Revised Income Topcodes for the Annual Social and Economic

Survey (ASEC) Public Use Files

**Background**

The Bureau of the Census is restricted by law as specified in Title 13 from:

**“making any publication whereby the data furnished by any particular establishment or individual under this title can be identified”**

In order to meet the requirement imposed by Title 13, to protect the confidentiality of respondents, and to meet the needs of the community of users and analysts who have legitimate and widely varying interests in the data collected in Census surveys, the Census Bureau has developed various protocols for the release of these data, both as summary statistical reports and micro-data contained on “computer files”. These micro-data files may include detailed information for individual sample persons participating in a survey where Title 13 protects the information.

In its long history of releasing public use microdata files based on the March CPS Annual Social and Economic Supplement (ASEC), the Census Bureau has censored the release of “high income” amounts in order to meet the requirements of Title 13. During the period prior to the survey of March 1996 this censorship was applied by limiting the values for income amounts to be no greater than a specified maximum value. Values above the maximum were replaced by the maximum. Maximum values varied by income source and year, and, for the most part, assigned as the amount at the 97th percentile for all nonzero income values. Beginning with the 1996 survey the censorship method was modified so that mean values were substituted for all amounts above the maximum value (the topcode).

**Extending the Mean Value Replacement Method**

Difference in the methods used to censor high-income amounts over time made it difficult to generate consistent time series for important measures of income distribution such as the Gini Coefficient of Income Concentration (GINI), decile shares, etc. In an effort to alleviate this problem and improve the overall usefulness of the ASEC, a team of researchers at Cornell University, led by Jeff Larrimore, applied the post-March 1996 censorship methods to generate new topcode values for the period from March 1976 to March 1995. This work was undertaken with the cooperation of the Census Bureau and funding from the National Science Foundation and the National Institute for Disability Research and Rehabilitation. Their work is documented in a paper entitled “Consistent Cell Means for Topcoded Incomes in the Public Use March CPS (1976-2007)”.

Rank Proximity Swapping Methods for Income Topcoding

While the method that assigns the mean value above the topcode was an important step forward to improving measures of inequality based on the ASEC public use data files, the distribution above the topcode is not preserved. In 2011, the Census Bureau sponsored additional research on methods that both meet Title 13 requirements and preserve the income distribution above the topcode. This research has led to implementation of rank proximity swapping methods that switch income amounts above the topcode for respondents that are of similar income rank. Swapped amounts are rounded following the swapping process to provide additional disclosure avoidance.

The rank proximity swapping method was implemented beginning with the 2011 ASEC. Extract files containing swapped income values for survey years 1975 to 2010 are now available on the FTP site:

http://www.census.gov/housing/extract\_files/.

**Accessing and Using the Revised Income Topcodes**

The new income topcodes are available on blank-delimited text files, one file for each ASEC year; “pu\_swaptopcodes1975ASEC” thru “pu\_swaptopcodes2010ASEC”. All 36 files are included in the file ASEC\_IncomeTopCodes\_SwappingMethod.zip. The names of the income sources differ between those data sets prior to 1988 and those 1988 and later. The layout of the all topcode files is:

**Data Dictionary for New Topcode Data Files**



Each data file contains one line for each income source amount requiring topcoding. If a person has two income amounts that require topcoding then there will be two lines with the same household sequence and person number but differing income source names and amounts.

The new topcodes can be matched to the existing public use data files using household sequence number and person number. The person-number field values follow the numbering convention in use for the later ASEC years. Using this convention, the person numbers within a household begin with 41 (variable” PPPOS” on Census public use documentation). Adjustments may be required to person numbers from the older public use data sets depending on the source.

Income Source Names Specified for March 1975 to March 1987

* Wsal\_val – total wage and salary income
* Semp\_val – total nonfarm self-employment income
* Frse\_val – total farm self-employment income
* I51a\_val – total social security and railroad retirement income
* Ssi\_val - Supplemental Security Income
* Paw\_val – total cash public assistance income
* Int\_val – total interest income
* I53c\_val - total dividends, net rents and royalties, estate and trusts
* I53d\_val - total unemployment benefits, workers’ compensation, and veterans’ payments
* I53e\_val – total private pensions and annuities, military pensions, federal, state, and local pensions
* I53f\_val – total alimony, financial contributions from outside the household, any other money income

Income Source Names Specified for March 1988 to March 2010

* Ern\_val - earnings from primary job last year
* Ws\_val – wage and salary income from all other jobs
* Se\_val- net self-employment income from all other nonfarm businesses
* Frm\_val - net farm self-employment income from all other farm businesses
* Ss\_val – total social security income
* Ssi\_val – total Supplemental Security Income
* Int\_val – total interest income
* Div\_val – total dividend income
* Rnt\_val – total net income from rents and royalties
* Vet\_val – total disability and pension payment s and educational assistance
* csp\_val – child support payments from separated and former spouses
* Alm\_val – alimony payments from former spouse
* Ed\_val – total educational assistance
* Ret\_val1 – retirement income source specified in ret\_sc1
* Ret\_val2 – retirement income source specified in ret\_sc2
* Sur\_val1 – survivor income source specified in sur\_sc1
* Sur\_val2 – survivor income source specified in sur\_sc2
* Dis\_val1 – disability income source specified in dis\_sc1
* Dis\_val2 - disability income source specified in dis\_sc2
* Paw\_val – total cash public assistance
* Uc\_val – total unemployment compensation
* Wc\_val – total worker compensation
* Fin\_val – total financial contributions from persons not living in the household
* Oi\_val – total income not specified elsewhere