Appendix C

Data Quality

Because the estimates in this report are based on observations from randomly chosen subsets of the entire population of commercial buildings and residential households and these observations span a 10-year period, several factors arise, which affect the quality of the data. This appendix briefly discusses nonresponse adjustment procedures, computation of relative standard errors (RSE), development of a common unit of analysis (i.e., the building), the energy price index used to maintain constant purchasing power, and the undercoverage issue associated with the early CBECS.

Nonresponse Adjustment

There are two major types of nonresponse: unit nonresponse and item nonresponse. Most unit nonresponse occurs when a respondent refuses to cooperate or is unavailable. Item nonresponse occurs when the respondent does not know or, less frequently, refuses to give the answer to a particular question or when the interviewer does not ask the question or does not record the answer during the interview. The next two sections provide details on the procedures followed to deal with each type of nonresponse.

Adjustments for Item Nonresponse

Nonresponses to several items in otherwise completed questionnaires were treated by a technique known as "hot-deck imputation." In hot-decking, when a certain response is missing for a given case, another case (called a "donor") is randomly selected to furnish its reported value for that missing item. That value is then assigned to the case with the item nonresponse (the nonrespondent, or "receiver"). To serve as a donor, a case must be similar to the nonrespondent in characteristics correlated with the missing item.

In some cases, the energy supplier did not provide the consumption or expenditures data for some or all billing periods or deliveries in the survey year, as requested in the Supplier Survey. Reasons for missing data have included energy supplier refusal; archived, lost, or destroyed billing records; and, waiver refusal on the part of the building or household respondent. In other cases, the energy supplier provided data, but either the building data were combined with those of nonsampled buildings and could not be disaggregated or the consumption or expenditures or both were not completed enough to be treated as missing.

The general approach taken to impute annual consumption and expenditures was to annualize the complete or partial sets of bills first and then to use these annualized bills in regression equations to develop imputed values for the data that were totally missing. The regression imputation approach was chosen because data from the Building or Household Survey were already available for the buildings (households) lacking energy supplier data. The first step to correcting this problem was to estimate missing consumption data on the basis of regression equations developed using building characteristics in cases where the consumption was already known. After the consumption had been imputed, missing expenditures were estimated based on the reported or imputed consumption. For a more in-depth discussion of the adjustment process for item nonresponse of consumption and expenditures, see the appendices of previous publication of the RECS or CBECS.[19]

[19]See Appendix E, "Related EIA Energy Consumption Publications."

Adjustments for Unit Nonresponse

Weight adjustment is the method used to reduce unit nonresponse bias in the CBECS survey statistics. The CBECS was designed so that survey responses can be used to estimate characteristics of the entire stock of nonresidential buildings in the contiguous United States. The method of estimation is to calculate basic sampling weights that relate the sampled buildings to the entire stock of nonresidential buildings. To reduce bias for unit nonresponse in the survey statistics, the base weights of respondent buildings are adjusted upward, so that the respondent buildings will represent not only unsampled buildings but also nonrespondent buildings. For a more in depth discussion of the adjustment process for unit nonresponse, see the appendices of previous publications of the CBECS.[20]

Similarly, the RECS also uses weight adjustment to reduce unit nonresponse bias. An additional adjustment is performed to improve the representation of the population by the RECS sample as a whole (regardless of response levels). Ratio estimation is used to adjust selected estimates of household counts to official population values. The ratio adjustment is arrived at in several stages and ultimately benchmarks the RECS estimates to the Bureau of the Census population estimates at a regional level. For a more in-depth discussion, see the appendices of previous publications of the RECS.[21] No similar adjustment is used for CBECS because no independent benchmark totals are available for the commercial buildings population.

Relative Standard Errors

For some surveys, a convenient algebraic formula for computing variances can be obtained. However, both the CBECS and the RECS use a multistage area sample design of such complexity that it is virtually impossible to construct an exact algebraic expression for estimating variances. Due to the complexity of the sample designs, the CBECS uses a jackknife replication method for variance estimation and the RECS uses the balanced half-sample replication method (also termed BRR). For more details about the jackknife replication method, see Appendix C, "Nonsampling and Sampling Errors," of the 1992 CBECS reports. For more details about BRR variance estimation method, see Appendix B, "Quality of the Data," of the 1990 RECS reports.

The relative standard error (RSE) is the square root of the mean square error, expressed as a percent of the estimate. The RSE was used for any statistical tests or confidence intervals given in the text. Estimates with RSE greater than 50 percent are withheld from the published tables due to their lack of precision.

Space limitations prevent publishing the complete set of RSE's with this report. Instead, a generalized variance technique is provided, by which the reader can compute an approximate RSE for each of the estimates in the tables. For an estimate in the i^{th} row and j^{th} column of a particular table, the approximate RSE is found by multiplying the row factor in the rightmost column of row i by the column factor given at the top of column j. See Figure C1.

[20]See Appendix E, "Related EIA Energy Consumption Publications."

[21]See Appendix E, "Related EIA Energy Consumption Publications."

Figure C1. Use of RSE Row and Column Factors

Table C1. Commercial Building Characteristics, 1989

Building Characteristics	Total Floorspace (million square feet)	Buildings (thousand)	Floorspace per Building (square feet)	RSE Row Factors
RSE Column Factors:	1.2	1.1	0.8	
All Buildings	63,184	4,528	13,955	3.4
Census Region				
Northeast	13,569	783	17,320	8.3
Midwest	15,955	1,046	15,255	7.8
South	22,039	1,847	11,932	5.9
West	11,620	851	13,651	7.3

R(South) = 5.9
C(Buildings) = 1.1
Approximate RSE(South, Buildings)
= (5.9) * (1.1) = 6.49 percent.
Approximate Standard Error(South, Buildings)
= (0.0649) * (1,847) = 119.9 thousand.

Source: Energy Information Administration, Office of Energy Markets and End Use, Energy End Use and Integrated Statistics Division, the 1989 Commercial Buildings Energy Consumption Survey.

For more details about the derivation of the row and column RSE factors, see Appendix C, "Nonsampling and Sampling Errors," of the 1992 CBECS reports.

Residential Buildings as Units of Analysis

The RECS is a survey of United States households. All of the estimates presented in the RECS reports are based on the housing unit as the primary unit of measurement. Sometimes, as in this report, the population of interest is not the housing unit but the building itself. Hence, the problem arises of recasting the estimates to reflect the new unit of analysis (i.e., the residential building).

Definitions

To demonstrate the underlying differences between a housing unit and a residential building, consider the following definitions.

> Housing Unit: A house, an apartment, a group of rooms, or a single room if it is either occupied or intended for occupancy as separate living quarters by a family, an individual, or a group of one to nine unrelated persons.

> Residential Building: A structure used primarily as a dwelling for one or more households that is totally enclosed by walls that extend from the foundation to the roof.

By this definition, a single-family attached housing unit (such as a single side of a duplex) would be considered a single building. In other words, the number of single-family attached housing units is equivalent to the number of single-family attached buildings. This one-to-one correspondence between housing units and residential buildings also exists between single-family detached housing units and mobile homes and the buildings which they represent; however, it does not exist for multi-unit apartment buildings. These are the RECS units whose data must be recast into the building frame of reference. Procedures for adjusting these data are given in the next subsection, "Derivation of Weights."

If only one housing unit is represented in a residential building, then the housing unit immediately corresponds to the CBECS definition of a building:

> Building: A structure totally enclosed by walls that extend from the foundation to the roof and intended for human access.

After the multi-unit data are adjusted to the building level, residential and commercial sector data can be compared or combined because both are based on the unit definition of building.

Two significant problems arise in deriving residential building estimates from RECS housing unit estimates: the number of buildings in the total population are likely to be undercounted and the size of multi-unit buildings is likely to be underestimated.

- Buildings are undercounted because RECS interviews are not conducted in vacant housing units. It is likely that the amount of underestimation is similar to the rate of housing unit vacancy, which the 1989 American Housing Survey estimated at about 9 percent.

- The size of the multi-unit buildings is understated because the floorspace of common areas, such as hallways, stairwells, elevators, or lobbies, is not accounted for.

We recognize these problems, but we do not attempt to account for them when deriving building estimates from housing unit estimates.

Derivation of Weights

In order to reduce bias in the survey statistics, the sample units are weighted to reflect the selection probability, P_i, and to adjust for unit nonresponse. The base weight for the i^{th} unit is:

$$w_i = \frac{1}{P_i}.$$

The adjustment for unit nonresponse is designed to spread the effects of nonresponse over the entire responding sample. It is equal to:

$$a_i = \frac{\sum_{i \subseteq S} w_i}{\sum_{i \subseteq R} w_i} ;$$

where S is the entire sample and R is the respondent subset of S. The adjusted housing unit weight is the product of the weight and the adjustment:

$$hu_{adj} = w_i \; a_i .$$

This process of weight adjustment has been used for all cycles of the RECS.

The calculations for deriving building estimates were performed in the *Household Energy Consumption and Expenditures, 1990*. In multi-unit buildings, the specific number of housing units in the building containing the i^{th} sampled housing unit, u_i, is needed to convert housing unit data to building data. Viewing $1/u_i$ as a proportion of the residential building,

$$b_i = \frac{1}{u_i},$$

which makes the adjusted building weight equivalent to:

$$bld_{adj} = w_i \; a_i \; b_i .$$

In this way, estimates based on housing units can be transformed into building estimates.

Example. Suppose the specific estimate of interest is the count of the number of housing units, HU, or buildings, BLD. Then with *n* equal to the number of sample units, the number of housing units in the population is estimated by

$$HU = \sum_{i=1}^{n} (a_i \; w_i) ,$$

and the estimated number of buildings is

$$BLD = \sum_{i=1}^{n} (a_i \; w_i \; b_i).$$

Let $sfhu_i$ be the floorspace in square feet for the i^{th} housing unit in a particular building. Because total building floorspace data are not collected, the true proportion of the building floorspace that the i^{th} housing unit occupies is unknown. Assuming that the individual unit chosen in the sample is representative of the other units in the same building, the floorspace of the building containing the i^{th} unit, $sfbld_i$, is estimated to be the product of the sampled unit's floorspace, $sfhu_i$, and the number of units in the i^{th} building, u_i. In this way, each sampled unit is a proportionally smaller representative of a larger separate single building,

$$sfhu_i = sfbld_i * b_i$$

so that,

Unless the number of housing units in a building is one, it is expected that changes in count among the categories will occur because, generally, in multi-unit buildings,

$$sfbld_i = (sfhu_i * u_i) .$$

$$\text{Category}(sfbld_i) \neq \text{Category}(sfhu_i).$$

However, the **sum** of the weighted estimates of the amount of floorspace in all sample housing units and the **sum** of all subsequent building floorspace estimates are equivalent. The estimates are equivalent because the proportion of floorspace a housing unit occupies in a building is assumed to be the reciprocal of the number of units in the building. If the true proportion were known, then the two estimates would not necessarily be equivalent.

The total floorspace in all units is

$$SFHU = \sum_{i=1}^{n} (a_i w_i * sfhu_i) ,$$

and the total floorspace in all residential buildings is

$$SFBLD = \sum_{i=1}^{n} (a_i w_i b_i * sfhu_i u_i) .$$

Rearranging and using the definition of b_i, this becomes

$$SFBLD = \sum_{i=1}^{n} (a_i w_i * sfhu_i * \frac{1}{u_i} u_i).$$

That is, the total square feet of occupied floorspace in residential buildings does indeed equal the total square feet of floorspace in housing units. Estimates of square feet per unit of analysis, however, will be different between housing units and buildings. In general, estimates of **totals** as well as any **ratio of totals to totals** (e.g., total floorspace, total consumption, consumption per square foot, etc.) will not differ between the units of analysis. However, estimates of **totals per unit of analysis** (e.g., square feet per building/household, consumption per building/household, etc.) will necessarily be different because the total number of each unit of analysis is different.

The Choice of an Energy Price Index

In this report, data are presented for consumption and expenditures for both the residential and commercial sectors with a common unit of analysis, the building. For consistency, the energy price indices that were developed as part of the *State Energy Price and Expenditure Report 1992* (SEPER 1992) can be used as the deflators in the energy expenditure tables of the report. The following section describes these energy price indices.

State Energy Price and Expenditure Report 1991 (SEPER 1992): Real Fixed-Weight Energy Price Index

In order to compare price data between years, the SEPER uses fixed-weight price indices. The nominal fixed-weight energy price index is a measure of the average price of net energy consumption in 1987, the weight-base year. The composition of net energy consumption is held constant at 1987 weights for each year. The 1987 weights consist of detailed energy source and end-use sector categories for each State and the relative distribution of net energy consumption among the various States. The real fixed-weight energy price index is then obtained by dividing the nominal fixed-weight energy price index by the gross domestic product purchases (GDP) benchmark-years-weighted price index (See Appendix D in SEPER 1992).

Using the *State Energy Price and Expenditure Data System 1991*, fixed-weight energy price indices were developed by sector for total energy and for each of the major energy sources. The following are the nominal and real fixed-weight energy price indices for the commercial and residential sectors (Tables C1, C2, and C3). The deflator that was used in this report to adjust current dollars to 1987 dollars was the gross domestic purchases (GDP) deflator. The nominal fixed-weight energy price index can be used to analyze changes in expenditures with the energy product mix benchmarked to 1987.

Table C1. Energy Price Indices, 1979 to 1991 (1987 = 100)

Year	Energy Price Indices				Gross Domestic Purchases (Benchmark Years)
	Residential		Commercial		
	Nominal	Real	Nominal	Real	
1979	62.8	94.2	67.7	92.6	66.6
1980	75.2	102.2	81.3	110.5	73.6
1981	87.5	109.1	94.3	117.6	80.2
1982	97.5	115.0	103.3	121.8	84.8
1983	103.4	117.8	106.9	121.7	87.8
1984	103.4	113.9	108.0	119.0	90.8
1985	105.1	111.9	108.7	115.8	93.9
1986	101.6	105.2	103.6	107.3	96.6
1987	100.0	100.0	100.0	100.0	100.0
1988	100.0	96.2	98.8	95.1	103.9
1989	103.7	95.5	101.6	93.5	108.6
1990	107.8	94.9	104.9	92.3	113.6
1991	109.6	93.0	106.3	90.2	117.9

Note: Calculations are based on unrounded numbers.

Sources: **Nominal Fixed-Weight Energy Price Index--**Based on net energy consumption (1987 fixed-quantity weights) and end-use price data from EIA, *State Energy Price and Expenditure Data System 1991*. **Real Fixed-Weight Price Index--**Nominal energy fixed-weight price index/Gross Domestic Purchases benchmark-years-weighted price index. **Gross Domestic Purchases, Price Index, 1987 = 100, Benchmark-Years-, 1980-1991--**U.S. Department of Commerce, Bureau of Economic Analysis, *Survey of Current Business*, April 1993, Volume 73, No. 4., p. 35, and *National Income and Product Accounts of the United States*, Volume 2, 1959-88, pp. 266-267.

Table C2. Residential Energy Price Indices for Specific Energy Sources, 1979 to 1990 (1987 = 100)

Year	Residential Energy Price Indices					
	Nominal			Real		
	Electricity	Natural Gas	Fuel Oil	Electricity	Natural Gas	Fuel Oil
1979	62.5	55.5	83.5	94.0	83.8	125.4
1980	72.4	67.8	116.1	98.2	92.0	157.5
1981	84.0	78.4	142.9	104.8	97.7	178.2
1982	93.1	94.6	138.8	109.8	111.5	163.7
1983	96.9	110.2	134.1	110.4	125.5	152.7
1984	96.5	111.1	136.4	106.3	122.4	150.2
1985	99.8	111.1	131.3	106.2	118.3	139.8
1986	99.9	105.4	105.3	103.4	109.1	109.0
1987	100.0	100.0	100.0	100.0	100.0	100.0
1988	100.4	98.9	101.0	96.6	95.2	97.2
1989	102.7	101.8	111.7	94.6	93.7	102.8
1990	105.3	104.8	132.5	92.7	92.2	116.6

Note: Calculations are based on unrounded numbers.
Sources: **Nominal Fixed-Weight Energy Price Index--**Based on net energy consumption (1987 fixed-quantity weights) and end-use price data from EIA, *State Energy Price and Expenditure Data System 1991*. **Real Fixed-Weight Price Index--**Nominal energy fixed-weight price index/Gross Domestic Purchases benchmark-years-weighted price index. **Gross Domestic Purchases, Price Index, 1987 = 100, Benchmark-Years-Weights, 1980-1990-**U.S. Department of Commerce, Bureau of Economic Analysis, *Survey of Current Business*, April 1993, Volume 73, No. 4., p. 35, and *National Income and Product Accounts of the United States*, Volume 2, 1959-88, pp. 266-267.

Table C3. Commercial Energy Price Indices for Specific Energy Sources, 1980 to 1991 (1987 = 100)

Year	Commercial Energy Price Indices					
	Nominal			Real		
	Electricity	Natural Gas	Fuel Oil	Electricity	Natural Gas	Fuel Oil
1979	67.2	59.2	100.6	100.9	88.9	151.1
1980	78.6	72.7	142.6	106.7	98.6	193.5
1981	90.7	85.0	175.7	113.1	105.9	219.9
1982	98.8	102.1	169.2	116.5	120.4	199.5
1983	101.1	117.8	154.0	115.1	134.2	175.4
1984	102.5	117.7	152.8	112.9	129.6	168.3
1985	104.5	115.8	142.6	111.3	123.3	151.8
1986	103.4	107.0	101.3	107.0	110.8	104.9
1987	100.0	100.0	100.0	100.0	100.0	100.0
1988	99.5	97.6	95.5	95.8	93.9	91.9
1989	101.6	99.3	111.4	93.5	91.5	102.6
1990	103.9	101.0	132.9	91.4	86.9	116.3
1991	106.7	101.0	121.0	90.5	85.6	102.6

Note: Calculations are based on unrounded numbers.
Sources: **Nominal Fixed-Weight Energy Price Index--**Based on net energy consumption (1987 fixed-quantity weights) and end-use price data from EIA, *State Energy Price and Expenditure Data System 1991*. **Real Fixed-Weight Price Index--**Nominal energy fixed-weight price index/Gross Domestic Purchases benchmark-years-weighted price index. **Gross Domestic Purchases, Price Index, 1987 = 100, Benchmark-Years-, 1980-1991--**U.S. Department of Commerce, Bureau of Economic Analysis, *Survey of Current Business*, April 1993, Volume 73, No. 4. p. 35, and *National I Income and Product Accounts of the United States*, Volume 2, 1959-88, pp. 266-267.

The CBECS Undercoverage Issue

The 1979 CBECS sample design was based on an adaptation of a nonbuilding national survey sample design. By means of using this approach, data could be gathered, assimilated, and analyzed much faster. However, EIA did recognize that readapting this design was an interim solution. The 1979 CBECS sampling frame is suspected of undercovering the building population. Undercoverage could have been the result of exclusion of specific types of buildings (e.g., farm buildings, outbuildings, buildings on military bases), mistakes by listers, or the handling of buildings under construction at the time of listing. Regardless of the source of coverage problems, evidence suggests that an undercounting did occur and introduced nonsampling error into the survey estimates.

The 1983 CBECS sample had two components: (1) a follow-up of all the buildings sampled in 1979 and (2) an update sample from the files of F.W. Dodge[22] to represent all buildings constructed since 1979. Any deficiencies present in the original 1979 sample were also present in the 1983 CBECS estimates for buildings constructed through 1979. Furthermore, the update frame appeared biased towards large buildings, with severe undercoverage of smaller buildings and possible overcoverage of larger buildings.

For the 1986 sample, CBECS constructed a frame designed specifically for a national survey of commercial buildings. An independently drawn sample from the 1986 CBECS frame was used to represent buildings constructed through 1986 in the 1989 CBECS. The 1989 CBECS had a two-part update for new construction. The F.W. Dodge lists were used to update for large buildings, and roughly half of the area sample segments were relisted to update for smaller buildings.

The undercoverage problem associated with the early CBECS surveys can be readily seen in Table C4, which contains estimates of the number of commercial buildings by year of construction for the successive survey years.

Table C4. Number of Commercial Buildings by Year of Construction and Survey Year

Year of Construction	Number of Buildings (thousand)			
	1979 CBECS	1983 CBECS	1986 CBECS	CBECS 1989
Total	3,073	3,185	4,154	4,528
Year of Construction				
1979 or Before	3,073	3,055	3,495	3,667
1980 to 1983	N/A	131	350	317
1984 to 1986	N/A	N/A	309	329
1987 to 1989	N/A	N/A	N/A	215

N/A Not applicable.
Sources: Commercial Buildings Energy Consumption Surveys (CBECS).

The 1979 and 1983 CBECS estimates of the number of buildings constructed before 1980 are very close. This is expected since both estimates are based on the same sample of buildings. Estimates from the 1986 and 1989 CBECS differ more because each was based on independent samples drawn from the same sampling frame. The most notable difference among the estimates is the large increase in the number of buildings built between the 1979/1983 CBECS and the 1986/1989 CBECS. It is believed that this increase reflects the improved sampling frame introduced in 1986. Because the later CBECS estimates more accurately reflect the commercial building stock, this report uses the 1989 CBECS estimates of the building stock constructed in 1979 or before as the lower bound for the true value.

[22]F.W. Dodge, National Information Services Division, McGraw-Hill Information Systems Company, New York, NY. Figures reported currently in *Dodge Construction Potentials.*

It is a lower bound because some buildings, which existed in 1979 were demolished, or because of some type of physical conversion, no longer fit into the definition of a commercial building during the 1980's.

The adjustment was carried out as follows. For building counts and floorspace estimates, the 1989 estimates of pre-1980 construction were substituted for the 1979 estimates, both across the population and for individual subgroups (e.g., regions, size categories, and building types). Consumption estimates for the various fuels in 1979 were created by multiplying the 1989 values of floorspace by the consumption per square foot intensities based on 1979 data. Expenditures for 1979 were computed by multiplying these consumption values by 1979 energy prices adjusted to 1987 dollars.

Primary Electricity Consumption

Electricity consumption can be expressed in terms of either physical units (e.g., kilowatthours), or thermal units (most commonly British thermal units (Btu)). Conversion of electricity use from kilowatthours to Btu can be given in terms of either site electricity or primary electricity. Site electricity is the amount of electricity delivered to the building, while primary electricity is site electricity plus the amount of energy lost during the generation, transmission, and distribution of the electricity. Conversion of site electricity from kilowatthours to Btu is at the universal value of 3,412 Btu per kilowatthour. Because of energy losses, primary electricity is about three times that of site electricity. The exact conversion value varies by year (e.g., 10,388 Btu per kilowatthour in 1980 and 10,335 Btu per kilowatthour in 1990). Conversion values for all survey years are given in the Glossary under **Btu Conversion Factors**.