Using internal CPS data to reevaluate trends in labor-earnings gaps

The Current Population Survey provides data that are used to compare gaps in the labor earnings of women and men, people of different races, and people of different levels of education; this article presents a data series that uses cell means and more accurately measures gaps and trends in earnings than do other publicly available series

Richard V. Burkhauser and Jeff Larrimore

The Current Population Survey (CPS) is a large, nationally representative sample of households collected each month since 1942 by the U.S. Census Bureau.[1] This article focuses on data from the surveys conducted in March because the March survey includes an extensive income questionnaire. The data that are publicly available from the CPS are the primary tool used to investigate yearly trends in United States average labor earnings and their distribution. However, to protect the confidentiality of its respondents, the Census Bureau topcodes the highest values from each source of income that it collects when it reports the income in the public-use CPS data. Topcoding is the replacement of a datum representing part or all of a person's true income with a lower value. One of the challenges that topcoding presents for those using the public-use data to examine labor-earnings levels and trends is that the topcodes vary over time, which leads to artificial increases or decreases in earnings (when the term "earnings" appears alone in this article, it still refers to "labor earnings") at the top of the earnings distribution as different fractions of the population are subject to topcoding each year.[2] Although the public-use data are used extensively to measure the earnings gaps between men and women and Blacks and Whites,[3] until now little was known about how topcoding affects comparisons of labor earnings across these subsets of the population.[4]

This article finds that gaps between the earnings of men and women, Blacks and Whites, and people of various education levels are all sensitive to topcoding. Ratios of these earnings as well as trends in the gaps and ratios also are sensitive to topcoding. The article arrives at these findings by analyzing 1975–2007 CPS data and comparing the values of gaps and ratios obtained using the public-use CPS data with values found using the internal CPS data.

This article presents an extended cell mean series that will be explained in more detail in a later section. The earnings gaps calculated using the extended cell mean series in conjunction with public-use CPS data are found to closely approximate those obtained with the Census Bureau's internal CPS data. Additionally, this article finds that women, Blacks, and the less-educated are relatively worse off compared with men, Whites, and the more-educated, respectively, than previously reported using the public-use CPS data. Although the trends for all of the aforementioned earnings gaps are sensitive

The results and conclusions presented in this article are those of the authors and do not necessarily reflect the views of the U. S. Census Bureau. This article has been screened to ensure that no confidential data are disclosed.

Richard V. Burkhauser is a professor in the Department of Policy Analysis and Management at Cornell University. Jeff Larrimore is a Ph.D. candidate in the Department of Economics at Cornell University. E-mail: rvb1@cornell.edu, jhl42@cornell.edu

to topcoding, the impact that attempting to correct for topcoding has on trends differs by year.[5]

Calculating earnings gaps

To calculate gaps in earnings between men and women, between Blacks and Whites, and among people of various levels of education, this article examines the annual labor earnings from wages and salaries, self-employment, and farm earnings of full-time, full-year workers in the CPS.[6] Prior to 1987 these "earnings sources" were reported as three separate values. Since then a fourth source—primary labor earnings (regardless of source)—has been added. The earnings sources and their names in the public and internal CPS data files are listed in table A–1 of the appendix. Much of the previous work exploring earnings gaps between men and women, between or among races, and among people of various levels of education focuses solely on wage and salary earnings and excludes self-employment and farm earnings, primarily because of concerns about the accuracy of self-employment earnings in the CPS. However, as Theresa J. Devine demonstrates, earnings gap data are sensitive to the inclusion or exclusion of self-employment earnings since the earnings gap between men and women is larger among full-time self-employed workers than among full-time wage earners.[7] Because the aim is to compare groups of people on the basis of *all* their labor market earnings, farm and self-employment earnings must be included along with wages.

An additional detail to consider is whether to analyze annual earnings or to instead recalculate the statistics as weekly or hourly wages. For this article a choice has been made to use annual earnings. The results are similar no matter which of these three methods is used; however, since women tend to work fewer weeks per year, using a weekly or hourly measure does generate a slightly smaller earnings gap between men and women.[8]

Another question is how best to calculate group earnings when calculating earnings gaps. To limit the impact of outliers on the earnings gap between men and women, the Census Bureau uses median rather than mean earnings when reporting the earnings gap between men and women in its *Income, Poverty, and Health Insurance Coverage in the United States* series.[9] The Census Bureau does not calculate earnings gaps between people of different races or levels of education in this report. The gap in median earnings between men and women that is presented by the Census Bureau is regularly reproduced in factsheets by policy institutes and has been widely used as background information in the literature on the pay gap between men and women.[10] However, using median earnings comes at the cost of focusing only on the midpoint of the earnings distribution. As a result of the use of median earnings, if women make substantial gains compared with men at either tail of the distribution, a simple comparison of the median over time will probably understate these gains. Additionally, since earnings distributions are positively skewed in all years, mean earnings give relatively more weight than median earnings to changes in the upper tail of the distribution. So for researchers interested in this portion of the distribution, the mean is better able to capture differences between groups and changes over time. Because this article focuses on the upper tail of the distribution, where most topcoding occurs, it evaluates mean earnings, which better reflect changes occurring throughout the entire earnings distribution and are better able to capture the impact of topcoding on earnings gaps.

Despite these differences in calculating earnings gaps, the general trends in earnings gaps in the literature have generally been consistent. Most previous literature has found that the earnings gap between men and women was largely unchanged for much of the 20th Century. It was not until the 1980s that women made substantial gains. In the 1990s, however, these gains subsided and the gap remained stable for much of the decade.[11]

While the consensus among researchers is that the earnings gap between Blacks and Whites also has been shrinking, the timing of its decline differs greatly from the timing of the decline in the earnings gap between women and men. The earnings gap between Blacks and Whites declined rapidly from the mid-1960s until the middle of the 1970s before stagnating or increasing slightly through much of the 1980s.[12] There is some disagreement on the direction of the earnings gap between Blacks and Whites during the 1990s, with David Card and John E. DiNardo finding the gap more or less constant and Kenneth Couch and Mary C. Daly and Chinhui Juhn reporting a decline.[13] The next section of the article shows the sensitivity of such earnings trends to four methods of dealing with topcodes in the CPS data.

Topcoding CPS data

To protect the confidentiality of respondents, the Census Bureau topcodes each source of income that respondents report in the public-use CPS data. The full list of labor-earnings topcoding thresholds over time is presented in tables A–2 and A–3 of the appendix. In addition to

topcoding each income source in the March CPS, the Census Bureau topcodes earnings reported in CPSs from other months, such as the usual weekly earnings reported in the surveys filled out by outgoing rotation groups.[14] The further topcoding prevents researchers from obtaining additional earnings information from other questions in the CPS. Because topcodes vary over time, they can affect both the sizes of earnings gaps and their trends over time.

Prior to 1995, the Census Bureau simply replaced the value for each source of an individual's income that was topcoded with the level of income at the threshold for topcoding. Starting with 1995 data, the Census Bureau instead began replacing the income figure with a cell mean—the mean value of all topcoded data from the source of income in question. For labor earnings, each cell contains earnings figures from workers who are all of the same sex and race and who all either work both full time and year round or do not. Because the Census Bureau has not provided cell means retroactively for years prior to 1995, using the public-use CPS data without taking this major change in reported earnings values into account results in a sizable increase in measured earnings in 1995 and beyond. Hence, while the use of cell means starting in 1995 causes the public-use CPS data to conform better to the internal CPS data, not taking the improvement in measurement into account will overestimate actual increases in labor earnings from any year before 1995 to 1995 or any year after.[15]

Topcoding also has important implications for measuring the relative labor earnings of subsamples of the population and measuring gaps in earnings among subsamples. For example, if the distributions of labor earnings of women and men were identical, individuals' earnings in both groups would be topcoded at the same rate. So, topcoding would reduce the mean earnings of both men and women by the same percentage, leaving intergroup inequality unchanged.

However, if individuals in the two groups have different probabilities of being topcoded or if the mean suppressed labor earnings of those who are topcoded differ between the two groups, topcoding will influence the earnings gap measure. Because a larger percentage of women than men are below the topcoding threshold, women are less likely to be topcoded; it can be expected that topcoding will artificially raise the ratio of women's mean earnings to men's mean earnings, because the women's observed mean earnings will be less artificially depressed from the topcodes than those of men and hence will be closer to their true mean. Similar results will occur even if the probability of topcoding is the same across both groups, provided that the amount of suppressed earnings is higher for men than for women. The same holds for Blacks relative to Whites and those with less education relative to those with more education.

Prevalence of topcoding

Table 1 shows, for the trough year of each business cycle since 1975, the percentages of various groups of full-time, full-year workers who have had earnings from at least one source topcoded in the public-use CPS data.[16] The groups of people are organized by sex (men and women), race (Blacks and Whites), and level of education attained (less than a high school degree, a high school degree but no higher education, and education beyond high school). The three business cycles run from 1975 to 1982, from 1982 to 1992, and from 1993 to 2004. The method for selecting the starting points and endpoints of business cycles in this article has been chosen somewhat arbitrarily. Rather than define business cycles directly by changes in macroeconomic growth, this article uses troughs in income, which in general lag behind macroeconomic growth. Choosing slightly different trough years would not have a significant effect on this article's findings. Although it is not a trough year, 1992 is included in the table. As will be discussed in more detail later, Census Bureau data collection procedures were redesigned after 1992. This reduces the ability to compare 1992 data with 1993 data. So 1993 represents both the trough year of the 1993–2004 business cycle and the first year of the new procedures. Like 1992, the year 2007 is not a trough year, but it is included in the table because it is the most recent year for which data are available. The business cycles are measured from trough to trough.

As can be seen in table 1, although the percentage of people whose earnings are topcoded varies by sex, race, and level of education, the overall incidence of topcoding has increased greatly over the past 30 years for every group of workers in the table. For example, virtually no women or black full-time, full-year workers had topcoded labor earnings in 1975, but close to 1 percent of each group had topcoded earnings in 2007.

While topcoding has been rising among the earnings of men, women, Blacks, Whites, and people of all three levels of education, in any given year there are noticeable differences in topcoding rates among these groups. Because women's earnings are less likely to be topcoded than those of men, one expects to find a larger difference between men's observed labor earnings and their true mean labor earnings than one expects to find for women's observed

Table 1. Percentages of various groups of full-time, full-year workers whose labor earnings are topcoded, and ratios of selected percentages; by year, selected years,1975–2007

Year	Women	Men	Ratio	Blacks	Whites	Ratio	Less than a high school degree	High school degree	Education beyond high school	Ratio	Ratio
	(1)	(2)	(1)/(2)	(3)	(4)	(3)/(4)	(7)	(8)	(9)	(8)/(7)	(9)/(8)
1975	0.02	1.18	0.02	0.00	0.91	0.00	0.09	0.28	1.73	3.14	6.24
1982	.16	1.76	.09	.33	1.30	.26	.07	.34	2.18	4.70	6.44
1992	.39	2.98	.13	.37	2.22	.17	.22	.35	3.24	1.59	9.39
1993	.66	3.51	.19	.80	2.68	.30	.30	.56	3.78	1.91	6.70
2004	.57	2.23	.26	.61	1.84	.33	.31	.59	2.23	1.88	3.80
2007	.86	2.59	.33	.85	2.30	.37	.22	.64	2.66	2.84	4.18

SOURCE: Authors' calculations made by use of public and internal CPS data.

and true earnings. Correcting for topcoding should show that the gap between women's and men's earnings is wider than previously reported. For the same reasons, one can expect that correcting for topcoding will show that the gap between the earnings of Blacks and those of Whites is wider than previously reported and that the gap between the earnings of people with a high school degree or less and the earnings of those in higher education groups also is wider than previously reported.

As can be seen in the table, topcoding ratios also have changed over time. In 2007, women were topcoded 33 percent as much as men, up from only 2 percent as much in 1975. In 2007, Blacks were topcoded 37 percent as much as Whites, compared with 1975 when no Blacks were topcoded. On the whole, from 1975 to 2007 the less-educated showed larger increases in topcoding than did the more-educated. Hence, trends in earnings gaps between the sexes, between Blacks and Whites, and among people of varying levels of education are expected to be affected by topcoding.

Methods of managing topcoding problems

The issue of topcoding can be handled in various ways. A first approach—referred to for the purposes of this article as "Unadjusted Public Use"—is to simply ignore topcoding issues and use the unadjusted public-use CPS data as released by the Census Bureau. However, as discussed earlier, doing so will result in a series whose labor-earnings levels are suppressed prior to 1995, because of topcoding, and are much higher thereafter, primarily because of the Census Bureau's introduction of cell means in 1995. This shift to cell means in 1995 is further complicated by changes to topcoding thresholds made by the Census Bureau at the same time. For instance, the topcode for primary earnings rose from $99,999 to $150,000, thus reducing the share of full-time male workers whose primary labor earnings were topcoded from 3.93 percent to 1.35 percent, but the use of cell means increases the average reported primary labor earnings of those men who were still topcoded to $305,989.

A second approach—referred to as "No Cell Mean Public Use"—is to ignore the introduction of cell means into the public-use CPS data and to produce a labor-earnings series in which all topcoded values are assigned the value of the topcoding threshold, even those values which date from after the introduction of cell means in 1995. While this approach removes the large artificial jump in labor earnings due to the introduction of cell means in 1995, it does not address the problem of inconsistent changes in topcoding thresholds over time (such as the change in the primary labor earnings topcode from $99,999 in 1994 to $150,000 in 1995) or the variation in topcoding rates across groups within the U.S. population.[17]

A third approach, used by Richard V. Burkhauser, J. S. Butler, Shuaizhang Feng, and Andrew J. Houtenville for labor earnings and by Burkhauser, Couch, Houtenville, and Ludmila Rovba for household income, is to create a consistent topcode series—an approach referred to as "Consistent Topcode Public Use."[18] For each earnings source, this series finds the year in which the topcoding threshold cuts most deeply into the source's earnings distribution and then for every other year applies whatever topcoding threshold cuts into the source's earnings distribution by the same percentage. This approach is preferable to both the Unadjusted Public Use

approach and the No Cell Mean Public Use approach in that it consistently measures a given percentage of the distribution of the earnings from the source in question in all years of the study. However, this consistency over time in topcoding rates comes at the cost of losing information by topcoding a larger fraction of the population in almost every year. In this article, which analyzes labor earnings for full-time, full-year workers, the Consistent Topcode Public use approach cuts into the data by anywhere from 2.5 to 3.8 percent. The public-use CPS data reflect a cut (due to topcoding) that ranges from 0.6 to 2.7 percent, depending on the year.

Just as the existence of topcoding in the public-use CPS data can distort gaps in earnings and trends in earnings inequality across groups, increasing the fraction of the population that is topcoded can exacerbate the problem. Because more individuals are topcoded with the Consistent Topcode Public Use approach than they are in the public data, the observed mean labor earnings of each group within the population will be lower. But, because most of the people who are captured by the reduction in the topcodes are men, white, or more educated, using this approach will reduce the mean earnings of these groups more than it will reduce the mean labor earnings of women, Blacks or the less-educated. Hence, the Consistent Topcode Public Use method will consistently overestimate the mean earnings of workers with the former set of characteristics relative to workers with the latter characteristics by disproportionately excluding the top part of the labor-earnings distribution.

Given the limitations of consistent topcoding in providing a consistent comparison of the economic well-being of subpopulations, a new method for controlling for topcoding in the public-use CPS data is needed. As mentioned earlier, the Census Bureau began using cell means in 1995. Cell means from before 1995 are what is necessary to create an unbroken series that is based on cell means. Jeff Larrimore, Burkhauser, Feng, and Laura Zayatz have employed approximately the same method the Census Bureau used to create its cell means from 1995 onward in order to generate cell means that date back to 1975.[19] With these cell means, it is possible to create an unbroken cell-means-based data series that can be used with the public-use CPS data. The earnings distributions in this series better match those found in the internal CPS data for each of the population subgroups examined.

To create the extended cell mean series for each source of labor earnings, the population is divided by sex, race, and employment status, the same categories the Census Bureau uses to produce its cell means. The topcoded earnings value is then replaced with the weighted mean earnings—from the source of earnings in question—of all individuals with the same set of demographic characteristics for whom the source of earnings in question is topcoded in the public-use CPS data. To protect the confidentiality of respondents' identities, when fewer than 5 individuals are topcoded from an earnings source, those individuals' earnings are combined with the earnings of individuals from a similar earnings source in order to obtain a cell size of 5 or more and generate a cell mean. (This procedure for preserving confidentiality is the same as that used by the Census Bureau.)

Although this new approach for correcting the effects of topcoding—an approach referred to as "Cell Mean Public Use"—has significant advantages over consistent topcoding because it allows one to better understand changes at the high end of the earnings distribution, it still does not capture the full distribution. In addition to topcoding income in the public-use CPS data, the Census Bureau censors high-income values for each source of income in the internal CPS data. The full list of points beyond which labor earnings are not released internally—termed "censoring points" in this article—is reported in tables A–2 and A–3 of the appendix. Since the internal CPS data are censored, values at the very top of the distribution for each source of income cannot be observed in these data.[20] This poses a potential problem when creating a cell mean series for the public-use CPS data from the internal CPS data, because at best the trends in the series will match those found in the internal data from which the cell means are created. If changes in the censoring points in the internal CPS data affect earnings gaps, ratios, or trends in the Internal series, the same gaps, ratios, and trends will be affected in the Cell Mean Public Use Series.

While this is a limitation of the cell mean series in measuring the "true" trends in labor earnings, the problem is not as serious as it could be because the censoring points in the internal CPS data are much higher than the topcodes in the public-use CPS data. As a result, the fraction of individuals who are affected by censoring points is lower than the fraction affected by the public-use CPS topcodes. Thus, although some censoring does occur in the internal CPS data, the results calculated using the extended cell mean series with the public-use CPS data (that is, using the Cell Mean Public Use approach) are much closer to the results that would be obtained using data that consistently captures the full earnings distribution.

Additionally, the censoring points tend to be more stable than their counterparts used for the public-use CPS

data, the topcoding thresholds. Since the Census Bureau switched from reporting three sources of labor earnings to four sources in 1987, the only years in which changes were made to censoring points were 1992 and 1993.

Problems with data from the years 1992 and 1993 are not limited to the internal data. In 1993 the Census Bureau also implemented a substantial redesign of its collection procedures, a redesign that included the implementation of computer-assisted data collection.[21] The change in procedures increased the ability of the Census Bureau to observe earnings near the top of the distribution; since those high earnings are observed in the internal data but are topcoded in the public-use data, the use of internal data exacerbates the observed break in the series. Therefore, although the use of cell means with public-use CPS data allows for consistent trends before and after these years—trends that closely match the internal CPS data—researchers should take caution when using the cell mean series, or any CPS-based earnings series, to compare the year 1992 or any year before with the year 1993 or any year after.

Accuracy in capturing mean labor earnings

As was explained in the previous section, men's and women's mean labor earnings were calculated using four methods of dealing with topcoding. Each cell in panel 1 of table 2 is the ratio of a datum from one of the four series to its corresponding figure from the internal CPS data. There are separate columns for men and women. A ratio of 1.000 indicates that the method perfectly captures the mean earnings observed in the Internal data series. The lower the ratio, the more earnings are missed as a result of topcoding.

As can be seen when looking at the data for 2007, because of the cell means provided by the Census Bureau, the mean earnings of full-time, full-year male and female workers captured in the Unadjusted Public Use data since 1995 are very close to the mean earnings in the Internal data series. So, for people only interested in years since 1995 (the year cell means were first provided by the Census Bureau), the men's and women's earnings statistics in the Unadjusted Public Use data and the Cell Mean Public Use data come very close to matching the corresponding statistics in the Internal series.

But for those also interested in years prior to 1995, the Unadjusted Public Use data series is flawed because it does not provide cell means for earnings that are above the threshold for topcoding. Hence, its mean values are smaller for both men's and women's earnings. In contrast, the Cell Mean Public Use data provide yearly means very close to those from the Internal series for both men and women in all years back to 1975, coming within 0.2 percent of the internal mean values for both men and women in each of the trough years.

Unlike the Unadjusted Public Use and Cell Mean Public Use series, the No Cell Mean Public Use and the Consistent Topcode Public Use series understate the mean earnings of both men and women in *all* years. Additionally, the amount by which earnings are understated through the use of these series has grown over time. For example, the mean earnings that are calculated using the Consistent Topcode Public Use series understate the results in the Internal series by 4.9 percent for men and 0.2 percent for women in 1975. By 2007 the gap between the Consistent Topcode Public Use series and Internal series rises to 9 percent for men's earnings and 4 percent for women's earnings.

As is seen in panels 2 and 3 of table 2, the methods for managing topcoding have effects on the calculations of mean earnings of black and white workers and of workers with different levels of education that are similar to the methods' effects on the calculation of men's and women's earnings. Mean earnings computed using the Cell Mean Public Use series in all years or the Unadjusted Public Use series after 1995 closely match the mean earnings calculated using the Internal series. Use of the Consistent Topcode Public Use or the No Cell Mean Public Use series understates mean earnings (in relation to the Internal series), doing so more for white than for black workers and more for more highly educated workers than for less-educated workers.

Accuracy in capturing earnings gaps

Having shown that mean earnings of men, women, Blacks, Whites, and people of three levels of education are influenced by the height of topcoding thresholds, the article now focuses in this section on differences among the No Cell Mean Public Use, Consistent Topcode Public Use, Cell Mean Public Use, and Internal series in order to explain how topcoding affects earnings gaps. The Unadjusted Public Use series is excluded from further discussions because its data from prior to 1995 are identical to the No Cell Mean Public Use series and its data from 1995 onward are nearly identical to the Cell Mean Public Use series. In addition, the Unadjusted Public Use series has a clear artificial jump in 1995 that makes it inferior to either the No Cell Mean Public Use series or the Cell

Table 2. The ratio of mean labor earnings according to each of four publicly available data series to mean labor earnings according to internal CPS data, selected years, 1975–2007

Panel 1. Ratios involving the mean labor earnings of women and men

Year	No Cell Mean Public Use		Unadjusted Public Use		Consistent Topcode Public Use		Cell Mean Public Use	
	Women	Men	Women	Men	Women	Men	Women	Men
1975	1.000	0.986	1.000	0.986	0.998	0.951	1.000	1.000
1982	.998	.988	.998	.988	.993	.955	1.000	.999
1992	.992	.958	.992	.958	.988	.940	1.000	1.000
1993	.970	.914	.970	.914	.966	.901	.999	1.000
2004	.973	.929	1.001	1.000	.965	.902	1.001	1.000
2007	.970	.935	1.000	1.000	.960	.910	1.000	1.000

Panel 2. Ratios involving the mean labor earnings of Blacks and Whites

Year	No Cell Mean Public Use		Unadjusted Public Use		Consistent Topcode Public Use		Cell Mean Public Use	
	Blacks	Whites	Blacks	Whites	Blacks	Whites	Blacks	Whites
1975	1.000	0.988	1.000	0.988	0.998	0.957	1.000	1.000
1982	.997	.990	.997	.990	.989	.962	1.000	.999
1992	.993	.966	.993	.966	.990	.951	1.000	1.000
1993	.961	.927	.961	.927	.957	.916	1.000	1.000
2004	.978	.939	1.003	1.002	.972	.915	1.003	1.002
2007	.961	.944	1.001	1.002	.953	.921	1.001	1.002

Panel 3. Ratios involving the mean labor earnings of people of each of three levels of education

Year	No Cell Mean Public Use			Unadjusted Public Use		
	Less than a high school degree	High school degree	Education beyond high school	Less than a high school degree	High school degree	Education beyond high school
1975	0.999	0.994	0.982	0.999	0.994	0.982
1982	.999	.997	.986	.999	.997	.986
1992	.992	.993	.957	.992	.993	.957
1993	.966	.967	.915	.966	.967	.915
2004	.967	.970	.934	.982	.996	1.003
2007	.987	.973	.937	.994	.996	1.002

Year	Consistent Topcode Public Use			Cell Mean Public Use		
	Less than a high school degree	High school degree	Education beyond high school	Less than a high school degree	High school degree	Education beyond high school
1975	0.991	0.982	0.935	1.000	0.999	1.001
1982	.996	.987	.947	1.000	1.000	.999
1992	.989	.990	.938	.999	.999	1.000
1993	.964	.963	.902	.979	.989	1.006
2004	.964	.962	.908	.982	.996	1.003
2007	.982	.967	.913	.994	.996	1.002

SOURCE: Authors' calculations made by use of public and internal CPS data.

Mean Public Use series alone.

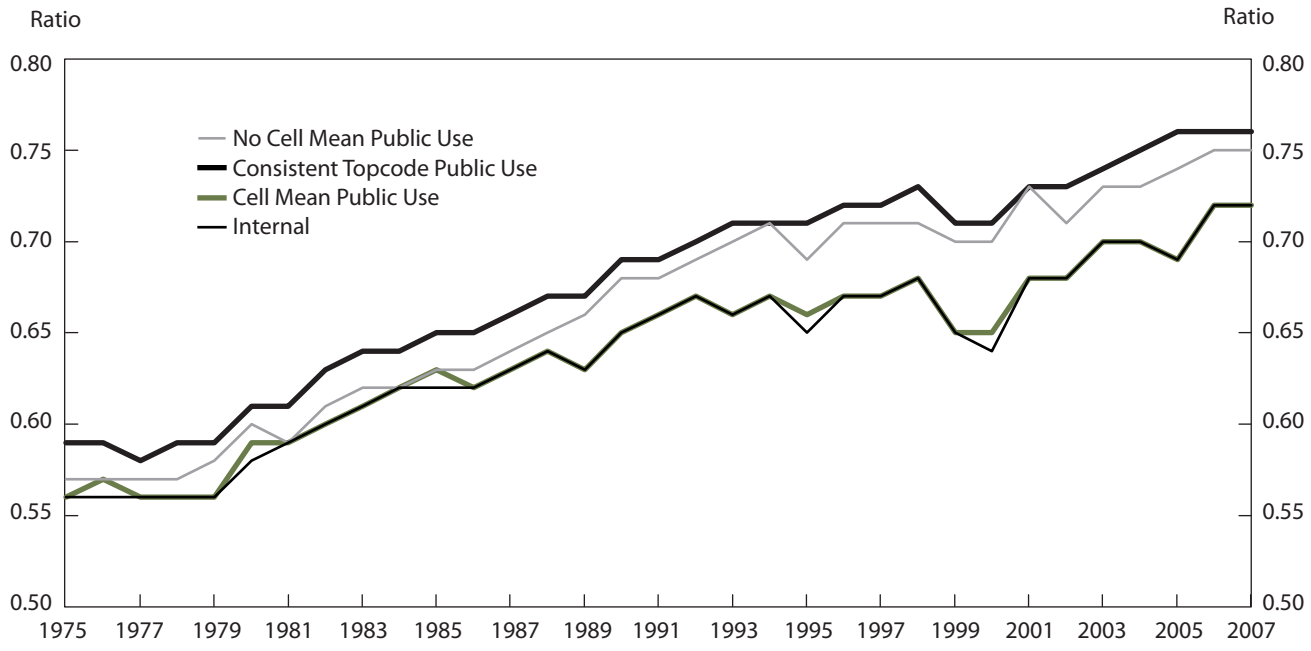
The gap in earnings between women and men. Because the No Cell Mean Public Use and Consistent Topcode Public Use series consistently understate the labor earnings of both men and women, the true ratio of women's earnings to men's earnings could in principal be greater or less than the ratio in the Cell Mean Public Use and Internal series. But as tables 1 and 2 have shown, men are more likely than women to be topcoded, and the average man who is topcoded has a higher wage or salary than the average woman who is topcoded. One therefore expects the ratio of women's earnings to men's earnings to be higher in the No Cell Mean Public Use and Consistent Topcode Public Use series than in the Cell Mean Public Use and Internal series, especially in the years for which cell means were not calculated.

The expectation proves to be true, as can be seen in chart 1, which compares the ratio of mean women's earnings to mean men's earnings as calculated using each of the four data series. In all years, the ratio of women's earnings to men's earnings is larger according to the No Cell Mean Public Use series than according to the Internal series. This difference is relatively small in the first year of the sample, but grows over time. In 1975 it was under 1 percentage point—female workers earned 56.6 percent of what male workers earned according to the No Cell Mean Public Use series, and they earned 55.8 percent of what male workers earned according to the Internal series—in 1989 it was over 2 percent, and in 2007 it was 2.8 percent. Thus, using the public-use CPS data without cell means will cause researchers to overstate the decline in the earnings gap between men and women over these years.

This overstatement is even greater when the Consistent Topcode Public Use method is used, since this approach further suppresses values at the top of the earnings distribution and topcodes even more men's earnings relative to women's earnings. Using consistent topcoding overstates the ratio of women's earnings to men's mean earnings by 2.8 percentage points in 1975, and the overstatement rises to 4.0 percentage points by 2007. In contrast, as can also be seen in chart 1, the Cell Mean Public Use series nicely approximates the women-to-men earnings ratios found using the internal CPS data.

The chart shows that the gap between the earnings ratio calculated using the No Cell Mean Public Use series and

Chart 1. Ratio of women's mean labor earnings to men's mean labor earnings, according to four data series, 1975–2007



SOURCE: Authors' calculations made by use of public and internal CPS data.

that calculated using the Internal series widens over time. The same happens for the Consistent Topcode Public Use series relative to the Internal series. Because of the widening of the gaps between the ratio calculated using the Internal series and the ratios calculated using the other two series, it might be assumed that using either of the other two series will overstate the earnings gains made by female workers relative to male workers for each of the three business cycles occurring during the 1975–2004 period. However, it will be shown that this is not the case.

Panel 1 of table 3 shows the percentage change in the ratio of women's mean earnings to men's mean earnings over each of the three business cycles that have occurred since 1975. As was done previously, direct comparisons across 1992–93 are excluded from the analysis because of the Census redesign.

When the years from 1975 to 2004 are grouped into the business cycles of 1975–82, 1982–92, and 1993–2004, one finds that in each of the three business cycles the percentage change calculated with the Cell Mean Public Use series closely matches that calculated with the Internal series. In contrast, both the Consistent Topcode Public Use and the No Cell Mean Public Use series understate the percentage change that occurred in the 1975–82 business cycle and, to a lesser extent, also understate the change that occurred during the 1993–2004 business cycle. However, for the 1982–92 business cycle, these two series overstate the relative earnings gains of women. Thus, while each of these two series slightly misstates the relative earnings gains of

Table 3. Percentage change in four ratios during the 1975–82, 1982–92, and 1993–2004 periods, according to four CPS data series

Panel 1. Percentage change in the ratio of women's mean labor earnings to men's mean labor earnings

Timespan	No Cell Mean Public Use	Consistent Topcode Public Use	Cell Mean Public Use	Internal
1975–1982	7.76	7.12	8.29	8.16
1982–1992	13.65	12.20	10.77	10.92
1993–2004	4.17	5.28	5.60	5.47

Panel 2. Percentage change in the ratio of Blacks' mean labor earnings to Whites' mean labor earnings

Timespan	No Cell Mean Public Use	Consistent Topcode Public Use	Cell Mean Public Use	Internal
1975–1982	1.60	0.55	2.20	2.14
1982–1992	3.04	2.32	.78	.90
1993–2004	4.51	–3.50	–4.87	–5.00

Panel 3. Percentage change in the ratio of the mean labor earnings of workers with a high school degree but no higher education to the mean labor earnings of wokers without a high school degree

Timespan	No Cell Mean Public Use	Consistent Topcode Public Use	Cell Mean Public Use	Internal
1975–1982	3.33	3.20	3.29	3.16
1982–1992	4.79	5.38	4.55	4.43
1993–2004	5.31	4.99	5.47	5.06

Panel 4. Percentage change in the ratio of the mean labor earnings of workers with education beyond high school to the mean labor earnings of workers with a high school degree but no higher education

Timespan	No Cell Mean Public Use	Consistent Topcode Public Use	Cell Mean Public Use	Internal
1975–1982	1.70	2.37	1.24	1.58
1982–1992	5.63	7.04	8.66	8.41
1993–2004	6.14	5.18	3.39	4.33

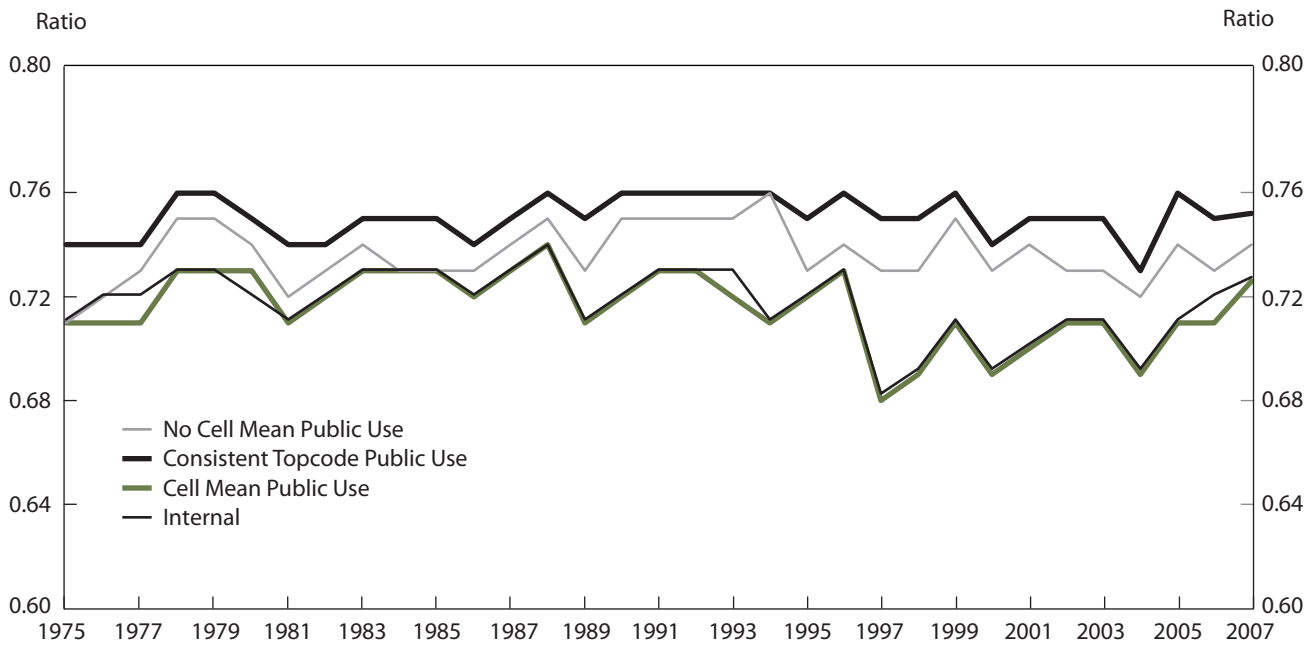
SOURCE: Authors' calculations made by use of public and internal CPS data.

women in all three business cycles, the direction of the misstatement is specific to the time period analyzed.

The gap in earnings between Blacks and Whites. Chart 2 shows the ratio of Blacks' mean earnings to Whites' mean earnings during the 1975–2007 period, according to the Internal series and each of the three methods of correcting for topcoding. Similar to the case of the ratio of women's mean earnings to men's mean earnings, using the No Cell Mean Public Use series overstates the relative earnings of black workers; the extent of this overstatement grows over time from 0.9 percentage points in 1975 to 2.9 percentage points in 2004 before falling back to 1.3 percentage points in 2007. In another parallel to the ratio of women's earnings to men's earnings, the Consistent Topcode Public Use series overstates the relative earnings of black workers by even more than the No Cell Mean Public Use series, as white workers are more likely to be near the top of the earnings distribution and thus have additional earnings suppressed by consistent topcoding. However, the earnings ratio calculated from year to year with the Cell Mean Public Use series again closely matches the ratio from the Internal series, and it is the best available method of replicating the earnings gap seen in the Internal series.

Panel 2 of table 3 displays the percentage change in the ratio of Blacks' mean earnings to Whites' mean earnings for each of the three business cycles. For every business cycle, the relationships among trends in the ratios of Blacks' mean earnings to Whites' mean earnings are similar to the relationships among trends in the ratios of women's mean earnings to men's mean earnings. Again, the Cell Mean Public Use series closely matches the trends in the Internal series for all three business cycles. Additionally, one also can see that during the 1975–82 business cycle, the Consistent Topcode Public Use and No Cell Mean Public Use series both slightly understate the relative *gain* in earnings made by black workers, as compared with the Internal series. For the 1993–2004 business cycle, the Consistent Topcode Public Use and No Cell Mean Public Use series understate the relative *decline* in Blacks' earnings in relation to Whites' earnings. For the 1982–92 business cycle the No Cell Mean Public Use and the Consistent Topcode Public Use series slightly overstate the earnings gains made by black workers. As was the case regarding men's and women's earnings, although these two series slightly misstate the percentage change in the ratio of Blacks' mean earnings to Whites' mean earnings,

Chart 2. Ratio of Blacks' mean labor earnings to White's mean labor earnings, according to four data series, 1975–2007



SOURCE: Authors' calculations made by use of public and internal CPS data.

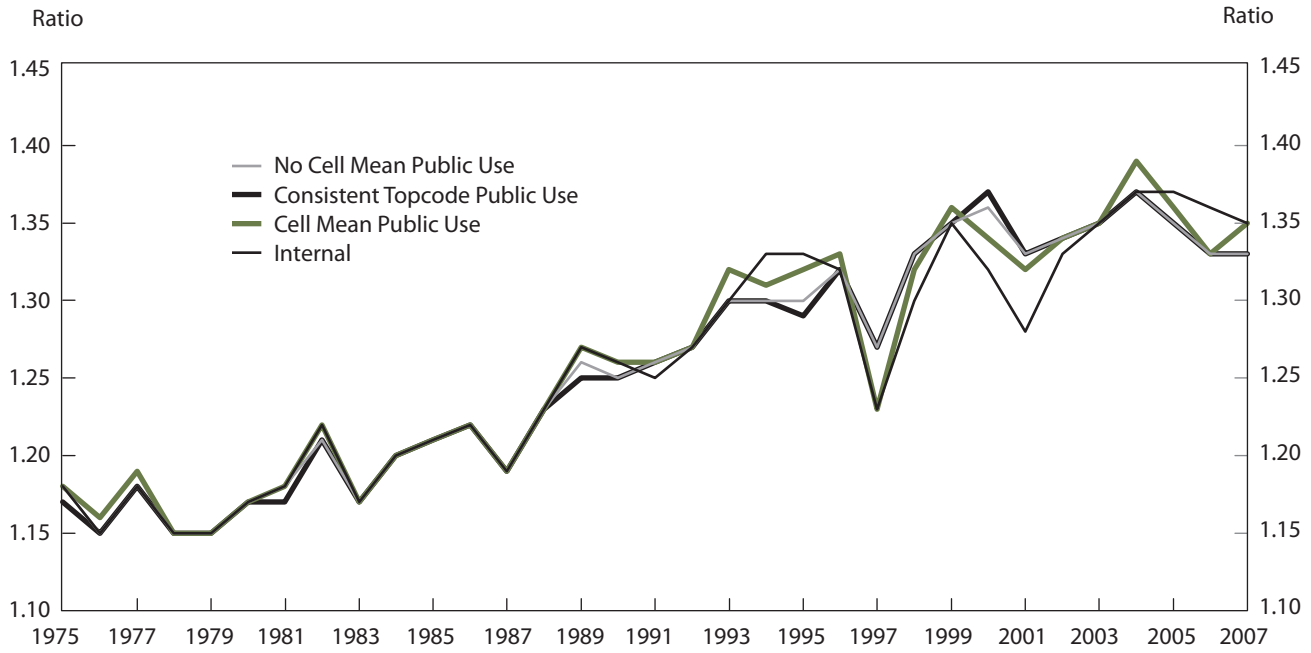
the direction of this misstatement varies over the three business cycles.

It may not come as a surprise that the Cell Mean Public Use series is nearly able to replicate the results from the Internal series in generating comparisons of women with men and Blacks with Whites, since sex and race were two of the conditioning criteria used when generating the cell means for each earnings source. Thus, a natural question is whether the Cell Mean Public Use approach is as successful at replicating the Internal series for subsets of the population that do not match the conditioning criteria.

Education mean earnings gaps. Mean earnings were calculated for the three levels of education previously mentioned: no high school degree, a high school degree but no higher education, and education beyond high school. For the 1975–2007 period, chart 3 displays the ratio of the mean earnings of workers with a high school degree but no higher education to the mean earnings of those without a high school degree. Chart 4 shows the ratio of the mean earnings of workers with education beyond high school to those of workers with only a high school degree. Both the charts present their respective ratios as calculated using data from the Internal series and each of the three methods of correcting for topcoding. In the creation of cells, level of education was not controlled for like sex and race were; therefore, the cells contain earnings figures from people of various levels of education. Nevertheless, as was seen with the earnings gaps between men and women and between Whites and Blacks, the "education earnings gaps" that are calculated using the Cell Mean Public Use series very closely match those calculated with the Internal series. Thus, it does not seem that the benefits of using cell means are confined to data calculated using the conditioning criteria of sex, race, and employment status.

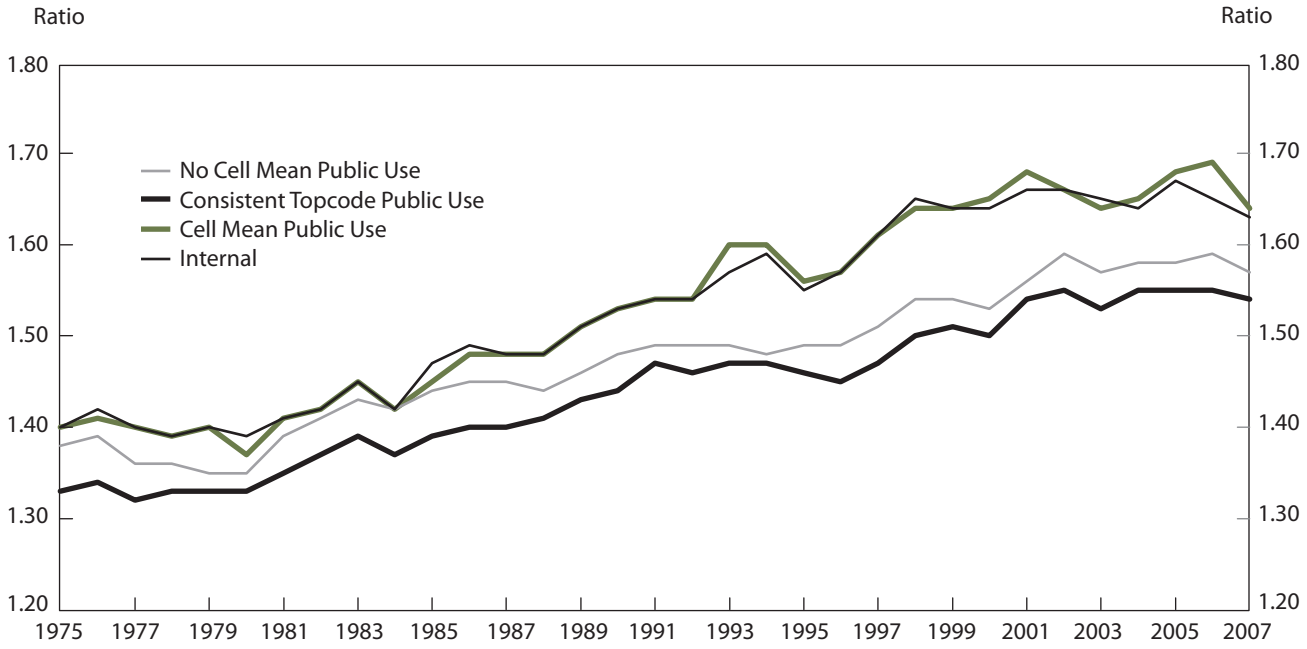
Additionally, this article finds that the degree to which labor earnings are understated when one uses the No Cell Mean Public Use or Consistent Topcode Public Use series increases with education because those with education beyond high school are more likely to have higher labor earnings and thus are more likely to have earnings suppressed by topcoding. Among the lower two education groups, there actually are some years in which the workers without a high school degree have earnings suppressed at a slightly higher rate than those with a high school degree, which causes the ratio of the mean

Chart 3. Ratio of the mean labor earnings of workers with a high school degree but no higher education to the mean labor earnings of workers without a high school degree, according to four data series, 1975–2007



SOURCE: Authors' calculations made by use of public and internal CPS data.

Chart 4. Ratio of the mean labor earnings of workers with education beyond high school to the mean labor earnings of workers with a high school degree but no higher education, according to four data series, 1975–2007



SOURCE: Authors' calculations made by use of public and internal CPS data.

earnings of the group with more education to the mean earnings of the group with less education to be higher in the No Cell Mean Public Use Series and the Consistent Topcode Public Use series than in the Internal series. In contrast, among the higher two education groups, in all years earnings are suppressed at a higher rate among those with some higher education than those with just a high school degree; therefore, not appropriately correcting for topcoding will lead to an understatement of the returns to higher education.

Panels 3 and 4 of table 3 present percentage changes in ratios of mean earnings for the business cycles of 1975–82, 1982–92, and 1993–2004, as calculated using data from the Internal series and the three other data series. The subject of panel 3 is the ratio of the mean earnings of workers with a high school degree but no higher education to the mean earnings of workers without a high school degree; the subject of panel 4 is the ratio of the mean earnings of workers with education beyond high school to those of workers with a high school degree but no higher education. Panels 3 and 4 take the same approach as panels 1 and 2 except that in panels 3 and 4, the ratio is of the group with the higher earnings to the group with the lower earnings. (The ratio is the other way around in panels 1 and 2).

In each of the first two business cycles, there is a similar pattern to that seen for the mean earnings ratios of women to men and Blacks to Whites: the percentage changes calculated using the Cell Mean Public Use series are quite similar to those calculated the Internal series. Considering all three business cycles, the No Cell Mean Public Use series and Consistent Topcode Public Use series are less accurate in capturing trends, but, as is the case in panels 1 and 2, the direction of the misstatement is not systematic; the percentage change is understated in some years and overstated in others.

In contrast to the findings concerning the earnings ratios of women to men and Blacks to Whites, in panels 3 and 4 the trends in data calculated using the Cell Mean Public Use series do not closely match the trends in data calculated using the Internal series in all three business cycles. In the 1993–2004 period, the Cell Mean Public Use series somewhat overstates the relative increase in the earnings of workers with a high school diploma (but no higher education) in relation to the earnings of workers without a high school diploma. This misstatement of the

trend occurs primarily because the cells do not control for education, thereby causing variations in how closely cell means represent the individual components of the cells. Nonetheless, in calculating the relative earnings of the lower two education groups, the Cell Mean Public Use series still approximates the Internal series better than do the other series.

For the 1993–2004 period the Cell Mean Public Use series somewhat understates the relative increase in the earnings of workers with some higher education in relation to workers with a high school diploma but no further education. Upon closer inspection, however, it can be seen that this understatement results mainly from the choice of 1993 as the first year in the timespan in question. In 1993 the difference (of 0.026) between the Internal and the Cell Mean Public Use series values for the earnings gap between those with some higher education and those with only a high school diploma is at its second largest amount over the entire 1975–2007 period. When 1994 is used as the base year, the Cell Mean Public Use values are much closer to the Internal series values. Thus, it is not that the Cell Mean Public Use series is unable to capture the trends in the Internal series in recent years, but rather that it does a poor job when 1993 is the anchor year.

TOPCODING IS A WELL-DOCUMENTED PROBLEM for the CPS, but until recently, the only available strategy for mitigating the problem has been to place further restrictions on the data, either by using consistent topcoding or by discarding the cell means provided by the Census Bureau from 1995 onward. As a result, calculations have tended to understate true mean earnings in the United States. When comparing earnings across two groups within the population that are topcoded at different rates, all previously available topcode correction schemes may lead to a misstatement of the earnings gap between the groups.

The authors of this article were able to partially lift the constraints of topcoding by obtaining access to the internal CPS data files. Although these internal data also are topcoded, the topcoding thresholds (censoring points) are substantially higher and more stable over time than those in the public-use CPS data. The key to this article is the extension of the cell mean series provided by the Census Bureau. The extension of cell means back to 1975 allows researchers using the public-use CPS data to estimate the earnings of individuals above the topcode threshold. Using the Cell Mean Public Use series with the public-use CPS data makes it possible to closely match the results found using internal CPS values from 1975 to 2007. Although the Cell Mean Public Use series best approximates the earnings statistics in the internal CPS data for groups based on race, sex, or employment status—because these characteristics are controlled for in the creation of cells—the cell mean series also is very useful for approximating the internal data for groups formed on the basis of other criteria, such as education level. Since the Cell Mean Public Use series is now available to the general public, researchers who are interested in exploring not just trends in earnings gaps and ratios but also more detailed questions about the underlying causes of gaps in pay can use the series to answer their questions with a precision similar to that obtained with access to the internal CPS files.

For this article, four data series were used to calculate earning gaps between women and men, between Blacks and Whites, and among people of three levels of education—all who worked full time year round. Using the Cell Mean Public Use series resulted in earnings gaps that, on the whole, were moderately larger than those calculated using the No Cell Mean Public Use series. According to the public-use data without cell means, in 2007 the mean earnings of women who worked full time year round were 75.1 percent of those of their male counterparts. The figure drops to 72.3 percent when topcoding is accounted for through the use of cell means. Similarly, in 2007 the mean earnings of Blacks were 74.0 percent of those of Whites without the use of cell means, compared with 72.6 percent with the use of cell means. The largest change, however, occurs for groups based on educational attainment. For the year 2007, the mean earnings of workers with some postsecondary education were 64 percent more than the mean earnings of those with only a high school degree as calculated with data from the Cell Mean Public Use series, compared with 57 percent as calculated using the No Cell Mean Public Use series. Thus, the returns to higher education are understated substantially if cell means are not used.

Sizes of individual earnings gaps and trends in earnings gaps both are sensitive to the choice of method of correcting for topcoding. Ignoring cell means and the earnings of individuals above the topcoding thresholds will distort the measured trends in earnings ratios between women and men, between Blacks and Whites, and among groups of different levels of education. However, unlike the case of earnings gaps, the direction of the distortion is not consistent and is sensitive to the years chosen for calculating the trends. Using public-use data without cell means will overstate relative changes in the earnings of women, Blacks, and the less-educated in some years but will understate relative changes in their earnings in other years. □

NOTES

ACKNOWLEDGMENTS: Support for the research in this article came from the National Science Foundation and the National Institute for Disability and Rehabilitation Research. The authors thank Lisa Marie Dragoset, 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.

[1] Each year the U.S. Census Bureau uses March CPS data to calculate yearly average income and poverty rates, and it releases these rates to the public; see **www.census.gov/prod/2008pubs/p60-235.pdf** (visited July 27, 2009) for more details. The March CPS data that the Census Bureau uses in its calculations are not available, except under certain conditions, to researchers outside of the Census Bureau.

[2] For an early review of this problem in the earnings-inequality literature, see Frank Levy and Richard J. Murnane, "U.S. Earnings Levels and Earnings Inequality: A Review of Recent Trends and Proposed Explanations," *Journal of Economic Literature*, September 1992, pp. 1333–81. For a more recent discussion see Shuaizhang Feng, Richard V. Burkhauser, and J.S. Butler, "Levels and Long-Term Trends in Earnings Inequality: Overcoming Current Population Survey Censoring Problems Using the GB2 Distribution," *Journal of Business and Economic Statistics*, January 2006, pp. 57–62.

[3] See, among other sources, Chinhui Juhn, Kevin M. Murphy, and Brooks Pierce, "Accounting for the Slowdown in Black-White Wage Convergence," in Marvin Kosters, ed., *Workers and their Wages* (Washington, DC, AEI Press, 1991); David Card and John E. DiNardo, "Skill-Biased Technological Change and Rising Wage Inequality: Some Problems and Puzzles," *Journal of Labor Economics*, October 2002, pp. 733–83; Kenneth Couch and Mary C. Daly, "Black-White Wage Inequality in the 1990s: a Decade of Progress," *Economic Inquiry*, January 2002, pp. 31–42; and Chinhui Juhn, "Labor Market Dropouts and Trends in the Wages of Black and White Men," *Industrial and Labor Relations Review*, July 2003, pp. 643–62.

[4] For a discussion of the impact of topcoding on the income gap between men with and without disabilities, see Richard V. Burkhauser and Jeff Larrimore, "Trends in the Relative Household Income of Working-Age Men with Work Limitations: Correcting the Record using Internal Current Population Survey Data," *Journal of Disability Policy Studies*, forthcoming article, see **http://dps.sagepub.com** (visited July 27, 2009).

[5] The research in this article was conducted while the authors were Special Sworn Status researchers of the U.S. Census Bureau at the New York Census Research Data Center at Cornell University. The article was completed while Richard V. Burkhauser was a Visiting Scholar at the American Enterprise Institute.

[6] In order to reduce the impact of changes in hours worked on the analysis of labor earnings, the sample used in this analysis is restricted to individuals over the age of 15 who work full time (35 hours or more per week) and year round (50 or more weeks per year). The Census Bureau uses the same restrictions for their annual analysis of earnings. (See page 10 of **www.census.gov/prod/2008pubs/p60-235.pdf**.) For this article, the sample is restricted also to individuals who are not in the military and do not reside in group quarters. These additional restrictions do not substantially affect the results.

[7] Theresa J. Devine, "Characteristics of self-employed women in the United States," *Monthly Labor Review*, March 1994, pp. 20–34.

[8] Francine D. Blau, and Lawrence M. Kahn, "Gender Differences in Pay," *Journal of Economic Perspectives*, Fall 2000, pp. 75–99.

[9] Carmen DeNavas-Walt, Bernadette D. Proctor, and Jessica Smith, *Income, Poverty, and Health Insurance Coverage in the United States: 2006*, Current Population Reports P60-233 (U.S. Census Bureau, 2007).

[10] See "The Paycheck Fairness Act: Helping to Close the Gap for Women," National Women's Law Center, 2006, on the Internet at **www.pay-equity.org/PDFs/PaycheckFairnessActApr06.pdf** (visited July 27, 2009); and "The Gender Wage Ratio: Women's and Men's Earnings," Institute for Women's Policy Research, IWPR # C350, 2008, on the Internet at **www.iwpr.org/pdf/C350.pdf** (visited July 27, 2009) for examples of policy factsheets that use data from the Census Bureau. See Blau and Kahn, "Gender Differences in Pay"; and June O'Neill, "The Gender Wage Gap, circa 2000," *American Economic Review: AEA Papers and Proceedings*, May 2003, pp. 309–14, for examples of using Census data for background information on the pay gap between men and women.

[11] Francine D. Blau and Lawrence M. Kahn, "Swimming Upstream: Trends in the Gender Wage Differential in the 1980s," *Journal of Labor Economics*, January 1997, pp. 1–42; Card and DiNardo, "Skill-Biased Technological Change and Rising Wage Inequality"; and O'Neill, "The Gender Wage Gap, circa 2000."

[12] Juhn and others, "Accounting for the Slowdown in Black-White Wage Convergence"; John Bound and Richard B. Freeman, "What Went Wrong? The Erosion of Relative Earnings and Employment Among Young Black Men in the 1980s," *Quarterly Journal of Economics*, February 1992, pp. 201–32.

[13] Card and DiNardo, "Skill-Biased Technological Change and Rising Wage Inequality"; Couch and Daly, "Black-White Wage Inequality in the 1990s"; and Juhn, "Labor Market Dropouts and Trends in the Wages of Black and White Men."

[14] Outgoing rotation groups are groups of people who are in their fourth or sixteenth month as part of the sample. The survey of outgoing rotation groups contains questions on usual weekly and hourly earnings. However, unlike the income supplement in the March CPS, this survey does not contain detailed income questions asking about sources of income other than earnings.

[15] Feng and others, "Levels and Long-Term Trends in Earnings Inequality."

[16] Complete annual statistics on topcoding rates and income by group as well as earnings ratios for all years from 1975 to 2007 for both the public use and internal use are available on request from the authors.

[17] A common refinement to the No Cell Mean Public Use approach is to assign topcoded individuals earnings that are a fixed multiple of the topcoding threshold—usually between 1.3 and 1.5. (See, for example, Blau and Kahn, "Gender Differences in Pay."). While the addition of this refinement comes closer to capturing levels of earnings gaps, the trends are nearly identical to those seen in the No Cell Mean Public Use series, and the refinement does not account for changes in the distribution of earnings above the topcoding thresholds over time. For the sake of brevity, the results that were calculated through the use of this method are not included in this article, but they are available from the authors upon request.

[18] Richard V. Burkhauser, J.S. Butler, Shuaizhang Feng, and Andrew J. Houtenville, "Long term trends in earnings inequality: what the CPS

can tell us," *Economics Letter*s, February 2004, pp. 295–99; and Richard V. Burkhauser, Kenneth A. Couch, Andrew J. Houtenville, and Ludmila Rovba, "Income Inequality in the 1990s: Re-Forging a Lost Relationship," *Journal of Income Distribution*, Winter 2004, pp. 8–35.

[19] Jeff Larrimore, Richard V. Burkhauser, Shuaizhang Feng, and Laura Zayatz, "Consistent Cell Means for Topcoded Incomes in the Public Use March CPS (1975-2007)," *Journal of Economic and Social Measurement, 2008*, pp. 89–128.

[20] For a more detailed discussion of internal censoring, see Edward J. Welniak, "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, Internal Revenue Service, 2003); and Richard V. Burkhauser, Shuaizhang Feng, and Stephen Jenkins, "Using the P90/P10 ratio to measure U.S. inequality trends with the Current Population Survey: a view from inside the Census Bureau vaults," *The Review of Income and Wealth*, February 2009, pp. 166–85.

[21] For details on the redesign of the Census Bureau's collection procedures, see Paul Ryscavage, "A surge in growing income inequality?" *Monthly Labor Review*, August 1995, pp. 51–61; and Arthur F. Jones and Daniel H. Weinberg, *The Changing Shape of the Nation's Income Distribution*, Current Population Reports P60-204 (U.S. Census Bureau, 2000).

Appendix A–1. Sources of labor earnings that are reported in the Current Population Survey

Name	Name in public files	Name in internal files	Definition
	1975–86		
Wages and salaries	I51A	WSAL_VAL	Wages and salaries
Self-employment	I51B	SEMP_VAL	Earnings from self-employment
Farm	I51C	FRSE_VAL	Farm earnings
	1987–2007		
Primary earnings	ERN_VAL	ERN_VAL	Primary earnings
Wages and salaries	WS_VAL	WS_VAL	Wages and salaries—second source
Self-employment	SE_VAL	SE_VAL	Self-employment earnings—second source
Farm	FRM_VAL	FRM_VAL	Farm earnings—second source

SOURCES: Current Population Survey Annual Demographic File Technical Documentation, 1976–2002; Current Population Survey Annual Social and Economic Supplement Technical Documentation, 2003–08.

Appendix A–2. Topcoding thresholds used for public CPS data and those used for internal data, by earnings source, selected years, 1975–86

Year or years	Topcoding thresholds used for public data			Topcoding thresholds used for internal data		
	Wages and salaries	Self-employment	Farm earnings	Wages and salaries	Self-employment	Farm earnings
1975–80	50,000	50,000	50,000	99,999	99,999	99,999
1981–83	75,000	75,000	75,000	99,999	99,999	99,999
1984	99,999	99,999	99,999	99,999	99,999	99,999
1985–86	99,999	99,999	99,999	250,000	250,000	250,000

SOURCES: The topcoding thresholds used for public data come from Current Population Survey Annual Demographic File Technical Documentation. The topcoding thresholds used for internal data come from the authors' calculations, which were made by use of internal CPS data.

Appendix A–3. Topcoding thresholds used for public CPS data and those used for internal data, by income source, selected years, 1987–2007

Year or years	Topcoding thresholds used for public data				Topcoding thresholds used for internal data			
	Primary earnings	Wages and salaries	Self-employment	Farm earnings	Primary earnings	Wages and salaries	Self-employment	Farm earnings
1987–92	99,999	99,999	99,999	99,999	299,999	99,999	99,999	99,999
1993	99,999	99,999	99,999	99,999	999,999	999,999	999,999	999,999
1994	99,999	99,999	99,999	99,999	1,099,999	1,099,999	999,999	999,999
1995–2001	150,000	25,000	40,000	25,000	1,099,999	1,099,999	999,999	999,999
2002–07	200,000	35,000	50,000	25,000	1,099,999	1,099,999	999,999	999,999

SOURCES: The topcoding thresholds used for public data come from the Current Population Survey Annual Demographic File Technical Documentation, 1987–2002, and from the Current Population Survey Annual Social and Economic Supplement Technical Documentation, 2003–08. The topcoding thresholds used for internal data come from the authors' calculations, which were made by use of internal CPS data.