

6. Data Processing and Imputation

This and the next chapter of the Quality Profile cover the processing operations that follow the collection of RECS data from households, rental agents, and energy suppliers. This chapter is about the quality aspects of initial manual reviews of questionnaires, coding, data entry, computer-assisted edits, and imputation. Chapter 7 covers the weighting procedures used to develop sample estimates, the model-based procedures for allocating consumption of each fuel to specific end uses, and the estimation of sampling errors.

The primary outputs of the processing operations discussed in this chapter are the data files delivered by the survey contractor to EIA. These data files are of three kinds:

1. *Household Files*, which contain all information collected for every *interviewed* household
2. The *HURS* (Household Unit Record Sheet) *File*, which contains information on type of living quarters and occupancy, sampling information, and data collection outcomes for *all* sample housing units, including those that did not respond to the survey
3. *Utility Bill Data Files*, which contain individual household billing data, annualized consumption and expenditures data, and associated weather data for each of the five major fuels.

Unlike the Household and Utility Bill Files, the HURS File does not include the weights that would be needed to produce estimated totals and weighted unit response rates.

Prior to delivering these files to EIA, the survey contractor removes specific identifiers and related information that would make it possible for EIA to identify individual respondents. For the same reason, individual bill data are inoculated with random errors.

An overview of the processing operations covered in this chapter was provided in the first section of Chapter 2, under "Data Processing and Imputation." As noted there, the procedures are intricate and detailed, consisting of a large number of distinct processing steps. Initially, data from each of the six major sources--Household Survey interview questionnaires, Household Survey telephone and mail questionnaires, Housing Unit Record Sheets, Rental Agent Survey questionnaires, Supplier Survey questionnaires and billing records, and weather data--are processed independently to make them computer-readable, eliminate as many errors as possible, and impute values for some of the items that are missing or incorrect (other kinds of imputation require matching with records from other sources to provide donor information). Then data from the six sources are compared and combined in various ways to produce the three major output files.

This chapter describes the general structure of the data processing operations, with emphasis on features that affect the quality of the final output. Quantitative data are presented when available. Most of the material focuses on the procedures used in the 1990 and 1993 RECS; however, some relevant data from earlier surveys are presented and significant procedural changes over the

history of RECS are discussed. Readers who want a step-by-step detailed description of all processing activities should consult the *Data Editing and Manipulation Procedures Manuals* for the 1990 and 1993 RECS (Response Analysis Corporation 1992b, 1995c).

Several steps in data processing, especially the manual and computer edits, are designed to detect and, insofar as possible, to reduce or eliminate errors. The purpose of imputation is to replace missing entries or those believed to be in error with values that are closer, at least on the average, to the correct ones. However, errors can also be introduced at each stage of data processing. The overall processing system is designed to optimize the quality of the final product. The underlying philosophy guiding this effort has been that because the RECS sample of households is relatively small, the use of substantial manual and computer resources to make the final files as "clean" as possible is justified.

Data Processing Other than Imputation

In order to establish a context for discussing the quality aspects of data processing, this section begins with a general overview of the nature and flow of the main processing steps. The overview is followed by a review of the quality-related features of manual operations and computer-assisted operations. Finally, some results of a special analysis of changes resulting from computer-assisted data processing operations in the 1984 RECS are presented, along with some summary data for these kinds of changes in the 1987 and 1990 RECS.

Structure of the Processing Operations

Figures 6.1 and 6.2 show the main processing steps, including imputation, for the RECS data sources other than the weather data obtained from the National Oceanic and Atmospheric Administration (NOAA). As noted above, questionnaires and other inputs from each of these five sources are first processed independently of each other. These processing steps are shown for each data source in Figure 6.1. For the sake of completeness, imputation, which will be discussed in the following subsection of this chapter, is included in the figure. The first three steps--receipt and check-in, manual coding and editing, and data entry--are primarily clerical operations. The edits listed under Step 4 are computer-assisted. Computer-generated listings of suspect variable values and associated information are reviewed manually by editors and, when necessary, by supervisors or specialists, to determine whether and how to change the suspect values. In some instances, respondents are recontacted by telephone.

Steps 5 and 6 are computer procedures that apply only to the Household Survey data. Computer-generated updates are used for Household Survey interview questionnaires to correct certain common interviewer or respondent errors without manual examination of the questionnaire. These updates are used only in a few instances where a preliminary review has shown that a particular kind of correction is virtually certain to be appropriate for all households that have a specified combination of entries.

Figure 6.1. Steps in Internal Processing of Data from Each Major Source^a

Processing Step	Data Source				
	Household Survey			Rental Agent Survey	Supplier Survey
	Interview	Phone, Mail	HURS ^b		
1. Receipt, check-in	X	X	X	X	X
2. Manual editing and coding	X	X	X	X	X
3. Data entry	X	X	X	X	X
4. Computer-assisted edits					
a. Range checks	X	X	X	X	X
b. Internal consistency checks	X	X	X	X	X
c. Special reports	X				
5. Computer updates	X				
6. Same-source imputations	X	X			

^aProcessing of weather data, which follows a different pattern, is excluded from this exhibit.

^bHousing Unit Record Sheet.

Source: Response Analysis Corporation, *1990 RECS Data Editing & Manipulation Procedures Manual* (September 1992).

Same-source imputations for the Household Survey questionnaires (step 6) are those which do not require inputs from other sources. They may be based on the values of other variables for the same household or on data for other sample households.

The initial processing of weather data, not shown in Figure 6.1, consists of the extraction and manipulation of temperature data from data tapes obtained from NOAA. The first step is to associate a weather station with each secondary sampling unit (SSU, see Chapter 2, Figure 2.3), taking into account physical proximity and the completeness of data available for the stations. For each weather station associated with one or more SSU's, temperature data are extracted from the NOAA tapes, missing data are imputed, and long-term and reference year values of heating and cooling degree-days are developed.

Figure 6.2 lists the principal processing operations that require comparing or merging data from more than one source:

- A. The annualization of Supplier Survey bill data for individual housing units starts with data reported by suppliers for billing periods, most commonly months, and uses these data to develop estimates of total consumption and expenditures for each fuel for the 12-month survey reference period. As part of this process, which is described later in this chapter, the degree-day information developed from the NOAA tapes is used to adjust data for billing periods that overlap the start or end of the calendar-year reference period and to impute data for parts of the year for which no billing data are available.

Figure 6.2. Processing Operations Involving More than One Data Source

-
- A. Annualizing billing data
 - B. Comparison of Household and Rental Agent Survey responses
 - C. Imputation of missing data for telephone and mail households
 - D. Model-based outlier checks of Supplier Survey data
 - E. Creation of output files
-

Source: Response Analysis Corporation, *1990 RECS Data Editing & Manipulation Procedures Manual* (September 1992).

- B. As discussed in previous chapters, Rental Agent Survey responses for such variables as main heating and cooling equipment and fuels are usually considered to be more reliable than those of the occupants of rental units in multiunit structures and are substituted for the latter when disagreements exist.
- C. The imputation of data items not included on the telephone and mail versions of the Household Survey questionnaire requires matching the two sets of sample households--those that responded by interview and those that responded by telephone or mail--on variables that are common to all questionnaire versions to find interview households that are suitable to serve as "donors" for imputation.
- D. The model-based outlier checks of estimates of annual consumption, which were discussed in Chapter 5, require merging of Household and Supplier Survey data.
- E. The ultimate goal of all processing steps up to this point is the creation of the output files, especially the Household Files, which, in their final form, include data from all six sources.

Manual Operations: Quality Considerations

Questionnaires from each of the three surveys, following receipt and control operations, are subjected to an extensive manual combined editing and coding operation. For each type of questionnaire, the first step is to verify the accuracy of the basic identification information. Then the editors check each questionnaire item for completeness and logical consistency with responses to closely related items. In preparation for data entry, they enter codes next to card column numbers on the questionnaire.

The extensive and detailed nature of the editing/coding operations may be seen by examining the detailed instructions for processing interview questionnaires from the 1993 Household Survey (EIA 1995b). In addition to the basic tasks of consistency checking and coding, editors are instructed to:

- Convert numerical entries for such items as amount of wood burned to the desired dimensions, rounding when necessary.
- Keep lists of write-in responses to "other" categories for several questions which have this response option. These lists are used later in processing to recode some of the "other" responses to other existing or newly created categories.
- Refer unusual types of entries and other problems to a special coding and editing section or to designated technicians.

Editing and coding of the housing unit measurement section of the Household Survey questionnaire require certain particularly complex tasks, such as dealing with measurements for floors having shapes that are not simple rectangles.

There is no formal verification system for the manual editing and coding operations. The subsequent computer-assisted edits provide an opportunity to detect some of the errors that may have been overlooked or introduced by the editors. One example of findings from such checks is provided in the next subsection.

Data keying is performed by a separate EIA contractor. Batch tapes of keyed data are transmitted to EIA and loaded to its main computer, where they are used by Response Analysis Corporation, the main contractor, to create unedited data files for each source and perform the computer edits. For the 1981, 1982, and 1984 surveys, key Household Survey questionnaire items were 100-percent verified and the remaining items verified for a 25-percent sample of households. However, a review of the changes that had been made during processing operations for the 1984 Household and Supplier Surveys showed that keying errors were leading to substantial numbers of computer edit rejects (Jabine 1987). The costs of processing these rejects were deemed to exceed the savings from sample verification of data entry and there was also no guarantee that the computer edits and special reports would detect all keying errors. Consequently, beginning with the 1987 RECS, all keying has been 100-percent verified.

Computer-Assisted Edits

Figure 6.3 shows the different kinds of computer-assisted edit checks used in RECS. *Range checks* are applied to values for individual variables from all of the survey components of RECS. Simple range checks ensure that no illegal or impossible variable values are included in the final records. Outlier checks identify, for clerical review, values for continuous variables that may be correct but are unusually high or low for that variable.

Internal consistency checks are also applied to data from all of the RECS survey components. Most commonly, these checks examine relationships of responses for different variables for the same household. Some checks involve comparisons of data for the same household from the current and prior rounds of RECS. For example, housing unit area measurements may be compared for the current and immediately preceding round. Such checks can be used only in those survey years for which the sample includes a longitudinal component and only for the housing units that were in the sample both times.

Figure 6.3. Computer-Assisted Edits Used in RECS

-
- A. Range checks
 - 1. Simple range checks
 - 2. Outlier checks

 - B. Consistency checks
 - 1. Internal to source
 - a. Same unit, same survey
 - b. Same unit, prior survey
 - c. Different units, same survey
 - 2. Across sources

 - C. Special reports
-

Source: Response Analysis Corporation, *1990 RECS Data Editing & Manipulation Procedures Manual* (September 1992).

Another kind of internal consistency check compares data for different households from the same survey component. For example, during the processing of the Supplier Survey records, values for each household are compared with those for some of the other sample households. Some of these comparisons are with all households using the same fuel and others are with all households obtaining that fuel from the same company. In these between household checks, large differences are treated as outliers and are reviewed to determine whether changes are necessary.

The primary consistency checks across survey components are items B and D shown in Figure 6.2. Item B, Comparison of Household and Rental Agent Survey responses, is the final step in processing data collected in the Rental Agent Survey. For the 1990 RECS, Memo #951 provides instructions for resolving the differences appearing on computer-generated listings. In addition to reviewing the Household and Rental Agent questionnaires for units with differences, editors are instructed to review other sources of information, including Supplier Survey data and questionnaires for neighboring households in the same building.

Item D, Model-Based outlier checks of Supplier Survey data, requires the use of Household and Supplier Survey data for the same units. A nonlinear model, with parameter values based on the previous survey, uses data on housing and household characteristics from the Household Survey to predict fuel consumption for each unit. The predicted values are compared with annualized estimates of consumption based on bill data obtained in the Supplier Survey. In the 1990 RECS, whenever the model-based estimate was more than three times or less than one-third of the annualized consumption estimate from the bill records, the data from the two sources were checked.

For the Household Survey records, in addition to the range and consistency checks, a series of computer-generated *special reports* lists information for households with unusual responses or combinations of responses on several different topics. Figure 6.4 lists the topics for which special reports were prepared in the 1990 RECS. Each of these reports identifies the households that qualify for inclusion and gives their values for variables that are relevant to the topic of the report.

Figure 6.4. 1990 RECS: List of Topics for Special Reports

1. Households with Wood-burning Inconsistencies	14. Swamp Coolers
2. Households Not Using a Heating Fuel	15. Households Using Gas Air-Conditioners
3. Households Reporting That Their Main Heating Fuel Changed in November 1987 or Later	16. Air-Conditioner Section Skip Patterns
4. Missing Supplemental Fuels and Equipment	17. Unusable Measurements
5. Incompatible Fuel/Equipment Combinations	18. Households with Unknown Heating Equipment and Heating Fuel
6. Family Members' Ages and Relationships	19. Households That Selected Underground Gas as an Alternative Heating Fuel
7. Central Fuel Inconsistencies	20. Households with Marital Inconsistencies
8. Inconsistencies in the Number of Rooms Cooled	21. Basement Insulation vs. Basement Heating
9. Inconsistencies with the Family Grid	22. Recoded Variables
10. Inconsistencies with the Foldout Page	23. Changes Made to the Mail Questionnaire Donor Selection Variables
11. Households That Use Heat Pumps	24. Various Miscellaneous Checks
12. Lighting	25. Households That Used the Answer "Other"
13. Households with No Windows, Doors, or Electricity	

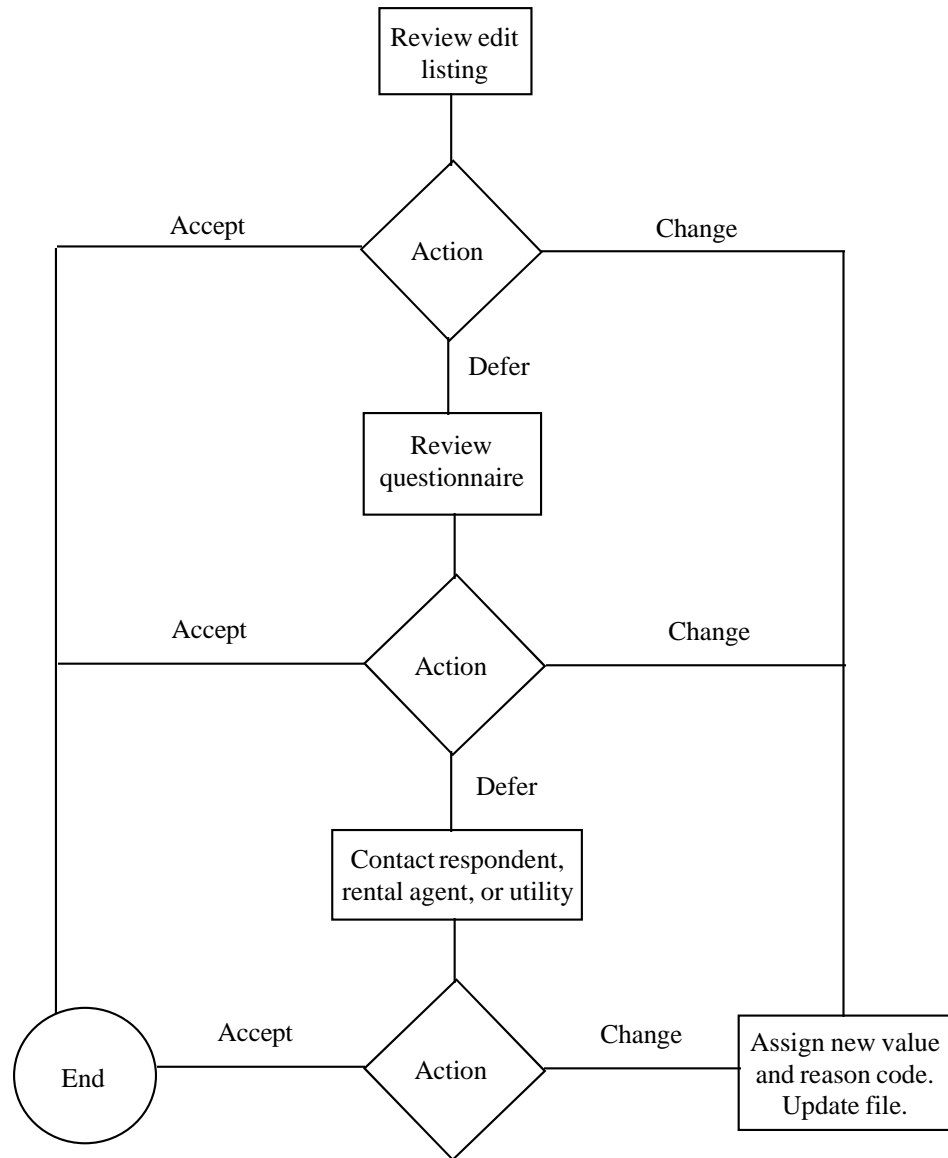
Source: Response Analysis Corporation, *1990 RECS Data Editing & Manipulation Procedures Manual* (September 1992).

Processing Computer Edit Outputs

A more or less standard approach, shown in Figure 6.5, was used to resolve rejects and questionable values included in the computer-generated outputs of range checks, consistency checks, and special reports. In some instances, editors can decide whether to accept or change a response on the basis of a review of the information in the computer listing and the original questionnaire. Failing this, decisions require examination of other relevant questionnaires. At this stage, an editor may be able to determine, for example, that a coding or keying error led to the inclusion of an incorrect value in the computer record. As a final resort, in a small proportion of cases, editors may attempt to contact a household, rental agent or supplier. In order to maintain an audit trail, whenever an editor decides to change a value in the record, information about the nature of the change and the basis for making it is recorded in an archive file.

For the 1990 RECS, instructions for processing each of the special reports are contained in a series of RECS-90 Memos. Most of the memoranda consist of general rules for making changes or for referring certain types of problems to a supervisor. In a few instances, however, the memoranda include quantitative information about the number of households included in a report and the manner in which apparent inconsistencies and other possible errors were dealt with.

Figure 6.5. Processing Computer Edit Outputs



Source: Response Analysis Corporation, 1992b.

A special report was run to examine consistency and skip checks for the "family information grid" (questions K-1 and K-3). This report generated lists of households with 14 different kinds of potential errors--for example, households reporting more than one spouse or households with inconsistencies between the number of persons listed in the chart and the variable for the number of persons in the household. RECS-90 Memo #602 provides information about the disposition of the 292 households that appeared on the 14 listings. Changes were made for 227 (78 percent) of these households, in most instances because the editor determined that there had been coding or keying errors. The largest number of changes, 127, were made in cases where the family grid showed the householder to be married but a separate question on marital status (K-6) indicated otherwise. Most of the discrepancies were the result of coding errors: a person identified as partner or fiancée in the family grid was incorrectly coded as spouse. The relevant questionnaire items were modified in the 1993 RECS in an effort to reduce the frequency of this kind of error.

Another set of special reports dealt with households that reported using gas air conditioners. Because the use of gas as a fuel for air conditioning is rare, all of the 117 sample households that reported it were listed and their data were reviewed. Some cases were resolved by reviewing data for the same household from the 1987 RECS, where similar checks had been undertaken, or from the Rental Agent or Supplier Surveys. Respondent contacts were attempted with 53 of the 117 households. As a result of all of these efforts, the weighted estimate of the number of households using gas air-conditioners was reduced from 1.3 million to 0.4 million (Response Analysis Corporation 1992b, RECS-90 Memo #306 and EIA 1993a, p. 150).

The percentages of sample households for which telephone contacts have been successfully completed at *any* stage of the manual and computer edit procedures have been as follows (*Consumption and Expenditure* reports for years shown):

<u>Survey Year</u>	<u>Percent Contacted</u>
1981	14
1982	10
1984	6
1987	1
1990	4

The gradual reduction in telephone contacts with Household Survey respondents during data processing, especially from 1987 on, reflected concerns that such contacts may have been hurting response rates for the Residential Transportation Energy Consumption Survey and for households in the longitudinal panel for RECS.

Analysis of Processing Changes in the 1984 RECS

As noted above, information about changes made to individual records in the initial unedited Household and Supplier data files is systematically maintained in archival files. The archival files for the 1984 RECS provided the basis for a detailed analysis of processing changes (Jabine 1987). The initial portion of the study was based on tables that had been produced as a matter

of routine, showing the number of changes and their distribution by "reason" for each of 562 Household Survey and 59 Supplier Survey variables. Summary findings for all variables combined for each survey are shown in Table 6.1 (the Billing Files contained Supplier Survey data for the five fuels).

Table 6.1. 1984 RECS: Changes to the Household and Billing Files, by Reason

Reason	Changes to Household File		Changes to Billing Files	
	Number	Percent	Number	Percent
Keying error ^a	1,868	9.1	2,066 ^a	50.0 ^a
Coding error ^a	3,699	18.0		
Clerical error (prior to coding)	NA	NA	374	9.0
Interviewer error	1,118	5.5	NA	NA
Respondent error	236	1.2	122	3.0
Interviewer or respondent error	422	2.1	NA	NA
Data processing error (after keying)	202	1.0	1	*
Phone call to respondent household	514	2.5	20	0.5
Phone call to utility/supplier	256	1.3	496	12.0
Other phone call or information	143	0.7	14	0.3
Rental agent (master meter) information	1,251	6.1	--	--
Kerosene survey information	NA	NA	--	--
Editor's judgement	9,807	47.8	1,016	24.6
Additional information from questionnaire	545	2.7	25	0.6
None of the above	411	2.0	--	--
Total	20,472	100.0	4,134	100.0

^aChanges due to keying errors that could not be distinguished from changes due to coding errors.

-- = None in this category.

NA = Not Applicable.

* = Less than 0.05 percent

Source: Jabine, *Review of Computer Edit & Update Performance Statistics for the RECS, Final Report* (December 1987).

Knowledge of certain limitations and other features of the data shown in Table 6.1 is necessary for an informed interpretation:

- When review of a computer edit reject led to changes in a string of two or more consecutive variables in the same portion of the record, an archival file record was created only for the first variable. Consequently, the numbers in the table are undercounts of the total changes made.
- Not all of the changes made during processing were made because of errors; some of them were built into the processing system. About one-tenth of the changes to the Household File were recodes for questions that included an open-end "other" category. Recodes of the "other" responses were necessary because the final code structures could not be established until after the coding and keying operations had been completed and an analysis of the nature of the open-end responses was possible. Another one-tenth resulted from a special computer program that was used to insert leading zeros in square footage measurement variables for those households in which data from the prior (1982) survey were being used.
- The reason codes that were entered in the archival files are a mixture of two different dimensions: the source of the (presumably) incorrect value that was changed and the source of the information that allowed the editor to determine the (presumably) correct value. The reason "editor's judgment," which was assigned to nearly half of the Household File changes and one-quarter of the Billing File changes, does not provide much useful information in isolation, but its meaning becomes clearer when it is associated with a code showing the specific processing step in which the change was made (see Table 6.2).

Notwithstanding these limitations, it was clear from the data that keying and coding errors were a major source of computer edit rejects, accounting for more than one-fourth of the changes to the Household File and one-half of the changes to the Billing Files. The 193 Household Survey "key" variables for which data entry was verified 100 percent had an average of only 0.44 changes, but the remaining 369 variables that were subject to sample verification averaged 4.83 changes per variable. These findings prompted the decision to revert to 100 percent verification of data entry for the 1987 and subsequent surveys. Use of 100 percent verification has led to a substantial reduction in keying errors, although not necessarily to their complete elimination.

Overall, changes to the Household File averaged 3.6 per household and changes to the Billing Files averaged 0.5 per household (a household was counted once for each fuel for which Supplier Survey data for that household were available). Analysis of variables with large numbers of changes showed that 25 of the 562 Household Survey variables accounted for 42 percent of all changes. Most of these 25 variables were located in two areas of the questionnaire: the portion dealing with main and secondary heating equipment (11 variables) and the portion in which area measurements of the housing unit were recorded (6 variables). For the Billing Files, 10 of 59 variables accounted for 81 percent of the changes. Many of these changes were made to beginning dates for billing periods.

In the second part of the 1984 study of changes, codes were added to the archival records for the Household File to show the stage of data processing at which each change was made and the identification number of the interviewer for the household. Table 6.2 shows a distribution of the Household File changes, by reason, for each step of the processing operations. The nature of most of the processes shown in the table is self-evident. The recode changes were made primarily for two reasons: to assign initial "other" responses to the final set of categories adopted for an item and to make the area measurement data carried over from the 1982 RECS consistent with the 1984 format for these variables. Initial imputations of missing data were not included as changes in the archive files. The "imputation-related" changes shown in Table 6.2 are changes that were necessary to make imputed variables consistent with related variables. The table updates reflect changes that were made to eliminate anomalies detected in preliminary tabulations. Utility and final updates shown in the last two columns of the table were made at the final stages of computer processing for a variety of reasons.

Many features of the processing operations are clear from Table 6.2, for example:

- Nearly all of the changes made to items rejected by range checks were to correct coding and keying errors;
- Consistency checks frequently required contacts with respondents and other sources of information or exercise of an editor's judgment to determine the correct values;
- Special reports followed a pattern similar to that of consistency checks, but fewer changes were made to correct coding and keying errors, presumably because most of them had already been detected by range and consistency checks.

The study report also includes analyses of changes by process for individual variables with large numbers of changes. Analyses of changes attributed to interviewer errors were used to guide the training of 1984 RECS interviewers who were scheduled to serve as interviewers or supervisors in the 1987 RECS.

The report included some recommendations for refinements to the archive files: (1) include a separate record for each variable changed; (2) include a code to show at what stage of processing the change was made; (3) include both old and new values for each change; and (4) replace the 1984 RECS reason code with two codes, one showing the source of information on which the decision to change the value was based and one showing the probable source of the (presumed) incorrect value.

Archive files have been created for all subsequent survey years, but the "reason" codes were replaced by a set of "level-of-effort" codes, putting less emphasis on the source of the error and more on the level of effort required to correct it. Each change was assigned a single code reflecting the greatest level of effort needed to reach a decision. Thus, for example, a change based both on examination of questionnaires and recontact with respondents would be assigned a code reflecting the type(s) of respondent contacts. All decisions about changes required reviewing both edit outputs and the original questionnaire on which the edited record was based. Hence, codes for examination of questionnaires were assigned only when a decision required

Table 6.2. 1984 RECS: Percent Distribution of Changes for Each Process, by Reason

Reason	Process									
	Range Checks	Consistency Checks		Recodes	Special Reports	Imputation Related	Rental Agent	Table Updates	Utility Updates	Final Updates
		Round 1	Round 2							
Interviewer Error	1.0	10.1	16.3	0.1	6.7	0.2	--	--	--	--
Respondent Error	--	0.1	2.3	--	6.5	--	--	--	--	--
Interviewer or Respondent Error	0.1	4.4	4.1	--	3.1	0.9	--	--	--	--
Coding Error	38.6	50.8	23.5	0.1	6.3	0.2	--	--	--	0.7
Keying Error	58.5	14.5	7.5	*	0.4	0.2	--	--	--	--
Data Processing Error	0.4	0.1	5.1	--	0.9	0.1	--	--	--	9.2
Respondent Call	--	3.4	5.2	0.1	4.5	0.5	--	--	61.3	0.7
Utility Call	--	0.2	1.2	--	6.6	1.0	--	--	1.6	0.7
Other Call	--	0.5	0.9	--	2.9	--	--	--	19.4	--
Rental Agent Information	0.2	0.7	9.9	--	2.4	5.9	100.0	--	3.2	4.2
Editor's Judgement	1.1	14.4	22.4	87.2	46.8	90.8	--	100.0	14.5	84.5
Other Questionnaire Information	--	*	--	12.5	--	0.2	--	--	--	--
Other	0.1	0.8	1.6	--	12.9	--	--	--	--	--
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Number of Changes	1,727	4,234	2,998	4,370	2,539	3,664	634	40	124	142

* = Less than 0.05 percent.

-- = None in this category.

Source: Jabine, *Review of Computer Edit & Update Performance Statistics for the RECS, Final Report* (December 1987).

Table 6.3. Changes to the Household File by Source of Information or by Reason: 1987 and 1990 RECS

Source of Information or Reason	Number of Changes		Changes per 100 Households ^a		
	1987	1990	1987	1990	Change
Contacts with:					
Household.....	129	37	2.2	0.8	-1.4
Fuel Supplier.....	81	6	1.4	0.1	-1.3
Rental Agent or Interviewer.....	67	196	1.1	4.1	3.0
Multiple Sources.....	22	3	0.4	0.1	-0.3
Examination of:					
Supplier Data.....	86	202	1.5	4.2	2.7
Rental Agent Data.....	1,399	568	23.9	11.8	-12.1
Other Information.....	275	29	4.7	0.6	-4.1
Multiple Sources.....	668	133	11.4	2.8	-8.6
Application of Inference Editing Rules:					
Less Than 5 Minutes.....	5,338	6,828	91.2	141.4	50.2
5 Minutes or More.....	342	164	5.8	3.4	-2.4
Other					
Correction of Prior Processing Error.....	243	210	4.1	4.3	0.2
Post-Imputation Change.....	64	b	1.1	b	b
Recode of Open-End Response or Special Rule.....	3,625	433	61.9	9.0	-52.9
Dependent Change ^c	13,130	351	224.2	7.3	-216.9
Total					
With Dependent Changes.....	25,469	9,160 ^d	434.9	189.7	-245.2
Excluding Dependent Changes....	12,339	8,809 ^d	210.7	182.5	-28.2

^aNumber of completed personal interviews was 5,856 in 1987 and 4,828 in 1990.

^bCategory not used in 1990.

^cA different definition may have been used in 1990, see text.

^dExcludes one change in category "Dummy Editor II Update."

Source: Reason for Change Tabulations, RECS Personal Interview Editing: 1987 and 1990.

looking at a questionnaire other than the one that the record to be changed was derived from. Codes assigned to changes based solely on reviews of the outputs of range and consistency checks or special reports identified separately those instances where more than 5 minutes time was needed to reach a decision. The level-of-effort codes used in the 1987 and 1990 RECS are listed and explained in RECS Memo #007.

Summary tabulations of changes to the Household File records based on personal interviews for 1987 and 1990 are shown in Table 6.3. In order to account for differences in sample size for the two years, the data are presented in terms of changes per 100 households.

The differences between 1987 and 1990 in the figures for changes of all kinds are dominated by a precipitous drop in the number of "dependent changes." This probably resulted from a change in the procedures for generating archival records and assigning level-of-effort codes, but it has not been possible to determine the exact nature of the change.

Aside from the effect of the different treatment of dependent changes in 1987 and 1990, the overall frequency of changes on a per-household basis was lower by about 13 percent in the latter year. Overall, the frequency of recontacts with respondents was about the same in both years, with a decline in the number of contacts with household respondents and suppliers being balanced by an increase in contacts with rental agents and interviewers. There was a substantial increase in changes based on the application of inference and editing rules, but a decline in the number requiring 5 minutes or more to reach a decision. There was a substantial decline in 1990 in the number of changes representing recodes of open-end and "other" responses or the application of special rules. Much of the decline appears to have been associated with changes made because of inconsistencies between entries in the part of the questionnaire that summarized uses and methods of payment for each fuel and earlier questions on these same topics.

Imputation

At several stages during data processing, interviewers' or respondents' initial entries on questionnaires are changed or deleted, or values (including 0) are supplied for items initially left blank. In RECS, most of this imputation follows completion of the initial round of computer-assisted edits and special reports. Following imputation, consistency checks are repeated to ensure that imputed values are consistent with other related variables.

This section discusses the imputation procedures used for the Household Survey, the Supplier Survey, and the weather information obtained from NOAA. Items missing on Rental Agent Survey questionnaires are not imputed; in general, the Household Survey responses for these missing items are accepted. The model-based allocation of energy consumption and expenditures to end uses, such as space heating, water heating, and appliances, is considered to be estimation, not imputation, because respondents and suppliers are not asked directly for this information. These allocation procedures are discussed in Chapter 7 in the section on "End-use Estimation."

Imputation for the Household Survey: Interview Questionnaires

The frequency of item nonresponse, based on unedited data files from the 1990 RECS, was discussed in Chapter 4. Excluding variables related to household measurements, 51 of 416 variables based on questionnaire entries had item nonresponse rates of 5.0 percent or more. Of the 10 variables with the highest nonresponse rates (see Table 4.4), only 3--age of hot water heater and two items related to household income--required entries for more than 10 percent of the households.

The treatment of each missing item requires two decisions. The first is whether or not to impute a value for it. Since 1982 missing values have been imputed for roughly two-thirds of the Household Survey variables in each survey year (see Table 6.4). Items not imputed are those for which it is judged that there is not enough information for related variables to provide the basis for an imputation procedure that is likely to reduce the effects of nonresponse bias. For the 1990 RECS, items not imputed included questions on the presence, type and amount of attic and floor insulation, indoor temperatures, and the presence of wall insulation (EIA 1992, p.200).

For each item that is to be imputed, a choice of the most appropriate method is required. Except in 1987, hot-deck imputation, in which the missing value is obtained from a household that matches on variables related to the missing item, has been the most commonly used method. Based on an intensive review of imputation procedures prior to data processing for the 1987 survey, some variables were shifted from hot-deck to other methods of imputation. For example, the entire household grid (demographic characteristics of household members) was shifted from hot-deck to an allocation method and deductive imputation methods were adopted for some variables related to main heating fuel and equipment. Some items were shifted back from deductive to hot-deck for the 1990 RECS and hot-deck imputation was used for most of the new variables in 1993.

Table 6.4. Imputation Methods Used for Household Survey Variables: 1981-1993

Imputation Method	Percent of Items Subject to Imputation					
	1981	1982	1984	1987	1990	1993
Not imputed	23	35	32	36	32	32
Imputed	77	65	68	64	68	68
Hot-Deck	58	52	56	27	42	51
Random	13	9	9	15	13	7
Other ^a	6	4	3	22	13	9
Total	100	100	100	100	100	100
Number of Items ^b	356	443	447	422	429	559

^aIncludes regression, deductive, allocation, and modal imputation methods.

^bExcludes items for which missing values, if any, were determined by explicit editing rules during the early stages of processing.
Source: Energy Information Administration, *Housing Characteristics* (1981-1993).

In 1993, hot-deck imputation was the only procedure used for the 26 most frequently imputed variables (EIA 1995a, Table A8). The use of other procedures is limited to variables for which there is relatively little item nonresponse.

Other imputation methods that are or have been used include regression, random, deductive, allocation, and modal techniques:

- A *regression* equation, developed from questionnaires with usable data, is used to estimate the total square footage of each sample housing unit for which actual measurements are not obtained or are unusable. Variables used to predict the area of the unit include such housing unit characteristics, as type of housing unit, year built, number of rooms, number of bathrooms, and type of heating equipment, and such household characteristics as income and number of persons. A full statement of the equation and the variable definitions used in the 1993 RECS is given in the *Housing Characteristics 1993* report (EIA 1995a, pp. 230-231).
- *Random-selection* imputation procedures are used for two purposes: to supply missing dates, such as the year and month a housing unit was occupied, and to supply missing values that are conditional on other known values, such as the number of storm windows in a house with a known total number of windows. A value is assigned at random from the appropriate distribution of values for households that report fully.
- *Deductive* procedures are used when the amount of missing data is small and other available information provides reasonably conclusive evidence of what the missing value should be. These procedures are used primarily when information is missing on fuels used for specific purposes and methods of payment for fuels used.
- *Allocation* procedures are used for imputation of missing information on household members, such as age, sex, and relation to householder. Rules for assigning missing values are based on the configuration of known information on these variables for other household members.
- *Modal* imputation procedures, which were used in RECS only in the earlier survey years, assign the most commonly reported value to the missing variable. Typically, a negative response would be assigned to a question on the presence of a relatively rare item, such as the use of secondary water-heating equipment. In the more recent survey years, modal imputation procedures have been replaced by hot-deck or random methods in order to obtain better variance estimates.

Imputation for the Household Survey: Telephone and Mail Questionnaires

Nearly all household survey variables are imputed for the relatively small proportion of households for which questionnaires are obtained by telephone or mail rather than by personal interview (see Chapter 4, Table 4.1). These imputed values of housing unit and household characteristics are combined with the actual Supplier Survey and weather data obtained for these

households and their localities. Thus, the main purposes of obtaining mail and telephone questionnaires for households not responding to interviews are to identify the energy suppliers for the unit, to obtain vehicle information needed to apply sample selection procedures for the Residential Transportation Energy Consumption Survey, and to provide the data needed to select a donor interview questionnaire for imputation of housing unit and household characteristics.

Figure 6.6 shows the steps in the modified hot-deck procedure that was used to impute data for telephone and mail questionnaires in the 1993 RECS. Most of the procedures were computerized. The procedure ensured that no interview questionnaire was used more than once as a donor for a telephone or mail questionnaire. Donor questionnaires were selected manually only for the small proportion of questionnaires for which a suitable donor was not identified by the computerized scoring rules. In the 1990 RECS, only 3 of the mail questionnaires required donors that did not match on all of the sorting variables used in Operation 1; donors were selected from other Census regions for these donees (Response Analysis Corporation 1992b, p.4-22).

Figure 6.6. Imputation Procedure for Household Survey Mail and Telephone Questionnaires: 1993 RECS

Operation	Sorting/Matching Variables
1. Sort both donor (interview) and donee (mail and telephone) questionnaires into groups based on basis of specified variables.	Census region Type of housing unit structure Space-heating fuel Hot-water fuel Presence of air-conditioning Type of air-conditioning
2. For each donee, pick the best donor from the corresponding sort group, using a scoring procedure based on specified additional variables.	Income Number of persons in household Number of vehicles Age of householder Tenure (owned/rented) Number of rooms Model year of newest vehicle Household type (married couple/other)
3. Assign donor values for all Household Survey variables, except number of vehicles, to the donee household.	

Source: Energy Information Administration, *Housing Characteristics* (1993).

Imputation of Supplier Survey Data: Procedures

The level of imputation for Household Survey questions on housing and household characteristics is relatively low, but it has been somewhat higher for consumption and expenditure data for the 5 major fuels, for which the preferred source of information is the actual bill data obtained from energy suppliers. Two kinds of imputation are required. In a process called "annualization," bill data obtained from suppliers are used to arrive at estimates of consumption for a 1-year period.

For those households for which usable bill data are not obtained, annual consumption is imputed on the basis of housing unit and household characteristics. The extent of need for the latter type of imputation depends on eligibility and completion rates for the Supplier Survey. As can be seen in Table 4.7, Chapter 4, Supplier Survey completion rates have been relatively high for electricity and natural gas, somewhat lower for fuel oil and LPG, and quite low for kerosene. Table 4.8 shows that imputation of consumption of electricity and natural gas is seldom needed for single family houses but is needed more often for housing units in multiunit structures.

The procedures for estimating annual consumption differ by fuel type. For electricity and natural gas, the reporting unit for each household in the Supplier Survey is the *billing period*, and an "annualization" process is used to convert the data by billing period to an estimate for a 365-day period. For the 1993 RECS, utilities were asked to provide data for the sample households for all billing periods starting on or after December 1, 1992, and ending prior to the date at which they were asked to complete the form, generally in the late winter or early spring of 1994. They were asked to report the beginning and ending date for each billing period, the amount consumed, the cost, and whether the amount was based on a reading by the customer or on a reading or estimate by the company. Suppliers were instructed to provide bill data only for the specific account for which a waiver was obtained in the Household Survey. Thus, if the sample household did not occupy the housing unit for all of 1993, bill data would, in most instances, be obtained only for the portion of the year during which they occupied it.

For electricity and natural gas, estimation of annual consumption from billing period data was attempted only in the following circumstances:

1. The household paid for some or all uses of the fuel and had 146 or more days of bill data; or
2. The household paid for appliance and/or water heating use, did not pay for space heating or space cooling use, and had 60 or more days of bill data.

When these criteria were met, the procedure for annualization of a household's bill data followed the steps shown in Figure 6.7. First, an annualization period was defined, consisting of consecutive billing periods with a start date as close as possible to January 1, 1993 (the first day of the reference year for consumption), and with the total number of days as close as possible to 365. Second, a consumption year was defined, containing exactly 365 days and matching the annualization period as closely as possible. Third, *predicted* values of consumption for the annualization period and the consumption year were derived from a nonlinear model by using available information about the household's uses of the fuel, the number of days in the period, and the number of heating and cooling degree-days in the period. Finally, the actual consumption reported for the annualization period was adjusted by the ratio of the values of predicted consumption for the consumption year and the annualization period. Expenditures for the consumption year were estimated by applying the unit cost for the annualization period to the estimated value of consumption for the consumption year. A special adjustment procedure was applied to consumption and cost estimates in those instances where the household paid for some but not all uses of electricity or natural gas. In 1993, such households accounted for 0.7 percent of total annual electricity consumption and 1.9 percent of natural gas consumption (EIA 1995d, Table B7).

Figure 6.7. Imputation of Annual Household Consumption from Billing Period Data for Electricity and Natural Gas

Step	Criteria/Procedure
1. Select billing periods for use in estimation. These billing periods cover the "annualization period" (AP).	(1) Continuous data. (2) Start date close as possible to January 1, 1993. (3) Total days close as possible to 365.
2. Define the consumption year (CY).	(1) Must contain 365 days. (2) Match AP as closely as possible.
3. Calculate predicted consumption (C_p) for AP and CY.	Prediction model based on prior survey. Inputs include household uses of fuel, number of days in period, and number of heating and cooling degree days in period.
4. Calculate imputed consumption C_i for CY.	C_i (CY) = Actual consumption for AP x $\frac{\text{Predicted consumption for CY}}{\text{Predicted consumption for AP}}$

Source: Response Analysis Corporation, *1990 RECS Data Editing & Manipulation Procedures Manual* (September 1992).

The inclusion of information on heating and cooling degree-days as part of the prediction models for the annualization procedure started in the 1990 RECS. Prior to 1990, the prediction models relied only on the total number of days in the annualization period and the consumption year (Response Analysis Corporation 1992b, p. A-195).

For fuel oil, LPG, and kerosene, the reporting unit in the Supplier Survey is the *delivery*. Suppliers were asked to report all deliveries from October 1, 1992, through the date at which they completed the form. For each delivery they were asked to report the type of fuel, the amount, the price per unit of volume, and the total price. They were also asked to report the beginning and ending dates of the period covered by the recorded deliveries. If the beginning and ending dates covered all of calendar year 1993, only those deliveries occurring during 1993 were included as part of consumption. If the data on deliveries did not cover a full year, the Supplier Survey data for that household for fuel oil, LPG, or kerosene were not used. It would be possible to develop an imputation procedure that made use of part-year data on deliveries, but the number of households with part-year data is so small that the addition of such a procedure would have a low payoff.

The Household Survey questionnaire included some questions on deliveries and use of fuel oil, LPG, and kerosene. When no usable Supplier Survey data were available, these responses could sometimes be used to estimate annual consumption. This occurred frequently for kerosene, but rarely for fuel oil and LPG.

As explained more fully in Chapter 7 in the section on End-Use Estimation, a separate nonlinear regression model was developed for each fuel, based on data for sample households that had a

full or nearly full year of acceptable bill data, and was used to allocate total consumption of each fuel to 5 use categories: space heating, water heating, air-conditioning, refrigerators, and other appliances. This same model was used to estimate total consumption of a fuel when the sample household used it and did not have usable bill data from the Supplier Survey or, for kerosene, from the Supplier or Household Survey. The regression imputation procedure included the addition of a random error component, making it possible to calculate estimates of sampling error without separating imputed and unimputed data.

Finally, an imputation adjustment was made for each fuel for any household reporting in the Household Survey that some of its bills covered non-household uses of that fuel, for example, for a farm or home business or another household. In such instances, total consumption was reduced by a scale factor developed on the basis of responses to a Household Survey question about the proportion of the bills for that fuel which covered the non-household uses. The percentages of estimated total annual consumption for such households for each fuel in 1990 and 1993 were:

<u>Fuel</u>	<u>1990</u>	<u>1993</u>
Electricity	2.8	4.9
Natural gas	0.6	0.6
Fuel oil	1.3	2.0
Kerosene	0.1	1.0
LPG	3.2	4.4

Imputation of Supplier Survey Data: Quality Implications

As noted in Chapter 3, the goal of RECS is to collect data, for a sample of households, on energy consumption of each major fuel used for residential purposes during a specified time period (for the 1993 RECS, calendar year 1993). Chapters 4 and 5 have revealed several factors that pose problems for the acquisition of precise data for each household, most of them related to nonresponse or incomplete response to the Supplier Survey. The primary factors, the procedures used to deal with them, and their effects on the accuracy of consumption estimates are summarized in Figure 6.8.

For electricity and natural gas, the ideal situation would be to have, for each household, metered values of total consumption, for household uses only, for the calendar year covered by the survey. Because the metering and billing practices of utilities seldom meet these precise requirements, various kinds of compromises and approximations are required. For fuel oil, LPG and kerosene, direct records of consumption do not exist, so information about delivered amounts during the consumption reference period is used as a proxy.

Figure 6.8. Sources of Error in RECS Consumption Data

Fuels Affected and Source of Error	Estimation Procedures
<p>Electricity and natural gas</p> <p>No separate metering for household</p> <p>Billing periods do not coincide with reference year.</p> <p>Billing periods cover only part of reference year.</p> <p>Estimated bills.</p> <p>Household pays for some but not all uses.</p>	<p>Model-based imputation based on household survey data.</p> <p>Annualization procedure.</p> <p>Annualization procedure.</p> <p>Influences choice of annualization period.</p> <p>If bill amounts are annualized, the results are adjusted upwards.</p>
<p>Fuel oil, LPG, and kerosene</p> <p>Data available for deliveries, not actual consumption.</p>	<p>Estimation based on delivered amounts.</p>
<p>All fuels</p> <p>Changes in occupancy during reference year.</p> <p>No supplier survey data obtained.</p> <p>Nonresidential uses included in bills.</p>	<p>Household generally treated as if occupied for full year.</p> <p>Model-based imputation based on household survey data.</p> <p>Amounts scaled back based on household survey estimate of proportion nonresidential.</p>

For all fuels, bill data from the Supplier Survey normally only cover the period during which the sample housing unit was occupied by the household that was present at the time of the Household Survey interview. In cases of part-year occupancy, the imputation procedures treat such housing units as though they were occupied and consumed fuels at the same rate for the entire reference year. Because some of these units were actually not occupied or even not ready for occupancy for part of the year, this approach to imputation overstates their consumption. However, this overstatement may be at least partly offset by the failure to obtain consumption data for units that were occupied for part of the consumption year but were vacant during the interview period for the Household Survey.

Table 6.5 shows, for 1990 and 1993, the proportion of total annual consumption of each fuel that was estimated or imputed by the various methods just described. For fuels other than kerosene, bill data for all or most of a year were the basis for roughly two-thirds to four-fifths of the estimated amounts. For electricity and natural gas, less complete bill data accounted for about 8 or 9 percent of the total amounts. The proportion of consumption based on regression estimates varied from one-tenth for electricity to slightly more than one-third for fuel oil.

Table 6.5. Basis of Estimates of Annual Consumption: 1990 and 1993 RECS (Percent of Total Consumption of Each Energy Source)

Source of Consumption Estimates	Fuel and Year									
	Electricity		Natural Gas		Fuel Oil		Kerosene		LPG	
	1990	1993	1990	1993	1990	1993	1990	1993	1990	1993
Actual Billing Records										
Covering All Uses										
330 or more days ^a	80.5	80.2	74.3	74.1	64.7	67.4	28.5	27.8	71.6	79.1
146 to 329 days	7.5	8.5	6.4	6.3	NA	NA	NA	NA	NA	NA
60 to 145 days	0.2	0.1	0.1	*	NA	NA	NA	NA	NA	NA
Covering Some Uses	0.7	0.7	2.0	1.9	NA	NA	NA	NA	NA	NA
Estimate from Supplier or Household ^b	NA	NA	NA	NA	0.1	0.9	43.0	34.4	0.3	0.1
Regression Estimate	11.0	10.5	17.3	17.7	35.3	31.7	28.5	37.8	28.1	20.9
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Percent of Total Consumption Accounted for by Fuel	32.5	32.8	53.1	52.7	10.6	10.2	0.7	0.5	3.0	3.8

^aFor fuel oil, kerosene, and LPG, billing records were used only if they covered 365 days.

^bFor kerosene, the estimate was supplied by the household.

* = less than 0.05 percent.

NA = Not Applicable.

Note: Because of rounding, percents may not sum to 100.0.

Sources: Energy Information Administration, special tabulations of RECS data files for 1990 and 1993.

For kerosene only, a substantial proportion of total consumption (more than two-fifths in 1990 and about one-third in 1993) was estimated from information on deliveries and purchases supplied by household respondents. Slightly more than one-fourth was based on bill data from the Supplier Survey and the remainder was based on regression estimates.

The basis for estimates varied substantially by type of living quarters. For electricity in 1993, for example, the percent of consumption based on regression estimates by type of structure was:

<u>Type of Structure</u>	<u>Percent Based on Regression Estimate</u>
Mobile home	10.8
One-family detached	7.5
One-family attached	12.6
2 to 4 housing units	25.7
5 or more housing units	22.9

This kind of variation occurs because many of the households in multiunit structures are living in rental units for which some of the utilities are included in the rent and are therefore not eligible for the Supplier Survey. Similar patterns with more pronounced differences by type of structure can be observed for natural gas, fuel oil, and LPG.

As shown on the last line of Table 6.5, the percent of total consumption accounted for by each fuel varies substantially, from 52.7 percent for natural gas in 1993 to 0.5 percent for kerosene in the same year. Using these percents and the data in the body of Table 6.5, it is possible to calculate the basis for estimates of annual consumption for all fuels combined:

<u>Source of consumption estimate</u>	<u>1990</u>	<u>1993</u>
Bill data for all or most of year	74.8	75.4
Partial bill data	7.3	7.4
Estimate from supplier or household	0.3	0.3
Regression estimate	17.5	17.0

Imputation of Missing Weather Data

Weather data obtained from the National Oceanic and Atmospheric Administration (NOAA) for use in RECS include daily temperature data for each of NOAA's weather stations. These data, which are used primarily to estimate heating and cooling degree-days for sample housing units, are sometimes missing for one or more days. Through the 1984 RECS, average temperatures for all weather stations in a NOAA division were used for this purpose. Starting with the 1987 RECS and subsequently, an individual weather station has been selected to provide temperature data for each cluster of sample housing units. The extent of missing temperature data is one of the factors considered in selection of a weather station to be associated with each cluster; data quality is considered acceptable if data are missing for fewer than 15 days of the consumption reference year. Once the stations are selected, missing temperature data are imputed by making use of the relationship between division temperatures and station temperatures for the previous year. For each survey year, for those clusters that remain in the sample from earlier years, the

selection of the associated weather stations is reviewed to take into account changes in the extent of missing temperature data (Response Analysis Corporation 1992b, pp. 8-31 to 8-35).

Evaluation of Imputation Procedures

Because of concern about the high proportion of imputed consumption data for housing units in apartment buildings, a special study was undertaken as part of the 1981 RECS. Permission was obtained from selected apartment building managers to obtain actual fuel records for their buildings, each including one or more RECS sample housing units. Total consumption for each building was allocated equally to the apartment units in that building, and the estimates for the sample units were compared with imputed values assigned by the regression modeling procedures used for units lacking Supplier Survey data. These comparisons indicated biases in some imputed values (EIA 1983b, p.102):

<u>Households Using</u>	<u>Imputed Values Are:</u>	<u>Adjust. Factor</u>
Electricity with air-conditioning	Too low by 50 percent	1.84
Electricity, no air-conditioning	Too high by 10 percent	None
Natural gas for space heating	About right	None
Natural gas, but not for space heating	Too low by 50 percent	2.04

Records of use of fuel oil and LPG in apartments were insufficient in number to make reliable estimates of bias in their imputed values. As a result of this study, the adjustment factors shown above were applied to imputed values of electricity and natural gas consumption in apartments. The same adjustment factors were used in the 1982 RECS. For the 1984 RECS, the regression imputation model was revised to reflect differences between apartments and other units more explicitly, so that these final adjustments were no longer necessary.

RECS questions about temperature setting behavior in the household have been among those for which missing values have not been imputed. Battles and Harrison (1992), using 1990 RECS data, experimented with several regression models in an attempt to identify some of the household and housing unit characteristics that relate to temperature setback behavior when natural gas, electricity, or fuel oil is used for space heating. They also hoped to develop discrete temperature models that could be used to impute missing temperatures. Some of their findings were:

- Low income was a significant factor in a household's decision to reduce temperatures in homes heated with fuel oil, but not in those heated with natural gas or electricity.
- For all three fuels, homes with higher daytime temperatures were likely to have higher setback temperatures as well. Homes that had recently installed a clock thermostat were likely to have lower setback temperatures.
- The colder the climate, the higher the proportion of households that set back temperatures.

In spite of these and other significant findings, the fit of the models developed was not thought to be good enough to use them to impute missing values, so they were not imputed for these items in the 1993 RECS.