

Consistent cell means for topcoded incomes in the public use march CPS (1976–2007)

Jeff Larrimore^{a,*}, Richard V. Burkhauser^a, Shuaizhang Feng^b and Laura Zayatz^c

^a*Cornell University, NY, USA*

^b*Shanghai University of Finance and Economics, China*

^c*US Census Bureau, USA*

Using the internal March CPS, we create and in this paper distribute to the larger research community a cell mean series that provides the mean of all income values above the topcode for any income source of any individual in the public use March CPS that has been topcoded since 1976. We also describe our construction of this series. When we use this series together with the public use March CPS, we closely match the yearly mean income levels and income inequalities of the US population found using the internal March CPS data.

Keywords: Inequality, income, earnings, Current Population Survey, Gini coefficient, topcoding, and cell means

1. Introduction

The public use version of the March Current Population Survey (CPS) is the primary data source used by public policy researchers and administrators to investigate yearly trends in average income and its distribution in the United States. The public use March CPS is a large nationally representative sample of households collected each March, since 1942, by the US Census Bureau.¹ Its detailed questions on the employment and income sources of household members make it an extremely valuable resource for tracking long term trends in the economic well-being of Americans. However, to protect the confidentiality of its respondents, the US Census Bureau censors the income of individuals above specified topcoded levels in the public use March CPS data that it makes available to the outside research community. The impact of US Census Bureau topcoding procedures on measured wage earnings and income inequality for the population as a whole have been explored by Feng, Burkhauser, and Butler [4] and Burkhauser, Feng, and Jenkins [2].

*Corresponding author: Jeff Larrimore, Department of Economics, 408 Uris Hall, Cornell University, Ithaca, NY 14853, USA. Tel.: +1 610 715 7704; Fax: +1 607 255 2818; E-mail: jhl42@cornell.edu.

¹Each year the US Census Bureau releases its yearly average income and poverty rates from the March CPS using these data in “Income Poverty, and Health Insurance Coverage in the United States” [9].

In 2006 we were granted permission by the US Census Bureau to use the internal March CPS to test the sensitivity of measured income and income inequality in the public use CPS to alternative methods of providing information on topcoded values. We then provide researchers without access to the internal March CPS data with our best method of capturing information on these topcoded values from the internal data, without unduly risking the confidentiality of CPS respondents. Using the internal March CPS data we create and in this paper distribute to the larger research community a cell mean series extending back to 1976 that provides the mean of all income values above the topcode for any income source of any individual in the public use March CPS that has been topcoded.² Our cell mean series, which includes years from 1976–2002, can then be used in conjunction with cell means provided by the Census Bureau for later years to create a complete set of cell means from 1976–2007.

We show that when we use this extended cell mean series together with the public use March CPS, we are able to closely match the yearly mean income of the US population found with the internal March CPS data over the period 1976–2002. We then show that this corrected public use March CPS does a better job of matching income inequality trends found in the internal March CPS than those previously available in the literature.

2. Topcoding in the public use March CPS

In the March CPS, a respondent in each household is asked a series of questions on the sources of income of the household. Starting in 1976 they reported income from 11 sources and since 1988 they have done so for income from 24 sources. These income sources are listed in Appendix Table 1. While this is an effective way of capturing total household income, it complicates efforts to protect the confidentiality of respondents and their households. Rather than simply topcoding high total household income values in the public use March CPS, the US Census Bureau must topcode high values for each source of household income. Otherwise the income of high-income individuals and their households could be reconstructed by summing each individual source of income. The full list of topcode values over time in the public use CPS are reported in Appendix Tables 2 and 3.

Prior to 1996, the US Census Bureau assigned the topcode value from that source of income to all topcoded income values, regardless of the total income of the individual or the individuals' household. For example, in 1990 the topcode for Social Security income was \$29,999. Individuals reporting Social Security income above that value

²Each CPS survey captures income from the previous year. In this paper all references are to the survey year, so when we discuss the year 1976, this refers to income from various sources that members of the household received in 1975 reported on the March 1976 Current Population Survey.

were assigned that value, even if that relatively low value in the distribution of total household income in the United States was the only source of their household's income.

In 1996, rather than simply assigning the topcode value of that income source to individuals in the public use March CPS, the US Census Bureau used values from the internal March CPS to assign a cell mean value for each topcoded labor earnings value in the public use March CPS. In 1999, the Census Bureau started to provide cell mean values to non-governmental non-labor earnings income sources in a similar fashion. For income from non-labor earnings sources, these cell means provide the mean income of all topcoded individuals from that source. For income from labor earnings sources, the US Census Bureau provides more detailed information, providing cell means conditional on the topcoded individual's race, gender, and employment status.

Researchers using these cell means in conjunction with the public use March CPS will obtain mean income values that more closely mirror those in the internal March CPS data. Hence the US Census Bureau's cell means series has since 1996 added useful information to the public use March CPS.

Obviously, cell means do not provide all the information on the income of the population topcoded in the public use data available in the internal data, since cell means can't be used to capture the actual dispersion of income among those at the highest end of each source of income in the March CPS data. But cell means not only allow researchers to better track mean income with the public use data, but also show how much income is being topcoded from each source of income.

Because the US Census Bureau cell means series starts in 1996, using the public use March CPS data without correcting for this major change in the reported income values of those at the highest end of each of its sources of income results in a significant increase in measured income in 1996 and beyond simply due to more accurate reporting of their income. Hence while the use of cell means after 1996 makes the public use March CPS better conform to the internal March CPS, not taking this improvement in measurement into account will grossly overestimate how much actual income increased in 1996 among those at the highest income levels. And because these topcodes occur throughout the income distribution, attempting to account for them by simply using a 90/10 ratio will not fully control for their impact on measured income inequality. (See Burkhauser, Feng, and Jenkins [2]).

The major increase in the accuracy of the public use March CPS in 1996 is a specific example of the more general problem that income topcoding presents researchers interested in measuring levels and trends in the income and income distribution of the US population – inconsistent topcode levels lead to artificial increases or decreases in mean incomes as different fractions of the population are subject to topcoding each year. (See: Levy and Murnane [8] for an early review of the income distribution literature and a more formal statement of this problem.)

This is a legitimate concern since topcodes in the public use March CPS have, in general, non-systematically increased over the past 30 years and some part of the measured increase in income over this time in the uncorrected data is caused by

topcoding capturing a larger portion of the income distribution. Using cell means substantially alleviates this problem after 1996, since cell means provide all the information available in the internal data on total income. As a result, even if public topcodes are inconsistently adjusted in the public use March CPS, mean income for the entire population in the internal March CPS can be matched with the public use data using cell means.

Despite the US Census Bureau's attempt to alleviate the problem of topcoding, their cell means have generally been ignored by researchers studying United States long-term income trends, since to do otherwise exacerbates time-inconsistencies by using unadjusted data before 1996 and cell mean data thereafter. Instead, researchers analyzing data that includes years prior to 1996 use other methods of controlling for inconsistent topcoding. These methods include measuring inequality with 90/10 ratios to avoid complications from topcoding at the top of the distribution, truncating the data by removing the top 2.5% of wage earners (Lerman [7]), or artificially lowering the topcodes to create a series with a constant percentage of people topcoded in every year (Burkhauser, Butler, Feng, and Houtenville [1]). While each of these methods is preferable to using the unadjusted data, they also have drawbacks.

Burkhauser, Feng, and Jenkins [2] show that using 90/10 ratios may not alleviate the problem, since individuals are topcoded by income source and not total income and therefore topcoded individuals may fall below the 90th percentile and impact 90-10 ratios. Additionally, any analysis of inequality based on just 2 points in the distribution will be less informative than analyses using a Gini or Theil coefficient that incorporates more of the available data.

As can be seen by the growing number of people in Tables 1 and 2 who are topcoded in the public use CPS data for each income source from 1976–2007, using a truncated dataset or dropping the topcodes to provide a consistent income series is becoming increasingly problematic.³ Part of this increase in topcode counts was by design as the US Census Bureau attempted to ensure that enough people were topcoded in the public data to provide meaningful cell means, while still protecting respondent confidentiality. The number of individuals topcoded for secondary wages, farm, and self-employment earnings increased significantly in 1996, the year the US Census Bureau lowered the topcodes for all secondary labor income sources and began providing cell means. Similarly, in 1999 when cell means were first provided for non-labor income and the topcode values for those income sources dropped, the number of individuals topcoded from these sources spiked. However, since 1999 topcode levels have remained unchanged but increases in income over time have led to further increases in the number of individuals topcoded. As illustrated in Fig. 1, in 2007, almost 6% of individuals in the public use CPS had at least one income source

³All calculations for this paper were performed using Stata Statistical Software, version 10.0, and SAS software, Version 9.1. Stata and all other StataCorp products or service names are registered trademarks or trademarks of StataCorp LP, College Station, TX. SAS and all other SAS Institute Inc. products or service names are registered trademarks or trademarks of SAS Institute Inc., Cary, NC.

Table 1
Count of topcoded individuals in the public use March CPS by income source (1976–1987)

All Sources	Wages (I51A)	Self Employment (I51B)	Farm (I51C)	Social Security (I52A)	Supplemental Security (I52B)	Public Assistance (I53A)	Interest Dividends (I53B)	Rentals (I53C)	Veterans and Workers Comp (I53D)	Retirement (I53E)	Other (I53F)	Other Household Members
1976	1189	204	96	11	0	0	6	14	0	1	0	868
1977	1786	302	143	19	3	0	1	28	0	1	1	1314
1978	2018	362	163	7	3	0	4	40	0	0	3	1450
1979	2709	476	227	20	23	0	9	33	0	6	3	1939
1980	4028	772	243	33	63	1	14	47	4	11	13	2868
1981	4914	1003	237	37	121	3	10	41	8	4	4	3508
1982	1502	322	91	8	1	1	4	12	3	5	2	1057
1983	2047	441	114	8	0	4	3	26	14	6	4	1448
1984	2117	461	144	13	1	23	11	10	8	8	6	1458
1985	1293	257	87	2	6	2	20	25	6	6	5	907
1986	1404	317	76	0	4	0	6	21	16	6	6	975
1987	1829	413	101	8	3	2	14	25	18	5	0	1269

Note: In parentheses below each variable name is the mnemonic for the income source from the March Current Population Survey Technical Documentation. Source: Author's calculations using internal March CPS data.

Table 2
Count of topcoded individuals in the public use March CPS by income source (1988-2005)

Year	All Sources	Primary Earnings (ERN_VAL)	Wages (WS_VAL)	Self Employment (SE_VAL)	Farm (FRM_VAL)	Social Security (SS_VAL)	Supplemental Security (SSL_VAL)	Public Assistance (PAW_VAL)	Interest (INT_VAL)	Dividends (DIV_VAL)	Rental (RNT_VAL)	Alimony (ALM_VAL)	Child Support (CSP_VAL)
1988	1813	523	2	4	1	1	4	0	14	5	5	0	0
1989	2095	613	0	6	1	0	1	0	16	4	11	0	0
1990	3022	903	5	8	0	6	6	4	37	7	7	2	0
1991	2929	866	0	16	0	4	5	0	25	7	10	0	0
1992	2931	874	4	6	0	2	8	0	30	13	15	0	0
1993	3371	980	2	7	0	1	10	2	31	6	15	0	0
1994	4102	1188	19	10	1	13	49	1	35	13	20	0	0
1995	5115	1373	29	17	3	80	55	9	24	18	30	1	0
1996	2807	408	309	65	17	47	2	11	30	10	32	2	1
1997	2736	492	229	51	5	71	8	7	52	24	24	1	1
1998	3697	549	452	67	28	64	5	1	70	52	47	1	2
1999	6193	576	429	88	20	73	4	0	312	685	253	5	38
2000	7030	718	513	107	26	39	1	1	391	700	260	5	62
2001	6771	732	581	86	44	65	1	1	351	520	201	10	64
2002	12010	1410	859	156	96	81	4	4	452	646	388	12	122
2003	8496	819	383	88	64	63	7	6	543	439	231	15	151
2004	9553	780	442	138	102	43	5	0	523	621	232	13	148
2005	9684	760	439	122	77	80	9	2	672	660	238	20	164
2006	10635	866	482	118	107	52	25	2	737	877	222	10	155
2007	11779	974	541	107	52	92	13	5	1013	883	279	12	168

Note: In parentheses below each variable name is the mnemonic for the income source from the March Current Population Survey Technical Documentation. Source: Author's calculations using internal March CPS data.

Table 2, continued

Year	Unemployment (UC_VAL)	Workers Comp (WC_VAL)	Veterans (VET_VAL)	Retirement 1st source (RET_VAL) (RET_VAL)	Retirement 2nd source (RET_VAL) (RET_VAL)	Survivors 1st Source (SUR_VAL) (SUR_VAL)	Survivors 2nd Source (SUR_VAL) (SUR_VAL)	Disability 1st Source (DIS_VAL) (DIS_VAL)	Disability 2nd Source (DIS_VAL) (DIS_VAL)	Education Assistance (ED_VAL) (ED_VAL)	Financial Assistance (FIN_VAL) (FIN_VAL)	Other (OL_VAL) (OL_VAL)	Household Members
1988	0	0	4	9	0	6	2	1	0	0	0	0	1258
1989	0	0	4	7	0	9	0	0	0	0	0	4	1463
1990	0	1	5	10	0	11	0	1	0	0	0	5	2077
1991	0	2	6	4	0	19	2	0	0	0	0	0	2016
1992	0	0	12	9	1	11	0	1	0	1	0	1	2019
1993	0	0	9	8	1	22	0	0	0	2	3	4	2332
1994	0	0	0	6	0	20	0	1	0	0	0	0	2828
1995	2	2	0	29	2	13	0	2	0	0	9	5	3543
1996	2	3	1	16	1	16	1	3	0	0	2	5	1918
1997	0	0	2	31	0	14	0	3	0	1	1	3	1816
1998	1	4	0	17	2	16	0	7	0	0	0	4	2522
1999	0	1	2	205	9	39	1	25	5	60	13	16	4158
2000	1	2	0	214	7	30	3	20	1	69	11	29	4772
2001	2	2	5	203	6	41	5	20	1	70	19	20	4487
2002	4	0	2	297	17	45	1	26	0	91	31	41	8387
2003	5	4	5	372	11	54	4	28	0	143	30	29	5890
2004	2	4	1	385	7	56	3	36	0	216	32	40	6622
2005	2	0	1	474	9	50	4	61	4	185	20	29	6742
2006	1	2	2	478	22	53	2	64	1	224	38	40	7398
2007	1	1	4	554	7	107	1	50	0	217	27	37	8152

Note: In parentheses below each variable name is the mnemonic for the income source from the March Current Population Survey Technical Documentation. Source: Author's calculations using internal March CPS data.

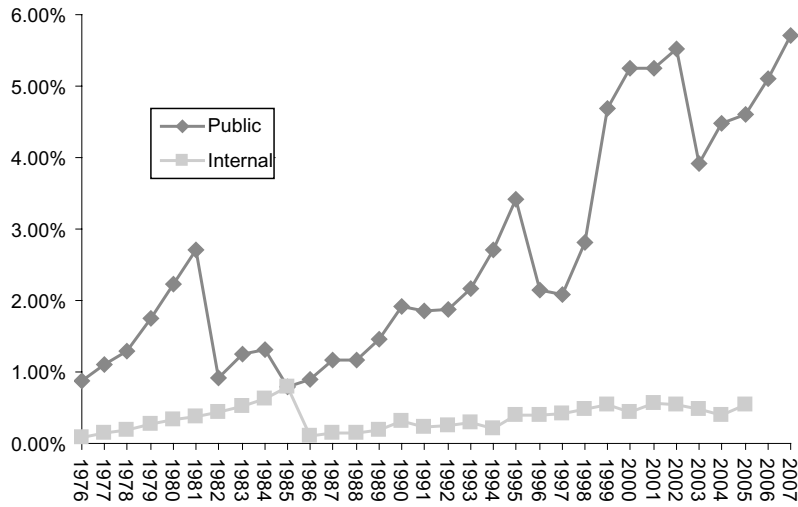


Fig. 1. Percentage of individuals topcoded/censored in the March CPS by year. Note that internal data was not available for years after 2005.

that was topcoded, the highest rate of topcoding since 1976. Unless public use CPS topcodes are systematically increased as income increases, these topcodes will cut deeper into the income distribution. In that case consistent topcoding methods will do an increasingly poorer job of capturing trends at the top of the income distribution as they remove larger fractions of the population to maintain their consistency.

While each of these previously employed methods improve the unadjusted public use March CPS data, the lack of publicly available data on individuals above the topcode thresholds prior to 1996 makes it impossible to obtain an accurate picture of long-term income trends across the population captured in the internal March CPS. Recognizing the benefit of cell means for replicating internal March CPS, we obtained access to the internal March CPS data and received permission to extend the cell mean series back to 1976.

3. Procedures for Obtaining and using the Internal March CPS

To gain access to restricted access US Census Bureau data like the internal March CPS, a researcher must submit a proposal to one of the nine US Census Bureau Research Data Centers (RDCs) describing the nature of the proposed research, the required data set and the expected results. The research must be beneficial to the US Census Bureau by, for example, improving data quality or methods of collecting, measuring, or tabulating a survey, census, or estimate. If the proposal is accepted, the researcher and any associates working at the RDC must obtain *Special Sworn*

Status to do so. All are bound by law to maintain confidentiality, like any other US Census Bureau employee.

In 2006, we were granted permission through this process to use the internal March CPS to test the sensitivity of measured income inequality in the public use CPS to alternative methods of providing information on topcoded persons and to provide researchers without access to the internal March CPS data with our best method of capturing information on these topcoded values from the internal data, without unduly risking the confidentiality of CPS respondents.

We were given access to internal March CPS records from 1976–2005. These data records include information on income above the public use topcode thresholds for respondents in the March CPS survey. However, since our research project was limited to how topcoding impacts measures of inequality in the United States, the files do not contain the responses to all of the non-income related questions in the March CPS dataset. These omissions occur because the US Census Bureau limits internal data availability to data within the scope of the project.

Additionally, even the internal March CPS data, which are not subject to top coding, have been censored to various degrees over time.

This censoring was initially implemented in the internal March CPS due to data-storage limitations in the computing systems of the 1970s which necessitated truncating written records to 5 digits. While these data storage limitations are no longer a concern, the internal censoring of income values has continued, partially due to concerns about data reliability of individuals who report an extremely high income value and partially due to continuing concerns over confidentiality. Thus, the US Census Bureau does not maintain any versions of the internal March CPS data that are not subject to this form of censoring. The internal March CPS data we were provided is identical in this respect to the internal March CPS data the US Census Bureau uses to produce all official United States income statistics. The extent to which censoring exists in this internal data can be seen in Appendix Tables 4 and 5, which provide the censoring levels in the internal data for each income source over time. (See Feng, Burkhauser, and Butler [4] for a discussion of the problem of censorship in the internal March CPS labor earnings series and Welniak [11] and Burkhauser, Feng, and Jenkins [2] for a similar discussion for all sources of income.)

While our data agreement allows us to freely explore the internal March CPS data in our file within the purview of our project, US Census Bureau data confidentiality requirements limit the level of detail of our findings that we can release to the broader research community. Our results, like almost all publicly released data products were reviewed by the US Census Bureau Disclosure Review Board (DRB) to ensure that they were consistent with its disclosure policies. Six members of the DRB represent the US Census Bureau's demographic, decennial, and economic directorates, and its RDCs. An additional three members represent the research and policy areas. Although our data went through the standard review procedure and had to be approved by this board before its release, occasionally exceptions are made to this review policy, which are outlined in Zayatz [12].

In order to release data, we were required to send a memo to the chair of the DRB accompanied by the DRB checklist, the questionnaire from the survey or census, a list of variables of interest, a record layout (if microdata), table outlines (if tabular data), and on occasion some cross tabulations of the variables of interest. The DRB checklist asks basic questions about the content of the data file to be released and also has sections for microdata, frequency count data, and magnitude data. This procedure helps to ensure consistency in the DRB's decision making process.

Topcoding of the public use March CPS is one of the DRB's main tools for protecting the confidentiality of respondents. While the DRB is not responsible for setting the actual topcode levels in the public use March CPS, they do provide the guidelines that dictate what fraction of individuals must be topcoded from each income source before confidentiality becomes compromised. These guidelines state that either 3% of individuals with positive income from an income source must be topcoded or 0.5% of all individuals including those without income from the source must be topcoded – whichever results in fewer individuals topcoded. In practice, for all income sources other than primary labor earnings, topcoding 3% of individuals with positive income from an income source will result in fewer individuals topcoded – so this is the rule that is used most often. However for some income sources these guidelines have been waived thus permitting a smaller fraction of the population to be topcoded.

Since topcoding is an important element of the US Census Bureau's confidentiality policies, in our requests to release cell mean information about topcoded individuals, it was always necessary to ensure that no personal income information could be identified from the data we released. The general guideline for releasing information on topcoded individuals is that any means must be based on at least 3 topcoded responses, and in many cases a minimum threshold of 5 is used. These means can be released at the national level, state level, or even for a subgroup of the population with a given geographic level as long as the threshold is met. The DRB has stated that the mean, median, and standard deviation of the topcoded values may be released if there are at least 5 topcoded values and in doing so, the mean and median should be rounded to 2 significant digits. In addition, quantiles of the topcoded values may be released if there are at least 10 topcoded values in each quantile and if the quantile bounds are rounded to 2 significant digits.

These guidelines prohibited us from releasing cell means for income sources with less than 5 individuals. However, in this case, we were permitted to combine such an income source with another one in order to create a cell mean based on 5 or more individuals.

Even when our cell means were based on 5 or more observations, the DBR further limited their release if they could be combined with other previously released cell means to induce the actual income of a specific person. For example, this could occur if we had previously released a file including the mean of 6 topcoded values and requested the release of the mean of 5 of these values. Given those two means,

the exact value of the remaining respondent could be deduced and hence release of the mean of 5 would be prohibited based on the DRB guidelines.

Since 1996, the US Census Bureau started to provide cell means for earnings sources in the public use March CPS. To mitigate potential disclosure concerns associated with our cell mean series, for these years we use their identical cell definitions, taking particular note of which cells are combined to achieve groups of 5 or more individuals when less than 5 people are topcoded from an income source. Prior to 1996, such issues were of less concern since no comprehensive cell mean information had been released.

4. The extended public use March CPS cell mean series

Since the US Census Bureau topcodes individuals in the public use March CPS on each separate income source, we calculated cell means for each source of income separately – 11 prior to 1988 and 24 since then. In addition, mirroring US Census Bureau practices in their cell means releases since 1996, our cell means are weighted mean incomes of all individuals above the topcode value for the source for all non-labor earnings income sources. And for all labor earnings sources, the cell means are the weighted mean incomes of individuals with income above the topcoded, conditional on the individual's race, gender, and employment status. The cell mean values for 1976–1987 are provided in Tables 3 through 6 and for 1988–2002 in Tables 7 through 11. Although we only provide cell means through 2002, a comparison of our cell means since the US Census Bureau began releasing them in 1996 illustrate that our cell means series is virtually identical to those they release. Thus, the cell means we provide can be used in conjunction with cell means produced by the US Census Bureau in years since 2002 to provide a full set of cell means dating back to 1976.

While our cell mean creation methods described above closely follow those of the US Census Bureau there are small differences. In contrast to the US Census Bureau, we adjusted the treatment of individuals who have income exactly at the topcode level in order to better conform to the public use March CPS data. In the US Census Bureau cell mean series, individuals who have income exactly at the topcode level are not considered to have been topcoded because the topcode is not binding. However, prior to 1996, individuals whose income is exactly at the topcode level are listed in the public use March CPS data as having the same level of income as those whose income is above the topcode and hence are bound by it.

As a result, when using just the public use March CPS income records, it can't be determined who among those with income listed at the topcode threshold had income that was bound by it and whose actual income was exactly at the topcode value. If we had used the current US Census Bureau method of only including individuals who have income above the topcode value in their cell means, this would have limited the ability of researchers to use our extended cell mean series with the public use March CPS data.

Table 3
Primary and secondary wage earnings cell means (1976–1987)

Year	Full-Time (35+ hours per week)			Full-Year (50+ weeks)			Workers			Not Full-Time, Full-Year Workers		
	Male, Black or Hispanic	Not Male, Black or Hispanic	Female, Black or Hispanic	Male, Black or Hispanic	Not Male, Black or Hispanic	Female, Black or Hispanic	Male, Black or Hispanic	Not Male, Black or Hispanic	Female, Black or Hispanic	Male, Black or Hispanic	Not Male, Black or Hispanic	Female, Black or Hispanic
1976	65013	N/A	60612 ^A	60612 ^A	N/A	N/A	68598	N/A	N/A	N/A	N/A	N/A
1977	64197	N/A	62351 ^A	63207	N/A	N/A	70337	62351 ^A	N/A	N/A	N/A	N/A
1978	65436	70047 ^A	70047 ^A	59625	70047 ^A	N/A	67600	N/A	70047 ^A	70047 ^A	N/A	N/A
1979	66128	64292 ^A	73370	63549	64292 ^A	N/A	72495	N/A	64292 ^A	65016	N/A	N/A
1980	67737	61925	64371	62874	72125 ^A	72125 ^A	68818	N/A	70963	72125 ^A	N/A	N/A
1981	66210	58990	68661	60852	55631 ^A	55631 ^A	70415	65397	55631 ^A	85119	N/A	55631 ^A
1982	91610	N/A	90491 ^A	83489	N/A	N/A	94607	90491 ^A	N/A	N/A	N/A	N/A
1983	89485	87647 ^A	96915	92340	87647 ^A	N/A	88228	N/A	87647 ^A	87647 ^A	N/A	N/A
1984	90220	N/A	92530 ^A	88528	N/A	N/A	95586	N/A	N/A	92530 ^A	N/A	92530 ^A
1985	99999	99999	99999	99999	N/A	N/A	99999	N/A	N/A	99999	N/A	N/A
1986	136613	170804 ^A	124324	133348	N/A	N/A	137028	N/A	N/A	170804 ^A	N/A	N/A
1987	140359	119934	150042 ^A	125434	169047	N/A	137893	150042 ^A	N/A	150042 ^A	N/A	N/A

^A – Indicates that there are not enough observations to report a cell mean for this population group in this year. The reported cell mean is a combined cell mean with the other footnoted population groups in the same year.

N/A – Indicates that no individual with these demographic characteristics were topcoded in this year from the specified income source.

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

Table 4
Primary and secondary self-employment earnings cell means (1976–1987)

Year	Full-Time (35+ hours per week)			Full-Year (50+ weeks)			Not Full-Time, Full-Year Workers			
	Male, Not Black or Hispanic	Male, Not Black or Hispanic	Female, Not Black or Hispanic	Male, Not Black or Hispanic	Male, Not Black or Hispanic	Female, Not Black or Hispanic	Male, Not Black or Hispanic	Female, Not Black or Hispanic	Black or Hispanic	
1976	69286	N/A	62217 ^A	N/A	N/A	N/A	62217 ^A	N/A	N/A	N/A
1977	65763	79290 ^A	79290 ^A	79290 ^A	N/A	N/A	68452	N/A	N/A	N/A
1978	67885	N/A	56602 ^A	56602 ^A	N/A	N/A	62707	56602 ^A	N/A	56602 ^A
1979	68168	56976 ^A	56976 ^A	53503	N/A	N/A	71237	N/A	N/A	56976 ^A
1980	69954	72753 ^A	72753 ^A	55589	72753 ^A	N/A	70074	N/A	72753 ^A	N/A
1981	71645	56602 ^A	73083	81331	N/A	N/A	63109	N/A	56602 ^A	56602 ^A
1982	89700	N/A	95061 ^A	95061 ^A	N/A	N/A	92898	N/A	95061 ^A	95061 ^A
1983	88987	90964 ^A	90964 ^A	90608	N/A	90964 ^A	87585	N/A	N/A	N/A
1984	92506	86400 ^A	86400 ^A	89529	N/A	N/A	89588	N/A	N/A	86400 ^A
1985	99999	N/A	99999	99999	N/A	N/A	99999	N/A	99999	99999
1986	136144	108836 ^A	N/A	108836 ^A	N/A	108836 ^A	106879	N/A	N/A	108836 ^A
1987	130751	N/A	170968 ^A	170968 ^A	N/A	N/A	144319	N/A	170968 ^A	170968 ^A

^A – Indicates that there are not enough observations to report a cell mean for this population group in this year. The reported cell mean is a combined cell mean with the other footnoted population groups in the same year.

N/A – Indicates that no individual with these demographic characteristics were topcoded in this year from the specified income source.

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

Table 5
Primary and secondary farm earnings cell means (1976–1987)

Year	Full-Time (35+ hours per week), Full-Year (50+ weeks) Workers			Not Full-Time, Full-Year Workers		
	Male, Not Black or Hispanic	Female, Not Black or Hispanic	Hispanic	Male, Not Black or Hispanic	Female, Not Black or Hispanic	Hispanic
1976	67970 ^A	N/A	N/A	N/A	N/A	67970 ^A
1977	61813 ^A	N/A	N/A	N/A	N/A	61813 ^A
1978	60590	N/A	N/A	N/A	N/A	N/A
1979	63208 ^A	N/A	N/A	63208 ^A	N/A	N/A
1980	64447 ^A	N/A	N/A	64447 ^A	N/A	N/A
1981	61356 ^A	N/A	N/A	N/A	N/A	61356 ^A
1982	91015 ^A	N/A	N/A	91015 ^A	N/A	N/A
1983	82381	N/A	N/A	N/A	N/A	N/A
1984	83154 ^A	N/A	N/A	83154 ^A	N/A	N/A
1985	99999	N/A	N/A	N/A	N/A	N/A
1986	N/A	N/A	N/A	N/A	N/A	N/A
1987	122398 ^A	N/A	N/A	122398 ^A	N/A	N/A

^A – Indicates that there are not enough observations to report a cell mean for this population group in this year. The reported cell mean is a combined cell mean with the other footnoted population groups in the same year.

N/A – Indicates that no individual with these demographic characteristics were topcoded in this year from the specified income source. Source: Author's calculations using internal March CPS data.

Table 6
Non-labor cell means (1976–1987)

Year	Wages (I51A)	Self Employment (I51B)	Farm Security (I51C)	Social Security (I52A)	Supplemental Security (I52B)	Public Assistance (I53A)	Interest Dividends (I53B)	Rentals (I53C)	Veterans and Workers Comp (I53D)	Retirement (I53E)	Other (I53F)
1976	65193	68022	67970	N/A	N/A	N/A	73035 ^A	74448	N/A	73035 ^A	N/A
1977	64562	66620	61813	9999 ^B	N/A	N/A	71651 ^A	71651	N/A	71651 ^A	71651 ^A
1978	65687	66665	60590	9999 ^B	5999 ^B	N/A	62534 ^A	74304	N/A	N/A	62534 ^A
1979	66514	67764	63208	9999 ^B	5999 ^B	N/A	66269 ^A	70785	N/A	88167	66269 ^A
1980	67561	69583	64447	12893	5999 ^B	N/A	71163	74089	61135 ^A	63786	61135 ^A
1981	66367	70528	61356	11589	5999 ^B	N/A	72861	71240	33808	70004 ^A	70004 ^A
1982	91534	90562	91015	19999 ^B	5999 ^B	N/A	70242 ^A	93788	70242 ^A	81636	70242 ^A
1983	89595	88947	82381	N/A	5999 ^B	19999	97565 ^A	92724	42207	91843	97565 ^A
1984	90447	91829	83154	19999 ^B	6520	N/A	94024 ^C	87201	N/A	92638	89513
1985	99999 ^B	99999 ^B	99999 ^B	19999 ^B	9999 ^B	N/A	99999 ^B	99999 ^B	43336	99999 ^B	99999 ^B
1986	137113	129996	N/A	19999 ^B	N/A	N/A	99999 ^B	99999 ^B	41302	99999 ^B	99999 ^B
1987	140026	135346	122398	19999 ^B	9999 ^B	N/A	99999 ^B	99999 ^B	48662	99999 ^B	N/A

^A – Indicates that there are not enough observations to report a cell mean for this population group in this year. The reported cell mean is a combined cell mean with the other footnoted population groups in the same year.

^B – Indicates that the internal censoring point is identical to the public cell mean so no additional information can be obtained from the internal data.

^C – Interest income in 1984 does not properly match between the internal and the public data. This cell mean is based on the 25 individuals with interest income at or above \$75,000 topcode threshold in the internal data, not just the 11 people who are listed as topcoded for interest income in the public data.

N/A – Indicates that no individual with these demographic characteristics were topcoded in this year from the specified income source. Note: In parentheses below each variable name is the mnemonic for the income source from the March Current Population Survey Technical Documentation.

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

Table 7
Primary labor earnings cell means (1988–2002)

Year	Full-Time (35+ hours per week), Full-Year (50+ weeks) Workers				Not Full-Time, Full-Year Workers			
	Male, Not Black or Hispanic	Male, Black	Female, Not Black or Hispanic	Female, Black or Hispanic	Male, Not Black or Hispanic	Male, Black	Female, Not Black or Hispanic	Female, Black or Hispanic
1988	148852	136582	151838	153098 ^A	124539	153098 ^A	124581	143082
1989	143204	138971	154412	137250 ^A	160834	137250 ^A	N/A	125994
1990	153067	159309	153072	124782 ^A	173779	124782 ^A	124782 ^A	115220
1991	151763	144161	135010	132453	148513	160432	132230 ^A	127641
1992	142991	133707	136560	121099 ^A	140743	N/A	121099 ^A	128059
1993	148241	144800	143657	114123 ^A	141543	114123 ^A	N/A	139295
1994	188027	232995	205449	215571	188154	179478 ^A	179478 ^A	202593
1995	187347	180854	179894	160143	173157	458721 ^A	458721 ^A	154367
1996	302536	464791	257394	283521	268563	404595 ^A	404595 ^A	576398
1997	318985	391150	384160	357895	325794	454812 ^A	454812 ^A	222550
1998	330658	204326	309943	306469	330092	267650 ^A	267650 ^A	442032
1999	306732	266285	419044	402202	348516	N/A	492661 ^A	390509
2000	300974	257525	362315	256384	284124	244810 ^A	244810 ^A	284141
2001	335049	307007	281859	288962	321704	N/A	337247 ^A	195780
2002	320719	326982	331937	361315	319740	432873 ^A	432873 ^A	270371

^A – Indicates that there are not enough observations to report a cell mean for this population group in this year. The reported cell mean is a combined cell mean with the other footnoted population groups in the same year.

N/A – Indicates that no individual with these demographic characteristics were topcoded in this year from the specified income source. Source: Author's calculations using internal March CPS data.

Table 8
Secondary wage earnings cell means (1988–2002)

Year	Full-Time (35+ hours per week), Full-Year (50+ weeks) Workers				Not Full-Time, Full-Year Workers			
	Male, Not Black or Hispanic	Male, Black or Hispanic	Female, Not Black or Hispanic	Female, Black or Hispanic	Male, Not Black or Hispanic	Male, Black or Hispanic	Female, Not Black or Hispanic	Female, Black or Hispanic
1988	N/A	N/A	N/A	N/A	99999 ^B	N/A	N/A	99999 ^B
1989	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
1990	99999 ^B	N/A	N/A	N/A	99999 ^B	N/A	N/A	N/A
1991	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
1992	99999 ^B	N/A	N/A	99999 ^B	N/A	N/A	N/A	N/A
1993	99999 ^B	N/A	N/A	N/A	N/A	N/A	N/A	N/A
1994	158174	125569 ^A	N/A	125569 ^A	125569 ^A	N/A	N/A	N/A
1995	207148	109775 ^A	109775 ^A	109775 ^A	109775 ^A	N/A	202541	N/A
1996	64541	29777	183740	56978	35661 ^A	35661 ^A	41501	35661 ^A
1997	45749	62044 ^A	62044 ^A	48634	257107	62044 ^A	54580	N/A
1998	61345	51706	39942	48753	47529 ^A	35078	39354	47529 ^A
1999	59925	51139	52682	35583	34826	36827	59300	57830 ^A
2000	50037	35625	39676	51469	67776	50770	65966	31178 ^A
2001	56861	76598	39968	41433	39816	37788	39320	37721 ^A
2002	60672	49155	50535	43388	40556	65480	44255	48817

A – Indicates that there are not enough observations to report a cell mean for this population group in this year. The reported cell mean is a combined cell mean with the other footnoted population groups in the same year.

B – Indicates that the internal censoring point is identical to the public cell mean so no additional information can be obtained from the internal data.

N/A – Indicates that no individual with these demographic characteristics were topcoded in this year from the specified income source. Source: Author's calculations using internal March CPS data.

Table 9
Secondary self-employment earnings cell means (1988–2002)

Year	Full-Time (35+ hours per week)				Full-Year (50+ weeks)				Not Full-Time, Full-Year Workers				
	Male, Black or Hispanic	Male, Black	Female, Black or Hispanic	Female, Black	Male, Black or Hispanic	Male, Black	Female, Black or Hispanic	Female, Black	Male, Black or Hispanic	Male, Black	Female, Black or Hispanic	Female, Black	Hispanic
1988	99999 ^B	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
1989	99999 ^B	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
1990	99999 ^B	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	99999 ^B	N/A	N/A
1991	99999 ^B	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
1992	99999 ^B	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
1993	99999 ^B	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
1994	157513 ^A	N/A	N/A	N/A	157513 ^A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
1995	305001	N/A	357471 ^A	N/A	357471 ^A	N/A	N/A	N/A	N/A	N/A	357471 ^A	N/A	357471 ^A
1996	154533	82232 ^A	N/A	64058	N/A	64058	N/A	N/A	82232 ^A	N/A	82232 ^A	N/A	N/A
1997	128473	152704 ^A	152704 ^A	152704 ^A	152704 ^A	N/A	152704 ^A	N/A	N/A	N/A	152704 ^A	152704 ^A	N/A
1998	101769	N/A	104340 ^A	53482	N/A	104340 ^A	N/A	N/A	104340 ^A	104340 ^A	104340 ^A	104340 ^A	N/A
1999	123543	N/A	103546 ^A	52835	N/A	103546 ^A	N/A	N/A	103546 ^A	N/A	131515	N/A	N/A
2000	119583	N/A	64058 ^A	63258	N/A	64058 ^A	N/A	N/A	64058 ^A	64542	N/A	N/A	64058 ^A
2001	119739	59954 ^A	59954 ^A	59954 ^A	N/A	59954 ^A	N/A	N/A	88916	N/A	61946	N/A	59954 ^A
2002	127593	108081	79681	56935	49520 ^A	49520 ^A	49520 ^A	49520 ^A	98045	N/A	48880	N/A	49520 ^A

A – Indicates that there are not enough observations to report a cell mean for this population group in this year. The reported cell mean is a combined cell mean with the other footnoted population groups in the same year.

B – Indicates that the internal censoring point is identical to the public cell mean so no additional information can be obtained from the internal data.

N/A – Indicates that no individual with these demographic characteristics were topcoded in this year from the specified income source. Source: Author's calculations using internal March CPS data.

Table 10
Secondary farm earnings cell means (1988–2002)

Year	Full-Time (35+ hours per week), Full-Year (50+ weeks) Workers				Not Full-Time, Full-Year Workers			
	Male, Not Black or Hispanic	Male, Black or Hispanic	Female, Not Black or Hispanic	Female, Black or Hispanic	Male, Not Black or Hispanic	Male, Black or Hispanic	Female, Not Black or Hispanic	Female, Black or Hispanic
1988	9999 ^B	N/A	N/A	N/A	N/A	N/A	N/A	N/A
1989	N/A	N/A	N/A	N/A	9999 ^B	N/A	N/A	N/A
1990	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
1991	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
1992	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
1993	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
1994	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
1995	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
1996	53068	45717 ^A	N/A	45717 ^A	N/A	N/A	45717 ^A	N/A
1997	38782 ^A	N/A	N/A	38782 ^A	N/A	N/A	N/A	N/A
1998	90173	N/A	N/A	61128	N/A	92604 ^A	92604 ^A	92604 ^A
1999	65336	N/A	N/A	44560 ^A	N/A	44560 ^A	N/A	N/A
2000	87162	N/A	51354 ^A	54785	N/A	51354 ^A	62400	51354 ^A
2001	133766	186210 ^A	186210 ^A	186210 ^A	186210 ^A	186210 ^A	186210 ^A	186210 ^A
2002	44548	303613 ^A	49415	431361	303613 ^A	45603	303613 ^A	360924

^A – Indicates that there are not enough observations to report a cell mean for this population group in this year. The reported cell mean is a combined cell mean with the other footnoted population groups in the same year.

^B – Indicates that the internal censoring point is identical to the public cell mean so no additional information can be obtained from the internal data.

N/A – Indicates that no individual with these demographic characteristics were topcoded in this year from the specified income source. Source: Author's calculations using internal March CPS data.

Table 11
Non-labor earnings cell means (1988–2002)

Year	Primary Earnings (ERN_VAL)	Wages (WS_VAL)	Self Employment (SE_VAL)	Farm (FRM_VAL)	Social Security (SS_VAL)	Supplemental Security (SSI_VAL)	Public Assistance (PAW_VAL)	Interest (INT_VAL)	Dividends (DIV_VAL)	Rental (RNT_VAL)	Alimony (ALM_VAL)	Child Support (CSP_VAL)
1988	147389	99999 ^B	99999 ^B	99999 ^B	29999 ^B	9999	N/A	99999 ^B	99999 ^B	99999 ^B	N/A	N/A
1989	145322	N/A	125624	99999 ^B	N/A	9999	N/A	99999 ^B	99999 ^B	99999 ^B	N/A	N/A
1990	153438	99999 ^B	99999 ^B	N/A	29999 ^B	9999	19999 ^B	99999 ^B	99999 ^B	99999 ^B	99999 ^B	N/A
1991	150805	N/A	99999 ^B	N/A	29999 ^B	9999	N/A	99999 ^B	99999 ^B	99999 ^B	N/A	N/A
1992	141202	99999 ^B	99999 ^B	N/A	29999 ^B	9999	N/A	99999 ^B	99999 ^B	99999 ^B	N/A	N/A
1993	147511	99999 ^B	99999 ^B	N/A	29999 ^B	9999	19999 ^B	99999 ^B	99999 ^B	99999 ^B	N/A	N/A
1994	192210	144956	153328	153328	49999 ^B	14933	24999 ^B	99999 ^B	99999 ^B	99999 ^B	N/A	N/A
1995	188180	177066	319060	319060	50000 ^B	14154	24999 ^B	99999 ^B	99999 ^B	99999 ^B	99999 ^B	N/A
1996	308691	65394	125234	49392	50000 ^B	25000 ^B	25000 ^B	99999 ^B	99999 ^B	99999 ^B	99999 ^B	99999 ^B
1997	326907	63032	134769	38782	50000 ^B	25000 ^B	25000 ^B	99999 ^B	99999 ^B	99999 ^B	99999 ^B	99999 ^B
1998	326155	61078	98663	86012	50000 ^B	25000 ^B	25000 ^B	99999 ^B	99999 ^B	99999 ^B	99999 ^B	99999 ^B
1999	325365	53368	117226	60370	50000 ^B	25000 ^B	25000 ^B	99999 ^B	99999 ^B	99999 ^B	99999 ^B	99999 ^B
2000	295041	52527	97311	62726	50000 ^B	25000 ^B	25000 ^B	60820	36877	57453	93114	26592
2001	325563	51599	104848	157047	50000 ^B	25000 ^B	25000 ^B	63005	36962	55220	54009	23918
2002	325382	55636	99362	167762	50000 ^B	25000 ^B	25000 ^B	61107	38907	54644	62078	27947
								64853	38962	57416	63554	25657

^B – Indicates that the internal censoring point is identical to the public cell mean so no additional information can be obtained from the internal data.

N/A – Indicates that no individual with these demographic characteristics were topcoded in this year from the specified income source.

Note: In parentheses below each variable name is the mnemonic for the income source from the March Current Population Survey Technical Documentation.

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

Table 11, continued
Non-labor earnings cell means (1988–2002)

Year	Unemployment (UC_VAL)	Workers Comp (WC_VAL)	Veterans (VET_VAL)	Retirement 1st source (RET_VAL1)	Retirement 2nd Source (RET_VAL2)	Survivors 1st Source (SUR_VAL1)	Survivors 2nd Source (SUR_VAL2)	Disability 1st Source (DIS_VAL1)	Disability 2nd Source (DIS_VAL2)	Education Assistance (ED_VAL)	Financial Assistance (FIN_VAL)	Other (OL_VAL)
1988	N/A	N/A	29999B	99999B	N/A	99999B	99999B	99999B	N/A	N/A	N/A	N/A
1989	N/A	N/A	29999	99999B	N/A	99999B	N/A	N/A	N/A	N/A	N/A	99999B
1990	N/A	99999B	29999	99999B	N/A	99999B	N/A	99999B	N/A	N/A	N/A	99999B
1991	N/A	99999B	29999	99999B	N/A	99999B	99999B	N/A	N/A	N/A	N/A	N/A
1992	N/A	N/A	29999	99999B	99999B	99999B	N/A	99999B	N/A	99999B	N/A	99999B
1993	N/A	N/A	29999	99999B	99999B	99999B	N/A	N/A	N/A	99999B	99999B	99999B
1994	N/A	N/A	N/A	99999B	N/A	99999B	N/A	99999B	N/A	N/A	N/A	N/A
1995	99999B	99999B	N/A	99999B	99999B	99999B	N/A	99999B	N/A	N/A	99999B	99999B
1996	99999B	99999B	99999B	99999B	99999B	99999B	99999B	99999B	N/A	N/A	99999B	99999B
1997	N/A	N/A	99999B	99999B	N/A	99999B	N/A	99999B	N/A	99999B	99999B	99999B
1998	99999B	99999B	N/A	99999B	99999B	99999B	N/A	99999B	N/A	N/A	N/A	99999B
1999	N/A	99999B	99999B	62277	49551	88985C	88985C	55327	54856	27809	40717	44265
2000	99999B	99999B	N/A	63210	62139	83427C	83427C	73486D	73486D	25585	54559	50217
2001	99999B	99999B	99999B	64547	62457	82077	84408	60657D	60657D	31902	48059	40865
2002	99999B	N/A	99999B	64501	65080	90937C	90937C	52859	N/A	34876	46155	47179

B – Indicates that the internal censoring point is identical to the public cell mean so no additional information can be obtained from the internal data.
 C – Indicates that there are not enough individuals topcoded for INCSI2 to report a cell mean in this year, so the cell mean reported is a combined cell mean for INCSI1 and INCSI2.

D – Indicates that there are not enough individuals topcoded for INCDS2 to report a cell mean in this year, so the cell mean reported is a combined cell mean for INCDS1 and INCDS2.

N/A – Indicates that no individual with these demographic characteristics were topcoded in this year from the specified income source.

Note: In parentheses below each variable name is the mnemonic for the income source from the March Current Population Survey Technical Documentation. Source: Author's calculations using internal March CPS data.

Therefore, for years prior to 1996 we instead treat individuals who have income exactly at the topcode as topcoded and include their income in the cell means. For many years there is no difference between these two approaches as no individuals have reported incomes exactly at the topcode value. Hence this distinction is relevant only in years where topcodes are round numbers so clustering in reported values occurs at the topcode level. The actual number of individuals whose reported income is the same as the topcode is provided in Appendix Table 6.

Starting in 1996, when the US Census Bureau introduced cell means, income for individuals who face a binding topcode constraint are listed with their cell mean income in the public use data, so they can be distinguished from individuals for whom the topcode does not bind by users of the public use data. Thus, starting in 1996 we follow the census technique and only include individuals with income greater than the topcode level in our cell means.

The second deviation from the method described above occurs when less than 5 individuals are topcoded from a non-labor income source or from a demographic group within that income source for a labor earnings source. For confidentiality reasons, the US Census Bureau would not allow us to release cell means for these groups. Instead, we combined 2 or more cells to achieve a group of at least 5 individuals. We then provide the mean of topcoded incomes from this combined cell. This is the same procedure for handling small cells used by the US Census Bureau in the cell means they have released in recent years (2007 CPS Annual Social and Economic Supplement Technical Documentation).

For labor earnings sources we used the current US Census Bureau convention of combining demographic groups with less than 5 topcoded individuals into a single cell for that income source. If there were still less than 5 individuals, we combined it with the next smallest cell from the same income source to achieve a cell with more than 5 people.

For non-labor income sources, it was necessary to combine observations across sources of income. For these, we first attempted to combine similar sources of income, such as 1st and 2nd source disability income. If there was no clear matching income source, we combined the small cells into a single group, as was done for the labor income sources. In all cases, when we combine cells due to the small cell size, we note it in the relevant tables.

There are also several missing values (“N/A”) in the extended cell mean tables we provide. Before the introduction of US Census Bureau-provided cell means in 1996, missing values indicate that no one has been topcoded from the specified source of income or fits that demographic group. After 1996, it indicates either: (1) no one is topcoded from the income source or (2) that the internal censoring point is equal to the public topcode so no one has income in the internal dataset that is above the public topcode. In both cases, no new information can be gained using internal data or the cell means series.⁴

⁴Most of the US Census Bureau procedures for creating cell means for topcoded values in the public use

5. Limitation of using the extended cell mean series

While the extended cell mean series is useful for studies using income data from the public use March CPS, it also has limitations. As we will show it matches the mean values in the internal March CPS quite closely but in doing so, it reflects the same problems with censorship of those at the highest end of each source of income and their households in the internal March CPS. However, since this is the same data used to produce official US Census Bureau statistics, these remaining censorship problems for users of the public use March CPS data are no worse than those used by the US Census Bureau to produce its official income statistics with the internal March CPS.

Because our cell mean series, as well as the US Census Bureau cell mean series, are based on the internal March CPS data, these internal censoring issues limit the additional information that the cell means provide. Specifically, one should pay careful attention to changes in the internal censoring points which introduce trend-breaks in our cell mean series. Since 1976, there were 3 years where internal censoring points were adjusted substantially and thus trends at these years should be considered with caution.

The first is 1986, when the internal censoring points for labor income sources increased from \$99,999 to \$250,000. The second is 1988 when the income sources were redefined and expanded from 11 to 24, redefining censoring points in the process. The third is 1994 when the internal censoring points were increased from \$299,999 for primary labor earnings and \$99,999 for secondary labor earnings to \$999,999 for all labor earnings sources. A fourth increase occurred in 1995, when internal censoring points for primary and secondary wage earnings were increased to \$1,099,999. However in terms of consistency, very few individuals fall into this range, so this increase is much less troublesome than the other three adjustments. (See Appendix Tables 2–5 for the topcoded value for all sources of income.)

In addition, because cell means are based on the internal March CPS, which is subject to some censoring at the very top of the distribution, the cell means series does not reveal the full scope of the US income distribution. However, while the public use March CPS generally provides complete information for only around 94–98% of the sample population, the internal March CPS is able to provide income information for 99.0 to 99.8% of the sample distribution.

The percentages of the population that are topcoded in the public use March CPS and censored in the internal March CPS are illustrated in Fig. 1. Since the public use March CPS with our extended cell means series incorporates information about the incomes for all individuals below the internal March CPS censoring points, results

March CPS data can be found in the 2007 Current Population Survey Annual Demographic File Technical Documentation [10], but in some cases we learned about them via conversations with various US Census Bureau employees charged with creating the cell means.

using this series can provide information to researchers on the 1 to 5% of the sample population that is unobserved in the public use March CPS without cell means but are not impacted by the internal March CPS censoring points. Additionally, for the small fraction of individuals who are censored in the internal March CPS, their cell mean value will be significantly closer to their true income than their public topcode value.

6. Impact of using the extended cell mean series

Below we show the added value of using our cell mean series with the public use March CPS. We do so by first comparing mean income in the United States since 1976 using our cell mean series with other available topcoding methods. We then do the same comparing Gini coefficients for each year since 1976 using the other topcoding methods.

In Table 12 we calculate mean household income using our extended cell mean series (Cell Mean) and compare it to the public use March CPS data including the US Census Bureau cell means since 1996 series (Unadjusted), the public use March CPS data removing the US Census Bureau provided cell means since 1996 series (Adjusted), and the internal March CPS data series (Internal).

Thanks to cell means, the mean income of individuals in 2002 captured in the Unadjusted data is very close to our Cell Mean data and both are very close to the values in the 2002 Internal data. So for those only interested in evaluating the mean household income of individuals in the United States in 2002, the current Unadjusted data or our Cell Mean data nicely capture the means in the Internal data. And this is true for all years since 1996 when cell means were first provided by the US Census Bureau.

But for those interested in the trends in these series prior to 1996 the Unadjusted data is flawed because it does not provide cell means for persons above the topcoded values. Hence its mean values are significantly smaller than those produced using the Internal data. In 1995, for example, using the Unadjusted data will understate mean household income in the United States by 5.8% when compared to mean household income using Internal data. In contrast, our Cell Mean data provide yearly means very close to those from the Internal data for all years.

Since topcoding may have drastically different impacts when looking at different income sources and populations, in addition to examining the change in mean household income, we also looked at mean labor earnings at the individual level for full-time/full-year workers. These results are presented in Table 13. The patterns are similar to those discussed in Table 12. When analyzing only the years after 1996, the Unadjusted and Cell Mean series both do a good job of capturing the mean labor earnings found in the Internal data. However, prior to 1996 there is again a significant difference between the results using the Unadjusted series and the Internal series. In fact, the percentage difference between these series is larger than those for

Table 12
 Mean Household Income over time using different topcode series^a

	Public Unadjusted	Public Adjusted	Public Cell means	Internal
1976	15,535	15,535	15,691	15,696
1977	16,861	16,861	17,047	17,044
1978	18,162	18,162	18,389	18,392
1979	19,947	19,947	20,259	20,251
1980	22,017	22,017	22,450	22,457
1981	23,545	23,545	24,012	24,012
1982	25,614	25,614	25,782	25,778
1983	27,124	27,124	27,331	27,331
1984	28,266	28,266	28,680	28,682
1985	30,723	30,723	30,725	30,725
1986	32,290	32,290	32,609	32,610
1987	34,153	34,153	34,610	34,609
1988	36,069	36,069	36,637	36,636
1989	37,535	37,535	38,225	38,255
1990	40,064	40,064	41,061	41,078
1991	41,087	41,087	42,015	42,016
1992	41,716	41,716	42,505	42,503
1993	42,752	42,752	43,806	43,815
1994	44,077	44,077	46,588	46,716
1995	45,970	45,970	48,752	48,802
1996	50,697	48,409	50,697	50,704
1997	53,011	50,439	53,011	52,960
1998	56,019	53,077	56,019	56,006
1999	58,534	54,820	58,535	58,564
2000	60,725	57,903	61,982	61,895
2001	65,115	60,181	65,115	65,064
2002	66,272	60,901	66,272	66,167
2003	66,209	62,419	66,209	66,218
2004	67,558	63,805	67,558	67,597
2005	69,131	65,000	69,131	69,052
2006	72,409	67,680	72,409	
2007	76,262	70,663	76,262	

Note: Through 2002, the Public Cell means series uses our generated cell means that are provided in Tables 3–11. After 2002, the series uses the Census Provided cell means which can be used to extend the cell mean series beyond 2002.

Source: Author’s calculations using internal March CPS data.

household income since many of the individuals whose income is significantly over the public topcodes are topcoded for labor earnings. In 1995, for example, using the Unadjusted data will understate labor earnings by 7% compared to labor earnings using the Cell Mean or Internal data. In contrast to these three series, the Adjusted series understates mean household income in every year since it does not use cell means to adjust for top coding.

More interesting than looking at the impact of using cell means on mean income is how the choice of topcoding procedure impacts various measures of inequality. In

Table 13
Mean labor earnings income using different topcode series^a

Year	Public Unadjusted	Public Adjusted	Public Cell Means	Internal
1976	11,879	11,879	12,011	12,011
1977	12,714	12,714	12,870	12,871
1978	13,561	13,561	13,754	13,753
1979	14,638	14,638	14,899	14,900
1980	15,893	15,893	16,249	16,250
1981	17,072	17,072	17,470	17,470
1982	18,729	18,729	18,883	18,883
1983	19,937	19,937	20,125	20,124
1984	20,776	20,776	20,988	20,988
1985	22,112	22,112	22,112	22,112
1986	23,157	23,157	23,450	23,453
1987	24,174	24,174	24,585	24,585
1988	25,070	25,070	25,583	25,583
1989	26,074	26,074	26,636	26,636
1990	27,472	27,472	28,342	28,342
1991	28,016	28,016	28,853	28,853
1992	29,014	29,014	29,725	29,725
1993	29,972	29,972	30,925	30,925
1994	30,647	30,647	32,890	32,890
1995	31,680	31,680	34,069	34,065
1996	34,777	32,963	34,782	34,777
1997	36,501	34,265	36,504	36,501
1998	37,788	35,460	37,790	37,788
1999	39,096	36,715	39,096	39,096
2000	39,738	38,211	40,734	40,747
2001	43,048	39,705	43,039	43,048
2002	44,809	41,051	44,799	44,809
2003	45,485	42,960	45,485	45,490
2004	46,408	44,008	46,408	46,408
2005	47,313	44,636	47,313	47,288
2006	49,054	46,092	49,054	
2007	50,989	47,611	50,989	

Note: Through 2002, the Public Cell means series uses our generated cell means that are provided in Tables 3–11. After 2002, the series uses the Census Provided cell means which can be used to extend the cell mean series beyond 2002.

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

Burkhauser and Larrimore [3], we examined the impact of topcoding on one type of cross-group inequality and find that without correcting for topcoding through the use of cell means you will understate the drop in the economic well being of individuals with disabilities over the past two decades.

Using cell means also impacts measures of population level inequality such as the Gini Coefficient. In Fig. 2 we report the levels and trends in size-adjusted household income for 1976–2007 based on Gini values estimated from our Cell Means, Unadjusted, Adjusted, and Internal data series. As is standard when looking at household income inequality, we controlled for household size by examining the

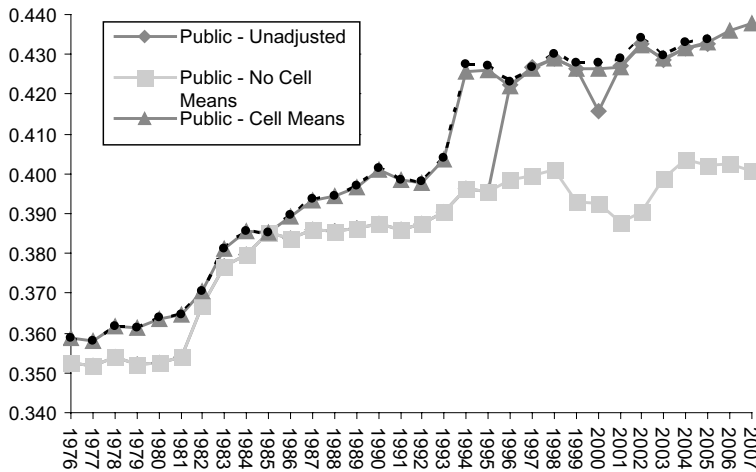


Fig. 2. Comparing Gini-trends using four different topcode methods. Note that internal data was not available for years after 2005.

size-adjusted household income of individuals instead of simply their household's total income.⁵ To adjust for the number of individuals in the household, we divided the total income in the household by the square root of the number of household members.

Because the public use March CPS is topcoded, unless cell means are used, mean household income and its distribution across the population (Gini values) will be understated relative to those estimated from the internal March CPS. Since cell means provide more information about individuals at the top of the income distribution in this topcoded data set, using the cell mean series with the public use March CPS data will increase measured inequality and hence increase measured inequality in the public use March CPS data.

In all years the Adjusted series that reports only the topcode value for all those above the topcode is below the Internal data. To correct for this difference between inequality measured using the Adjusted and Internal Gini values, the US Census Bureau introduced topcoding in 1996. The result is reflected in the difference between the Unadjusted and Adjusted values in Fig. 2. Prior to 1996 they are identical. After 1996 the Unadjusted values now equal the Internal values and are dramatically greater than the Adjusted values. While they now better represent the values found in the Internal data, anyone uncritically using this series to explain changes in United States income inequality would greatly exaggerate its increase in 1996.

⁵We follow the same procedure for generating size-adjusted household income as discussed in Burkhauser and Larrimore [3]. Also see Gottschalk and Smeeding [5] for a review of the general income distribution literature and Gottschalk and Danziger [6] for a more detailed discussion of these usual assumptions made in this literature.

Our Cell Mean series corrects this problem by accurately matching the Internal series Gini values in all years. Hence it provides a consistent way of capturing inequality as measured using the Internal data. But anyone uncritically using this series to explain changes in United States income inequality would also slightly exaggerate its increase in 1986 and in 1988 when there were increases to the internal censoring points and would greatly exaggerate its increase in 1994 when major changes occurred in the way the March CPS was collected including significant changes in the internal censoring points.. All these issues are discussed in more detail in Burkhauser, Feng, and Jenkins [2]. Hence while our cell means series provides researchers without access to the internal March CPS data with a useful way to approximate its means and distribution, it does not solve the problems created by censoring of the internal March CPS data itself.

7. Conclusions and recommendations

The fraction of individuals who are topcoded is growing in the public use March CPS, limiting the ability of researchers to examine trends at the top of the income distribution as well as trends to the income distribution as a whole. While cell means can help alleviate this problem, because the US Census Bureau-provided cell mean series began in 1996, its usefulness is restricted to income trend studies since then. Here we not only describe how we created a cell means series beginning in 1976 for the public use March CPS data using internal March CPS data but provide that series to all researchers. We also demonstrate that this series can be used in conjunction with the public use March CPS to provide results in a variety of contexts that closely match those using the internal March CPS data and do so better than other available methods in the income literature.

Prior to the introduction of cell means, because the US Census Bureau did not systematically adjust its topcodes in the public use March CPS to the same percentile in the income distribution each year, measured mean income and its distribution in the public use March CPS did not consistently match such measured values using the internal March CPS data in level or trend. The introduction of cell means has dramatically reduced this problem. All income that is reported below the internal censoring levels of the internal March CPS data can now be captured in income statistics using the public use March CPS data, thus mitigating or eliminating the inconsistencies caused by changing public topcodes, depending on the income statistics observed.

However, the full implications of the introduction of cell means on US Census Bureau policies with respect to where to place topcodes each year, has not been fully considered. Prior to the provision of cell means, it might have been reasonable to believe that researchers would prefer that topcodes not be systematically changed each year. But that is no longer the case. Because systematically updating topcodes each year will provide more accurate data to users of the public use March CPS, the US Census Bureau should annually update these values. In doing so rather than

maintaining consistency in the nominal value of topcodes from previous years, new topcodes should be set at the highest possible level that both maintain confidentiality requirements and keeps enough individuals above the topcode threshold to produce separate cell means for each non-labor income source.

The US Census Bureau's effort to maintain consistency in the nominal values of its topcodes has resulted in information being censored in the public use March CPS far in excess of the DRB's requirements for protecting the confidentiality of respondents. In recent years, over 5% of respondents with positive income in the public use March CPS have been topcoded for secondary wage, labor, and farm earnings as well as from some non-labor income sources. This is the case even though the DRB's confidentiality requirements only require 3% of these individuals to be topcoded. Given that there is no significant cost to providing full information about individuals' incomes up to the 3% threshold set by the DRB, the US Census Bureau should increase the topcodes for these income sources to this threshold.

A second factor the US Census Bureau should consider in updating its topcodes is the effect that these changes will have on its ability to provide cell means for each income source. An increase in topcodes allows researchers to observe the true income of more individuals. Additionally, as the remaining group who are topcoded shrinks, their cell mean better estimates each of their reported incomes. However, this gain in information is offset to some degree if the new topcode reduces the remaining group below 5, since this cell mean must be merged with the cell mean from another income source to protect data confidentiality.

For non-labor earnings, adjusting the topcode for one source of income has no effect on the topcode level for other income sources. Thus, it is relatively easy to lower the topcode for a single source of income to ensure that at least 5 individuals are topcoded and a separate cell mean can be produced. Since there are benefits to having source-specific cell means, we urge the US Census Bureau to consider adjusting the public topcodes for each income source annually with the topcode set at the highest possible income point for that income source without lowering the population who are topcoded to below 5, except when this could unduly risk the confidentiality of respondents.

For labor income sources where demographic sub-cells are provided, avoiding combinations across sub-cells is still beneficial but the cost of doing so is much greater. While the topcode for a non-labor income source can be dropped just enough to topcode an additional one or two individuals and hence produce a separate cell mean for a given non-labor income source, if the topcode is lowered for a labor income source to provide a separate cell mean for one demographic group, the topcode must be lowered for all demographic groups from that labor source. As a result, obtaining this separate cell mean for labor income demographic groups may come at the cost of topcoding dozens of additional people. Thus, for labor income sources, the US Census Bureau should focus on setting topcodes at the highest possible levels even if it requires combining additional demographic groups for that labor income source when producing cell means.

The US Census Bureau should also make the general public more aware of the problem of inconsistent censoring in the internal March CPS, now that the public use March CPS is better able to match internal March CPS income levels and trends. Even using our cell mean series, the public use March CPS will still reflect time-inconsistencies caused by different levels of censoring in the internal March CPS data over time. The US Census Bureau should consider keeping the point where these internal censoring points hit the distribution of income for a given income source consistent over time. This is especially true since the lack of publicly available March CPS data around these internal censoring points limits researchers' ability to self-generate consistent trends when topcodes increase. Additionally, given their increased importance, it would be prudent to include the internal censoring levels in the documentation files for the public use March CPS along with the public topcode levels. Doing so in a more prominent way would better alert users of the public use March CPS to changes in these levels that could influence the consistency of these data over time.

Despite these caveats, the cell means series provided to researchers in this paper through the cooperation of the US Census Bureau is of great value to the research community. When used with the public use March CPS it will allow researchers to closely match the mean and the distribution of income of the internal March CPS data for all March CPS years since 1976. No other publicly available data set provides such a long term look at trends in income and income distribution in the United States.

Acknowledgements

The research in this paper was conducted while the first three authors were Special Sworn Status researchers of the US Census Bureau at the New York Census Research Data Center at Cornell University and the fourth author was an employee of the US Census Bureau. Research results and conclusions expressed are those of the authors and do not necessarily reflect the views of the US Census Bureau. This paper has been screened to ensure that no confidential data are disclosed. Support for this research from the National Science Foundation (award nos. SES-0427889 SES-0322902, and SES-0339191) and the National Institute for Disability and Rehabilitation Research (H133B040013 and H133B031111). We thank Lisa Marie Dragoset, Arnie Reznick, the Cornell Census RDC Administrators, and all their US Census Bureau colleagues who have helped with this project. We also thank Robert Weathers and Stephen Jenkins for their comments on earlier drafts of this paper.

References

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

- [2] R.V. Burkhauser, S. Feng and S. Jenkins, *Using the P90/P10 Ratio to Measure US Inequality Trends with Current Population Survey Data: A View from Inside the Census Bureau Vaults*, ISER Working Paper 2007-14. Colchester, UK: Institute for Social and Economic Research, University of Essex, 2007, <http://www.iser.essex.ac.uk/pubs/workpaps/pdf/2007-14.pdf>.
- [3] R.V. Burkhauser and J. Larrimore, *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, 2008, <http://www.ces.census.gov/index.php/ces/cespapers>.
- [4] S. Feng, R.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* **24**(1) (2006), 57–62.
- [5] P. Gottschalk and T.M. Smeeding, Cross-National Comparisons of Earnings and Income Inequality, *Journal of Economic Literature* **35** (June 2006), 633–687.
- [6] P. Gottschalk and S. Danziger, Inequality of Wage Rates, Earnings and Family Income in the United States, 1975–2002, *Review of Income and Wealth* **51**(2) (2005), 231–254.
- [7] R.I. Lerman, US Wage-Inequality Trends and Recent Immigration, *American Economic Review* **89** (1999), 23–28.
- [8] F. Levy and R.J. Murnane, US Earnings Levels and Earnings Inequality: A Review of Recent Trends and Proposed Explanations, *Journal of Economic Literature* **30**(3) (1992), 1333–1381.
- [9] US Census Bureau, *Income, Poverty, and Health Insurance Coverage in the United States: 2006. (August)*, Current Population Reports, Consumer Income, GPO, Washington DC, 2007.
- [10] US Census Bureau, *Current Population Survey Annual Social and Economic Supplement Technical Documentation*, GPO, Washington DC, 2007.
- [11] E.J. Welniak, Measuring Household Income Inequality Using the CPS, in: *Special Studies in Federal Tax*, J. Statistics, Dalton and B. Kilss, eds, Statistics of Income Directorate, Inland Revenue Service, Washington DC, 2003.
- [12] L. Zayatz, *Disclosure Review, Census Bureau Standard*, 2004 http://www.census.gov/quality/quality_standards.htm.

Appendix Table 1. Income items reported in the current population survey

Name	Name in Public Files	Name in Internal Files	Definition
1976–1987			
<i>Labor Earnings</i>			
Wages	I51A	WSAL_VAL	Wages and Salaries
Self Employment	I51B	SEMP_VAL	Self employment income
Farm	I51C	FRSE_VAL	Farm income
<i>Other Sources</i>			
Social Security	I52A	I52A_VAL	Income from Social Security and/or Railroad Retirement
Supplemental Security	I52B	SSL_VAL	Supplemental Security Income
Public Assistance	I53A	PAW_VAL	Public Assistance
Interest	I53B	INT_VAL	Interest
Dividends Rentals	I53C	I53C_VAL	Dividends, Rentals, Trust Income
Veterans	I53D	I53D_VAL	Veteran's, unemployment, worker's compensation
Retirement	I53E	I53E_VAL	Pension Income
Other	I53F	I53F_VAL	Alimony, Child Support, Other income
1988–2007			
<i>Labor Earnings</i>			
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
<i>Other Sources</i>			
Social Security	SS_VAL	SS_VAL	Social Security Income
Supplemental Security	SSL_VAL	SSL_VAL	Supplemental Security Income
Public Assistance	PAW_VAL	PAW_VAL	Public Assistance & Welfare Income
Interest	INT_VAL	INT_VAL	Interest
Dividends	DIV_VAL	DIV_VAL	Dividends
Rental	RNT_VAL	RNT_VAL	Rental income
Alimony	ALM_VAL	ALM_VAL	Alimony income
Child Support	CSP_VAL	CSP_VAL	Child Support Income
Unemployment	UC_VAL	UC_VAL	Unemployment income
Workers Comp	WC_VAL	WC_VAL	Worker's compensation income
Veterans	VET_VAL	VET_VAL	Veteran's Benefits
Retirement – Source 1	RET_VAL1	RET_VAL1	Retirement income – source 1
Retirement – Source 2	RET_VAL2	RET_VAL2	Retirement income – source 2
Survivors – Source 1	SUR_VAL1	SUR_VAL1	Survivor's income – source 1
Survivors – Source 2	SUR_VAL2	SUR_VAL2	Survivor's income – source 2
Disability – Source 1	DIS_VAL1	DIS_VAL1	Disability income – source 1
Disability – Source 2	DIS_VAL2	DIS_VAL2	Disability income – source 2
Education assistance	ED_VAL	ED_VAL	Education assistance
Financial assistance	FIN_VAL	FIN_VAL	Financial Assistance
Other	OI_VAL	OI_VAL	Other income

Sources: Current Population Survey Annual Demographic File Technical Documentation [10], 1976–2002.

Current Population Survey Annual Social and Economic Supplement Technical Documentation [10], 2003–2007.

Appendix Table 2. Public use CPS censoring points for each income source in Dollars (1976–1987)

Year	Wages (I51A)	Self Employment (I51B)	Farm Security (I52A)	Social Security (I52B)	Supplemental Security (I53A)	Public Assistance (I53B)	Interest Dividends (I53C)	Rentals (I53D)	Veterans Workers Comp (I53E)	Retirement (I53F)	Other (I53G)
1976	50,000	50,000	9,999	5,999	19,999	50,000	50,000	50,000	29,999	50,000	50,000
1977	50,000	50,000	9,999	5,999	19,999	50,000	50,000	50,000	29,999	50,000	50,000
1978	50,000	50,000	9,999	5,999	19,999	50,000	50,000	50,000	29,999	50,000	50,000
1979	50,000	50,000	9,999	5,999	19,999	50,000	50,000	50,000	29,999	50,000	50,000
1980	50,000	50,000	9,999	5,999	19,999	50,000	50,000	50,000	29,999	50,000	50,000
1981	50,000	50,000	9,999	5,999	19,999	50,000	50,000	50,000	29,999	50,000	50,000
1982	75,000	75,000	19,999	5,999	19,999	75,000	75,000	75,000	29,999	75,000	75,000
1983	75,000	75,000	19,999	5,999	19,999	75,000	75,000	75,000	29,999	75,000	75,000
1984	75,000	75,000	19,999	5,999	19,999	75,000	75,000	75,000	29,999	75,000	75,000
1985	99,999	99,999	19,999	9,999	19,999	99,999	99,999	99,999	29,999	99,999	99,999
1986	99,999	99,999	19,999	9,999	19,999	99,999	99,999	99,999	29,999	99,999	99,999
1987	99,999	99,999	19,999	9,999	19,999	99,999	99,999	99,999	29,999	99,999	99,999

Source: Current Population Survey Annual Demographic File Technical Documentation

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

Appendix Table 3. Public use CPS censoring points for each income source in Dollars (1988–2007)

Year	Primary Earnings (ERN_VAL)	Wages (WS_VAL)	Self Employment (SE_VAL)	Farm (FRM_VAL)	Social Security (SS_VAL)	Supplemental Security (SSL_VAL)	Public Assistance (PAW_VAL)	Interest (INT_VAL)	Dividends (DIV_VAL)	Rental (RNT_VAL)	Alimony (ALM_VAL)	Child Support (CSP_VAL)
1988	99,999	99,999	99,999	99,999	29,999	9,999	19,999	99,999	99,999	99,999	99,999	99,999
1989	99,999	99,999	99,999	99,999	29,999	9,999	19,999	99,999	99,999	99,999	99,999	99,999
1990	99,999	99,999	99,999	99,999	29,999	9,999	19,999	99,999	99,999	99,999	99,999	99,999
1991	99,999	99,999	99,999	99,999	29,999	9,999	19,999	99,999	99,999	99,999	99,999	99,999
1992	99,999	99,999	99,999	99,999	29,999	9,999	19,999	99,999	99,999	99,999	99,999	99,999
1993	99,999	99,999	99,999	99,999	29,999	9,999	19,999	99,999	99,999	99,999	99,999	99,999
1994	99,999	99,999	99,999	99,999	49,999	9,999	24,999	99,999	99,999	99,999	99,999	99,999
1995	99,999	99,999	99,999	99,999	49,999	9,999	24,999	99,999	99,999	99,999	99,999	99,999
1996	150,000	25,000	40,000	25,000	49,999	25,000	24,999	99,999	99,999	99,999	99,999	99,999
1997	150,000	25,000	40,000	25,000	49,999	25,000	24,999	99,999	99,999	99,999	99,999	99,999
1998	150,000	25,000	40,000	25,000	49,999	25,000	24,999	99,999	99,999	99,999	99,999	99,999
1999	150,000	25,000	40,000	25,000	49,999	25,000	24,999	35,000	15,000	25,000	50,000	15,000
2000	150,000	25,000	40,000	25,000	49,999	25,000	24,999	35,000	15,000	25,000	40,000	15,000
2001	150,000	25,000	40,000	25,000	49,999	25,000	24,999	35,000	15,000	25,000	40,000	15,000
2002	150,000	25,000	40,000	25,000	49,999	25,000	24,999	35,000	15,000	25,000	40,000	15,000
2003	200,000	35,000	50,000	25,000	49,999	25,000	24,999	25,000	15,000	40,000	45,000	15,000
2004	200,000	35,000	50,000	25,000	49,999	25,000	24,999	25,000	15,000	40,000	45,000	15,000
2005	200,000	35,000	50,000	25,000	49,999	25,000	24,999	25,000	15,000	40,000	45,000	15,000
2006	200,000	35,000	50,000	25,000	49,999	25,000	24,999	25,000	15,000	40,000	45,000	15,000
2007	200,000	35,000	50,000	25,000	49,999	25,000	24,999	25,000	15,000	40,000	45,000	15,000

Source: Current Population Survey Annual Demographic File Technical Documentation (1988–2002), Current Population Survey Annual Social and Economic Supplement Technical Documentation (2003–2007).

Appendix Table 3. (Continued)

Year	Unemployment (UC_VAL)	Workers Comp (WC_VAL)	Veterans (VET_VAL)	Retirement 1st source (RET_VAL1)	Retirement 2nd Source (RET_VAL2)	Survivors 1st Source (SUR_VAL1)	Survivors 2nd Source (SUR_VAL2)	Disability 1st Source (DIS_VAL1)	Disability 2nd Source (DIS_VAL2)	Education Assistance (ED_VAL)	Financial Assistance (FIN_VAL)	Other (OL_VAL)
1988	99,999	99,999	29,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999
1989	99,999	99,999	29,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999
1990	99,999	99,999	29,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999
1991	99,999	99,999	29,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999
1992	99,999	99,999	29,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999
1993	99,999	99,999	29,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999
1994	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999
1995	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999
1996	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999
1997	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999
1998	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999
1999	99,999	99,999	99,999	45,000	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999
2000	99,999	99,999	99,999	45,000	45,000	50,000	50,000	35,000	35,000	20,000	30,000	25,000
2001	99,999	99,999	99,999	45,000	45,000	50,000	50,000	35,000	35,000	20,000	30,000	25,000
2002	99,999	99,999	99,999	45,000	45,000	50,000	50,000	35,000	35,000	20,000	30,000	25,000
2003	99,999	99,999	99,999	45,000	45,000	50,000	50,000	35,000	35,000	20,000	30,000	25,000
2004	99,999	99,999	99,999	45,000	45,000	50,000	50,000	35,000	35,000	20,000	30,000	25,000
2005	99,999	99,999	99,999	45,000	45,000	50,000	50,000	35,000	35,000	20,000	30,000	25,000
2006	99,999	99,999	99,999	45,000	45,000	50,000	50,000	35,000	35,000	20,000	30,000	25,000
2007	99,999	99,999	99,999	45,000	45,000	50,000	50,000	35,000	35,000	20,000	30,000	25,000

Source: Current Population Survey Annual Demographic File Technical Documentation (1988–2002), Current Population Survey Annual Social and Economic Supplement Technical Documentation (2003–2007).

Appendix Table 4. Internal CPS censoring points for each income source in Dollars (1976–1987)

Year	Wages (I51A)	Self Employment (I51B)	Farm Income (I51C)	Social Security (I52A)	Supplemental Security (I52B)	Public Assistance (I53A)	Interest Dividends (I53B)	Rentals and Comp (I53C)	Veterans and Workers (I53D)	Retirement (I53E)	Other (I53F)
1976	99,999	99,999	99,999	9,999	9,999	19,999	99,999	99,999	99,999	99,999	99,999
1977	99,999	99,999	99,999	9,999	9,999	19,999	99,999	99,999	99,999	99,999	99,999
1978	99,999	99,999	99,999	9,999	9,999	19,999	99,999	99,999	99,999	99,999	99,999
1979	99,999	99,999	99,999	9,999	9,999	19,999	99,999	99,999	99,999	99,999	99,999
1980	99,999	99,999	99,999	9,999	9,999	19,999	99,999	99,999	99,999	99,999	99,999
1981	99,999	99,999	99,999	9,999	9,999	19,999	99,999	99,999	99,999	99,999	99,999
1982	99,999	99,999	99,999	9,999	9,999	19,999	99,999	99,999	99,999	99,999	99,999
1983	99,999	99,999	99,999	9,999	9,999	19,999	99,999	99,999	99,999	99,999	99,999
1984	99,999	99,999	99,999	9,999	9,999	19,999	99,999	99,999	99,999	99,999	99,999
1985	99,999	99,999	99,999	9,999	9,999	19,999	99,999	99,999	99,999	99,999	99,999
1986	250,000	250,000	250,000	9,999	9,999	19,999	99,999	99,999	99,999	99,999	99,999
1987	250,000	250,000	250,000	9,999	9,999	19,999	99,999	99,999	99,999	99,999	99,999

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

Appendix Table 5. Internal CPS censoring points for each income source in Dollars (1988–2005)

Year	Primary Earnings (ERN_VAL)	Wages (WS_VAL)	Self Employment (SE_VAL)	Farm (FRML_VAL)	Social Security (SS_VAL)	Supplemental Security (SSI_VAL)	Public Assistance (PAW_VAL)	Interest (INT_VAL)	Dividends (DIV_VAL)	Rental (RNT_VAL)	Alimony (ALML_VAL)	Child Support (CSP_VAL)
1988	299,999	99,999	99,999	99,999	29,999	9,999	19,999	99,999	99,999	99,999	99,999	99,999
1989	299,999	99,999	99,999	99,999	29,999	9,999	19,999	99,999	99,999	99,999	99,999	99,999
1990	299,999	99,999	99,999	99,999	29,999	9,999	19,999	99,999	99,999	99,999	99,999	99,999
1991	299,999	99,999	99,999	99,999	29,999	9,999	19,999	99,999	99,999	99,999	99,999	99,999
1992	299,999	99,999	99,999	99,999	29,999	9,999	19,999	99,999	99,999	99,999	99,999	99,999
1993	299,999	99,999	99,999	99,999	29,999	9,999	19,999	99,999	99,999	99,999	99,999	99,999
1994	999,999	999,999	999,999	999,999	49,999	25,000	24,999	99,999	99,999	99,999	99,999	99,999
1995	1,099,999	1,099,999	999,999	999,999	50,000	25,000	25,000	99,999	99,999	99,999	99,999	99,999
1996	1,099,999	1,099,999	999,999	999,999	50,000	25,000	25,000	99,999	99,999	99,999	99,999	99,999
1997	1,099,999	1,099,999	999,999	999,999	50,000	25,000	25,000	99,999	99,999	99,999	99,999	99,999
1998	1,099,999	1,099,999	999,999	999,999	50,000	25,000	25,000	99,999	99,999	99,999	99,999	99,999
1999	1,099,999	1,099,999	999,999	999,999	50,000	25,000	25,000	99,999	99,999	99,999	99,999	99,999
2000	1,099,999	1,099,999	999,999	999,999	50,000	25,000	25,000	99,999	99,999	99,999	99,999	99,999
2001	1,099,999	1,099,999	999,999	999,999	50,000	25,000	25,000	99,999	99,999	99,999	99,999	99,999
2002	1,099,999	1,099,999	999,999	999,999	50,000	25,000	25,000	99,999	99,999	99,999	99,999	99,999
2003	1,099,999	1,099,999	999,999	999,999	50,000	25,000	25,000	99,999	99,999	99,999	99,999	99,999
2004	1,099,999	1,099,999	999,999	999,999	50,000	25,000	25,000	99,999	99,999	99,999	99,999	99,999
2005	1,099,999	1,099,999	999,999	999,999	50,000	25,000	25,000	99,999	99,999	99,999	99,999	99,999

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

Appendix Table 5. (Continued)

Year	Unemployment (UC_VAL)	Workers Comp (WC_VAL)	Veterans (VET_VAL)	Retirement		Survivors		Disability		Education Assistance (ED_VAL)	Financial Assistance (FIN_VAL)	Other (OL_VAL)
				1st source (RET_VAL1)	2nd Source (RET_VAL2)	1st Source (SUR_VAL1)	2nd Source (SUR_VAL2)	1st Source (DIS_VAL1)	2nd Source (DIS_VAL2)			
1988	99,999	99,999	29,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999
1989	99,999	99,999	29,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999
1990	99,999	99,999	29,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999
1991	99,999	99,999	29,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999
1992	99,999	99,999	29,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999
1993	99,999	99,999	29,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999
1994	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999
1995	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999
1996	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999
1997	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999
1998	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999
1999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999
2000	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999
2001	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999
2002	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999
2003	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999
2004	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999
2005	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999	99,999

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

Appendix Table 6: Number of individuals reporting income exactly at the topcode threshold

Year	Wages (I51A)	Self Employment (I51B)	Farm (I51C)	Social Security (I52A)	Supplemental Security (I52B)	Public Assistance (I53A)	Interest (I53B)	Dividends (I53C)	Veterans and Workers Comp (I53D)	Retirement (I53E)	Other (I53F)
	WAG	SE	FRM	SEC	SS	PA	INT	DIV	OMP	RET	ALC
1976	60	26	6	0	0	0	0	3	0	***	0
1977	102	38	9	0	0	0	***	1	0	***	***
1978	102	58	***	1	0	0	***	6	0	0	***
1979	109	69	9	***	0	0	5	4	0	0	***
1980	186	60	12	***	0	0	6	6	0	1	2
1981	196	62	19	***	0	0	3	4	0	0	***
1982	37	9	0	***	***	0	0	0	***	1	***
1983	83	26	4	***	0	0	***	1	0	1	0
1984	69	21	3	0	0	0	1	1	0	0	0
1985	0	0	0	0	0	0	0	0	0	0	0
1986	1	6	0	0	0	0	0	0	0	0	0
1987	2	2	0	0	0	0	0	0	0	0	0

*** indicates that there are under 5 individuals topcoded from this source of income in the public data, so to protect the confidentiality of respondents we cannot report the number of individuals exactly at the threshold or above the threshold. Source: Author's calculations using internal March CPS data.

Appendix Table 6 (Continued)

	Primary Earnings (ERN_VAL)	Wages (WS_VAL)	Self Employment (SE_VAL)	Farm (FRM_VAL)
1988	0	0	0	0
1989	2	0	0	0
1990	6	0	0	0
1991	2	0	0	0
1992	0	0	0	0
1993	0	0	0	0
1994	4	1	0	***
1995	12	0	2	***

Non-labor income sources are not listed after 1988 as no individual reports income exactly at the topcode threshold for non-labor income sources after this year.

*** indicates that there are under 5 individuals topcoded from this source of income in the public data, so to protect the confidentiality of respondents we cannot report the number of individuals exactly at the threshold or above the threshold.

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