Long-Term Residential Load Forecasting

EPRI

Keywords:
Load Forecasting
Time-of-Day
Peak-Load Pricing
Residential Electricity Demand

EPRI EA-584
Project 943-1
Final Report
Volume 1
February 1978

Prepared by
University of California
Econometrics Research Associates

MASTER

ELECTRIC POWER RESEARCH INSTITUTE

DISCLAIMER

Portions of this document may be illegible in electronic image products. Images are produced from the best available original document.

Long-Term Residential Load Forecasting

EA-584, Volume 1
Research Project 943-1

Final Report, February 1978

Prepared by

UNIVERSITY OF CALIFORNIA AT SAN DIEGO 950 6627
Department of Economics
La Jolla, California 92037

and

ECONOMETRIC RESEARCH ASSOCIATES 950 6552
1321 Stratford Court
Del Mar, California 92014

Principal Investigators
Clive W. J. Granger
Robert F. Engle
Ramu Ramanathan
Allan Andersen

Prepared for

Electric Power Research Institute
3412 Hillview Avenue
Palo Alto, California 94304

EPRI Project Manager
Anthony G. Lawrence
Energy Analysis and Environment Division

eb

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

LEGAL NOTICE

This report was prepared by the University of California, San Diego (UCSD) and Econometric Research Associates (ERA), as an account of work sponsored by the Electric Power Research Institute, Inc. (EPRI). Neither EPRI, members of EPRI, UCSD, ERA, nor any person acting on behalf of either: (a) makes any warranty or representation, express or implied, with respect to the accuracy, completeness, or usefulness of the information contained in this report, or that the use of any information, apparatus, method, or process disclosed in this report may not infringe privately owned rights; or (b) assumes any liabilities with respect to the use of, or for damages resulting from the use of, any information, apparatus, method, or process disclosed in this report.

FOREWORD

The Demand and Conservation Program at EPRI is sponsoring an ongoing research effort to develop new methodological approaches for load forecasting. This research effort concentrates on developing statistical models that elucidate the long-run behavioral and technological determinants of the hourly and seasonal load patterns of electricity consumption.

This research report is the first of several studies developing new econometric and statistical methods for modeling household level load patterns. The study develops a two-step scheme for estimating both fifteen-minute and hourly household electricity demands. These microload curves can be aggregated into a total residential load curve.

Load modeling studies can only be as strong as the data that support them. For the present study the Connecticut Peak-Load Pricing Experiment provided a rich and fertile proving ground for alternative approaches to household-level load modeling. Indeed, the only two significant shortcomings of the Connecticut experiment were its relatively short duration and the use of only one experimental peak load rate schedule. The former limits the ability to make statistical inferences about the load pattern's long-run response to time-of-day rates as households change their appliance and space-conditioning equipment to take advantage of the bargain priced off-peak electricity. The latter shortcoming precludes estimation of the contemporaneous and noncontemporaneous price elasticities of demand unless highly restrictive assumptions are imposed on the econometric demand model specification. Thus the nature of available data precludes the study from adequately measuring the price-induced responsiveness of the residential load pattern. The model structure would be fully applicable to data with a variety of time-of-day rates were they available. Applied to such data, it would provide estimates of the price-induced responsiveness of the load pattern to peak-load pricing rate structures. As such experimental data become available, this report should be a valuable foundation upon which further work can be developed. Even without time-of-day price data, the study has accomplished a useful and inexpensive empirical specification for estimating how typical household load curves are affected by the household's appliance ownership and other socioeconomic variables. This, by itself, is a major advance.

Further research projects in progress or planned will deal with alternative econometric and statistical methods for modeling residential load patterns. Research on commercial and industrial loads by establishment is also under way, as is research on potential transportation loads. When combined with the residential load studies, this research will provide a firm basis for both long-term load forecasting and conservation analysis. Since the models under development are price responsive and have end-use detail, they will be useful for analyzing the effects of load management alternatives, including peak-load pricing, load shifts due to new electricity-utilizing devices, effectiveness of conservation regulations and efficiency standards, and the impact of changing economic and social variables.

In the not too distant future these microeconometric models of electricity demand will become part of the core of advanced simulation models describing energy utilization with end-use detail and enabling both EPRI R&D planners and industry forecasters to rigorously analyze and project future electricity uses and load patterns.

Copies of Volume 2 of this report, the Statistical Appendix, may be obtained from my office at Electric Power Research Institute, 3412 Hillview Avenue, Palo Alto, California 94304.

Anthony G. Lawrence
Project Manager
Energy Demand and Conservation

ABSTRACT

The main objective of this study was to isolate and evaluate the importance of various factors, many of which are household characteristics and weather conditions, that determine the demand for electricity at different times of day. A second purpose was to investigate one of the factors in detail, namely, prices, which was feasible because half of the households in the sample were subjected to time-of-day pricing.

Substantial differences between the load curves of the experimental and control groups were found. Households in the experimental group significantly decreased electricity usage when its price was high, the consumption being shifted partly into the early morning hours but more heavily into the evening. The importance of certain appliances in shifting the load curve is also clearly brought out. For example, households with a dishwasher or electric heating appeared to change the timing of use of these appliances under peak-load pricing. Other appliances were also important in determining the load curve for both groups. Swimming pool pumps and air conditioning, for instance, were important determinants in the summer, whereas in the winter, electric heating and dishwashers substantially increased consumption levels.

CONTENTS

Chapter 1.	Summary of Results	1
Chapter 2.	The Problem and the Approach	7
Chapter 3.	Data Used in the Study	10
Chapter 4.	Analysis Using Quarter-hourly Data	17
	Formulation of the Time-Series Models	17
	Time-Series Results	25
	Cross-Sectional Analysis	31
	Using the Model to Forecast	42
Chapter 5.	Analysis Using Hourly Data	48
	Econometric Considerations	49
	Time-Series Regressions	54
	Cross-Sectional Analysis	60
Chapter 6.	Implications for Peak-Load Pricing	72
References		79

Chapter 1

SUMMARY OF RESULTS

Two complete analyses of the data on residential electricity usage were undertaken. One used quarter-hourly observations for the entire 10-month data period and assumed that various parameters were constant throughout the year. The other analysis used hourly observations and was computed separately for winter and summer. Only the latter results are summarized in this chapter because it is felt that the differences between winter and summer are sufficient to make this approach more appropriate. The quarter-hourly results are, however, rather similar in many respects and are described in detail in Chapter 4. Volume II of this report contains the detailed statistical tables.

Using the hourly observations, time-series regressions were run for winter and summer months for 135 households in Connecticut, of which 85 faced a peak pricing schedule. The parameters of these estimated household load curves were then explained on the basis of demographic and appliance-stock data for the households. The procedure is therefore a two-step analysis of a cross section of time series. Considering the difficulty of predicting the timing of personal habits, the fits were moderate to encouraging. Salient details of the final estimates are given below.

Summer Results

The importance of certain appliances is clearly brought out in the estimates. For instance, households with an electric water heater demand significantly more electricity in hours 7 through 24 (hour 1 is from midnight to 1 a.m.). The dishwasher is most significant in the 19th hour, but it also has a significant impact during hours 20 through 23. The product of a dummy variable for central air conditioning and the living area of the house in square feet was used as an explanatory variable. This becomes significant at hours 12 and at hours 14 through 23. The electric dryer is significant at hour 24, and in the preceding two hours it is marginally insignificant. The swimming pool pump is important from 4 to 5 p.m. (hour 17). Among other variables, the number of people under 18 is important during almost all the hours, the only exception being hours 5 through 8. The number of people in the age group 18-64 is significant in the morning from 8 to 9 a.m. and during the rest of the day from 5 p.m. to 1 a.m. (hours 17 through 1). The number of older people (65 or over) doesn't seem to be a significant factor. The dummy variable for the experimental group was generally not significant, although it came close during the 23d hour. The remaining variables did not seem to have any significant effect on the regression coefficients for the hourly dummies. During the high-use period of 9 a.m. to 9 p.m., (hours 9 through 21) the significant variables were number of people under 18, number of people 18-64, electric water heating, and central air conditioning times area (square feet) being cooled.

To economists, the most interesting variable is the peak price. Not surprisingly, the experimental households showed a strongly significant negative effect, indicating that they would use less electricity during the peak pricing period than the control group. The heated pool also has an important negative effect; and households with a pool pump are more likely to shift out of the peak and reduce electricity consumption during that period. The central air conditioning and area seem to interact to yield a negative, but not strong, effect. The number of people under 18 also reacts negatively; (there is a significant shift out of the peak). The remaining variables do not have much effect on peak period usage.

In the case of the weather variables (wind speed, temperature moving average, and temperature-humidity index), only air conditioning (both central and window) times area had a significant impact, especially on the temperature variables. For wind speed, the dehumidifier had a strong effect. The experimental group showed a significant effect on wind speed.

Winter Results

Usage during the winter exhibits interesting patterns similar to those in the summer. Electric heating times area is important during hours 5 through 10, 15, 18, and 21. The effect is generally positive, except from 2 to 3 p.m., suggesting that less is used than between 1 and 4 a.m. The dishwasher has a significant effect from 9 to 11 a.m. and again from 6 p.m. to midnight. The electric water heater has a noticeable effect from 6 to 7 a.m. and again from 9 p.m. to 1 a.m. The number of people under 18 has a significant effect between noon and 1 a.m. (This variable was important at all hours after 8 a.m. in the summer.) Presumably the people under 18 tend to be in school during the winter, accounting for less consumption during those mornings when other variables are held constant. People in the age group 18-64 generally had no effect on the usage at specific hours, the only exception being from 4 to 5 a.m. , when there was a significant decrease in consumption compared with that of the preceding three hours. Here there is a strong contrast between summer and winter. It will be recalled that the age group 18-64 was significant from 8 to 9 a.m. and from 5 p.m. to 1 a.m. The activities of this group involving electricity consumption have indeed been curtailed in winter.

The experimental group did shift out of the peak hours, as evidenced by the significant positive effect for the survey group from 9 p.m. to 1 a.m. This also is in clear contrast to the summer behavior, when the survey group variable was usually insignificant. Presumably the extraordinary expense of winter heating provided the incentive for the survey households to postpone their consumption to the off-peak hours. During the high-use period of 9 a.m. to 9 p.m., the dishwasher and the number of people under 18 were the significant variables, although the electric water heater came close.

The peak price effect was significant for electric heating times area, for dishwasher use, and for the survey group, all of which has the expected negative sign, that is the higher the price the less the usage.

Electric heating times area is significant in all the weather variables, and the survey group had a significant effect only on temperature squared. The electric water heater came close to being significant for wind speed and for temperature moving average. The sum of the temperature variables (temperature now = temperature moving average) had a very good fit, but only electric heating was significant, although supplementary electric heating was nearly significant.

Implications for Peak Load Pricing

As mentioned earlier, the survey group substantially decreased its usage at the peak hours. There was a spreading of the load, particularly in the evening, but also into the morning period before the peak prices are charged. There appears to be a shift out of the shoulder-priced hours, particularly in the evening. The pattern of cross-elasticities appears to be rather complicated. For example, consumption between 8 and 9 a.m. in the winter is above that of the control group even though it is in a shoulder period. The explanation is that the large shift from the peak period, which begins at 9 a.m. , exceeds the decrease that would ordinarily be expected from households shifting usage to the period before 8 a.m. , when the shoulder price begins.

In general, the households facing peak pricing respond in the same way to weather conditions as do those with flat-rate schedules. The exceptions are the quadratic term in temperature in the winter and the response to wind speed in the summer. Ordinarily the temperature-squared variable would be expected to enter with a positive sign, indicating a convex function, with usage increasing faster as the temperature decreases. The survey variable enters the explanation of temperature squared with a significant negative sign, thereby indicating that the households facing peak prices may be more likely to have a concave response to temperature than the control group. In the summer the experimental households apparently decrease their consumption when the wind speed increases. Presumably this is due to a discretionary shutdown of air conditioners or perhaps other appliances.

In the winter the electric heating variable is significantly different between the experimental and control groups, indicating that households with electric heating that face peak pricing shift out of the peak period more than those without electric heat. Another way in which experimental households shift usage out of the peak period seems to be to change the time at which they use the dishwasher. In the summer only the use of the swimming pool pump appears to differ significantly between the experimental and control groups. As this is a regular use of appliances that can easily be shifted, it seems reasonable that it appears to be a mechanism for the shift. Notably lacking any differences is the effect of air conditioners. Not only is the use of air conditioners not significantly different between experimental and control groups, but the point estimates suggest that the experimental group actually uses them more during the peak hours.

In conclusion, there appears to be a very strong response to peak load pricing, with a substantial decrease in the peak hours and the load being shifted partly to the early morning but more heavily into the evening. This shift is particularly associated with households with electric heating but also with those having dishwashers and swimming pool pumps.

Bearing in mind the difficulty of predicting personal habits, the results seem sensible and persuasive. Further analyses along these lines could well help distinguish still further the critical determinants of residential load curves and their response to time-of-day pricing.

Chapter 2

THE PROBLEM AND THE APPROACH

The future demand for electricity by residential users clearly depends on a large number of factors. The main objective of this study is to isolate and evaluate the importance of many of these factors and thereby to produce a model that, it is hoped, will be useful in forming long-term forecasts of time-of-day demand. In so doing, we also believe that a model useful for short-term prediction will have been produced. The study was made possible by the availability of a comprehensive data set of good quality from Northeast Utilities for parts of Connecticut, and was initiated and funded by the Electric Power Research Institute. The basic approach is outlined in this chapter. Later chapters describe the data used and the stages of the modeling procedure in more detail and outline the results obtained.

If the causes of variations in electricity usage by a single household are considered, three main effects can be distinguished.

(1) First there are those resulting from the lifestyle of the family. The time that a husband or wife has to be at work or a child at school helps determine the start of the active part of the day for the household. The types of activities enjoyed in the evenings or weekends determine the amount of the house that has to be heated, lighted, and so forth. Such factors are called lifestyle effects.

(2) Then there are the reactions to changes in the environment, either social or physical, such as levels of temperature or humidity, seasonal changes in the number of hours

of darkness, or the occurrence of public holidays. These reactions are typically short-term in nature.

(3) The above two sets of factors interact with the basic characteristics of the household, such as the number of family members and their age distribution; the type, age, and size of the house or apartment; and the appliances used. Since most of these variables change infrequently, they will both interact with the short-run causes and will help explain long-term fluctuations in electricity usage.

The modeling approach we use is based explicitly on this three-tier breakdown of factors, but it is also conditioned by the data available to form the model. At the first stage of our analysis, electricity usage values for a household at different times of a day were regressed on a group of variables designed to capture the basic daily usage shape, plus changes in this shape from one day of the week to another, together with a number of short-run variables, including weather variables and the timing of school vacations. Some experimenting was undertaken to help determine precisely what variables should be included in these time-series regressions. These models were estimated for each household in a sample of 200. To capture the effect of individual household characteristics, the parameters of all of these time-series regressions were then regressed on the household variables. Thus, the final stage of the modeling process consists of cross-sectional regressions for each parameter of the time-series models, suitably normalized to reduce heteroscedasticity problems. The study has been carried out in two parts. In the first part, quarter-hourly data were used in the first stage of the analysis. The second part deals with the same data but aggregated to the hour and using more sophisticated econometric techniques.

The resulting models can be used for forecasting by (1) inserting the household characteristics of a sufficiently large random sample of residential users, thereby producing estimates of the time-series models for these households, (2) inserting predicted values of the explanatory variables in each equation, such as forecast weather values; and then (3) aggregating over the whole sample. For short-term forecasting, the household characteristics of an actual sample can be used, together with actual weather forecasts. For long-term forecasting, a prediction is needed of the kind of household characteristics. Data on the appliances used by each type of household and typical weather values for each day of the year can be inserted to achieve an aggregate forecast of future daily shape for any day and any region. It is assumed, of course, that households in regions other than Connecticut will behave in the same way as those actually observed in our sample. Some biasing effects may be noticed, so the model's forecasting ability would have to be first evaluated on past data before true forecasts are attempted. The method by which forecasts are formed from the model is described in more detail later.

Chapter 3

DATA USED IN THE STUDY

The data used in this study were collected by Northeast Utilities for approximately 400 households in Connecticut during the ten months from November 1975 through August 1976. Electricity usage data were collected every quarter hour for each household, together with matching weather data that included temperature (both wet and dry bulb), wind speed, and a measure of solar radiation. A fuller description of the data set is given in Burbank [2]. One of the most attractive features of these data (to an economist) is that half of the sample belonged to an experimental group who were exposed to a peak pricing method. The rest of the sample were a control group that had the usual methods of pricing. Thus, one of the results of our modeling will be to throw further light on the implications of a utility implementing a peak pricing billing method.

To an econometrician, the most noticeable feature of the data was its quantity, approaching 12 million numbers. This abundance meant that complicated regressions could be attempted, but it also produced unfamiliar problems. The size of the data bank meant that it was not possible to experiment with a large proportion of it. To reduce the computational burden, we selected several members of each of the experimental and control groups to try out various formulations of the models for the time-series regressions on portions of the time period. Even then, the length of the time series limited our ability to experiment because of cost and time constraints. We were also doubtful about the level of confidence to use in deciding whether an estimated parameter was significant or not. Somehow, using a 95% confidence interval with a time series of almost 30,000 terms did not seem completely appropriate. Kendall and Stuart [7] recommend using higher levels of confidence with large amounts of data, but as the statistical literature offers no precise recommendations on this point, we eventually retained the 95% confidence bands for t-statistics to help our decision making about which variables to retain in a model.

There were a few problems with the data. Some of the weather values were missing, and whenever this occurred the whole time period in the regression was ignored. In some months this led to a 5% or 6% reduction in the sample. As there was a certain amount of attrition of the households included in the sample, they could not all be used in the modeling process. This problem was much worse for the control group, where less than 50% of the households had electricity usage data for eight or more of the ten months and only 40% or less of the original sample were still producing data at the end of the period. The experimental group was very much better in this respect, with very few dropping out and only a few months of data missing, although it seems likely that the "households" involved did not necessarily contain the same family throughout. Because of these problems we eventually selected 85 households from each group, preferring those households with the most complete data. This may have resulted in a bias toward the less-mobile families living in the region, but it was hoped that the increased accuracy of the model would compensate for this acknowledged bias. Both the experimental and control groups were divided into five subgroups for the experiment, these subgroups being based on the level of electricity usage in the year prior to the start of the experiment. In our work we used 17 households from each of these subgroups to ensure a balanced model. As each subgroup does not necessarily represent exactly 10% of the total population, some care is needed in adding our models to get an aggregate forecast. A correctly weighted sum would be more appropriate.

Apart from these problems, the data seem to be of good quality, and most variables needed to build a sound model were available, except information about when a household was on vacation.

It would have been possible to reduce the amount of data used by converting to hourly usage and weather series. However, it was felt that the reaction to changes in variables such as sunlight or temperature could take place within a short time, and so time-aggregation of the data could result in somewhat misspecified relationships. The decision to first use quarter-hourly data does have the implication that the resulting R^2 statistics may well be lower for the time-series regression, as much of the very short-term movements in usage may be due to causes (such as decisions by family members to turn on or turn off various appliances) that may not be reflected by the explanatory variables used. As pointed out earlier, the second part of the analysis used hourly data. The data used in this part of the analysis are described later.

A number of sample plots of the usage data were made, and three of them are shown in Figures 3-1 through 3-3. The first figure shows the usage for four quarter hours equally spaced throughout the data for each day from November 1, 1975, to April 1, 1976, for a selected household (Household A). Figure 3-2 shows daily maximum quarter-hourly usage for the same time period but using a different household (Household B). Figure 3-3 shows the plot of usage for each quarter hour during the first seven days in November for Household B. Although some time patterns can be seen, the noise-to-signal ratio seems very high and so large R^2 values are not expected to arise from the time-series models constructed for each household.

In using the data, care was taken to allow both for missing values and for the hour change due to the start of summer season in Connecticut on April 25, 1976, at 2:00 a.m.

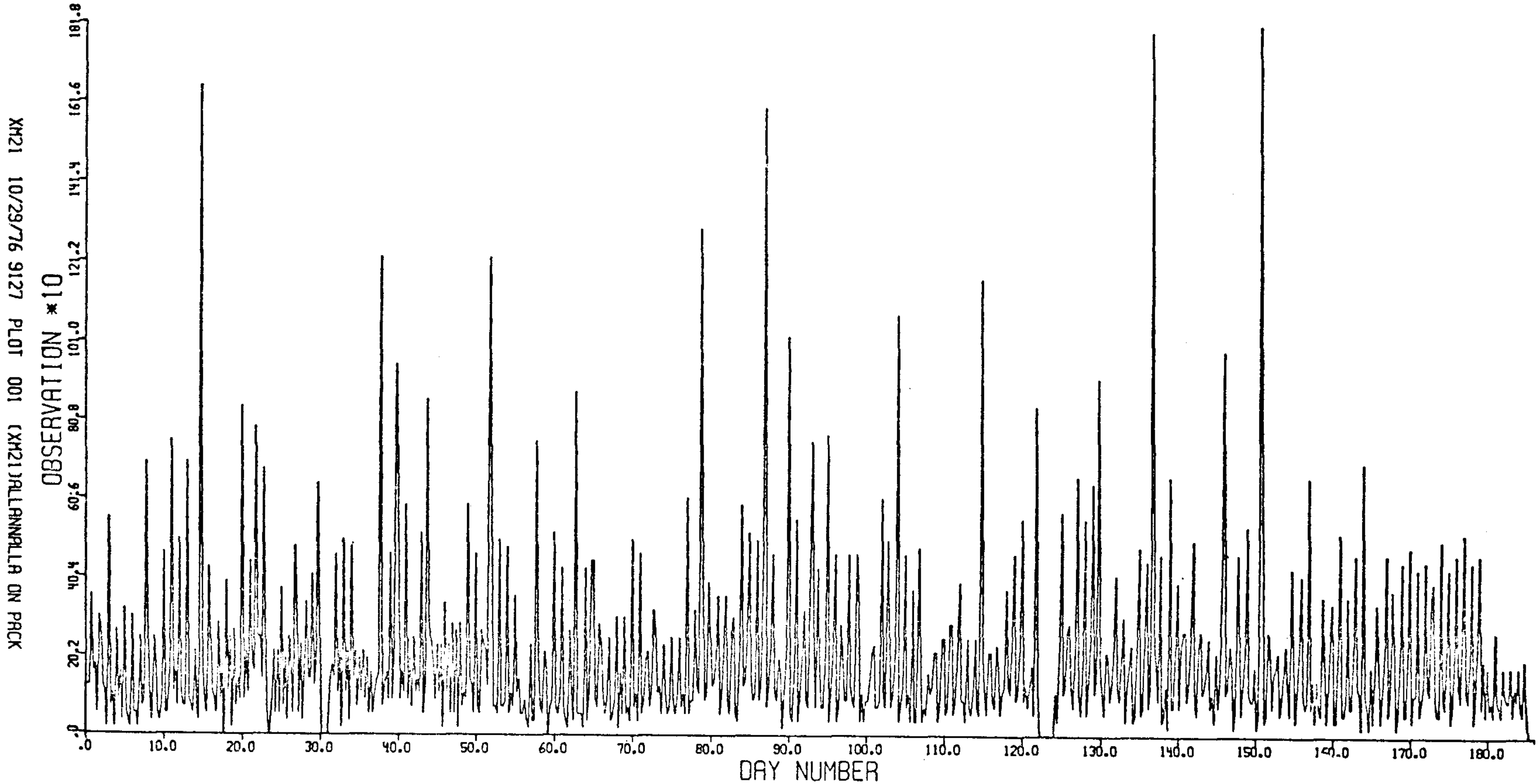


Figure 3-1. Quarter-hourly consumption, November 1975 - April 1976, Household A.

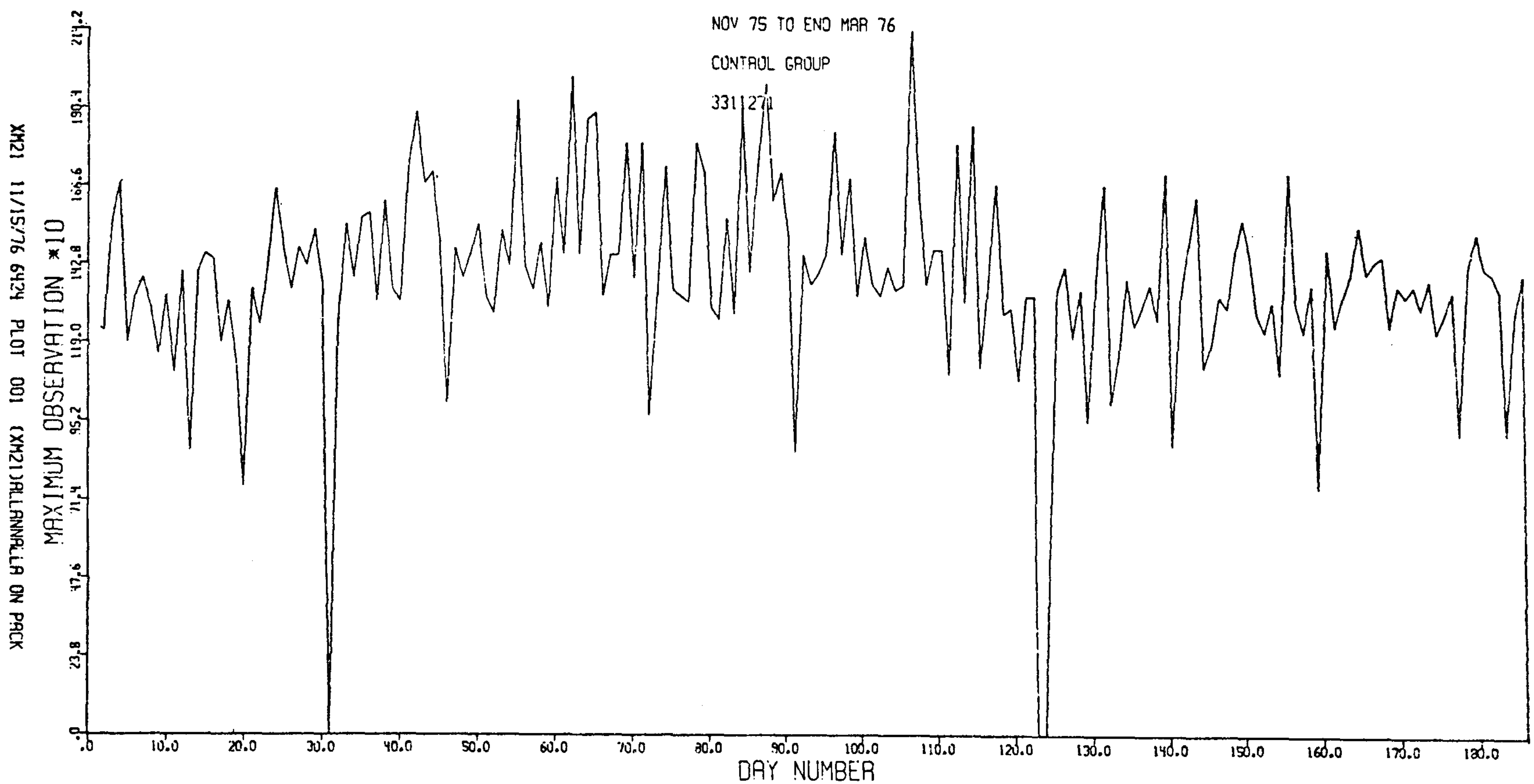


Figure 3-2. Maximum quarter-hourly consumption, November 1975 - April 1976, Household B.

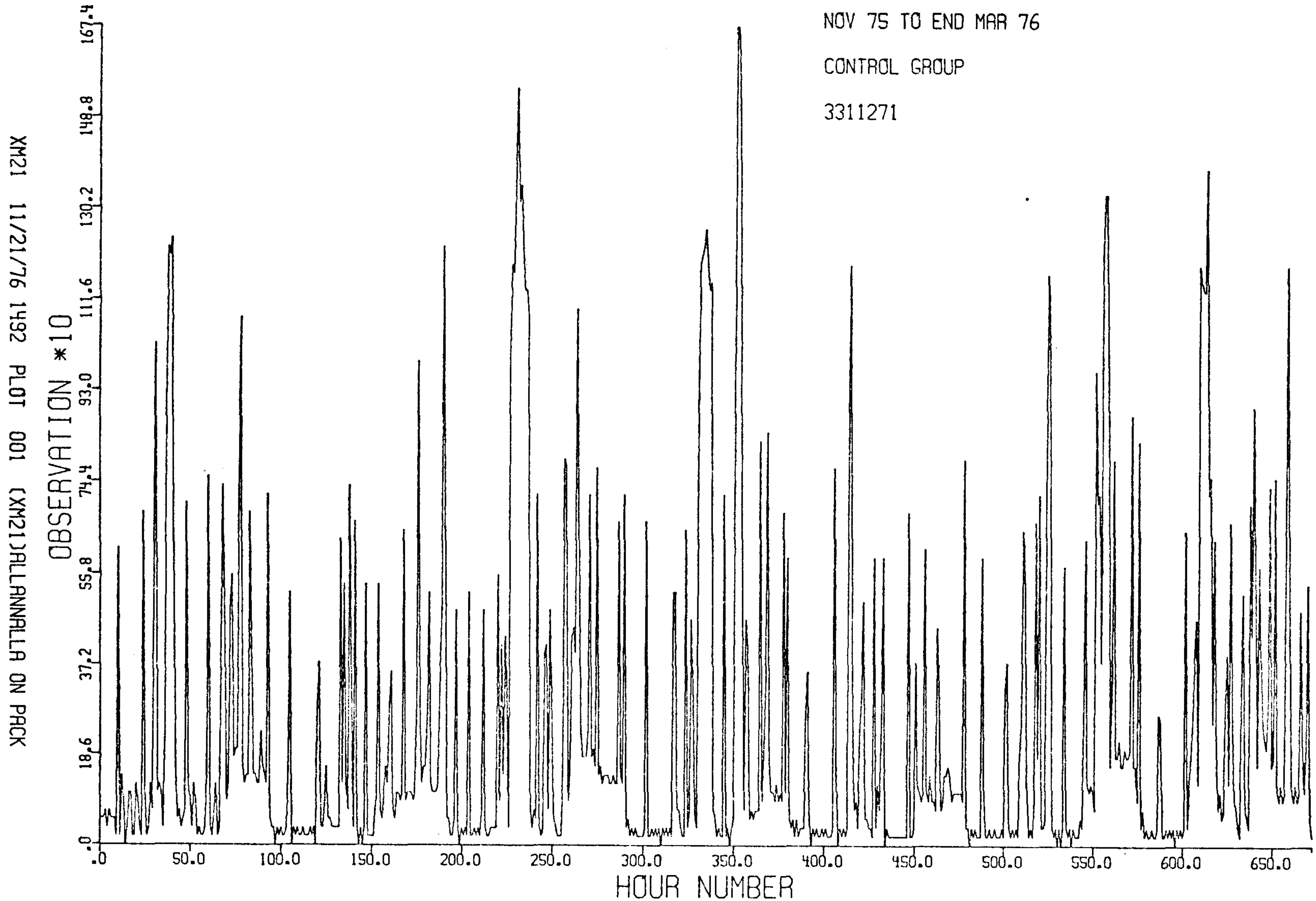


Figure 3-3. Quarter-hourly wage; November 1-7, 1975; Household B.

Chapter 4

ANALYSIS USING QUARTER-HOURLY DATA

Formulation of the Time-Series Model

The variables considered for inclusion in the time-series models to be fitted to the quarter-hourly usage data for each of the 200 households are described below.

Lifestyle Effects

To try to model the major regularities in electricity usage displayed by a household due to the habits or lifestyle requirements of its members, a set of daily dummies were used together with two alternative methods of representing movement with the day. One method is to use sine and cosine terms with 24-hour periods together with the first few harmonics, and the alternative is to use dummies for each hour of the day. The advantage of the first method is that it involves fewer parameters; but on the other hand, it may not be flexible enough to pick up all the details of the daily curve. Furthermore, the coefficients for hourly dummies are easily interpretable. The full list of variables considered here is then:

Variable	Description
CONST	Constant
SIN1	Sin wt, $w = 2\pi/12$
SIN2	Sin $\frac{wt}{2}$
COS1	Cos wt
COS2	Cos $\frac{wt}{2}$
MON-SAT	Daily dummies for Monday through Saturday, so that Tuesday dummy=1 on Tuesday, =0 on other days
HR2-HR24	Hourly dummies for each hour of the day except the first

It was necessary to leave out dummies for Sunday and for the first hour of the day to prevent exact colinearity, since these effects are included in the constant term. The coefficients for the hourly dummies measure the differences in usage from the first hour (midnight to 1 a.m.). The hourly dummies and the sines/cosines have to be alternatives, again to prevent colinearity. It would have been possible to consider using 96 quarter-hourly dummies, but the computing task was too formidable for our resources. The possibility of reducing the number of hourly dummies by leaving out those not significantly different from zero was also considered in the experiments discussed later. The weekends are allowed to exhibit a different shape through four harmonics, which are zero except for weekends.

Besides the daily dummies listed above, two other dummies were added, one for public holidays and one for school vacations. School vacations included Christmas break, Washington's birthday, and Easter break. (In New England these are prime times for families to brave the ice and frigid weather to go skiing.) Monthly dummies were not used because there seems to be no obvious reason why electricity usage should depend on which month one is in, apart from reaction to causes such as the timing of public holidays and school vacations and changes in temperature and other weather variables.

<u>Short-term Causal Variables</u>

The first causal variable considered was a sunlight dummy, taking the value 0 between one hour after sunrise and one hour before sunset and the value 1 otherwise. Two further dummies were used to pick up the effects of the pricing experiment being conducted on half of the sample households. The prices used in the experiment and the times and days the different prices were in effect are given in Burbank [2].

The two dummies used to represent this price effect were:

Peak dummy	= 1	when peak prices (16¢/kWh) were operating
	= 0	at other times
High dummy	= 1	when high-use prices (3¢/kWh) were operating
	= 0	at other times

These dummies were used for both the experiment and the control groups. If found to be significant for the control group, this would not indicate anything about price but would reflect an interaction between the daily shape and weekdays or holidays.

The quarter-hourly weather data made available to us were as follows:

WINDSP	- average wind speed (m/s)
AIRTEMP	- average air temperature (°C)
DEWPT	- average dew point (°C)
SOLRAD	- solar radiation average (cal/cm^2/min)

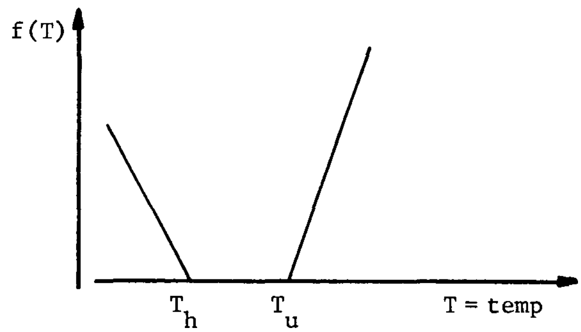
A measure of wind direction was also available, but we did not use it. The weather data were collected at two different weather stations in Connecticut. For each household the values allocated were those of the nearer station, with a constant correction factor added to the temperatures on the basis of isotherm maps for the state. The weather variables can be expected to affect electricity usage in a nonlinear fashion, so a review of the literature concerning the effects of weather on aggregate electricity demand for a utility is not entirely appropriate for a single household. However, the most satisfactory simple function for temperature appears to be of the form:

$$f(T) = k_u(T-T_u) \qquad \text{for } T > T_u$$

$$= 0 \qquad \text{for } T_h \le T \le T_u$$

$$= k_h(T-T_h) \qquad \text{for } T < T_h$$

which has the shape:



if $k_u > 0$ and $k_h < 0$, as expected. Thus, electricity usage will increase both when temperature gets low (due to heating) and when it is high (due to air conditioning). To capture this function, two variables were defined: TMIN, which is the upper leg of a piecewise linear temperature function; and TMAX, which is the lower leg. The dividing points were taken to be 10°C and 21.1°C (TMIN is bounded below and TMAX is bounded above).

The official temperature-humidity index is designed to measure human discomfort in the summer that results from the combined effects of temperature and humidity; it is defined as:

$$\text{THI} = \alpha(\text{wet-bulb} + \text{dry-bulb temperatures}) + \beta$$

for $T > 75°F$. Thus, adding DEWPT to the list of variables together with TMAX should allow for this effect. The widely used wind-chill factor (WCF) takes the form

$$WCF = (\alpha\sqrt{V} + \beta - V)\ (\gamma - T)$$

where V is wind speed. This measures the human discomfort of combined low temperature and wind speed, but is unclear whether a similar function is equally appropriate for both dwelling and the human body. The form of the function nevertheless suggests the possible use of the following variables in addition to WINDSP:

WINDHALF - the square root of WINDSP
TEMPWIND - the product of TMAX and WINDSP
MIX - product of TMAX and the square root of WINDSP

Although it was thought unlikely that all of these variables would be useful in the final model, they were included in our initial modeling experiments.

Some of the literature emphasized the cumulative effect of exceptional temperatures. We would have preferred to include a variable such as average temperature for the previous two days; but since the computing costs would have been too much, the values of TMIN and TMAX at the same time on the previous day were considered instead.

The above characterization of the weather is rather flexible, but it leads to confusion among several of the variables. Dew point, for example, behaves sometimes like humidity and sometimes like temperature. Furthermore, the specific form of the piecewise linear temperature function differs somewhat from the more attractive specification described earlier. In these regressions, the temperature is not constrained to be continuous. Thus we feel that although we are picking up the weather effects in the regressions, the coefficients are difficult to interpret, and we plan to change some of these variables in the next phase of the analysis.

All of these variables were first used in forming linear regressions to try to explain quarter-hourly variations in electricity usage for a single household. One obvious problem with doing this is that many possible interaction terms are ignored. Some interactions seem intuitively to be important, whereas others are less plausible. The linear addition of the daily shape represented by the hourly dummies, say, with the shape represented by the daily dummies means that it is assumed that the swings in usage within the day retain precisely the same <u>shape</u> each day but that different <u>levels</u> could occur from one day to another. However, it certainly seems likely that the shape will be different on weekdays (Monday through Friday) than on weekends and public holidays. This is an example of a possibly important interaction between the previously defined variables. On the other hand, it is not practical to allow complete freedom for the daily shape by defining hourly dummies differently for each day, say, since the resulting number of independent variables would be impractically large. As a compromise, the four sine and cosine variables were multiplied by a dummy that takes the value 1 on Saturdays, Sundays, and public holidays and 0 at other times. When used in conjunction with the hourly dummies, these new variables allow for a significant change in daily shape for days when members of the household do not go to work or to school. Some other interactions are less intuitively important, such as that between day dummies and the weather variables. Will a household react differently to a particularly low temperature on a Monday than on a Thursday? Possibly yes, but the number of interactions involving the previously defined variables occurring in pairs or in groups of three or more soon becomes too large to handle. It was therefore decided to exclude all other interactions from the model as presently constituted. The number of possible hypotheses and alternative models that could be considered is immense because of the plentiful data. Opportunities may arise to extend the present work, but for now we continue to concentrate on a long-term forecasting model.

A certain amount of experimentation was undertaken to test the relevance of the explanatory variables listed above. To cut down costs and time, five households each were selected from the experimental and control groups, with one household picked at random from each of the subgroups on the basis of previous usage levels. Regressions were fitted using a variety of explanatory variables both for the first two months of the full sample and for all Mondays in the first six months of the sample. The significance of individual parameter estimates was considered, together with some groups of estimates. The residual series were estimated and the residual autocorrelation sequence were examined together with averages for each quarter hour in the day, these averages being formed over the days used in the experiment. These experiments allowed a number of decisions to be made, hopefully strengthening the model eventually fitted to the full sample of households by improving its specification. The important decisions were as follows:

(1) It was decided to use the hourly dummies rather than trying to explain the within-day usage shape by sine and cosine

terms. The latter could not properly represent the shape in its full detail. Most of the hourly dummies were significantly different from one another, although the first few hours of the day were usually similar. It was decided to use all of the hourly dummies but also to use the weekend-daily shape interaction variables involving sines and cosines as described above.

(2) The weather variables were simplified by leaving out WINDHALF (square root of wind speed) and MIN (TMAX times WINDHALF); these variables were so colinear with other variables which were retained that those discarded seemed to be of lesser importance.

(3) Sunlight was left out because it was highly colinear with solar radiation, which was retained as being a more flexible variable.

These changes produced a final list of 46 explanatory variables, including a constant. The residuals from experiments using these variables generally had quarter-hourly averages that were not significantly different from 0, indicating that the daily shape in usage was being well captured, as might be expected. The autocorrelations of the residuals suggested low-order autoregressive models, usually AR(1) with parameters (first autocorrelation) in the region 0.3 to 0.6. In a few instances, the higher-order autocorrelations were small but showed an inclination to be positive, suggesting a (relatively) low frequency such as a trend, a long cycle, or possibly a weekly cycle. This could have been due to a learning process by the experimental group at the start of the pricing equipment, to missing interaction variables, or to changes in the household or its appliances. Although the results suggest some model misspecification, it should not be of overwhelming importance, and further consideration will be given to this property of the data. For some households there was a small, but possibly significant, autocorrelation at lag 96, corresponding to one day. The values of these correlations were under 0.1 and typically were about 0.007. They again suggest a minor misspecification in the model, probably due to some interaction term being left out, but they are likely to be of small economic significance. Although the residuals are autocorrelated, this was not allowed for in the regressions, which were estimated by ordinary least squares. We took comfort in the fact that these estimates should be unbiased; and although some efficiency was lost, this was of little importance given the length of the series being used. To improve the estimation procedure, it would have been necessary to include a lagged residual in the model, but the extra computing cost was prohibitive. The addition of a lagged residual would have no long-term forecasting implications, in fact virtually no implications beyond an hour or so, and thus it was thought reasonable to exclude such a term.

Time-Series Results

The final regression model chosen for the time-series regressions had 46 explanatory variables to describe the time of day, type of day, instantaneous and lagged weather effects, and some interactions. These regressions, which generally had more than 25,000 observations, were run for 155 households approximately evenly divided between the control and experimental groups, where the experimental group faced a three-tier, peak pricing structure. Because we chose to use such a large set of observations, the actual calculation of these regressions severely taxes our computational facilities, both through the construction of the cross-product matrix and the matrix inversion required for the regression coefficients and standard errors. For subsequent calculations we are investigating several more carefully optimized numerical procedures and also intend to reduce the scope of the sample.

An example of one of the regressions for a moderately large user in the control group who does not face peak prices is presented in Table 4-1. Notice first that both the peak and high variables are negative, suggesting that the user consumes less during the peak periods than indicated by the other variables in the regression. In particular, this individual uses less at the peak hours during the week than on the weekends, when the peak variable is turned off. It is, however, not surprising that the simple correlations of these variables with demand are positive.

The school vacation variable is significantly negative, suggesting that the household probably went on a vacation. Tuesday and Thursday are days on which significantly less electricity is used than on Sunday, while more is used on Saturday. Three of the harmonics are quite significant.

The hours tell a reasonable story. The household gets up before seven but doesn't achieve a morning peak until noon. The evening peak occurs at six o'clock and then tapers off until midnight, with a secondary peak before nine. The same pattern can be seen in the simple correlations between these variables and the dependent variable.

Table 4-1

REGRESSION COEFFICIENTS FOR A TYPICAL HOUSEHOLD

IN THE CONTROL GROUP*

Variable	Coefficient	t-stat	Variable	Coefficient	t-stat
PEAK	-0.034	-3.8	HR2	0.0007	0.12
HIGH	-0.011	-1.4	HR3	0.008	1.4
CONSTANT	0.050	7.5	HR4	0.023	4.0
PUBLIC HOLIDAY	0.001	0.2	HR5	0.024	4.0
SCHOOL VACATION	-0.042	-14.2	HR6	0.024	4.0
MONDAY	-0.004	-0.7	HR7	0.055	8.8
TUESDAY	-0.011	-2.2	HR8	0.055	8.8
WEDNESDAY	0.004	0.7	HR9	0.062	6.5
THURSDAY	-0.012	-2.3	HR10	0.050	5.1
FRIDAY	0.009	1.7	HR11	0.083	7.9
SATURDAY	0.014	4.5	HR12	0.105	9.8
SIN1	0.029	6.9	HR13	0.035	3.3
SIN2	-0.000	-0.01	HR14	0.020	1.9
COS1	-0.042	-8.8	HR15	0.048	4.5
COS2	0.046	17.9	HR16	0.052	5.0
WIND SPEED	0.011	35.6	HR17	0.112	10.7
DEW POINT	0.001	10.0	HR18	0.261	25.1
SOLAR RADIATION	0.029	7.0	HR19	0.169	16.5
TMAX	0.009	20.8	HR20	0.169	17.2
TMIN	0.0005	2.3	HR21	0.174	18.4
WIND•TMAX	-0.002	-20.7	HR22	0.126	20.4
TMAX (-96)	0.002	8.5	HR23	0.106	17.7
TMIN (-96)	-0.002	-9.6	HR24	0.021	3.5

$\bar{R}^2 = 0.0257$ SER = 0.136

---

*Temperature variables defined slightly different than in text.

The weather variables are frequently significantly different from 0; but, as pointed out above, it is hard to interpret them.

The analysis of 155 regressions, each with 46 variables, is a substantial job. Many of the results follow intuition, but there are also a variety of surprises. For example, the daily shape of some households is dramatically different from that of others, and some have strong peaks on certain days of the week while others do not. Whether these differences are systematic or merely sampling errors must await the cross-sectional regressions where these differences can be explicitly tested.

There are several characteristics of the time-series regressions that should be mentioned at this point. The expectation that the explanatory power is quite low is upheld, especially for the small users. In Table 4-2, the mean $\bar{R}^2$ are given for each use class. Class 1 consists of the smallest users in the previous year and Class 5 the largest. The explanatory power of the regressions for the small users is substantially lower than for the large users. This suggests that the nonrepeating or unpredictable component of electricity usage does not increase proportionately with use. Large energy-using appliances are likely to be used at a systematic time, and therefore the demand can be better explained by the regression.

In Table 4-3, the means of each of the regression coefficients across individuals are tabulated for the control and experimental groups. Notice first that for both groups the peak and high variables have negative coefficients. While this is to be expected for the experimental group, it might appear

Table 4-2

MEAN $\bar{R}^2$ OF TIME-SERIES REGRESSIONS FOR GROUPS

Size Class	Control	Experimental
1	0.156	0.216
2	0.176	0.146
3	0.260	0.237
4	0.256	0.305
5	0.351	0.394

Table 4-3

MEANS OF REGRESSION COEFFICIENTS IN BOTH GROUPS

Variable	Control	Experimental	Variable	Control	Experimental
PEAK	-0.031	-0.150	HR2	-0.021	-0.058
HIGH	-0.034	-0.089	HR3	-0.013	-0.085
CONSTRANT	0.138	0.342	HR4	-0.027	-0.089
PUBLIC HOLIDAY	-0.014	-0.050	HR5	-0.018	-0.078
SCHOOL VACATION	-0.082	0.016	HR6	-0.0003	-0.046
MONDAY	0.025	0.017	HR7	0.056	-0.020
TUESDAY	0.012	0.018	HR8	0.084	0.116
WEDNESDAY	0.003	0.017	HR9	0.110	0.174
THURSDAY	0.013	0.021	HR10	0.106	0.130
FRIDAY	0.002	0.020	HR11	0.096	0.130
SATURDAY	-0.001	0.0007	HR12	0.094	0.157
SIN1	0.004	0.007	HR13	0.080	0.140
SIN2	0.027	0.036	HR14	0.060	0.114
COS1	-0.004	-0.020	HR15	0.053	0.094
COS2	0.038	0.041	HR16	0.074	0.111
WIND SPEED	0.033	0.007	HR17	0.129	0.161
DEW POINT	-0.004	-0.008	HR18	0.194	0.212
SOLAR RADIATION	0.035	-0.092	HR19	0.224	0.210
TMAX	0.009	-0.002	HR20	0.225	0.221
TMIN	0.001	0.002	HR21	0.221	0.251
WIND•TMAX	-0.002	-0.0003	HR22	0.183	0.229
TMAX (-96)	0.002	-0.0002	HR23	0.129	0.198
TMIN (-96)	-0.008	-0.007	HR24	0.059	0.095

surprising for the control group. The implication, however, is simply that the peak is higher on weekends than on weekdays for both groups, and thus the peak variable is negative. The coefficients are more negative for the experimental group, as anticipated. The simple correlations between peak and use are generally, but not always, positive, indicating that use in peak periods is generally above the average. It is surprising that this is not more pronounced. One must wonder if a substantial portion of the system load peak is in fact due to residential customers.

The daily load shape is mainly captured by the hourly dummies. For both groups, this decreases from midnight until seven in the morining, reaches a peak between eight and nine, and then declines at midday. The evening peak rises at six and peaks at eight or nine and then falls off again toward midnight. The fact that the two daily shapes are so similar supports the notion that all the shift due to the peak pricing is captured in the peak and high variables.

Other interesting features are the nearly 0 means of the daily dummies, indicated that, on the average, not only are all weekdays roughly similar but so are weekends, at least in levels. Although some individuals have strong patterns among days, these average out in this sample. Public holidays, and especially school vacations, are generally negative, probably because the family goes away and electricity usage is drastically decreased.

The other variables in the regression have less-clear interpretations for their coefficients. The full explanation for the load pattern, however, rests on the ability of the cross-sectional regression to explain the variations in these coefficients across individuals.

Cross-Sectional Analysis

For each household we were given data on 60 household-specific variables, as recorded in September 1976, which was just before the complete set of electricity usage data started to be recorded. These variables included information on whether the house was in the experimental pricing group or not, how many of various appliances were owned, the number of members and age structure of the household, the type of heating system used, the type of structure, and the age and square footage of the home. Many of these variables were redundant or unusable because of colinearity or very rare occurrence in the sample. As an example of colinearity, dummy variables were given for "heating system replaced in past year" and "heating system not replaced in past year." Care had to be taken in selecting an appropriate set of household characteristic variables. The list we decided to use in the cross-sectional regressions was:

Variable	Description
0	1 if in experimental pricing group, 0 otherwise [For the following variables, the number owned by the household was used.]
1	Electric range
2	Electric self-cleaning oven
3	Electric dryer
4	Self-defrosting refrigerator
5	Manual-defrosting refrigerator
6	Freezer (self- or manual-defrosting)
7	Dishwasher
8	Black-and-white television
9	Color television

10	Humidifier
11	Dehumidifier
12	Window air conditioner [The following three variables also are measured as numbers.]
13	People in household (age under 18)
14	People in household (age 18-64)
15	People in household (age 65 and over) [The following three variables are 0-1 dummies; 1 is yes.]
16	Main heating system electric
17	Supplementary heating system electric
18	Electric water heating
19	Age of home
20	Square footage of home
21	Type of strucutre: single-family

For members of the experimental group, data were also available for the answers given to a comprehensive survey, containing a potential of almost a thousand questions, conducted in August 1975. The survey included detailed questions about the consumer's attitude to electricity usage and prices, how 30 different appliances were used, and other topics. Because these data were not recent, were very detailed, and were available for only half of our sample, we made no attempt to include them in the cross-sectional regressions at this point in the project.

A number of different strategies could be taken with the cross-sectional analysis, depending on how heteroscedasticity of the data is dealt with, which explanatory variables are used, and whether or not the experimental and control groups are pooled. Since the dependent variables are estimated coefficients from regression equations, they can be expected to have different standard deviations, so heteroscedasticity becomes important. The estimated standard deviation of a particular coefficient can be 10 times greater for some households than for others in our sample. The problem can be reduced by dividing both dependent and independent variables by the estimated standard errors while performing the cross-sectional regression, but a superior two-stage procedure can be based on an article by Hanushek [5]. The first stage forms the ordinary least-squares regression estimate, and the results can be used to form an Aitken generalized least-squares estimate. This procedure was used in Chapter 5, but the results presented below involve just ordinary least squares. The initial regressions in this part of the project use all of the 23 independent variables listed above for each set of coefficients coming from the time-series regressions. At a later stage we investigate subsets of the explanatory variables, including weighted averages of the household appliances. The results presented also pool the two types of customer, so that the experimental pricing effects appear only as an additive dummy. After the set of explanatory variables have been further condensed, it is hoped to repeat the cross-sectional regression for both groups of customers separately so that more sophisticated effects of the pricing experiment on electricity usage can be investigated.

Table 4-4 summarizes the results of the first set of cross-sectional regressions for selected variables, the dependent variables being the sets of time-series regression coefficients and the independent variables being the list of household characteristics given above. The table shows the adjusted R^2

Table 4-4

CROSS-SECTIONAL RESULTS FOR SELECTED VARIABLES

Dependent Variable	Adj. R^2	d	Main Significant Independent Variables
CONSTANT	0.53	1.66	Survey,[a] people 18-64, people over 65, electric heat, square footage
SIN1	-0.05	2.12	Electric heat (?), square footage
SIN2	0.32	1.91	Survey, electric stove, dishwasher, color television, humidifier (-), people under 18, people 18-64, electric water heater
COS1	0.09	1.94	Electric dryer (-), electric heat (-), supplemental electric heat
COS2	0.16	1.8	Electric stove (?), air conditioner (?), people under 18, electric heat
MONDAY[b]	-0.012	2.18	---
TUESDAY	0.01	2.07	Dishwasher, supplemental electric heat (?)
WEDNESDAY	0.00	2.12	Dishwasher, supplemental electric heat (?), square footage (-)
THURSDAY	0.04	2.15	Dishwasher, electric heat (?), supplemental electric heat
FRIDAY	0.00	2.24	Dishwasher
SATURDAY	0.05	1.84	Dishwasher, electric heat, square footage (-)
HR2[c]	0.196	2.12	Survey (-), people under 18 (-), people 18-64 (-1), people over 65 (-?)

[a]Survey = 1 if in experimental group, 0 otherwise.

[b]It should be remembered when interpreting these results that the time-series coefficient on the day dummies are all relative to Sunday.

[c]When interpreting hourly dummies, it should be remembered that they are all relative to electricity usage in the suppressed variable hour 1 (midnight to 1:00 a.m.). There is also some interaction with the sine and cosine terms on weekends and with the peak and high price variables on weekdays.

Table 4-4 (continued)

Dependent Variable	Adj. R^2	d	Main Significant Independent Variables
HR3	0.22	2.13	Survey (-), people under 18 (-), perople 18-64 (-), people over 65 (-)
HR4	0.20	2.18	Same as for HR3
HR5	0.16	2.16	Survey (-), people under 18 (-), people 18-64 (-1), people over 65 (-?), supplemental electric heat
HR6	0.13	2.17	Survey (-), people 18-64 (-), supplemental electric, electric water heater (?)
HR7	0.12	2.03	Survey (-), Man.-def. refrigerator (?), people 18-64 (-), electric water heater
HR8	0.14	2.64	Electric stove, people 18-64 (-), electric heat, electric water heater
HR9	0.18	1.89	Dishwasher (?), humidifier (-?), electric heat, electric water heater
HR10	0.16	1.97	Humidifier (-), electric heat (?), electric water heater
HR11	0.17	1.87	Humidifier (-), electric heat (?), supplemental electric heat (?), electric water heater
HR12	0.15	1.91	Electric heat, electric water heater, one-family dwelling
HR13	0.12	1.86	Home (-?), people over 65 (-?), electric water heater
HR14	0.08	1.89	People over 65 (-?), supplemental electric heat (?), electric water heat
HR15	0.04	1.89	People over 65 (-), supplemental electric heat, electric water heater
HR16	0.05	1.97	Same as for HR15
HR17	0.08	2.00	Same as for HR15
HR18	0.13	2.03	Electric range (?), dishwasher, electric water heater

Table 4-4 (continued)

Dependent Variable	Adj. R^2	d	Main Significant Independent Variables
HR19	0.22	2.09	Dishwasher, color television, people over 65 (-?), electric water heater
HR20	0.30	2.10	Electric stove, dishwasher, air conditioner (-?), people over 65 (-), electric heat, electric water heater, square footage
HR21	0.31	2.02	Survey (?), dishwasher, people under 18, supplemental electric heat (?), electric water heater, square footage
HR22	0.25	1.96	Survey, electric dryer, dishwasher, dehumidifier (?), people under 18 (?), electric water heater, square footage
PUBLIC HOLIDAY	0.23	1.94	Survey (-), electric heat (-), supplemental electric heat (-)
SCHOOL VACATION	0.49	1.12	Survey, self-def. refrigerator
PEAK PRICE	0.24	2.07	Survey (-), dishwasher (-), people 18-64 (-), electric heat (-)
HIGH PRICE	0.14	1.97	Survey (-), dishwasher (-), electric heat (-)
WIND SPEED	0.62	1.07	Survey (-), dehumidifier (-), electric heat, square footage
DEW POINT	0.79	1.70	Survey (-), color television, air conditioner, electric heat (-), supplemental heat (-), square footage (-)
SOLAR RADIATION	0.44	1.27	Survey (-), freezer (?), electric heat (-)

value (henceforth R^2 is always adjusted for degrees of freedom), the Durbin-Watson statistic (d) and those independent variables which appear to be significant, that is, those with t-values over 1.96. A variable is underlined in the table if $|t| > 3$ and is given an query if the t-value is suggestive but not strictly significant (i.e., $|t| > 1.6$). A negative sign in parentheses indicates the coefficient in the cross-sectional regression is negative. Although the data are not in time-series form, the Durbin-Watson statistic does have some interpretative value, as the households were ranked approximately in order of level of electricity usage, first in the experimental group and then in the control group. Thus, a significant level for d would suggest a relationship with level of usage that has not been picked up by the independent variables.

In general, the cross-sectional results seem to be very promising. For some important variables the R^2 are adequate to good, the d-values do not indicate any serial correlation, and most of the significant explanatory variables make economic sense. Some of the more important interpretations possible from the table are as follows:

(1) The sines and cosines come in with mixed significance. There is an indication that the within-day shape of aggregate electricity usage is different for weekends than for the rest of the week, but in our final model we expect to try to pick this up by variables specified differently.

(2) The day dummies come in with very small R^2 values. Although the means for these variables suggest that there might be day-to-day differences in the aggregate, and there certainly are significant differences for individual families, the household

characteristics being used cannot explain these differences.

(3) The hour dummies contain many interesting results. Although the R^2 values are not very high in some cases, there is evidence that certain appliances are important at different times of the day. Other appliances that might be thought to be of importance but did not appear might be due to relatively low electricity use by some appliances or to relative low saturation levels for some and very high saturation levels for others.

(4) The results for school vacations are difficult to explain, and further analysis is planned. The remarkably low d-value suggests effects due to level of electricity usage that have not been accounted for.

(5) The weather variables often have very respectable R^2 values, but some have low d-values. As we were not satisfied with the specification of the temperature variable used in our preliminary model, because of multicolinearity and nonlinear effects, the time-series coefficients of some of these variables are difficult to interpret, as they are for the cross-sectional regressions. Alternative specifications were considered and incorporated into our final model.

To economists, the most interesting variables are probably peak and high prices, and so the cross-sectional results for these variables plus "constant" will be discussed in more detail than the rest. The results from the full-scale cross-sectional regression, as summarized above, suggested that some amalgamation or respecification of the explanatory variables was worth considering.

To study some possible important cross effects, new variables were defined as follows:

If survey ≡ 1, if household is in experimental group

0, if household is in control group

then:

Sur sq heat ≡ survey × electric heat dummy × square footage of house

Sur dry ≡ survey × no. of electric dryers

Sur dish ≡ survey × no. of dishwashers

Sur wat ≡ survey × electric water heating dummy

Sur sup ≡ survey × supplementary electric heating dummy

Sq heat ≡ square footage of house × electric heat dummy

People ≡ total size of household

Appliance ≡ weighted average of number of appliances in household, weights given by national average electricity usage of appliances (in annual kilowatt hours)

The detailed definition of this last variable is:

appliance = (electric range × 1200) + self-defrosting refrigerator × 1620) + (manual-defrosting refrigerator × 1200) + (freezer 1500) + (black-and-white television × 140) + (color television × 350)

Table 4-5 shows the estimated coefficients and |t|-statistics for the peak price, high price, and constant values from the time-series regression for both (1) the original list of household characteristics and (2) the revised list after eliminating some characteristics.

It is seen that the second formulation of the independent variables both improves the R^2 values and explains the methods by which members of the experimental group react to the peak and high prices. As expected, they reduce appliance use, particularly electric heating and dishwashing. The second set of regressions indicate that the independent variables could be

Table 4-5

CROSS-SECTIONAL RESULTS WITH INTERACTIONS

Independent Variable	Peak Price		High Price		Constant	
	(1)	(2)	(1)	(2)	(1)	(2)
CONSTANT	0.16 (1.92)	0.12 (1.73)	0.07 (1.33)	0.05 (0.95)	-0.23 (2.67)	-0.17 (2.20)
SURVEY	-0.14* (4.83)	0.006 (0.10)	-0.06* (3.37)	0.02 (0.38)	0.22* (7.4)	0.12 (1.83)
ELECTRIC RANGE	-0.003 (0.08)		-0.01 (0.38)		0.03 (0.75)	
ELECTRIC STOVE	0.05 (1.22)	0.02 (0.35)	0.02 (0.80)	0.01 (0.16)	-0.04 (1.1)	-0.02 (0.41)
SELF-DEF. REFRIG.	-0.04 (0.75)		-0.01 (0.40)		-0.01 (0.12)	
MAN.-DEF. REFRIG.	-0.05 (1.1)		-0.02 (0.77)		-0.07 (1.54)	
FREEZER	-0.01 (0.23)		-0.006 (0.2)		0.004 (0.09)	
DISHWASHER	-0.07 (-2.28)	-0.02 (0.45)	-0.04* (2.0)	0.01 (0.34)	0.02 (0.47)	0.02 (0.43)
BLACK AND WHITE TELEVISION	-0.01 (0.9)		0.003 (0.18)		-0.005 (0.20)	
COLOR TELEVISION	-0.03 (0.22)		-0.01 (0.59)		-0.007 (0.21)	
HUMIDIFIER	0.008 (0.22)		0.004 (0.18)		0.01 (0.25)	
DEHUMIDIFIER	-0.03 (1.1)		-0.02 (1.1)		0.005 (0.17)	
AIR CONDITIONING	-0.01 (0.62)	-0.01 (0.75)	-0.006 (0.51)	-0.01 (0.78)	-0.01 (0.66)	-0.01 (0.54)
PEOPLE < 18	-0.02 (1.66)	0.02 (1.32)	-0.006 (0.80)	0.01 (0.93)	0.02 (1.32)	0.03 (1.51)
PEOPLE 18-64	-0.04* (2.18)		-0.01 (1.2)		-0.05* (2.80)	
PEOPLE > 65	0.01 (0.37)	0.03 (1.39)	-0.02 (0.64)	0.03 (1.24)	0.08 (2.02)	0.03 (0.88)

Table 4-5 (continued)

Independent Variable	Peak Price (1)	Peak Price (2)	High Price (1)	High Price (2)	Constant (1)	Constant (2)
ELECTRIC HEAT	-0.09* (2.13)		-0.06* (2.07)		0.32* (7.31)	
SUPPLEMENTAL ELECTRIC HEAT	-0.09 (1.47)	-0.02 (0.28)	-0.07 (1.5)	-0.05 (0.84)	-0.04 (0.54)	0.02 (0.24)
ELECTRIC WATER HEATER	-0.05 (1.60)	-0.03 (0.76)	-0.03 (1.35)	-0.01 (0.46)	0.01 (0.24)	0.01 (0.23)
AGE OF HOUSE	-0.0005 (1.14)		-0.0004 (1.22)		0.0001 (0.23)	
AREA OF HOUSE (SQUARE FOOTAGE)	0.26E-04 (1.04)		0.86E-05 (0.50)		0.91E-04* (3.34)	
SINGLE-FAMILY HOUSE	0.05 (0.96)		0.03 (0.83)		0.04 (0.77)	
SUR SQ HEAT		-0.14E-03* (3.45)		-0.89E-04* (3.24)		0.20E-03* (4.55)
SUR DRY		0.02 (0.27)		0.01 (0.18)		0.04 (0.54)
SUR DISH		-0.12* (2.18)		-0.07 (1.75)		-0.01 (0.16)
SUR WAT		-0.03 (0.54)		-0.03 (0.70)		0.01 (0.21)
SUR SUP		-0.17 (1.46)		-0.05 (0.66)		-0.10 (0.79)
SQ HEAT		0.17E-04 (0.57)		0.86E-05 (0.44)		0.92E-04* (2.91)
PEOPLE		-0.03* (2.07)		-0.01 (1.22)		0.05 (3.25)
APPLIANCE		-0.24E-04 (1.70)		0.15E-04 (1.6)		0.34E-04* (2.26)
R^2	0.35	0.43	0.26	0.35	0.60	0.64
Adj. R^2	0.24	0.36	0.14	0.27	0.54	0.60
d	2.07	2.42	1.97	2.28	1.66	2.02

much reduced in number without any appreciable effect on the degree of explanation achieved.

Figure 4-1 shows the variation in effect of peak prices on the experimental group, as usage levels increase. It shows the plots of the estimated time-series regression coefficients for peak prices against customer identification number in both the experimental and control groups. The identification numbers approximately indicate usage levels: low numbers correspond to low usage and high numbers to the highest usage levels. The dependent variables clearly exhibit heteroscedasticity, and the figure also indicates a great variability in the reaction of households to the peak pricing system.

Using the Model to Forecast

The model has been constructed in two stages, which may be characterized as follows.

Stage 1:

$$U_{j,t} = \sum_k \beta_{j,k} X_{k,t} + e_{j,t} \tag{4-1}$$

where $U_{j,t}$ is the electricity usage by family j at time t and the $X_{k,t}$ are the explanatory time series, such as the hour dummies and the weather variables described earlier.

Stage 2:

$$\beta_{jk} = \sum_i \gamma_i H_{i,j} + g_{j,k} \tag{4-2}$$

where the $H_{i,j}$ are the household characteristics for family j (such as the number of children, number of color television sets, size of the house); they are also dummy variables, taking value 1 if they have electric heating and 0 otherwise; g_{jk} is a

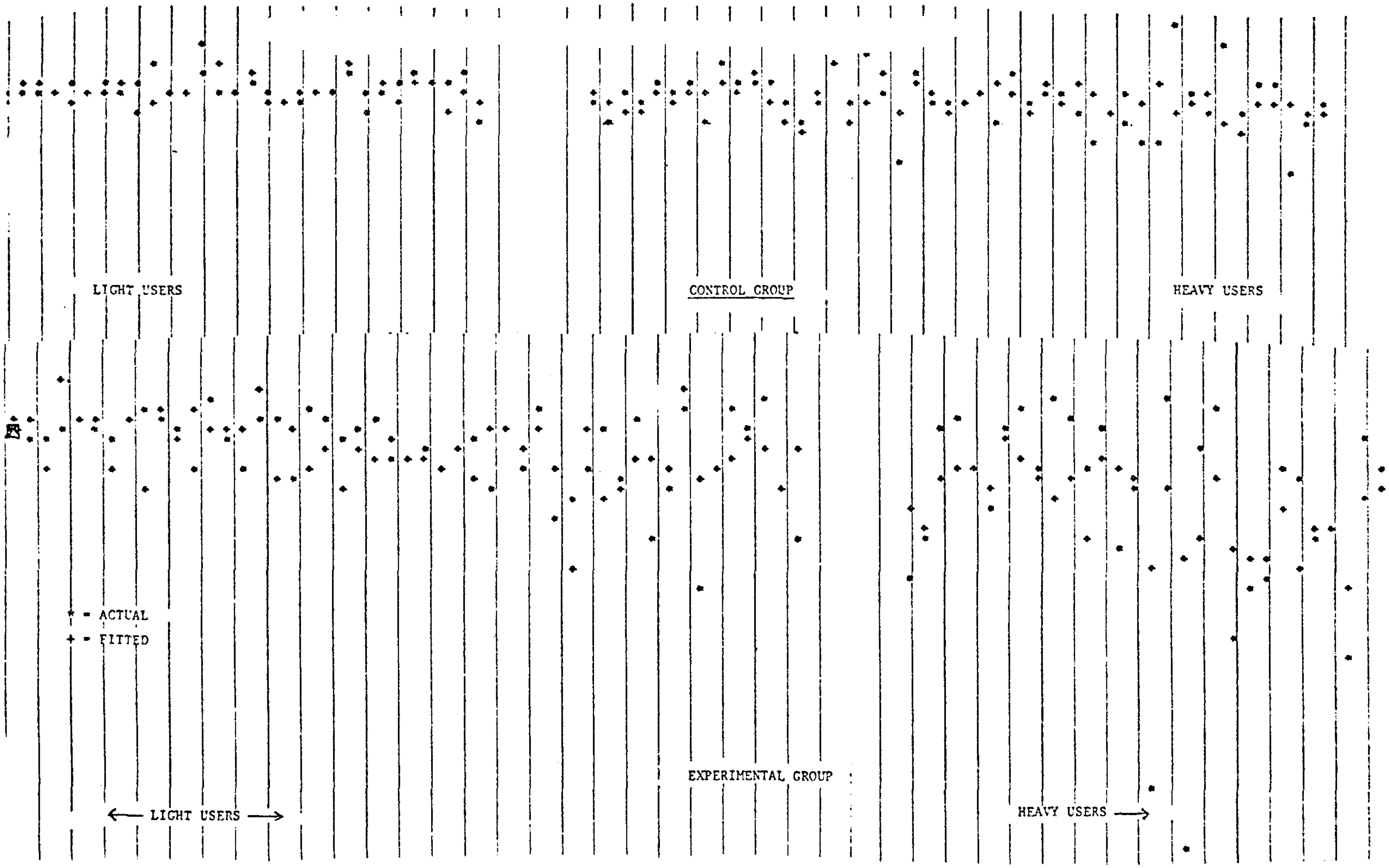


Figure 4-1. Peak price time-series coefficients versus household identification number.

disturbance.

The total usage for the sample will then be

$$U_t^{(s)} = \sum_j U_{j,t}$$

$$= \sum_k (\sum_j \beta_{j,k}) X_{k,t} + e_t$$

where $e_t = \sum_j e_{j,t}$

and so

$$U_t^{(s)} = \sum_k [\sum_i \gamma_{ik} (\sum_j H_{ij})] X_{kt} + \sum_k (\sum_j g_{j,k}) X_{k,t} + e_t$$

As the conditional expectation of $g_{j,k}$ and e_t will be 0, the best forecast will be obtained by setting the last two terms on the right-hand side equal to 0. If the sample contains N_S families and $\bar{H}_i$ represents the average value for the $i^{\underline{th}}$ household characteristics variable, the estimated sample usage formula becomes

$$U_t^{(s)} = N_S [\sum_k (\sum_i \gamma_{ik} \bar{H}_i) X_{k,t}] \qquad (4\text{-}3)$$

The corresponding estimate for the total residential electricity use in region $U_t^{(\gamma)}$ will be as, in Equation 4-3, but with N_S replaced by N_T, the total number of families in the region.

For short-term forecasting, a sample of families in the region will supply the values for the $\bar{H}_i$, the results of the previous section will supply estimates of the γ_{ik}, and so only forecasts of the $X_{k,t}$ need to be inserted to obtain a forecast of residential usage. Most of the $X_{k,t}$ can be forecast without error, such as the daily and hourly dummy values for peak and high prices, the school and public holiday dummies, and the weekend cosine terms. Forecasts of weather values will be required, but they are usually available from the local weather office. It is possible, however, that weather data are not available for every quarter hour, as some weather offices supply only forecasts of daily low and high temperatures without any indication of when in the day they may occur. This leads to a weather forecasting question that is outside the scope of this paper but not unsolvable. The error series may also contain some time-structure and so would need analysis by the usual Box-Jenkins single-series modeling techniques, from which forecasts can be easily obtained (see, for example, Granger and Newbold [3]).

For longer-run forecasting, projected values for the $\bar{H}_i$ need to be obtained and inserted in the model, possibly weighted to allow for any changes in efficiency of appliances, say, together with long-run forecasts of the time-series variables. For the weather variables this will involve using "normal" values for each day and hour of the day.

It is planned to evaluate many aspects of the model in the final stages of our project. There is, however, one obvious problem with using the model for forecasting in regions other than that for which the data were collected. It is quite possible, for example, that residents in the Midwest react differently to low temperatures than those in New England or in California, say, because their houses may be insulated more efficiently or because they are more used to extreme temperatures. This could lead to a bias in the coefficient on temperature in the forecasting model, but analysis of "forecasting" errors based on past data should allow compensation for such a bias. A further problem arises if a peak-load-pricing scheme is implemented in a region using a different pricing structure than that used in the Northeast Utilities experiment. There are a number of other studies looking into aspects of this question, and it is hoped that their results can be incorporated into this model.

It might be worthwhile concluding with a discussion of why we feel that the modeling procedure that we have used is at least potentially better for long-term forecasting than some alternatives. The method used attempts to take into account many of the causes of changes in residential electricity usage, especially over several seasons, and so it is both sophisticated and comprehensive, although doubtless many improvements are possible and further development is required. The cross-effects of appliances on short-term causal variables have been explicitly modeled. It would not have been possible to estimate a model such as Equation 4-3 by using just a single aggregate use series, so the availability of household data has been explicitly recognized. An alternative procedure would have been to try to model the regular components of the household demand series by using Box-Jenkins techniques, say, and then to have related the coefficients and residuals from these models to the data on household characteristics. Although this may well prove to be a feasible and valuable method for short-run forecasting, we doubt its usefulness over the longer term. The time-series models would try to pick up the daily and weekly usage shapes by daily and weekly differencing, or by using autoregressive models with very long lags (see, for instance, the paper by Uri and Parzen in Boyd [1]). The actual daily and weekly shapes in these models are determined by a set of startup values acting together with the model. However, our experience is that complicated but fairly stable shapes are not well forecast by this technique over the middle or long-run, since the forecast shape is inclined to drift away from the true shape. The procedure is also more disturbed by exceptional periods of usage than would be our model and its forecasts. Although by no means perfect, we do believe that the model analyzed in this section is a sensible one making use of most of the very considerable amount of data made available to us.

Chapter 5

ANALYSIS USING HOURLY DATA

Introduction

This chapter develops a model of hourly demand for electricity by individual households. As before, the main objective of the model is to isolate and evaluate the importance of various factors and to produce a model that will be useful in forming medium-term forecasts of time-of-day demand. A medium-term forecast is understood to be for a time period when the demographic and appliance stock variables are known and the mean weather for that period is also known.

A second purpose of the model is to investigate in detail one of the factors: prices. Half of the sample population was subjected to time-of-day pricing with very high prices during two hours in the morning and two hours in the evening. The model endeavors to determine the extent of shift in the load curve, whether there is evidence for cross-elasticities between different hours, and whether it is possible to distinguish between households in their propensity to shift.

It is clear that many of the behavior patterns will not be easily described as a function of the demographic variables. For example, the time of rising in the morning is a habit that one might never expect to predict from the demographic variables. That a household does the laundry on Thursday would similarly be unpredictable. Furthermore, even if all mealtimes are the same every day, the fact that for some meals a full dinner is prepared (with hours of oven time) while other meals consist of takeout foods, casts doubt on the ability of the model to distinguish between households with electric and gas stoves. Only events that are very regular and that are predictable from the demographic variables can be observed. In short, the expectations for the fits in both the first and the second stage are not very high.

In this second part, two separate analyses were undertaken, one for three winter months and one for two summer months. These periods were analyzed separately because the determining factors are very different. First, the peak pricing times are different, and it would be surprising if the response were the same for each period. Second, the major response to weather in the winter is the use of heating equipment; while in the summer, it is cooling. By separating the two it is possible to study conditions more like the experiment of the Southwest in the summer and that of the North and the East in the winter. Third, summer is almost entirely school vacation; while there are only two weeks of school vacation in the winter. Finally, swimming pool pumps, dehumidifiers, and air conditioners are used only in the summer; while electric heating and supplemental electric heaters are used only in the winter.

Econometric Considerations

As described earlier, the prediction and estimation of the use of electricity by a residential household proceeds in two steps, each of which is a linear regression. In the first stage, hourly usage for household i, denoted y_i, is regressed upon K time-series causal variables that reflect the time of day and the weather, X_i. There are T observations in each vector. The time-series properties of the load curve are therefore summarized by the behavior of the mean vector that depends on the K unknown regression coefficients β_i. These regressions can be written as

$$y_i = X_i\beta_i + \varepsilon_i \qquad i=1,\ldots,N \qquad (5\text{-}1)$$

for the N households in the sample.

The second regression relates the K parameters of the load curve to its demographic determinants. Let β_j denote the vector that is made up of the value of the j^{th} regression coefficient for all N households. Similarly, let Z_j represent the N×L matrix of demographic determinants of β_j. Then the second-stage cross-sectional relationship can be written

$$\beta_j = Z_j\gamma_j + \eta_j \qquad j=1,\ldots,K \qquad (5\text{-}2)$$

Once the γ_j are known, then the β_i can be constructed for a household with known demographic characteristics. This then allows prediction of usage by this household over time, since the X_i are known constants or weather variables.

Several econometric issues arise when estimating such a set of regression equations. The particular solution of preference depends on the distributional properties of the disturbances and on the dimensions of the problems If none of disturbances can be assumed to have zero variance, the problem can be seen to be a very large linear regression. The dependent variable would be the stacked vector of N households with T observations on each. This generates a vector of dimensions NT which, for our winter sample period, would be more than 300,000 observations. If there are L variables in each Z_j, then there would be L composite variables of dimension NT, each of which is found by multiplication of X's and Z's. While this regression problem is complicated, the computation difficulty is exacerbated by the nonscalar covariance matrix of the disturbances. There would surely be different variances in the household regressions and there would be components related to the X's.

Because of the particular nature of the variables and the dimensions of our problem, a two-step procedure was more attractive. This involved estimating Equation (5-1) for each household using ordinary least sqaures (OLS) and then using the regression coefficients as dependent variables in Equation (5-2). Model (5-1) is therefore a random-coefficients model for which OLS is unbiased and consistent but not asymptotically efficient (see, for example, Hildreth and Houck [6]). This loss of efficiency and bias of standard errors may have some repercussions, but for our sample sizes these costs seem small.

In the analysis with quarter-hourly data, even the evaluation of the least-squares regressions in Equation (5-1) posed a substantial computational burden. For each household large matrices $(X_i'X_i)$ had to be constructed and inverted using standard least-squares algorithms. Two important by-products of the decision to look at winter and summer separately and a computational improvement enormously decreased the computational effort.

By examining winter and summer separately, it was possible to diminish the size of the X_i matrices because weather variables could be tailored to the particular season. Also, some variables were omitted because of insignificant effects in the previous analysis. More important, however, all households included had complete records for the sample period and all weather variables were complete for the period. Linear adjustments to the reported weather were the only corrections necessary to produce weather appropriate to a specific location. Hourly dummies, school vacations, and peak prices were common; and the X_i matrices became the same for all households (i.e., $\equiv X$). The inversion of $X'X$ was needed only once for all 140 regressions. For each household, only the $X'y_i$ needed to be constructed. Drawbacks to this procedure are : (1) some restriction of the sample to households with continuous records and located near the principle weather station; (2) some difficulty interpreting the constant term, since this would include the temperature adjustments; and (3) failure of the nonlinear temperature terms, which may be attributed to the ommission of recommended locational adjustments.

A substantial computational improvement was achieved by solving the normal equations $X'X\beta_i = X'y_i$ using the Cholesky square-root decomposition of the matrix $X'X$ (Graybill [4]). The Cholesky decomposition consists of finding a unique triangular matrix T such that $T'T = X'X$. The solution of the normal equations, $T'T\beta_i = X'y_i$, can then be done in two steps: get $T\hat{\beta}_i = T'^{-1}X'y_i$, and then $\hat{\beta}_i = T^{-1}T'^{-1}X'y$. The advantage of this procedure over the conventional method is that inversion of a triangular matrix is substantially simpler and more accurate than inverting the full matrix $X'X$.

Estimation of the cross-sectional regressions in Equation (5-2) also presents some econometric questions. Even if the η_j are assumed to have scalar covariance matrices, the dependent variables will be measured with error since estimated coefficients are used rather than the true coefficients. Since the variances of these measurement errors are known from the least-squares estimates of Equation (5-1), a generalized least-squares procedure can be implemented.

Letting $\beta_j - \hat{\beta}_j = -\xi_j$, Equation (5-2) becomes

$$\hat{\beta}_j = Z_j \gamma_j + \eta_j + \xi_j \qquad j=1,\ldots,K \tag{5-3}$$

Assuming $E(\eta_j \xi_j') = 0$, $E(\eta_j \eta_j') = \sigma_j^2 I$, and $E(\xi_j \xi_j') = D_j$ where D_j is a diagonal matrix with the estimated variance of regression coefficient j for each of the N households on the diagonal, the covariance matrix of the disturbances is simply $\Omega = \sigma_j^2 I + D_j$. The one unknown parameter σ_j^2 can be estimated from the least-squares residuals of Equation (5-3) following Hanushek [5]. Letting e_j be the least-squares residuals from Equation (5-3), it can be shown that

$$E(e_j' e_j) = (N-1)\sigma_j^2 + tr[D_j + (Z_j{,}Z_j)^{-1}(Z_j{'}D_j Z_j)] \tag{5-4}$$

and therefore an unbiased estimator of σ_j^2 is

$$\sigma_j^2 = \frac{e_j' e_j - tr[D_j + (Z_j' Z_j)^{-1}(Z_j{'}D_j Z_j)]}{N-L} \tag{5-5}$$

This procedure is easy to follow, although evaluation of the second term in the trace is not computational trivial.

In this study there are approximately 30 cross-sectional regressions in each of the two seasons; hence this procedure requires substantial computational effort. However, because of the fact that each time-series regression has the same matrix of regressors, X, and each cross-sectional regression has been chosen to have the same demographic determinants, Z, the procedure can be vastly simplified. The i^{th} element of D_j can be rewritten as

$$(D_j)_{ii} = \alpha_j s_i^2 \qquad (5\text{-}6)$$

where α_j is $(X'X)^{-1}_{jj}$ and s_i is the standard error of the regression for the i^{th} household. Placing s_i^2 on the diagonal of a matrix S, the estimator (Equation (5-5)) can be rewritten as

$$\hat{\sigma}_j^2 = \frac{e_j{}'e_j - \alpha_j[\sum_i s_i^2 + \mathrm{tr}(Z'Z)^{-1}(Z'SZ)]}{N-L} \qquad (5\text{-}7)$$

For each choice of Z and X, the trace is only evaluated once and the estimation of $\hat{\sigma}_j^2$ is simply accomplished.

The matrix $\hat{\sigma}_j^2 I + D_j = \hat{\Omega}_j$ is a consistent estimator of the disturbance covariance matrix, and thus generalized least squares (GLS) will be asymptotically efficient.

$$\hat{\gamma}_j = (Z'\hat{\Omega}_j^{-1}Z)^{-1}Z'\hat{\Omega}_j^{-1}\hat{\beta}_j \qquad j=1,\ldots,K \qquad (5\text{-}8)$$

The importance of the GLS correction will differ from equation to equation, depending on the size of $\hat{\sigma}_j^2/\alpha_j$ relative to S.

<u>Time-Series Regressions</u>

The data available for the time-series regressions consist of quarter-hourly usage figures from 140 households over a 10-month period, together with matching weather variables. As described in Chapter 4, regressions were run using all of the data, which ran to several million terms. In the present version, all variables have been aggregated to hourly figures and separate regressions are run for a winter period, consisting of December 1973, January and Februray 1976, and a summer period of July and

and August 1976, excluding periods immediately around major public holidays. Samples of the hourly usage data were plotted and a very high noise-to-signal ratio was apparent, so high R^2 values for the time-series regressions were not expected, or achieved.

The explanatory variables

To model the major regularities in electricity usage displayed by a household due to habits or lifestyle requirements of its members, hourly dummies were used, but omitting the hours 1 a.m. to 4 a.m. The hours omitted are typically minimum-use and act as a base period against which other hours can be compared. In the earlier part, day dummies had also been used, but except for weekends the typical daily shape appeared to be constant for most households, or at least any differences could not be explained by the household characteristics that were available. To allow for shifts in the curve on weekends, two variables were introduced, sine 2 (SIN2) and cosine 2 (COS2), defined to be sin $\pi t/12$ and cos $\pi t/12$ respectively, at weekends but 0 on other days (the earlier results suggested that additional sine and cosine terms were not necessary). To allow for differences in level on workdays, a dummy WORK was used, being 0 on weekends or public holidays and 1 otherwise.

Since the pricing experiment for part of the sample should also alter the shape of the household-demand curve, a further dummy was defined to investigate this effect. The variable PEAK is defined as follows:

PEAK = 1 9 - 11 a.m. and 5 - 7 p.m. for winter weekdays
= 1 10 a.m. to noon and 1 - 3 p.m. for summer weekdays
= 0 otherwise

Families in the peak-load pricing experiment were required to pay 16¢/kWh during peak periods, 3¢/kWh from 8 a.m. to 9 p.m. (other than the peak hours), and 1¢/kWh for all other hours. On weekends and public holidays the peak period was priced at the intermediate rate. Since this dummy varies between weekends and weekdays, it is not completely colinear with the hour dummies. It may also pick up part of the weekend change of shape for families not in the experiment, so its interpretation is not necessarily straightforward.

For the winter period a dummy variable SCHVAC was used to investigate the effect of school vacations, taking the value 1 during the vacation and 0 otherwise. The entire summer period was during a school vacation, so this variable was not necessary.

The most important short-term causes of movements away from the typical load curve are, of course, the weather variables. The process whereby outside temperatures, humidity, and wind speed change the temperature inside a house or apartment is likely to involve a complicated transfer function, possibly including nonlinear terms. In an attempt to pick up the delayed effect of exceptional temperatures, the moving average of temperatures over the preceding 26 hours (denoted TEMPMA) was included as an explanatory variable. For the winter period both present temperature and this temperature value squared were used (denoted TEMPNOW and TEMPSQ) to see if a nonlinear effect was discernible. For summer, it is expected that temperature by itself is less important than is a measure of discomfort involving both temperature and humidity. The measure used was the average of wet- and dry-bulb temperatures, which is a convenient proxy for the well-known human discomfort index (denoted TEMPHUM). All temperatures are measured in degrees Celsius. After some experimentation with alternative formulations it was decided to include wind speed (denoted WINDSPEED) simply as an additive term in both the winter and summer periods. Two other weather variables were available, solar radiation ($cal/cm^2/min$), and wind direction, but earlier results did not find these variables to be of significance.

The time-series regressions were calculated for a three-month winter period, December 1975, January and February 1976, but excluding the Christmas and New Year periods, which were considered to be exceptional. The summer months were July and August 1976, excluding the July 4 holiday. To list the variables used in these regressions and to illustrate the type of results obtained for an individual household, Table 5-1 shows the regression parameters and t-values for household number 3412317, which is just above average in the total amount of electricity demanded and belongs to the control group. The hour dummy parameters show the household beginning to stir in hour 9, a midmorning peak in use, an afternoon lull, and the most intensive use in the evening. The significance of SIN2 in the winter suggests a deviation between weekday and weekend patterns for that period. Temperature-humidity (TEMPHUM) is seen to be significant in summer, as are several temperature variables during winter.

The analysis and interpretation of almost 280 regressions, each involving 28 or more explanatory variables, is no simple task. Most of the results follow intuition, but there are also some surprises. For example, the daily shape of some households

Table 5-1

TIME-SERIES REGRESSION RESULTS FOR PARTICULAR HOUSEHOLD

Variable	Summer		Winter		Variable	Summer		Winter	
HR1*	0.17	(0.7)	-0.29	(-1.8)	WORKDAY	0.19	(2.0)	-0.04	(-0.5)
HR5	0.27	(1.2)	-0.02	(-0.1)	PEAK	0.21	(0.8)	-0.26	(-1.5)
HR6	0.21	(0.9)	-0.21	(-1.3)	WINDSPEED	-0.016	(-0.75)	-0.006	(-0.6)
HR7	0.22	(1.0)	-0.11	(-0.6)	TEMPMA	0.18	(5.9)	-0.02	(-2.4)
HR8	0.18	(0.8)	0.41	(2.5)	TEMPHUM	0.08	(3.6)		
HR9	1.00	(4.3)	1.20	(6.6)	TEMPSQ	-		0.002	(2.4)
HR10	1.80	(7.6)	2.20	(9.9)	TEMPNOW	-		0.013	(1.6)
HR11	1.30	(4.7)	2.40	(11.1)	SIN2	0.06	(0.4)	0.27	(2.9)
HR12	0.93	(3.2)	1.60	(9.7)	COS2	-0.01	(-0.1)	-0.05	(-0.5)
HR13	0.79	(3.4)	1.10	(6.8)	CONSTANT	-4.68	(-8.8)	7.90	(6.8)
HR14	0.31	(1.1)	0.68	(4.3)					
HR15	0.54	(1.8)	0.55	(3.3)	Adj. R^2	0.203		0.331	
HR16	0.80	(3.6)	0.66	(4.1)					
HR17	0.86	(3.7)	0.90	(5.3)					
HR18	1.20	(5.1)	2.00	(9.2)					
HR19	1.10	(4.9)	2.30	(10.4)					
HR20	1.10	(4.7)	2.00	(12.1)					
HR21	1.70	(7.3)	2.60	(15.5)					
HR22	1.20	(5.3)	1.20	(7.3)					
HR23	0.88	(3.9)	0.55	(3.6)					
HR24	0.52	(2.4)	0.56	(3.3)					
SCHVAC	-		0.35	(4.7)					

Note: t-values are given in parentheses

*HR1 is midnight to 1 a.m.

Table 5-2

PERCENTAGE OF SIGNIFICANT COEFFICIENTS

	Experimental Group				Control Group			
Variable	Summer		Winter		Summer		Winter	
PEAK	57	(8)	56	(6)	39	(39)	49	(21)
TEMPNOW			47	(0)			55	(12)
TEMPSQ			29	(32)			24	(66)
TEMPHUM	33	(78)			50	(79)		
TEMPMA	49	(93)	57	(5)	50	(83)	55	(11)
WINDSPEED	28	(14)	40	(94)	29	(64)	39	(94)
SCHVAC			36	(81)			41	(80)
SIN2	36	(77)	63	(96)	23	(82)	67	(94)
COS2	38	(76)	48	(85)	35	(76)	55	(81)
WORK	32	(32)	48	(27)	38	(33)	36	(11)

Note: Numbers in parentheses indicate percent of these positive.

is dramatically different from that of others. Whether these differences are systematic or are merely sampling errors and whether or not these shapes can be explained by the available data on household characteristics must await the cross-sectional regressions (reported in Chapter 6 where these differences are explicitly tested).

Table 5-2 summarizes some of the time-series regression results. For all variables other than the hourly dummies and the constant term, the percentage of coefficients having t-values with absolute values greater than 1.9 is shown, and in parentheses is given the percentage of these "significant" coefficients that have a positive sign.

Many of the results are easily explained, such as electricity usage being negatively correlated with temperature (TEMPNOW) in winter but positively related in summer (TEMPHUM) and, similarly, higher usage being correlated with higher wind speed in winter. Other results have less clear-cut explanations. For example, the square of present temperature, although infrequently significant, is predominantly negative for the experimental group but is usually positive for households in the control group. The results for SIN2 and COS2 suggest that the daily shape varies between weekends and other days to a greater extent during winter than summer, which indicates that analyzing winter and summer separately was a correct strategy.

The households were classified into five equal subsamples according to the previous year's total usage level; class 1 represented those in the lowest 20% of users and class 5 were the 20% heaviest users. Table 5-3 shows the average corrected R^2

values achieved in the time-series regressions for the various groups.

Table 5-3

AVERAGE R^2 VALUES

	Class 1	2	3	4	5	Total
Experimental Group						
Winter	0.315	0.281	0.360	0.352	0.507	0.363
Summer	0.259	0.235	0.281	0.260	0.272	0.261
Control Group						
Winter	0.345	0.301	0.307	0.379	0.404	0.353
Summer	0.281	0.236	0.225	0.214	0.309	0.245

Although the very heaviest users, who typically used electricity for heating, have the highest R^2 values, in general there is no systematic relationship between the level of usage and R^2. However, the R^2 values are consistently higher in the winter than the summer, suggesting either more consistency of the daily shape (as picked up by the hourly dummies) in the winter or the greater explanatory power of the weather variables during the cold months.

Compared with the results achieved when using quarter-hourly data, the R^2 values are now somewhat higher, as might be expected; this observation was particularly true for the lower-usage classes of households.

Cross-Sectional Analysis

The variables used in the cross-sectional regressions for this part are the numbers of electric ranges, self-cleaning ovens, electric dryers, self-defrosting refrigerators, manual-defrosting refrigerators, freezers, dishwashers, color TV sets, black-and-white televisions, humidifiers, dehumidifiers, window air conditioners, number of persons in the household in each of three age groups (under 18, 18-64, 65 or over), age of house, square footage (area) of house, and the following dummy variables: one for experimental pricing group, central air conditioning, electric main heating system, electric supplementary heating system, electric water heater and swimming pool pump; and two dummy variables for type of structure, one for single-family and the other for mobile homes. Based primarily on the earlier experience, a subset was chosen for each season.

Ordinary least squares and generalized least squares regressions were estimated for each of the time-series regression coefficients in both the summer and the winter. The GLS correction was different for each regression, but invariably it had little effect. The weights constructed varied by a maximum of a factor of 2 over the households, but they frequently varied only by a few percent. This occurred because the homoscedastic component of the error term that is due directly to the "noise" in the cross-sectional relation was much larger than the component due to uncertainty in the dependent variable.

Out of 60 regressions, only 13 showed any change in the list of significant variables, as measured by t-statistics; and these generally were just small changes that made the variable appear marginally significant rather than marginally insignificant. The R^2 fell slightly in the GLS regressions; but R^2 here has a dubious interpretation, particularly since there is no constant. Because the results are so similar, only the OLS regressions are presented. We have, however, indicated the places where the GLS results differed.

Tables 5-4 and 5-5 present, respectively, the summer and winter cross-sectional results using the OLS estimation procedure. The dependent variables are the series of time-series regression coefficients, and the independent variables are the household characteristics listed above. The tables also show, for each variable, the adjusted R^2, Durbin-Watson statistic (d), and those independent variables that appear to be significant (those with t-values over 1.96 in absolute terms). A variable is underlined if $|t| \geq 3$ and is given a query if the t-statistic is suggestive but not strictly significant (i.e., $|t| \geq 1.6$). A negative sign in parentheses indicates that the coefficient in the cross-sectional regression is negative. Although the data are cross-sectional, as the households were approximately ranked in order of level of electricity usage (first in the experimental group and then in the control group), the Durbin-Watson statistic does have some interpretive value. None of the d-values, however, indicate any serial correlation.

In general, the cross-sectional results seem promising. For some important variables the R^2 are adequate to good, and most of the significant explanatory variables make economic sense. Some of the more important interpretations possible from the tables are listed below, by season.

<u>Summer Results</u>

The adjusted R^2 values vary from -0.07 to 0.28. For the more important variables (hourly dummies, peak price, etc.), these values range from 0.10 to 0.28 which is not unreasonable for a cross-sectional study, though certainly not spectacular. The results for individual variables are summarized below.

Table 5-4

SUMMER CROSS-SECTIONAL RESULTS FOR SELECTED VARIABLES (OLS)

Dependent Variable	Adj. R^2	d	Main Significant Household Characteristics
CONSTANT	0.15	1.72	Central air conditioning × square footage (-), window air conditioning × square footage (-)
HR5[a]	0.03	2.02	Dehumidifier (-), window air conditioning × square footage, people under 18 (-?)
HR6	-0.03	2.02	Electric water heater (?)
HR7	0.02	1.72	Electric water heater
HR8	0.05	1.64	Electric water heater
HR9	0.06	1.69	Electric water heater, people 18-64
HR10	0.10	2.01	Electric water heater, people under 18, dehumidifier (-?), dishwasher (+?)
HR11	0.18	1.82	Electric water heater, people under 18, central air conditioning × square footage (+?)
HR12	0.18	1.61	Electric water heater, people under 18, central air conditioning × square footage
HR13	0.13	2.11	Electric water heater, people under 18
HR14	0.19	1.65	Electric water heater, people under 18, central air conditioning × square footage
HR15	0.22	1.61	Electric water heater, people under 18, central air conditioning × square footage
HR16	0.18	1.95	Electric water heater, people under 18, central air conditioning × square footage
HR17	0.22	2.00	Electric water heater, people under 18, central air conditioning × square footage, heated pool, people 18-64 (+?)
HR18	0.24	1.95	Electric water heater, people 18-64, people under 18, heated pool, central air conditioning × square footage, people 65 or over (+?), dishwasher (+?)
HR19	0.28	2.02	Electric water heater, people under 18, people 18-64, central air conditioning × square footage, dishwasher

Dependent Variable	Adj. R^2	d	Main Significant Household Characteristics
HR20	0.28	1.86	Electric water heater, people under 18, people 18-64, central air conditioning × square footage, dishwasher
HR21	0.26	2.04	Electric water heater, people under 18, people 18-24, central air conditioning × square footage, dishwasher
HR22	0.25	2.05	Electric water heater, people under 18, people 18-64, central air conditiioning × square footage, dishwasher, electric dryer (+?)
HR23	0.26	1.93	Electric water heater, people under 18, people 18-64, central air conditioning × square footage, dishwasher, electric dryer, survey dummy (+?)[b]
HR24	0.17	1.94	Electric water heater, people under 18, people 18-64, electric dryer
HR4	0.12	1.96	People under 18, people 18-64
PEAK PRICE	0.20	2.01	People under 18 (-), swimming pool pump (-), survey dummy (-), central air conditioning × square footage (-?)
WEEKDAY	-0.07	1.84	None
WIND SPEED	0.06	2.07	Dehumidifier (-), survey (-)
TEMPERATURE MOVING AVERAGE	0.08	1.73	Central air conditioning × square footage, window air conditioning × square footage
TEMPERATURE HUMIDITY INDEX	0.16	2.09	Central air conditioning × square footage, window air conditioning × square footage
SIN2	-0.01	1.80	None
COS2	-0.01	2.05	People 18-64
HIGH HOURS	0.26	1.76	People under 18, central air conditioning × square footage, electric water heater, people 18-64
SUMMER TEMP	0.19	1.79	Central air conditioning × square footage, window air conditioning × square footage

[a]When interpreting hourly dummies it should be remembered that they are relative to electricity usage in the suppressed hours 2 through 4 (1 to 4 a.m.).

[b]Survey dummy = 1 if in experimental group, 0 otherwise.

Table 5-5

WINTER CROSS-SECTIONAL RESULTS FOR SELECTED VARIABLES (OLS)

Dependent Variable	Adj. R^2	d	Main Significant Household Characteristics
CONSTANT	0.68	1.88	Electric water heater, electric heat × square footage
HR5	0.12	2.07	Electric heat × square footage, people 18-64 (-)
HR6	0.11	2.15	Electric heat × square footage, supplementary electric heat, electric water heater (+?)
HR7	0.14	2.16	Electric heat × square footage, electric water heater
HR8	0.17	2.33	Electric heat × square footage, survey dummy (?)
HR9	0.18	2.34	Electric heat × square footage, dishwasher (?)
HR10	0.18	2.14	Electric heat × square footage, dishwasher
HR11	0.09	2.25	Electric heat × square footage (?), dishwasher (?), electric water heater (?), people under 18 (?)
HR12	-0.02	2.36	None
HR13	0.01	2.05	People under 18
HR14	0.02	2.10	People under 18
HR15	0.05	2.31	Electric heat × square footage (-), people under 18
HR16	0.03	2.29	People under 18 (?)
HR17	0.07	2.41	People under 18
HR18	0.19	2.28	Electric heat × square footage, people under 18, dishwasher
HR19	0.22	2.17	People under 18, dishwasher, electric water heater (?)
HR20	0.17	2.28	People under 18, dishwasher, electric water heater (?)

Dependent Variable	Adj. R^2	d	Main Significant Household Characteristics
HR21	0.31	2.06	People under 18, dishwasher, electric heat × square footage, electric water heater (?)
HR22	0.31	2.04	People under 18, dishwasher, electric water heater, survey, electric heat × square footage (?)
HR23	0.28	2.19	People under 18, dishwasher, electric water heater, survey
HR24	0.18	2.08	People under 18, dishwasher, electric water heater, survey
HR1	0.10	2.12	People under 18, electric water heater, survey
PEAK PRICE	0.31	1.92	Electric heat × square footage (-), dishwasher (-), survey (-)
SCHVAC	-0.02	1.66	None
WEEKDAY	-0.05	2.26	None
WIND SPEED	0.54	2.39	Electric heat × square footage, electric water heater (?)
TEMPERATURE MOVING AVERAGE	0.77	1.83	Electric heat × square footage (-), electric water heater (-?)
TEMPERATURE SQUARED	-0.00	2.20	Survey (-)
TEMPERATURE NOW	0.74	2.24	Electric heat × square footage
SIN2	0.32	2.30	Electric heat × square footage, people under 18, people 18-64, dishwasher, electric water heater (?)
COS2	0.15	2.48	Electric heat × square footage, people under 18 (?)
HIGH HOURS	0.12	2.24	People under 18, electric water heater (?), dishwasher
SUM OF TEMPERATURES	0.83	2.09	Electric heat × square footage (-), supplementary electric heater (?)

The hourly dummies contain many interesting results. The importance of certain appliances is clearly brought out in these variables. For instance, the electric water heater becomes a significant appliance in determining the regression coefficients for the hours 7 through 24 (Table 5-4). With GLS estimates, hour 6 is also significant, with hour 5 coming close. The dishwasher is most significant at the 19th hour, but during hours 20 through 23 also it has a significant impact. The product of a dummy for central air conditioning and the living area of the house in square feet was used as an explanatory variable. This becomes significant at hours 12 and 14 through 23. The electric dryer is significant from 11 p.m. to midnight, but it has a t-value between 1.6 and 1.96 in the preceding two hours. If some of the insignificant variables causing multicolinearity problems were eliminated, the dryer might become important in these hours too. The pool pump is important from 4 to 5 p.m. Among other variables, the number of people under 18 is important during almost all the hours, the only exception being hours 5 through 8. The number of people in the age group 18-64 is significant in the morning only from 8 to 9 a.m. The number of older people (65 or over) doesn't seem to matter much, except possibly from 5 to 6 p.m., during which time the t-value lies between 1.6 and 1.96. The dummy variable for the experimental group was generally not significant, although it came close during the 23d hour. The remaining variables do not seem to have any significant effect on the regression coefficients for the hourly dummies. During the high-hour period of 9 a.m. to 9 p.m. the significant variables were people under 18, people 18-64, electric water heaters, and central air conditioning times square footage (area).

To economists, the most interesting variable is the peak price, with an R^2 of 0.2. Not surprisingly, the experimental households showed a strongly significant negative regression coefficient, indicating that they would use less electricity during the peak pricing period as compared with the control group. The heated pool also has an important negative effect, and households with a pool pump are more likely to shift out of the peak and reduce electricity consumption during that period. Central air conditioning and area seem to interact to yield a negative effect, but its t is only between -1.6 and -1.96. The number of people under 18 also reacts negatively, that is, it shifts out of the peak in a significant manner (with GLS estimates this variable is marginally insignificant). The remaining variables do not have much effect on the regression coefficient for the peak price.

The weather variables (wind speed, temperature moving average, temperature-humidity index, and sum of temperature-which is the sum of the other two temperature variables) had R^2 values ranging from 0.06 to 0.19. Air conditioning (both central and window) times area was the only variable that had any significant impact on the temperature variables. For wind speed, the dehumidifier had a strong negative effect. The experimental group showed a negative effect on wind speed.

The adjusted R^2 values for the sine and cosine terms are negative, indicating a poor fit in both cases. In the case of the sine term, none of the explanatory variables were significant;

but in the case of cosine wt/2, the number of people in the age group 18-64 was significant.

Winter Results

The values of R^2 ranged, for the winter cross-sectional regressions, from -0.05 to 0.83 (Table 5-5). As in the case of summer, the hourly dummies had reasonable R^2 values, except in a few cases. In this case, also, there is no serial correlation. The specific results are described below.

The hour dummies exhibit interesting patterns similar to those in the summer. Electric heating times area is important during hours 5 through 10, 15, 18, and 21. The regression coefficient is generally positive except during 2 and 3 p.m., which suggests that less is used then than from 1 to 4 a.m. The dishwasher has a significant effect from 9 to 11 a.m. and again from 6 p.m. to midnight. The GLS estimate was marginally insignificant for hour 1. The electric water heater has a noticeable effect from 6 to 7 a.m. and again from 9 p.m. to 1 a.m. In addition, the GLS estimates are marginally significant for hours 8 and 10. During the hours 5 to 6 a.m. and 6 to 9 p.m., the t-values for this variable range from 1.6 to 1.96, suggesting possible, although not strict, statistical significance. The number of people under 18 has a significant impact on the regression coefficients for the hours 12 noon through 1 a.m. It will be remembered that this variable was important at all hours after 8 a.m. in the summer. Presumably the people under 18 tend to be in school during the winter, accounting for less consumption during those mornings when other variables are held constant. People in the age group 18-64 generally had no effect on the hourly dummy regression coefficients, the only exception being from 4 to 5 a.m. when they had a significant decrease in consumption relative to 1 to 4 a.m. In this respect, summer and winter exhibit a strong constrast. It will be recalled that the age group 18-64 significantly influenced the hourly regression coefficients from 8 to 9 a.m. and from 5 p.m. to 1 a.m. This group's activities involving electricity consumption have indeed been curtailed in winter.

The experimental group did shift out of the peak hours, as evidenced by the significant positive coefficient for the survey dummy during the hours 9 p.m. to 1 a.m. This also is in clear contrast to the summer behavior, when it was usually insignificant. Presumably the extraordinary expense of winter heating provided the incentive for the survey households to postpone their consumption to the off-peak hours. During the high-hour period of 9 a.m. to 9 p.m., the dishwasher and the number of people under 18 were the significant variables, although the electric water heater came close.

The regression coefficient for the peak price was significantly affected by electric heating times area, dishwasher, and the survey dummy, all of which had the expected negative sign. The R^2 for this variable was a respectable 0.31. In addition, the GLS estimate indicated a t-value of 1.61, suggesting near significance of people 18-64.

The weather variables generally had a very high value for R^2 (0.54 to 0.77), with the exception of the regression for current temperature, for which it was nearly 0. Electric heating times area is significant in all these cases. The survey had a significant effect only on temperature squared.

The electric water heater came close to being significant (for GLS it was significant) for the wind speed and temperature moving average. The sum of the temperature variables (temperature now + temperature moving average) had a very high R^2 (0.83), but only the electric heating was significant; although supplementary electric heating was nearly significant, with a t-value between 1.6 and 1.96.

Unlike in the summer, the sine and cosine terms had R^2 values of 0.32 and 0.15 respectively. Electric heating times area, people under 18, people 18-64, and the dishwasher significantly affected the regression coefficient for sine wt/2, and the electric water heater was close. For cosine wt/2 only the electric heating times area was significant, although the number of people under 18 was close.

Chapter 6

IMPLICATIONS FOR PEAK-LOAD PRICING

The previous chapter presented results which indicated that households subject to peak load pricing used significantly less electricity during the four hours per day during which the peak price was charged, in both winter and summer. While it would not be sensible to attempt to derive a price elasticity from this observation (since there are only two price schedules and thus two points on a demand curve), it is possible to infer which appliance uses are responsible for the shift and whether activities are shifted into neighboring time periods or are merely curtailed.

In Figures 6-1 and 6-2 the coefficient of the survey variable from the cross-sectional regression is plotted against the hours of the day for the summer and winter periods. On the same graphs are plotted the price levels at these hours of the day. Assuming that the day is a weekday, the peak coefficient is included for the relevant hours. These series represent the difference in use by two households having the same appliances and demographic variables but different pricing. This interpretation is correct if the weather remains constant over the day or if the survey coefficient is 0 in the weather equations.

It can be seen from these graphs that there is substantial decrease in usage at the peak hours. There is also a spreading of the load, particularly into the evening but also into the morning period before the peak prices are charged. There appears to be a shift out of the shoulder-priced hours, again particularly

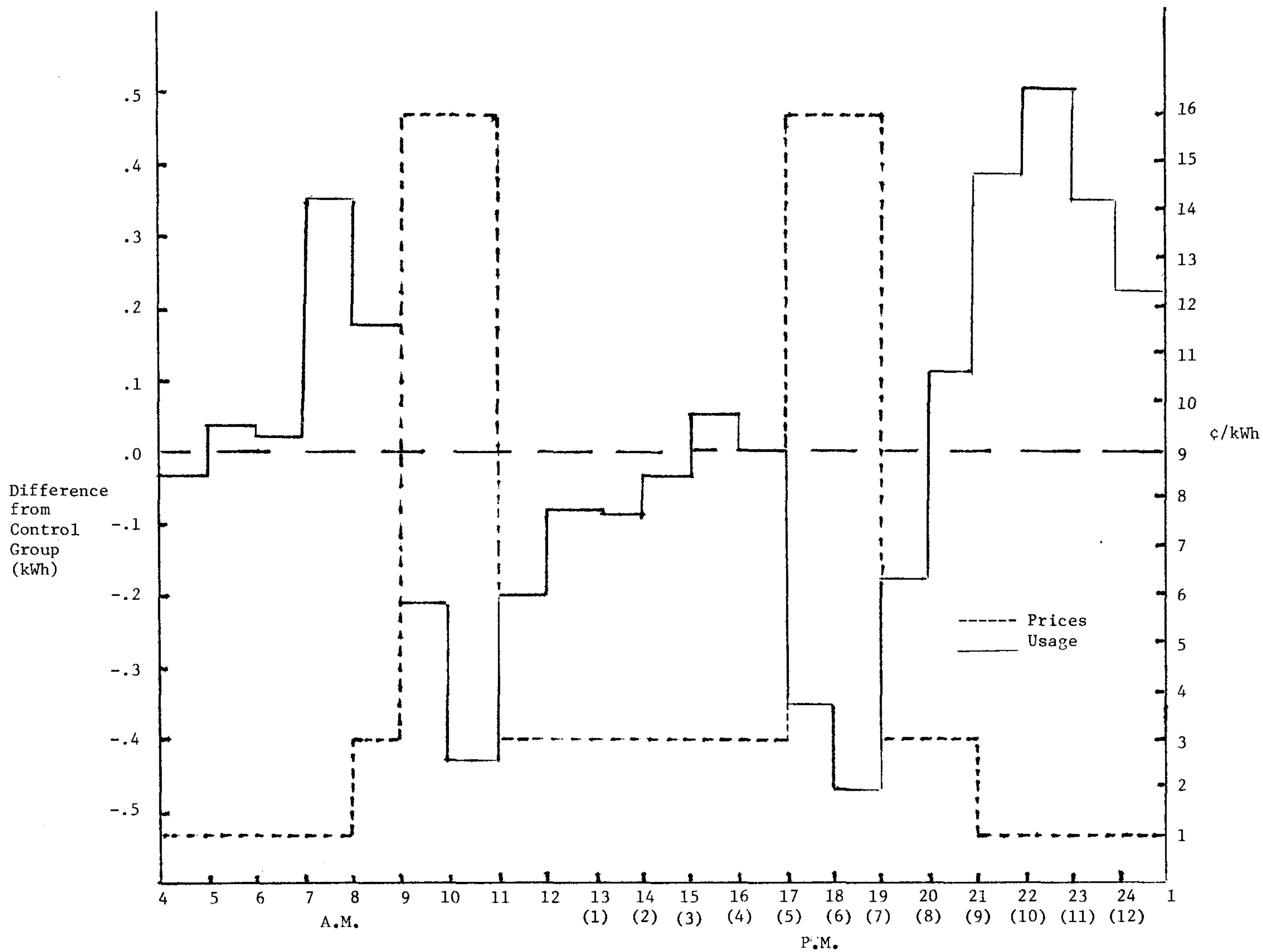


Figure 6-1. Difference between consumption by experimental and control groups, by time of day, in winter.

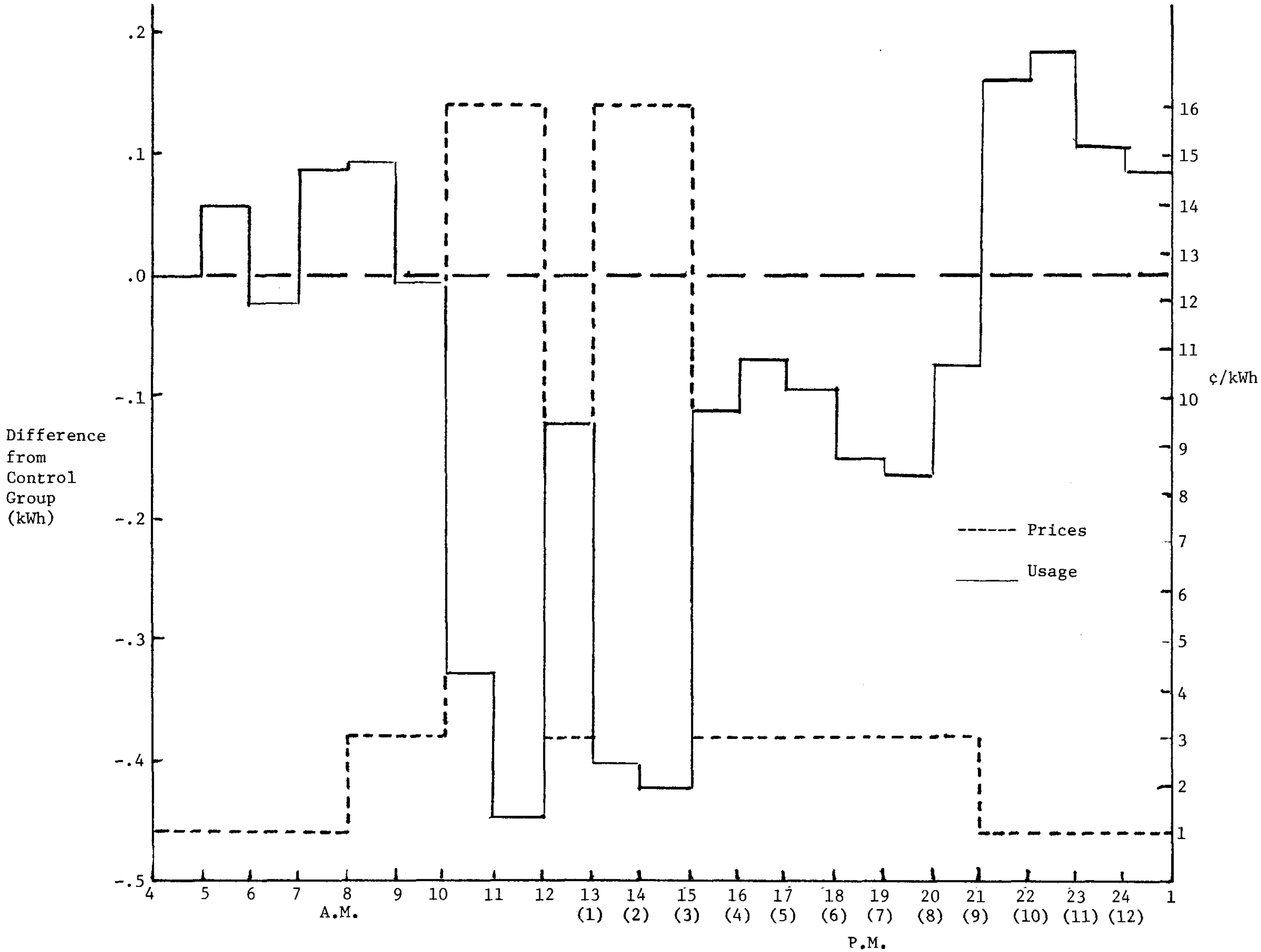


Figure 6-2. Difference between consumption by experimental and control group, by time of day, in summer.

in the evening. The pattern of cross-elasticities appears to be rather complicated. For example, the hour from 8 to 9 a.m. in the winter is above the usage of the control group even though it is a shoulder period. The explanation is that the large shift from the peak period, which begins at 9 a.m., exceeds the decrease that would ordinarily be expected from households shifting usage to the period before 8 when the shoulder price begins.

In general, the households having peak pricing responded in the same way to weather conditions as those with flat-rate schedules. The exceptions are the quadratic term in temperature in the winter and the response to wind speed in the summer. Ordinarily one would expect the temperature-squared variable to enter with a positive sign, indicating a convex function with usage increasing faster as the temperature becomes lower. The survey variable enters the explanation of temperature squared with a significant negative sign, thereby indicating that the households apparently decrease their consumption when the wind speed increases. Presumably this is due to a discretionary shutdown of the air conditioners, or perhaps other appliances.

It may be possible to infer how the experimental group decreases their consumption. If, for example, it is from shifting the time when the electric dishwasher is used, then separate regressions for the experimental and control groups should exhibit the different responses. This is particularly useful information, since the elasticity of response to peak pricing will therefore depend upon the appliance mix in a region.

The regressions for the peak period are presented separately for the experimental and control groups in Table 6-1. A Chow test for equality of all coefficients except the constant is

Table 6-1

REGRESSIONS FOR PEAK BY SUBGROUP

Variable	Winter Control	Winter Experimental	Summer Control	Summer Experimental
ELECTRIC DRYER	0.022 (0.125)	0.219 (0.199)	-0.057 (0.145)	-0.178 (0.116)
DISHWASHER	-0.064 (0.094)	-0.410* (0.178)	0.166 (0.107)	-0.015 (0.103)
ELECTRIC WATER HEATER	-0.096 (0.109)	-0.183 (0.189)	-0.006 (0.092)	-0.137 (0.098)
PEOPLE UNDER 18	-0.051 (0.052)	0.008 (0.066)	-0.009 (0.040)	-0.082 (0.039)
PEOPLE 18-64	-0.063 (0.049)	-0.120 (0.085)	-0.139 (0.076)	-0.009 (0.050)
PEOPLE OVER 65	-0.088 (0.139)	0.013 (0.226)	-0.132 (0.138)	0.074 (0.131)
ELECTRIC HEAT × SQUARE FOOTAGE × 1000	-0.066 (0.091)	-0.511 (0.124)		
SUPPLEMENTARY ELECTRIC HEAT	0.224 (0.216)	-0.076 (0.378)		
TEMPORARY STRUCTURE	-0.196 (0.345)	-0.116 (0.559)		
DEHUMIDIFIER			0.056 (0.094)	0.047 (0.105)
WINDOW AIR CONDITIONERS × SQUARE FOOTAGE × 1000			-0.012 (0.023)	0.001 (0.036)
CENTRAL AIR CONDITIONING × SQUARE FOOTAGE × 1000			-0.223 (0.144)	-0.180 (0.124)
SWIMMING POOL PUMP			0.005 (0.138)	-0.636 (0.163)
CONSTANT	0.184 (0.159)	0.089 (0.227)	0.220 (0.202)	0.018 (0.159)
R^2	0.150	0.370	0.180	0.300

easily accepted, with F values of 1.22 and 1.24 in winter and summer. There are, however, some interesting differences in the coefficients. In the winter the electric heating variable is significantly different, judging by the difference in the coefficients divided by the square root of the sum of their variances. This asymptotically normal ratio is 2.9, indicating that peak-pricing households having electric heat shift out of the peak period more than those without electric heat. For dishwashers, the statistic is 1.7, indicating that a mechanism for experimental households to shift usage out of the peak is to alter the time at which they use the dishwasher. In the summer, only the swimming pool pump appears to differ significantly between the two groups, with a statistic of 3. As this is a regular use of appliances that can easily be shifted, it seems reasonable that it appear to be a mechanism for the shift. The dishwasher has a statistic of 1.2 in the summer. Notably lacking any difference is the effect of air conditioners; not only are air conditioners not significantly different between experimental and control groups, but the point estimates suggest that the experimental group actually uses them more during the peak hours.

In conclusion, there appears to be a very strong response to peak load pricing, with the substantial decrease in the peak hours being shifted partly to the early morning but more heavily into the evening. This shift is associated particularly with households having electric heating but also with dishwashers and swimming pool pumps.

Bearing in mind the difficulty of predicting personal habits, the results seem sensible and persuasive. Further analyses along these lines could well help distinguish still further the critical determinants of residential load curves and their response to time-of-day pricing.

REFERENCES

1. Boyd, James W. (editor). Proceedings on Forecasting Methodology for Time-of-Day and Seasonal Electric Utility Loads. Palo Alto, Calif.: Electric Power Research Institute, 1976.

2. Burbank, Donald. "The Connecticut Peak-Load Price Experiment" in Lawrence, Anthony, ed., Forecasting and Modeling Time-of-Day and Seasonal Electricity Demands, EA 578-SR, Electric Power Research Institute, Palo Alto, CA. 1977.

3. Granger, C.W.J., and P. Newbold. Forecasting Economic Time Series. Academic Press, 1977.

4. Graybill, Franklin. Theory and Application of the Linear Model. Duxbury Press, 1976.

5. Hanushek, Eric A. "Efficient Estimators for Regressing Regression Coefficients." Journal of the American Statistical Association. 1974.

6. Hildreth, C., and J.P. Houck. "Some Estimates for a Linear Model with Random Coefficients." Journal of the American Statistical Association, Vol. 63, (1968), pp. 584-595.

7. Kendall, M.G., and A. Stuart. Advanced Theory of Statistics. Hofner, 1967.