Survey Methods and Reliability of Data

INTRODUCTION

This appendix describes the data sources, sample design, and estimation procedures used to develop quarterly estimates of expenditures for the improvement and repairs to residential properties. This description refers to the revised survey methods effective with fourth-quarter 1984 data. A description of the earlier methods appears in Construction Reports C50-84-A, issued April 1985.

SOURCES OF DATA

The data presented in these tables are compiled from two sources:

1. Household survey of a sample of households.
2. Mail survey of owners of a sample of rental or vacant properties.

Household Survey

Description of survey.

Data based on personal interviews are obtained from household members as part of the Consumer Expenditure (CE) Surveys conducted by the U.S. Census Bureau for the Bureau of Labor Statistics (BLS). The CE surveys are designed to collect data on major items of consumer expense, household characteristics, and income. The expenditures covered by the survey are those that respondents can be expected to recall fairly accurately for three months or less, including expenditures for maintenance and repairs and improvement of properties. Each sample household is interviewed once per quarter for five consecutive quarters.

For the initial interview, information is collected on demographic and family characteristics and on the inventory of major durable goods of each consumer unit. Construction expenditure information is also collected in this interview, using a 1-month recall, and is used solely for bounding purposes: that is, to prevent the reporting of expenditures outside the reference period in subsequent interviews.

The second through fifth interviews use uniform questionnaires to collect expenditures for the previous months and the current month to date. Six months of data collection are required to account for all the expenditures for a quarter.

Households that move from their sample address between interviews are dropped from the survey. New households that move into the sample address are screened for eligibility and included in the survey, if found qualified.

Sample design.

The sample for the CE survey is a national probability sample of households designed to be representative of the total U.S. civilian noninstitutional population. The U.S. Census Bureau selected a sample of primary sampling units (PSUs) that consist of counties (or parts thereof), groups of counties, or independent cities. Within these PSUs, the Census Bureau selects a sample of addresses from the 2000 census address file. Excluding the bounding interviews and nonresponse (including vacancies), about 7,500 households are targeted for interviews per quarter.

Estimation and data adjustment procedures.

Each sample household included in the survey represents a given number of households in the United States. This number is the weight of the household.

The following components determine the weight for each interviewed household for purposes of estimating

expenditures for improvements and repairs:

1. The base weight, the inverse of the probability of selecting the sample household.
2. A weighting control factor to adjust for subsampling households in the field.
3. A noninterview adjustment factor that adjusts for interviews that could not be collected from occupied housing units because of refusals or because no one was home. The adjustment is performed as a function of region, tenure (owner or renter), family size, and race.
4. A factor that adjusts the weight to independently derived census population controls.

Mail Survey

Description of the survey.

Nonresident owners of rental or vacant properties with one to four housing units and owners of rental or vacant properties containing five housing units or more, as identified in the CE household survey, are mailed a questionnaire to report detailed maintenance and repairs and improvement expenditures for their entire property. The monthly sample size is about 4,000 owners.

Sample design.

The sample consists of owners of the properties identified in the household survey as being one to four unit properties with no resident owner and all properties (excluding owner occupied condominiums) with five housing units or more. The probability of selection of a property is proportional to the number of housing units in the property.

Estimation and data adjustment procedures.

The data collected are adjusted for unreturned or unusable forms by region and metropolitan area status. The weights are adjusted so that sample counts of renter occupied and vacant housing units agree with independently derived controls from the Current Population Survey.

RELIABILITY OF DATA

The statistics in these tables are based on sample surveys and may differ from statistics that would have been obtained from a complete census using the same forms and procedures. An estimate based on a sample survey is subject to both sampling error and nonsampling error. The reliability of a survey result is determined by the joint effect of these errors.

Measures of sampling errors.

Sampling error reflects the fact that only a particular sample was surveyed rather than the entire population. The selected sample for the CE survey is one of a large number of probability samples that, by chance, might have been selected under the same specifications. Estimates derived from the different samples would differ from each other. The standard error, or sampling error, of a survey estimate is a measure of the variation among the estimates from all possible samples and, thus, is a measure of the precision with which an estimate from a particular sample approximates the average from all possible samples.

Estimates of standard errors have been computed from the sample data for statistics in these tables. They are presented in the tables in the form of relative standard errors. The relative standard error equals the standard error divided by the estimated value to which it refers.

The sample estimate and an estimate of its standard error allow us to construct interval estimates with prescribed confidence that the interval includes the average result of all possible samples with the same size and design. A 90-percent confidence interval is defined to be from 1.6 standard errors below the estimate to 1.6 standard errors

above the estimate. If all possible samples were selected and surveyed under essentially the same conditions and all the respective 90-percent confidence intervals were generated, then approximately 90-percent of the intervals would include the average value of all sample estimates and approximately 10-percent would not include this estimate. For example, suppose these tables show that residential property owners spent $38.6 billion for improvements and repairs in a particular quarter and the relative standard error of this estimate is 8 percent. Multiplying $38.6 billion by .08, we obtain $3.1 billion as the standard error. To obtain a 90-percent confidence interval, multiply $3.1 billion by 1.6, yielding limits of $33.6 billion and $43.6 billion ($38.6 billion plus or minus $5.0 billion).

The sampling errors of some estimates are too great to allow meaningful comparisons among these estimates. The sampling errors should be regarded as orders of magnitude rather than absolute measurements.

Nonsampling errors and other limitations.

As calculated for these tables, the estimated relative standard errors measure certain nonsampling errors, but do not measure any systematic biases in the data. Bias is the difference, averaged over all possible samples with the same size and design, between the estimates and the true value being estimated. Nonsampling errors can be attributed to many sources: inability to obtain information about all cases in the sample; definitional difficulties; differences in interpretation of questions; inability or unwillingness of respondents to provide correct information; and errors made in processing the data.

These nonsampling errors also occur in complete censuses. Although no direct measurements of the biases have been obtained, it is believed that most of the important response and operational errors were detected in the course of reviewing the data for reasonableness and consistency. Potential sources of bias include:

1. Nonresponse resulting from interview refusal or because no one was home.
2. Undercoverage in the sampling list with respect to demographic and family characteristics of sample units.
3. Memory failure, including failure to remember exact values, and the reporting of information for an earlier or later time period than when it actually occurred (telescoping).
4. Problems in classifying the types of jobs performed.
5. Adjustment for extreme values (‘‘outliers’’) where an unusually large weighted value is contributed by a single job.