

Surveying Parts to Construct the Whole: Sampling and Estimation Issues

Eugene M. Burns, Jay K. Olsen, and Katie Joseph
Energy Information Administration, EI-63, Washington, DC 20585

Abstract

One challenge faced by the Commercial Buildings Energy Consumption Survey (CBECS) has been that of surveying strip shopping malls. These malls consist of unrelated establishments sharing a common building. Respondents knowledgeable about energy and related characteristics of the entire mall are difficult to find. This presentation describes the multilevel data collection strategy adopted by the 2007 CBECS. Interviewers compiled rosters of mall tenants, which were used to draw a PPS systematic sample of establishments. Interviews were conducted with the mall manager, as well as with the selected establishments. Estimation involves developing a strategy for combining these different pieces of information to reconstruct the single mall building. Auxiliary information, from the roster and from other malls, assisted in the reconstruction; inconsistency and nonresponse were problems.

Key Words: establishment survey methodology, PPS systematic sampling, auxiliary information

1. Background

The Commercial Buildings Energy Consumption Survey (CBECS) is a national survey of commercial buildings conducted by the Energy Information Administration. CBECS collects information on the stock of U.S. commercial buildings, their energy-related building characteristics, and their energy consumption and expenditures. One challenge faced by the CBECS has been that of obtaining complete building characteristics and energy consumption data for strip shopping centers.

Strip shopping centers are more difficult to survey than other types of buildings. Respondents knowledgeable about the entire building, including structural characteristics, equipment, operating hours, energy use, and energy cost, are difficult to find. While the owner or manager of a strip shopping center can answer questions about the overall structure, the individual store managers can best provide information about items such as operating hours, number of employees, equipment, and energy bills.

To deal with this data collection problem, EIA adopted a multipronged data collection strategy for the 2007 CBECS (which is currently being fielded):

- Field interviewers compiled rosters of mall tenants, which were used to draw a PPS systematic sample of establishments.
- Field interviewers conducted interviews with the selected establishments, as well as with the mall manager.

By focusing on the building (the strip shopping center as a whole), using separate mall manager and establishment data collection instruments merely formalizes the common business survey practice of having multiple respondents. A built-in overlap between managers and establishments is designed to maximize the chance of obtaining a knowledgeable respondent for the required items.

The challenge addressed in this presentation is that of developing a strategy for combining these different pieces of information together to reconstruct the single mall building.

2. Sampling

Sampling was required to control fieldwork costs; otherwise the strip mall sample could be as large as that for all other building types combined. Data from the 2003 CBECS on the number of establishments per strip shopping center was used to estimate sample sizes. While the number of establishments within a strip shopping center can vary from two up to forty or more, the median number of establishments per strip center was nine.

The sample size, n , at any particular strip center would depend on the number of establishments in the center.

- For centers with two establishments, take both.
- If there are three to nine establishments, take two.
- If there are ten or more establishments, take three.

Assuming that the distribution of number of establishments at strip shopping centers in 2007 is similar to that obtained by the 2003 CBECS, the 2007 CBECS would obtain a sample of two establishments at 188 centers and a sample of three establishments at 161 centers, for an average of 2.46 establishments per center.

The measures of size (MOS) to be used in sampling should be proportional to each establishment’s energy consumption. The true energy consumption may be unknown beforehand, but a MOS can be assigned to establishments based on a combination of interviewer observation and prior 2003 CBECS data.

While compiling the roster of establishments in a strip center, the field interviewer was asked to classify each establishment into one of ten establishment types and one of two relative sizes.

Table 1: Establishment Type and Relative Establishment Size Classifications Used by Field Interviewers	
<i>Types of Establishments</i>	<i>Relative Establishment Sizes</i>
<ul style="list-style-type: none"> • Retail shop (includes movie rentals) • Food sales • Food service • Dry cleaner/Laundromat • Office/Financial • Medical/Dental • Movie theatre/Cinema • Beauty salon/Tanning salon • Vacant • Other 	<ul style="list-style-type: none"> • Average (for this shopping center) • Large (more than twice the average size)

To estimate establishment MOS,

- The establishment types were matched to corresponding building activities.
- Building sizes were used as surrogates for relative establishment sizes,
 - Buildings of 1,001 to 5,000 square feet were used for average size, and
 - Buildings of 5,001 to 25,000 square feet were used for large size.
- The 2003 CBECS data were used to estimate energy consumption for all fuels, and the resulting estimates were rounded to units of 50 million Btu.

To sample establishments, a systematic PPS sample selection algorithm was adopted. This strategy was originally proposed by Madow (Cochran 1977, pp. 265-266). Let

M_i' = measure of size for the i^{th} establishment and

$$M_0' = \sum_i M_i'$$

To sample n establishments,

- Form a column of cumulative totals, T_i , of nM_i' , and assign a range of size nM_i' to establishment i . When forming the column, randomize the order of establishments.
- For each establishment, the beginning of the assigned range should be the cumulative total of nM_j' for all previously placed establishments. The end of the assigned range is the cumulative total plus nM_i' for the current establishment.
- Finally, draw a random number, r , between 1 and M_0' and select the n establishments for which the assigned ranges include the numbers $r, r + M_0', \dots, r + (n - 1)M_0'$.

3. Estimation

The overall goal of the strip center estimation is to obtain better building-level information for the strip center as a whole. Obtaining establishment-level estimates, mainly for use in developing future samples, is a lower priority.

The data available to construct building-level estimates include:

- Establishment-level data collected from the sampled establishments,
- Building-level data collected from the mall managers,
- The mall rosters compiled by interviewers for sampling purposes, which list
 - the activity type,
 - the size (floorspace) category, and
 - the posted operating hours
 of all mall establishments, both sampled and non-sampled.

In addition, the data obtained by interviewing mall managers and establishments at other strip shopping centers is also available to help with the reconstruction. Ideally, the reconstructions would make the best use all available information to produce building-level estimates, and to produce those estimates in a timely fashion.

In estimating the data values for the building as a whole, values are needed both for the nonsampled establishments and for nonrespondents among the sampled establishments. The last time CBECS was fielded, in 2003, CBECS had a response rate above 80 percent. Assuming that the 2007 CBECS achieves an 80 percent response from mall managers and establishments (and that managers and establishments respond independently), a shopping center with nine or fewer establishments would be expected to have an a full response rate of about 50 percent ($0.512 = 0.8^3$) and a larger center would be expected to have a rate of only 40 percent ($0.4096 = 0.8^4$). Nonresponding sampled establishments can be treated along with nonsampled establishments. However, the amount of complete and partial response will be noted in the survey documentation.

The following general procedures are proposed for estimation, and will be tested when the final 2007 CBECS data file is ready:

- Initialize the building-level record, using the data reported by the mall manager.
- Define establishment activity profiles, using data reported by establishments at all strip shopping centers.
 - Activity profiling is motivated by the working assumption that establishments will be more similar to same-activity establishments at other centers than to other establishments at the same strip center.
 - The similarity will be greater for interior (within the establishment's space) characteristics than for externals.
 - Data items included in the profile will include: uses of energy (air-conditioning, water heating, cooking, etc.), energy-using equipment (lighting, refrigeration, electronics, etc.), number of employees (per 1,000 square feet), energy consumption (per square foot), and other items as appropriate.
- Reconcile floorspace estimates from the mall manager (building-level), establishments (establishment-level), and the roster (establishment-level, but relative).
 - Floorspace estimates will be used to assist in combining establishment information into building-level data estimates.

- In the ideal case, all three estimates will be consistent, and the difference between the floorspace reported by the mall manager and the sampled establishments can be prorated among the nonsampled and nonresponding establishments.
- Combine establishment-level estimates into building-level data. The method used will vary by the type of data.
 - For continuous items, such as the number of employees, the building-level value is the sum of the establishment-level values.
 - For binary items (e.g., whether energy is used for cooking), if any establishment indicates use, then the building uses.
 - For percents (e.g., percent lit), first convert to a quantitative item (e.g., floorspace lit), then sum and recalculate the percent.
 - Items not fitting the above types will be handled individually.

Throughout the process, the CBECS staff members need to learn from the data and document their findings, so that the 2007 CBECS experience can be used to direct future surveys of strip shopping centers.

References

Cochran, W.G. 1977. *Sampling Techniques*, 3rd ed. New York: Wiley.