Appendix A. Sample Design

INTRODUCTION

In the CBECS, the individual building is the basic sample unit. The sample design for the 1992 CBECS was based on the 1986 CBECS sample. For the 1992 survey, 10,171 buildings were selected for inclusion in the sample. A total of 7,699 sample buildings were selected by use of multistage area probability methods. A supplementary sample of 2,472 buildings was obtained by sampling from lists of large and specialized buildings. All of the buildings sampled for the CBECS were selected to complete the Census Supplement. Of the 10,171 buildings selected, 2,889 were dropped when determined ineligible for the CBECS interview. The data shown in these tables are based on the remaining 7,282 buildings. The following discussion of the sample design was extracted from a description of the CBECS survey found in appendix B of CBECS Characteristics.

Multistage Area Probability Sample

The area component of the 1992 CBECS sample used a four-stage cluster sampling design. In the first stage, 129 primary sampling units (PSU's) were selected. A PSU typically consists of one or more contiguous counties, such as a metropolitan area with surrounding suburban counties, or a set of one or more rural counties. To prepare for the first stage sample, the approximately 3,100 counties and independent cities of the United States were grouped into 1,799 PSU's. PSU's with similar characteristics were grouped to form 129 strata. Characteristics used to define the strata were Census division, Metropolitan Statistical Area (MSA) or non-MSA status, the predominant residential heating fuel in 1980, and climate zone. Within each stratum, one PSU was selected with probability proportional to its 1980 Census population.

Probability-proportional-to-size (PPS) sampling is commonly used to take advantage of knowledge about the sample units, that is, knowledge about a measure of size (MOS) such as population, to improve the reliability of survey estimates. For quantities roughly proportional to these MOS's, estimates based on PPS sampling have lower variances than estimates based on equal-probability sampling. The 1980 population of a PSU was a useful MOS because of its relationship with commercial activity and energy consumption.

Thirty-two PSU's had populations large enough for each of these PSU's to formed a stratum by itself, so that each was selected with certainty. For the noncertainty PSU's, the Keyfitz method was used to assign selection probabilities. This method enhanced the probability of inclusion of specific PSU's that had been selected for DOE's previous Residential Energy Consumption Survey (RECS), while ensuring that the current RECS selection probabilities were still proportional to 1980 population levels. Controlled selection was used to improve the geographic coverage of the sample by maximizing the number of different States represented by the sampled PSU's. Finally, 10 non-MSA PSU's were randomly deleted from the initial sample of PSU's to reduce survey costs.

To form second-stage sampling units for CBECS, each sampled PSU was divided into areas corresponding to 5-digit ZIP codes. ZIP codes covering small areas or representing individual buildings or post office boxes were grouped together with larger area ZIP codes. All second- stage sample units are referred to as ZIP "groups." A total of about 3,900 ZIP groups were formed within the sampled PSU's. Of these, 444 were selected, using probabilities proportional to a second-stage MOS. This MOS, designed to reflect the level of commercial activity, was the estimated number of buildings in the ZIP group. It was computed for each ZIP group using

employment data from the Bureau of the Census' 1983 County Business Patterns (CBP) reports, and employee occupancy rates in different building types obtained from the 1979 CBECS.

The third-stage sampling unit was the segment, which was a geographically compact area containing roughly 100 nonresidential buildings. Sampled ZIP groups were divided into segments based on field maps and rough counts of nonresidential buildings. A total of 509 segments were selected from within sampled ZIP groups. If the field mapping and counting procedures were performed in all PSU's and ZIP groups nationwide, approximately 43,260 potential segments would result. Thus, the 509 segments actually selected represented a sampling rate of roughly 1 in 85 segments nationwide. Within PSU's and ZIP groups, the segments were selected such that 509 of the 43,260 potential segments nationwide were sampled with equal overall probabilities. However, due to the subsampling of PSU's mentioned earlier, segments in the non-MSA PSU's in the 119 PSU's designated for the 1992 CBECS had overall probabilities of selection equal to approximately three-fourths of the probabilities of selection of segments in the MSA PSU's.

Once segments were selected, preparations were made for the fourth stage of sampling, selecting commercial buildings from within segments. With a few exceptions, a building is defined as a structure totally enclosed by walls extending from the foundation to the roof. A commercial building is one that houses some type of commercial activity. Since the 1992 CBECS is a longitudinal revisit of the 1986 sample, the 1992 sample deliberately maximized overlap with the earlier sample. That is, the buildings selected in 1986 were reselected in 1992, with the exception of the 1986 buildings that were in the 10 PSU's that were dropped from the 1989 and 1992 surveys. These 1986 buildings in the dropped PSU's were excluded from the 1992 building selection. In 1986, field workers canvassed on foot each sampled segment, identifying and listing the addresses of all commercial buildings. Field workers also estimated the square footage and apparent principal usage of listed building. This information was subsequently used to assign buildings to strata for sampling.

Buildings were sampled within size/usage strata with equal probability. However, sampling fractions varied between strata so that strata containing large buildings were sampled more intensively than strata containing small buildings. For example, while the stratum of office buildings with less than 10,000 square feet was sampled at an overall rate of only 1 in 1,400, the stratum of office buildings with 50,000 square feet or more was sampled at a rate of 1 in 204. This stratified sampling is similar to PPS sampling in that each uses MOS's (but in a different way) to increase the reliability of estimates of square footage and energy consumption.

Approximately 16 buildings were sampled from each segment. If during the interview a sample selection turned out to be a facility (for example, a campus or complex) of two or three buildings rather than a single building, all buildings in the facility were taken into the sample. Facilities of four or more buildings were subsampled. A final total of 7,699 buildings were selected into the multistage area probability sample.

Supplementary Sample from Lists of Large and Specialized Buildings

To ensure adequate coverage of buildings that were significant energy users, the multistage area probability sample was supplemented within each selected PSU by a sample from a list of "large" buildings or facilities. In addition, to improve the precision of energy consumption estimates for certain types of buildings, a supplementary sample was also drawn from seven lists of special buildings.

In PSU's that were MSA's, the list of large buildings contained buildings with 250,000 or more square feet of enclosed floor space. In the non-MSA PSU's, this list contained buildings of 100,000 square feet or more. The list was compiled through inquiries with Chambers of Commerce and other local sources, and special directories.

The seven lists of specialized buildings were limited to certain types of buildings or facilities with 50,000 square feet or more. These lists were as follows:

1. Hospitals
2. Colleges and universities

3. Elementary and secondary schools
4. Post offices
5. Federal Government buildings
6. Reports for "small" new construction projects (50,000 to 250,000 square feet) which are available from the F.W. Dodge Division of the McGraw-Hill Information Systems Company
7. F.W. Dodge reports for "large" new construction projects (over 250,000 square feet)

These lists of specialized buildings were used for three reasons. First, they contained many large buildings and, thus, helped ensure accurate coverage of significant energy users. Second, they ensured good coverage for certain building types that are distinguished separately in CBECS reports, such as "health care" and "education." Third, they compensated for inadequacies in the MOS's developed for ZIP groups using the 1983 CBP reports. The CBP reports do not cover employees exempt from the Social Security System, such as the majority of the Federal workforce. The weighting procedure used for the final sample does not require that the supplemental lists be comprehensive to produce unbiased estimates. However, the more complete these lists are, the more efficient the sample design.

The lists within each sampled PSU were stratified by building size and general usage, and buildings were sampled with equal probability within strata. (In many cases, building size in square feet was estimated from available data such as the number of beds for hospitals, or the number of students for education buildings.) As in the area sample, strata containing large buildings were sampled more intensively than strata of small buildings. Also, as with the area probability sample, if a selected unit turned out to be a facility with three or fewer buildings, all were taken into the sample. Otherwise, the facility was subsampled.

The eight lists (large building list and seven specialized building lists) were sampled independently. The problem of overlap was handled by unduplicating the large buildings list to the extent possible, and by using a "priorities" approach. The priorities of the lists, in descending order, were as follows:

1. Hospitals
2. Colleges and universities
3. Elementary and secondary schools
4. Post offices
5. Large buildings list
6. Federal Government buildings
7. Dodge reports for "large" new construction projects (over 250,000 square feet)
8. Dodge reports for "small" new construction projects (50,000 to 250,000 square feet)

For example, if a given building was sampled from the hospitals list, then its selection from another list was disregarded.

For the Dodge reports on large projects (over 250,000 feet), a complete list of projects in each sampled PSU was obtained, and a sample was drawn from that list. Thus, it was possible to determine if a building sampled from some other source was also included in this Dodge list. For small Dodge projects (between 50,000 and 250,000 square feet) only a sample was obtained. Therefore, there was no way to verify whether a building that, by definition, should have been covered by this list was in fact included in the list from which that sample was drawn. For this reason, this "conceptual list" was given lowest priority.

There was also a problem of overlap between the supplemental list sample and the multistage area probability sample. Computation of joint probabilities of selection would be somewhat intractable in the complex design. Instead, a less efficient, but unbiased, procedure was adopted where buildings were made self-representing if they were sampled from an area segment and also appeared on one of the list frames. A new building sampled from an update segment of the area sample and between 50,000 and 250,000 square feet in size was assumed to appear on the (unverifiable) Dodge list for that size range. Smaller new buildings were assumed not to appear on Dodge lists, and larger ones were checked against the complete lists that were obtained for that size range.

In the 1992 CBECS, in addition to the regular supplemental list sample of Dodge reports of New Construction, another sample of approximately 150 buildings was selected from the Dodge list of large (250,000 or more square feet) office buildings. This sample was limited to buildings with a construction start date after February 1, 1989, and was included to permit special study of energy conservation issues in office buildings. The final weights for the sample were adjusted to compensate for this oversampling. The overall sampling rates for the Dodge lists were the same as for other list samples except for the newer large office buildings. The sampling rates by size class for the newer large office buildings were as follows:

Size of Office Building	Sampling Rate
100,000 to 249,999 square feet	.208333
250,000 to 399,999 square feet	.75
400,000 square feet or more	1.0

These rates achieved the desired supplemental list sample of newer large office buildings.

A total of 1,871 list entries were sampled. Because some entries were multibuilding facilities, the final list sample comprised 2,472 individual buildings.