USING INTERNAL CURRENT POPULATION SURVEY DATA TO REEVALUATE TRENDS IN LABOR EARNINGS GAPS BY GENDER, RACE AND EDUCATION LEVEL

by

Richard V. Burkhauser *
Cornell University

and

Jeff Larrimore *
Cornell University

CES 08-18   July, 2008

Abstract

Most empirical studies of trends in labor earnings gaps by gender, race or education level are based on data from the public use March Current Population Survey (CPS). Using the internal March CPS, we show that inconsistent topcoding in the public use data will understate these gaps and inaccurately capture their trends. We create a cell mean series beginning in 1975 that provides the mean of all values above the topcode for each income source in the public use March CPS and better approximate earnings gaps found in the internal March CPS than was previously possible using publically available data.

Introduction

The March Current Population Survey (CPS) is a large, nationally representative sample of households collected each March since 1942 by the U.S. Census Bureau.[1] The public use version of the March CPS is the primary data source used to investigate yearly trends in United States average labor earnings and its 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 them in the public use March CPS data. One of the challenges this presents for those using the public use March CPS to examine labor earnings levels and trends over time is that these topcodes are time-inconsistent, leading to artificial increases or decreases in earnings at the top of the income distribution as different fractions of the population are subject to topcoding each year. See Levy and Murnane (1992) for an early review of this problem in the earnings inequality literature. For a more recent discussion, see: Feng, Burkhauser, and Butler (2006). While the public use March CPS is used extensively to measure female-male and black-white earnings gaps (see, among others, Juhn, et al. 1991; Blau and Kahn 2000; Card and Dinardo 2002; Couch and Daly 2002; Juhn 2003), little is known about how topcoding impacts comparisons of labor earnings across these subsets of the population.[2]

In this paper we show that levels and trends in the female-male, black-white, and various education level ratios in labor earnings, or the earnings gaps between these groups, are sensitive to topcoding by comparing their values using alternative methods of controlling for topcoding in the public use March CPS to values found using the internal March CPS data between 1975 and 2006.

We find that the earnings gaps calculated using our extended cell mean series in conjunction with public use March CPS data closely approximate those obtained with the Census

Bureau's internal CPS data. Additionally, we find that women, blacks, and the less-educated are relatively worse off compared to men, whites, and the more-educated than previously seen using the public use CPS data. We also show that the time-trends for each of these earnings gaps are sensitive to topcoding, although the impact of topcoding corrections on trends differs based on the years examined.

Calculating the Earnings Gaps by Gender, Race, and Education Level

To calculate earnings gaps, we examine the annual labor earnings from wages, self-employment, and farm income of full-time, full-year workers in the March CPS.[3] Prior to 1987 these income sources are reported as three separate values. Since then a fourth source—primary labor income (regardless of source) has been added. The income sources and their names in the public and internal March CPS data files are listed in Appendix Table 1. Much of the previous work exploring earnings gaps by gender, race, or education level focuses solely on wage income and excludes self-employment and farm income, primarily because of concerns about the accuracy of self-employment income in the March CPS. However, as Devine (1994) demonstrates self-employment income is relevant to earnings gaps, since the gender earnings gap is larger among full-time self employed workers than among full-time wage earners. Since we are interested in comparisons of the full level of financial resources that individuals are generating for their households through their labor market work, we include all labor earnings in our analysis – including those from farm and self-employment.

An additional measurement detail is whether to analyze annual earnings or whether to re-scale the results to use weekly or hourly wages instead. While we use annual earnings, this choice should not greatly influence the results. However, since women tend to work fewer

weeks per year, using a weekly or hourly measure does generate a slightly smaller gender earnings gap (Blau and Kahn 2000).

A third measurement question is how best to calculate group earnings when calculating earnings gaps. To limit the impact of outliers on the female-male earnings gap, the Census Bureau uses median rather than mean earnings when reporting the gender earnings gap in their *Income, Poverty, and Health Insurance Coverage in the United States* series (U.S. Census Bureau, 2007a). The Census Bureau does not calculate earnings gaps by race or education in this report. The gender gap in median earnings presented by the Census Bureau is regularly reproduced in fact sheets by policy institutes (National Women's Law Center 2006; Institute for Women's Policy Research 2008) and has been widely presented as background information in the literature on the gender pay gap (Blau and Kahn 2000; O'Neill 2003). However, using median earnings comes at the cost of only using a single point in the income distribution. As a result, if women made substantial gains compared to men at either tail of the distribution, a simple comparison of the median over time would miss these gains that do not occur around the midpoint of the distribution. Since our focus is on the upper tail of the distribution where most topcoding occurs, we instead evaluate mean earnings, which will better reflect changes occurring throughout the entire earnings distribution.

Despite these differences in calculating the earnings gaps, the general trends in the earnings gap in the literature have generally been consistent. Most previous literature has found that the gender earnings gap was largely unchanged for much of the 20th Century until the 1980s, at which point women made substantial gains. In the 1990s, however, these gains subsided and the gap in earnings between women and men was stable for much of the decade (Blau and Kahn 1997; Blau and Kahn 2000; Card and DiNardo 2002; O'Neill 2003).

While the consensus opinion is that the black-white earnings gap has also been shrinking, the timing of its decline differs greatly from that observed for the female-male earnings gap. The black-white earnings gap declined rapidly from World War II until the middle of the 1970s before stagnating or increasing slightly through much of the 1980s (Juhn, et al. 1991; Bound and Freeman 1992; Card and Dinardo 2002; Couch and Daly 2002, Juhn 2003). In the 1990s, there is some disagreement on the direction of the black-white earnings gap, with Card and Dinardo (2002) finding the gap more or less constant while Couch and Daly (2002) and Juhn (2003) saw a return to a declining earnings gap. Below we show the sensitivity of these kinds of earnings trends to alternative methods of controlling for topcodes in the March CPS.

Topcoding in the March CPS

To protect the confidentiality of respondents, the Census Bureau topcodes each source of income respondents report in the public use March CPS survey. The full list of topcode values over time is presented in Appendix Tables 2 and 3. In addition to topcoding each income source in the main income supplement, the Census Bureau topcodes the usual weekly earnings in the Outgoing Rotation Group survey, preventing researchers from obtaining additional income information from other questions in the CPS surveys. Because these topcodes are time inconsistent they can impact both the level of the earnings gap and it trends over time.

In income year 1995 the Census Bureau began providing cell means for topcoded individuals in the public use March CPS – the mean income of all individuals who are topcoded from that topcoded source of income. Prior to 1995, the Census Bureau simply replaced the incomes of topcoded individuals with the topcode value. Since cell means were not provided

retroactively in years prior to 1995, using the public use March CPS data without taking this major change in reported income values into account results in a sizable increase in 1996 and beyond in their measured income due to more accurate reporting of their incomes since then. Hence, while the use of cell means after 1995 causes the public use March CPS to conform better to the internal March CPS, not taking this improvement in measurement into account will grossly overestimate actual increases in labor earnings after 1995 compared to prior years. (See Feng, Burkhauser, and Butler, 2006).

Topcoding has important implications for measuring the relative labor earnings, or earnings gap, of different subsamples of the population. For example, if the distribution of labor earnings of women and men were identical, individuals in both groups will be topcoded at the same rate. So, topcoding would reduce the mean incomes 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. If women are concentrated at lower earnings levels where they are less likely to be topcoded, we would expect topcoding to artificially raise the ratio of their mean earnings relative to men, because their observed mean earnings will be less artificially depressed from the topcodes than that 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 if the amount of suppressed earnings is higher for men than for women. Under these circumstances, we would find an artificial increase in the earnings gap between women and men resulting from the topcodes.

Prevalence of Topcoding by Gender, Race and Education Level

Table 1 shows the percentage of full time, full year workers in the public use March CPS topcoded for one or more sources of labor income. While its prevalence varies substantially across our gender, race, and education level groups, as can be seen in Table 1 topcoding is not only found but has increased substantially over the past 30 years for each group. For example, virtually no women (column 1) or black (column 4) full-time and full-year workers had topcoded labor earnings in 1975, but almost 1 percent of each had such topcodes in 2006.

While topcoding has been rising among all gender, race and education level groups, in any given year there are substantial differences in topcoding rates between these groups. Women are less likely to be topcoded than men (column 2). So, we expect differences between their observed and true mean labor earnings will be greater. Hence correcting for topcoding will show that women's labor earnings relative to men's are lower than previously reported.

Similarly, as can be seen in Table 1, topcoding is much less prevalent among blacks than whites (5). So, we expect differences between their observed and true mean labor earnings will be greater. Hence correcting for topcoding will show that blacks' labor earnings are relatively lower compared to whites' than previously reported.

Finally, as can be seen in Table 1 when exploring the returns to education by comparing those with at least some college education (9) to those with only a high school degree (8) or less (7), a greater portion of those with some higher education are topcoded. So, we expect differences between their observed and true mean labor earnings will be greater. Hence correcting for topcoding will show that the labor earnings of those with a high school degree or less are relatively lower than previously reported.

As can be seen in columns (3), (6), (10) and (11), the relative ratios of topcoding have also changed over time. In 2006, women were topcoded 35 percent as much as men,, up from only 2 percent as much in 1975. In 2006, blacks were topcoded 39 percent as much as whites, compared with 1975 when no blacks were topcoded. And it is also the case that the less-educated are topcoded at higher rates compared to the more-educated in 2006 than they were in 1975. Hence earnings gaps trends by gender, race and education level are also likely to be affected by topcoding.

Methods to Correct for Topcoding Problems

Topcoding in the public use March CPS data can be controlled in various ways. A first approach (Unadjusted Public Use) is to just use the unadjusted public use March CPS data as released by the Census Bureau. However, as discussed above, this will result in a series whose labor earnings levels are suppressed prior to 1995, due to topcoding, mixed with much higher earnings levels thereafter due, to some degree, to the Census Bureau's introduction of cell means in 1995. This shift to cell means in 1995 is further complicated by changes to the topcode level made by the Census Bureau at the same time. For instance, the topcode for primary earnings income rose from $99,999 to $150,000 thus reducing the share of full-time male workers who were topcoded on their own primary labor earnings from 3.93 to 1.35 percent, but the use of cell means increased the average reported primary labor earnings of those men who were still topcoded to $305,989.

A second approach (No Cell Mean Public Use) is to simply ignore the introduction of cell means into the public use March CPS, and produce a labor earnings series where all topcoded values are assigned a value at the topcode level even after the introduction of cell means in 1995.

While this will remove the large artificial jump in labor earnings due to the introduction of cell means in 1995, it does not address the basic problems of inconsistent topcode level changes over time (such as the change in primary labor earnings topcoding from $99,999 to $150,000 between 1994 and 1995) or the different topcoding rates across subgroups of the population.[4]

A third approach (Consistent Top Coded Public Use) is to create a consistent topcoded series. (See: Burkhauser, Butler, Feng, and Houtenville, 2004, for labor earnings and Burkhauser, Couch, Houtenville and Rovba, 2005, for household income.) For each income source, this series takes the topcode that cuts most deeply into that source's income distribution in a given year and then chooses a topcode value that cuts that deeply into that source's income distribution in all other years. This approach is preferable to either the Unadjusted Public Use data or the No Cell Public Use means data in that it consistently measures a given percentage of the income distribution of that income source in all years of the study. However, this time-consistency in topcode rates comes at the cost of losing information by topcoding a larger fraction of the population in all other years. In our case, where we are looking at labor earnings for full-time, full-year workers, the cut into the data using consistent topcoding ranges from 2.5 to 3.8 pecent. This compares to topcoding in the public use March CPS on all income sources for all individuals that ranges from 0.6 to 2.7 percent depending on the year.

Just as the existence of topcoding in the public use March CPS data can distort the levels and trends in earnings inequality across groups, increasing the fraction of the population who are topcoded can exacerbate the problem. Because more individuals are topcoded with this approach than in the public data, the observed mean labor earnings of each subgroup in the population will be lower. But, because most of the people who are captured by our reduction in the topcodes are men, white or more educated, using this approach will reduce their mean income more than it

will the mean labor earnings of women, blacks or the less educated. Hence we will consistently overestimate the mean income 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, we provide a new method for controlling for topcoding in the public use March CPS data. Using the internal March CPS data, we use approximately the same methodology the Census Bureau used to create its cell means after 1995 and extend the series back to 1975. With our cell means, which are now publicly available in Larrimore, Burkhauser, Feng, and Zayatz (2008), it is possible to create a consistent cell mean series that can be used with the public use March CPS, which better matches the income distributions found in the internal March CPS data for each of the population subgroups that we examine.[5]

While our cell mean approach has significant advantages over consistent topcoding because it allows us to better understand changes at the top of the income distribution, it does not capture the full distribution. In addition to topcoding income in the public use March CPS data, the Census Bureau censors high income values for each source of income in the internal March CPS data (See Welniak, 2003, Feng et al. 2006, and Burkhauser et al. 2007 for a fuller discussion). The full list of internal censoring points is reported in Appendix Tables 2 and 3. Since the internal March CPS data is censored, values at the very top of the distribution for each source of income will not be observed in these data. This poses a potential problem in creating a cell mean series for the public use March CPS from the internal March CPS data since at best it will match the trends found in the internal data from which the cell means are created. If changes

in the censoring points in the internal March CPS data result in inconsistencies, our cell mean series used with the public use March CPS data will retain those inconsistencies.

While this is a limitation of our cell mean series in measuring the "true" trends in labor earnings, the problem is mitigated because censoring points in the internal March CPS data are more stable than their public use March CPS counterparts. Since the Census Bureau switched from reporting three sources of labor earnings to four sources in 1987, the only changes in the internal March CPS censor levels occurred in 1992 and 1993. As a result, while there is a disconnect in the internal March CPS between these years, using our cell means with the public use March CPS allows for consistent trends before and after these years that closely match the internal March CPS data.

Additionally, since the censoring points in the internal March CPS data are much higher than the topcodes in the public use March CPS data, the fraction of individuals who are impacted by them is lower than the fraction impacted by the public use March CPS topcodes. Thus, while some censoring does occur in the internal March CPS data, the results we provide using the extended cell mean series with the public use March CPS data (Cell Means Public Use) are much closer to results that would be obtained using data that consistently captured the full income distribution. The additional information gained by using our cell means series with the public use March CPS justifies using the extended cell mean series despite the cost of accepting a trend-break in 1993 in our analysis.

Comparison of Mean labor earnings by Gender, Race, and Education Level

In Table 2 we compare the mean labor earnings of men and women who work full-time and full-year from 1975-2006 using our extended cell mean series together with the public use

March CPS data (Cell Mean Public Use) to those using the unadjusted public use March CPS data (Unadjusted Public Use), the public use March CPS data without cell means (No Cell Mean Public Use), the public use March CPS data consistently topcoded (Consistent Topcode Public Use), and the internal March CPS data used by the Census Bureau (Internal). For each series, the first column presents the mean labor earnings of women and the second column presents the mean labor earnings of men. The third column is the ratio of these two values. It measures the average economic well-being of women who work full-time and full-year relative to full-time, full-year male workers.

Thanks to Census Bureau provided cell means, the mean income of full-time, full-year male and female workers captured in the Unadjusted Public Use data since 1995 is very close to our Cell Mean Public Use means and both are very close to the mean values in the 2004 Internal data. So for those simply interested in comparing the relative income of those with and without work limitations in years since 1995 when cell means were first provided by the Census Bureau, the current Unadjusted Public Use data or our Cell Mean Public Use data nicely capture the mean earnings in the Internal data.

But for those interested in these series prior to 1995 the Unadjusted Public Use data is flawed because it does not provide cell means for persons above the topcoded values. Hence its mean values are smaller for both men and women. In contrast, our Cell Mean Public Use data provide yearly means very close to those from the Internal data for both men and women.

Because the Unadjusted Public Use series consistently understates the labor earnings of both men and women, the ratio of these two values could in principal be greater or less than the ratio in the Cell Mean Public Use and Internal series. But as we have shown in Table 1, men are more likely than women to be topcoded so their amount of suppressed labor earnings from

topcoding is likely to be greater. We therefore expect the ratio to be higher in the Unadjusted Public Use series than in the Cell Mean Public Use and Internal series in the years where cell means were not calculated. This is the case, as can be seen by comparing the ratios for 1975-1994 in the three series. In 1990, for example, the Unadjusted Public Use series ratio is 0.677, greater than 0.655 and 0.654 in the other two series. Since the Unadjusted Public Use series uses different topcoding adjustments before and after 1995, most yearly comparisons of before-1995 years to after-1995 years using this data series will show a greater decline in the relative labor earnings of women after 1995 than when using the more accurate Cell Means Public Use series.

In contrast to these three series, both 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. As predicted both series miss less of the labor earnings of men than women, so their female-male earnings ratios are always above those of our Cell Mean Public Use series and the Internal series. Finally, in 2006, researchers using the Consistent Topcoding Public Use series to calculate the gender earnings gap will understate the size of the earnings gap by over 4.3 percentage points relative to the Cell Mean Public Use and Internal series.

In Table 3, we provide a similar analysis comparing the labor earnings of blacks and whites under each of the four methods of controlling for topcoding. As was the case when examining how labor earnings differ between women and men, for researchers only interested in the relative labor earnings of blacks and whites since 1995 when the Census Bureau first provided cell means, using the Unadjusted Public Use series very closely approximate the mean earnings levels observed in the Cell Mean Public Use or Internal data. However, for researchers interested in earnings ratios prior to 1995, because whites are more likely to be topcoded the Unadjusted Public Use data will understate the mean earnings of white workers by a greater

13

percentage than it understates the mean earnings of black workers. As a result, prior to 1995 black workers appear comparatively better off than they actually are. Thus, when examining trends over time using the Unadjusted Public data, one will observe a dramatic artificial decline in the economic well being of black workers in 1995.

While the No Cell Mean Public Use data and Consistent Topcode Public Use series avoid this large artificial increase in the earnings gap between black and white workers in 1995, they also understate the earnings gap in each year since 1975. In contrast to these series which consistently understate the earnings of both races and overstate the relative earnings of black workers, the Cell Means Public Use series closely replicates the mean earnings levels of both races seen in the Internal data.

It should not come as a great surprise that the Cell Mean Public Use series based on our cell means used in conjunction with the public use March CPS is able to replicate the results from Internal data for comparisons across race and genders, since race and gender were two of the conditioning criterion used when generating the cell means for each income source. Thus, a natural question is whether our Cell Mean Public Use series is as successful at replicating the labor earnings obtained using Internal data for subsets of the population that do not match its conditioning criteria.

In Table 4 we compare labor earnings across different education levels, which were not conditioned for when calculating cell means. As was seen with the gender and racial earnings gaps, the mean labor earnings using the Cell Means Public Use series for those with less than a high-school degree, a high-school degree, and greater than a high-school degree very closely match the labor earnings observed using the Internal data. Thus, it does not seem that the

benefits of using cell means are confined to questions relating to the conditioning criteria of race, gender, and employment status.

Once again, after 1995 when cell means were first provided by the Census Bureau the Unadjusted Public Use series successfully captures trends in education level earnings ratios found in the Internal data. However, the Unadjusted Public Use series is less accurate for calculating prior earnings gaps and contains a large artificial increase in the earnings of individuals with education beyond high school in 1995, reflecting the change in how the Census Bureau reported the labor earnings of individuals with topcoded income.

While the No Cell Mean Public Use and Consistent Topcode Public Use series avoid this large artificial increase in the education earnings gap in 1995, they also understate the mean earnings of workers with all three levels of education when compared to both our Cell Mean Public Use series and the Internal data between individuals with a high-school degree and those with some higher education in each year since 1975. And, the degree to which labor earnings are understated increases with education levels since individuals with greater than a high school degree are more likely to have higher labor earnings and thus are more likely to have income suppressed by topcoding. As a result, the returns to receiving higher education that can be found using the Consistent Topcode Public Use or No Cell Public Use series are smaller than would be found using data with more complete information about the upper tail of the income distribution.

Comparison of Trends in Mean labor earnings by Gender, Race, and Education Level

In addition to exploring the impact of topcoding on levels of the female-male, black-white, and education earnings gaps, we also examined how topcoding impacts the time-trends for these earnings gaps. Unfortunately, because of the trend break in 1975 in the Cell Mean Public

Use and Internal series, we cannot observe trends across the entire 32 year sample. Instead, we look at the trends from 1975-1992 and 1993-2006 separately.

When looking at the trends we exclude the Unadjusted Public Use series since before 1995 it is identical to the No Cell Mean Public Use series and after 1995 it is nearly identical to our Cell Means Public Use series. Because the Unadjusted Public Use series is a combination of these two series, it cannot provide additional information about trends in the earnings gaps, and has a clear artificial jump in 1995 that makes it inferior to either of its component series individually. Therefore, we do not include this series in our analysis of trends in earnings ratios.

Figure 1 shows the trend in the female-male earnings ratio from 1975 through 1992 using each of the other four topcoding series: No Cell Mean Public Use, Consistent Topcode Public Use, Cell Mean Public Use, and Internal. In order to focus on the trends in the earnings ratios and not the levels, the earnings ratio in 1975 is normalized to 1. The gender earnings gap trends during this period are similar across all four series—women's relative earnings improved dramatically. However, the No Cell Mean Public Use series slightly overstates this gain over the period.

The trend in the gender earnings gap from 1993-2006 can be seen in Figure 2 with 1993 normalized to 1. As was the case prior to 1992, the Cell Mean Public Use and the Internal series have very similar trends. The Consistent Topcode Public Use series and the No Cell Mean Public Use series, however, have somewhat different trends. For example, the relative mean earnings of women is 0.711 that of men in 1993 using the Consistent Topcode Public Use series but is 0.663 in the Cell Mean Public Use series. By 2006 these values are 0.760, a rise of 6.93 percent in the No Cell Mean Public Use series, and 0.717, a rise of 8.15 percent in the Cell Mean Public Use series.

16

While the trends seem to differ across the series, the direction and magnitudes of the difference between the trends varies greatly based on the years chosen. This is evident from Figure 2, where the No Cell Mean Public Use series is below the Internal series from 1995-1998 and from 2002-2004, indicating that the No Cell Mean Public Use series is understating the gains of women since 1993. In 1999-2000, the Internal series is lower, indicating that the No Cell Mean Public Use series is overstating the gains of women since 1995. Thus, unlike the levels of the earnings gap where the No Cell Mean and the Consistent Topcode Public Use series consistently overstated the relative earnings of women, this is not always the case for their trends.

A similar analysis can be seen for the trend in the black-white earnings ratio in Figures 3 and 4. As was the case for the female-male earnings ratio, the trend in the black-white earnings ratio using the Cell Mean Public Use series closely matches those in the Internal series. The No Cell Mean Public Use and Consistent Topcode Public Use series, however, less accurately captures trends in the Internal data – especially in the years since 1989. While the trends differ based on the topcoding series used, there do not appear to be any clear patterns where one series consistently finds that black workers are improving at a substantially faster or slower pace than is seen in the other series. Instead, depending on the year chosen the No Cell Mean Series and Consistent Topcode Series sometimes understate and sometimes overstate the relative decline in relative earnings of black workers since 1993.

Comparing the earnings ratio between individuals with post-secondary education and individuals with just a high school degree, in most years prior to 1992 (Figure 5) a similar pattern emerges to that seen for the female-male and black-white earnings ratios. The trend when using the Cell Mean Public Use series closely matches the Internal series. The No Cell Mean Public

17

Use series and Consistent Topcode Public Use series are less accurate in capturing trends in the Internal data.

After 1992, the trend in the earnings gap between those with post-secondary education and those with only a high-school degree (Figure 6) shows a different pattern to those seen before or in the female-male or black-white earnings gap values in both time periods. In this period, the Cell Mean Public Use series consistently overstates the earnings gains compared to the earnings gains in the internal data. Thus, none of the three series using publicly available data closely match the trend found using the internal data. Upon closer inspection, however, this lack of accuracy using the Cell Means Public Use series after 1992 comes mainly from the choice of initial reference years. If 1994 rather than 1993 was the base year, the Cell Mean Public Use trend values would much more closely approximate the Internal series trend values. The reason is that the difference (0.026) between the Internal and the Cell Mean Public Use series values is greatest over the entire 32 year period in 1993. Thus, it is not that the Cell Mean Public Use series is unable to capture the trends in all years since 1993 in the Internal series, but rather that it does its poorest job of doing so when 1993 is the anchor year.

Conclusion

Even though topcoding is a well documented problem in the public use March CPS, until now the only available recourse to correct for time inconsistencies resulting from topcoding has been to place further restrictions on the data, either by using consistent topcoding or discarding the cell mean information provided by the Census Bureau after 1995. As a result, calculations have tended to understate true mean earnings in the United States. When comparing incomes across two groups within the population that are topcoded at different rates, all previously

18

available topcode correction schemes will therefore lead to a misstatement of the earnings gap between the groups.

We were able to partially lift these topcoding constraints by obtaining access to the internal March CPS data files. While this internal data is also topcoded, the topcodes are substantially higher and more stable over time than those in the public use March CPS data. As a result, we were able to evaluate the impact of topcoding not by reducing the amount of data examined, but by expanding available data above the public topcode limits. We do so by extending the cell mean series provided by the Census Bureau to allow researchers using the public use March CPS data to estimate the incomes of individuals above the topcode threshold. Using this cell mean series with the public use March CPS data, we are able to closely match the results found using internal March CPS values from 1975-2004. Although the cell mean series best approximates the earnings levels in the internal March CPS for groups based on the conditioning criteria used when calculating cell means, earnings levels in the internal data for groups formed on other criteria such as education level can also be well approximated using the cell means series. Since our extended cell mean series is now available to the general public, researchers interested in exploring not just the trends in earnings gaps and ratios, but also more detailed questions about the underlying causes of these pay differences can use this extended cell mean series to answer these questions with precision similar to those obtained with access to the internal March CPS files.

Using the Cell Mean Public Use series to evaluate the levels of the female-male, black-white, and education earnings gaps, we found that the earnings gaps are substantially larger than those observed using the No Cell Mean Public Use series. In 2006, when using public use data without cell means, the mean earnings of women who work full-time, full-year made 74.9

percent of that made by men, compared with just 71.7 percent of men's earnings once accounting for topcoding through the use of cell means. Similarly, in 2006 blacks mean earnings were just 71.2 percent of the mean earnings for whites when controlling for topcoding using cell means, compared with 73.4% without the use of cell means. And the mean earnings of individuals with some post-secondary education were 69 percent more than the mean earnings of individuals with just a high-school degree when cell means are used to control for topcoding, compared with 59 percent higher mean earnings when cell means are not used.

In addition to greatly differing levels for the earnings gaps, the trends in earnings gaps are also sensitive to the choice of topcode controls. We find that ignoring cell means and the earnings of individuals above the topcode thresholds will lead to an inaccurate perception of the trends in the female-male, black-white, and education earnings ratios. However, unlike what was seen for the levels of the earnings gap, the direction of this error in trends is not consistent across the years and is sensitive to the precise years chosen when calculating the trends. As a result, using public use data without cell means will overstate the relative increase in earnings of women, blacks, and the less-educated in some years, but will understate the relative increase in their earnings in other years.

ENDNOTES

[1] Each year the U.S. Census Bureau releases its yearly average income and poverty rates from the March CPS using these data (U.S. Census Bureau, 2007). As we will discuss in some detail below these official values are based on the internal March CPS data that is not available, except under certain conditions, to researchers outside of the U.S. Census Bureau.

[2] For a discussion of how topcoding impacts comparisons of income between males with and without work limitations, see Burkhauser and Larrimore (2008)

[3] In order to reduce the impact of labor supply decisions on our analysis of labor earnings, we follow the convention used by the U.S. Census Bureau in their annual analysis of earnings income and restrict our sample to individuals over the age of 15 who work full time (35+ hours per week) and full year (50+ weeks) for the income year (U.S. Census Bureau 2007). We also restrict our sample to individuals who are not in the military and do not reside in group quarters. Additionally, to report income consistently over time, all income has been adjusted to 2006 dollars using the CPI-U-RS reported by the Bureau of Labor Statistics.

[4] A common refinement on the No Cell Mean approach is to assign topcoded individuals income that is a fixed multiple of the topcode level—e.g. 1.3 to 1.5. (See: Blau and Kahn 2000). While this comes closer to capturing levels of earnings gaps, the trends are nearly identical to those seen in the No Cell Mean series and does not account for changes in the distribution of incomes above the topcode thresholds over time. For the sake of brevity we do not include the set of results using this method, but they are available from the authors upon request.

[5] In 2006 the Census Bureau granted us permission to use the internal March CPS to test the sensitivity of measured income inequality to alternative methods of providing additional information on topcoded persons in the public use March CPS. In addition, we were allowed to provide researchers without access to the internal March CPS data with this information, as long as in doing so, we do not unduly risk the confidentially of CPS respondents. In creating our extended cell mean series for each source of labor earnings, we divide the population by gender, race, and employment status, the same categories the Census Bureau uses to produce their cell means. We then replace the topcoded income value with the weighted mean-income of all individuals with the same set of demographic characteristics who are topcoded from that source of income in the public use March CPS. Like the Census Bureau, to protect the confidentiality of respondents, when less than five individuals are topcoded from an income source;

21

we combine those individuals with individuals from a similar source to obtain a cell-size of five or more to generate a cell mean. See Larrimore, Burkhauser, Feng, and Zayatz (2008) for a more detailed discussion of our extended cell mean series and the procedures we followed to protect the confidentiality of respondents.

REFERENCES

Blau, Francine D. and Lawrence M. Kahn. 2000. "Gender Differences in Pay." *Journal of Economic Perspectives* 14(4): 75-99

Blau, Francine D. and Lawrence M. Kahn. 1997. "Swimming Upstream: Trends in the Gender Wage Differential in the 1980s." *Journal of Labor Economics* 15(1): 1-42

Bound, John and Richard B. Freeman. 1992. "What Went Wrong? The Erosion of Relative Earnings and Employment Among Young Black Men in the 1980s." *Quarterly Journal of Economics* 107(1): 201-232

Burkhauser, Richard V., J.S. Butler, Shuaizhang Feng, and Andrew J. Houtenville. 2004. "Long Term Trends in Earnings Inequality: What the CPS Can Tell Us." *Economic Letters* 82 (2): 295-299.

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

Burkhauser, Richard V., Shuaizhang Feng, and Stephen Jenkins. 2007. "Using a P90/P10 Ratio to Measure Inequality Trends with the Public Use Current Population Survey: A View from Inside the Census Bureau Vaults." ISER WP 2007-14, June, 2007.

Burkhauser, Richard V and Jeff Larrimore. 2008. "Trends in the Relative Household Income of Working-Age Men with Work Limitations: Correcting the Record using Internal Current Population Survey Data." Center for Economic Studies Working Paper Series CES-WP-08-05, <http://www.ces.census.gov/index.php/ces/cespapers>.

Card, David and John E. DiNardo. 2002. "Skill-Biased Technological Change and Rising Wage Inequality: Some Problems and Puzzles." *Journal of Labor Economics* 20(4).

Couch, Kenneth and Mary C. Daly. 2002. "Black-white wage inequality in the 1990s: A decade of progress." *Economic Inquiry* 40(1): 31-42.

DeNavas-Walt, Carmen, Bernadette D. Proctor, and Jessica Smith, 2007. U.S. Census Bureau, Current Population Reports P60-233, *Income, Poverty, and Health Insurance Coverage in the United States: 2006*. U.S. Government Printing Office, Washington DC,

Devine, Theresa J. 1994. "Characteristics of Self-Employed Women in the United States." *Monthly Labor Review* 117(3): 20-34.

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.

Institute for Womens Policy Research. 2008. "The Gender Wage Ratio: Women's and Men's Earnings." IWPR # C350 <http://www.iwpr.org/pdf/C350.pdf>.

Juhn, Chinhui, Kevin M. Murphy, and Brooks Pierce. 1991. "Accounting for the Slowdiwn in Black-White Wage Convergence." in Marvin Kosters (Ed), *Workers and their Wages*, AEI Press, Washington DC.

Juhn, Chinhui. 2003. "Labor Market Dropouts and Trends in the Wages of Black and White Men." *Industrial and Labor Relations Review* 56(4): 643-662.

Larrimore, Jeff, Richard V. Burkhauser, Shuaizhang Feng and Laura Zayatz, 2008. "Consistent Cell Means for Topcoded Incomes in the Public Use March CPS (1975-2007)." NBER Working Paper 13941.

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.

National Women's Law Center. 2006. "The Paycheck Fairness Act: Helping to Close the Gap for Women." < http://www.pay-equity.org/PDFs/PaycheckFairnessActApr06.pdf>.

O'Neill, June. 2003. "The Gender Wage Gap, circa 2000." *American Economic Review: AEA Papers and Proceedings* 93(2): 309-314.

U.S. Census Bureau. 2007. "Current Population Survey Annual Social and Economic Supplement Technical Documentation." Washington DC: GPO.

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.

Table 1. Percentage of Full-Time and Full-Year Workers who are Topcoded on their Labor Earnings by Gender, Race, and Education Level (1987-2006).

Income Year	Female (1)	Male (2)	Ratio (1)/(2)	Black (3)	White (4)	Ratio (3)/(4)	Less than a High School Degree (7)	High School Graduate (8)	Education Beyond High School (9)	Ratio (8)/(7)	Ratio (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
1976	0.11%	1.41%	0.08	0.06%	1.13%	0.06	0.16%	0.24%	2.16%	1.50	8.84
1977	0.06%	1.63%	0.04	0.09%	1.30%	0.07	0.11%	0.27%	2.42%	2.38	8.99
1978	0.11%	2.20%	0.05	0.18%	1.74%	0.10	0.36%	0.41%	3.03%	1.13	7.44
1979	0.19%	2.75%	0.07	0.29%	2.12%	0.14	0.35%	0.59%	3.68%	1.67	6.26
1980	0.24%	3.37%	0.07	0.45%	2.54%	0.18	0.29%	0.79%	4.36%	2.74	5.55
1981	0.08%	1.27%	0.06	0.00%	0.98%	0.00	0.15%	0.22%	1.64%	1.45	7.33
1982	0.16%	1.76%	0.09	0.33%	1.30%	0.26	0.07%	0.34%	2.18%	4.70	6.44
1983	0.13%	1.90%	0.07	0.03%	1.45%	0.02	0.25%	0.34%	2.26%	1.35	6.64
1984	0.08%	1.05%	0.08	0.03%	0.80%	0.03	0.15%	0.19%	1.24%	1.31	6.47
1985	0.08%	1.18%	0.07	0.18%	0.89%	0.20	0.16%	0.10%	1.44%	0.65	14.09
1986	0.11%	1.59%	0.07	0.25%	1.21%	0.21	0.17%	0.16%	1.89%	0.95	11.50
1987	0.20%	1.49%	0.13	0.36%	1.12%	0.32	0.22%	0.27%	1.74%	1.23	6.54
1988	0.25%	1.83%	0.14	0.55%	1.34%	0.41	0.31%	0.32%	2.10%	1.05	6.47
1989	0.33%	2.53%	0.13	0.20%	1.94%	0.10	0.19%	0.44%	2.92%	2.30	6.67
1990	0.28%	2.49%	0.11	0.28%	1.87%	0.15	0.26%	0.36%	2.85%	1.35	7.99
1991	0.43%	2.53%	0.17	0.31%	1.95%	0.16	0.28%	0.41%	2.84%	1.48	6.96
1992	0.39%	2.98%	0.13	0.37%	2.22%	0.17	0.22%	0.35%	3.24%	1.59	9.39
1993	0.66%	3.51%	0.19	0.80%	2.68%	0.30	0.30%	0.56%	3.78%	1.91	6.70
1994	0.83%	3.98%	0.21	0.94%	3.13%	0.30	0.37%	0.65%	4.34%	1.75	6.67
1995	0.58%	2.21%	0.26	0.48%	1.80%	0.27	0.26%	0.59%	2.35%	2.25	3.97
1996	0.59%	2.27%	0.26	0.69%	1.88%	0.37	0.24%	0.63%	2.40%	2.69	3.78
1997	0.88%	2.74%	0.32	0.52%	2.31%	0.22	0.78%	0.66%	2.96%	0.85	4.47
1998	0.76%	2.83%	0.27	0.79%	2.32%	0.34	0.39%	0.64%	3.05%	1.62	4.80
1999	0.86%	3.38%	0.25	1.06%	2.83%	0.37	0.43%	0.78%	3.59%	1.80	4.62
2000	1.07%	3.58%	0.30	1.03%	3.01%	0.34	0.99%	0.74%	3.83%	0.75	5.17
2001	1.10%	3.94%	0.28	0.93%	3.30%	0.28	0.85%	1.02%	4.01%	1.20	3.92
2002	0.55%	2.30%	0.24	0.78%	1.85%	0.42	0.27%	0.64%	2.27%	2.36	3.57
2003	0.73%	2.29%	0.32	0.71%	1.96%	0.36	0.42%	0.68%	2.33%	1.63	3.42
2004	0.57%	2.23%	0.26	0.61%	1.84%	0.33	0.31%	0.59%	2.23%	1.88	3.80
2005	0.85%	2.48%	0.34	0.85%	2.21%	0.38	0.32%	0.55%	2.67%	1.72	4.85
2006	0.93%	2.62%	0.35	0.91%	2.34%	0.39	0.45%	0.57%	2.81%	1.27	4.93

Source: Author's calculations using public use March CPS data.

Table 2. Mean Labor Earnings and Earnings Gap among Full-Time, Full-Year workers by gender (1975-2006).

							Public Use									
	No Cell Mean			Unadjusted			Consistent Topcode			Cell Mean			Internal		
Income year	Female (1)	Male (2)	Ratio (1)/(2)	Female (4)	Male (5)	Ratio (4)/(5)	Female (7)	Male (8)	Ratio (7)/(8)	Female (10)	Male (11)	Ratio (10)/(11)	Female (13)	Male (14)	Ratio (13)/(14)
1975	26831	47512	0.565	26831	47512	0.565	26782	45797	0.585	26836	48177	0.557	26914	48233	0.558
1976	27487	48026	0.572	27487	48026	0.572	27390	46541	0.589	27547	48749	0.565	27540	48753	0.565
1977	27432	48409	0.567	27432	48409	0.567	27378	46965	0.583	27450	49449	0.555	27459	49278	0.557
1978	27861	48613	0.573	27861	48613	0.573	27803	47453	0.586	27891	49642	0.562	27889	49719	0.561
1979	27887	48105	0.580	27887	48105	0.580	27822	47292	0.588	27960	49655	0.563	27953	49480	0.565
1980	27799	46391	0.599	27799	46391	0.599	27759	45751	0.607	27871	47111	0.592	27869	47790	0.583
1981	27710	46702	0.593	27710	46702	0.593	27636	45222	0.611	27733	47205	0.588	27736	47203	0.588
1982	28571	46869	0.610	28571	46869	0.610	28428	45299	0.628	28627	47379	0.604	28627	47436	0.603
1983	28999	46681	0.621	28999	46681	0.621	28893	45233	0.639	29036	47316	0.614	29036	47317	0.614
1984	29519	47812	0.617	29519	47812	0.617	29362	45910	0.640	29519	47812	0.617	29519	47812	0.617
1985	30305	48177	0.629	30305	48177	0.629	30171	46426	0.650	30348	48543	0.625	30341	48997	0.619
1986	31292	49396	0.633	31292	49396	0.633	31131	47674	0.653	31376	50515	0.621	31380	50514	0.621
1987	31777	49457	0.643	31777	49457	0.643	31579	47799	0.661	31961	50768	0.630	31963	50800	0.629
1988	32236	49519	0.651	32236	49519	0.651	32053	48135	0.666	32441	50909	0.637	32458	50933	0.637
1989	32769	49900	0.657	32769	49900	0.657	32589	48504	0.672	32974	52033	0.634	32982	52086	0.633
1990	32567	48136	0.677	32567	48136	0.677	32411	46857	0.692	32807	50078	0.655	32795	50122	0.654
1991	32909	48183	0.683	32909	48183	0.683	32718	47142	0.694	33117	49793	0.665	33104	49835	0.664
1992	33514	48351	0.693	33514	48351	0.693	33395	47404	0.704	33791	50459	0.670	33791	50482	0.669
1993	33701	47889	0.704	33701	47889	0.704	33572	47211	0.711	34742	52387	0.663	34774	52402	0.664
1994	34240	48326	0.709	34240	48326	0.709	34174	47914	0.713	35297	53029	0.666	35340	53040	0.666
1995	34295	49550	0.692	34843	53217	0.655	33997	47684	0.713	34911	53182	0.656	34859	53221	0.655
1996	35187	49858	0.706	36333	53927	0.674	34797	48046	0.724	36326	53937	0.673	36323	53963	0.673
1997	35734	50525	0.707	36667	54840	0.669	35312	48745	0.724	36690	54828	0.669	36638	54857	0.668
1998	36713	51373	0.715	37886	55564	0.682	36386	49895	0.729	37914	55545	0.683	37887	55578	0.682
1999	36941	52918	0.698	37412	55767	0.671	36685	51553	0.712	37707	57637	0.654	37662	57693	0.653
2000	37321	53093	0.703	38250	59158	0.647	37090	51957	0.714	38248	59141	0.647	38203	59300	0.644
2001	38362	52825	0.726	40269	58863	0.684	38203	52096	0.733	40118	58952	0.681	40308	58858	0.685
2002	39097	54718	0.715	40118	58893	0.681	38690	52718	0.734	40118	58893	0.681	40152	58887	0.682
2003	39635	54502	0.727	40759	58251	0.700	39172	52827	0.742	40759	58251	0.700	40780	58266	0.700
2004	39293	53575	0.733	40416	57685	0.701	38959	52016	0.749	40416	57685	0.701	40373	57686	0.700
2005	39460	53438	0.738	40344	58088	0.695	39070	51742	0.755	40344	58088	0.695			
2006	39811	53180	0.749	41457	57800	0.717	39299	51684	0.760	41457	57800	0.717			

Source: Author's calculations using public and internal March CPS data.

Table 3. Mean Labor Earnings and Earnings Gap among Full-Time, Full-Year workers by race (1975-2006).

	Public Use												Internal		
	No Cell Mean			Unadjusted			Consistent Topcode			Cell Mean					
Income Year	black (1)	white (2)	ratio (1)/(2)	black (4)	white (5)	ratio (4)/(5)	black (7)	white (8)	ratio (7)/(8)	black (10)	white (11)	ratio (10)/(11)	Black (13)	White (14)	Ratio (13)/(14)
1975	30324	42363	0.716	30324	42363	0.716	30274	41028	0.738	30324	42876	0.707	30342	42970	0.706
1976	31043	42941	0.723	31043	42941	0.723	30928	41771	0.740	31098	43514	0.715	31137	43517	0.716
1977	31215	43045	0.725	31215	43045	0.725	31052	41927	0.741	31297	43845	0.714	31296	43715	0.716
1978	32153	43077	0.746	32153	43077	0.746	32028	42181	0.759	32232	43868	0.735	32194	43933	0.733
1979	31915	42666	0.748	31915	42666	0.748	31826	42038	0.757	32019	43858	0.730	32017	43731	0.732
1980	30512	41315	0.739	30512	41315	0.739	30454	40834	0.746	30582	41844	0.731	30565	42359	0.722
1981	29829	41578	0.717	29829	41578	0.717	29785	40452	0.736	29829	41959	0.711	29829	41959	0.711
1982	30253	41656	0.726	30253	41656	0.726	30002	40492	0.741	30336	42030	0.722	30342	42071	0.721
1983	30670	41614	0.737	30670	41614	0.737	30585	40537	0.754	30677	42087	0.729	30676	42086	0.729
1984	31070	42706	0.728	31070	42706	0.728	30934	41282	0.749	31070	42706	0.728	31070	42706	0.728
1985	31747	43356	0.732	31747	43356	0.732	31507	42043	0.749	31931	43620	0.732	31897	43971	0.725
1986	32385	44534	0.727	32385	44534	0.727	32169	43226	0.744	32567	45375	0.718	32567	45379	0.718
1987	32938	44581	0.739	32938	44581	0.739	32690	43323	0.755	33130	45576	0.727	33102	45606	0.726
1988	33357	44734	0.746	33357	44734	0.746	33062	43706	0.756	33689	45777	0.736	33705	45805	0.736
1989	33280	45407	0.733	33280	45407	0.733	33092	44341	0.746	33435	47021	0.711	33426	47045	0.711
1990	32899	44101	0.746	32899	44101	0.746	32756	43129	0.760	33059	45591	0.725	33045	45609	0.725
1991	33046	44025	0.751	33046	44025	0.751	32909	43229	0.761	33162	45232	0.733	33091	45254	0.731
1992	33208	44328	0.749	33208	44328	0.749	33108	43625	0.759	33424	45900	0.728	33406	45904	0.728
1993	33407	44461	0.751	33407	44461	0.751	33265	43932	0.757	34760	47980	0.724	34785	47971	0.725
1994	34070	45093	0.756	34070	45093	0.756	33968	44769	0.759	34890	48902	0.713	34889	48907	0.713
1995	33859	46220	0.733	34991	48937	0.715	33592	44759	0.751	35094	48932	0.717	34997	48889	0.716
1996	34524	46777	0.738	36488	49990	0.730	34259	45302	0.756	36478	49996	0.730	36430	49907	0.730
1997	34651	47550	0.729	34973	51059	0.685	34434	46080	0.747	34973	51062	0.685	34889	51101	0.683
1998	35706	48622	0.734	36047	52164	0.691	35532	47394	0.750	36046	52164	0.691	36045	52235	0.690
1999	37108	49648	0.747	37818	51830	0.730	36837	48558	0.759	37818	53391	0.708	37773	53454	0.707
2000	36577	50045	0.731	37742	54817	0.689	36408	49120	0.741	37729	54805	0.688	38014	54862	0.693
2001	37398	50343	0.743	38887	55535	0.700	37286	49745	0.750	38814	55521	0.699	38725	55438	0.699
2002	38188	52153	0.732	39579	55435	0.714	37651	50535	0.745	39579	55435	0.714	39550	55431	0.713
2003	38323	52298	0.733	39556	55475	0.713	37933	50873	0.746	39556	55475	0.713	39429	55503	0.710
2004	37158	51741	0.718	38102	55233	0.690	36900	50449	0.731	38102	55233	0.690	37972	55123	0.689
2005	38235	51594	0.741	39333	55412	0.710	37898	50183	0.755	39333	55412	0.710			
2006	37969	51725	0.734	39689	55718	0.712	37600	50379	0.746	39689	55718	0.712			

Source: Author's calculations using public and internal March CPS data.

Table 4. Mean Labor Earnings and Earnings Gap among Full-Time, Full-Year workers by educational attainment (1975-2006).

	No Cell Mean Public Use					Unadjusted Public Use					Consistent Topcode Public Use				
Income year	Less than a High School Degree (1)	High School Graduate (2)	Education Beyond High School (3)	Ratio (2)/(1)	Ratio (3)/(2)	Less than a High School Degree (6)	High School Graduate (7)	Education Beyond High School (8)	Ratio (6)/(7)	Ratio (7)/(8)	Less than a High School Degree (11)	High School Graduate (12)	Education Beyond High School (13)	Ratio (12)/(11)	Ratio (13)/(12)
1975	31108	36568	50423	1.18	1.38	31108	36568	50423	1.18	1.38	30870	36123	48017	1.17	1.33
1976	31829	36743	51008	1.15	1.39	31829	36743	51008	1.15	1.39	31621	36429	48837	1.15	1.34
1977	31382	37119	50613	1.18	1.36	31382	37119	50613	1.18	1.36	31261	36798	48579	1.18	1.32
1978	32066	36940	50231	1.15	1.36	32066	36940	50231	1.15	1.36	31888	36693	48669	1.15	1.33
1979	31800	36489	49408	1.15	1.35	31800	36489	49408	1.15	1.35	31692	36302	48329	1.15	1.33
1980	30141	35295	47512	1.17	1.35	30141	35295	47512	1.17	1.35	30093	35160	46682	1.17	1.33
1981	29379	34573	48208	1.18	1.39	29379	34573	48208	1.18	1.39	29250	34288	46338	1.17	1.35
1982	28264	34323	48301	1.21	1.41	28264	34323	48301	1.21	1.41	28156	33993	46411	1.21	1.37
1983	28892	33783	48228	1.17	1.43	28892	33783	48228	1.17	1.43	28703	33544	46508	1.17	1.39
1984	28875	34731	49222	1.20	1.42	28875	34731	49222	1.20	1.42	28645	34342	47019	1.20	1.37
1985	28657	34751	50023	1.21	1.44	28657	34751	50023	1.21	1.44	28460	34477	47962	1.21	1.39
1986	28999	35432	51404	1.22	1.45	28999	35432	51404	1.22	1.45	28833	35196	49351	1.22	1.40
1987	29838	35556	51371	1.19	1.44	29838	35556	51371	1.19	1.44	29607	35268	49437	1.19	1.40
1988	28965	35570	51381	1.23	1.44	28965	35570	51381	1.23	1.44	28758	35299	49824	1.23	1.41
1989	28311	35693	52139	1.26	1.46	28311	35693	52139	1.26	1.46	28207	35414	50562	1.26	1.43
1990	27396	34363	50726	1.25	1.48	27396	34363	50726	1.25	1.48	27288	34155	49277	1.25	1.44
1991	26801	33817	50502	1.26	1.49	26801	33817	50502	1.26	1.49	26674	33641	49342	1.26	1.47
1992	26628	33907	50390	1.27	1.49	26628	33907	50390	1.27	1.49	26543	33792	49375	1.27	1.46
1993	25813	33645	50024	1.30	1.49	25813	33645	50024	1.30	1.49	25763	33517	49303	1.30	1.47
1994	26267	34184	50517	1.30	1.48	26267	34184	50517	1.30	1.48	26188	34084	50109	1.30	1.47
1995	26475	34517	51559	1.30	1.49	26730	35392	55241	1.32	1.56	26383	34077	49662	1.29	1.46
1996	26598	35021	52232	1.32	1.49	27120	36035	56638	1.33	1.57	26302	34651	50319	1.32	1.45
1997	27616	34989	52927	1.27	1.51	29064	35712	57406	1.23	1.61	27187	34657	51043	1.27	1.47
1998	26501	35332	54392	1.33	1.54	27209	35984	59060	1.32	1.64	26327	35139	52773	1.33	1.50
1999	26592	35902	55339	1.35	1.54	26926	36457	58185	1.35	1.60	26421	35630	53959	1.35	1.51
2000	26610	36290	55406	1.36	1.53	27751	37243	61417	1.34	1.65	26379	36092	54253	1.37	1.50
2001	27031	35918	55939	1.33	1.56	28294	37302	62334	1.32	1.67	26892	35765	55206	1.33	1.54
2002	27175	36370	57781	1.34	1.59	27864	37318	61984	1.34	1.66	26936	35976	55780	1.34	1.55
2003	27246	36724	57576	1.35	1.57	27858	37481	61527	1.35	1.64	27062	36455	55787	1.35	1.53
2004	26343	36164	57066	1.37	1.58	26757	37118	61331	1.39	1.65	26247	35858	55474	1.37	1.55
2005	26560	35914	56903	1.35	1.58	26944	36731	61567	1.36	1.68	26378	35599	55170	1.35	1.55
2006	26791	35753	56764	1.33	1.59	27493	36589	61830	1.33	1.69	26607	35515	55115	1.33	1.55

Source: Author's calculations using public and internal March CPS data.

Table 4 (continued):

Income year		Cell Mean Public Use						Internal			
	Less than a High School Degree (16)	High School Graduate (17)	Education Beyond High School (18)	Ratio (17)/(16)	Ratio (18)/(17)		Less than a High School Degree (20)	High School Graduate (21)	Education Beyond High School (22)	Ratio (21)/(20)	Ratio (22)/(21)
1975	31155	36731	51379	1.18	1.40		31175	36799	51546	1.18	1.40
1976	31899	36864	52125	1.16	1.41		31905	36840	52146	1.15	1.42
1977	31429	37273	52151	1.19	1.40		31441	37224	51931	1.18	1.40
1978	32226	37112	51642	1.15	1.39		32234	37175	51705	1.15	1.39
1979	31956	36775	51513	1.15	1.40		31871	36736	51310	1.15	1.40
1980	30218	35458	48448	1.17	1.37		30265	35519	49366	1.17	1.39
1981	29441	34645	48855	1.18	1.41		29449	34644	48853	1.18	1.41
1982	28280	34430	48928	1.22	1.42		28280	34435	48999	1.22	1.42
1983	28964	33881	48984	1.17	1.45		28963	33875	48989	1.17	1.45
1984	28875	34731	49222	1.20	1.42		28875	34731	49222	1.20	1.42
1985	28723	34771	50486	1.21	1.45		28796	34795	51019	1.21	1.47
1986	29133	35538	52751	1.22	1.48		29130	35521	52765	1.22	1.49
1987	30046	35755	52937	1.19	1.48		30127	35725	52986	1.19	1.48
1988	29218	35845	52947	1.23	1.48		29148	35869	52990	1.23	1.48
1989	28475	36083	54522	1.27	1.51		28443	36085	54577	1.27	1.51
1990	27533	34652	52953	1.26	1.53		27573	34664	52982	1.26	1.53
1991	26985	34054	52256	1.26	1.53		27195	33984	52296	1.25	1.54
1992	26795	34127	52684	1.27	1.54		26855	34150	52678	1.27	1.54
1993	26163	34408	54988	1.32	1.60		26717	34795	54704	1.30	1.57
1994	26642	34935	55723	1.31	1.60		26432	35037	55730	1.33	1.59
1995	26736	35375	55261	1.32	1.56		26789	35617	55113	1.33	1.55
1996	27120	36037	56642	1.33	1.57		27296	36009	56649	1.32	1.57
1997	29063	35722	57405	1.23	1.61		29113	35707	57399	1.23	1.61
1998	27207	36001	59051	1.32	1.64		27524	35886	59077	1.30	1.65
1999	27048	36713	60118	1.36	1.64		27174	36693	60136	1.35	1.64
2000	27744	37238	61404	1.34	1.65		28380	37383	61347	1.32	1.64
2001	28264	37227	62361	1.32	1.68		29253	37505	62085	1.28	1.66
2002	27864	37318	61984	1.34	1.66		28157	37341	61940	1.33	1.66
2003	27858	37481	61527	1.35	1.64		27660	37362	61649	1.35	1.65
2004	26757	37118	61331	1.39	1.65		27241	37274	61135	1.37	1.64
2005	26944	36731	61567	1.36	1.68						
2006	27493	36589	61830	1.33	1.69						

Source: Author's calculations using public and internal March CPS data.

Figure 1: Trend in the female-male earnings ratio 1975-1992 (1975 normalized to 1)
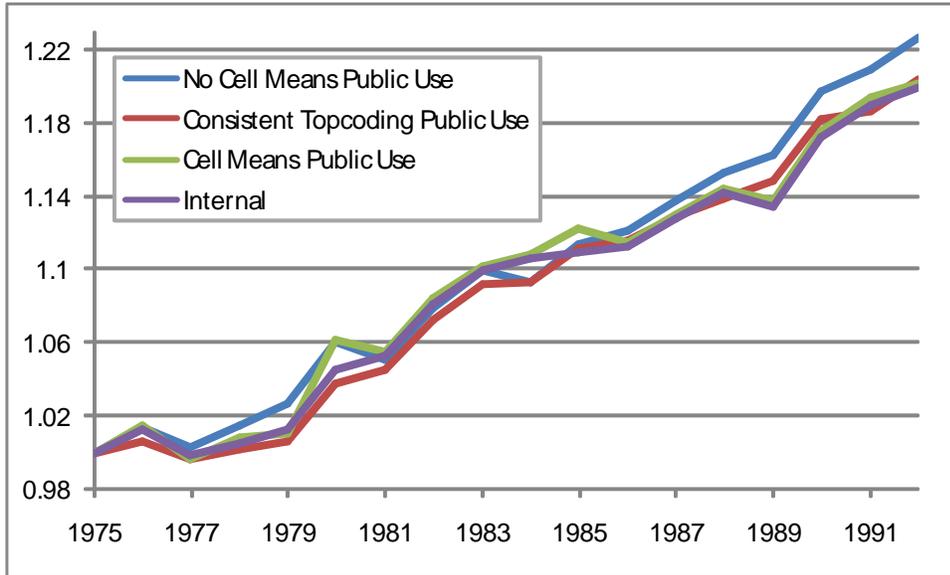
Figure 2: Trend in the female-male earnings ratio 1993-2006 (1993 normalized to 1)
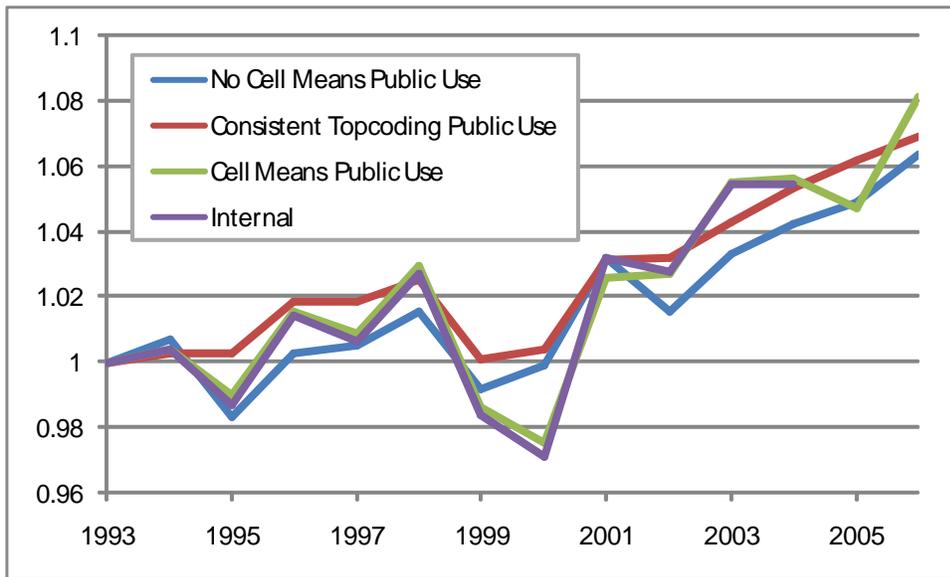
Figure 3: Trend in the black-white earnings ratio 1975-1992 (1975 normalized to 1)
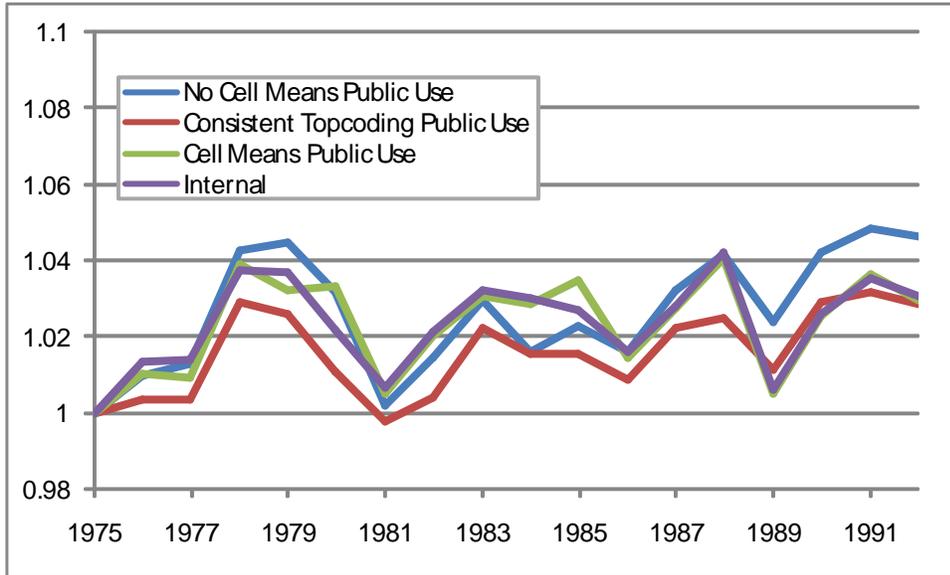
Figure 4: Trend in the black-white earnings ratio 1993-2006 (1993 normalized to 1)
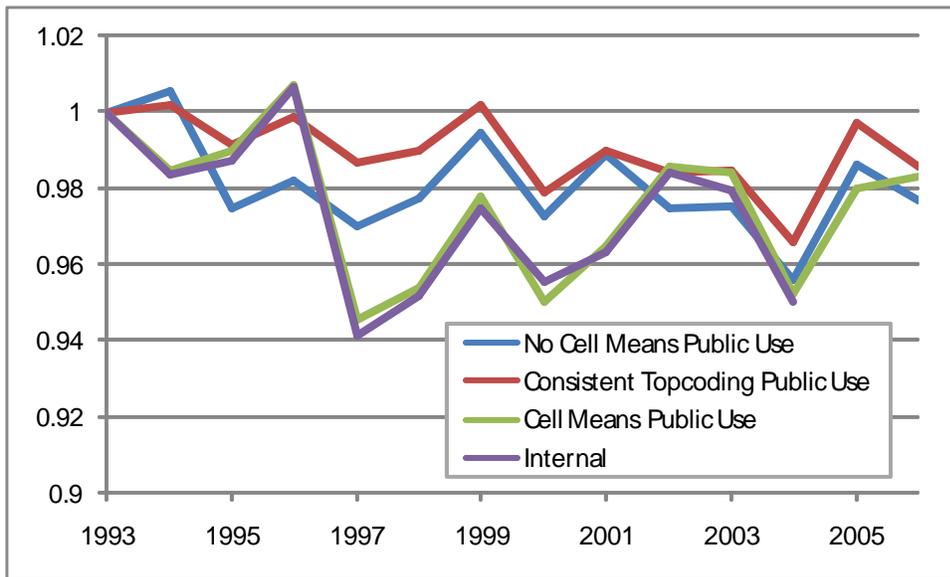
Figure 5: Trend in the earnings ratio between workers with post-secondary education and workers with only a high-school degree 1975-1992 (1975 normalized to 1)
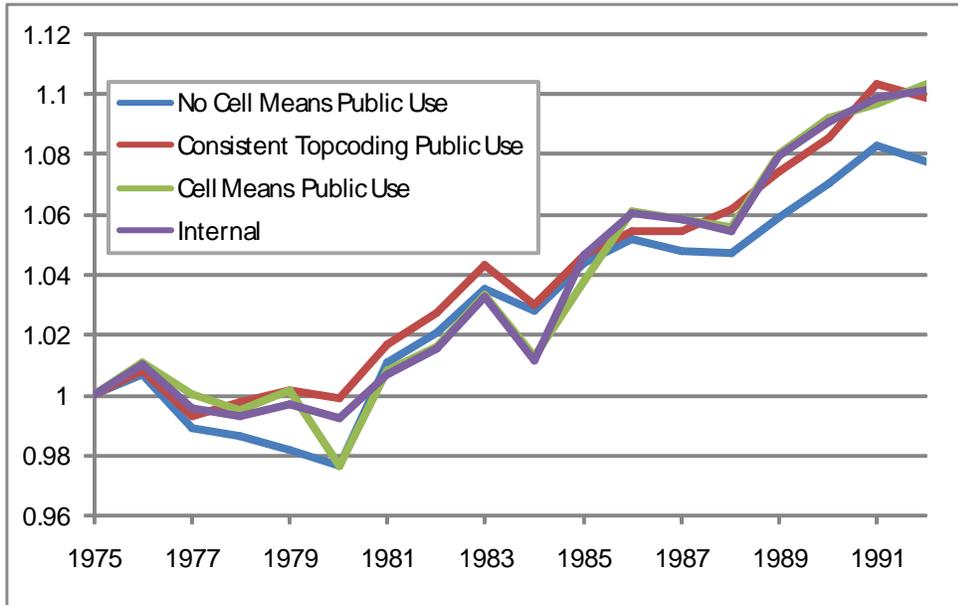
Figure 6: Trend in the earnings ratio between workers with post-secondary education and workers with only a high-school degree 1993-2006 (1993 normalized to 1)
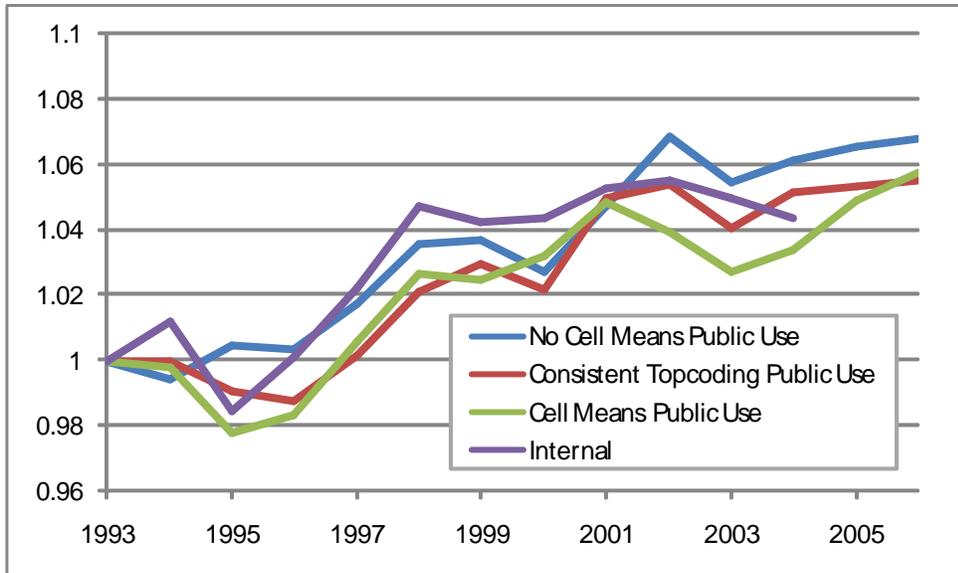
Appendix Table 1. Labor Earnings Items Reported in the Current Population Survey

Name	Name in Public Files	Name in Internal Files	Definition
1976–1987			
Wages	I51A	WSAL_VAL	Wages and Salaries
Self Employment	I51B	SEMP_VAL	Self employment income
Farm	I51C	FRSE_VAL	Farm income
1988–2007			
Primary earnings	ERN_VAL	ERN_VAL	Primary Earnings
Wages	WS_VAL	WS_VAL	Wages and Salaries - Second Source
Self Employment	SE_VAL	SE_VAL	Self employment income - Second Source
Farm	FRM_VAL	FRM_VAL	Farm income - Second Source

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

Appendix Table 2: Public and Internal Topcoding Points for each Labor Earnings Source in Dollars (1975-1986)

Income Year	Public-Use Data Topcode Thresholds			Internal Data Topcode Thresholds		
	Wages (I51A)	Self Employment (I51B)	Farm (I51C)	Wages (I51A)	Self Employment (I51B)	Farm (I51C)
1975	50,000	50,000	50,000	99,999	99,999	99,999
1976	50,000	50,000	50,000	99,999	99,999	99,999
1977	50,000	50,000	50,000	99,999	99,999	99,999
1978	50,000	50,000	50,000	99,999	99,999	99,999
1979	50,000	50,000	50,000	99,999	99,999	99,999
1980	50,000	50,000	50,000	99,999	99,999	99,999
1981	75,000	75,000	75,000	99,999	99,999	99,999
1982	75,000	75,000	75,000	99,999	99,999	99,999
1983	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	99,999	99,999	99,999	250,000	250,000	250,000
1986	99,999	99,999	99,999	250,000	250,000	250,000

Source: Public Use Topcode points from Current Population Survey Annual Demographic File Technical Documentation (1976-1987) and Current Population Survey Annual Social and Economic Supplement Technical Documentation (2003-2007). Internal Topcode points from Author's calculations using internal March CPS data

Appendix Table 3: Public and Internal Topcoding Points for each Labor Earnings Source in Dollars (1987-2006)

| | Public-Use Data Topcode Thresholds | | | | Internal Data Topcode Thresholds | | | |
| | Primary | | Self | | Primary | | Self | |
Income Year	Earnings (ERN_VAL)	Wages (WS_VAL)	Employment (SE_VAL)	Farm (FRM_VAL)	Earnings (ERN_VAL)	Wages (WS_VAL)	Employment (SE_VAL)	Farm (FRM_VAL)
1988	99,999	99,999	99,999	99,999	299,999	99,999	99,999	99,999
1989	99,999	99,999	99,999	99,999	299,999	99,999	99,999	99,999
1990	99,999	99,999	99,999	99,999	299,999	99,999	99,999	99,999
1991	99,999	99,999	99,999	99,999	299,999	99,999	99,999	99,999
1992	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	299,999	99,999	99,999	99,999
1994	99,999	99,999	99,999	99,999	999,999	999,999	999,999	999,999
1995	99,999	99,999	99,999	99,999	1,099,999	1,099,999	999,999	999,999
1996	150,000	25,000	40,000	25,000	1,099,999	1,099,999	999,999	999,999
1997	150,000	25,000	40,000	25,000	1,099,999	1,099,999	999,999	999,999
1998	150,000	25,000	40,000	25,000	1,099,999	1,099,999	999,999	999,999
1999	150,000	25,000	40,000	25,000	1,099,999	1,099,999	999,999	999,999
2000	150,000	25,000	40,000	25,000	1,099,999	1,099,999	999,999	999,999
2001	150,000	25,000	40,000	25,000	1,099,999	1,099,999	999,999	999,999
2002	150,000	25,000	40,000	25,000	1,099,999	1,099,999	999,999	999,999
2003	200,000	35,000	50,000	25,000	1,099,999	1,099,999	999,999	999,999
2004	200,000	35,000	50,000	25,000	1,099,999	1,099,999	999,999	999,999
2005	200,000	35,000	50,000	25,000	1,099,999	1,099,999	999,999	999,999
2006	200,000	35,000	50,000	25,000				
2007	200,000	35,000	50,000	25,000				

Source: Public Use Topcode points from Current Population Survey Annual Demographic File Technical Documentation (1988-2002) and Current Population Survey Annual Social and Economic Supplement Technical Documentation (2003-2007). Internal Topcode points from Author's calculations using internal March CPS data