Survey Procedures

Technical details of the survey include the sample design, cooperation levels, sample weighting factors, and data collection and processing.

Sample Design

Samples for the CE are national probability samples of households designed to be representative of the total U.S. civilian population. The eligible population includes all civilian noninstitutional persons.

The first step in the sampling is the selection of primary sampling units (PSUs), which consist of counties (or parts thereof) or groups of counties. The design classifies the PSUs into four categories:

- 31 "A" PSUs are Metropolitan Statistical Areas (MSAs) with a population greater than 1.5 million.
- 46 "B" PSUs are medium-sized MSAs.
- 10 "C" PSUs are nonmetropolitan areas that are included in the CPI.
- 18 "D" 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) is generated from the 1990 Population Census 100-percent-detail 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 nonpermit-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 variance within the PSU and, as a result, within 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.

Cooperation Levels

The annual target sample size at the U.S. level for the diary survey is 7,800 participating sample units. To achieve this target, the Census Bureau issues a total estimated workload of 11,275 sample units. This size allows for refusals, vacancies, or nonexistent sample unit addresses.

The Bureau asks each participating sample unit selected 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 2004 diary survey was 68.9 percent.

Sample Weighting Factors

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. Since the unit of analysis for the CE is a CU, the weighting is performed at the CU level. Several factors help determine the weight for each CU for which a diary is obtained. The weighting procedure comprises four basic steps: (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; and (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 both on 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 14 age/race categories, 4 regions, and 4 region/urban categories. The CU weight is also adjusted to control for the total number of CUs and the total number of CUs who own their own living quarters. The weighting procedure uses an iterative process to ensure that the sample estimates will meet all population controls.

Data Collection and Processing

In addition to collecting data, the Census Bureau performs field editing and coding, consistency checking, quality control, and data transmittal to BLS. BLS performs additional review and editing to prepare the data for publication and release.

The Census Bureau has conducted data collection on a continuing basis since October 1979. The Bureau collects data from the diary survey and the interview survey separately, due to differences in format and design. Preliminary diary survey data processing carried out by the Bureau includes keying the data from the questionnaires, clerical data editing, and correcting for inconsistencies in the collected data.

Regional Census Bureau offices send completed surveys to the Census National Processing Center (NPC) in Jeffersonville, IN. NPC staff apply codes to the surveys to identify demographic characteristics, expenditures, and inconsistencies, and to identify and correct errors. In 2004, the processing of surveys changed with the introduction of Computer Assisted Personal Interviewing (CAPI).

After processing the data at the NPC, the Bureau transmits them 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 transmitted to BLS.

Upon receiving the data, BLS performs 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, BLS imputes demographic and work experience items (as well as income in 2004) when these data are missing or invalid. All data changes and imputations are identified with flags on the interview database.

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 verification of the various data transformations.

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 missing or invalid entries among selected CU characteristics and expenditure fields. Allocation routines are applied when respondents provide 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.