CONSUMER EXPENDITURE DIARY SURVEY
PUBLIC USE MICRODATA
2013 Users' Documentation
September 9, 2014

U.S. Department of Labor
Bureau of Labor Statistics
Division of Consumer Expenditure Survey

Table of Contents

I. Introduction

The Consumer Expenditure Survey (CE) program provides a continuous and comprehensive flow of data on the buying habits of American consumers. These data are used widely in economic research and analysis, and in support of revisions of the Consumer Price Index. To meet the needs of users, the Bureau of Labor Statistics (BLS) produces population estimates (for consumer units or CUs) of average expenditures in news releases, reports, and articles. Tabulated CE data are also available on the internet (see Section XV. Appendix 4). The microdata are available on the public BLS website for free download.

These microdata files present detailed expenditure and income data for the Diary component of the CE. They include weekly expenditure (EXPD), annual income (DTBD), and imputed income (DTID) files. The data in EXPD, DTBD, and DTID files are categorized by a Universal Classification Code (UCC). The advantage of the EXPD and DTBD files is that with the data classified in a standardized format, the user may perform comparative expenditure (income) analysis with relative ease. The FMLD and MEMD files contain data on the characteristics and demographics of CUs and CU members. The summary level expenditure and income information on the FMLD files permits the data user to link consumer spending, by general expenditure category, to household characteristics and demographics on one set of files.

Estimates of average expenditures from the Diary survey, integrated with data from the Interview survey, are published online in the CE annual reports.. A number of recent publications containing data from the CE are available on the public website as well.

The microdata files are in the public domain and, with appropriate credit, may be reproduced without permission. A suggested citation is: "U.S. Department of Labor, Bureau of Labor Statistics, Consumer Expenditure Survey, Diary Survey, 2013."

II. Changes from the 2012 Microdata Files

A. FMLD Files

Variable Additions

Beginning in 2013Q1 the following variables will be added to the FMLD files

Variable name	Description	Format
EITC	During the past 12 months, did you claim an Earned Income Tax credit on your federal income tax return?	CHAR(1)
EITC_	EITC Flag	CHAR(1)
FSMPFRMX	Family level summation for new variable SEMPFRMX and SMPFRMBX.	NUM(8)
FSMP_RMX	FSMPFRMX Flag	CHAR(1)
FSMPFRX1-5	imputation iterations - FSMPFRXn	NUM(8)
FSMPFRXI	Indicator/descriptor variable for income imputation.	NUM(3)
FSMPFRXM	Total amount of income received from self-employment income Income Imputation by family grouping, mean of imputation iterations FSMPFRXM	NUM(8)
INTRDVB	Range that best reflects the amount you received in interest or dividends during the past 12 months	CHAR(2)
INTRDVB_	INTRDVB Flag	CHAR(1)
INTRDVBX	Median value of INTRDVB bracket range	NUM(6)
INTR_VBX	INTRDVBX Flag	CHAR(1)

INTRDVX	What was the amount received in interest or dividend during the past 12 months?	NUM(8)
INTRDVX_	INTRDVX Flag	CHAR(1)
INTRDVX1-5	Amount of income received from interest and dividends, Income Imputation iterations	NUM(8)
INTRDVXI	Indicator/descriptor variable for income imputation.	NUM(3)
INTRDVXM	Amount of income received from interest and dividends, mean Income Imputation of the iterations	NUM(8)
NETRENT1-5	Amount of income received from net rental income or loss, Income Imputation iterations	NUM(8)
NETRENTI	Indicator/descriptor variable for income imputation.	NUM(3)
NETRENTM	Amount of income received from net rental income or loss, mean of the iterations NETRNTXM	NUM(8)
NETRENTB	Could you tell me which range that best reflects the total net rental income or loss during the PAST 12 MONTHS?	CHAR(2)
NETR_NTB	NETRENTB Flag	CHAR(1)
NETRENTX	What was the amount in net rental income received or lost in the past 12 months?	NUM(8)
NETR_NTX	NETRENTX Flag	CHAR(1)
NETRNTBX	Median value of NETRENTB bracket range	NUM(6)
NETR_TBX	NETRNTBX Flag	CHAR(1)
OTHREGB	Range that best reflects the total amount received in Veteran's Administration (VA) payments, unemployment compensation, child support, or alimony during the past 12 months	CHAR(2)
OTHREGB_	OTHREGB Flag	CHAR(1)
OTHREGBX	Median value of OTHREGB bracket range	NUM(6)
OTHR_GBX	OTHREGBX Flag	CHAR(1)
OTHREGX	What was the amount from all sources?	NUM(8)
OTHREGX_	OTHREGX Flag	CHAR(1)
OTHREGX1-5	Amount of income received from any other source such as Veterans Administration (VA) payments, unemployment compensation, child support, or alimony, mean of the iterations	NUM(8)
OTHREGXI	Indicator/descriptor variable for income imputation.	NUM(3)
OTHREGXM	Amount of income received from any other source such as Veterans Administration (VA) payments, unemployment compensation, child support, or alimony, mean of the iterations	NUM(8)
RETSRVBX	Median value of RETSURVB bracket range	NUM(6)
RETS_VBX	RETSRVBX Flag	CHAR(1)
RETSURVB	Range that best reflects the total amount received in retirement, survivor, or disability pensions during the past 12 months	CHAR(2)
RETS_RVB	RETSURVB Flag	CHAR(1)
RETSURVX	What was the amount received for retirement, survivor, or disability pensions?	NUM(8)
RETS_RVX	RETSURVX Flag	CHAR(1)
RETSURV1-5	Amount of income received from retirement, survivor, or disability pensions, imputation iterations	NUM(8)
RETSURVI	Indicator/descriptor variable for income imputation	NUM(3)
RETSURVM	Amount of income received from retirement, survivor, or disability pensions, mean of the iterations	NUM(8)

ROYESTB	Range that best reflects the total amount received in royalty income or income from estates and trusts during the past 12 months	CHAR(2)
ROYESTB_	ROYESTB Flag	CHAR(1)
ROYESTBX	Median value of ROYESTB bracket range	NUM(6)
ROYE_TBX	ROYESTBX Flag	CHAR(1)
ROYESTX	What was the amount received from royal income or income from estates and trusts?	NUM(8)
ROYESTX_	ROYESTX Flag	CHAR(1)
ROYESTX1-5	Amount of income received from royalty income or income from estates and trusts, imputation iterations	NUM(8)
ROYESTXI	Indicator/descriptor variable for income imputation.	NUM(3)
ROYESTXM	Amount of income received from royalty income or income from estates and trusts, mean of the iterations	NUM(8)

Variable Deletions

Beginning in 2013Q1 the following variables will no longer be available in the FMLD files

Variable Name	Description
ALIOTHB	Could you tell me which range best reflects the total amount received in alimony and other sources during the last 12 months?
ALIOTHB_	ALIOTHB Flag
ALIOTHBX	Median of ALIOTHB bracket range
ALIO_HBX	ALIOTHBX Flag
ALIOTHX	During the past 12 months, what was the total amount of income from regular contributions from alimony and other sources, such as from persons outside the CU received by ALL CU members?
ALIOTHX_	ALIOTHX Flag
ALIOTHX1-5	Imputation iteration 1-5
ALIOTHXI	Indicator/descriptor variable for income imputation.
ALIOTHXM	During the past 12 months, what was the total amount of income from regular contributions from alimony and other sources such as from persons outside the CU received by ALL CU members?
CHDLMPB	Could you tell me which range best reflects the total amount received in lump sum payments for child support during the last 12 months?
CHDLMPB_	CHDLMPB Flag
CHDLMPBX	Median of CHDLMPB bracket range
CHDL_PBX	CHDLMPBX Flag
CHDLMPX	During the past 12 months, what was the total amount received from a onetime lump sum payment for child support by ALL CU members?
CHDLMPX_	CHDLMPX Flag
CHDOTHB	Could you tell me which best reflects the total amount received in child support payments, other than lump sum amounts, by all CU members during the last 12 months?
CHDOTHB_	CHDOTHB Flag
CHDOTHBX	Median of CHDOTHB bracket range
CHDO_HBX	CHDOTHBX Flag
CHDOTHX	During the past 12 months, what was the total amount of income from child support payments in other than a lump sum amount received by ALL CU members?
CHDOTHX_	CHDOTHX Flag

CHDOTHX1-5	Imputation Iterations - CHDOTHX
CHDOTHXI	Indicator/descriptor variable for income imputation.
CHDOTHXM	During the past 12 months, what was the total amount of income from child support payments in other than a lump sum amount received by ALL CU members?
DIVB	Could you tell me which range best reflects the total amount of income from dividends, trusts, estates or royalties during the last 12 months?
DIVB_	DIVB Flag
DIVBX	Median of DIVB bracket range
DIVBX_	DIVBX Flag
DIVX	During the past 12 months, what was the total amount of income from dividends, royalties, estates, or trusts received by ALL CU members?
DIVX_	DIVX Flag
DIVX1-5	Imputation iteration for DIVX
DIVXI	Indicator/descriptor variable for income imputation.
DIVXM	During the past 12 months, what was the total amount of income from dividends, royalties, estates, or trusts received by ALL CU members?
FBSNSX	Amount of income or loss from nonfarm business, partnership or professional practice received by all CU members in past 12 months
FBSNSX1-5	Imputation iteration for FBNSX
FBSNSXI	Indicator/descriptor variable for income imputation.
FBSNSXM	Amount of income or loss from nonfarm business, partnership or professional practice received by all CU members in past 12 months
FD_STMPS	Have any members of your CU received any Food Stamps, in the past month?
FD_S_MPS	FD_STMPS Flag
FFARMX	Amount of income or loss from own farm received by all CU members in past 12 months
FFARMX_	FFARMX Flag
FFARMX1-5	Imputation iteration for FFARMX
FFARMXI	Indicator/descriptor variable for income imputation.
FFARMXM	Amount of income or loss from own farm received by all CU members in past 12 months
INSREFX	During the past 12 months, what was the total amount of refund received from insurance policies by ALL CU members?
INSREFX_	INSREFX Flag
INTB	Could you tell me which range best reflects the total amount of interest received by all CU members during the last 12 months?
INTB_	INTB Flag
INTBX	Median of INTB bracket range
INTBX_	INTBX Flag
INTX	During the past 12 months, what was the total amount of income from interest on savings accounts or bonds received by ALL CU members?
INTX_	INTX Flag
INTX1-5	Imputation iteration for INTX
INTXI	Indicator/descriptor variable for income imputation.
INTXM	During the past 12 months, what was the total amount of income from interest on savings accounts or bonds received by ALL CU members?
OTHLOSBX	Median of OTHLOSSB bracket range
OTHL_SBX	OTHLOSBX Flag

OTHLOSSB	Could you tell me which range best reflects your net income or loss from other rental units during the last 12 months?
OTHL_SSB	OTHLOSSB Flag
OTHREFX	During the past 12 months, what was the total amount of refund received from other sources, including any other taxes, by ALL CU members?
OTHREFX_	OTHREFX Flag
OTHRNTX	During the past 12 months, how much net income or loss was received from payments from other rental units?
OTHRNTX_	OTHRNTX Flag
OTHRNTX1-5	Imputation iteration for OTHRNTX
OTHRNTXI	Indicator/descriptor variable for income imputation.
OTHRNTXM	During the past 12 months, how much net income or loss was received from payments from other rental units?
PENSION1-5	Imputation iteration for PENSIONX
PENSIONI	Indicator/descriptor variable for income imputation.
PENSIONM	During the past 12 months, what was the total amount of income from pensions or annuities from private companies, military, Government, IRA, or Keogh received by ALL CU members?
PENSIONX	During the past 12 months, what was the total amount of income from pensions or annuities from private companies, military, Government, IRA, or Keogh received by ALL CU members?
PENS_ONX	PENSIONX Flag
PNSIONB	Could you tell me which range best reflects the total amount of retirement pensions and annuities during the last 12 months?
PNSIONB_	PNSIONB Flag
PNSIONBX	Median of PNSIONB bracket range
PNSI_NBX	PNSIONBX Flag
PTAXREFX	During the past 12 months, what was the total amount of refund received from property taxes by ALL CU members?
PTAX_EFX	PTAXREFX Flag
ROOMLOSB	Could you tell me which range best reflects your net income or loss from roomers or boarders?
ROOM_OSB	ROOMLOSB Flag
ROOMLSBX	Median of ROOMLOSB bracket range
ROOM_SBX	ROOMLSBX Flag
ROOMX	During the past 12 months, how much net income or loss was received from roomers or boarders?
ROOMX_	ROOMX Flag
ROOMX1-5	Imputation iteration for ROOMX
ROOMXI	Indicator/descriptor variable for income imputation.
ROOMXM	During the past 12 months, how much net income or loss was received from roomers or boarders?
SALEB	Could you tell me which range best reflects the total amount received from these sales during the last 12 months?
SALEB_	SALEB Flag
SALEBX	Median of SALEB bracket range
SALEBX_	SALEBX Flag

SALEX	During the past 12 months, what was the total amount received from the sale of household furnishings, equipment, clothing, jewelry, pets or other belongings, excluding the sale of vehicles or property by ALL CU members?
SALEX_	SALEX Flag
SSREFX	During the past 12 months, what was the total amount of refund received from overpayment on Social Security by ALL CU members?
SSREFX_	SSREFX Flag
TAXPROPX	During the past 12 months, what was the total amount PAID for personal property taxes for vehicles not reported elsewhere by ALL CU members?
TAXP_OPX	TAXPROPX Flag
UNEMPB	Could you tell me which range best reflects the total amount received in unemployment compensation during the last 12 months?
UNEMPB_	UNEMPB Flag
UNEMPBX	Median of UNEMPB bracket range
UNEMPBX_	UNEMPBX Flag
UNEMPX	During the past 12 months, what was the total amount of income from unemployment compensation received by ALL CU members?
UNEMPX_	UNEMPX Flag
UNEMPX1-5	Imputation iteration of UNEMPX
UNEMPXI	Indicator/descriptor variable for income imputation.
UNEMPXM	During the past 12 months, what was the total amount of income from unemployment compensation received by ALL CU members?
WRKRSB	Could you tell me which range best reflects the total amount of income from worker's compensation during the last 12 months?
WRKRSB_	WRKRSB Flag
WRKRSBX	Median of WRKRSBX bracket range
WRKRSBX_	WRKRSBX Flag
WRKRSX	During the past 12 months, what was the total amount of income from workers' compensation or veterans' benefits, including education benefits, but excluding military retirement, received by ALL CU members?
WRKRSX_	WRKRSX Flag
WRKRSX1-5	Imputation iteration of WRKRSX
WRKRSXI	Indicator/descriptor variable for income imputation.
WRKRSXM	During the past 12 months, what was the total amount of income from workers' compensation or veterans' benefits, including education benefits, but excluding military retirement, received by ALL CU members?

Variable Changes

Beginning in 2013Q1 the following variables' definitions will be changed in the FMLD files

Variable name	Description
LUMPX	Now includes expenditures on Child Supports Lump Sum payments
LUMPBX	Now includes expenditures on Child Supports Lump Sum payments
EDUC_REF	Some codes have been collapsed Old codes: 16 Master's degree 17 Professional/doctorate degree New codes: 16 Master's, professional or doctorate degree

B. MEMD Files

Variable Additions

Beginning in 2013Q1 the following variables will be added to the MEMD files

Variable name	Description	Format
PAYSTUB	Does the respondent have a paper or electronic pay check record present for their last paycheck?	CHAR(2)
PAYSTUB_	PAYSTUB Flag	CHAR(1)
SEMPFRM	During the past 12 months did you receive any self-employment income or have a loss?	NUM(8)
SEMPFRM_	SEMPFRM Flag	CHAR(1)
SEMPFRMX	What was the amount of self-employment income or loss?	NUM(8)
SEMP_RMX	SEMPFRMX Flag	CHAR(1)
SMPFRMB	Range that best reflects the income or loss from self-employment during the past 12 months	CHAR(2)
SMPFRMB_	SMPFRMB Flag	CHAR(1)
SMPFRMBX	Median value of SMPFRMB bracket range	NUM(6)
SMPF_MBX	SMPFRMBX Flag	CHAR(1)
SEMPFRM1-5	Amount of income received from self-employment, mean of the Income Imputation iterations	NUM(8)
SEMPFRMI	Indicator/descriptor variable for income imputation.	NUM(3)
SEMPFRMM	Amount of income received from self-employment, mean of the Income Imputation iterations	NUM(8)
SOCSRRET	During the past 12 months – Did you name receive any social security or railroad retirement benefits?	CHAR(1)
SOCS_RET	SOCSRRET Flag	CHAR(1)
WKSTATUS	This variable replaces PWRKSTAT because of changes to data collection.	NUM(8)

Variable Deletions

Beginning in 2013Q1 the following variables will no longer be available in the MEMD files

Variable name	Description
ANYRAIL	During the past 12 months, did the member receive from the U.S. Government any money from Railroad Retirement checks?
ANYRAIL_	ANYRAIL Flag
ANYSSINC	During the past 12 months, did the member receive from the U.S. Government any money from Social Security checks?
ANYS_INC	ANYSSINC Flag
BSNSB	Could you tell me which range best reflects the member's income or loss from the member's own nonfarm business, partnership or professional practice during the last 12 months?
BSNSB_	BSNSB Flag
BSNSBX	Median of BSNSB bracket range
BSNSBX_	BSNSBX Flag
BSNSX	Amount of income or loss received from nonfarm business
BSNSX_	BSNSX Flag
BSNSX1-5	Imputation iteration of BSNSX
BSNSXI	Indicator/descriptor variable for income imputation.
BSNSXM	Amount of income or loss from nonfarm business, partnership or professional practice received in past 12 months
FARMB	Could you tell me which range best reflects the member's income or loss from the member's own farm during the last 12 months?
FARMB_	FARMB Flag
FARMBX	Median of FARMB bracket range
FARMBX_	FARMBX Flag
FARMX	Amount of income or loss received from own farm
FARMX_	FARMX Flag
FARMX1-5	Imputation iteration of FARMX
FARMXI	Indicator/descriptor variable for income imputation.
FARMXM	During the past 12 months, what was the amount of income or loss from the member's own farm after expenses?
STA_SUPP	During the past 12 months, did the member receive any Supplemental Security Income checks from the State or local government?
STA__UPP	STA_SUPP Flag

Variable Changes

Beginning in 2013Q1 the following variables' definitions will be changed in the MEMD files

Variable name	Description
EDUCA	Old codes: 0. Never attended, preschool, kindergarten 1. 1st grade 2. 2nd grade 3. 3rd grade 4. 4th grade 5. 5th grade 6. 6th grade

	7. 7th grade 8. 8th grade 9. 9th grade 10. 10th grade 11. 11th grade 38. 12th grade NO DIPLOMA 39. HIGH SCHOOL GRADUATE – high school DIPLOMA, or the equivalent 40. Some college but no degree 41. Associate degree in college – Occupational program 42. Associate degree in college – Academic program 43. Bachelor's degree (Example: BA, AB, BS) 44. Master's degree (Example: MA, MS, MEng, MSW, MBA) 45. Professional School Degree (Example: MD, DDS, DVM, LLB, JD) 46. Doctorate degree (Example: PhD, EdD) <u>New codes</u>: 1. No schooling completed, or less than 1 year 2. Nursery, kindergarten, and elementary (grades 1-8) 3. High school (grades 9-12, no degree) 4. High school graduate – high school diploma or the equivalent (GED) 5. Some college but no degree 6. Associate's degree in college 7. Bachelor's degree (BA, AB, BS, etc.) 8. Master's professional, or doctorate degree (MA, MS, MBA, MD, JD, PhD, etc.)
OCCULIST	Some codes have been combined, others have been recoded <u>Old codes:</u> 11 Machine Operator, assembler, inspector 12 Transportation operator [combined into 11] 13 Handler, helper, laborer [combined into 11] 14 Mechanic, repairer, precision production [recoded to 12] 15 Construction, mining [combined into 12] 16 Farming [recoded] 17 Forestry, fishing, grounds-keeping [recoded] 18 Armed Forces [recoded] <u>New codes:</u> 11 Machine or transportation operator, laborer 12 Construction workers, mechanics 13 Farming 14 Forestry, fishing, grounds-keeping 15 Armed Forces

C. EXPD Files

UCC Additions

Beginning in 2013Q1 the following UCCs will be added to the EXPD files

UCC	Description
280140	KITCHEN/DINING ROOM/OTHR LINENS
300218	WASHERS AND DRYERS
310316	RADIOS/SPEAKERS/SOUND COMP SYSTMS
320221	LAMPS/LIGHT FIXTURES/CEILING FANS
320345	DISHES/CUPS/GLASSES/SERVING PIECS
320627	FLOORING INSTALL/REPAIR/REPLACE
360420	MENS SWEATERS/SHIRTS/VESTS
370125	BOYS SWEATERS/SHIRTS/VESTS
380315	WOMENS SWEATERS/SHIRTS/TOPS
420115	SEWING/NDLWRK/QUILT MATRLS/ITEMS
450350	CAR/TRUCK LEASE PAYMENTS
490300	VEHICLE OR ENGINE REPAIRS
520516	AUTO/TRUCK RENTAL
620114	AUTO SERVICE CLUBS/GPS SERVICES
620213	TKTS TO PLAY/THEATR/OPERA/CONCERT
620214	TKTS TO MOVIE, PARK, MUSEUMS
950024	VEHICLE PERSONAL PROPERTY TAXES

UCC Deletions

Beginning in 2013Q1 the following UCCs will no longer be available in the EXPD files

UCC	Description
230120	INSTALLED HARD SURFACE FLOORIN
230130	INSTALLED WALL-TO-WALL CARPET
280130	KITCHEN AND DINING ROOM LINENS
280230	SEWING MATERIALS
280900	OTHER LINENS
300210	WASHERS
300220	DRYERS
310311	RADIOS
310313	TAPE RECORDERS AND PLAYERS
310320	COMPONENTS/COMPONENT SYSTEMS
320620	PERM HARD SURFACE FLR COVERING
320220	LAMPS AND LIGHTING FIXTURES
320310	PLASTIC DINNERWARE
320320	CHINA AND OTHER DINNERWARE
320340	GLASSWARE
320350	SILVER SERVING PIECES
320360	OTHER SERVING PIECES
360340	MENS SWEATERS AND VESTS
360410	MENS SHIRTS
370120	BOYS SWEATERS
370130	BOYS SHIRTS

380312	WOMENS VESTS AND SWEATERS
380313	WOMENS SHIRTS, TOPS,BLOUSES
420110	MATERIAL FOR MAKING CLOTHES
420120	SEWING NOTIONS, PATTERNS
450310	CAR LEASE PAYMENTS
450410	TRUCK LEASE PAYMENTS
490211	CLUTCH, TRANSMISSION REPAIR
490212	DRIVE SHAFT AND REAR-END REPAIR
490220	BRAKE WORK (DIARY)
490231	REPAIR TO STEERING OR FRONT END
490232	REPAIR TO ENGINE COOLING SYSTEM
490319	VEHICLE AIR CONDITION REPAIR
490411	EXHAUST SYSTEM REPAIR
490412	ELECTRICAL SYSTEM REPAIR
490413	MOTOR REPAIR/REPLACEMENT
520560	GLOBAL POSITIONING SERVICES
620113	AUTOMOBILE SERVICE CLUBS
620211	MOVIE, THEATER, OPERA, BALLET

UCC Changes

Beginning in 2013Q1 the following UCC's definition will be changed in the EXPD files

UCC	Description
340210	Now includes expenditures on Childcare

D. DTBD Files

No changes in 2013

E. DTID Files

No changes in 2013

III. File Information

The microdata are provided as SAS, STATA, SPSS data sets or ASCII comma-delimited files. The 2013 Diary release contains five sets of data files (FMLD, MEMD, EXPD, DTBD, DTID) and three processing files. The FMLD, MEMD, EXPD, DTBD, and DTID files are organized by the quarter of the calendar year in which the data were collected. There are four quarterly data sets for each of these files. The FMLD files contain CU characteristics, income, and summary level expenditures; the MEMD files contain member characteristics and income data; the EXPD files contain detailed weekly expenditures at the UCC level; the DTBD files contain the CUs' reported income values or the mean of the five imputed income values in the multiple imputation method; and the DTID files contain the five imputed income values.

The three processing files enhance computer processing and tabulation of data, and provide descriptive information on item codes. The three processing files are: an aggregation scheme file used in the published consumer expenditure tables (DSTUB), a UCC file that contains UCCs and their abbreviated

titles, identifying the expenditure, income, or demographic item represented by each UCC, and a sample program file that verifies CE estimates (see Section VII.). The processing files are further explained in Section III.F.6. Processing Files.

In addition to these processing files, there is a "User's Guide to Income Imputation in the CE," which includes information on how to appropriately use the imputed income data.

Note that the variable NEWID, the CUs' identification number, is the common variable among files by which matching is done. Values for NEWID have a leading "blank." Because of this, it appears the NEWID values are only 7 characters long, when actually they are 8.

A. Dataset Names

The file naming convention is listed in the table below.

\DIARY13\FMLD131.* (Diary FMLD file for first quarter, 2013)
\DIARY13\MEMD131.* (Diary MEMD file for first quarter, 2013)
\DIARY13\EXPD131.* (Diary EXPD file for first quarter, 2013)
\DIARY13\DTBD131.* (Diary DTBD file for first quarter, 2013)
\DIARY13\DTID131.* (Diary DTID file for first quarter, 2013)
\DIARY13\FMLD132.*
\DIARY13\MEMD132.*
\DIARY13\EXPD132.*
\DIARY13\DTBD132.*
\DIARY13\DTID132.*
\DIARY13\FMLD133.*
\DIARY13\MEMD133.*
\DIARY13\EXPD133.*
\DIARY13\DTBD133.*
\DIARY13\DTID133.*
\DIARY13\FMLD134.*
\DIARY13\MEMD134.*
\DIARY13\EXPD134.*
\DIARY13\DTBD134.*
\DIARY13\DTID134.*
\DIARY13\UCCD13.txt

Note: All data files are compressed. These files can be uncompressed using most unzip utilities.

B. Record Counts

The following are number of records in each data set.

Data Set	2013 Record Count
FMLD131	3,251
MEMD131	7,841
EXPD131	113,891
DTBD131	50,569
DTID131	66,495
FMLD132	3,324

Data Set	2013 Record Count
MEMD132	8,117
EXPD132	116,043
DTBD132	52,176
DTID132	67,190
FMLD133	3,036
MEMD133	7,471
EXPD133	107,498
DTBD133	47,265
DTID133	61,080
FMLD134	2,724
MEMD134	6,599
EXPD134	94,732
DTBD134	42,463
DTID134	54,490

C. Data Flags

Data fields on the FMLD and MEMD files are explained by flag variables following the data field. The names of the flag variables are derived from the names of the data fields they reference. In general the rule is to add an underscore to the last position of the data field name, for example WAGEX becomes WAGEX_. However, if the data field name is eight characters in length, then the fifth position is replaced with an underscore. If this fifth position is already an underscore, then the fifth position is changed to a zero, so that PENSIONX becomes PENS_ONX, EDUC_REF becomes EDUC0REF.

The flag values are defined as follows:

A flag value of "A" indicates a valid blank; that is, a blank field where a response is not anticipated.

A flag value of "B" indicates a blank resulting from an invalid nonresponse; that is, a nonresponse that is not consistent with other data reported by the CU.

A flag value of "C" refers to a blank resulting from a "don't know", refusal, or other type of nonresponse.

A flag value of "D" indicates that the data field contains a valid or good data value.

A flag value of "T" indicates topcoding has been applied to the data field.

Some Primary Sampling Units (PSUs) in some states are given "false" STATE codes for nondisclosure reasons. See Section IV.A.CU Characteristics and Income File (FMLD) on topcoding of CU characteristics and income for more detail.

D. Income Imputation

Beginning in 2004, the CE has implemented multiple imputation of income data. Imputation allows income values to be estimated when they are not reported. Many income variables and other income related variables will be imputed using a multiple imputation process. These imputed income values will

be included in the FMLD, MEMD, DTBD, and DTID files. The multiple imputation process derives five imputation values and a mean imputation value per income variable. More information on the imputation process and how to appropriately use the data are found in the document "User's guide to Income Imputation in the CE."

In the public-use microdata, not all of the imputed income variables will contain the derived imputation values. For some income variables, the five derived imputations are excluded and only the mean of those imputations is available. For these variables, there are 3 associated income variables in the FMLD and MEMD files (*INCOME*M, *INCOME*M_, and *INCOME*I). For all other imputed income variables, there are 7 associated variables in the FMLD and MEMD files:

*INCOME*1	the first imputed income value or the reported income value, if non-missing
*INCOME*2	the second imputed income value or the reported income value, if non-missing
*INCOME*3	the third imputed income value or the reported income value, if non-missing
*INCOME*4	the fourth imputed income value or the reported income value, if non-missing
*INCOME*5	the fifth imputed income value or the reported income value, if non-missing
*INCOME*M	the mean of the five imputed income values
*INCOME*M_	the flag variable for the imputed variable (see Section III.C. Data Flags)
*INCOME*I	the imputation indicator

Income variables that have imputed values as components (ex: FINCBEFM) will also have 5 imputed values and a mean based on each of the imputed components.

The imputation indicator variable is a 3 digit number that is coded as follows:

The first digit in the 3 digit code defines the imputation method. The meanings are:
1: No Imputation
2: Multiple imputation due to invalid blank only
3: Multiple imputation due to bracketing only
4: Multiple imputation due to invalid blanks and bracketing
5: Multiple imputation due to conversion of a valid blank to an invalid blank (this occurs only when initial values for all sources of income for the CU were valid blanks).

The meaning of the last two digits of the three digit code differs depending on whether you are looking at one of the components of overall income, like FWAGEXM, or you are looking at the summary level variable FINCBEFM. For the components, the last 2 digits represent the number of family members who had their data imputed for that source. For example, if a family had a value of 302 for FWAGEXI that would mean that 2 of the members in the family had their salary income imputed and that in both cases the imputation was due to bracketing only. For the summary level variable FINCBEFM which is a summation of all of the income components, the last 2 digits represent the number of income sources imputed for each member added together. For example, if a family had 3 members and 2 had salary income imputed due to invalid blank only, and 2 had self-employment income imputed due to bracketing only, and that was the only income data imputed for members of that family, then FSMPFRXI for the family would be 202, FBSNSXI would be 302, and FINCBEFI would be 404.

The DTBD file includes income UCCs mapped from the associated *INCOME*M variables and the income variables that are not imputed in the FMLD files. The DTID file includes UCCs mapped from income variables subject to income imputation, including the variable IMPNUM to indicate the imputation number 1 - 5.

E. File Notation

Every record from each data file includes the variable NEWID, the CU's unique identification number, which can be used to link records of one CU from several files. Data fields for variables on the microdata files have either numeric or character values. The format column in the diary data dictionary distinguishes

whether a variable is numeric (NUM) or character (CHAR) and shows the number of field positions the variable occupies. Variables that include decimal points are formatted as NUM(t,r) where t is the total number of positions occupied, and r is the number of places to the right of the decimal.

In addition to format, the diary data dictionary gives an item description, questionnaire source,and identification of codes where applicable for each variable.

An asterisk (*) is shown in front of new variables, those which have changed in format or definition, and those which have been deleted.

Some variables require special notation. The following notation is used throughout the documentation for all files:

*D(Yxxq) identifies a variable that is deleted as of the quarterly file indicated. The year and quarter are identified by the 'xx' and 'q' respectively. For example, the notation *D(Y131) indicates the variable is deleted starting with the data file of the first quarter of 2013.

*N(Yxxq) identifies a variable that is added as of the quarterly file indicated. The year and quarter are identified by the 'xx' and 'q' for new variables in the same way as for deleted variables.

*C(Yxxq) identifies a variable whose description has been changed. The year and quarter are identified by the 'xx' and 'q' for new variables in the same way as for new and deleted variables.

*L indicates that the variable can contain negative values.

F. Notes on Files

1. Consumer Unit (CU) Characteristics and Income File (FMLD)

The FMLD file, also referred to as the "Consumer Unit Characteristics and Income" file, contains CU characteristics, CU income, and characteristics and earnings of the reference person and of the spouse. The file includes weights needed to calculate population estimates and variances (see Sections V. Estimation Procedures and VI. Reliability Statement).

Summary expenditure variables in this file can be combined to derive weekly estimates for broad consumption categories. Demographic characteristics, such as family size, refer to the CU status on the date of the interview. Income variables contain annual values, covering the 12 months prior to the date of the interview. When there is a valid nonresponse, or where nonresponse occurs and there is no imputation, there will be missing values. The type of nonresponse is explained by associated data flag variables described in Section III.C. Data Flags.

Summary Expenditure Data

Some variables in the FMLD file contain summary expenditure data. They are all BLS derived. The UCCs comprising each summary expenditure variable are listed below the variable description. UCCs may not be represented in all Diary quarters. When UCCs are added to or deleted from the summary variable definition, the quarter in which the addition (deletion) to the summary expenditure variable occurs is denoted by a leading character directly after the UCC code in the "Changes to the 2012 Microdata" section. For example, N131<UCC> or D131<UCC> identifies a new or deleted UCC for a given summary expenditure variable beginning in Q131.

2. Member Characteristics and Income File (MEMD)

The "MEMD" file, also referred to as the "Member Characteristics and Income" file, contains selected characteristics for each CU member, including identification of relationship to reference person. Characteristics for the reference person and spouse appear on both the MEMD file and FMLD file. Demographic characteristic data, such as age of CU member, refer to the member status at the placement of each diary. Income data are collected for all CU members over 13 years of age. Income taxes withheld and pension and retirement contributions are shown both annually and as deductions from the member's last paycheck. Income variables contain annual values for the 12 months prior to the interview month. When there is a valid nonresponse, or where nonresponse occurs and there is no imputation, there will be missing values. The type of nonresponse is explained by associated data flag variables described in Section III C. Data Flags.

3. Detailed Expenditures File (EXPD)

In the "EXPD" file, each expenditure recorded by a CU in a weekly diary is identified by UCC, gift/nongift status, and day on which the expenditure occurred. UCCs are six digit codes that identify items or groups of items. (See the UCCd13.txt file within the DIARY13 folder for a listing of UCCs.) There may be more than one record for a UCC on a single day if that is what was reported in the diary. There are no missing values in this file. If no expenditure was recorded for the item(s) represented by a UCC, then there is no record for the UCC on file.

4. Income File (DTBD)

The "DTBD" file, also referred to as the "Income" file, contains CU characteristic and income data. This file is created directly from the FMLD file and contains the same annual and point-of-placement data. It was created to facilitate computer processing when linking CU income and demographic characteristic data with EXPD expenditure data. As such, the file structure is similar to EXPD. Each characteristic and income item is identified by UCC. (See the UCCd13.txt file within the DIARY13 folder for a listing of UCCs.) There are no records with missing values in DTBD. If the corresponding FMLD file variable contained a missing value, there is no record for the UCC.

5. Imputed Income File (DTID)

As a result of the introduction of multiply imputed income data in the Consumer Expenditure Survey, the Imputed DTID file is now on the Microdata. It is very similar to the DTBD file, except that the variable “IMPNUM” will indicate the number (1-5) of the imputation variant of the income variable and it only contains UCCs from variables subject to income imputation.

6. Processing Files

Dstub File

X:\FILEPATH\diary13\\Dstub2013.txt

The Dstub file shows the aggregation scheme used in the published consumer expenditure tables. It is formatted as follows:

DESCRIPTION	FORMAT
Type: represents whether information in this line contains aggregation data or not	CHAR(1)

Level: aggregation level (lowest number is highest level of aggregation)	CHAR(1)
Title: title of the line item	CHAR(60)
UCC: UCC number in the EXPD or DTBD file	CHAR(6)
Survey: Indicates survey source (D = Diary, G = Aggregated item)	CHAR(1)
Group: Indicates if the item is an expenditure, income, or asset	CHAR(7)

UCC File

X:\FILEPATH\diary13\UCCD13.TXT

The UCC file contains UCCs and their abbreviated titles, identifying the expenditure, income, or demographic item represented by each UCC. It is formatted as follows:

DESCRIPTION	FORMAT
UCC	CHAR(6)
UCC title	CHAR(50)

IV. Topcoding and Other Nondisclosure Requirements

Sensitive CU data are changed so that users will not be able to identify CUs who participated in the survey. Topcoding refers to the replacement of data in cases where the value of the original data exceeds prescribed critical values. Critical values for each variable containing sensitive data are calculated in accordance with Census Disclosure Review Board guidelines. Each observation that falls outside the critical value is replaced with a topcoded value that represents the mean of the subset of all outlying observations. All four quarters of data in the CE microdata release are used when calculating the critical value and topcode amount. If an observation is topcoded, the flag variable assigned to that observation is set to 'T.'

Since the critical value and the mean of the set of values outside the critical value may differ with each annual (four-quarter) release, the topcode values may change annually and be applied at a different starting point. By topcoding values in this manner, the first moment will be preserved for each four-quarter data release when using the total sample. This, however, will not be the case when means are estimated by characteristic, because topcode values are not calculated by characteristic.

A. CU Characteristics and Income File (FMLD)

The following table lists FMLD file variables that are subject to topcoding as well as their associated critical values and topcode values. For multiply imputed income variables, it is possible for an upper topcode value to be less than the upper critical value or for a lower topcode value to be greater than the lower critical value.

Variable	Description	2013 Upper Critical Value	2013 Lower Critical Value	2013 Upper Topcode Value	2013 Lower Topcode Value
ADDFEDX	Amount of federal tax paid by any CU member in addition to that withheld from earnings in past 12 months	32,000	-	86,300	-
ADDOTHX	Amount of other taxes paid which were not reported elsewhere during past 12 months	6,500	-	23,133.33	-

Variable	Description	2013 Upper Critical Value	2013 Lower Critical Value	2013 Upper Topcode Value	2013 Lower Topcode Value
ADDSTAX	Amount of state and local tax paid in addition to that withheld from earnings during past 12 months	8,000	-	22,930.43	-
FEDREFX	Amount of refund received from federal income tax	9,000	-	37,834.03	-
INTRDVX	Amount of income received from interest and dividends, imputation iterations	23,000	-	45,026.67	-
INTRDVXM	Amount of income received from interest and dividends, imputation iterations	23,000	-	29,548.58	-
LUMPX	Lump sum payments from insurance, estates, trusts, royalties, child support, alimony, prizes or games of chance, or from people who are not on your list	160,000	-	500,000	-
NETRENTM	Amount of income received from net rental income or loss	75,000	-20,000	87,980.26	-23,832.54
NETRENTX	Amount of income received from net rental income or loss	75,000	-20,000	134,093.20	-78,000
OCCEXPNX	Amount of payment by CU for occupational expenses such as union dues, tools, uniforms, business or professional association dues, licenses or permits in past 12 months	6,041	-	44,562.07	-
OTHINX	Amount received in other money income including money received from care of foster children, cash scholarships and fellowships, or stipends not based on working	37,000	-	96,307.69	-
OTHINXM	Amount received in other money income including money received from care of foster children, cash scholarships and fellowships, or stipends not based on working	37,000	-	69,090.26	-
OTHREGX	Amount of income received from any other source such as Veteran's Administration (VA) payments, unemployment compensation, child support, or alimony, imputation iterations	26,000	-	37,237.40	-
OTHREGXM	Amount of income received from any other source such as Veteran's Administration (VA) payments, unemployment compensation, child support, or alimony, imputation iterations	26,000	-	25,522.34	-

Variable	Description	2013 Upper Critical Value	2013 Lower Critical Value	2013 Upper Topcode Value	2013 Lower Topcode Value
RETSURVM	Amount of income received from retirement, survivor, or disability pensions	72,000	-	63,526.93	-
RETSURVX	Amount of income received from retirement, survivor, or disability pensions	72,000	-	86,956.46	-
ROYESTX	Amount of income received from royalty income or income from estates and trusts	160,000	-	312,000	-
ROYESTXM	Amount of income received from royalty income or income from estates and trusts	160,000	-	282,052.53	-
STATREFX	Amount of refund received from state and local income tax	2,331	-	46,33.18	-

Some income variables that are subject to topcoding are constructed by summing up the values of "lower level" MEMD or FMLD file component variables. These variables are not topcoded by the conventional method of replacement with a topcode value. Instead the variables' components are summed normally and the variables are flagged as topcoded if one of their component variables is topcoded.

Following are the income variables that are calculated using values of their component variables. (See the descriptions of each variable in the diary data dictionary for a list of component variables.)

FSMPFRXM FSMPFRX1-5 FSMPFRMX	Amount of income received from self-employment income
FFEDTXX	Amount of Federal tax deducted from last pay, annualized for all CU members
FGVXM, FGVX1-5 FGVX	Amount of government retirement deducted from last pay, annualized for all CU members
FINCAFTM FINCAFT1-5 FINCAFTX	Amount of CU income after taxes
FINCBEFM FINCBEF1-5 FINCBEFX	Amount of CU income before taxes
FIRAX	Amount of money placed in individual retirement plan
FJSSDEDM FJSSDED1-5 FJSSDEDX	Estimated amount of annual Social Security contribution
FPVTXM FPVTX	Amount of private pension fund deducted from last pay, annualized for all CU members
FRRXM FRRX	Amount of Railroad Retirement deducted from last pay, annualized for all CU members
FSTATXXM FSTATXX1-5 FSTATXX	Amount of State and local income taxes deducted from last pay, annualized for all CU members
FWAGEXM FWAGEX1-5 FWAGEX	Amount received from wage and salary income before deduction

OTHRECX	Amount of other money receipts excluded from family income
PERSTAXM	Amount of personal taxes paid
PERSTAX1-5	
PERSTAX	

Here are some examples of situations that may occur. The value for the variable FSMPFRMX (family self-employment income) is computed as the sum of the values reported for the variable SEMPFRMX (member self-employment income) from the MEMD file. SEMPFRMX is subject to topcoding beyond the critical value of $135,000 (-$30,000). The topcode value for SEMPFRMX is $266,848 (-$68,000).

		SEMPFRMX		*FSMPFRMX*	
CU		**REPORTED**	**AFTER TOPCODING**	**VALUE**	**FLAGGED AS TOPCODED?**
CU 1:	Member 1	$125,000	$125,000	370,000	No
	Member 2	125,000	125,000		
	Member 3	120,000	120,000		
CU 2:	Member 1	125,000	276,000	647,848	Yes
	Member 2	140,000	266,848		
	Member 3	105,000	105,000		
CU 3:	Member 1	300,000	266,848	336,848	Yes
	Member 2	70,000	70,000		
CU 4:	Member 1	140,000	140,000	212,000	Yes
	Member 2	140,000	140,000		
	Member 3	-300,000	-68,000		

While CUs 1 and 2 each originally report a total of $370,000 for all members in SEMPFRMX, topcoding is done only on the values reported by the members of CU 2. Thus, the value for FSMPFRMX for CU 2 is higher than for CU 1 and is flagged as topcoded while CU 1 is not. By using the mean of the subset of observations that are above (below) the critical value as the topcode amount, values on the public use data can be either below or above the actual reported value. Note that while CU 2 has a topcoded value above the reported value, CU 3's topcoded FSMPFRMX value is lower than the amount that is reported. The case of CU 4 demonstrates that the reported value for FSMPFRMX can be negative, while the topcoded value can be positive. The reverse can also occur.

The value of the variable, STATE, which identifies state of residence, must be suppressed for some observations to meet the Census Disclosure Review Board's criterion that the smallest geographically identifiable area have a population of at least 100,000. STATE data were evaluated in conjunction with POPSIZE, REGION, and BLS_URBN, which show the population size of the geographic area that is sampled, the four Census regions, and the urban/rural status, respectively. Some STATE codes were suppressed because, in combination with these variables, they could be used to identify areas of 100,000 or less. On approximately 13 percent of the records on the FMLD files the STATE variable is blank.

A small proportion of STATE codes are replaced with codes of states other than the state where the CU resides. By re-coding in this manner, suppression of POPSIZE and REGION may be avoided. (In past releases selected observations of POPSIZE and REGION also required suppression.)

RR01	Alabama	*28	Mississippi
02	Alaska	29	Missouri
04	Arizona	*30	Montana
*05	Arkansas	31	Nebraska
**06	California	32	Nevada
**08	Colorado	33	New Hampshire
09	Connecticut	34	New Jersey
R10	Delaware	**36	New York
11	District of Columbia	*37	North Carolina
12	Florida	**39	Ohio
RR**13	Georgia	40	Oklahoma
15	Hawaii	**41	Oregon
16	Idaho	42	Pennsylvania
**17	Illinois	44	Rhode Island
**18	Indiana	45	South Carolina
**20	Kansas	*46	South Dakota
RR21	Kentucky	**47	Tennessee
22	Louisiana	**48	Texas
**23	Maine	49	Utah
RR 24	Maryland	**51	Virginia
25	Massachusetts	53	Washington
**26	Michigan	**54	West Virginia
R27	Minnesota	RR**55	Wisconsin

* indicates that the STATE code has been suppressed for all sampled CUs in that state.
** indicates that the STATE code has been suppressed for some sampled CUs in that state.
R indicates that either all observations from this state have been re-coded or all strata[1] of observations from this state include “re-codes” from other states.
RR indicates that either some observations from this state have been re-coded or at least one stratum[1] of observations from this state includes “re-codes” from other states.
R* indicates that the STATE code has been suppressed for some sampled CUs in that state, and either STATE has been re-coded or the state includes “re-codes” from other states in all strata[1].
RR** indicates that the STATE code has been suppressed for some sampled CUs in that state and, either STATE has been re-coded or the state includes “re-codes” from other states in at least one stratum[1].

[1] A STATE stratum is a unique POPSIZE and BLS_URBN combination.

States not listed are not in the CE sample.

B. Member Characteristics and Income File (MEMD)

The following table lists MEMD file variables that are subject to topcoding as well as their associated critical values and topcode values. For multiply imputed income variables, it is possible for an upper topcode value to be less than the upper critical value or for a lower topcode value to be greater than the lower critical value.

Variable	Description	2013 Upper Critical Value	2013 Lower Critical Value	2013 Upper Topcode Value	2013 Lower Topcode Value
AGE	Age of member	82	-	88	-
ANFEDTXM	Annual amount of Federal income tax deducted from pay	26,390	-	39,637.50	-

Variable	Description	2013 Upper Critical Value	2013 Lower Critical Value	2013 Upper Topcode Value	2013 Lower Topcode Value
ANFEDTXX	Annual amount of Federal income tax deducted from pay	26,390	-	39,207.85	-
ANGVX	Annual amount of government retirement deducted from pay	9,893	-	16,018.17	-
ANGVXM	Annual amount of government retirement deducted from pay	9,893	-	16,018.17	-
ANPVTX	Annual amount of private pension fund deducted from pay	20,077	-	28,771.05	-
ANPVTXM	Annual amount of private pension fund deducted from pay	20,077	-	29,166.24	-
ANSTATXM	Annual amount of state and local income taxes deducted from pay	9,289	-	12,803.27	-
ANSTATXX	Annual amount of state and local income taxes deducted from pay	9,289	-	12,803.27	-
FEDTXX	Amount of Federal income tax deducted from last pay	1,235	-	2,889.71	-
GROSPAYX	Amount of last gross pay	6,601	-	10,810.96	-
GVX	Amount of government retirement deducted from last pay	965	-	3,962.50	-
IRAX	Amount of money placed in an individual retirement plan	27,000	-	53,107.44	-
JSSDEDX	Estimated annual Social Security contribution	9,429	-	13,536.91	-
JSSDEDXM	Estimated annual Social Security contribution	9,429	-	10,511.55	-
PVTX	Amount of private pension fund deducted from last pay	1,900	-	5,327.14	-
SEMPFRMM	Amount of self-employment income	135,000	-30,000	128,819.20	-25,648.62
SEMPFRMX	Amount of self-employment income	135,000	-30,000	266,818.18	-68,000
SLFEMPSM	Amount of self-employment Social Security contributions	19,957	-	14,706.52	-
SLFEMPSS	Amount of self-employment Social Security contributions	19,957	-	24,659.83	-
STATXX	Amount of state and local income taxes deducted from last pay	393	-	937.24	-

Variable	Description	2013 Upper Critical Value	2013 Lower Critical Value	2013 Upper Topcode Value	2013 Lower Topcode Value
WAGEX	Amount received from wage and salary income before deductions	150,000	-	250,778.89	-
WAGEXM	Amount received from wage and salary income before deductions	150,000	-	199,461.89	-

Special suppression for MEMD file variables

The five MEMD file variables—FEDTXX, GVX, PVTX, RRX, and STATXX—describe deductions from the most recent pay. These variables are used in conjunction with GROSPAYX (amount of last gross pay) and WAGEXM (annual wage and salary income) to derive ANFEDTXM, ANGVXM, ANPVTXM, ANRRXM, and ANSTATXM, which represent the estimated annual deductions for each of these income deduction categories. For example, the estimated annual Federal income tax deduction from pay is calculated as

(1) ANFEDTXM = (WAGEXM (FEDTXX/GROSPAYX)).

Note that WAGEX can be estimated by using the above terms and rearranging such that

(2) WAGEXM = (ANFEDTXM (GROSPAYX/FEDTXX)).

In the above example, a problem with disclosure may arise when neither ANFEDTXM, GROSPAYX, nor FEDTXX (calculation components) are topcoded, *but WAGEXM is*. In this situation WAGEXM can be recalculated to obtain its original value by inserting the non-topcoded values into equation (2) and solving it. In order to prevent this, the non-topcoded terms in equation (2) will be suppressed (blanked out) and their associated flags will be assigned a value of 'T.'

The following chart describes in detail the specific rules that are applied to prevent the potential disclosure outlined above.

If WAGEXM is greater than the critical value but ANFEDTXM, GROSPAYX, and FEDTXX are not, then the values for ANFEDTXM, GROSPAYX, and FEDTXX are suppressed and their flag variables are assigned a value of 'T.'

If WAGEXM is greater than the critical value but ANGVXM, GROSPAYX, and GVX are not, then the values for ANGVXM, GROSPAYX, and GVX are suppressed and their flag variables assigned a value of 'T.'

If WAGEXM is greater than the critical value but ANPVTXM, GROSPAYX, and PVTX are not, then the values for ANPVTXM, GROSPAYX, and PVTX are suppressed and their flag variables assigned a value of 'T.'

If WAGEXM is greater than the critical value but ANRRXM, GROSPAYX, and RRX are not, then the values for ANRRXM, GROSPAYX, and RRX are suppressed and their flag variables assigned a value of 'T.'

If WAGEXM is greater than the critical value but ANSTATXM, GROSPAYX, and STATXX are not, then the values for ANSTATXM, GROSPAYX, and STATXX are suppressed and their flag variables assigned a value of 'T.'

The same special suppression for MEMD file variables occurs with the original (pre-income imputation) variables that correspond to the variables noted above (WAGEX, ANFEDTXX, etc.).

C. Detailed Expenditure File (EXPD)

The following table lists UCCs for which the EXPD variable COST is subject to topcoding as well as their associated critical values and topcode values (rounded to the nearest dollar). If the value of COST is greater (less) than the designated critical values for the above UCCs, COST is set to the topcode value and the associated flag variable, COST_, is set to 'T.'

Variable	Description	2013 Upper Critical Value	2013 Lower Critical Value	2013 Upper Topcode Value	2013 Lower Topcode Value
001000	Purchase price of stocks, bonds, mutual funds	162.75	-	223.21	-
009000	Mortgage payment including coop	3,356.41	-	4,124.46	-
210110	Rent of dwelling, includes parking fees	2,034.38	-	2,673.20	-
210210	Lodging away from home	651.6	-	1,054.88	-
210310	Housing for someone at school	433	-	895.44	-
220400	Purchase of property	93.94	-	92,770.31	-
550320	Medical equipment for general use	99.97	-	137.66	-
550330	Supportive convalescent or medical equipment	80	-	112.22	-
560110	Physicians' services	246.49	-	561.31	-
560210	Dental services	1,172	-	2,263.40	-
560310	Eyecare services	198.4	-	555.44	-
560330	Lab tests and x-rays	140	-	165.93	-
560400	Service by professionals other than physicians	187.25	-	303.30	-
570000	Hospital care not specified	708.82	-	2,299.73	-
570220	Nursing or convalescent home care	1,688	-	5,115.67	-
570230	Other medical care service	150	-	243.41	-
570901	Rental of medical equipment	25	-	49.75	-

D. Income File (DTBD)

The DTBD variable AMOUNT is subject to topcoding for some UCCs. The AMOUNT variable is not topcoded by the conventional method of replacement with a topcode value. First, variables are topcoded in the FMLD files. Then those variables are mapped to their appropriate UCC. If the variable was topcoded in the FMLD files, then the associated UCC will have a topcoded AMOUNT value, and the value of AMOUNT_ is set to 'T.' All the FMLD variables that are topcoded are listed in Section IV, A of this documentation. To obtain the concordance file that lists what FMLD variables are mapped to which UCC,

please contact the Consumer Expenditure Survey via the phone number or email address listed on the last page of this documentation.

Note: For some UCCs multiple topcode values should be expected based on where the original value is mapped from.

V. Estimation Procedure

This section provides users of the CE Diary microdata files with procedures for estimating means and variances of data associated with any U.S. subpopulation. The production of *Consumer Expenditures in 2013* used an integration methodology which incorporated information from *both* Diary and Interview Surveys. Diary data users will not be able to match published CE estimates because of this. In addition, users will not be able to match all values because of suppression of some values, due to topcoding. See the topcoding and other nondisclosure requirements in Section IV.

A. Definition of Terms

Consider the following general situation. We wish to estimate expenditures on certain food items for a special group (subpopulation) of U.S. CUs; for example, all CUs of three persons. Our specific objective is to estimate the expenditures for item *k* over a period of *q* months, where data collected over *r* months are used in the estimate. The following definitions will be helpful in formulating the above type of estimate.

Definition of Terms:

Let

S = all CUs in the subpopulation of interest
k = expenditure item(s) of interest
q = number of months for which estimate is desired
r = number of months in which expenditures were made to be used in calculating the estimate
D = number of days in each of the months in which expenditures were made
j = individual CU in subpopulation S
t = month of expenditure

Then

$X_{(j,k,t)}$ = the amount of money $CU_{(j)}$ spent on item *k* for a week during month *t*
$W_{(j,t,F21)}$ = the weight assigned to $CU_{(j)}$ during month *t*

The F21 denotes FINLWT21 which is used for population estimates.

NOTE: The CUs on the Diary Survey microdata files represent the U.S. population. Some CUs represent more of the population than others; and hence carry more weight. The weight, $W_{(j,t,F21)}$, is a complex estimate of this representation. Refer to Section X.C. Weighting for an explanation of weights. The weights have been adjusted so that the sum of all CU weights for one month approximates one third of the U.S. population. Consequently, the weights for three months (one quarter) of data approximate the total U.S. population.

Using the above terminology, we may define:

$X_{(S,k)(q,r)}$ as an estimate for the expenditures of subpopulation *S* on item *k* over a period of *q* months, where data collected over *r* months are used.

and

$\overline{X}_{(S,k)(q,r)}$ as an estimate of the mean expenditures of subpopulation *S* on item *k* over a period of *q* months, where data collected over *r* months are used.

B. Estimation of Total and Mean Expenditures

As an example, let us estimate total expenditures on milk (item *k*) of subpopulation *S* over a 12-month period. Data collected over 6 months will be used to make the estimate. Users may use less than 12 months of data to perform seasonal calculations. In the notation described above, the estimate is $X_{(S,k)(12,6)}$.

$$X_{(S,k)(12,6)} = 3\left(12/6\right)\sum_{t=1}^{6}\left(\sum_{j=1}^{n}\left(\frac{D_{(t)}}{7}\right)W_{(j,t,F21)}X_{(j,k,t)}\right)_t \qquad (1a)$$

where the inner summation sums expenditures for all *j* in *S*, indexed from *j* = 1 through *n* and the outer summation sums over months *t* = 1 through 6. The factor "3" compensates for the fact that the weights for the CUs visited in one month have been adjusted to represent one third of the U.S. population. The factor "12" reflects our desire to estimate expenditures over a 12-month period; and the "6" is the adjustment made because data for 6 months are used. Since the data $X_{(j,k,t)}$ are in terms of weekly expenditures, the factors, (number of days in the month)/7, are used to convert weekly expenditures into their monthly equivalents.

The above formula can be generalized to estimate the total expenditures of subpopulation *S* on item *k* for *q* months, but using data collected over *r* months. The generalization is

$$X_{(S,k)(q,r)} = 3\left(q/r\right)\sum_{t=1}^{r}\left(\sum_{j=1}^{n}\left(\frac{D_{(t)}}{7}\right)W_{(j,t,F21)}X_{(j,k,t)}\right)_t \qquad (1b)$$

where the inner summation sums expenditures for all *j* in *S*, indexed from *j* = 1 through *n* and the outer summation sums over months *t* = 1 through *r*.

An estimate for the expenditures for two or more items may be obtained by summing those expenditures at the CU level and then proceeding as before.

The next example will give an estimate, $\overline{X}_{(S,k)(12,6)}$, of mean expenditures over twelve months (*q*), on item *k*, of CUs in subpopulation *S*, where data collected over a six month period (*r*) are used. The result is

$$\overline{X}_{(S,k)(12,6)} = \frac{3\left(12/6\right)\sum_{t=1}^{6}\left(\sum_{j=1}^{n}\left(\frac{D_{(t)}}{7}\right)W_{(j,t,F21)}X_{(j,k,t)}\right)_t}{\frac{3\sum_{t=1}^{6}\left(\sum_{j=1}^{n}W_{(j,t,F21)}\right)_t}{6}} \qquad (2a)$$

where the numerator is an estimate of aggregate expenditures as formulated in equation (1a), and where the denominator is an estimate of the population of CUs in the U.S. during the six-month period for which the expenditure data are collected. The inner summation in the denominator of (2a) sums FINLWT21 for a given month (*t*), for all *j* in *S*, indexed from *j* = 1 through *n,* and the outer summation in the denominator of (2a) sums over months *t* = 1 through 6. As in the estimate of aggregate expenditures, the factor "3" to the left of the outer summation in the denominator of equation (2a) adjusts FINLWT21 to represent the entire population for each month of data used. The proper U.S. population count is arrived at by dividing the denominator by *r*, or in this case "6", (representing the 6 month period of collected data in this example).

The above formula generalizes to $\overline{X}_{(S,k)(q,k)}$, (i.e., the estimate of the mean expenditure by subpopulation *S* on item *k* for *q* months using data collected over *r* months). In detail:

$$\overline{X}_{(S,k)(q,r)} = \frac{q\sum_{t=1}^{r}\left(\sum_{j=1}^{n}\left(\frac{D_{(t)}}{7}\right)W_{(j,t,F21)}X_{(j,k,t)}\right)_t}{\sum_{t=1}^{r}\left(\sum_{j=1}^{n}W_{(j,t,F21)}\right)_t} \qquad (2b)$$

Note: The factors "3" (adjustment of FINLWT21 to one U.S. population) and "6", (number of months, *r*, for which the data are collected), which appear both in the numerator and the denominator of (2a), cancel. These scalars are dropped from the general form of $\overline{X}_{(S,k)(q,r)}$.

The estimates for total ($X_{(S,k)(q,r)}$) and mean expenditures ($\overline{X}_{(S,k)(q,r)}$) are based on all CUs; not just the CUs with positive expenditures for item *k*. Consider the calculation for the mean expenditure of tobacco. The formula $\overline{X}_{(S,k)(q,r)}$ includes all CUs, both smoking and nonsmoking. One might be more interested in the mean expenditures on tobacco but only for those CUs that actually have expenditures. This can be accounted for by properly defining the initial subpopulation *S* so as to restrict it to CUs with positive tobacco expenditures.

C. Estimation of Mean Annual Income

Let $\overline{Z}_{(S,r)}$ be an estimate of the mean annual income of CUs in subpopulation S, where income data collected over r months is to be used.

Let $Z_{(j,t)}$ = the annual income reported by CU(j) in month *t*. Then the estimated mean annual income is

$$\overline{Z}_{(S,r)} = \frac{\sum_{t=1}^{r}\left(\sum_{j=1}^{n} W_{(j,t,F21)} Z_{(j,t)}\right)_t}{\sum_{t=1}^{r}\left(\sum_{j=1}^{n} W_{(j,t,F21)}\right)_t}$$

VI. Reliability Statement

A. Description of Sampling Error and Non-Sampling Error

Sample surveys are subject to two types of errors, sampling and non-sampling. Sampling errors occur because observations are not taken from the entire population. The standard error, which is the accepted measure for sampling error, is an estimate of the difference between the sample data and the data that would have been obtained from a complete census. The sample estimate and its estimated standard error enable one to construct confidence intervals.

Assuming the Normal Distribution applies to the means of expenditures, the following statements can be made:

(1) The chances that an estimate from a given sample would differ from a complete census figure by less than one standard error are approximately 68 out of 100.

(2) The chances that the difference would be less than 1.6 times the standard error are approximately 90 out of 100.

(3) The chances that the difference would be less than two times the standard error are approximately 95 out of 100.

Non-sampling errors can be attributed to many sources, such as definitional difficulties, differences in the interpretation of questions, inability or unwillingness of the respondent to provide correct information, mistakes in recording or coding the data obtained, and other errors of collection, response, processing, coverage, and estimation for missing data. The full extent of the non-sampling error is unknown. Estimates using a small number of observations are less reliable. A small amount of non-sampling error can cause a small difference to appear significant even when it is not. It is probable that the levels of estimated expenditure obtained in the Diary Survey are generally lower than the "true" level due to the above factors.

B. Estimating Sampling Error

1. Variance Estimation

Variance estimation can be done in many ways. The method illustrated below (a pseudo-replication technique) is chosen because it is accurate yet simple to understand. The basic idea is to artificially construct several "subsamples" from the original sample data. This construction is done in a manner so that the variance information of the original data is preserved in these subsamples. These subsamples (or pseudo-replications) can then be used to obtain approximate variances for the estimates.

The Diary microdata files contain information that facilitates this form of variance estimation procedure. Specifically, 45 weights are associated with each CU. The forty-fifth weight, called FINLWT21 at BLS,

(which is the weight for the total sample) is used for estimations of total or mean expenditures. The other weights (replicates 1 through 44) are used for variance estimation of the totals or means. Note that half of the weights in each replicate are zero. This reflects the fact that in this technique only half the CUs are used in each of the 44 pseudo-replicates. Recall that $X_{(S,k)(q,r)}$ is an estimate for the expenditures of subpopulation *S* on item *k* over a period of *q* months, where data collected over *r* months are used. This notation does not reveal the fact that 45 replicate weights are to be used for estimation of variance. We expand the notation to include this information. Specifically, let $X_{(S,k)(q,r),a}$ = an estimate of the same quantity as $X_{(S,k)(q,r)}$, but using the weights of the a^{th} replicate.

That is $X_{(S,k)(q,r),a}$ is an estimate of the total expenditures by CUs in subpopulation *S* on item *k* over *q* months using *r* months of collection data, and where the weights from the a^{th} replicate are used. Note that the estimate using any one of the first 44 replicate weights only uses part of the data; hence in general $X_{(S,k)(q,r),a}$ is not equal to $X_{(S,k)(q,r)}$.

An estimate for the variance of $X_{(S,k)(q,r)}$ (denoted by $V(X_{(S,k)(q,r)})$ can be calculated using the following formula:

$$V(X_{(S,k)(q,r)}) = \frac{1}{44}\sum_{a=1}^{44}\left(X_{(S,k)(q,r),a} - X_{(S,k)(q,r)}\right)^2$$

Estimates for the variances of $\overline{X}_{(S,k)(q,r)}$ and $\overline{Z}_{(S,r)}$ are similar and are given below.

$$V\left(\overline{X}_{(S,k)(q,r)}\right) = \frac{1}{44}\sum_{a=1}^{44}\left(\overline{X}_{(S,k)(q,r),a} - \overline{X}_{(S,k)(q,r)}\right)^2$$

and

$$V\left(\overline{Z}_{(S,r)}\right) = \frac{1}{44}\sum_{a=1}^{44}\left(\overline{Z}_{(S,r),a} - \overline{Z}_{(S,r)}\right)^2$$

where $\overline{X}_{(S,k)(q,r),a}$ and $\overline{Z}_{(S,r),a}$ are estimates similar to $\overline{X}_{(S,k)(q,r)}$ and $\overline{Z}_{(S,r)}$ except weights of the a^{th} replicates are used.

2. Standard Error of the Mean

The standard error of the mean, *S.E.*($\overline{X}$), is defined as the square root of the variance of the mean. *S.E.*($\overline{X}$), is used to obtain confidence intervals that evaluate how close the estimate may be to the true population mean. A 95 percent confidence interval can be constructed around an estimate, bounded by values 1.96 times the standard error less than and greater than the estimate. For example, the average weekly expenditure for food away from home for All CUs in 2013 was $47.53. The standard error for this estimate is $3.20. Hence, the 95 percent confidence interval around this estimate is from $41.26 to $53.80. Therefore, we could conclude with 95 percent confidence that the mean weekly expenditures for food away from home for all CUs in 2013 lies within the interval $41.26 to $53.80.

3. Standard Error of the Difference between Two Means

Standard errors may also be used to perform hypothesis testing, a procedure for distinguishing between population parameters using sample estimates. The most common types of hypotheses are: 1) the population parameters are identical; versus 2) they are different.

For example, in 2013 the estimated average weekly expenditure for food away from home for CUs in the *Managers and professionals* occupation category is $62.76 and the estimate for CUs in the *Construction workers and mechanics* category is $45.44 The apparent difference between the two mean expenditures is $62.76 – $45.44 = $17.32. The standard error on the estimate of $62.75 is $1.61 and the estimated standard error for the $45.44 estimate is $3.81. The standard error (S.E.) of a difference is approximately equal to

$$S.E.(\overline{X}_1, \overline{X}_2) = \sqrt{\left(V(\overline{X}_1) + V(\overline{X}_2)\right)}$$

where

$$V(\overline{X}_i) = \left(S.E.(\overline{X}_i)\right)^2$$

This assumes that $\overline{x}_1$ and $\overline{x}_2$ are disjoint subsets of the population. Hence, the standard error of the difference in food away from home expenditures between CUs in the *Managers and professionals* occupation group and in the *Construction workers and mechanics* group is about

$$\sqrt{\left((1.61)^2 + (3.81)^2\right)} = 4.14$$

This means that the 95 percent confidence interval around the difference is from $9.21 to $25.43. Since this interval does not include zero, we can conclude with 95 percent confidence that the mean weekly food away from home expenditures for the *Managers and professionals* occupation group is more than the mean weekly food expenditures for the *Construction workers and mechanics* group.

Analyses of the difference between two estimates can also be performed on non-disjoint sets of the population, where one is a subset of the other. The formula for computing the standard error (S.E.) of the difference between two non-disjoint estimates is

$$S.E.(\overline{X}_1, \overline{X}_2) = \sqrt{\left(V(\overline{X}_1) + V(\overline{X}_2) - 2r\left(V(\overline{X}_1) * V(\overline{X}_2)\right)\right)}$$

where

$$V(\overline{X}_i) = \left(S.E.(\overline{X}_i)\right)^2$$

and where *r* is the correlation coefficient between $\overline{x}_1$ and $\overline{x}_2$. The correlation coefficient is generally no greater than 0.2 for CE estimates.

VII. Microdata Verification and Estimation Methodology

Sample programs available for download on the PUMD homepage, illustrate the methodology CE uses in producing publication tables, and offers an example of coding to access the data and produce a sample table. The programs are written in SAS and STATA, and shows usage of these data sets available online. (Note: CE data published by BLS may not match some values estimated using the microdata due to

topcoding of data and CE publication programming methodology.) All variables and ranges referred to in the program are described in detail in the diary data dictionary (available alongside this documentation online).

It should be emphasized that these programs have been written solely for the verification of the microdata and as an illustration of the CE estimation methodology. They should not be used for any other purpose.

VIII. Description of the Survey

The CE program consists of two separate components, each with its own questionnaire and independent sample:

1) A Diary or recordkeeping survey completed by the sample CUs for two consecutive 1-week periods; the sample is surveyed across a 12-month period.

2) An Interview panel survey in which each CU in the sample is interviewed once every 3 months over five consecutive quarters to obtain a year's worth of data. New panels are initiated every month of the year.

Data are collected by the Bureau of the Census under contract with BLS. All data collected in both surveys are subject to The U.S. Census Bureau confidentiality requirements, which prevent the disclosure of the CU member's identity.

The Diary survey collects expenditure data for items purchased each day over two one-week periods. This survey is designed to collect expenditure data for small, frequently purchased items such as food, beverages, food consumed away from home, gasoline, housekeeping supplies, nonprescription drugs and medical supplies, and personal care products and services. Respondents are not limited to recording expense for these items only.

A Household Characteristics Questionnaire is completed to record demographic and family characteristics data pertaining to age, sex, race, marital status, and CU relationships each CU member. Income information, such as wage, salary, unemployment compensation, child support, and alimony, as well as information on the employment of each CU member age 14 and over is collected. The expenditure collection instrument is a self-reporting, product-oriented diary on which respondents record all expenses for two consecutive one-week periods. It is divided by day of purchase and by broad classification of goods and services, a format designed to aid the respondents when recording daily purchases.

At the beginning of the two-week collection period, the interviewer uses the Household Characteristics Questionnaire to record demographic and characteristics information pertaining to CU members. Also at this time, a diary for the first week is left with the participating CU. At the completion of the first week, the interviewer picks up the diary, reviews the entries, clarifies any questions, and leaves a second diary for the following week. At the end of the second week, the diary is picked up and reviewed. At this point, the interviewer again uses the Household Characteristics Questionnaire to collect information on CU income, employment and earnings of CU members. These data, along with the other household characteristics information, permit data users to classify sample units for research purposes, and allow BLS to adjust population weights for CUs who do not cooperate in the survey.

IX. Data Collection and Processing

In addition to its data collection duties, the U.S. Census Bureau is responsible for field editing and coding, consistency checking, quality control, and data transmittal to BLS. BLS performs additional review and editing procedures in preparing the data for publication and release.

A. The US Census Bureau Activities

Data collection activities have been conducted by the U.S. Census Bureau on a continuing basis since October 1979. Due to differences in format and design, the Diary Survey and the Interview Survey data are collected and processed separately. Preliminary Diary survey data processing carried out by the U.S. Census Bureau includes programming the Computer Assisted Personal Interview (CAPI) instrument used to collect household characteristics, keying the expenditure data from the diary questionnaire, clerical data editing, and correcting for inconsistencies in the collected data.

The data collected on household characteristics using CAPI are sent directly to the Census Demographic Surveys Division (DSD). Upon completion of the written questionnaire by respondents, the diaries are sent from the regional offices to the Census National Processing Center (NPC) in Jeffersonville, IN. At the NPC, the expenditure data are keyed and codes are applied. The keyed expenditure data are sent to DSD, where they are merged with the household characteristic data. Inconsistencies and errors in the combined data are identified and corrected.

After clerical processing at the NPC, the data are transmitted to the Census Processing Center in Suitland, MD, where they pass through basic quality checks of control counts, missing values, etc. The data are then electronically transmitted to BLS in Washington, DC.

B. Bureau of Labor Statistics Activities

Upon receipt from the U.S. Census Bureau, the data undergo a series of computer edits that identify and correct irregularities and inconsistencies. Other adjustments apply appropriate sales taxes and derive CU weights based on BLS specifications. In addition, demographic and work experience items are imputed when missing or invalid. All data changes and imputations are identified with flags on the Interview data base.

Next, BLS conducts an extensive review to ensure that severe data aberrations are corrected. The review takes place in several stages: a review of counts, weighted means, and unweighted means by region; a review of family relationship coding inconsistencies; a review of selected extreme values for expenditure and income categories; and a verification of the various data transformations.

Cases of extreme data values are investigated by reviewing images of the questionnaires. Errors discovered through this procedure are corrected prior to release of the data.

Two major types of data adjustment routines--imputation and allocation--are carried out to improve and classify the estimates derived from the Diary Survey. Data imputation routines correct for missing or invalid entries among selected CU characteristic fields. Allocation routines are applied when respondents provided insufficient expenditure detail to meet tabulation requirements. For example, reports of combined expenditures for fuels and utilities are allocated among gas, electricity, and other items in this group. To analyze the effects of these adjustments, tabulations are made before and after the data adjustments.

X. Sampling Statement

A. Survey Sample Design

Samples for the CE are national probability samples of households designed to be representative of the total U. S. civilian population. Eligible population includes all civilian non-institutionalized persons.

The first step in sampling is the selection of primary sampling units (PSUs), which consist of counties (or parts thereof) or groups of counties. The set of sample PSUs used for the 2013 sample is composed of 91 areas. The design classifies the PSUs into four categories:

21 "A" certainty PSUs are Metropolitan Statistical Areas (MSA's) with a population greater than 1.5 million.
38 "X" PSUs, are medium-sized MSAs.
16 "Y" PSUs are nonmetropolitan areas that are included in the CPI.
16 "Z" PSUs are nonmetropolitan areas where only the urban population data will be included in the CPI.

The sampling frame (that is, the list from which housing units were chosen) for the 2013 survey is generated from the 2000 Population Census file. The sampling frame is augmented by new construction permits and by techniques used to eliminate recognized deficiencies in census coverage. All Enumeration Districts (EDs) from the Census that fail to meet the criterion for good addresses for new construction, and all EDs in non-permit-issuing areas are grouped into the area segment frame.

To the extent possible, an unclustered sample of units is selected within each PSU. This lack of clustering is desirable because the sample size of the Diary Survey is small relative to other surveys, while the intraclass correlations for expenditure characteristics are relatively large. This suggests that any clustering of the sample units could result in an unacceptable increase in the within-PSU variance and, as a result, the total variance.

Each selected sample unit is requested to keep two 1-week diaries of expenditures over consecutive weeks. The earliest possible day for placing a diary with a household is predesignated with each day of the week having an equal chance to be the first of the reference week. The diaries are evenly spaced throughout the year.

B. Cooperation Levels

The annual target sample size at the United States level for the Diary Survey is 7,050 participating sample units. To achieve this target the total estimated work load is 12,100 sample units. This allows for refusals, vacancies, or nonexistent sample unit addresses.

Each participating sample unit selected is asked to keep two 1-week diaries. Each diary is treated independently, so response rates are based on twice the number of housing units sampled.

The response rate for the 2013 Diary Survey is 60.8% as shown below.

			Eligible housing unit interviews			
Year	Number of Diaries Designated for the Survey	Type B or C Ineligible Cases	Number of Potential Diaries	Type A Non-Responses	Total Respondent Interviews	Response Rate
2009	25,211	5,187	20,024	5,400	14,624	73.0%
2010	25,158	5,170	19,988	5,692	14,296	71.5%
2011	25,258	5,435	19,823	5,898	13,925	70.2%
2012	25,356	5,058	20,298	6,537	13,761	67.8%
2013	25,362	5,066	20,296	7,961	12,335	60.8%

Type B or C cases are housing units that are vacant, nonexistent, or ineligible for diary placement. Type A non-responses are housing units which the interviewers were unable to contact or the respondents refused to participate in the survey. The response rate stated above is based only on the eligible housing units (i.e., the designated sample cases less type B and type C ineligible cases).

C. Weighting

Each CU included in the CE represents a given number of CUs in the U.S. population, which is considered to be the universe. The translation of sample families into the universe of families is known as weighting. However, since the unit of analysis for the CE is a CU, the weighting is performed at the CU level. Several factors are involved in determining the weight for each CU for which a diary is obtained. There are four basic steps in the weighting procedure:

1) The basic weight is assigned to an address and is the inverse of the probability of selection of the housing unit.

2) A weight control factor is applied to each diary if subsampling is performed in the field.

3) A noninterview adjustment is made for units where data could not be collected from occupied housing units. The adjustment is performed as a function of region, housing tenure, family size and race.

4) A final adjustment is performed to adjust the sample estimates to national population controls derived from the Current Population Survey. The adjustments are made based on both the CU's member composition and on the CU as a whole. The weight for the CU is adjusted for individuals within the CU to meet the controls for the 14 age/race categories, 4 regions, and 4 region/urban categories. The CU weight is also adjusted to meet the control for total number of CUs and total number of CU who own their living quarters. The weighting procedure uses an iterative process to ensure that the sample estimates will meet all the population controls.

NOTE: The weight for a consumer unit (CU) can be different for each week in which the CU participates in the survey as the CU may represent a different number of CUs with similar characteristics.

D. State Identifier

Since the CE is not designed to produce state-level estimates, summing the consumer unit weights by state will not yield state population totals. A CU's basic weight reflects its probability of selection among a group of primary sampling units of similar characteristics. For example, sample units in an urban nonmetropolitan area in California may represent similar areas in Wyoming and Nevada. Among other adjustments, CUs are post-stratified nationally by sex-age-race. For example, the weights of consumer units containing a black male, age 16-24 in Alabama, Colorado, or New York, are all adjusted equivalently. Therefore, weighted population state totals will not match population totals calculated from other surveys that are designed to represent state data.

To summarize, the CE sample was not designed to produce precise estimates for individual states. Although state-level estimates that are unbiased in a repeated sampling sense can be calculated for various statistical measures, such as means and aggregates, their estimates will generally be subject to large variances. Additionally, a particular state-population estimate from the CE sample may be far from the true state-population estimate.

XI. Interpreting the Data

Several factors should be considered when interpreting the expenditure data. The average expenditure for an item may be considerably lower than the expenditure by those CUs that purchased the item. The less frequently an item is purchased, the greater the difference between the average for all consumer units and the average of those purchasing (see Section V.B. Estimation of Total and Mean Expenditures). Also, an individual CU may spend more or less than the average, depending on its particular characteristics. Factors such as income, age of family members, geographic location, taste and personal

preference also influence expenditures. Furthermore, even within groups with similar characteristics, the distribution of expenditures varies substantially.

Expenditures reported are the direct out-of-pocket expenditures. Indirect expenditures, which may be significant, may be reflected elsewhere. For example, rental contracts often include utilities. Renters with such contracts would record no direct expense for utilities, and therefore, appear to have no utility expenses. Employers or insurance companies frequently pay other costs. CUs with members whose employers pay for all or part of their health insurance or life insurance would have lower direct expenses for these items than those who pay the entire amount themselves. These points should be considered when relating reported averages to individual circumstances.

XII. Appendix 1—Glossary

Population
The civilian non-institutional population of the United States as well as that portion of the institutional population living in the following group quarters: Boarding houses, housing facilities for students and workers, staff units in hospitals and homes for the aged, infirm, or needy, permanent living quarters in hotels and motels, and mobile home parks. Urban population is defined as all persons living in a Metropolitan Statistical Area (MSA's) and in urbanized areas and urban places of 2,500 or more persons outside of MSA's. Urban, defined in this survey, includes the rural populations within MSA. The general concept of an MSA is one of a large population nucleus together with adjacent communities that have a high degree of economic and social integration with that nucleus. Rural population is defined as all persons living outside of an MSA and within an area with less than 2,500 persons.

Consumer unit (CU)
A consumer unit comprises either: (1) all members of a particular household who are related by blood, marriage, adoption, or other legal arrangements; (2) a person living alone or sharing a household with others or living as a roomer in a private home or lodging house or in permanent living quarters in a hotel or motel, but who is financially independent; or (3) two or more persons living together who use their income to make joint expenditures. Financial independence is determined by the three major expense categories: housing, food, and other living expenses. To be considered financially independent, at least two of the three major expense categories have to be provided entirely or in part by the respondent.

Reference person
The first member mentioned by the respondent when asked to "Start with the name of the person or one of the persons who owns or rents the home." It is with respect to this person that the relationship of other CU members is determined.

Income before taxes
The combined income earned by all CU members 14 years old or over during the 12 months preceding the interview. The components of income are: Wage and salary income, business income, farm income, Social Security income and Supplemental Security income, unemployment compensation, workmen's compensation, public assistance, welfare, interest, dividends, pension income, income from roomers or boarders, other rental income, income from regular contributions, other income, and food stamps.

Income after taxes
Income before taxes minus personal taxes, which includes Federal income taxes, state and local taxes, and other taxes.

Geographic regions
CUs are classified by region according to the address at which they reside during the time of participation in the survey. The regions comprise the following States:

Northeast - Connecticut, Maine, Massachusetts, New Hampshire, New Jersey, New York, Pennsylvania, Rhode Island, and Vermont

Midwest - Illinois, Indiana, Iowa, Kansas, Michigan, Minnesota, Missouri, Nebraska, North Dakota, Ohio, South Dakota, and Wisconsin

South - Alabama, Arkansas, Delaware, District of Columbia, Florida, Georgia, Kentucky, Louisiana, Maryland, Mississippi, North Carolina, Oklahoma, South Carolina, Tennessee, Texas, Virginia, and West Virginia

West - Alaska, Arizona, California, Colorado, Hawaii, Idaho, Montana, Nevada, New Mexico, Oregon, Utah, Washington, and Wyoming

XIII. Appendix 2—UCC Aggregation

The Dstub file found with the documentation and sample programs when you download the data shows the UCC aggregation used in the sample programs.

XIV. Appendix 3—Publications and Data Releases from the Consumer Expenditure Survey

Consumer Expenditure Survey Data on the Internet

CE reports and data tables can be found on-line at http://www.bls.gov/cex/home.htm.
The following One-Year, Mid-Year and Two-Year Tables of integrated Diary and Interview data are available under the Tables Created by BLS heading:

One-Year Tables
Standard Tables from 1984-2011
Expenditure Shares Tables from 1998-2011
Aggregate Expenditure Shares Tables from 1998-2013
Combined Expenditure, Share and Standard Error Tables from 2013

Mid-Year Tables (July 20xx – June 20xy, where xy= xx+1)
Combined Expenditure, Share and Standard Error Tables from 2011-13

Two-Year Tables
Cross-Tabulated Tables from 1986-2013
Metropolitan Statistical Area Tables from 1986-2013
Region Tables from 1998-2013
High Income Tables from 1998-2002
Multi-Year Tables for 1984-1992 and 1994-2013

CDs and Free Online Data
The data releases are to be made available online in reverse chronological order, starting with the 2010 data release in July 2012, with prior years appearing incrementally until the 1996 data release is posted. Post-1995 data releases will remain available on CD for purchase until posted online. Please see PUMD on CD for ordering information. Pre-1996 PUMD will continue to only be available on CD for purchase.

For information and downloading of past PUMD releases, please visit the links below. Multiple zip files can also be downloaded at one time. Please see Instructions for Downloading Consumer Expenditure Survey (CE) Microdata and Documentation for information on downloading the files.
Public Use Microdata that are not available online must be purchased through the Bureau of Labor Statistics Division of Financial Planning and Management. To purchase CDs by check or charge, print and complete the order form (PDF) and return it with payment to: Bureau of Labor Statistics Division of Financial Planning and Management, Room 4135, 2 Massachusetts Avenue, NE Washington, DC 20212-0001. Phone (202) 691-7794, Fax (202) 691-7796.

CE microdata on CD are available from the Bureau of Labor Statistics for 1972-73, 1980-81, 1990-91, 1992-93, and for each individual year after 1993 (excluding those years which are currently available for free download online). The 1980-81 through 2013 releases contain Interview and Diary data, while the 1972-73 CD includes Interview data only. The 1980-81, and the 1990 files (of the 1990-91 CD) include selected EXPN data, while the 1991 files (from the 1990-91 CD) and the 1992-93 CD do not. In addition to the Interview and Diary data, the CDs from 1994-2004 include the complete collection of EXPN files. A 1984-94 "multi-year" CD that presents Interview FMLI file data is also available. In addition to the microdata, the CDs also contain the same integrated Diary and Interview tabulated data (1984-2009) that are found on the Consumer Expenditure Survey web site (http://www.bls.gov/cex).

More information on the particular CDs available and the order form can be found on the Consumer Expenditure Survey web site: http://www.bls.gov/cex/pumdhome.htm#order.

XV. Inquiries, Suggestions and Comments

If you have any questions, suggestions, or comments about the survey, the microdata, or its documentation, please call (202) 691-6900 or email cexinfo@bls.gov.

Written suggestions and comments should be forwarded to:

Division of Consumer Expenditure Survey
Branch of Information and Analysis
Bureau of Labor Statistics, Room 3985
2 Massachusetts Ave. N.E.
Washington, DC. 20212-0001

The Bureau of Labor Statistics will use these responses in planning future releases of the microdata.