

Validation of Conditional Demand Estimates: Does It Lead to Model Improvements?

Stephanie J. Battles, Energy Information Administration'

The Energy Information Administration uses Conditional Demand Analysis (CDA) to produce residential end-use consumption estimates for each of the five major energy sources: natural gas, electricity, fuel oil, kerosene, and liquefied petroleum gas. Residential Energy Consumption Survey (RECS) data are used as variables in these CDA models. An earlier study, presented at the ACEEE 1990 Summer Study, tested the validity of 1987 RECS end-use estimates of electricity using submetered estimate comparisons obtained from households that were not part of the 1987 RECS. This study suggested a misallocation between space conditioning and water heating. A further study was warranted where RECS CDA end-use estimates were based on households which were also submetered.

In this paper, the validation study of the 1990 RECS CDA model is presented using data collected from five electric utilities with not only submetered electricity consumption data but also housing unit and household characteristic data. This study supports the earlier study. Based on the results of this study and the earlier study, model improvements to be incorporated into the 1993 RECS CDA end-use models are proposed such as taking into account regional differences and household behavior.

Introduction

One approach to testing the validity of end-use estimates from Conditional Demand Analysis (CDA) is by comparing the estimates with submetered end-use estimates obtained from individual households. In 1990, the validity of 1987 Residential Energy Consumption Survey (RECS) CDA electricity end-use estimates were compared with submetered estimates obtained from households that were not a part of the 1987 RECS.² This study suggested a misallocation between space conditioning and water heating. Therefore a further study was warranted where RECS CDA electricity end-use estimates were based on households which were also submetered. This new study of the validation of the 1990 RECS CDA end-use estimates also suggests a misallocation between space conditioning and water heating in the estimation of end-use consumption.³

Research Approach

The best test of the accuracy of statistically derived RECS end-use estimates would be to submeter the energy-using equipment in those RECS households. However, resources were not available to complete such a study. In the study presented in this paper, five utilities agreed to participate in this validation of the 1990 RECS CDA electric end-use

model. These five utilities furnished submetered end-use, household, and housing unit characteristic data for the participating households. Statistically-derived CDA estimates of end-use electricity consumption were developed for each of the households supplied by the participating utilities. Since the household data from each utility company were not as extensive as the 1990 RECS data, a modified CDA end-use model containing only variables from the utility data set was created for each utility. Wherever possible, variables from the utility dataset were used as substitutes for RECS variables, and these changes were reflected in the modified CDA model. Often, however, no adequate substitutes were available, and variables from the full 1990 RECS model were excluded in the modified models. The modified models, therefore, are abbreviated versions of the full 1990 RECS end-use model.

The modified models were run using the utility data. These statistically-defined estimates⁴ were then directly compared to the submetered end-use estimates. The quality of the end-use estimates from the modified models were evaluated to discover potential biases. Mean estimates using both the full RECS CDA model and the modified model were developed based on RECS data for

each end use by utility. These comparisons supported the use of the modified models but also led to the development of an adjustment index.

In the same manner as the 1990 study, RECS households were selected to match the utility households. These statistically derived end-use estimates using the full 1990 RECS model were compared to the submetered estimates. Also, the modified model for each utility was run using the selected RECS households. Examining these selected RECS households made it possible to reassess the modified RECS models compared to the full RECS model and to use these estimates to develop an adjustment index for the modified CDA model estimates.

RECS CDA Electric Model

The RECS, conducted by the Energy Information Administration (EIA), is a triennial national multistage probability sample survey concerning housing unit and household characteristics. The RECS data are collected via a Personal interview with the householder. Householders are asked to sign authorization forms allowing their suppliers of energy to release billing information about their household. A mail survey is used to collect household energy consumption and expenditure information from the energy suppliers.

One concern to policymakers and others is how total electricity consumption is partitioned among uses. 1990 RECS data were used to develop the 1990 RECS CDA electric model to estimate five end-use components for each RECS household. Total electricity consumption was partitioned into five end-use components:

Total Consumption = Space Heating
 + Water Heating
 + Air Conditioning
 + General Appliances
 + Refrigerators

The parameters for the model were estimated using nonlinear regression techniques. As described earlier in endnote 4, a normalization procedure was required to develop final estimates since the model does not predict total consumption without some level of error. Therefore,

$$Y_a = Y_e + \epsilon$$

where Y_a = actual total annual usage of electricity from whole house billing data,

and Y_e = estimated total annual usage of electricity from the sum of end-use estimates,

and ϵ = error term.

Electricity usage is modeled as a function of equipment and behavioral characteristics. The components of the model may not be completely independent. The terms used to predict one end use may be collinear with another end use. This is a major justification to undertake such a validity study of a statistical model as the one presented in this paper.

Electric Utility Submetered End-use, Household, and Housing-unit Data

The survey data from each of the five utilities—the City of Austin (AUS), Bonneville Power Administration (BPA), Pacific Gas and Electric (PGE), Santee Cooper, and Southern California Edison (SCE)—was first modified to match the 1990 RECS data as closely as possible before being merged with the submetered end-use data.

Those households missing information, either survey data or submetered data that were crucial to the RECS model were deleted from the final dataset and summarized in Table 1. Not all of the electric utilities had submetering data for each of the specific end uses (see Table 2).

Table 1. Validity Study of the 1990 RECS Electric CDA Model: Number of Observations

Electric Utility	Number of Observations	
	Originally Obtained from the Utility	Used in the Modified RECS CDA Model
City of Austin	26	13
Bonneville Power Administration	318	62
Pacific Gas and Electric	629	182
Santee Cooper	42	33
Southern California Edison	124	56

Table 2. Validity Study of the 1990 RECS CDA Electric Model: Availability of End-Use Submetering Data from Participating Utilities

End Use	Electric Utilities				
	City of Austin	BPA	PGE	Santee Cooper	SCE
Central Air Conditioning	Yes	No	Yes	Yes	Yes
Room Air Conditioning	No	No	Yes	No	Yes
Space Heating	Yes	Yes	No	Yes	No
Water Heating	Yes	Yes	Yes	Yes	No
Refrigerators	No	Yes	Yes	No	Yes
Total Appliances ^(a)	Derived	Derived	No	Derived	No

(a) Total appliances also includes refrigerators.

Modified Models

Although in general the modified models performed well using RECS data compared to the full RECS CDA model, there were some specific end uses for some of the utilities where use of the modified model clearly affected the comparisons (Table 3). As an example, the appliance and refrigerator mean estimates from the City of Austin’s modified model were significantly different from the full 1990 RECS CDA model’s mean estimates. The missing variable on the number of household members and variable on lighting may be one reason why the City of Austin’s appliance estimate was so low. The City of Austin’s dataset was missing one core refrigerator term that every other dataset contained—ownership of a third refrigerator and thus the variable had been excluded from the model. Additionally, the City of Austin’s model could not contain a variable on the size or age of the refrigerators. To correct for these and other missing variables in the City of Austin and the other modified models, adjustment indices were developed using selected RECS household-based median estimates from the full 1990 RECS and modified models as depicted in Table 4. Using median estimates’ an adjustment index was calculated for each end use per utility by dividing the median estimates from the two models using RECS data:

$$\text{Adjustment Index} = \frac{\text{full 1990 RECS CDA model median end-use estimate}}{\text{modified RECS CDA model median end-use estimate}}$$

The median end-use estimates used in developing adjustment indices are not based upon all the RECS households, but only those RECS households that are comparable to the utility households. The selection procedures attempted to find a reasonable sample of RECS households that matched utility households. Estimates were produced using the same selected RECS data in both the full RECS models and in the modified model as well (see footnote b in Table 4). The modified model end-use estimates developed using utility data were then multiplied by the appropriate adjustment index.

Comparisons of CDA End-use Estimates to Submetered Estimates

The first comparison examines the difference between the median submetered end-use estimates and the median CDA estimates developed using the utility data in the modified models. The percent difference between the median estimate and the median submetered estimate is calculated for all households for each utility and each end use where applicable.⁷

Assuming the modified RECS models did not differ significantly from the full RECS model, the differences between the estimates and submetered estimates are the true indicator of how well the 1990 RECS model allocates usage. For example, if the estimates for central air-conditioning are consistently higher than the submetered estimates, one can conclude that the 1990 RECS CDA

Table 3. Modified CDA Model Mean End-Use Estimates Compared to RECS Full CDA Model Mean End-Use Estimates Based on 1990 RECS Data

RECS/Electric Utility	Electric End Use				
	Appliance	Refrigerator	Main Water Heating	Air Conditioning	Main Space Heating
1990 RECS Full CDA Model Estimate^(a) (million Btu)	15.1	5.3	9.7	6.9	8.2
Electric Utility					
City of Austin					
Modified CDA Model Estimate (million Btu)	11.7	8.1	10.0	7.6	7.9
Percent Difference from RECS Full CDA Model	-22.5	52.8	3.1	10.1	-3.7
Santee Cooper					
Modified CDA Model Estimate (million Btu)	15.9	5.0	10.8	7.4	8.7
Percent Difference from RECS Full CDA Model	5.3	-5.7	11.3	7.2	6.1
Pacific Gas and Electric					
Modified CDA Model Estimate (million Btu)	15.0	5.0	9.0	7.5	8.7
Percent Difference from RECS Full CDA Model	-1.0	-5.7	-7.2	8.7	6.1
Bonneville Power Administration					
Modified CDA Model Estimate (million Btu)	14.4	5.8	9.6	7.0	8.5
Percent Difference from RECS Full CDA Model	-4.6	9.4	-1.0	1.0	3.7
Southern California Edison					
Modified CDA Model Estimate (million Btu)	14.6	5.6	9.5	7.4	8.6
Percent Difference from RECS Full CDA Model	-3.3	5.7	-2.1	2.8	4.9

(a) End-use estimates are calculated only for those households using the particular end use.

model is allocating too large a share of electricity to air conditioning for housing units that share characteristics with those that were submetered.

Secondly, index adjusted end-use estimates from the modified models are compared to the submetered estimates. Figure 1 presents these comparisons which are discussed in the following sections.

Central Air Conditioning

For the utilities, City of Austin, PGE, Santee Cooper, and SCE, the central air-conditioning estimates from the adjusted-modified RECS models are greater than the submetered estimates.⁸The modified models themselves do not have the distinction between having central air-conditioning and the amount of usage (not used, used a few days, used quite a bit, and used just about all summer). In 1990, only 46.9 percent of households with central air conditioning, used it just about all summer.⁹ This is important information when allocating a portion of

total energy consumption to central air conditioning. Nevertheless, it does appear, based on the adjusted-modified model estimates for central air conditioning, that the 1990 RECS electric model is overestimating central air conditioning. The utilities do cover a reasonable range of cooling climates—South Carolina, Texas, and California.

Room Air Conditioning

Both the modified model using utility data and the modified model using selected RECS data seem to overpredict room air conditioning consumption for both PGE and SCE. Knowledge of the actual usage duration is even more important when predicting room air conditioning consumption than it is for central air conditioning estimation. The absence of this information might inflate the estimates. The full RECS 1990 model can give these households an estimate of zero while the modified model estimates an end use amount for them even though the household may have the equipment but did not use it. Both of these utilities are in California where the

Table 4. Comparison of End-Use Median Estimates: RECS CDA Model and Submetered, 1990

End-Use By Utility	End-Use Estimate (Million Btu)				
	Utility Data Used			RECS Data Used	
	Utility Submetered	Modified RECS CDA Model	Adjusted Modified RECS CDA Model ^(a)	Full RECS CDA Model ^(b)	Modified RECS CDA Model
Central Air Conditioning					
City of Austin	15.6	17.4	21.8	21.9	17.5
Pacific Gas and Electric	.8	4.5	3.6	4.4	5.5
Santee Cooper	14.3	10.1	16.1	19.8	12.4
Southern California Edison	6.7	10.1	8.9	7.6	8.6
Room Air Conditioning					
Pacific Gas and Electric	.8	1.7	.5	.6	1.9
Southern California Edison	1.3	4.6	2.1	1.2	2.6
Space Heating					
City of Austin	5.8	12.9	8.9	7.6	11.0
Bonneville Power Adm.	21.5	21.1	19.6	20.4	22.0
Santee Cooper	7.1	10.0	9.3	5.7	6.1
Water Heating					
City of Austin	10.9	8.6	8.1	8.6	9.1
Bonneville Power Adm.	14.8	14.2	15.2	14.5	13.5
Pacific Gas and Electric	11.4	6.7	8.9	6.9	5.2
Santee Cooper	8.4	10.0	7.8	6.4	8.2
Refrigerator					
Bonneville Power Adm.	4.7	4.0	3.7	4.2	4.6
Pacific Gas and Electric	5.1	4.3	5.0	3.5	3.0
Southern California Edison	4.9	6.2	5.7	5.0	5.4
Total Appliance^(c)					
City of Austin	44.2	36.3	32.1	23.6	26.7
Bonneville Power Adm.	27.4	30.3	28.6	25.6	27.1
Santee Cooper	21.4	26.9	21.1	23.1	29.5

(a) The adjustment index = full RECS CDA model estimate based on selected RECS households/modified RECS CDA model estimate based on selected RECS households. Multiply this index by the estimate from the modified RECS CDA model estimate based on utility households to obtain the estimates for this column.

(b) The selection procedure attempted to match the utility households. The criteria for selection differed for each utility and end use but usually focused on Census division, climate zone, primary sampling unit, type of home, and type of equipment.

(c) Also includes refrigerator consumption.

temperature is temperate and the room air conditioners might be used less. In the 1990 RECS, the percentage of all room air conditioners “used just about all summer” was only 20 percent nationally. The percentage in the Pacific Census division was even lower, 11.3 percent. 10

Taking the limitations into consideration, the adjusted-modified model estimates for room air conditioning, though, when compared to the submetered estimates for both PGE and SCE, do suggest a misallocation of too much consumption to room air conditioning.

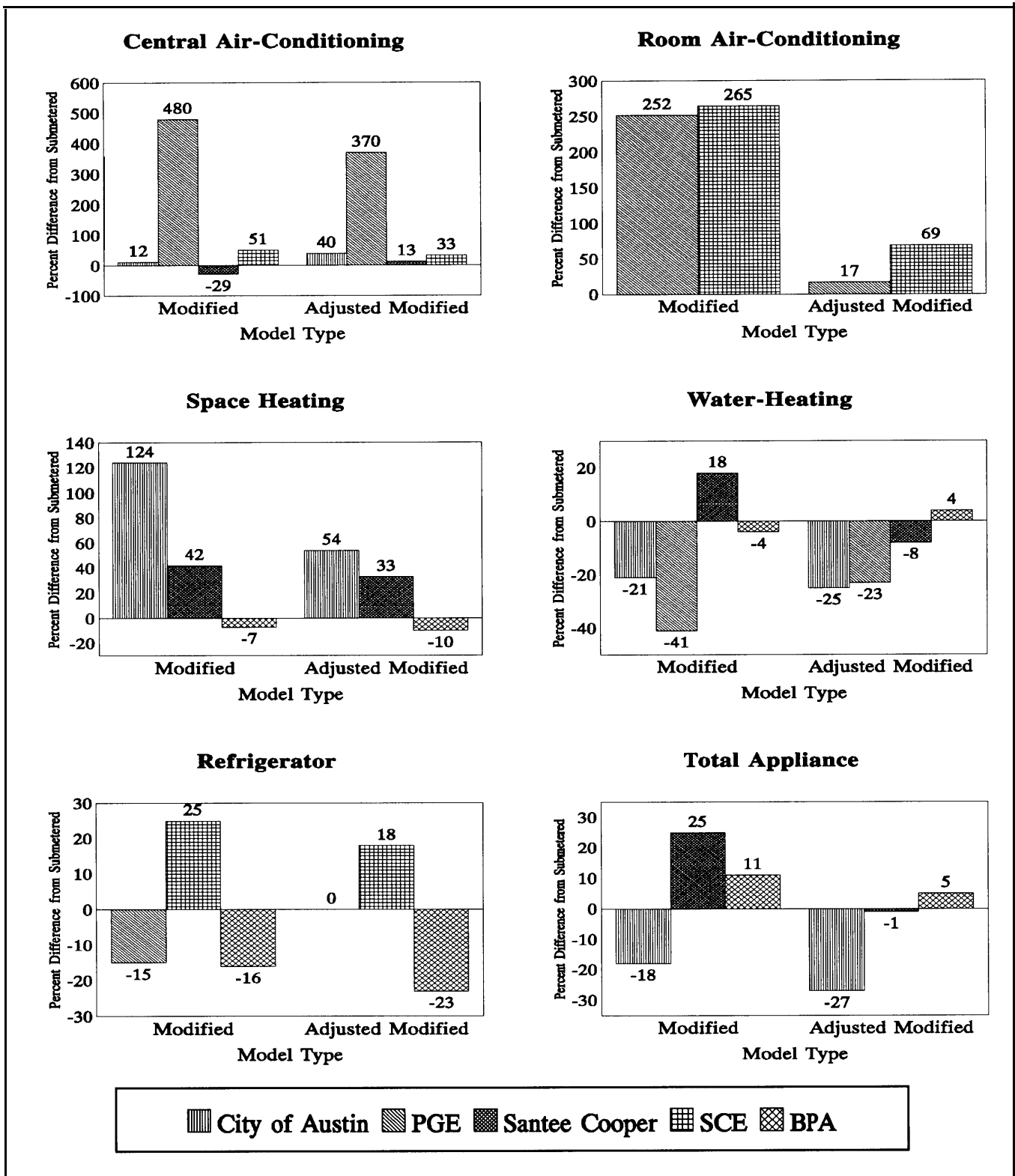


Figure 1. CDA Median Estimates Compared to Utility Submetered Estimates, 1990

Space Heating

The City of Austin's modified model's space-heating estimate was twice as large as the submetered median estimate. Even the median estimate from the modified model using RECS households was much greater than the full RECS model. This may give rise to the importance of missing variables from the modified model. Two missing variables were the age and race of the householder. The elderly use on average, 14.2 million Btus for electric space heating compared to 9.4 million Btus for all households. Black households use 9.7 million Btus as compared to 12.6 million Btus for white households.¹¹ After taking into account the limitations of the modified model and thus using the adjustment index, the adjusted estimate for space heating was still 54 percent higher than the submetered median estimate for space heating.

The BPA median estimates based on both the modified and adjusted-modified models were very similar to the submetered estimate for space heating. The BPA modified model contained most of the same variables as the 1990 RECS CDA model. Santee Cooper's space-heating median estimates based on both the modified and adjusted-modified models were greater than the submetered median estimate. Although the results seemed to be mixed, the RECS space-heating estimates do seem to be overestimated.

Water Heating

BPA's median estimates for water heating based on both the modified and adjusted-modified models are close to the submetered median estimate. City of Austin, PGE, and Santee Cooper median estimates are less than the submetered estimates, even after using the adjustment index on the modified estimates of water heating.

In any model like the 1990 RECS electric model, if one end use is being overestimated, there has to be an underestimation some where else in the model. The water-heating estimates do seem to be underestimated.

Refrigerators

Although refrigerators are normally considered part of appliances, end-use estimates for refrigerators were produced for the first time in 1990 using RECS data. Refrigerators use a lot of electricity—at least 17 percent of all household electricity in 1990.

There doesn't seem to be a clear pattern in the refrigerator comparisons. There was no difference between the PGE adjusted-modified model estimate and the submetered

estimate for refrigerators. The modified and adjusted-modified model estimates from the other two utilities, BPA and SCE produced mixed results.

Total Appliances¹²

The BPA and Santee Cooper adjusted-modified model estimates seem to be very close to the submetered estimates. The City of Austin sample, though, consisted of only 8 households.

Summary of Findings

Previous research on this topic has suggested that there are biases in the RECS end-use estimation techniques.¹³ The results from this study partially support those findings.

In this paper, it has been emphasized that the modified RECS CDA models used for this validity study were not as complete as the full 1990 RECS CDA model. The best of validity studies would be to compare the full RECS model estimates to submetered estimates from the same RECS households. Costs for such a study of this caliber would be beyond EIA's budgetary capabilities at this time.

Nevertheless, when using the adjusted-modified model estimates in the comparisons in this validity study, it does seem that RECS space-heating and cooling estimates are overestimated and RECS water-heating estimates seem to be underestimated. With the exception of BPA, the comparisons for the other four utilities support these conclusions.

Potential Sources of Modeling Biases

The earlier validation of the 1987 RECS model hypothesized that the high air-conditioning estimates might be caused by omitted variable bias. New variables such as: how often the central air conditioner and the room air conditioner was used; size of the most-used room air conditioner; presence of shade trees; and age of the air conditioners, were added variables to the 1990 RECS CDA electric model.

Another possibility is that there is a substantial amount of collinearity among the discrete end-use variables. For example, the size of the home is likely to be correlated with the number of people in the home. The size of the home is likely to be a good predictor of space and air conditioning. The 1990 RECS model has housing size in the space-heating and air-conditioning component and

household size in the water-heating component. However, since there is a relationship between the two independent variables, the parameters for the components are not independent and thus, violate one of the assumptions of the model.

A third possibility is suggested by the results from the BPA modified model. It may be true that the factors that determine space-heating and air-conditioning consumption in different climates are quite different. One obvious example is the degree-day base for heating and cooling degree-days—the bases may be different for different regions of the country.

Potential Improvements to the 1993 RECS CDA Electric Model

In the last section, three potential sources of error have been identified: omitted variable bias, collinearity among the end-use components; and shifts in model structure for different areas of the country.

During 1994, five 1993 RECS models will be developed using the 1993 RECS data and household total consumption billing data for the major energy sources: electricity, natural gas, LPG, fuel oil, and kerosene. Additional questions on the 1993 RECS Household Questionnaire and the results of preliminary investigations using the 1990 RECS electric model as a outcome of this validity study, should facilitate the development of robust 1993 RECS models.

The paper concludes with a discussion on potential improvements of a preliminary nature that may prove valuable in the development of the 1993 RECS CDA models for each of the major energy sources.

Omitted Variable Bias

The creation of modified models for this validity study has been very insightful as to the importance and unimportance of certain variables in modeling specific end-use consumption. The information obtained concerning the “usage” of air-conditioning equipment has been especially insightful. Available in the development of the 1993 RECS CDA models will be new usage variables for clothes washers and dryers, dishwashers, water usage for bathing, and number of meals cooked. New lighting data will also be available such as whether the outdoor lights are on a timer and the number and length of time incandescent and fluorescent lights are on.

Empirically, rural areas are cooler than city areas. EIA obtains the weather data from airports and therefore the

heating degree-days for city areas may be overestimated and underestimated for rural areas. Cooling degree-days may be underestimated for city areas and overestimated for rural areas. In the 1993 RECS, the interviewer was asked to observe the kind of area-city, town, suburbs, or rural that the housing unit was located. Using this information as adjustment terms for the space and air conditioning components may correct the estimations of degree-days depending upon the location of the household.

Collinearity Among the End-Use Components

Time permitting, in the development of the 1993 RECS CDA models as an offshoot of this validity study, the utility submetered data files may be useful for examining the potential for collinearity among the predictor variables. From the data files, a “bottom-up” approach could be undertaken. A model for each end use can be developed, with the submetered estimate as the dependent variable. Then a model for all total consumption can be developed, with annual consumption as the dependent variable.

Additionally, the new data involving the usage of hot water, namely the use of the dishwasher and clothes washer, and the number of baths or showers, may assist in the improvement of the water-heating component of the RECS CDA model. These new variables may lessen the problem of collinearity in that they are specific for hot water usage.

Model Different Areas of the Country Differently

Fundamental differences in the way an end use is employed would suggest that regional differences should be incorporated into the model. RECS models could be developed separately for a number of different regions of the country and then the results could be compared to the model for the entire RECS sample. This is a very time-consuming project and best to undertake when the actual 1993 RECS models are being developed. Possibly, a less time-consuming investigation could involve the effects of different aspects of climate. Three facets of climate: temperature, humidity, and the concept of variable degree-days could be explored. Preliminary investigations have been undertaken using the 1990 RECS CDA electric model.

Preliminary analysis demonstrates potential model improvements using: (1) the concept of a temperature-humidity index (THI), and (2) the notion of variable degree-days using different temperature bases.

Additionally, studies could be undertaken to estimate the magnitude of the affects of each of these aspects of climate on each separate end use.

Temperature-Humidity Index. A temperature-humidity index (THI) could be developed for each RECS household using temperature and humidity data from *Comparative Climatic Data For the United States Through 1990*. These data are not as area specific as would be ideal, but the THI could serve as a rough estimate of relative differences between area of the country.¹⁴ The THI should capture household reaction to the combination of heat and humidity or the stress (discomfort) level.

Degree-Days Using Variable Temperature Bases. An argument could be made to use variable degree-days temperature bases to handle the problems associated with the shoulder time periods. It would seem that households such as those in the Northeast who face a long, expensive heating season would be reluctant to use their heating equipment until necessary. Those in the South may turn on their heating equipment early-as soon as the climate turns cool. The opposite may take place with air conditioning. Households in the South face a long cooling season. It would seem reasonable that they would first use fans, open windows, etc. before using their air-conditioning equipment. Homes in areas such as the Northeast may use their equipment as soon as the climate turns warm.

Acknowledgments

The 1990 RECS validity study reported here was sponsored by the Energy Information Administration, U.S. Department of Energy under contract No. DE-AC01-92EI21948 with Response Analysis Corporation.

This validity study could not have been undertaken without the valuable cooperation of the utilities: City of Austin, Pacific Gas and Electric, Santee Cooper, Southern California Edison, and the Bonneville Power Administration.

Endnotes

1. The opinions and conclusions expressed herein are solely that of the author and should not be construed as representing the opinions or policy of any agency of the United States Government.
2. See "Comparison Between Residential End-Use Submetering and Conditional Demand Estimates For a National Survey" in Performance Measurement and

Analysis-ACEEE 1990 Summer Study on Energy Efficiency in Buildings, Volume 10, pp. 10.5-10.13. American Council for an Energy-Efficient Economy, Washington, D.C.

3. See "Appendix D. End-Use Estimation Methodology" in Household Energy Consumption and Expenditures 1990 (DOE/EIA-0314(90)). This appendix documents EIA's 1990 RECS CDA end-use models not only for electricity, but also for natural gas, fuel oil, liquefied petroleum gas, and kerosene.
4. The CDA end-use estimates were normalized so that the sum of the end-use estimates was equal to the actual yearly consumption of electricity by each household.
5. The appliance component also includes the electricity used for lighting.
6. Medians are used in the comparison because of small sample sizes, presence of outliers, and only households with both submetered data and household and housing characteristic data are included. These data are merely a sample of utilities that have undertaken submetering studies. The most meaningful statistic is the percentage difference in the medians, which give an indication of the direction and magnitude of possible biases in the 1990 RECS CDA electric model.
7. The percent difference for each household was calculated. Additionally, the distribution was examined graphically to observe how the end-use estimates differed from the submetered estimates. Since space is limited, these results have not been presented in this paper.
8. The very sharp disparity between the PGE submetered estimate and the CDA estimate for central air conditioning does suggest a potential problem associated with the submetered estimate.
9. See Table 54 in Housing Characteristics 1990 (DOE/EIA-0314(90)), pp. 180-184.
10. See the preceding endnote.
11. See Table 33 in Household Energy Consumption and Expenditures 1990 (DOE/EIA-0321(90)), pp. 90-92.
12. Includes refrigerators.
13. See endnote 2.

14. The formulation of the THI was obtained through personal communication on 8/19/93 from Joseph Eto of the Energy Program Analysis Group at the Lawrence Berkeley Laboratory, University of California.

References

Battles, Stephanie J. 1990. "Comparison Between Residential End-Use Metering and Conditional Demand Estimates for a National Survey." Performance Measurement and Analysis - Proceedings from the ACEEE 1990 Summer Study on Energy Efficiency in Buildings, Volume 10, pp. 10.5-10.14. American Council for an Energy Efficient Economy, Washington, D.C.

Energy Information Administration. 1992. Housing Characteristics 1990. DOE/EIA-0314(90), U.S. Department of Energy, Washington, D.C.

Energy Information Administration. 1993. Household Energy Consumption and Expenditures 1990. DOE/EIA-0321(90), U.S. Department of Energy, Washington, D.C.