Independent Statistics & Analysis
U.S. Energy Information Administration

Supporting Statement for Commercial Buildings Energy Consumption Survey (CBECS)

Part B: Collections of Information Employing Statistical Methods

Forms EIA-871 A, C, D, E, F, I, J, Commercial Buildings Energy Consumption Survey

OMB No. 1905-0145

December 2018

Independent Statistics & Analysis

www.eia.gov

U.S. Department of Energy

Washington, DC 20585

Table of Contents

B.1. Respondent Universe

CBECS provides estimates on energy-related building characteristics and energy usage data (consumption and expenditures) for U.S. commercial buildings with over 1,000 square feet of floor space. Commercial buildings include all buildings in which at least half of the floor space is used for a purpose that is not residential, industrial, or agricultural. By this definition, CBECS includes building types that might not traditionally be considered commercial, such as schools, hospitals, correctional institutions, and buildings used for religious worship, in addition to traditional commercial buildings such as stores, restaurants, warehouses, and offices buildings. The 2012 CBECS estimates that there are about 5.6 million commercial buildings in the United States.

B.2. Statistical Methods

Multi-stage Area Probability Frame

The CBECS sample design is a multi-stage area probability sample, supplemented by list frames of large buildings over 200,000 square feet. The general approach of the multi-stage area probability sample design is to sample and subsample successively smaller geographic areas until it is feasible to list all commercial buildings in the selected areas.

Primary Sampling Units (PSUs): The first stage of the multi-stage area probability sample consists of primary sampling units (PSUs), which are counties or groups of contiguous counties. The country is divided up into 687 total PSUs.

The PSUs in the frame are grouped into 65 strata that are homogeneous based on geographic characteristics such as similar climate and topography, but are not necessarily geographically contiguous.

Twenty of the 65 strata contained only one PSU and are sampled with certainty. Certainty PSUs are in major cities with substantial commercial activity. The remaining 45 non-certainty strata contain multiple PSUs. The PSUs are selected using probabilities proportional to a measure of size that relates to commercial activity. A measure of commercial business activity within each PSU is estimated using employee counts from the Census Bureau's County Business Patterns and employee per building ratio estimates from previous CBECS surveys. This measure of business activity is used as a measure of size to statistically select PSUs with probability proportionate to size (PPS) sampling. Of the 687 PSUs in the frame, 151 are selected for the sample PSUs – 20 selected from the certainty strata, 3 PSUs per strata in 43 of the 45 non-certainty strata, and 1 PSU each in the Alaska and Hawaii non-certainty strata.

Secondary Sampling Units (SSUs): The next step in the multi-stage area sampling procedure consists of selecting subareas within each sampled PSU. Within the 151 selected PSUs, 8,800 total SSUs are constructed by combining 23,117 census tracts. To construct the SSUs, first census tract tabulations were produced for the number of commercial buildings by building size and use classes from the Common Premises Location (CPL) database purchased from Dun & Bradstreet. A measure of size for each tract was calculated using as the expected number of interviews in the tract, which was obtained by applying the preliminary CBECS national building sampling rates to the tract-level CPL building counts

by size and use strata. Small tracts were combined to provide adequate numbers of expected interviews in each SSU. The final SSU measure of size was computed by aggregating the tract-level measures of size across all tracts defining a given SSU. A stratified PPS sample of SSUs was selected, with building size and use categories as strata. Of the 8,800 SSUs in the selected PSUs, 764 SSUs are sampled. Given the 151 selected PSUs, a targeted number of completed building interviews of 8,000 for the 2018 CBECS, a desire for cluster sample sizes of 10.5 completed building interviews per SSU (similar to 2012), and similar design effects due to clustering between the 2012 and 2018 CBECS, the number of SSUs to select for the 2018 CBECS was determined to be 764.

Buildings: The final step in the multi-stage area probability sample consists of selecting buildings within each sampled SSU. A combination of field listing and virtual listing is used to produce a list of all commercial buildings in the 764 sampled SSUs. For field listed segments, trained field staff walk every block in the segment (or drive in more suburban or rural areas) and record all commercial buildings within the designated segment boundaries. For virtually listed segments, buildings are listed using the web-based Virtual Listing System, which incorporates GIS, satellite imagery, points of interest, and other commercial databases. The Virtual Listing System allows the listing of buildings from computers in a centralized location, saving time and money compared to traditional field listing. Both listing methods produce the address, building size category, and building use category for each listed building.

Buildings are selected for inclusion in the sample from the field and virtual listings using a systematic PPS selection procedure. The overall building sample size is allocated to building size and use strata in such a way that the expected precision for the total Btu estimate is maximized. The optimal building sampling rates are calculated using the 2012 CBECS energy consumption data and the building population estimates. Using the 2012 CBECS data, the unit variances for total Btu among buildings within each of the use and size classes are computed. The building population size estimates and estimates of standard deviation of total Btu within each building size and use stratum determine the optimal sampling rates using a Neyman allocation.

List Frame of Large Buildings

It is desirable to sample large buildings at a higher rate than small buildings because of the relatively large amount of energy that is consumed in large buildings. However, the area probability sampling procedure cannot provide an optimally efficient mix of large and small buildings in order to guarantee that a sufficient number of very large buildings would be available for sampling.

To compensate for this inefficiency of the area sample, special lists of government buildings, college and university buildings, hospitals, airports, and other large buildings are used in the sample PSUs. To complete the full 2018 sample, buildings that are at least 200,000 square feet will be sampled from these lists at the PSU level and then combined with those selected by the area sampling procedure after checking the frames for duplication.

The list frame consists of: U.S. Government buildings that are at least 200,000 square feet (from the General Services Administration Real Property Inventory); four-year colleges and universities that are estimated to have at least 1,000,000 square feet of floorspace (from the Integrated Postsecondary Education Data System); hospitals that are estimated to have at least 200,000 square feet of floorspace

(purchased from IQVIA); a list of airports (from the Federal Aviation Administration) over 200,000 square feet; and a list of large buildings over 200,000 square feet (from the CPL file purchased from Dun & Bradstreet).

De-duplicating the frames

Because there are multiple frames, the area frame and 5 list frame components, it is possible that a building could be selected from more than one frame. The process of de-duplication ensures that sampled buildings have only one chance of selection. To control the overlap and the selection probabilities of buildings, an established hierarchy of frames is used, similar to prior rounds of CBECS. The frame hierarchy, from highest to lowest, is as follows: airport list, U.S. Government building list, college/university list, hospital list, CPL large building list, and the area frame. A selected building will be eligible for the sample only through the highest sampling frame in the hierarchy that contains the building. After selecting the samples from each frame at desired rates, the de-duplication process will identify any sample building that is on a higher frame. If the sample building is found to be in any higher frame it will be dropped from the sample.

Sample size

Approximately 14,830 buildings (including mall buildings) will be selected for the CBECS sample. Assuming an 83 percent in-scope rate and a 65 percent response rate, the 14,830 buildings selected will yield the target of 8,000 completed building interviews. The in-scope and response rate assumptions are based on data from previous CBECS cycles.

This overall sample size will target total energy consumption estimates with relative standard errors (RSEs) at the national level of 5% or less and at the census region level of 10% or less.

Imputation, weighting, and estimation procedures

Item nonresponse is filled in with imputation. Hot deck imputation is used to impute building characteristics variables. Imputation cells are formed using frame variables and survey questionnaire variables that are related to the variable being imputed. An engineering based regression model is used to impute energy consumption and expenditure variables.

Estimates of population parameters will be calculated with survey weights. For each sampled building, the base weight, which is the inverse of the probability of selecting the building, is first calculated. After excluding out of scope buildings, the base weights will be adjusted to account for nonresponse. Nonresponse adjustment cells are created with the goal of independence of nonresponse with survey variables of interest (i.e. building characteristics of interest, energy consumption and expenditures), which yield unbiased estimates. There are several methods of forming nonresponse adjustment cells to achieve this result. CBECS uses Chi-Square Automatic Interaction Detector, a decision tree method, to form the cells, using information available on the frame.

After forming the nonresponse adjustment cells, a nonresponse adjustment factor is calculated for each adjustment cell. The building nonresponse adjustment factor for the h-th nonresponse adjustment cell, δ_h, is:

$$\delta_h = \frac{\sum_{k \in A_h} W_{hk} + \sum_{k \in B_h} W_{hk} + \sum_{k \in C_h} W_{hk}}{\sum_{k \in A_h} W_{hk} + \sum_{k \in B_h} W_{hk}}$$

where,

W_{hk} = is the sampling weight (after adjusting for duplication) for the k-th building in the h-th nonresponse adjustment cell;

A_h = is the collection of buildings that were respondent and eligible in the h-th nonresponse adjustment cell;

B_h = is the collection of buildings that were declared ineligible in the h-th nonresponse adjustment cell; and

C_h = is the collection of buildings that were in-scope nonrespondents in the h-th nonresponse adjustment cell.

The nonresponse adjusted weight for respondent buildings is:

$$W_{hk}^{(ND)} = W_{hk} \delta_h \quad .$$

The resulting nonresponse adjusted weights represent the entire commercial building population.

Variance estimates and RSE estimates will be calculated by means of the Jackknife replication technique. The RSEs for the statistics in the cells of most CBECS report tables will be expressed in separate RSE tables.

B.3. Maximizing Response Rates

Over the years, the response rates to the CBECS have been quite high for voluntary reporting on Forms EIA-871 A, I, and J. However response rates have declined over 20 years for this survey. The response rates on Forms EIA-871 A, I, and J declined from 91% in 1992 to 82% in 2003, 81% in 2007, and 70% in 2012. Efforts will be made in 2018 in an attempt to slow that trend. The following pre-collection procedures will be used to maximize response rates:

- The survey contractor conducted a field test of contact procedures, as described further in *Section B.4.* The lessons learned from that test are being used to inform strategies to identify respondents, encourage survey completion on the web, and avoid refusals.
- A dedicated interviewer will be assigned to each sampled building, and will continue to monitor the case even if a respondent indicates that they will complete the interview via

web. The interviewer will follow-up with the respondent if the respondent does not complete the web survey within a given time.
- EIA has developed materials that emphasize the importance of the content of the survey, both for the interviewers and for the survey respondents.
- Letters signed by a Department of Energy official will be mailed to the respondent in advance of the first contact to inform them about the survey.
- The identified respondent will be provided with background materials consisting of information about the survey, a list of items for which data will be collected, a list of businesses and organizations who support the CBECS, and other information that highlights the importance of the CBECS and why their participation is vital.
- The respondent will be able to select their mode of completion, either in-person or over the phone with an interviewer, or on a self-administered web instrument. Debriefing with interviewers from the 2012 CBECS indicated that some respondents asked whether a web option was available.
- The survey contractor will provide a toll-free Help Center number for respondents to call with any questions or concerns. Respondents will also be provided with the phone number and email address of the survey manager at EIA.
- Nonresponding building owners and managers will receive reminder letters and telephone calls.

The response rates for the mandatory Energy Suppliers Survey (Forms EIA-871C, D, E, and F) are high. In the 2012 CBECS, 97% of suppliers provided data for at least one building. Most suppliers are asked to report data for more than one building. At the building-level, the response rate was 90%.

B.4. Test Procedures and Form Consultations

Cognitive pretesting of survey questions

Under *EIA-882T: Generic Clearance of Questionnaire Testing, Evaluating, and Research, OMB 1905-0186 (Expiration 3/31/2019)*, from January 9, 2017 to February 20, 2018, EIA's survey contractor conducted cognitive pretesting of survey questions in three CBECS content areas: building square footage, HVAC systems, and data centers and servers in buildings. This pretesting method served two main purposes: identify the variety of difficulties respondents encounter in understanding and answering survey questions, and exploring whether respondents interpret the question as intended. The survey contractor recruited 20 building respondents from the Washington DC metro area, which represented building types and sizes where previous data review had shown issues with the items in question. Each building respondent took part in an approximate one-hour interview at their place of work. The survey contractor summarized the results for EIA and respondent feedback was considered as the questions were revised. The feedback from the respondents was especially useful for updating the definitions of heating and cooling equipment; the updated examples should better align with respondents understanding of their energy systems.

Usability testing of web instrument

Because CBECS will be offered as a self-administered web survey for the first time, significant design effort has gone into the web interface. A well-designed, intuitive interface decreases interview time and reduces cognitive burden on the respondent. EIA's survey contractor developed a web prototype of a

subset of questions and tested its usability in a laboratory setting. Eight participants were recruited for the usability testing, which was held the week of September 17, 2018. Participants followed prescribed vignettes to ensure certain response paths were followed to test questions. In the interview, the participant was instructed to read the vignette aloud and then fill out the web instrument using a laptop computer. After the interview, the interviewer followed a testing protocol and asked the participant about their thoughts and impressions about the design of the instrument. Between going through the vignette and answering the follow-up questions, participants spent about 45 minutes with the cognitive interviewer. EIA and the survey contractor observed all of the interviews and used the respondent's feedback to improve the web survey design.

Contact protocol testing

Under *EIA-882T: Generic Clearance of Questionnaire Testing, Evaluating, and Research, OMB 1905-0186 (Expiration 3/31/2019)*, EIA's survey contractor conducted a test to help inform the development of the contact procedures and materials to be used in the main CBECS data collection. The focus of the work was on the most effective strategies to identify the best respondent, gain cooperation, and encourage participation via the web.

Field work took place in two stages from September 24 through October 19, 2018 in Rockville, MD and Reston, VA. Four interviewers worked in pairs and contacted respondents from a list of buildings that had been compiled by the contractor for these two areas. The sample represented a full range of CBECS building sizes and activities. The interviewers visited each building and attempted to find a respondent knowledgeable about energy use in the building. They finalized 215 cases, of which 111 had completed screeners and an appropriate respondent was identified. Interviewers were supplied with sample materials such as an authorization letter and a CBECS brochure. The goal was to have each respondent access a short (approximately 10 minute) web questionnaire, which consisted of a couple questions about the size and use of the building followed by some questions about the interaction with the interviewer and the respondent's preferences in how to complete the survey. The interviewers were also instructed to fill out a form detailing their experiences with each building. The survey contractor held debriefings with the interviewers twice during the first week of interviewing and once during the second to gain a full understanding of the types of issues they were encountering in the field. Three types of situations emerged as challenges for making contact and identifying respondents: (1) gatekeepers referring interviewers to off-site respondents; (2) high security/limited access buildings; and (3) single buildings owned by more than one entity. Procedures for overcoming these challenges are being developed for the main study.

B.5. Statistical Consultations

For questions or additional information regarding the statistical methods for this survey, contact Katie Lewis, Mathematical Statistician, at (202) 586-5138, email Katie.Lewis@eia.gov. For further information on the collection and processing, contact the CBECS Survey Manager, Joelle Michaels, at (202) 586-8952, e-mail Joelle.Michaels@eia.gov.