USING THE P90/P10 INDEX TO MEASURE US INEQUALITY TRENDS WITH CURRENT POPULATION SURVEY DATA: A VIEW FROM INSIDE THE CENSUS BUREAU VAULTS

Richard V. Burkhauser
(Cornell University)

Shuaizhang Feng
(Shanghai University of Finance and Economics)

Stephen P. Jenkins
(University of Essex)

ISER Working Paper
2007-14

Acknowledgements:

The research in this paper was conducted while the first two authors were Special Sworn Status researchers of the U.S. Census Bureau at the New York Census Research Data Center at Cornell University. Research results and conclusions expressed are those of the authors and do not necessarily reflect the views of the U. S. Census Bureau. This paper has been screened to ensure that no confidential data are disclosed. Support for this research from the National Science Foundation (award nos. SES-0427889 SES-0322902, and SES-0339191), and from ISER's core funding from the UK Economic and Social Research Council and the University of Essex, is gratefully acknowledged. We thank Mathis Schroeder, Ludmila Rovba and Jeff Larrimore for comments on earlier versions of this paper and Pinky Chandra and Lisa Marie Dragoset, the Cornell Census RDC Administrators, and all their Bureau of Census colleagues who have helped this project along, especially Edward J. Welniak, Jr. and Brian P. Holly.

Readers wishing to cite this document are asked to use the following form of words:

> **Burkhauser, Richard V., Feng, Shuaizhang, and Jenkins, Stephen P. (May 2007) 'Using the P90/P10 Index to Measure US Inequality Trends with Current Population Survey Data: A View from Inside the Census Bureau Vaults', ISER Working Paper 2007-14. Colchester: University of Essex.**

The on-line version of this working paper can be found at http://www.iser.essex.ac.uk/pubs/workpaps/

The Institute for Social and Economic Research (ISER) specialises in the production and analysis of longitudinal data. ISER incorporates

- MISOC (the ESRC Research Centre on Micro-social Change), an international centre for research into the lifecourse, and

- ULSC (the ESRC UK Longitudinal Studies Centre), a national resource centre to promote longitudinal surveys and longitudinal research.

The support of both the Economic and Social Research Council (ESRC) and the University of Essex is gratefully acknowledged. The work reported in this paper is part of the scientific programme of the Institute for Social and Economic Research.

Institute for Social and Economic Research, University of Essex, Wivenhoe Park,
Colchester. Essex CO4 3SQ UK
Telephone: +44 (0) 1206 872957 Fax: +44 (0) 1206 873151 E-mail: iser@essex.ac.uk
Website: http://www.iser.essex.ac.uk

ABSTRACT

The March Current Population Survey (CPS) is the primary data source for estimation of levels and trends in labor earnings and income inequality in the USA. Time-inconsistency problems related to top coding in theses data have led many researchers to use the ratio of the 90th and 10th percentiles of these distributions (P90/P10) rather than a more traditional summary measure of inequality. With access to public use and restricted-access internal CPS data, and bounding methods, we show that using P90/P10 does not completely obviate time-inconsistency problems, especially for household income inequality trends. Using internal data, we create consistent cell mean values for all top-coded public use values that, when used with public use data, closely track inequality trends in labor earnings and household income using internal data. But estimates of longer-term inequality trends with these corrected data based on P90/P10 differ from those based on the Gini coefficient. The choice of inequality measure matters.

JEL codes: D3; J3; C8

Key words: inequality, income, earnings, Current Population Survey, decile ratio, Gini coefficient

NON-TECHNICAL SUMMARY

The vast majority of research on trends in labour earnings and income inequality since the 1970s in the USA has been based on public use files of the March Current Population Survey (CPS). In the cross-national comparative literature, CPS data are also commonly used to compare both labour earnings and income inequality levels and trends in the USA with other industrialized countries. The most important source of standardized cross-sectional micro data on industrialized countries—the Luxembourg Income Study (LIS)—uses the public use version of the CPS data for the USA. The public use CPS data are also a major source of information about US inequality in the World Income Inequality Database.

To protect the confidentiality of CPS respondents, top codes are imposed on every source of income above a specific value, with the top code value differing by income source. For example, if someone reports earning a million dollars, then the wage and salary data for that respondent that researchers see in the data set is not one million dollars, but a lower value (the top coded value). Household income data is more likely to be top-coded than wages and salary income because it is the aggregation across individuals of a large number of income sources, each of which may be top-coded and, if one source is top coded, then so is the aggregated income variable.

Top coded data cause problems for inequality analysis because they censor the range of incomes that are observed. Inequality is underestimated because very high incomes appear as less-high incomes. This problem would be less of an issue when one is looking at inequality trends over time if the nature and extent of top coding were constant. However, CPS top codes have changed over time in a number of ways, leading to a potentially serious time-inconsistency problem for inequality analysis.

This time-inconsistency has led many researchers to use a measure of inequality that they believe will insulate them from the problem. The measure is the ratio of the 90^{th} and 10^{th} percentile of a distribution (P90/P10). If you lined every one up in ascending order of income, the 10^{th} percentile would be the income of the person one tenth of the way along the parade, and the 90^{th} percentile would be the income of the person nine-tenths along. The greater the difference between these two incomes – the larger that P90/P10 is – the greater the degree of inequality. P90/P10 contrasts with other commonly-used summary measures of inequality such as the Gini coefficient, Theil index, or coefficient of variation, each of which uses information about all income values, rather than only two. In the US labour economics literature, P90/P10 is the most commonly used measure of wage or labor earnings dispersion.. In the US income inequality literature, the P90/P10 is also a standard measure of inequality in the distributions of size-adjusted family or household income.

Researchers have implicitly assumed that P90/P10 is not affected by censoring, reasoning that the fraction of observations affected by censoring of total wages and salaries, labour earnings or income is less than 10 percent. While this is true, in the CPS data, censoring takes place at the level of each income source not for income totals, so some values below the 90^{th} percentile of total labor earnings and especially the 90^{th} income percentile are censored. As a result, even what are apparently modest amounts of censoring in the population as a whole may affect estimates of P90/P10.

To address the issues raised by censoring requires use of internal March CPS data, and we have been able to gain access to them for the very first time for this purpose. Our analysis considers data for income years 1975–2004. We examine three distributions of income that are commonly assessed in the labour and income inequality literatures: (i) wages and salaries income among individuals working full-time full-year for wages; (ii) total earnings income among full-time, full-year workers (wage and salaries plus farm and non-farm self-employment earnings); and (iii) household income among all individuals.

Our paper makes three contributions. First, using innovative bounding methods, we show that calculating P90/P10 with public use CPS data—even when Census Bureau cell means are used for top coded values—does not completely obviate the problem of time-inconsistency, especially for those interested in trends in the inequality of individuals' size-adjusted household income. Second, we offer a means by which researchers may reduce problems caused by censoring. Because we have access to the internal CPS data, we have been able to create consistent cell mean values for all top-coded values in all years of internal data made available to us (1975–2004) that offer a plausible correction for time inconsistency problems in the public use CPS data when integrated with them.

Our third contribution concerns the assessment of longer-term US inequality trends. When we compare estimates of P90/P10 based on our adjusted public use CPS data with estimates of Gini coefficients based on either the internal or public use CPS data consistently top-coded to control for time inconsistencies, we find that the trends in P90/P10 differ significantly from the trends in either of the two Gini coefficient series. Hence, researchers should be cautious in making inference about trends in the inequality of the distributions of wages and salaries income, labour earnings income, or size-adjusted household income over the last three decades based on changes in the relative position of only two points in each of those distributions.

1. INTRODUCTION

The vast majority of research on trends in labor earnings and income inequality since the 1970s in the USA has been based on public use files of the March Current Population Survey (CPS). Yet time-inconsistency problems related to top coding in these data have led many researchers to use the ratio of the 90th and 10th percentile of a distribution (P90/P10) rather than a more traditional summary measure of inequality such as the Gini coefficient, Theil index, or coefficient of variation, each of which uses information about all income values, rather than only two. In the US labor economics literature, P90/P10 is the most commonly used measure of wage or labor earnings dispersion: see e.g. Juhn *et al.* (1993), Danziger and Gottschalk (1993), DiNardo *et al.* (1996), Gottschalk and Smeeding (1997), Gottschalk and Joyce (1998), Katz and Autor (1999), Autor *et al.* (2005), Blau and Kahn (2005), Lemieux (2006) and Pencavel (2006). In the US income inequality literature, the P90/P10 is also a standard measure of inequality in the distributions of size-adjusted family or household income: see e.g. Danziger and Gottschalk (1993), Gottschalk and Smeeding (1997), Gottschalk and Danziger (2005), and Daly and Valletta (2006).

In the cross-national comparative literature, CPS data are also commonly used to compare both labor earnings and income inequality levels and trends in the USA with other industrialized countries. See Smeeding (2004) for a review of literature using the CPS. Other recent examples include Nielsen *et al.* (2005), Prus and Brown (2006), Atkinson (2007), Burkhauser *et al.* (2007), and Brandolini (forthcoming). The most important source of standardized cross-sectional micro data on industrialized countries—the Luxembourg Income Study (LIS)—uses the public use version of the CPS data for the USA. On its website (http://www.lisproject.org/keyfigures/ineqtable.htm), LIS provides summarizes income inequality using P90/P10 and Gini coefficient estimates that do not adjust for the top coding issues discussed below. The public use CPS data are also a major source of information about US inequality in the World Income Inequality Database (WIDER, 2007).

Other things being equal, any of the traditional summary measures of inequality are likely to be better measures of inequality of the entire distribution, and hence of its trends over time, than P90/P10 which only captures two points in that distribution. But other things are often not equal. The public use March CPS is the best source of annual information on trends in the labor earnings and income of US households available to the research community. However, all sources of income in the public use CPS are top coded, which makes accurate calculations of traditional summary measures of the distribution impossible and comparisons of these values over time difficult (Levy and Murnane, 1992; Gottschalk and Smeeding, 1997). Moreover, even the internal CPS data, which are not subject to top coding, have been censored to various degrees over time (Welniak, 2003).

The impact of censoring on Gini coefficients estimated with both the public use and internal CPS data has been documented in previous research: see e.g. Burkhauser *et al.* (2004), and Feng *et al.* (2006). But no similar scrutiny has been given to the impact of censoring on quantile ratio measures such as P90/P10. Researchers have implicitly assumed that P90/P10 is not affected by censoring, reasoning that the fraction of observations affected by censoring of total wages and salaries, labor earnings or income is less than 10 percent. While this is true, in the CPS data, censoring takes place at the level of each income source not for income totals, so some values below the 90th percentile of total labor earnings and especially the 90th income percentile are censored. As a result, even what are apparently modest amounts of censoring in the population as a whole may affect estimates of P90/P10.

To address the issues raised by censoring requires use of internal March CPS data, and we have been able to gain access to them for the very first time for this purpose.[1] Our analysis considers data for income years 1975–2004. We examine three distributions of income that are commonly assessed in the labor and income inequality literatures: (i) wages and salaries income among individuals working full-time full-year for wages; (ii) total earnings income among full-time, full-year workers (wage and salaries plus farm and non-farm self-employment earnings); and (iii) household income among all individuals.

Our paper makes three contributions. First, using innovative bounding methods, we show that calculating P90/P10 with public use CPS data—even when Census Bureau cell means are used for top coded values—does not completely obviate the problem of time-inconsistency, especially for those interested in trends in the inequality of individuals' size-adjusted household income. Second, we offer a means by which researchers may reduce problems caused by censoring. Because we have access to the internal CPS data, we have been able to create consistent cell mean values for all top-coded values in all years of internal data made available to us (1975–2004) that offer a plausible correction for time inconsistency problems in the public use CPS data when integrated with them.

Our third contribution concerns the assessment of longer-term US inequality trends. When we compare estimates of P90/P10 based on our adjusted public use CPS data with estimates of Gini coefficients based on either the internal or public use CPS data consistently top-coded to control for time inconsistencies, we find that the trends in P90/P10 differ significantly from the trends in either of the two Gini coefficient series. Hence, researchers should be cautious in making inference about trends in the inequality of the distributions of wages and salaries income, labor earnings income, or size-adjusted household income over the last three decades based on changes in the relative position of only two points in each of those distributions.

2. Censoring problems in the Current Population Survey

The Current Population Survey (CPS), based on a large representative sample of the US population, interviews about 57,000 households each month. Each March, the CPS collects detailed information about each source of income in the previous year for every household member. To protect the confidentiality of respondents, top codes are imposed on all sources of income above a specific value. Less well known to the research community is the fact that even the internal data the Census Bureau uses to calculate various official statistics including inequality measures, are also subject to censoring. In earlier years this was primarily because of restrictions on computer tape space. Although such constraints are substantially relaxed nowadays, CPS internal income data are still censored for various Census Bureau considerations, including minimizing the possible impact of recording (keying) errors, helping to maintain respondents' confidentiality, and preventing volatility and distortion of annual statistics (Welniak, 2003, Feng *et al.* 2006).

The precise Census Bureau variable names, and definitions of the three sources of income that we analyze, and how they have changed over time, are shown in Table 1. For income years 1975–1986, the Census Bureau reported three sources of labor earnings and eight other sources of income. From 1987 onwards they have used a finer categorization, reporting four sources of labor earnings and twenty other sources of income. For all income components, both the internal and public use CPS censoring points have changed over time. Public use CPS censoring points for income years 1975–1986 are shown in Appendix Table 1 and for 1987–2004 in Appendix Table 2. Corresponding internal CPS censoring points for the two periods are provided in Appendix Tables 3 and 4.

<Table 1 near here>

Because censored values start at different points in the distribution each year, any inequality estimate not taking account of this variation is time-inconsistent. This includes estimates published by the Census Bureau using internal CPS data. Past researchers have recognized this problem and, for the most part, used some rule-of-thumb adjustment procedure to control for it: see e.g. Juhn *et al.* (1993) and Trejo (1997). More recently, Burkhauser *et al.* (2004) consistently top coded values at the same point in the distribution (the highest common point in the distribution available for all years) and estimated Gini coefficients that, while lower in level, captured the long-term trends in inequality relatively well. They argued that their Gini coefficient estimates from the public use CPS data better captured long-term trends in labor earnings for this population than even Census Bureau estimates based on uncorrected internal CPS data.[2]

3. Bounds for P90/P10 and Seven Series of Estimates

In this section, first we describe our method for putting bounds on estimates of P90/P10 from censored data series and, second, we define seven series of CPS-based estimates that arise from application of the methods and from ignoring censoring.

3.1 Bounds on estimates of P90/P10 from top coded data

Let the true income distribution be denoted by the random variable x, which has a cumulative distribution function $F(x)$. The pth population income quantile ξ_p is defined by:

(1) $$p = F(\xi_p) = \Pr(x \le \xi_p)\ ,\ 0 \le p \le 1.$$

Suppose we have a random sample s comprising N income units, with the distribution of their incomes described by the vector $\boldsymbol{x} = \{x_1, x_2, x_3, \ldots, x_N\}$. The sample estimate of the pth quantile of the distribution is:

(2) $$\hat{\xi}_p = \sup\{\ x_i \in s \mid \hat{F}(x_i) \le p\ \},$$

derived by solving the equation $p = \hat{F}(\xi_p)$, where the sample estimate of the cumulative distribution function for $\boldsymbol{x}$ is:

(3) $$\hat{F}(x) = \Sigma_s\, w_i\, I(x_i \le x) \,/\, \hat{N}\ ,\ \text{with}\ \hat{N} = \Sigma_s\, w_i.$$

$I(.)$ is the indicator function and the sample weight for unit i is w_i.

The problem for researchers is that $\boldsymbol{x}$ is not fully observed. Top coding (or right censoring in general) means that some incomes at the top of the income distribution are not observed. Instead, two other vectors are observed in the sample data: censored incomes $\boldsymbol{y} = \{y_1, y_2, y_3, \ldots, y_N\}$ and censoring indicators $\boldsymbol{c} = \{c_1, c_2, c_3, \ldots, c_N\}$, with $y_i = x_i$ if $c_i = 0$ and $y_i < x_i$ if $c_i = 1$, for each $i = 1, \ldots, N$. In addition, because we are trying to model incomes that are aggregates of several income sources, but censoring occurs at the level of each individual income source, some lower-valued incomes might be censored while higher-valued ones are not censored.

The sample estimate of the proportion of censored observations is $\hat{\theta}$ where:

(4) $$\hat{\theta} = \Sigma_s\, w_i\, I(c_i = 1) \,/\, \hat{N}\ .$$

Although income values may be censored, we can place lower and upper bounds on the quantiles that we are trying to estimate (Manski, 1994). The lower bound is derived from distribution $\boldsymbol{y}$, assuming that the true (unobserved) value of each censored observation is equal to the observed censored value. The upper bound is derived by assuming that the true income value of each censored observation is equal to positive infinity, i.e. estimated from a distribution $\boldsymbol{z} = \{z_1, z_2, z_3, \ldots, z_N\}$, with $z_i = x_i$ if $c_i = 0$ and $z_i = +\infty$ if $c_i = 1$, for each $i = 1, \ldots, N$. In general, the ranking by income of units differs between distributions $\boldsymbol{y}$ and $\boldsymbol{z}$ and hence lower and upper bound estimates of the quantiles of the true distribution differ.

More formally, the estimate of the lower bound is:

$$\hat{\xi}_p^L = \sup\{\, y_i \in s \mid \hat{F}_y(y_i) \le p \,\}, \tag{5}$$

where the empirical CDF of the censored distribution $\boldsymbol{y}$ is:

$$\hat{F}_y(y) = \Sigma_s\, w_i\, I(y_i \le y) \,/\, \hat{N}\,. \tag{6}$$

The estimate of the upper bound is:

$$\hat{\xi}_p^U = \sup\{\, z_i \in s \mid \hat{F}_z(z_i) \le p \,\}, \tag{7}$$

where the empirical CDF of the distribution $\boldsymbol{z}$ is:

$$\hat{F}_z(z) = \Sigma_s\, w_i\, I(\mathrm{z}_i \le \mathrm{z}) \,/\, \hat{N}\,. \tag{8}$$

It is straightforward to show that $\hat{\xi}_p^L \le \hat{\xi}_p \le \hat{\xi}_p^U$ for $0 \le p \le 1$, because $y_i \le x_i \le z_i$ for each $i = 1, \ldots, N$. Moreover, when $p \le 1-\hat{\theta}$, the upper and lower bounds are both informative. If, instead, $p > 1-\hat{\theta}$, censoring bites: the pth quantile lies within the censored income range. In this case, the lower bound estimate of the pth quantile derived from $\boldsymbol{y}$ remains well-defined, but the upper bound estimate is uninformative—it is infinity.

To illustrate how the upper and lower bounds of order statistics such as quantiles are derived, we give a simple numerical example. Assume the distribution of observed incomes is {2,000, 1,000, 4,000, 5,000} and the first income is censored. (Recall that a censored value need not be the maximum value observed in sample data.) Suppose the aim is to estimate the income corresponding to the upper quartile (the income of the second highest earner in this simple case). Only one income is censored, and so we have the case corresponding to $p \leq 1-\theta$. The lower bound estimate of the upper quartile is 4,000, and the upper bound estimate is 5,000. Now suppose instead that income 4,000 is also censored. This takes us to the case $p > 1-\theta$. The lower bound estimate of the upper quartile is again 4,000, but the upper bound estimate is uninformative.

If the total income for any income-recipient unit (e.g. a household) is the aggregate of incomes across individuals belonging to the same unit, the same estimation methods may be applied. The greater the aggregation across income sources, or across individuals, the further down the distribution of total income that censoring is likely to occur. There is a range of top coded values interspersed along the range of non-top coded values. This dispersion means that the adjustment for top coding in the CPS proposed by Fichtenbaum and Shahidi (1988) for estimation of the Gini coefficient, based on fitting a Pareto distribution to incomes above a single critical value, is not practical in the current context.

3.2 Seven Series of P90/P10 Estimates

Using these bounding methods, we calculate upper and lower bound estimates for P90/P10 based on public use CPS data files, which we will call the *Public-Upper* and *Public-Lower* series respectively. Because we have access to the internal CPS data files, we are also able to calculate *Internal-Upper* and *Internal-Lower* series of P90/P10 estimates from the internal CPS data in a similar way. Because internal data contain more information than public use data (the internal censoring point is greater than or equal to the public use censoring point), the *Public-Upper* estimates will be greater than or equal to corresponding *Internal-Upper* estimates and *Public-Lower* estimates will be less than or equal to the corresponding *Internal-Lower* estimates.

We also calculate three other P90/P10 series from the CPS for comparison purposes. The first, labeled *Public*, is calculated from public use files using the top coded value assigned by the Census Bureau to individuals' sources of income for all years. For each income year before 1995, estimates are the same as *Public-Lower* estimates for the same year. They are greater thereafter because, from income year 1995 onwards, the Census Bureau assigned an estimated cell mean to each top coded value based on the person's characteristics rather than the top code cutoff value. For these years, because the *Public* series is based on a distribution in which income values are more accurately observed than in the distribution including top coded values, it should yield P90/P10 estimates that are closer to the estimates based on internal data.

The second additional series, labeled *Cell-Mean*, assigns a cell-mean that we consistently calculate over all the years of internal data available to us (1975–2004) for each person top coded. Because we were given permission to use the internal data, we were able to construct a data file similar to the one discussed below that the Census Bureau has, since 1995, used to assign cell means to top coded values in the public use files. For the same reasons discussed above, the P90/P10 estimates in this series should more closely track the estimates derived from the internal data in all years.

In income year 1995, the Census Bureau began providing cell mean values rather than the top coded cutoff value for wages and salaries, self-employment earnings, and farm earnings from sex/race/work experience cells. That is, rather than reporting the top code cutoff value, the public use file reports the average value for those with the same sex/race/work experience characteristics with values above the top code cutoff point. In income year 1998, the Census Bureau extended its provision of cell means to other non-governmental sources of income. However, to date the Census Bureau has not provided cell means based on this methodology for earlier years. Hence for reasons of consistency, researchers interested in comparing trends in labor earnings or income before 1995 with those after 1995 are not able to take advantage of the cell mean option available in the public use data. However, using our access to the internal data, we were able to create a consistent set of cell mean values for each income source for every person in the public use files for income years 1975–2004.[3]

The third additional series, labeled *Rule-of-Thumb*, assigns a value of 150 percent of the top code cutoff value to all top coded values. This popular rule-of-thumb approach to assigning top code values has been used in the labor economics literature by Katz and Murphy (1992), Autor *et al.* (2005), and Lemieux (2006).

4. Trends in wages and labor earnings inequality for full-time, full-year, workers

Seven series of P90/P10 estimates were calculated for the distribution of wages and salaries of full-time, full-year workers, the most typical definition of labor earnings and of a worker in the labor economics literature tracking the inequality of labor earnings.[4] See Table 2. The first five columns provide estimates based on public use CPS data (though note that column 5 is based on our cell means series that is not yet available to the public). The last two columns are derived from internal CPS data. We show below that, although censoring is a potential problem in estimating inequality trends for wage and salary income of this population, it is not a very important one, because there is no censoring problem in the internal data and only a small potential problem in the public use data.

<Table 2 near here>

Prior to income year 1987, wages and salaries income came from only one source (INCWAG): see Table 1. Hence top coding was not a problem since none of the workers with wage and salary top codes in these years had incomes below the 90th percentile of the wage and salary distribution. Since then, the 90th percentile value could be affected by top coding, at least in principle, since the Census Bureau began reporting wage and salary income from two sources, one primary (INCER) and one secondary (INCWG1). Hence it is possible that workers below the 90th percentile of the distribution of wages and salaries formed by the sum of these two sources could be top coded in one of them. As Columns 1 and 2 of Table 2 show, top coding is not a problem for estimation of P90/P10 for any income year prior to 1987 and is only a potential problem after 1995—where *Public-Upper* does not equal *Public-Lower*. And in none of these years is the difference between these two values very large.

Columns 6 and 7 of Table 2 show that the internal CPS data provide accurate P90/P10 estimates for all years since the *Internal-Upper* (column 6) values equal *Internal-Lower* values (Column 7) in all years and are, in fact, the same value as reported in columns 1 and 2 in all the years prior to 1996. Hence with respect to wage and salaries, P90/P10 estimates are relatively free of top coding problems. This pattern of no difference in values prior to 1996 and only small differences thereafter with the internal values holds for all of the other series in Table 2. The *Public* series values (column 3) and the *Cell-Mean* series values (column 5) are almost identical. This is the case prior to 1995 because top coding was not a problem for estimation of P90/P10 from either the internal or public use data, so not correcting for top coding by adjusting the cell means in the *Public* series in these years does not matter. Thereafter our consistently measured *Cell-Mean* series is so close to the *Public* series that there is almost no difference. Both the *Public* and *Cell-Mean* series are slightly higher than the *Internal-Upper* one in most years since 1995, showing that using either of the cell mean adjustments slightly overestimates values derived from internal data. In contrast, the *Rule-of-Thumb* series, already available to the public, yields virtually the same P90/P10 estimates as the series based on internal data.

Table 2 confirms that, whereas in theory top coding could affect both internal and public use P90/P10 estimates for wages and salaries income, in practice it has no effect on P90/P10 estimates from internal data and only a minor effect on estimates from public use data after 1995. The table also suggests that the rule-of-thumb method common in the wage and salaries literature is at least as effective as using cell means to control for the effects of inconsistent top coding.

Table 3 reports trends in P90/P10 for the distribution of the total earnings of full-time, full-year workers, for each of the seven series. Prior to income year 1987, the Census Bureau summed income from three different sources to create the total earnings variable: wages and salaries (INCWAG), self-employment earnings (INCSE), and farm earnings (INCFRM). Since then, four sources have been combined: primary earnings (INCER), second wages and salaries (INCWG1), secondary self-employment earnings (INCSE1), and secondary farm earnings (INCFR1). As was the case in Table 2, censoring does not matter for any year prior to 1987 or for the years up to 1996 in the public use data (columns 1 and 2). However, in more recent years, top coding has become more of a potential problem. But even in these years, the differences between the series of estimates are small. In the years for which we have access to the internal files, censoring has not been a problem, with *Internal-Upper* estimates equaling *Internal-Lower* estimates in all years (columns 6 and 7). Once again, the *Public* estimates (column 3) and the *Cell-Mean* estimates (column 5) produce series that differ little after 1995 because they use a similar cell mean strategy and are the same prior to 1995 because top coding problems in the data do not affect estimation of P90/P10. Both slightly overestimate the values found in the internal data series. The *Rule-of-Thumb* series, already available to the public, yields virtually the same P90/P10 estimates as the internal series. But, once again, because top coding of total earnings in both the public use and internal data is a relatively small problem for estimation of P90/P10, any of these methods of controlling for top coding in the public use CPS results in plausible approximations of the internal CPS series.

<Table 3 near here>

5. TRENDS IN SIZE-ADJUSTED HOUSEHOLD INCOME INEQUALITY FOR INDIVIDUALS

P90/P10 estimates for the distribution of size-adjusted household income of individuals are reported in Table 4 for all seven series.[5] There are far more sources of household income than for total labor earnings and, because household income is assumed to be shared, the size-adjusted household income of each household member depends on the income sources of every household member. Thus censoring is likely to be a more serious problem in this literature than was the case for income from wages and salaries or from total labor earnings. Prior to 1987, eleven sources of income were reported, and the number has increased to 24 since then (see Table 1). As Table 4 shows, P90/P10 estimates derived from the public CPS data are affected by top coding problems although, prior to the 1990s, the gap between the

Public-Upper and *Public-Lower* series is small. But the gap between the two series has risen steadily since then and especially since 1998. This is clear from Figure 1 which graphs the *Public-Upper* and *Public-Lower* series.

<Table 4 near here>

<Figure 1 near here>

A clue to the source of the divergence between the *Public-Upper* and *Public-Lower* series is provided by Figure 2. The top line shows, for each year, the percentage of all individuals affected by top codes in the public use CPS file. This percentage increased steadily in the early 1990s, declined a little in the middle 1990s, and then rose sharply after 1996. This is not a problem as long as censoring only occurs for individuals whose size-adjusted household income is above the 90^{th} percentile of the distribution. Thus in Figure 2 we also show the percentage of all individuals who had observed size-adjusted household incomes less than the 95^{th} percentile and whose income was affected by top coding, together with corresponding percentages for those with incomes below the 90^{th} and 85^{th} percentiles. Individuals with incomes below the 90^{th} percentile began to be affected by top coding in the early 1990s and have been more sharply affected since 1998. Note that measuring inequality in terms of the ratio of the 85^{th} percentile to the 10^{th} percentile rather than P90/P10 would reduce this problem somewhat but would not resolve it.

<Figure 2 near here>

Figure 3 focuses on the post-1987 period and shows the percentage of top coded values below the 90 percentile by income source: primary labor earnings, other labor earnings, and all other income. Figure 3 shows that the jump in the gap between *Public-Upper* and *Public-Lower* estimates was primarily driven by the sharp increase in the fraction of individuals below the 90 percentile whose non-labor earnings was top coded, which rose from 0.1 percent in 1997 to 1.0 percent in 1998 and increased to 1.6 percent by 2004. Appendix Table 2 shows that, in income year 1998 (corresponding to CPS survey year 1999), when the Census Bureau started to top code all non-governmental sources of non-labor income items, there was a substantial reduction in the top code values in the public use files. For example, the censoring point for interest income was $99,999 in 1997, but only $35,000 in 1998.

<Figure 3 near here>

Hence unlike P90/P10 estimates derived from internal CPS data, P90/P10 estimates derived from public use data have been substantially affected by censoring, and this is especially the case in recent years. But, as Table 4 also shows, censoring problems are not confined to public use data. As can be seen from columns 6 and 7, *Internal-Upper* and *Internal-Lower* values are not the same in each year, although in most cases the difference is relatively small. Hence, when compared to the top coding problems in the public use CPS, the differences between the *Internal-Upper* and *Internal-Lower* series are negligible relative to the differences between the *Public-Upper* and *Public-Lower* series: see Figure 1.

For income, as for wages and salaries and total labor income, there is very little difference between the *Public* series (column 3) and the *Cell-Mean* series (column 5) from 1995 onward: compare Table 4 columns 3 and 4 with Tables 2 and 3. But, the situation for income differs from the other variables before 1995. Because P90/P10 estimates of income inequality from both public use data and, to a lesser degree from internal data, are affected by censoring, our *Cell-Mean* series does a much better job of aligning P90/P10 estimates from public use data with the series estimated from internal data. In the years prior to 1995, the *Cell-Mean* series almost coincides with the internal series. But thereafter, like *Public* estimates values, *Cell-Mean* estimates tend to slightly overstate P90/P10 relative to corresponding internal values. Although the *Rule-of-Thumb* estimates fall within the range provided by the *Public-Upper* and *Public-Lower* series, they now consistently fall below the range provided by the *Internal-Upper* and *Internal-Lower* series. For researchers interested in capturing long term trends in income inequality, measured using P90/P10 and estimated from public use CPS data, Table 4 shows that top coding is a problem and that our *Cell-Mean* series values do the best job of offsetting it and capturing the P90/P10 trends derived from internal CPS data.

6. Longer-term trends in inequality using Adjusted public use CPS data: P90/P10 versus Gini estimates

Researchers in the labor and income inequality literature employing public use CPS data frequently summarize trends in inequality using the P90/P10 measure rather than more traditional summary measures of inequality such as the Gini coefficient, Theil indices, or the coefficient of variation, because of concerns about censoring in CPS data. We have demonstrated that P90/P10 estimates are also subject to censoring problems, especially when used to measure household income inequality. But we have also shown that, by using a consistent set of cell means created from internal CPS data, one can estimate a P90/P10 series that is quite close to the P90/P10 series estimated with internal CPS data. The issue that we turn to now is whether P90/P10 estimates provide a picture of inequality trends that is robust. Does P90/P10 provide the same picture of inequality trends as a picture based on a measure that uses information about all incomes in the distribution rather than focusing only on two points?

We compare trends in inequality (of wage and salaries income, labor earnings, and the size-adjusted household income of individuals) derived from our P90/P10 *Cell-Mean* series with trends derived from Gini coefficients based on public use and on internal data. We employ the Gini coefficient as it is the most commonly-estimated summary measure of inequality used in the income distribution literature. We use our *Cell-Mean* series for P90/P10 both because it more closely replicates the internal series than any other currently available to the general research community and because, in principle, the underlying cell means could be made available to the public.

We derive time-consistent Gini inequality values via a consistent top coding method that is applied to both the public use data and the internal data for the years 1975–2004. We calculate the percentage of individuals subject to top coding in every year for each income source. We determine the year in which the greatest percentage of the population was affected by the top code for that income source and then top code the income source for every year to yield this same percentage. This procedure ensures that a common and constant percentage of the upper tail distribution is affected in each year for each income source. In doing so, we adjust the top codes used for each subcomponent of first wage and salary earnings, then labor earnings and then household income. For a fuller discussion of this method, see Burkhauser *et al.* (2004) and Feng *et al.* (2006) for its application to labor earnings, and Burkhauser *et al.* (2004) and Burkhauser *et al.* (2006) for its application to size-adjusted household income.

We are interested in comparing trends in inequality based on our adjusted P90/P10 estimates with trends in inequality based on our consistently top coded public use and internal CPS Gini values, so all three series are normalized using year 1975 as the base. Normalized Gini coefficient and P90/P10 estimates for wages and salaries among full-time, full-year workers from 1975 to 2004 are displayed in Figure 4. The P90/P10 series shows a greater degree of variance from one year to the next. According to it, inequality increased less in the early years and more in the later years than is the case according to either of the Gini series, with the difference most pronounced in the last few years.

<Figure 4 near here>

The estimates for the distribution of total labor earnings among full-time, full-year workers are shown in Figure 5, derived using the same methods as in Figure 4. In this case, there is a much greater difference in the relative trends. Not only is there much greater variance in P90/P10 estimates but, after the first few years, there is also a much greater rise in inequality based on the P90/P10 series over time than that produced by the estimates of the Gini coefficient from either the consistently top coded public use data or internal data.

<Figure 5 near here>

Estimates for the distribution of size-adjusted household income among individuals, derived using the same methods, are shown in Figure 6. Once again there is much greater variance in P90/P10 estimates over time, and there is now an even greater rise in P90/P10-measured inequality over time. The increase is much greater in magnitude than that indicated by the two Gini coefficient series. (The statistics graphed in Figures 4–6 are reported in Appendix Tables 5 and 6, together with the ratios of the P90/P10 and Gini coefficient estimates.)

<Figure 6 near here>

To more formally test differences in linear trends, we use a regression technique similar that employed by Burkhauser *et al.* (2004) and Feng *et al.* (2006), and summarized by the specification in the equation below. The dependent variable (*Index*) is the normalized inequality measure: a public use data based Gini coefficient or P90/P10. There are six explanatory variables: a constant, which is the level of P90/P10; a time trend t (= 1, 2, ..., 30), capturing the trend in P90/P10; a dummy variable which controls for the difference between levels of Gini and P90/P10 ($d = 1$ if the dependent variable is the Gini, and 0 otherwise); dt (d and t interacted), which controls for the difference between the trends in the two inequality measures; a dummy variable that controls for whether the observation refer to the post-1992 period of not ($u = 1$ if post-1992, and 0 otherwise) that we include to account for substantial changes in CPS collection procedures in that year (Feng *et al.*, 2006); du (d and u interacted) to control for differences in the post-1992 levels. Each number in parenthesis is the absolute value of the ratio of the corresponding regression coefficient to its robust standard error.

We first report results for wages and salaries of full-time full-year workers. The estimated equation is as follows:

$$\begin{array}{lllllll} \text{Index} = & 0.975 & +\,0.0105\,t & +\,0.0265\,d & -\,0.0024\,dt & +\,0.0243\,u & -\,0.0143\,du \\ & (117) & (14.04) & (2.25) & (2.30) & (1.85) & (0.77) \end{array}$$

The statistically significant coefficient for t suggests that inequality measured using the P90/P10 rose over time. The Gini coefficient shows a significantly different trend, as suggested both by the positive and significant value of d and by the interaction of d and t. The rise in inequality according to the Gini coefficient is significantly greater in the early years of the data but then becomes less so and eventually increases at a smaller rate than the P90/P10 trend. The increase in level of inequality after 1992, captured in u, is not quite significant.

For total earnings of full-time full-year workers, the estimated equation is:

$$\begin{array}{lllllll} \text{Index} = & 1.0012 & +\,0.0101\,t & -\,0.0035\,d & -\,0.0046\,dt & -\,0.0006\,u & +\,0.019\,du \\ & (89) & (10.05) & (0.22) & (3.22) & (0.03) & (0.74) \end{array}$$

For the whole period 1975–2004, P90/P10 shows a positive and significant linear trend, as suggested by the coefficient on *t*. Again, the Gini coefficient shows a different trend, suggested by the significance of *dt*, with a slower rate of increase. Nevertheless, there is still a positive trend for the Gini coefficient, as the F-test of the hypothesis that $t + dt = 0$ is rejected at the 1 percent level. Again, there is no significant change in levels for either the Gini coefficient or the P90/P10 for the post-1992 period.

For the distribution of size-adjusted household income among individuals, the estimated equation is:

$$\begin{array}{lcccccc} \text{Index} = & 1.028 & +\,0.0159\,t & -\,0.0266\,d & -\,0.0090\,dt & -\,0.0750\,u & +\,0.061\,du \\ & (51) & (8.96) & (0.95) & (3.58) & (2.39) & (1.38) \end{array}$$

For the whole period, P90/P10 shows a positive and significant trend, as suggested by the coefficient of *t*. Again, the Gini coefficient estimates show a different trend, suggested by the significance of *dt*, with a slower rate of increase. The level in inequality post-1992 is significantly lower than in early years as suggested by the significance of *u*. Nevertheless, there is still a positive trend for the Gini, as the F-test of the hypothesis that $t + dt = 0$ is rejected at the 1 percent level.

In all the regressions, the Gini coefficient and P90/P10 estimates show different trends. (The robustness of this result to using internal data instead is shown in the Appendix.) Thus, researchers should be cautious about using the relative position of two points in the distributions of wages and salaries, labor earnings or income to draw conclusions about how overall inequality of each of these income sources changed over the last three decades. The choice of inequality measure matters.

7. Summary and Conclusions

We investigate how P90/P10 is affected by censoring when used to measure inequality in the distribution of wages and salaries, labor earnings and household income. We do so both with public use and internal CPS data. In all cases we found that top coding is less of a problem for researchers using P90/P10 to measure inequality in wages and salaries and labor earnings than it is for those assessing inequality of size-adjusted household income. And, it is far less of a problem in the internal data than in the public use data. Except for the case of the household income distribution, estimating P90/P10 using a rule-of-thumb method to control for top coding in the public use data does as good a job as using our consistently created cell mean series in estimating P90/P10 values calculated from internal data.

However, we found that the cell mean series we created for all years of public use CPS data yields superior estimates of internal data-estimates of P90/P10 than does either using no cell means or using the cell means that the Census Bureau has provided from 1995 onward. We urge the Census Bureau to allow us to provide our cell mean series to the general research community or to develop and distribute an alternative cell mean series for all years of the public use CPS data.

P90/P10 is only one measure of inequality. Our comparisons of P90/P10 and Gini coefficient series derived using consistently top coded public use or internal CPS data yield large and significant differences in longer term trends for all three of the income definitions considered, but the largest differences by far were for our size-adjusted household income series. Hence researchers should be cautious about inferring longer term trends in inequality on the basis of a single inequality measure.

Furthermore, because the United States Census Bureau is not alone among statistical agencies in top coding income values, it is important for researchers to carefully consider the potential impact of top coding practices on their estimates of trends in inequality even if they measure inequality using P90/P10.

References

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

Atkinson, Anthony B. 2007. "The Long Run Earnings Distribution in Five Countries: Remarkable Stability," The Review of Income and Wealth, 53 (1): 1–24.

Autor, David H., Lawrence F. Katz, and Melissa S. Kearney. 2005. "Trends in U.S. Wage Inequality: Re-Assessing the Revisionists," NBER Working paper 11627, National Bureau of Economic Research, Cambridge MA.

Blau, Francine D. and Lawrence M. Kahn. 2005. "Do Cognitive Test Scores Explain Higher U.S. Wage Inequality?" *Review of Economics and Statistics,* 87 (1) 184–193.

Brandolini, Andrea. Forthcoming. "Measurement of Income Distribution in Supranational Entities: The Case of the European Union," in Stephen P. Jenkins and John Micklewright (Eds.), *Inequality and Poverty Re-Examined*, Oxford University Press, Oxford.

Burkhauser, Richard V., J.S. Butler, Shuaizhang Feng, and Andrew Houtenville. 2004. "Long-Term Trends in Earnings Inequality: What the CPS Can Tell Us," *Economics Letters,* 82: 295–299.

Burkhauser, Richard V., Kenneth A. Couch, Andrew 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., Takashi Oshio, and Ludmila Rovba. 2007. "Winners and Losers over the 1990s Business Cycles in Germany, Great Britain, Japan, and the United States." *Schmollers Jahrbuch: Journal of Applied Social Science Studies*, 127(1):

Daly, Mary C. and Robert G. Valletta. 2006. "Inequality and Poverty in United States: The Effects of Rising Dispersion of Men's Earnings and Changing Family Behaviour," *Economica,* 73 (289): 75–98.

Danziger, Sheldon and Peter Gottschalk. (Eds.) 1993. *Uneven Tides: Rising Inequality in America*, Russell Sage Foundation, New York.

DiNardo, John, Nicole Fortin, and Thomas Lemieux. 1996. "Labor Market Institutions and the Distribution of Wages, 1973-1992: A Semi-Parametric Approach." *Econometrica* 64: 1001–1044.

Feng, Shuaizhang, Richard V. Burkhauser, and J.S. Butler. 2006. "Levels and Long-Term Trends in Earnings Inequality: Overcoming Current Population Survey Censoring Problems Using the GB2 Distribution," *Journal of Business and Economic Statistics* 24 (1): 57–62.

Fichtenbaum, Rudy, and Hushang Shahidi, 1988. "Truncation Bias and the Measurement of Income Inequality," *Journal of Business & Economic Statistics*, 6 (3): 335–337.

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.

Gottschalk, Peter, and Mary Joyce. 1998. "Cross-National Differences in the Rise in Earnings Inequality: Market and Institutional Factors," *The Review of Economics and Statistics*, 80 (4): 489–502.

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

Juhn, Chinhui, Kevin M. Murphy, and Brooks Pierce. 1993. "Wage Inequality and the Rise in Returns to Skill," *Journal of Political Economy* 101 (3): 410–442.

Katz, Lawrence F. and David H. Autor. 1999. "Changes in the Wage Structure and Earnings Inequality," in O. Ashenfelter and D. Card, eds., *Handbook of Labor Economics*, vol. 3A, North-Holland, 1463–1555.

Katz, Lawrence F. and Kevin M. Murphy. 1992. "Changes in Relative Wages, 1963-87: Supply and Demand Factors," *Quarterly Journal of Economics* 107 (February): 35–78.

Karoly, Lynn A. and Gary Burtless. 1995. "Demographic Changes, Rising Earnings Inequality, and the Distribution of Personal Well-Being, 1959-1989," *Demography,* 32 (3): 379–405.

Luxemburg Income Study. Income Inequality Measures. http://www.lisproject.org/keyfigures/ineqtable.htm accessed May 22, 2007.

Lemieux, Thomas. 2006. "Increasing Residual Wage Inequality: Composition Effects, Noisy Data, or Rising Demand for Skill?" *The American Economic Review*, 96(3): 461–498.

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

Manski, Charles F., 1994. "The Selection Problem", in C. Sims (ed.), *Advances in Econometrics, Sixth World Congress*, Cambridge, Cambridge University Press, 143–170.

Nielsen, Francois, Arthur Alderson, and Jason Beckfield, 2005. "Exactly How Has Income Inequality Changed? Patterns of Distributional Change in Core Societies," *International Journal of Comparative Sociology* 46: 405–423.

Pencavel, John. 2006. "A Life Cycle Perspective on Changes in Earnings Inequality among Married Men and Women," *The Review of Economics and Statistics*, 88 (2): 232–242.

Prus, Steven and Robert Brown, 2006, "Income Inequality over the Later-life Course: A Comparative Analysis of Seven OECD Countries," Working Paper 435, Luxembourg Income Study, Luxembourg. http://www.lisproject.org/publications/liswps/435.pdf

Smeeding, Timothy, M. 2004. "Twenty Years of Research on Income Inequality, Poverty, and Redistribution in the Developed World. *Socio-Economic Review*, 2: 149–163.

Trejo, Stephen J. 1997. "Why Do Mexican Americans Earn Low Wages?" *Journal of Political Economy,* 105(6): 1235–68.

Welniak, Edward J. 2003. "Measuring Household Income Inequality Using the CPS," in James Dalton and Beth Kilss (Eds.), *Special Studies in Federal Tax Statistics 2003*, Statistics of Income Directorate, Inland Revenue Service, Washington DC.

WIDER, 2007. *World Income Inequality Database. User Guide and Data Sources*, United Nations University – World Institute for Development Economics Research, Helsinki. http://www.wider.unu.edu/wiid/wiid.htm

Table 1. Income Items Reported in the Current Population Survey

Name	Name in Public Files	Name in Internal Files	Definition
			1975–1986
Labor Earnings			
INCWAG	I51A	WSAL_VAL	Wages and Salaries
INCSE	I51B	SEMP_VAL	Self employment income
INCFRM	I51C	FRSE_VAL	Farm income
Other Sources			
INCSS	I52A	I52A_VAL	Income from Social Security and/or Railroad Retirement
INCSEC	I52B	SSI_VAL	Supplemental Security Income
INCPA	I53A	PAW_VAL	Public Assistance
INCINT	I53B	INT_VAL	Interest
INCDIV	I53C	I53C_VAL	Dividends, Rentals, Trust Income
INCOMP	I53D	I53D_VAL	Veteran's, unemployment, worker's compensation
INCRET	I53E	I53E_VAL	Pension Income
INCALC	I53F	I53F_VAL	Alimony, Child Support, Other income
			1987–2004
Labor Earnings			
INCER	ERN_VAL	ERN_VAL	Primary Earnings
INCWG1	WS_VAL	WS_VAL	Wages and Salaries-Second Source
INCSE1	SE_VAL	SE_VAL	Self employment income -Second Source
INCFR1	FRM_VAL	FRM_VAL	Farm income -Second Source
Other Sources			
INCSS	SS_VAL	SS_VAL	Social Security Income
INCSEC	SSI_VAL	SSI_VAL	Supplemental Security Income
INCPA	PAW_VAL	PAW_VAL	Public Assistance & Welfare Income
INCINT	INT_VAL	INT_VAL	Interest
INCDV2	DIV_VAL	DIV_VAL	Dividends
INCRNT	RNT_VAL	RNT_VAL	Rental income
INCALM	ALM_VAL	ALM_VAL	Alimony income
INCHLD	CSP_VAL	CSP_VAL	Child Support Income
INCUC	UC_VAL	UC_VAL	Unemployment income
INCWCP	WC_VAL	WC_VAL	Worker's compensation income
INCVET	VET_VAL	VET_VAL	Veteran's Benefits
INCRT1	RET_VAL1	RET_VAL1	Retirement income - source 1
INCRT2	RET_VAL2	RET_VAL2	Retirement income - source 2
INCSI1	SUR_VAL1	SUR_VAL1	Survivor's income - source 1
INCSI2	SUR_VAL2	SUR_VAL2	Survivor's income - source 2
INCDS1	DIS_VAL1	DIS_VAL1	Disability income - source 1
INCDS2	DIS_VAL2	DIS_VAL2	Disability income - source 2
INCED	ED_VAL	ED_VAL	Education assistance
INCONT	FIN_VAL	FIN_VAL	Financial Assistance
INCOTH	OI_VAL	OI_VAL	Other income

Table 2. P90/P10 Estimates for Wages and Salaries of Full-time Full-year Workers

Income Year	Public-Upper	Public-Lower	Public	Rule-of-Thumb	Cell-Mean	Internal-Upper	Internal-Lower
1975	4.00	4.00	4.00	4.00	4.00	4.00	4.00
1976	3.97	3.97	3.97	3.97	3.97	3.97	3.97
1977	4.08	4.08	4.08	4.08	4.08	4.08	4.08
1978	4.17	4.17	4.17	4.17	4.17	4.17	4.17
1979	3.97	3.97	3.97	3.97	3.97	3.97	3.97
1980	4.12	4.12	4.12	4.12	4.12	4.12	4.12
1981	4.00	4.00	4.00	4.00	4.00	4.00	4.00
1982	4.33	4.33	4.33	4.33	4.33	4.33	4.33
1983	4.29	4.29	4.29	4.29	4.29	4.29	4.29
1984	4.36	4.36	4.36	4.36	4.36	4.36	4.36
1985	4.44	4.44	4.44	4.44	4.44	4.44	4.44
1986	4.49	4.49	4.49	4.49	4.49	4.49	4.49
1987	4.40	4.40	4.40	4.40	4.40	4.40	4.40
1988	4.50	4.50	4.50	4.50	4.50	4.50	4.50
1989	4.66	4.66	4.66	4.66	4.66	4.66	4.66
1990	4.55	4.55	4.55	4.55	4.55	4.55	4.55
1991	4.57	4.57	4.57	4.57	4.57	4.57	4.57
1992	4.50	4.50	4.50	4.50	4.50	4.50	4.50
1993	4.58	4.58	4.58	4.58	4.58	4.58	4.58
1994	4.92	4.92	4.92	4.92	4.92	4.92	4.92
1995	5.00	5.00	5.00	5.00	5.00	5.00	5.00
1996	4.89	4.81	4.85	4.81	4.89	4.81	4.81
1997	5.00	4.92	5.00	5.00	5.00	5.00	5.00
1998	5.00	4.81	4.96	4.89	4.96	4.89	4.89
1999	5.07	5.00	5.00	5.00	5.00	5.00	5.00
2000	5.03	4.90	5.03	5.00	5.03	5.00	5.00
2001	5.13	5.00	5.07	5.00	5.07	5.00	5.00
2002	5.33	5.33	5.33	5.33	5.33	5.33	5.33
2003	5.26	5.13	5.19	5.13	5.26	5.13	5.13
2004	5.25	5.19	5.25	5.25	5.25	5.22	5.22

Notes. The definitions of the series are provided in the main text.

Table 3. P90/P10 Estimates for Total Earnings of Full-time Full-year Workers

Income Year	Public-Upper	Public-Lower	Public	Rule-of-Thumb	Cell-Mean	Internal-Upper	Internal-Lower
1975	4.27	4.27	4.27	4.27	4.27	4.27	4.27
1976	4.40	4.40	4.40	4.40	4.40	4.40	4.40
1977	4.63	4.63	4.63	4.63	4.63	4.63	4.63
1978	4.18	4.18	4.18	4.18	4.18	4.18	4.18
1979	4.45	4.45	4.45	4.45	4.45	4.45	4.45
1980	4.29	4.29	4.29	4.29	4.29	4.29	4.29
1981	4.58	4.58	4.58	4.58	4.58	4.58	4.58
1982	4.61	4.61	4.61	4.61	4.61	4.61	4.61
1983	4.65	4.65	4.65	4.65	4.65	4.65	4.65
1984	5.00	5.00	5.00	5.00	5.00	5.00	5.00
1985	4.72	4.72	4.72	4.72	4.72	4.72	4.72
1986	4.86	4.86	4.86	4.86	4.86	4.86	4.86
1987	5.00	5.00	5.00	5.00	5.00	5.00	5.00
1988	4.84	4.84	4.84	4.84	4.84	4.84	4.84
1989	5.00	5.00	5.00	5.00	5.00	5.00	5.00
1990	5.00	5.00	5.00	5.00	5.00	5.00	5.00
1991	5.00	5.00	5.00	5.00	5.00	5.00	5.00
1992	4.92	4.92	4.92	4.92	4.92	4.92	4.92
1993	5.09	5.09	5.09	5.09	5.09	5.09	5.09
1994	5.45	5.45	5.45	5.45	5.45	5.45	5.45
1995	5.00	5.00	5.00	5.00	5.00	5.00	5.00
1996	5.19	5.17	5.17	5.17	5.19	5.17	5.17
1997	5.32	5.32	5.32	5.32	5.32	5.32	5.32
1998	5.38	5.20	5.31	5.23	5.31	5.23	5.23
1999	5.54	5.38	5.54	5.46	5.54	5.46	5.46
2000	5.36	5.36	5.36	5.36	5.36	5.36	5.36
2001	5.30	5.04	5.24	5.10	5.24	5.17	5.17
2002	5.33	5.33	5.33	5.33	5.33	5.33	5.33
2003	5.53	5.47	5.52	5.47	5.53	5.47	5.47
2004	5.67	5.55	5.67	5.60	5.66	5.60	5.60

Notes. As for Table 2.

Table 4. P90/P10 Estimates for the Size-adjusted Household Income of Individuals

Income Year	Public-Upper	Public-Lower	Public	Rule-of-Thumb	Cell-Mean	Internal-Upper	Internal-Lower
1975	6.15	6.15	6.15	6.15	6.15	6.15	6.15
1976	6.11	6.11	6.11	6.11	6.11	6.11	6.11
1977	6.24	6.23	6.23	6.23	6.23	6.24	6.23
1978	6.35	6.32	6.32	6.34	6.33	6.34	6.33
1979	6.44	6.38	6.38	6.42	6.41	6.41	6.41
1980	6.74	6.61	6.61	6.71	6.68	6.66	6.66
1981	6.84	6.84	6.84	6.84	6.84	6.84	6.84
1982	7.53	7.52	7.52	7.53	7.52	7.52	7.52
1983	7.60	7.59	7.59	7.59	7.63	7.63	7.63
1984	7.62	7.62	7.62	7.62	7.62	7.62	7.62
1985	7.68	7.67	7.67	7.67	7.67	7.68	7.67
1986	7.85	7.84	7.84	7.84	7.84	7.84	7.84
1987	7.87	7.86	7.86	7.87	7.87	7.87	7.87
1988	7.91	7.90	7.90	7.91	7.91	7.91	7.91
1989	7.75	7.70	7.70	7.74	7.74	7.75	7.73
1990	7.80	7.76	7.76	7.80	7.79	7.78	7.78
1991	8.01	7.94	7.94	8.00	8.00	8.00	7.98
1992	8.25	8.15	8.15	8.24	8.24	8.22	8.21
1993	8.69	8.55	8.55	8.65	8.65	8.62	8.62
1994	8.53	8.26	8.26	8.48	8.47	8.44	8.41
1995	8.21	8.01	8.10	8.07	8.10	8.09	8.06
1996	8.28	8.10	8.17	8.15	8.19	8.19	8.16
1997	8.48	8.23	8.32	8.28	8.33	8.31	8.29
1998	8.75	7.98	8.26	8.15	8.26	8.22	8.18
1999	8.68	7.74	8.05	7.91	8.05	7.98	7.96
2000	8.59	7.67	7.96	7.87	7.96	7.93	7.91
2001	8.80	7.78	8.07	7.96	8.08	8.04	8.02
2002	8.62	7.96	8.12	8.08	8.12	8.12	8.10
2003	9.04	8.26	8.49	8.40	8.50	8.47	8.43
2004	9.14	8.24	8.43	8.35	8.44	8.44	8.41

Notes: As for Table 2. Also, for year 1983, interest incomes are reported differently in the public and internal data files. The results reported here use numbers from the internal data file.

Figure 1. P90/P10 Estimates for Size-Adjusted Household Income of Individuals, by Year

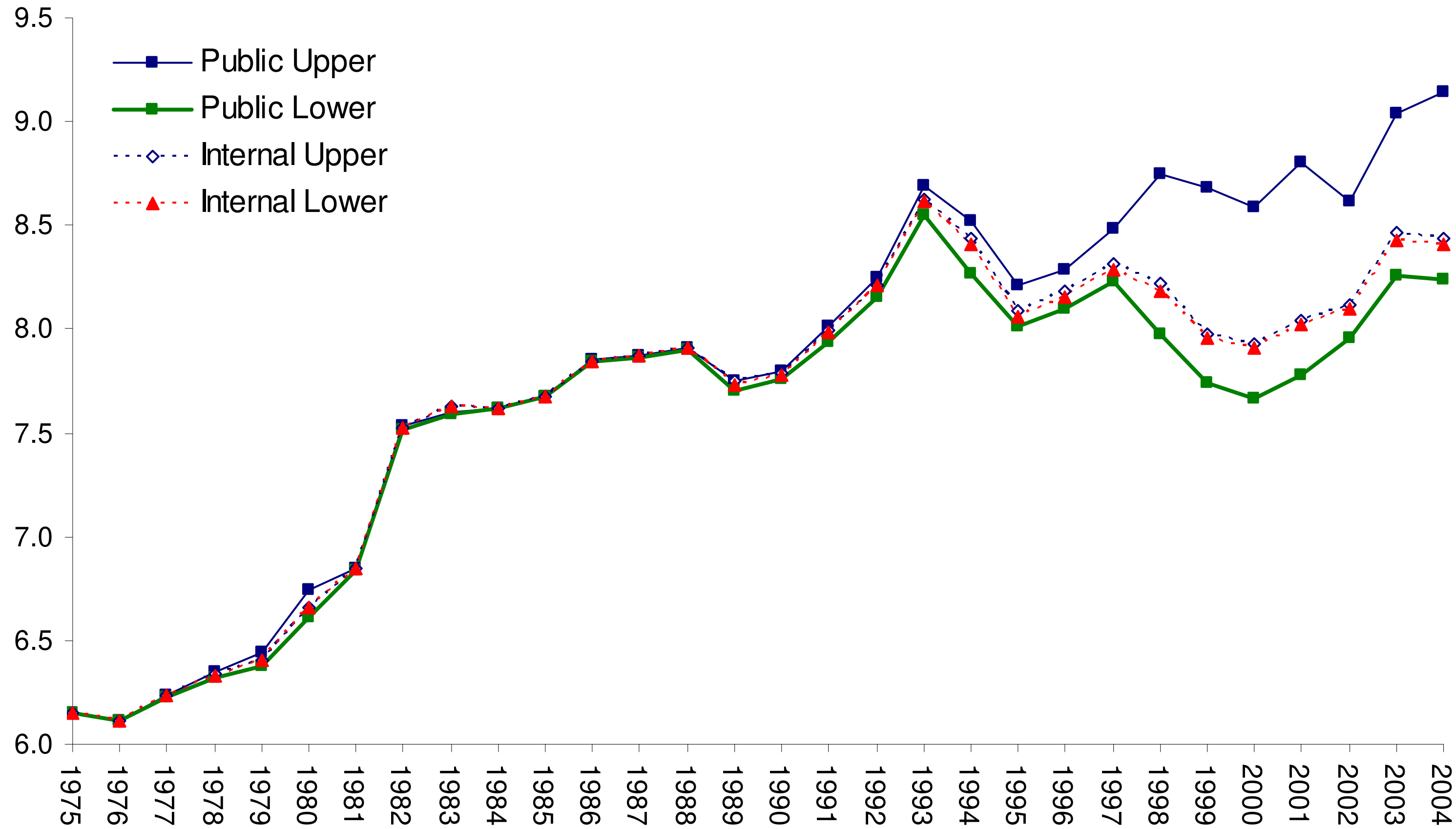


Figure 2. Percentage of Individuals with Size-Adjusted Household Income Censored in the Public Use CPS File

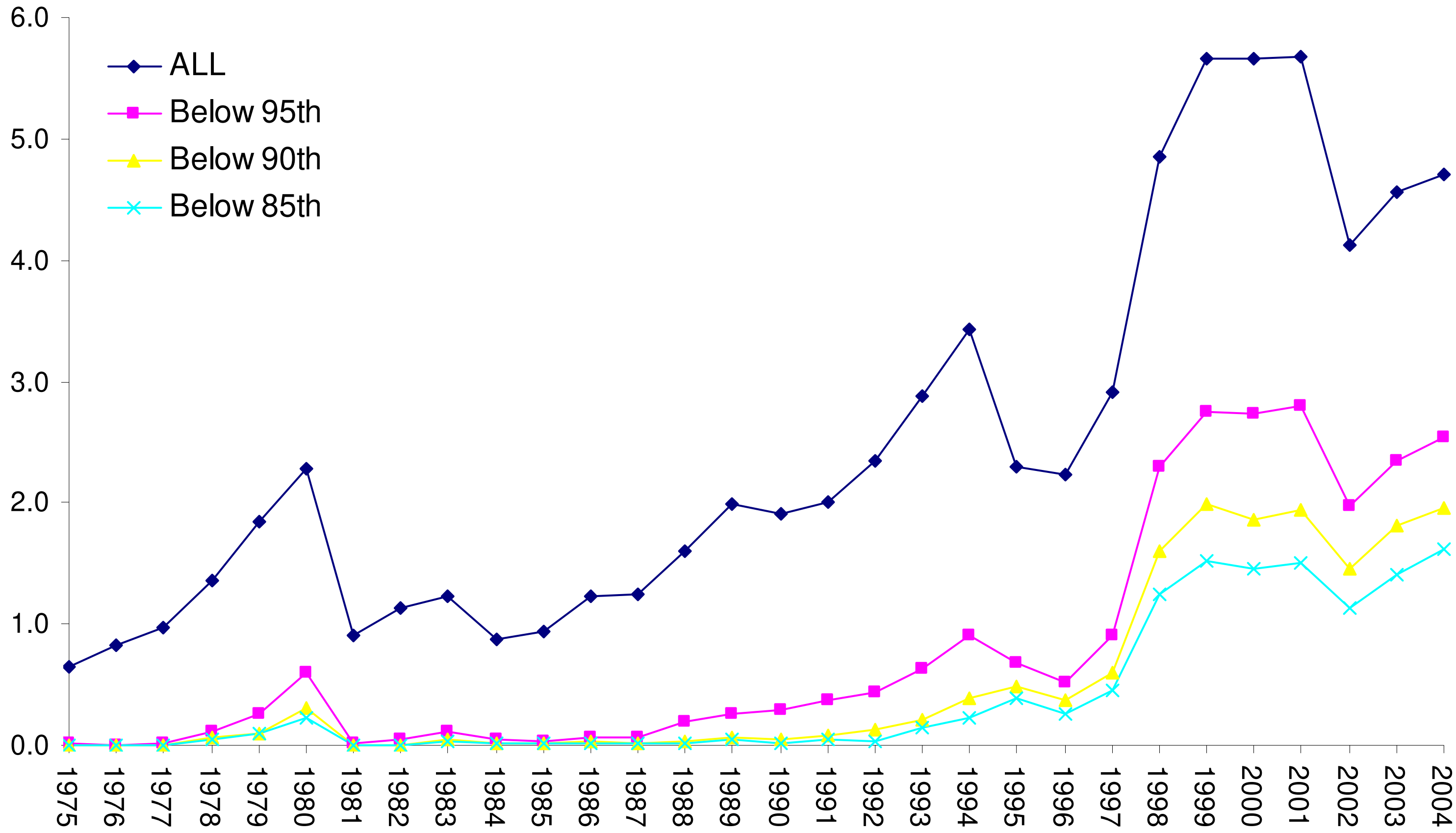


Figure 3. Percentage of Individuals with Censored Size-Adjusted Household Income Below the 90th Percentile by Income Source

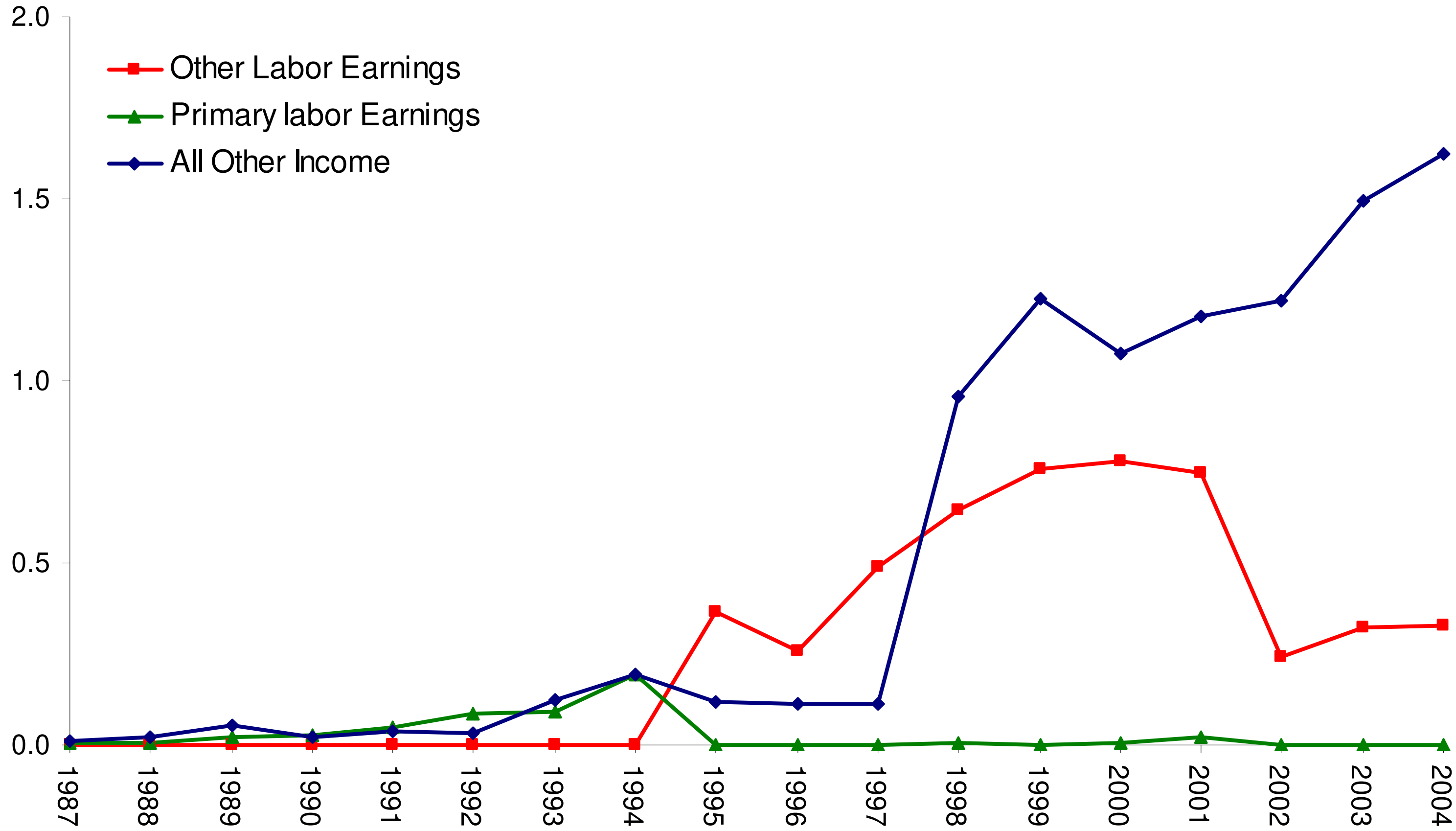
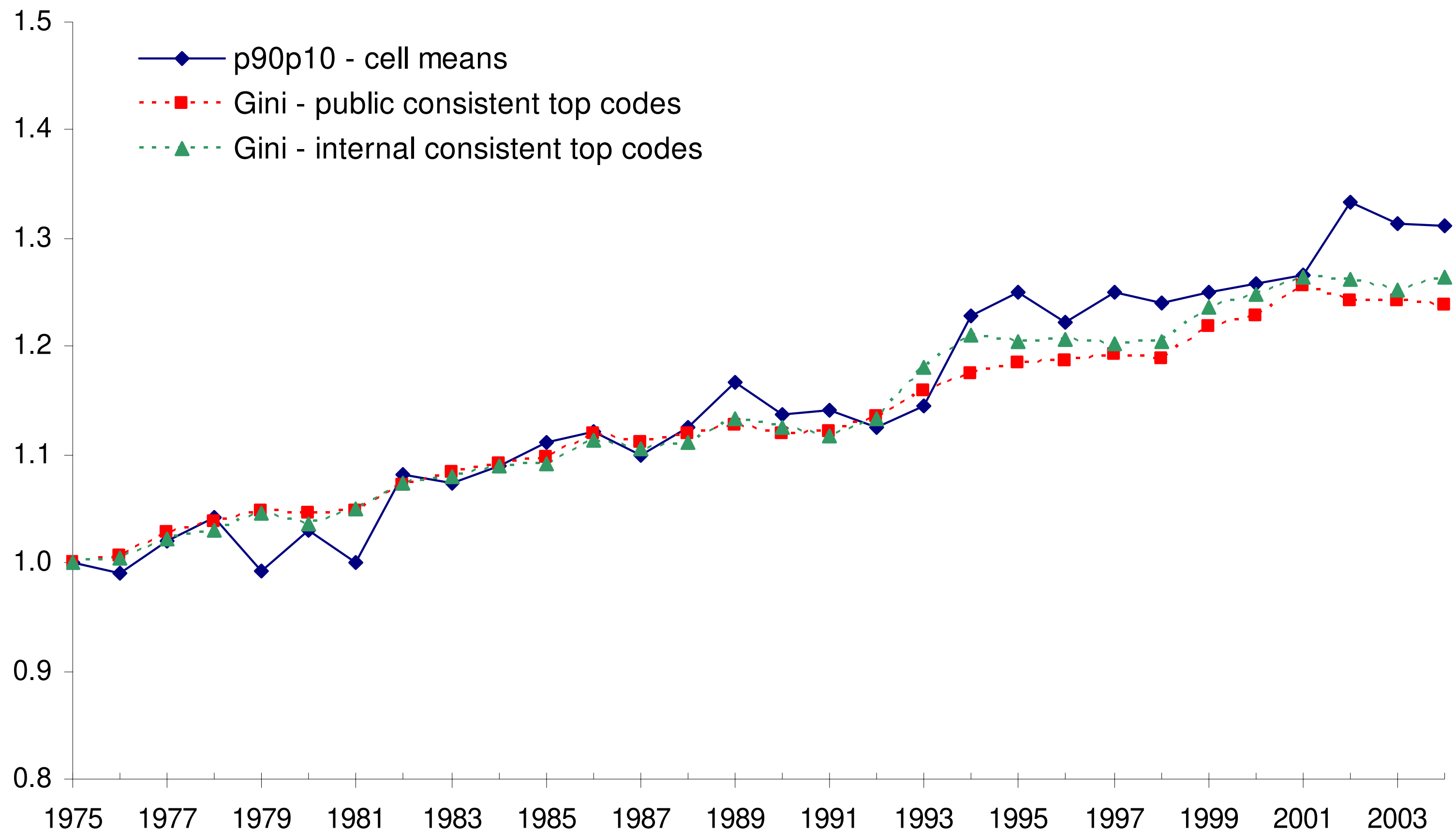
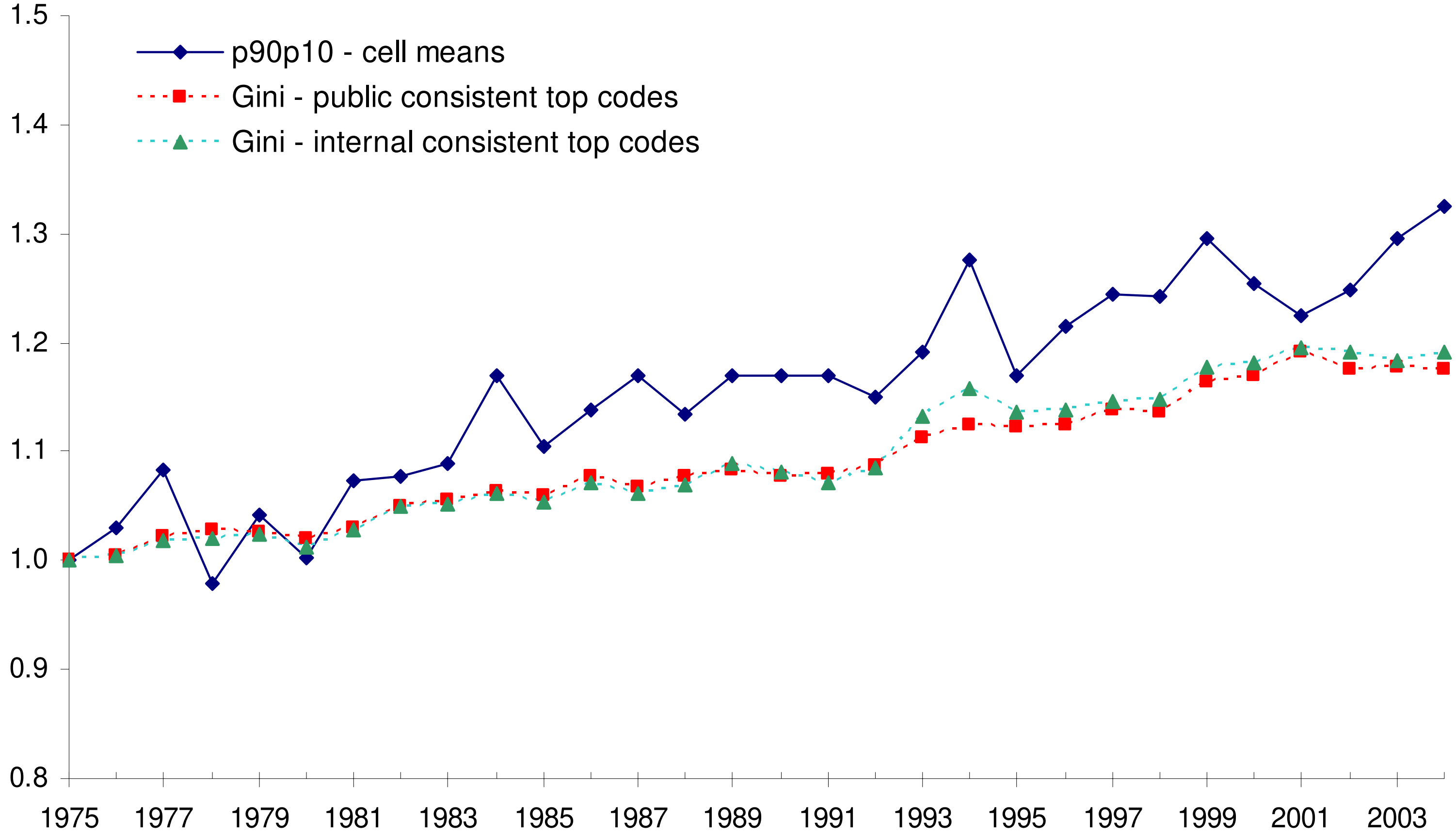


Figure 4. Trends in Consistently Top Coded Gini and Cell-mean adjusted P90/P10 Estimates for Wage and Salary Income of Full-time, Full-year Workers



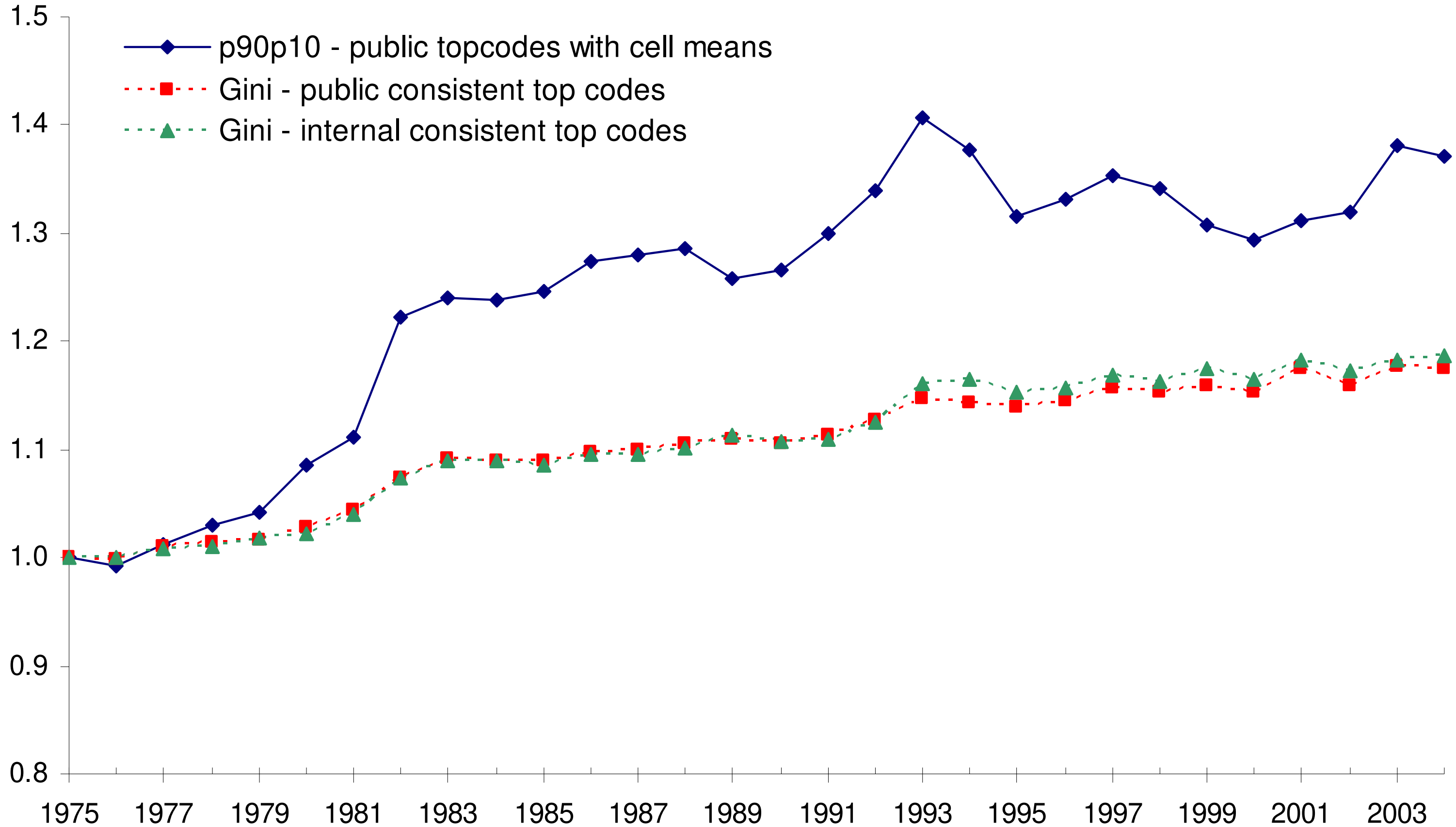
Note: The Gini and P90/P10 series are each normalized by their 1975 value.

Figure 5. Trends in Consistently Top Coded Gini Coefficient and Cell-mean adjusted P90/P10 Estimates for the Total Labor Earnings of Full-time, Full-year Workers



Note: The Gini and P90/P10 series are each normalized by their 1975 value.

Figure 6. Trends in Consistently Top Coded Gini Coefficient and Cell-mean adjusted P90/P10 Estimates for the Size-adjusted Household Income of Individuals



Note: The Gini and P90/P10 series are each normalized by their 1975 value.

Appendix Table 1. Public Use CPS Censoring Points for each Income Source in Dollars (1975–1986)

	INCWAG	INCSE	INCFRM	INCSS	INCSEC	INCPA	INCINT	INCDIV	INCALC	INCOMP	INCRET
1975	50,000	50,000	50,000	9,999	5,999	19,999	50,000	50,000	50,000	29,999	50,000
1976	50,000	50,000	50,000	9,999	5,999	19,999	50,000	50,000	50,000	29,999	50,000
1977	50,000	50,000	50,000	9,999	5,999	19,999	50,000	50,000	50,000	29,999	50,000
1978	50,000	50,000	50,000	9,999	5,999	19,999	50,000	50,000	50,000	29,999	50,000
1979	50,000	50,000	50,000	9,999	5,999	19,999	50,000	50,000	50,000	29,999	50,000
1980	50,000	50,000	50,000	9,999	5,999	19,999	50,000	50,000	50,000	29,999	50,000
1981	75,000	75,000	75,000	19,999	5,999	19,999	75,000	75,000	75,000	29,999	75,000
1982	75,000	75,000	75,000	19,999	5,999	19,999	75,000	75,000	75,000	29,999	75,000
1983	75,000	75,000	75,000	19,999	5,999	19,999	75,000	75,000	75,000	29,999	75,000
1984	99,999	99,999	99,999	19,999	9,999	19,999	99,999	99,999	99,999	29,999	99,999
1985	99,999	99,999	99,999	19,999	9,999	19,999	99,999	99,999	99,999	29,999	99,999
1986	99,999	99,999	99,999	19,999	9,999	19,999	99,999	99,999	99,999	29,999	99,999

Note: In the 1985 March CPS (income year 1984), six values for INCOMP exceeded $29,999 but were not top coded. In the calculations we did for this paper we corrected this error and top coded these values at $29,999.

Appendix Table 2. Public Use CPS Censoring Points for each Income Source in Dollars (1987–2004)

	INCER	INCWG1	INCSE1	INCFR1	INCSS	INCSEC	INCPA	INCINT	INCDV2	INCRNT	INCALM	INCHLD
1987	99,999	99,999	99,999	99,999	29,999	9,999	19,999	99,999	99,999	99,999	99,999	99,999
1988	99,999	99,999	99,999	99,999	29,999	9,999	19,999	99,999	99,999	99,999	99,999	99,999
1989	99,999	99,999	99,999	99,999	29,999	9,999	19,999	99,999	99,999	99,999	99,999	99,999
1990	99,999	99,999	99,999	99,999	29,999	9,999	19,999	99,999	99,999	99,999	99,999	99,999
1991	99,999	99,999	99,999	99,999	29,999	9,999	19,999	99,999	99,999	99,999	99,999	99,999
1992	99,999	99,999	99,999	99,999	29,999	9,999	19,999	99,999	99,999	99,999	99,999	99,999
1993	99,999	99,999	99,999	99,999	49,999	9,999	24,999	99,999	99,999	99,999	99,999	99,999
1994	99,999	99,999	99,999	99,999	49,999	9,999	24,999	99,999	99,999	99,999	99,999	99,999
1995	150,000	25,000	40,000	25,000	49,999	25,000	24,999	99,999	99,999	99,999	99,999	99,999
1996	150,000	25,000	40,000	25,000	49,999	25,000	24,999	99,999	99,999	99,999	99,999	99,999
1997	150,000	25,000	40,000	25,000	49,999	25,000	24,999	99,999	99,999	99,999	99,999	99,999
1998	150,000	25,000	40,000	25,000	49,999	25,000	24,999	35,000	15,000	25,000	50,000	15,000
1999	150,000	25,000	40,000	25,000	49,999	25,000	24,999	35,000	15,000	25,000	40,000	15,000
2000	150,000	25,000	40,000	25,000	49,999	25,000	24,999	35,000	15,000	25,000	40,000	15,000
2001	150,000	25,000	40,000	25,000	49,999	25,000	24,999	35,000	15,000	25,000	40,000	15,000
2002	200,000	35,000	50,000	25,000	49,999	25,000	24,999	25,000	15,000	40,000	45,000	15,000
2003	200,000	35,000	50,000	25,000	49,999	25,000	24,999	25,000	15,000	40,000	45,000	15,000
2004	200,000	35,000	50,000	25,000	49,999	25,000	24,999	25,000	15,000	40,000	45,000	15,000

Appendix Table 2. (Continued)

	INCUC	INCWCP	INCVET	INCRT1	INCRT2	INCSI1	INCSI2	INCDS1	INCDS2	INCED	INCONT	INCOTH
1987	99,999	99,999	29,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999
1988	99,999	99,999	29,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999
1989	99,999	99,999	29,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999
1990	99,999	99,999	29,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999
1991	99,999	99,999	29,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999
1992	99,999	99,999	29,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999
1993	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999
1994	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999
1995	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999
1996	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999
1997	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999
1998	99,999	99,999	99,999	45,000	45,000	50,000	50,000	35,000	35,000	20,000	30,000	25,000
1999	99,999	99,999	99,999	45,000	45,000	50,000	50,000	35,000	35,000	20,000	30,000	25,000
2000	99,999	99,999	99,999	45,000	45,000	50,000	50,000	35,000	35,000	20,000	30,000	25,000
2001	99,999	99,999	99,999	45,000	45,000	50,000	50,000	35,000	35,000	20,000	30,000	25,000
2002	99,999	99,999	99,999	45,000	45,000	50,000	50,000	35,000	35,000	20,000	30,000	25,000
2003	99,999	99,999	99,999	45,000	45,000	50,000	50,000	35,000	35,000	20,000	30,000	25,000
2004	99,999	99,999	99,999	45,000	45,000	50,000	50,000	35,000	35,000	20,000	30,000	25,000

Appendix Table 3. Internal CPS Censoring Points for each Income Source in Dollars (1975–1986)

	INCWAG	INCSE	INCFRM	INCSS	INCSEC	INCPA	INCINT	INCDIV	INCALC	INCOMP	INCRET
1975	99,999	99,999	99,999	9,999	9,999	19,999	99,999	99,999	99,999	99,999	99,999
1976	99,999	99,999	99,999	9,999	9,999	19,999	99,999	99,999	99,999	99,999	99,999
1977	99,999	99,999	99,999	9,999	9,999	19,999	99,999	99,999	99,999	99,999	99,999
1978	99,999	99,999	99,999	9,999	9,999	19,999	99,999	99,999	99,999	99,999	99,999
1979	99,999	99,999	99,999	19,999	9,999	19,999	99,999	99,999	99,999	99,999	99,999
1980	99,999	99,999	99,999	19,999	9,999	19,999	99,999	99,999	99,999	99,999	99,999
1981	99,999	99,999	99,999	19,999	9,999	19,999	99,999	99,999	99,999	99,999	99,999
1982	99,999	99,999	99,999	19,999	9,999	19,999	99,999	99,999	99,999	99,999	99,999
1983	99,999	99,999	99,999	19,999	9,999	19,999	99,999	99,999	99,999	99,999	99,999
1984	99,999	99,999	99,999	19,999	9,999	19,999	99,999	99,999	99,999	99,999	99,999
1985	250,000	250,000	250,000	19,999	9,999	19,999	99,999	99,999	99,999	99,999	99,999
1986	250,000	250,000	250,000	19,999	9,999	19,999	99,999	99,999	99,999	99,999	99,999

Appendix Table 4. Internal CPS Censoring Points for each Income Source in Dollars (1987–2004)

	INCER	INCWG1	INCSE1	INCFR1	INCSS	INCSEC	INCPA	INCINT	INCDV2	INCRNT	INCALM	INCHLD
1987	299,999	99,999	99,999	99,999	29,999	9,999	19,999	99,999	99,999	99,999	99,999	99,999
1988	299,999	99,999	99,999	99,999	29,999	9,999	19,999	99,999	99,999	99,999	99,999	99,999
1989	299,999	99,999	99,999	99,999	29,999	9,999	19,999	99,999	99,999	99,999	99,999	99,999
1990	299,999	99,999	99,999	99,999	29,999	9,999	19,999	99,999	99,999	99,999	99,999	99,999
1991	299,999	99,999	99,999	99,999	29,999	9,999	19,999	99,999	99,999	99,999	99,999	99,999
1992	299,999	99,999	99,999	99,999	29,999	9,999	19,999	99,999	99,999	99,999	99,999	99,999
1993	999,999	999,999	999,999	999,999	49,999	25,000	24,999	99,999	99,999	99,999	99,999	99,999
1994	1,099,999	1,099,999	999,999	999,999	50,000	25,000	25,000	99,999	99,999	99,999	99,999	99,999
1995	1,099,999	1,099,999	999,999	999,999	50,000	25,000	25,000	99,999	99,999	99,999	99,999	99,999
1996	1,099,999	1,099,999	999,999	999,999	50,000	25,000	25,000	99,999	99,999	99,999	99,999	99,999
1997	1,099,999	1,099,999	999,999	999,999	50,000	25,000	25,000	99,999	99,999	99,999	99,999	99,999
1998	1,099,999	1,099,999	999,999	999,999	50,000	25,000	25,000	99,999	99,999	99,999	99,999	99,999
1999	1,099,999	1,099,999	999,999	999,999	50,000	25,000	25,000	99,999	99,999	99,999	99,999	99,999
2000	1,099,999	1,099,999	999,999	999,999	50,000	25,000	25,000	99,999	99,999	99,999	99,999	99,999
2001	1,099,999	1,099,999	999,999	999,999	50,000	25,000	25,000	99,999	99,999	99,999	99,999	99,999
2002	1,099,999	1,099,999	999,999	999,999	50,000	25,000	25,000	99,999	99,999	99,999	99,999	99,999
2003	1,099,999	1,099,999	999,999	999,999	50,000	25,000	25,000	99,999	99,999	99,999	99,999	99,999
2004	1,099,999	1,099,999	999,999	999,999	50,000	25,000	25,000	99,999	99,999	99,999	99,999	99,999

Appendix Table 4. (Continued)

	INCUC	INCWCP	INCVET	INCRT1	INCRT2	INCSI1	INCSI2	INCDS1	INCDS2	INCED	INCONT	INCOTH
1987	99,999	99,999	29,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999
1988	99,999	99,999	29,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999
1989	99,999	99,999	29,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999
1990	99,999	99,999	29,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999
1991	99,999	99,999	29,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999
1992	99,999	99,999	29,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999
1993	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999
1994	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999
1995	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999
1996	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999
1997	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999
1998	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999
1999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999
2000	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999
2001	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999
2002	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999
2003	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999
2004	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999

Appendix Table 5. Adjusted Cell-Mean P90/P10 and Consistently Top Coded Gini Coefficient Estimates

Year	P90/P10 wages	P90/P10 earnings	P90/P10 income	Internal CTC Gini wages	Internal CTC Gini earnings	Internal CTC Gini Income	Public CTC Gini wages	Public CTC Gini earnings	Public CTC Gini income
1975	4.00	4.27	6.15	0.30	0.32	0.35	0.28	0.31	0.34
1976	3.97	4.40	6.11	0.30	0.32	0.35	0.29	0.31	0.34
1977	4.08	4.63	6.23	0.30	0.33	0.36	0.29	0.32	0.34
1978	4.17	4.18	6.33	0.31	0.33	0.36	0.30	0.32	0.35
1979	3.97	4.45	6.41	0.31	0.33	0.36	0.30	0.32	0.35
1980	4.12	4.29	6.68	0.31	0.33	0.36	0.30	0.32	0.35
1981	4.00	4.58	6.84	0.31	0.33	0.37	0.30	0.32	0.35
1982	4.33	4.61	7.52	0.32	0.34	0.38	0.30	0.32	0.37
1983	4.29	4.65	7.63	0.32	0.34	0.38	0.31	0.33	0.37
1984	4.36	5.00	7.62	0.32	0.34	0.38	0.31	0.33	0.37
1985	4.44	4.72	7.67	0.32	0.34	0.38	0.31	0.33	0.37
1986	4.49	4.86	7.84	0.33	0.35	0.39	0.32	0.33	0.37
1987	4.40	5.00	7.87	0.33	0.34	0.39	0.32	0.33	0.37
1988	4.50	4.84	7.91	0.33	0.35	0.39	0.32	0.33	0.38
1989	4.66	5.00	7.74	0.34	0.35	0.39	0.32	0.33	0.38
1990	4.55	5.00	7.79	0.33	0.35	0.39	0.32	0.33	0.38
1991	4.57	5.00	8.00	0.33	0.35	0.39	0.32	0.33	0.38
1992	4.50	4.92	8.24	0.34	0.35	0.40	0.32	0.34	0.38
1993	4.58	5.09	8.65	0.35	0.37	0.41	0.33	0.34	0.39
1994	4.92	5.45	8.47	0.36	0.37	0.41	0.33	0.35	0.39
1995	5.00	5.00	8.10	0.36	0.37	0.41	0.34	0.35	0.39
1996	4.89	5.19	8.19	0.36	0.37	0.41	0.34	0.35	0.39
1997	5.00	5.32	8.33	0.36	0.37	0.41	0.34	0.35	0.39
1998	4.96	5.31	8.26	0.36	0.37	0.41	0.34	0.35	0.39
1999	5.00	5.54	8.05	0.37	0.38	0.41	0.35	0.36	0.39
2000	5.03	5.36	7.96	0.37	0.38	0.41	0.35	0.36	0.39
2001	5.07	5.24	8.08	0.38	0.39	0.42	0.36	0.37	0.40
2002	5.33	5.33	8.12	0.37	0.39	0.41	0.35	0.36	0.39
2003	5.26	5.53	8.50	0.37	0.38	0.42	0.35	0.36	0.40
2004	5.25	5.66	8.44	0.38	0.39	0.42	0.35	0.36	0.40

Appendix Table 6. Adjusted Cell-Mean P90/P10 and Consistently Top Coded Gini Estimates Relative to their 1975 values

Year	P90/P10 wages	P90/P10 earnings	P90/P10 income	Internal CTC Gini wages	Internal CTC Gini earnings	Internal CTC Gini Income	Public CTC Gini wages	Public CTC Gini earnings	Public CTC Gini income
1975	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1976	0.99	1.03	0.99	1.00	1.00	1.00	1.01	1.00	1.00
1977	1.02	1.08	1.01	1.02	1.02	1.01	1.03	1.02	1.01
1978	1.04	0.98	1.03	1.03	1.02	1.01	1.04	1.03	1.01
1979	0.99	1.04	1.04	1.05	1.02	1.02	1.05	1.03	1.02
1980	1.03	1.00	1.08	1.04	1.01	1.02	1.05	1.02	1.03
1981	1.00	1.07	1.11	1.05	1.03	1.04	1.05	1.03	1.04
1982	1.08	1.08	1.22	1.07	1.05	1.07	1.07	1.05	1.07
1983	1.07	1.09	1.24	1.08	1.05	1.09	1.08	1.05	1.09
1984	1.09	1.17	1.24	1.09	1.06	1.09	1.09	1.06	1.09
1985	1.11	1.10	1.25	1.09	1.05	1.09	1.10	1.06	1.09
1986	1.12	1.14	1.27	1.11	1.07	1.10	1.12	1.08	1.10
1987	1.10	1.17	1.28	1.10	1.06	1.10	1.11	1.07	1.10
1988	1.13	1.13	1.29	1.11	1.07	1.10	1.12	1.08	1.10
1989	1.17	1.17	1.26	1.13	1.09	1.11	1.13	1.08	1.11
1990	1.14	1.17	1.27	1.13	1.08	1.11	1.12	1.08	1.11
1991	1.14	1.17	1.30	1.12	1.07	1.11	1.12	1.08	1.11
1992	1.13	1.15	1.34	1.13	1.08	1.12	1.13	1.09	1.13
1993	1.15	1.19	1.41	1.18	1.13	1.16	1.16	1.11	1.15
1994	1.23	1.28	1.38	1.21	1.16	1.17	1.17	1.13	1.14
1995	1.25	1.17	1.32	1.20	1.14	1.15	1.18	1.12	1.14
1996	1.22	1.21	1.33	1.21	1.14	1.16	1.19	1.12	1.15
1997	1.25	1.24	1.35	1.20	1.15	1.17	1.19	1.14	1.16
1998	1.24	1.24	1.34	1.20	1.15	1.16	1.19	1.14	1.15
1999	1.25	1.30	1.31	1.24	1.18	1.17	1.22	1.16	1.16
2000	1.26	1.25	1.29	1.25	1.18	1.16	1.23	1.17	1.15
2001	1.27	1.23	1.31	1.26	1.20	1.18	1.26	1.19	1.18
2002	1.33	1.25	1.32	1.26	1.19	1.17	1.24	1.18	1.16
2003	1.31	1.29	1.38	1.25	1.18	1.18	1.24	1.18	1.18
2004	1.31	1.32	1.37	1.26	1.19	1.19	1.24	1.17	1.18

Appendix: Robustness of Regression-based Inequality Trends Analysis

This Appendix substantiates our claim at the end of Section 4 that results are similar when using the Gini coefficient series based on internal consistent top coded CPS data instead of the consistently-top coded public use CPS data. Using the internal data, the wage regression is:

$$\underset{}{\text{Index}} = \underset{(117)}{0.975} + \underset{(14.04)}{0.0105\,t} + \underset{(2.25)}{0.0239\,d} - \underset{(2.30)}{0.0024\,dt} + \underset{(1.85)}{0.0243\,u} - \underset{(0.77)}{0.0143\,du}$$

The regression for labor earnings using the internal data is:

$$\text{Index} = \underset{(88)}{1.001} + \underset{(10.04)}{0.0101\,t} - \underset{(0.30)}{0.0048\,d} - \underset{(3.36)}{0.0048\,dt} - \underset{(0.03)}{0.0006\,u} + \underset{(1.53)}{0.0385\,du}$$

and the regression for size-adjusted household income using the internal data is:

$$\text{Index} = \underset{(51)}{1.028} + \underset{(14.04)}{0.0159\,t} - \underset{(2.25)}{0.0264\,d} - \underset{(2.30)}{0.0091\,dt} - \underset{(1.85)}{0.0750\,u} + \underset{(0.77)}{0.0766\,du}$$

Performing a similar analysis comparing the trends in the Gini coefficient estimates using the internal consistently-top coded data and the public consistently-top coded data shows that there is no noticeable difference in trends between these two series. For the three regressions below, p is a dummy variable that controls for whether the series uses internal or public top codes ($p = 1$ if public and 0 otherwise). The other variables are the same as above. The regression for trends in the Gini coefficient for wages is:

$$\text{Index} = \underset{(207)}{0.999} + \underset{(18.62)}{0.0080\,t} + \underset{(0.37)}{0.0025\,p} + \underset{(0.03)}{0.00001\,pt} + \underset{(4.20)}{0.0320\,u} - \underset{(2.04)}{0.0220\,pu}$$

The regression for the trend in the Gini coefficient for labor earnings is:

$$\text{Index} = \underset{(226)}{0.996} + \underset{(13.56)}{0.0053\,t} + \underset{(0.22)}{0.0013\,p} + \underset{(0.35)}{0.0002\,pt} + \underset{(5.46)}{0.0379\,u} - \underset{(2.03)}{0.0200\,pu}$$

and the regression for the trend in the Gini coefficient for size-adjusted household income is:

$$\text{Index} = \underset{(149)}{1.001} \; + \underset{(11.30)}{0.0068\,t} \; - \underset{(0.02)}{0.0002\,p} \; + \underset{(0.15)}{0.0001\,pt} \; + \underset{(0.15)}{0.0016\,u} \; - \underset{(1.03)}{0.0154\,pu}$$

Endnotes

1. To gain access to the internal CPS data, two of us (Burkhauser and Feng) became Special Sworn Status researchers of the U.S. Census Bureau at the New York Census Research Data Center, Cornell University, in 2005.
2. For examples of the use of consistent top coding to control for time inconsistency in the public use CPS data, see *inter alia* Burkhauser *et al*. (2003–2004), Burkhauser *et al.* (2007), Feng *et al.* (2006), Gottschalk and Danziger (2005), and Karoly and Burtless (1995).
3. For every income source, we calculate a single mean value for all top coded values. But we do not provide cell-means for subgroups of the population defined by e.g. sex, race, and experience. In contrast, the Census Bureau provides cell means based on sex/race/work experience cells for labor earnings but only single cell means for non-governmental sources of non-labor incomes and they do not provide cell means at all for governmental sources of non-labor income. In addition, our series provides consistent cell-mean values for earlier years, something the Census Bureau has not provided to the research community yet.
4. For our analysis of full-time, full year workers' income from wage and salaries, we excluded individuals who had non-positive income from wage and salaries or whose primary source of labor earnings income was farm income or non-farm self-employment income. For our analysis of full-time, full year workers' income from labor earnings and our analysis of all individuals' size-adjusted household income, we allow non-positive values for specific income sources but assign a value of $1 if the sum of all these income sources is non-positive to avoid including negative

incomes in the any of our calculations of labor earnings or size-adjusted household income.

5. We follow common conventions in the household income inequality literature by assuming that household resources are equally shared among all members and by capturing the economies of scale in their consumption of available resources using the 'square root' equivalence scale. We suppose that $Y = X/M^{0.5}$, where X is unadjusted total household income, M is the number of individuals in the household, and Y is the adjusted household income. See e.g. Atkinson *et al.* (1995), Burkhauser *et al.* (2003–2004), and Karoly and Burtless (1995).