619–F633 (2007)

44. Lemieux, T.: Increasing Residual Wage Inequality: Composition Effects, Noisy Data, or Rising Demand for Skill? Amer. Econ. Rev. **96(3)**, 461–498 (2006)

45. Levy, F., Murnane, R.J.: U.S. Earnings Levels and Earnings Inequality: A Review of Recent Trends and Proposed Explanations. J. Econ. Lit. **30(3)**, 1333–1381 (1992)

46. Lubotsky, D.: Chutes or Ladders? A Longitudinal Analysis of Immigrant Earnings. J. Polit. Economy **115(5)**, 820-866 (2007)

47. McDonald, J.B.: Some generalized functions for the size distribution of income. Econometrica **52**, 647–663 (1984)

48. Piketty, T.: Income Inequality in France, 1901–1998. J. Polit. Economy **111(5)**, 1004–1042 (2003)

49. Piketty, T., Saez, E.: Income Inequality in the United States, 1913–1998. Quart. J. Econ. **118(1)**, 1–39 (2003)

50. Pischke, J.S.: Individual Income, Incomplete Information, and Aggregate Consumption. Econometrica **63**, 805-840 (1995)

51. Reiter, J.P.: Inference for Partially Synthetic, Public Use Microdata Sets. Surv. Methodology **29**, 181–188 (2003)

52. Rubin, D.B.: Multiple Imputation for Nonresponse in Surveys. Wiley, New York (1987)

53. Reynolds, A.: Income and Wealth. Greenwood Press, Westport, CT (2006)

54. Ryscavage P.: A Surge in Growing Income Inequality? Monthly Labor Review, August, 51–61. http://www.bls.gov/opub/mlr/1995/08/art5full.pdf (1995)

55. Saez, E.: http://elsa.berkeley.edu/~saez/. Cited 24 Feb 2009 (2009)

56. Saez, E., Veall, M.R.: The Evolution of High Incomes in Northern America: Lessons from Canadian Evidence. Amer. Econ. Rev. **95(3)**, 831–49 (2005)

57. Shluter, C., Trede, M.: Tails of Lorenz curves. *J. Econometrics* **109**, 151–166 (2002)

58. Slemrod, J.: High-income Families and the Tax Changes of the 1980s: The Anatomy of Behavioral Response. In: Feldstein, M., Poterba, J.M. (eds.) Empirical Foundations of Household Taxation, pp. 193–226. University of Chicago Press (for NBER), Chicago, (1996)

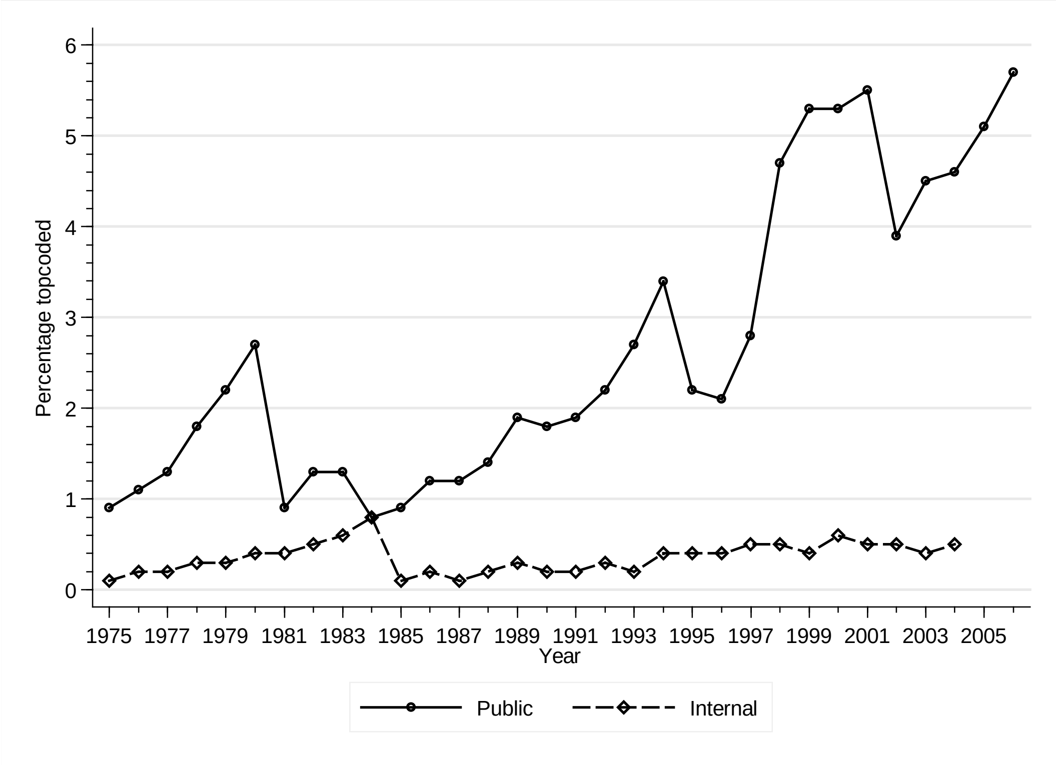
59. U.S. Census Bureau. Income, Poverty, and Health Insurance Coverage in the United States: various years. (August), Current Population Reports, Consumer Income. GPO, Washington DC. (Various years)

Table 1: Average Annual Percentage Change in Inequality by Subperiod, CPS Internal Data Series, and Inequality Index

	Average annual percentage change			
	1975–1980	1980–1992	1992–1993	1993–2004
Gini coefficient				
Internal-Unadjusted	0.33	0.89	5.85	0.14
Internal-MI	0.50	0.86	6.63	0.20
$I(0)$				
Internal-Unadjusted	1.39	1.42	12.48	1.20
Internal-MI	1.65	1.37	14.04	1.28
$I(1)$				
Internal-Unadjusted	0.36	2.05	22.08	0.26
Internal-MI	0.96	1.95	29.32	0.43
$I(2)$				
Internal-Unadjusted	–0.64	2.84	71.82	0.03
Internal-MI	0.40	2.90	147.17	0.45

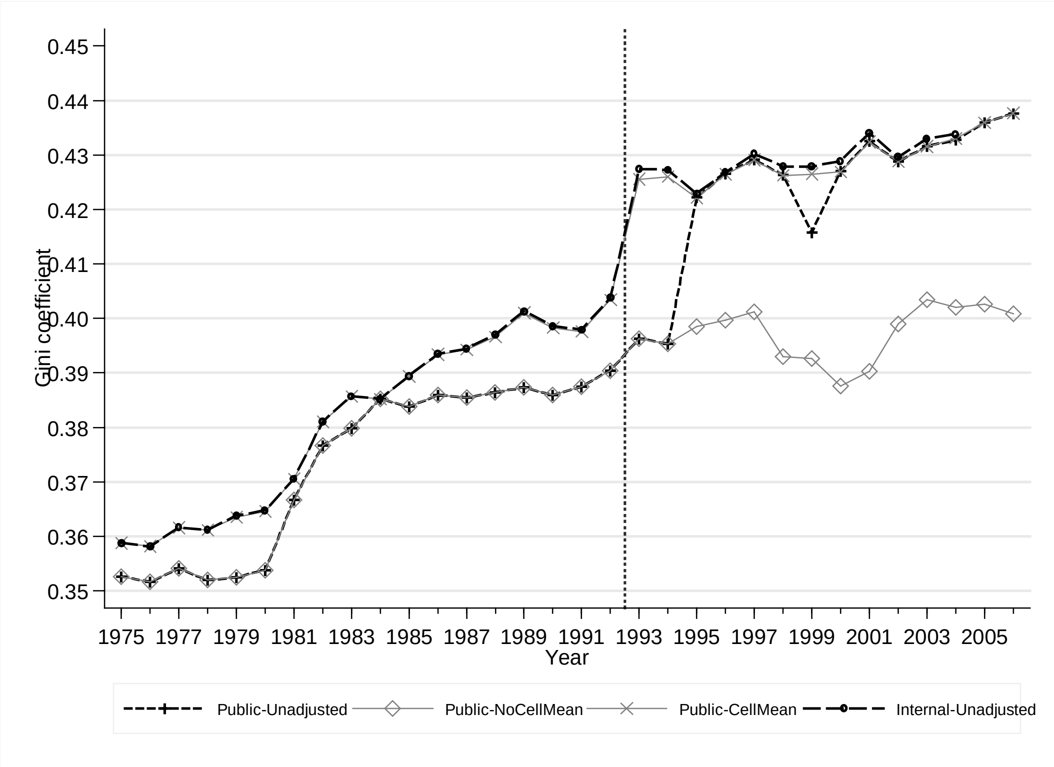
Source: Authors' calculations from internal and public use data files of the March CPS. $I(0)$, $I(1)$, and $I(2)$ are members of the Generalized Entropy class of inequality indices (see text). See text for definitions of the series.

Figure 1: Percentage of Individuals with Censored Household Income in March CPS, by Year



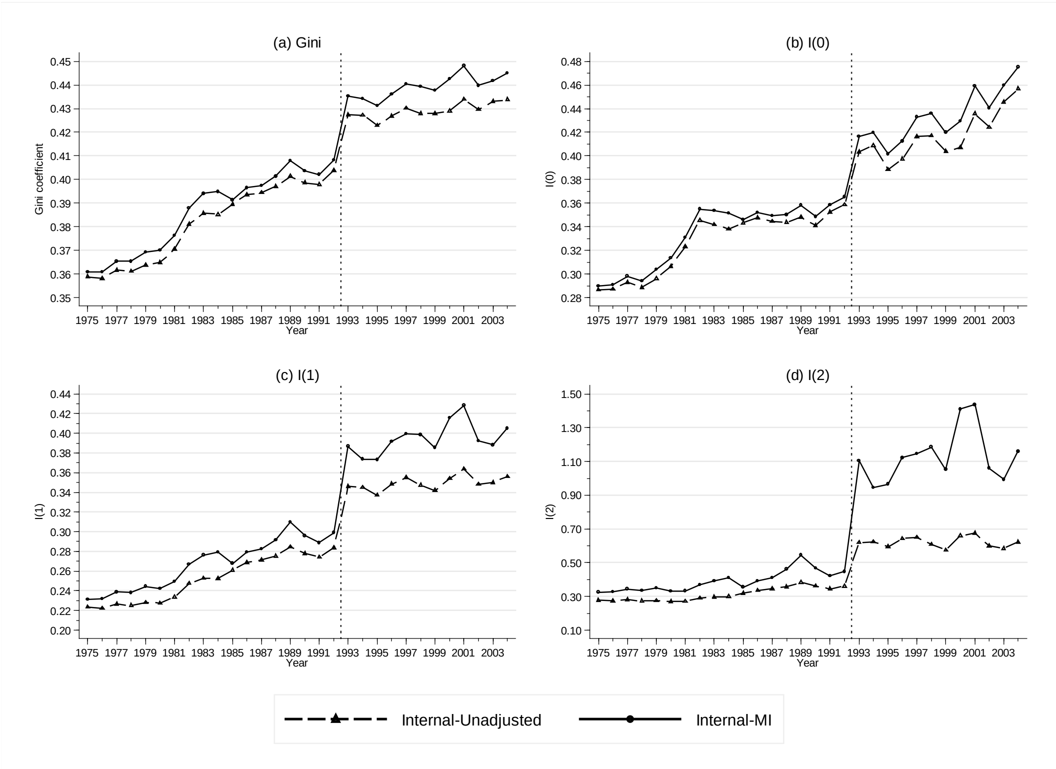
Source: authors' calculations from internal and public use data files of March CPS. Internal data were not available for years after 2004.

Figure 2: Gini Coefficient Estimates Derived Using Four Censoring Adjustment Methods



Source: authors' calculations from internal and public use data files of the March CPS. There was a major change in CPS data collection methods between 1992 and 1993. Internal data were not available for years after 2004. See text for definitions of the series.

Figure 3: Inequality Estimates Derived Using Two Internal Censoring Adjustment Methods, 1975–2004, by Index



Source: authors' calculations from internal and public use data files of the March CPS. There was a major change in CPS data collection methods between 1992 and 1993. $I(0)$, $I(1)$, and $I(2)$ are members of the Generalized Entropy class of inequality indices (see text). See text for definitions of the series.

Appendix Table 1. Income Items Reported in the Current Population Survey

Name	Name in Public Files	Name in Internal Files	Definition
		1975–1986	
Labor Earnings			
Wages	I51A	WSAL_VAL	Wages and Salaries
Self Employment	I51B	SEMP_VAL	Self employment income
Farm	I51C	FRSE_VAL	Farm income
Other Sources			
Social Security	I52A	I52A_VAL	Income from Social Security and/or Railroad Retirement
Supplemental Security	I52B	SSI_VAL	Supplemental Security Income
Public Assistance	I53A	PAW_VAL	Public Assistance
Interest	I53B	INT_VAL	Interest
Dividends Rentals	I53C	I53C_VAL	Dividends, Rentals, Trust Income
Veterans	I53D	I53D_VAL	Veteran's, unemployment, worker's compensation
Retirement	I53E	I53E_VAL	Pension Income
Other	I53F	I53F_VAL	Alimony, Child Support, Other income
		1987–2006	
Labor Earnings			
Primary earnings	ERN_VAL	ERN_VAL	Primary Earnings
Wages	WS_VAL	WS_VAL	Wages and Salaries-Second Source
Self Employment	SE_VAL	SE_VAL	Self employment income -Second Source
Farm	FRM_VAL	FRM_VAL	Farm income -Second Source
Other Sources			
Social Security	SS_VAL	SS_VAL	Social Security Income
Supplemental Security	SSI_VAL	SSI_VAL	Supplemental Security Income
Public Assistance	PAW_VAL	PAW_VAL	Public Assistance & Welfare Income
Interest	INT_VAL	INT_VAL	Interest
Dividends	DIV_VAL	DIV_VAL	Dividends
Rental	RNT_VAL	RNT_VAL	Rental income
Alimony	ALM_VAL	ALM_VAL	Alimony income
Child Support	CSP_VAL	CSP_VAL	Child Support Income
Unemployment	UC_VAL	UC_VAL	Unemployment income
Workers Comp	WC_VAL	WC_VAL	Worker's compensation income
Veterans	VET_VAL	VET_VAL	Veteran's Benefits
Retirement - Source 1	RET_VAL1	RET_VAL1	Retirement income - source 1
Retirement - Source 2	RET_VAL2	RET_VAL2	Retirement income - source 2
Survivors - Source 1	SUR_VAL1	SUR_VAL1	Survivor's income - source 1
Survivors - Source 2	SUR_VAL2	SUR_VAL2	Survivor's income - source 2
Disability - Source 1	DIS_VAL1	DIS_VAL1	Disability income - source 1
Disability - Source 2	DIS_VAL2	DIS_VAL2	Disability income - source 2
Education assistance	ED_VAL	ED_VAL	Education assistance
Financial assistance	FIN_VAL	FIN_VAL	Financial Assistance
Other	OI_VAL	OI_VAL	Other income

Sources: Current Population Survey Annual Demographic File Technical Documentation, 1976-2002. Current Population Survey Annual Social and Economic Supplement Technical Documentation, 2003-2007