

B. Collection of Information Employing Statistical Methods

1. **Sampling Method**

Consumer Units

There are approximately 119.6 million consumer units (CUs) in the potential Consumer Expenditure (CE) Surveys universe.¹ A CU is the unit from which we desire expenditure reports. It consists of all household members in a particular housing unit or other type of living quarters who are related by blood, marriage, adoption, or some other legal arrangement. The CU determination for unrelated persons is based on financial dependence. Unrelated persons are considered separate CU(s) if they are responsible for paying their own expenses for two out of three of the following expense categories: shelter, food, and all other expenses. Approximately 97 percent of all occupied living quarters are a single CU.

For an overview of the CE sample design and the CU selection process, please refer to the 2008 CE Anthology article, “Selecting a Sample of Households for the Consumer Expenditure Survey” by Susan King and Sylvia Johnson-Herring (attachment Q).

The following table shows estimated numbers of CUs in all 91 strata from which PSUs were selected.² (See the section below entitled “PSUs” for more information.)

¹¹ The number of CUs comes from dividing the Census Bureau’s 2008 estimate of the number of people in the civilian non-institutional population (299 million) by the average number of people per CU (2.5).

²² The number of CUs per stratum comes from allocating the nationwide total of 119.6 million CUs by each stratum’s proportion of the nationwide population in the 2000 Census.

Estimated Number of CUs in CE's 91 Strata

Stratum Code	Estimated Number of CUs in Stratum
A102	2,629,569
A103	3,016,199
A104	1,087,340
A109	3,402,836
A110	3,279,821
A111	2,830,676
A207	3,897,363
A208	2,318,517
A209	1,138,351
A210	1,251,727
A211	1,332,617
A312	2,136,385
A313	1,084,805
A316	2,241,817
A318	2,003,646
A319	1,785,160
A320	1,647,131
A321	997,881
A419	5,254,338
A420	1,703,066
A422	2,991,129
A423	1,510,470
A424	1,195,639
A425	1,076,881
A426	414,715
A427	151,280
A429	1,381,770
A433	1,244,860
X102	1,945,709
X104	1,130,986
X108	1,497,036
X210	1,510,998
X212	1,011,446
X218	1,153,093
X220	1,167,055
X222	1,280,832
X224	1,830,292
X226	1,034,214
X228	802,593
X232	742,678
X336	1,683,293
X338	1,170,101
X340	1,205,294
X342	1,566,851

X344	1,226,880
X346	1,111,246
X350	942,413
X352	1,203,472
X354	1,733,003
X356	1,223,329
X358	1,715,989
X362	1,566,804
X364	1,748,072
X366	886,473
X368	1,227,813
X472	1,745,624
X474	1,969,954
X482	1,474,859
X484	1,289,188
Y102	578,186
Y104	672,236
Y206	847,740
Y208	1,008,201
Y210	840,467
Y212	1,007,920
Y314	835,119
Y316	964,801
Y318	800,186
Y320	842,509
Y322	996,079
Y324	797,423
Y426	542,557
Y428	495,967
Y430	611,360
Y432	591,352
Z102	380,047
Z104	591,168
Z206	872,968
Z208	641,906
Z210	789,314
Z212	937,433
Z314	794,454
Z316	753,878
Z318	798,845
Z320	1,048,830
Z322	969,751
Z324	576,997
Z426	278,863
Z428	242,296

Z430	334,795
Z432	352,772
Total	119,600,000

Response Rates

The following table shows expected response rates for the Quarterly Interview Survey (CEQ) and Diary Survey (CED) based on 2008 response rates.

Category	Quarterly (quarter)	Diary (annual)
Total Sample Size	14,500	12,400
Total Type B and C Noninterviews (vacant, demolished, etc.):		
Number	2,900	2,600
Percent of Total Sample	20.0	21.0
Total Eligible Units	11,600	9,800
Total Type A Noninterviews		
Number	3,000	2,700
Percent of Total Eligible	25.9	27.6
Total Completed Interviews		
Number	8,600	7,100
Percent of Total Eligible	74.1	72.4

For more information on the calculation of response rates, please see the 2008 CE Anthology article “Response Rates in the Consumer Expenditure Survey,” by Sylvia Johnson-Herring and Sharon Krieger (attachment R).

2. Collection Methods

PSUs

The primary sampling units (PSUs) used in the CEQ and CED are small clusters of counties. The average number of counties in a PSU is approximately five. The set of sample PSUs used in the two CE surveys consist of 91 PSUs, 75 of which are also used in the Consumer Price Index (CPI). The 91 PSUs fall into four categories:

PSU “size class”	Number of PSUs	Description
A	28	Large Metropolitan CBSAs (self-representing PSUs)
X	31	Small Metropolitan CBSAs (non-self-representing PSUs)
Y	16	Micropolitan CBSAs (non-self-representing PSUs)
Z	16	Non-CBSA Areas (non-self-representing PSUs)

The BLS selected these PSUs from a stratified sample design in which one PSU was selected from each stratum. Stratification of the non-self-representing PSUs (the X, Y, and Z PSUs) used a 5-variable geographic model whose independent variables were latitude, longitude, latitude squared, longitude squared, and the percent of consumer units living in an urban area.

Sampling Within PSUs

Four non-overlapping sampling frames are used to select CUs for the expenditure surveys. The four frames are: Unit, Area, Group Quarters (GQ), and Permit. The first three frames, Unit, Area, and GQ, are called “old construction.” The Permit frame is called “new construction” and consists of housing units built after April 1, 2000. The Unit frame covers 80% of the sample and consists of addresses of housing units located in census blocks in areas that issue building permits and in which a high percentage of the addresses are “complete” (they contain a street name and house number). The Area frame, which covers 10% of the sample, contains addresses from the remaining census blocks that are not in permit-issuing areas, or where more than 4 percent of the addresses in the blocks are missing. These addresses are mostly in rural areas. The GQ frame, which covers 1% of the sample, includes boarding houses, hotel rooms, and institutions that are found in the decennial census but are not counted as housing units. These addresses are in census blocks that contain complete addresses and are covered by building permits. Since the census does not cover housing units built after April 1 of the census year, the list of addresses is supplemented by a sample of building permits from the Building Permit Survey which is conducted by the Census Bureau. The Permit frame is updated monthly and covers about 9% of the sample.

Within each PSU, a “systematic sample” of households is selected from each of the four frames. The objective is to minimize the within-PSU variance of expenditure estimates by grouping households with similar characteristics together. The households are sorted by variables that are correlated with their expenditures. Each frame has different sort variables. For the Unit frame each address is assigned to a domain based on whether the address is rental or owned property. Both renter and owner domains are further subdivided into quartiles of rental and property values. These quartiles of rental and property values are unique to a geographic areas based on PSU, FIPS State, FIPS County, Census Tract, Census Block, Basic Street Address, and Unit Sort Order code³. The eight domains are further classified as to whether the housing unit is vacant or occupied by 1, 2, 3, or more than 4 people, and each cell is assigned a stratification code value (see table 1).

Table 1. CE Unit Frame Stratification Code Values

Renter/Owner Quartile	Number of Occupants				
	Vacant	1 person	2 persons	3 persons	4+ persons
Renters 1 st Quartile	10	11	12	13	14
Renters 2 nd Quartile	25	24	23	22	21
Owners 1 st Quartile	30	31	32	33	34
Owners 2 nd Quartile	45	44	43	42	41
Renters 3 rd Quartile	50	51	52	53	54
Renters 4 th Quartile	65	64	63	62	61
Owners 3 rd Quartile	70	71	72	73	74
Owners 4 th Quartile	85	84	83	82	81
Residual Vacant	99				

³ The Unit Sort Order code was created by the Census Bureau to preserve geographic proximity between units in the sort.

All addresses in the Unit frame fall into one of these cells. When the addresses are placed in order, those whose rent is in the lowest quartile and have a small number of occupants are at one extreme and those whose property values are in the highest quartile with a small number of occupants are at the other extreme. The stratification code is a surrogate for sorting by expenditures. To draw a systematic sample, the Unit frame addresses are sorted by PSU, urban/rural classification, FIPS State code, FIPS County code, CE Unit frame stratification code, Census Tract Code, Census Block code, Basic Street Address, and Unit Sort Order code. A starting point for a systematic sample is located on the interval using a random number. Then after the initial household is selected, another household is selected every “k” households down the list where “k” is the inverse of the PSU’s probability of selection times the PSU sampling interval.

Samples are selected in the Area frame similar to the Unit frame but using different variables. A Combined Block Stratification Code is calculated using median household size and the proportion of owner-occupied housing units. To draw a systematic sample, the Area frame addresses are sorted by PSU, an urban/rural variable, Combined Block Housing Flag⁴, FIPS State code, FIPS County code, Combined Block Stratification Code, Census Tract, and a Combined Block code. Using the sorted sample, a sampling interval is calculated and a starting point randomly selected.

The GQ and Permit frames do not have a stratification code but have a within-PSU sort. As in the other frames, after the sort, a sampling interval is calculated and a starting point randomly selected. The sort variables in the GQ frame are: PSU, FIPS State code, FIPS County code, Census Tract, Combined Block Code, and a Within Combined Block Code.

The sampling points in the systematic sample are called “hits.” For CE, the “hits” are alternatively assigned to the CEQ and CED. In the unit frame, at each “hit” 24 households are selected surrounding the “hit.” These households are interviewed over several quarters. Instead of 24 units, the other three frames select a “measure,” which is typically 4 units.

For more information on sampling within PSUs for the CE Surveys, please refer to the 2008 CE Anthology article, “Selecting a Sample of Households for the Consumer Expenditure Survey” by Susan King and Sylvia Johnson-Herring (attachment Q).

Estimation

The estimation procedure for both the CED and CEQ follow well-established statistical principles. The final weight for each sample CU is the product of the inverse of the probability of selection; a weight adjustment to account for noninterviews; and a calibration adjustment that post-stratifies the weights to account for population undercoverage.

For additional information on the sample design and estimation methodology used in the CE surveys, please refer to “Chapter 16, Consumer Expenditures and Income” in the *BLS Handbook of Methods* (attachment S). Another source of additional information is Kenneth V. Dalton’s memo to Chester E. Bowie, “Specifications for the Selection of CE/CPI Samples in PSUs Based on the 2000 Census,” June 28, 2002 (attachment T); and Alan R. Tupek’s memo to

⁴ Combined Block College Housing Flag indicates whether the combined block contains college housing.

Kenneth V. Dalton, "Calculations of Within-PSU Sampling Intervals for the Census 2000-Based Redesign of the Consumer Expenditure Surveys and the CPI Permit New Construction Housing Sample," November 11, 2002 (attachment U).

3. **Methods to Maximize Response Rates**

In the CE Surveys, keeping the noninterview rate at a low level requires special efforts, particularly from the Census Bureau Field staff. For each refusal case, the regional office sends a special letter to the address and assigns the case for follow-up by the program supervisor, supervisory field representative, or senior interviewer, taking into account time and cost considerations.

To adjust for those noninterviews that the field staff cannot convert to interviews, the sample design provides for a noninterview adjustment in the estimation procedure. The computer processing employs special techniques in the CEQ to reference data provided in the previous interview, keeping recall problems and interview time to a minimum.

4. **Testing Plans**

Pending funding and resource availability, CE plans to conduct studies (prior to the expiration of the clearance) using the production sample on the topics listed below. An NCR will be submitted for all of the proposed studies should funding and resources become available.

CEQ Interview

- 1) \$40 CEQ Incentive – Continue to study the results of the prior CEQ Incentives test and determine if additional testing is needed.
- 2) Bounding Effect Study - A test of bounding effects, by removing sections of the first wave questionnaire in an effort to decrease respondent burden;
- 3) Telephone Friendly Questions - A field test of questions that are more neutral between personal visit and telephone interviewing. Current questions in the interview were designed for a personal visit setting, relying on long lists from the Information Booklet and do not convey well in telephone interviewing.
- 4) Promotional Materials Test - A test of providing promotional materials to respondents in order to encourage participation in subsequent waves.
- 5) CEQ Research Questions section – Test the feasibility of adding a new section to the CEQ questionnaire where research questions can be added ad hoc. These questions may be directed to the respondent or the interviewer.

CED Diary

- 1) Web Diary Test - Test of a web diary designed to decrease respondent burden by offering respondents an electronic way to report expenditures;
- 2) Individual Diaries Test – An individual diary test designed to increase reported expenditures by collecting data from all consumer unit members. This test will most likely be combined with the web diary test.

In addition, non-production samples will be used to test the following CEQ Interview topics:

- 1) TPOPS Test - A test to combine the CE Interview Survey and CPI's Point of Purchase Survey (TPOPS). This test is being undertaken in order to determine whether data collected on point of purchase can be combined with expenditure data in order to provide CPI with lower levels of nonresponse error for point of purchase data.
- 2) Expanded Validation Study - An expansion of the validation study designed to ascertain whether data is correctly reported by using validation mechanisms including debriefings, interviews with multiple CU members, and post-interview record collection. This study will also analyze financial record keeping software as a recall tool. Goals of the study include quantifying how much underreporting occurs, sources of underreporting, and ways to address the underreporting.

5. **Statistical Contacts**

The Census Bureau will collect the data. Within the Census Bureau, you may consult the following individuals and their area of expertise for further information.

Sample Design:	Stephen Ash	(301) 763-1974
Data Collection:	Howard McGowan	(301) 763-5342