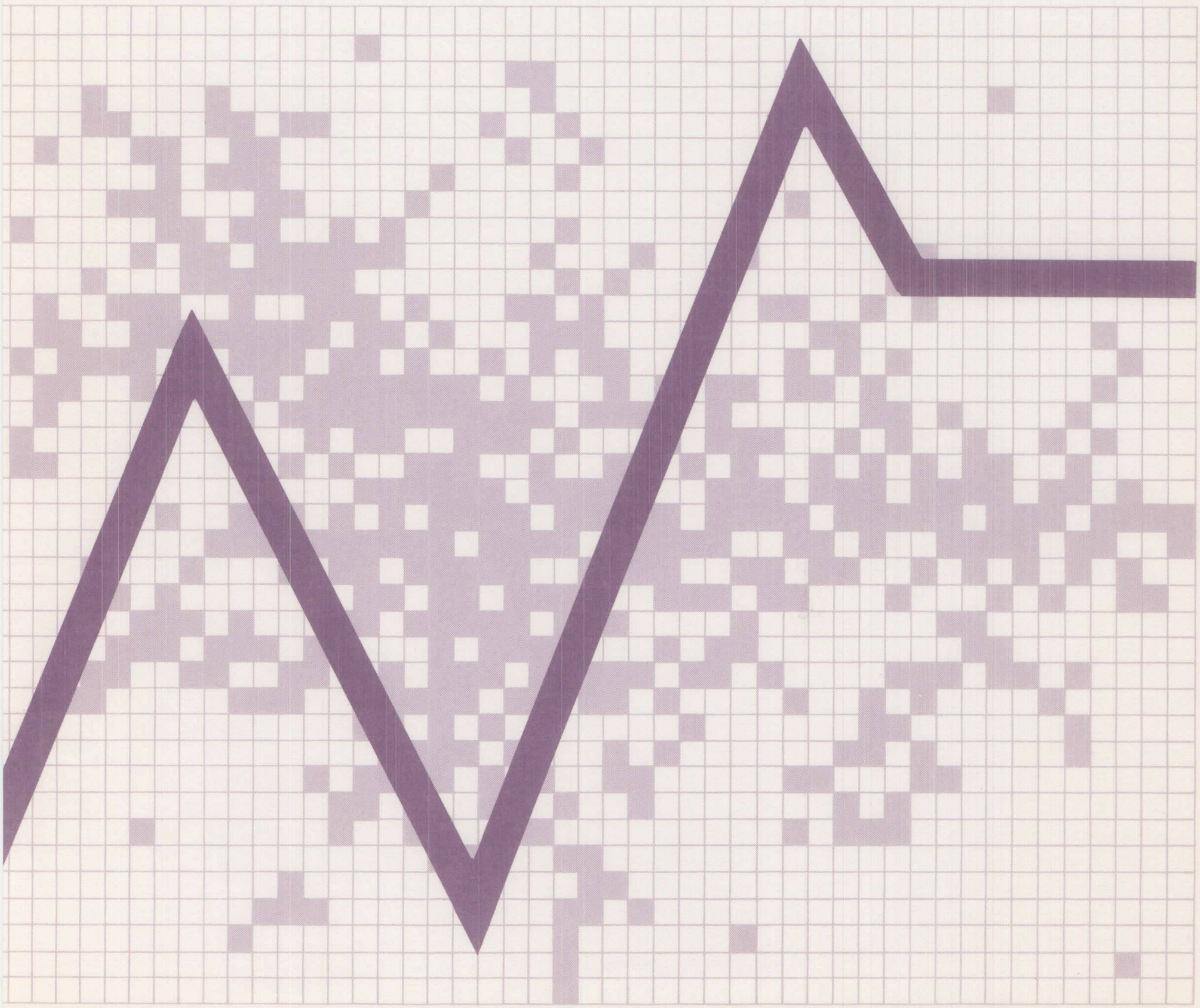
OE/EIA-0272

The National Interim Energy Consumption Survey:

Exploring the Variability in Energy Consumption

July 1981

U.S. Department of Energy
Energy Information Administration
Assistant Administrator for Program Development
Office of the Consumption Data System
Industrial Data Systems Division



This publication is available from the Superintendent of Documents, U.S. Government Printing Office, at the following address:

Superintendent of Documents
U.S. Government Printing Office
Washington, D.C. 20402

Order Desk: (202) 783-3238

Stock Number: 061-003-00205-6
Price: $4.25

For questions on energy statistics or information on availability of other EIA publications, contact:

National Energy Information Center, EI-20
Forrestal Building
U.S. Department of Energy
Washington, D.C. 20585

(202) 252-8800

For those living and working in the Mountain and Pacific time zones, you may call:

National Energy Information Center Affiliate
Albuquerque, N.M. 87131

(505) 846-2375

DOI/EIA-0272
UC-13

The National Interim Energy Consumption Survey:

Exploring the Variability in Energy Consumption

July 1981

U.S. Department of Energy
Energy Information Administration
Assistant Administrator for Program Development
Office of the Consumption Data System
Industrial Data Systems Division
Washington, D.C. 20585

Prepared by: Robert Latta, Thomas H. Woteki, Lynda Carlson,
Wendel Thompson, Kenneth Vagts

Released for printing August 18, 1981.

PREFACE

Authority for the National Interim Energy Consumption Survey (NIECS) is contained in Section 52 of the Federal Energy Administration Act of 1974, as amended, which charges the Energy Information Administration with creating and maintaining a National Energy Information System. The NIECS, a part this system, represents the first attempt at simultaneously collecting residential energy consumption data and household characteristics from a national, statistical sample. This NIECS publication is intended for use by representatives of Federal, State, and local governments as well as by representatives from the private sector. In addition, the Office of Management and Budget has plans to use the results in a report on the possible impact of natural gas deregulation.

This report is a description of some early efforts to model the variation in total energy consumption and consumption by end-use reported in the National Interim Energy Consumption Survey (NIECS). The analyses and estimates presented herein are very tentative; they are not presented as official estimates. Rather, it is our objective to point to some of the difficulties we have encountered in our analyses, to share some observations gleaned from the data, and to share ideas with other analysts who may be working on similar problems.

EXECUTIVE SUMMARY

In this report, separate models for the electricity and natural gas consumption in single-family detached dwellings were developed. The following are some of the highlights of the two models.

- In both models, the number of bathrooms in the dwelling was a major factor in the space-heating component. The number of bathrooms is probably an indication of the size of the house as well as the life-style of the occupants.

- In the natural gas model, the type of main space-heating equipment has a big effect on the intercept term. In particular, the intercept term is highest for dwellings that use natural gas for main space-heating via radiators or water pipes. The next highest is central forced-air equipment.

- By assuming a hypothetical household's main space-heating fuel is electricity we can obtain the modeled space-heating component from the electricity consumption model. Alternatively, by assuming the main space-heating fuel is natural gas, we can obtain the modeled space-heating component from the natural gas consumption model. By looking at a range of hypothetical households, it can be seen that the modeled space-heating component from the natural gas model tends to be approximately twice the modeled space-heating component of the electricity model.

- In the electricity consumption model, the interaction term between income level and the potential number of rooms that can be air conditioned is significant. In particular, higher income groups not only are more likely to have air conditioning, but also use the equipment they do have more than lower income groups.

- In the electricity consumption models, it was possible to use separate terms for most of the major electrical appliances. In the natural gas model, this was not possible. The variability in the space-and water-heating components is probably masking the effect of the natural gas appliances.

- The heating degree-days and cooling degree-days data used in the model were both computed using 65 degrees Fahrenheit as a base. A preliminary analysis using other bases indicated that a lower base for heating degree-days and a higher base for cooling degree-days would be preferable.

CONTENTS

ILLUSTRATIONS

TABLES

Page No.

INTRODUCTION

By December of 1979, fieldwork for EIA's first survey of energy consumption in the residential sector, the National Interim Energy Consumption Survey (NIECS), was largely concluded. Since then, a variety of reports [References 1,2,3,4] have been produced, which tabulate and summarize the data and survey operations. Concurrently, a variety of detailed methodological and subject matter analyses were begun concerning six interrelated sets of problems:

- Variation in energy consumption; How does energy consumption vary by region? By income group? By square footage? By household characteristics? By main space-heating fuel?

- Total energy consumption for the household; How should a household's housing unit consumption be linked with its vehicle consumption to determine total energy consumption by household?

- Energy consumption by end-use; How much natural gas is used for space-heating? How much electricity is used for air conditioning? How much of either is used for water-heating? What can we say about end-use consumption of fuel oil?

- Energy performance ratios; Can energy consumption be normalized to eliminate variation in physical factors (such as climate) and can such normalization form the basis for comparing consumption at different points in time? For example, does normalization by degree-days in the ratio Btu/degree-day substantially eliminate the variation in consumption associated with temporal and geographic variations in climate? At what level of aggregation or over what partition of the household sector should this normalization be performed? Can these ratios be used as indicators of the energy quality of the housing stock?

- Small area estimates: Can the data from a relatively small national sample be used to infer the energy consumption of "small" areas such as States? How should this be done?

- Imputing missing data: Every survey has missing data problems to one degree or another. What is the best way to handle them here? [References 5 and 6]. Can microaggregated consumption data assist in consumption imputations? Can consumption data be used to impute housing unit characteristics?

This progress report describes some early analyses of the first problem mentioned, the variability in energy consumption reported in the NIECS. It concentrates on regression models for describing the variability in natural gas and electricity consumption for households living in single-family detached houses. Section 1 provides a description of the NIECS sample data and the subsets of it that were used in our analyses. Several appendixes supplement this discussion. Section 2 presents a theoretical model for describing a household's energy consumption. Section 3 presents some graphical summaries of the data which serve to explain the specific models that were fit to the data and leads to some speculation about what contributes to their lack of fit. Section 4 describes the empirical models fit to the data, and the method of fitting, and relates them to the discussion in Section 2. Section 5 summarizes the results of the preceding sections and contains some speculation on interpretation of the model coefficients obtained in Section 4 and some potential uses of the model.

Dear Reader:

Please answer the following questions about the Energy Information Administration *National Interim Energy Consumption Survey: Exploring the Variability in Energy Consumption*. Your responses will enable us to continue to evaluate our publications.

1. Where did you obtain this publication?
   - ☐ GPO Superintendent of Documents
   - ☐ Energy Information Administration Subscription
   - ☐ National Energy Information Center
   - ☐ Energy Information Administration Training Course
   - ☐ Technical Information Center
   - ☐ National Technical Information Service
   - ☐ Other ______________________________
2. With which organization are you affiliated?
   - ☐ Federal Government
   - ☐ State/Local Government
   - ☐ Foreign Government
   - ☐ Industry/Commerce
   - ☐ Library
   - ☐ Nonprofit/Educational Institution
   - ☐ Consulting Firm
   - ☐ Publisher/Media
   - ☐ Other ______________________________
3. What is your occupation?
   - ☐ Manager
   - ☐ Technician
   - ☐ Researcher
   - ☐ Librarian
   - ☐ Journalist
   - ☐ Politician
   - ☐ Planner
   - ☐ Other ______________________________
4. For what do you use this publication?
   - ☐ Research
   - ☐ Legislation/Regulation
   - ☐ Instruction
   - ☐ Policy Formulation
   - ☐ Projection
   - ☐ Planning
   - ☐ General Information
5. How would you change the content of this report?
   - ☐ Combine it with another report
   - ☐ Change Section ______________ as follows: ______________
   - ☐ Change frequency of publication
   - ☐ Discontinue
   - ☐ Retain as currently published
6. How would you rate the general quality of the content of this publication?
   - ☐ Excellent ☐ Good ☐ Fair ☐ Poor
7. How would you rate the analysis of data on residential energy consumption?
   - ☐ Excellent ☐ Good ☐ Fair ☐ Poor
8. How relevant do you think the publication's information is for predicting residential energy use?
   - ☐ Valuable ☐ Useful ☐ Ineffective ☐ Inappropriate ☐ Superfluous
9. How do you rate the theoretical model used for residential energy consumption?
   - ☐ Excellent ☐ Good ☐ Fair ☐ Poor
10. In comparison to other sources for this information, how would you rate this volume?
    Source: ______________________________
    - ☐ Excellent ☐ Good ☐ Fair ☐ Poor
11. How would you rate the publication's readability?
    - ☐ Appropriate ☐ Too Technical ☐ Imprecise
12. How would you describe the scope of the data?
    - ☐ Appropriate ☐ Too Broad ☐ Incomplete ☐ Distorted
13. What specific additions, deletions, or corrections would you make to this publication?

____

____

____

Thank you for your assistance.

John E. Daniels
National Energy Information Center
Energy Information Administration

NO POSTAGE
NECESSARY
IF MAILED
IN THE
UNITED STATES

BUSINESS REPLY MAIL

FIRST CLASS PERMIT NO. 11301 WASHINGTON, D.C.

POSTAGE WILL BE PAID BY ADDRESSEE

U.S. Department of Energy
Energy Information Administration
National Energy Information Center, EI-20
Washington, D.C. 20585

Section 1. The NIECS: Sampling Plan and Data

The NIECS survey was designed as a probability sample of households using personal interviews to obtain energy-related characteristics of the housing unit, characteristics of appliances, information on the conservation activities and demographic characteristics of the household members, and data on energy consumption and expenditures. The latter data were obtained from the utilities serving the sampled households; permission to obtain these data was solicited during the interviews.

The availability of accurate consumption and expenditures data is a feature unique to the NIECS among Federal Surveys of the household sector. This section briefly describes the NIECS sampling plan and the subsets of the data used for preliminary analysis. The Appendixes provide greater detail.

The Sampling Plan

The NIECS was based on a multi-stage area probability sample. The selection technique was roughly equivalent to dividing the United States into small geographic segments--each consisting of a cluster of about 10 households--and making a systematic random selection of such clusters for the survey. Probability methods were used at each stage of sample selection. Interviewers had no choice in the selection of households for the survey.

Altogether, 4,849 housing units were selected for the national sample. Of these, 342 were determined to be vacant or seasonal at the time of the first interviewer contact, leaving 4,507 occupied housing units (households) in the sample for the survey.

Mail Questionnaire

A number of households were unavailable for a personal interview so a mail questionnaire was used to solicit a subset of the data from the interview questionnaire including the request for permission to seek utility data. The responses to the mail questionnaire were used primarily to investigate potential nonresponse biases in the consumption data and to impute household and housing unit characteristics.

Response Rates

Personal interviews were completed at 3,842 households (85.2 percent) of the 4,507 occupied housing units, and mailed questionnaires were completed by an additional 239 households (5.3 percent). The overall response rate, including both personal household interviews and mailed questionnaires was 90.5 percent.

Energy Consumption Data and Missing Data

The energy consumption data for a sampled household was obtained from the utilities and fuel suppliers serving the household, provided the household consented and the suppliers cooperated. When consent was obtained, a supplier was requested to provide 12 consecutive monthly billing records, or other periodic billing data, spanning the nominal survey year April 1978 to March 1979. The success of these procedures is summarized in Table B1 (Appendix B): data were received from suppliers for 85.6 percent of the households using electricity and 75.2 percent of the households using natural gas. Disaggregation of these percentages and references to the procedures used to impute the missing data are also given in the Appendix.

The energy data used here are annual figures derived from the utility billing data. In cases when 11 or more months of data were obtained for a household, the annual figure is essentially the total reported consumption prorated to a nominal 365-day survey year. The adjustment procedure is described in Reference 5 and an updated report in Reference 6 (See also Appendix B, "Degree-Day Data").

Both the heating degree-day data, base 65 degrees Fahrenheit, and the cooling degree-day data, base 65 degrees Fahrenheit, used here are included in the NIECS public use file. These data represent 40-year averages computed on a per household basis and are adjusted for departures from a 46-year average (1930-1975) on a regional basis. The adjustment procedures are also described in Appendix B.

Use of a Subset of the Data

The analysis of natural gas consumption discussed in later sections is based on data from only those single-family detached households that participated in the personal interviews, and consumed some natural gas and whose annualized natural gas consumption estimate was based on at least 11 months of reliable utility data. The households used in the analysis of electricity consumption were chosen in an analogous manner. Restricting the analysis to single-family detached dwellings emphasizes the importance of this type of dwelling in the NIECS data and simplifies the task of describing variability. Extending the results of this analysis to other housing types will be the subject of a later report.

Basing the analyses on only those households with at least 11 months of energy data was done to avoid confusion between systematic variation in the data, from variation introduced by the methods used to impute missing data. We have no reason to suspect systematic defects in our imputation methods [References 5 and 6]; nevertheless, we choose initially to explore the variation in energy consumption using only households with unimputed energy data. Later, we will extend our analyses to the entire data set.

Variable Names and Definitions

The names and definitions of the variables used here are documented in Tables C1 and C2 (Appendix C). The variables fall into two classes, those which are available directly from the NIECS users data (Table C) and those which are derivations or transformations of the former (Table C2). The names of the variables in the first category correspond to the names used for the public users data. The definitions of derived variables are presented in the text as appropriate.

Section 2. A Theoretical Model for Household Energy Consumption

Theory for a Single Household

In a continuing empirical study of energy consumption in occupied homes in Twin Rivers, New Jersey, (References 7-11) the amount of heating fuel (H_i) needed on a daily basis to maintain a house at a desired fixed temperature has been described by

$$\text{(Equation 1)} \qquad H_i = B(P_r - T_{ai})^+ + E_i$$

where i is an index for days, T_{ai} is the average outside temperature, P_r is a reference temperature, B is an overall performance index for the house which characterizes its response to cold weather, and E_i is a stochastic error term with expected value zero. The quantity $(P_r - T_{ai})^+$ is equal to $(P_r - T_{ai})$ when the difference is positive, and is equal to zero otherwise. We assume nothing more about E for now.

According to the theory underlying the model, the reference temperature P_r carries information about factors under the control of residents, such as thermostat setting and the added heat load from the use of appliances (electric and gas), and some information about the thermal properties of the house (Reference 8). A lowered thermostat setting or increased use of appliances should result in a lowered value of P_r. The overall interpretation of P_r is that it indicates the warmest outdoor temperature at which the heating system must supply heat to the house, or the effective temperature it must maintain. Although not independent of the physical properties of the house, it is dominated by the actions of the residents.

The value of B, in this theory, is independent of the outside temperature and internal and solar heat gains, and indexes the house's average rate of heat loss and heating system efficiency, factors determined primarily by the design, construction, and maintenance of the house (Reference 8). Adding insulation or weatherstripping or overhauling the heating system should decrease the value of B.

The empirical analyses carried out in the Twin Rivers program generally support these interpretations of T_r and B (References 8-11). Differences in B-values among houses were observed to correspond in a way consistent with theory to differences in design and construction characteristics, such as number of bedrooms, presence of double-pane windows and compass orientation of the house, while reductions in B were observed to coincide with improvements in insulation (References 8-10). Also confirming the theory, variations in P_r were associated with changes in the behavior of occupants: in one-owner homes, the value of P_r changed considerably less than it did in homes that changed owners (Reference 8). Additionally, the value of P_r was lower among those with more bedrooms, possibly because the smaller houses tend to be occupied by people without children and are therefore more frequently unoccupied, resulting in lower average-interior temperatures (Reference 9).

Extension to the NIECS Sample

The possibility of using these single-home observations in an analysis of the NIECS data is appealing. In order to do this, some modifications to the above model must be introduced.

First, H_i in Equation 1 represents heating fuel consumption, exclusive of other fuel uses, such as cooking and water-heating. Let us introduce the term A_o which we define to be the average amount of energy used for other purposes so that Equation 1 becomes:

(Equation 2) $$H_i + A_o = A_o + B\,(P_r - T_{ai})^+ + E_i \,.$$

Let $F_i = H_i + A_o$; F_i represents total consumption of a specified fuel. Approximate $(P_r - T_{ai})^+$ by $DD_i = (65 - T_{ai})^+$, the number of heating degree-days, base 65 degrees Fahrenheit, for the i^{th} day. Summing over the days in a year, Equation 2 then becomes:

(Equation 3) $$F = \Sigma F_i = \Sigma(A_o + B(P_r - T_{ai})^+) + \Sigma E_i$$

$$= A + B\ DD_i + \Sigma E_i$$

$$= A + B\ (DD) + E$$

where DD is the heating degree-day total for the year. Let us introduce a subscript to denote households.

Then, Equation 3 becomes

(Equation 4) $$F_h = A_h + B_h\ (DD_h) + E_h$$

where:

F_h is the total consumption of fuel F by the h^{th} household in a given year.

A_h summarizes the household's base load consumption, the amount of fuel needed for cooking, water-heating, etc.

B_h is a coefficient which summarizes the energy characteristics of the dwelling unit.

DD_h is the annual degree-day figure for the household, and E_h is a random error term, the structure of which we leave unspecified in this report.

The Modeling Strategy

According to Equation 4 and the above theory, A_h depends heavily on the behavior of the households or the characteristics of their appliances, and much less so on the physical properties of the dwelling unit. A_h remains constant over time. On the other hand, B_h is seen as depending on the thermal characteristics of the dwelling unit and is constant over time.

In practice, neither of the above statements is likely to be strictly true; one can expect a good deal of interaction between behavioral factors and physical factors affecting energy consumption, and both coefficients will change over time as the composition of the household and features of the dwelling unit change. Nevertheless, Equation 4 suggests the strategy for describing the variability in energy consumption which has been adopted for this report: the coefficients A_h and B_h are set up as simple functions of items in the NIECS data, and a methodology for fitting the model is developed. Thus we consider models of the form of Equation 4 with

$$A_h = \sum_{k=1}^{K} A_k X_{hk} + A_o$$

a linear combination of K coefficients A_k weighted by the value of an item X_{hk} from the NIECS data. Similarly, we express

$$B_h = \sum_{l=1}^{L} B_l Y_{hl}$$

Note that the coefficients A_k and B_l are constant across households.

Pursuant to our strategy, we emphasize characteristics X_{hk} of the household in the development of the coefficients A_h and characteristics, Y_{hl}, of the housing unit in the development of B_h.

We can introduce terms to model the fuel consumption due to air conditioning in a similar manner. Let G_h be the portion of F_h used for air conditioning. Model G_h by

$$G_h = \left(\sum_{m=1}^{M} C_m Z_{hm} \right) CC_h$$

where CC_h is the cooling degree-days for household h. The variables Z_{hm} are characteristics of the housing unit or household. The Z_{hm} are defined to be zero if the household does not have air conditioning. The coefficients C_m are constant across households.

Fitting Methodology

Substituting the expression for A_h and B_h in Equation 4 and adding in G_h we get

(Equation 5)

$$F_h = A_o + \sum_{k=1}^{K} A_k Y_{hk} + \left(\sum_{L=1}^{L} B_l Y_{hl} \right) DD_h$$

$$+ \left(\sum_{m=1}^{M} C_m Z_{hm} \right) CC_h + E_h$$

which, when fully expanded, is a model linear in the parameters A_k, B_l, and C_m. This suggests the use of ordinary least squares methodology to fit the parameters. A variation of this approach which discounts outlier observations is the subject of Section 4.

Summary

The Twin Rivers experiments and the preceeding discussion provide a theoretical underpinning amid a body of empirical evidence for the approach taken here. However, there are vast differences between the NIECS data and the Twin Rivers experiments, and significant omissions in our preliminary statistical analysis, which must be taken into account before the results of our analysis can be accepted as a definitive, or even adequate, description of the variability in household energy consumption. This report is a first effort to provide that description.

Section 3. Graphical Summaries of Consumption Data

In this section, we present seven plots that illustrate trends in residential consumption of natural gas and electricity as a function of degree-days. The plots anticipate the modeling strategy adopted in Section 4 and outlined in Section 2. They also illustrate the great amount of variability of energy consumption among households living in single-family detached dwellings. The following is a list of the seven plots:

- Figure 1 is a plot of total natural gas consumption versus heating degree-days for households that use natural gas as the main heating fuel. The figure shows total gas consumption, with heating and nonheating uses combined.

- Figure 2 is a plot of natural gas consumption versus heating degree-days for households that consume some natural gas, but do not use natural gas as the main space-heating fuel.

- Figure 3 is a plot of total electricity consumption versus heating degree-days for households that use electricity as the main space-heating fuel.

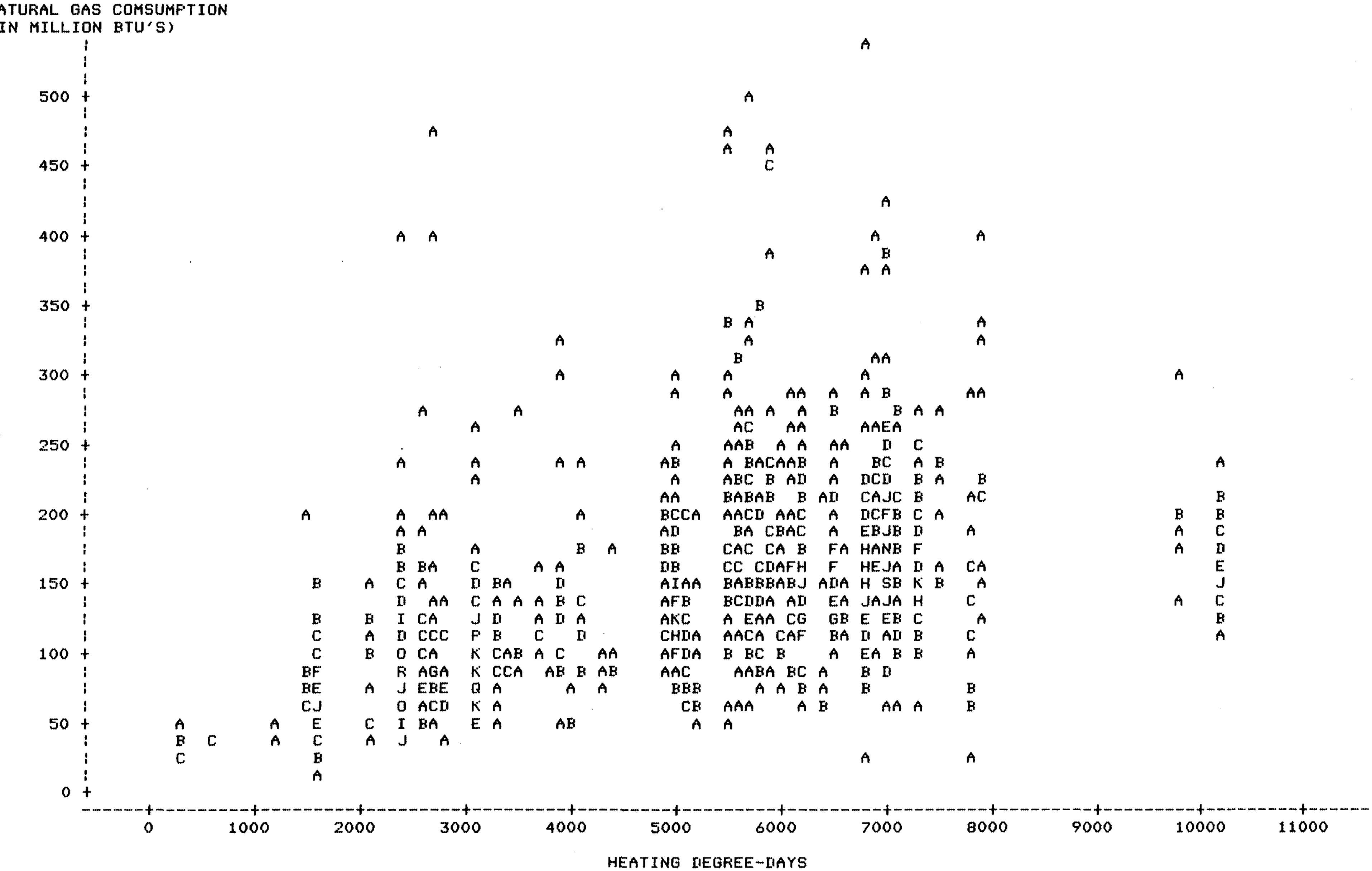
- Figure 4 is a plot of total electricity consumption versus heating degree-days for households that do not use electricity for air conditioning or as the main space-heating fuel.

- Figure 5 is a plot of total electricity consumption versus cooling degree-days for households that use electricity for air conditioning, but not as the main space-heating fuel.

- Figure 6 supplements Figure 1 and Figure 7 supplements Figure 3. The construction and meaning of these plots is discussed below.

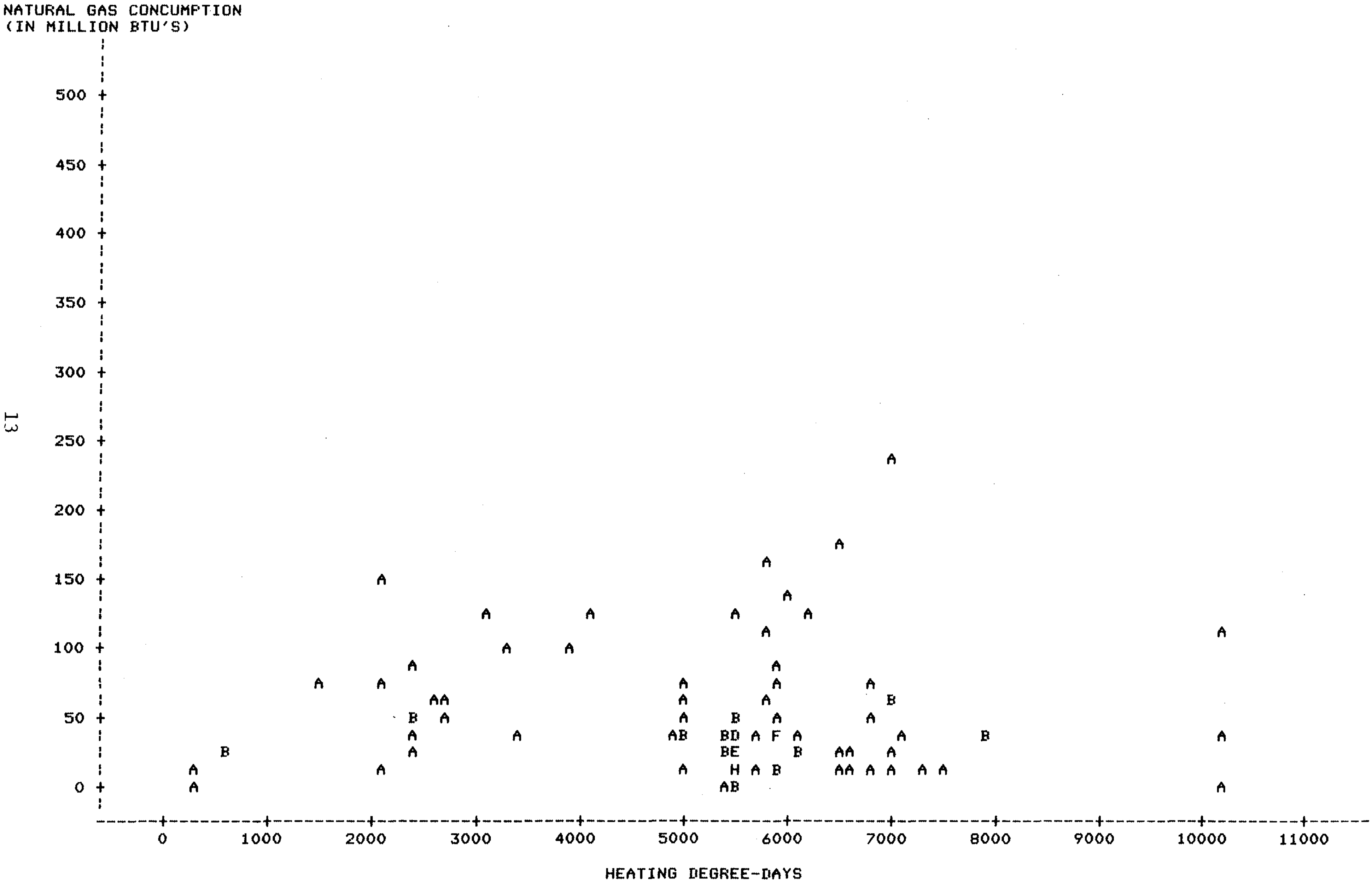
In Viewing Figures 1 and 3, it is obvious that consumption of the main heating fuel is positively correlated with degree-days. In comparing the figures, it is obvious that households using natural gas for heating tend to use one-and-a-half to two times as many Btu's as houses using electricity for the same purpose. The natural gas houses tend to center around a line between 150×10^6 Btu and 200×10^6 Btu whereas the electrically heated houses center around a line between 75×10^6 Btu and 100×10^6. However, this difference varies substantially over the range of degree-days considered. We will return to this point when we consider Figures 6 and 7 and the relationship between consumption and degree-days evidenced in the present figures. (These differences do not imply that residential space-heating is an inefficient use of natural gas. A powerplant would consume several Btu's of natural gas in generating and delivering one Btu of electricity to a residence.)

FIGURE 1. NATURAL GAS CONSUMPTION BY HEATING DEGREE-DAYS FOR HOUSEHOLDS THAT USE NATURAL GAS FOR THE MAIN SPACE-HEATING FUEL.



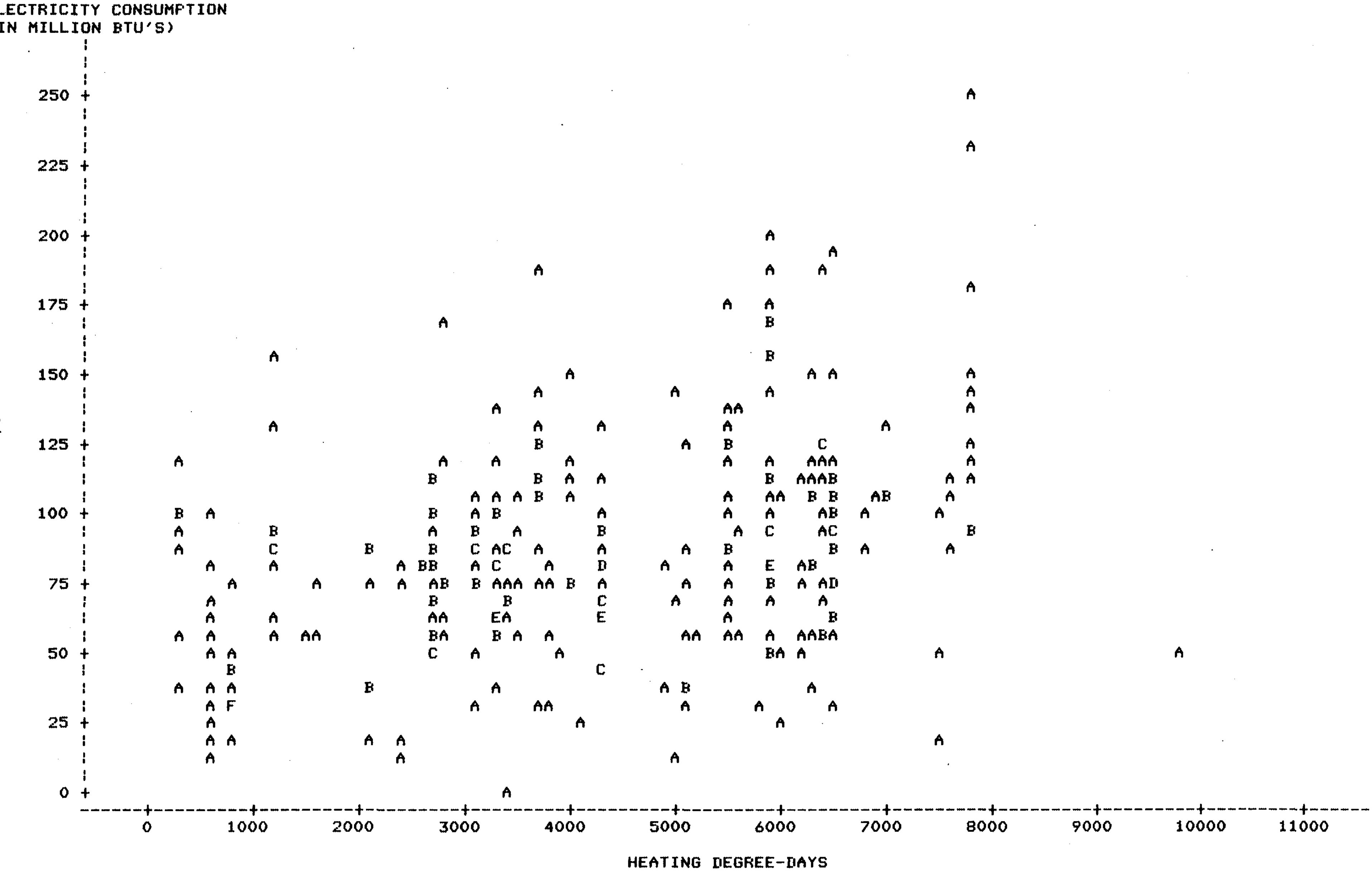
LEGEND: A = 1 OBSERVATION, B = 2 OBSERVATIONS, ETC.

FIGURE 2. NATURAL GAS CONSUMPTION BY HEATING DEGREE-DAYS FOR HOUSEHOLDS THAT DO NOT USE NATURAL GAS FOR THE MAIN SPACE-HEATING FUEL.



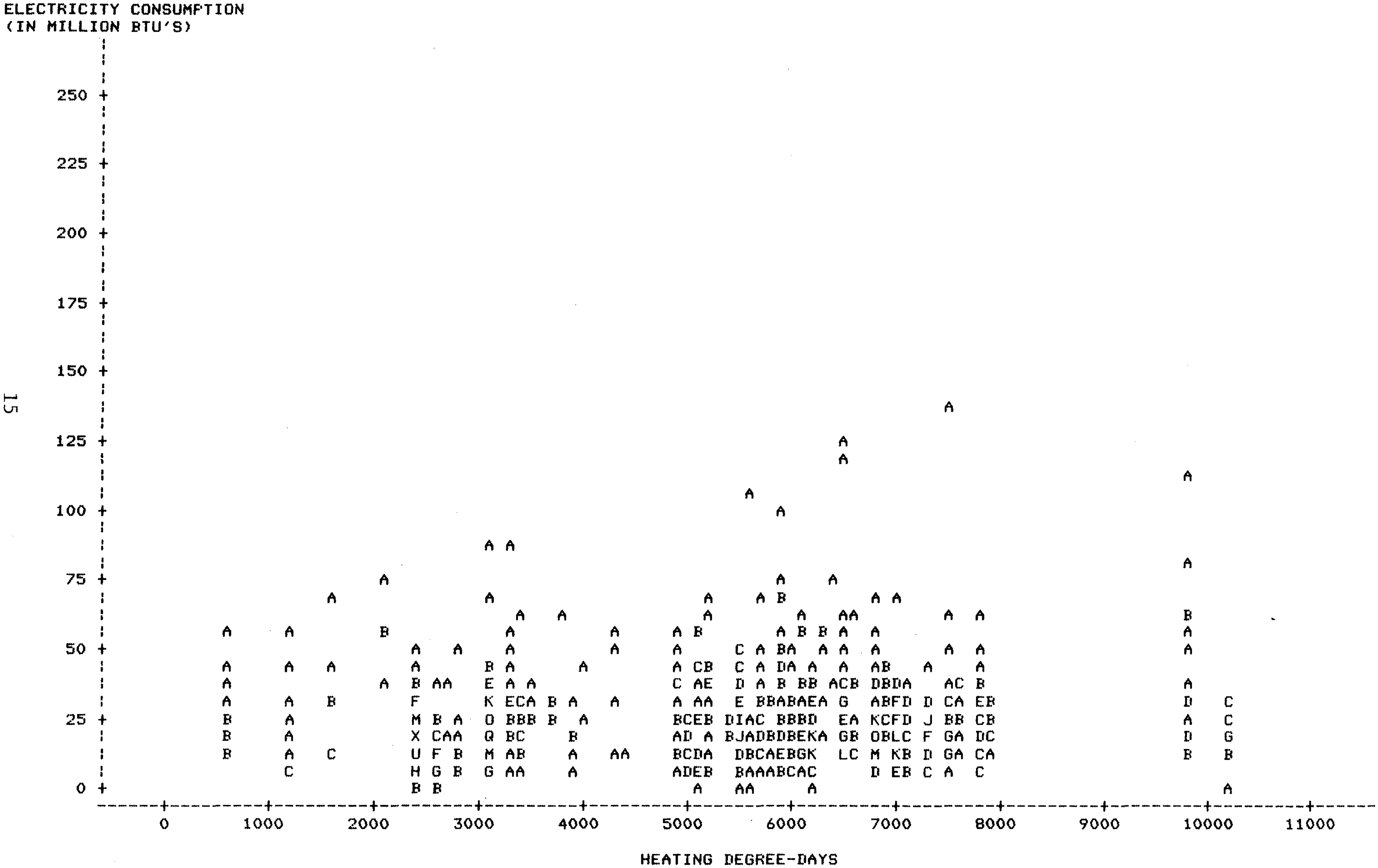
LEGEND: A = 1 OBSERVATION, B = 2 OBSERVATIONS, ETC.

FIGURE 3. ELECTRICITY CONSUMPTION BY HEATING DEGREE-DAYS FOR HOUSEHOLDS THAT USE ELECTRICITY FOR THE MAIN SPACE-HEATING FUEL.



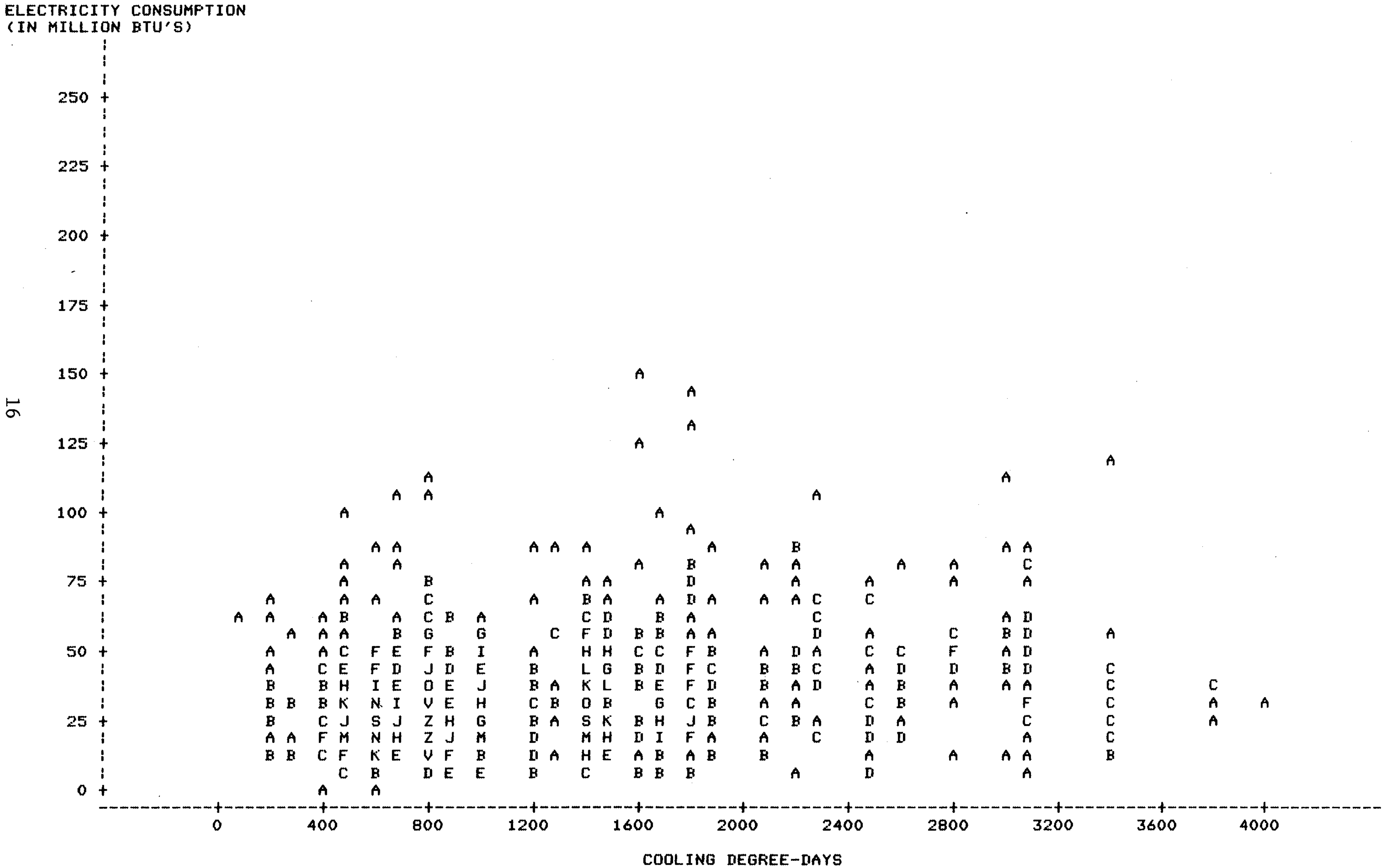
LEGEND: A = 1 OBSERVATION, B = 2 OBSERVATIONS, ETC.

FIGURE 4. ELECTRICITY CONSUMPTION BY HEATING DEGREE-DAYS FOR HOUSEHOLDS THAT DO NOT USE ELECTRICITY FOR AIR CONDITIONING OR FOR THE MAIN SPACE HEATING FUEL.



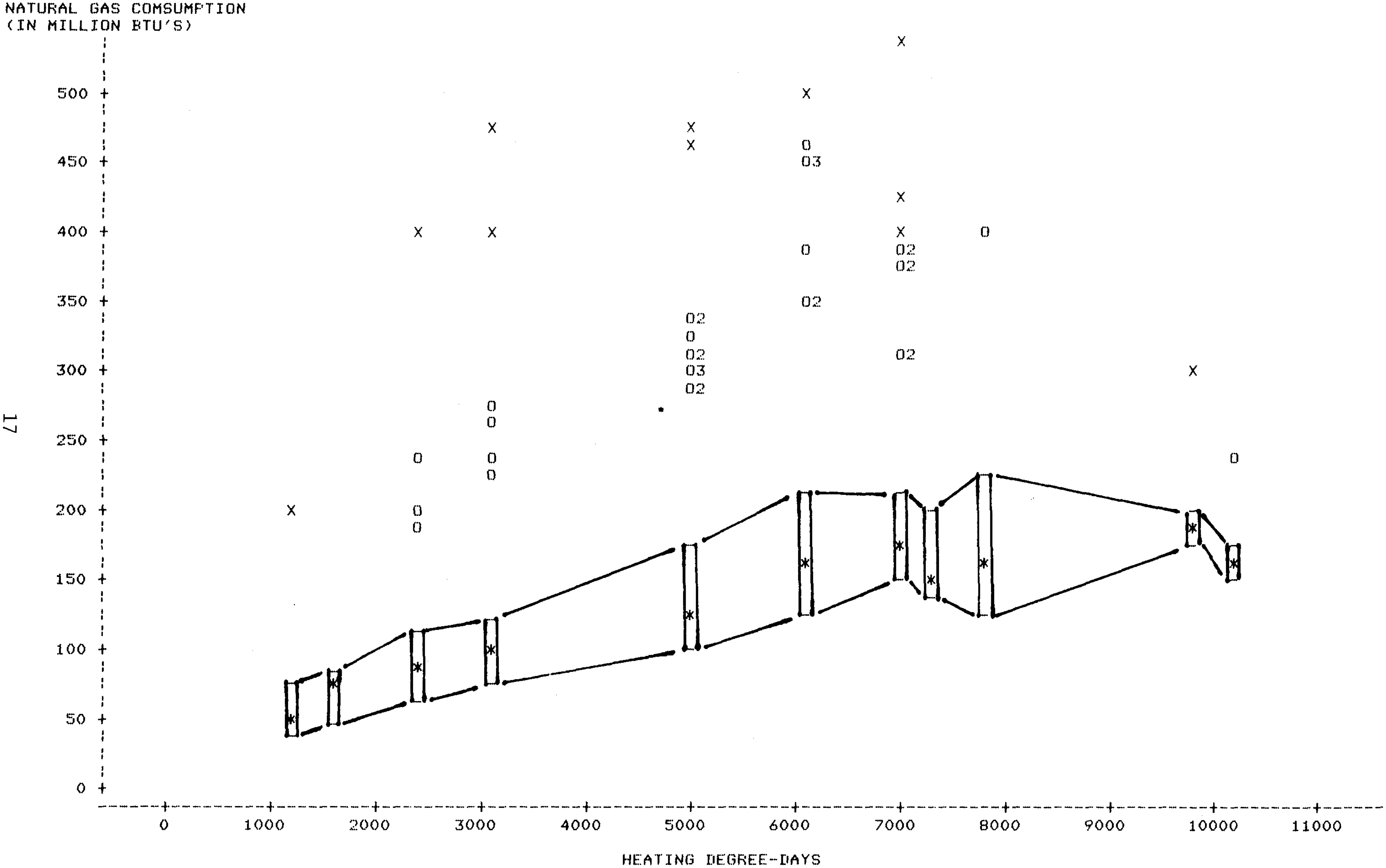
LEGEND: A = 1 OBSERVATION, B = 2 OBSERVATIONS, ETC.

FIGURE 5. ELECTRICITY CONSUMPTION BY COOLING DEGREE-DAYS FOR HOUSEHOLDS THAT DO NOT USE ELECTRICITY FOR THE MAIN SPACE-HEATING FUEL BUT DO USE ELECTRICITY FOR AIR CONDITIONING.



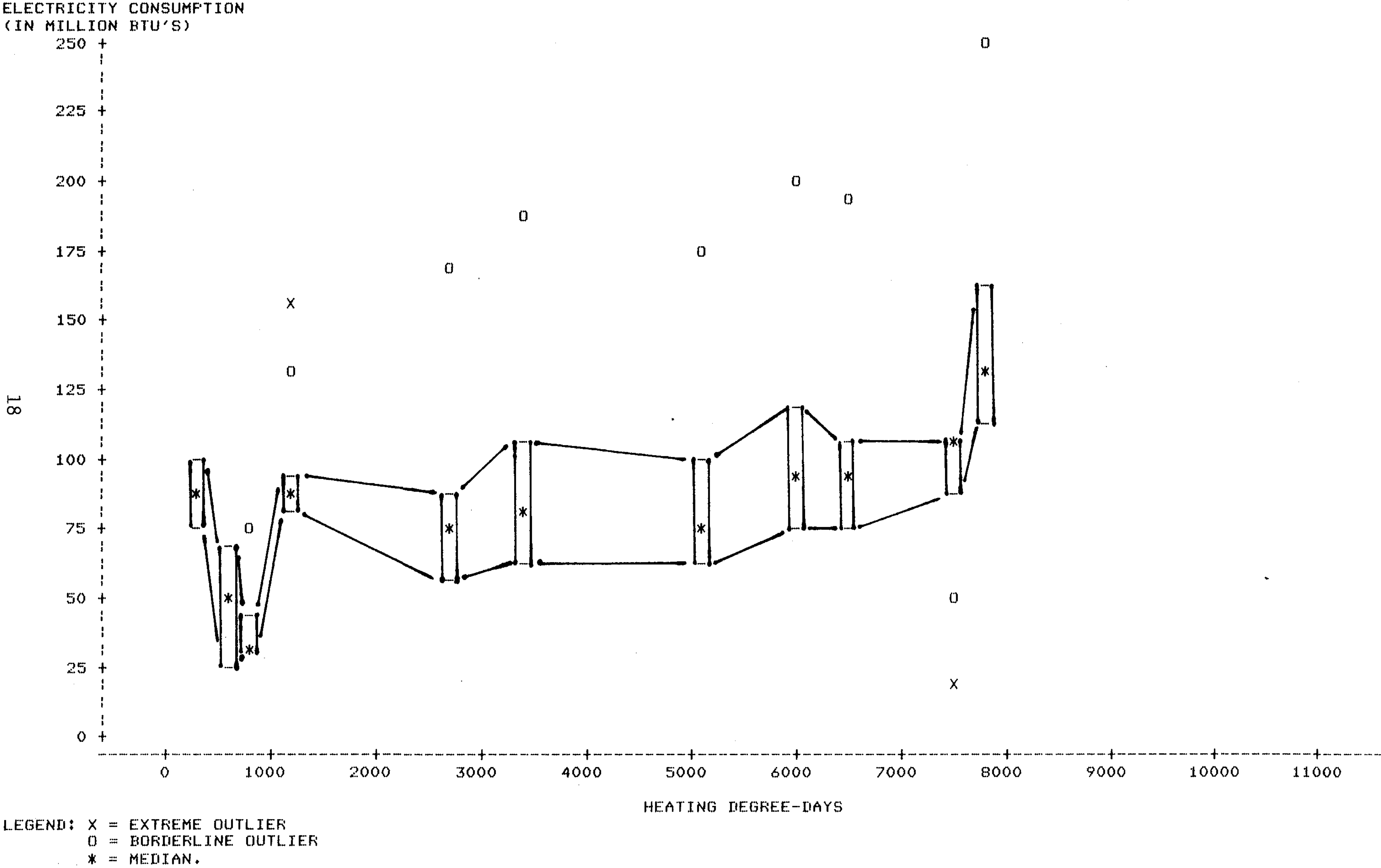
LEGEND: A = 1 OBSERVATION, B = 2 OBSERVATIONS, ETC.

FIGURE 6. WANDERING BOX PLOT OF NATURAL GAS COMSUMPTION BY HEATING DEGREE-DAYS FOR HOUSEHOLDS THAT USE NATURAL GAS FOR THE MAIN SPACE-HEATING FUEL.



LEGEND: X = EXTREME OUTLIER
O = BORDERLINE OUTLIER
* = MEDIAN.

FIGURE 7. WANDERING BOX PLOT OF ELECTRICITY CONSUMPTION BY HEATING DEGREE-DAYS FOR HOUSEHOLDS THAT USE ELECTRICITY FOR THE MAIN SPACE-HEATING FUEL.



Figures 2 and 4 provide a dramatic contrast to Figures 1 and 3 and emphasize the relationship between heating and degree-days observed in the latter. In addition, the large difference in consumption between the two energy types is gone. Natural gas consumption in Figure 2 tends to center around 25 x 10^6 Btu (although the small number of observations and large variability tend to obscure this) and so does electricity consumption in Figure 4.

Figure 5 is a plot of electricity consumption versus cooling degree-days among households that use electricity for air conditioning and other uses but not for heating. The figure is stunning for its lack of any discernible association between the two variables. On the face of it, the plot seems to indicate that the demand for cooling does not vary with climatic conditions. In our opinion, however, it more than likely expresses the need to correct for other variables in modeling the demand for cooling and the deficiencies of cooling degree-days, base 65 degrees Fahrenheit, as a measure of the warmth of a day or a region, or as an indicator of peoples' demand for cooling. Figures 6 and 7 are intended to supplement Figures 1 and 3 and reveal more of the relationship between energy consumption and heating degree-days. Construction of these figures is described in detail in Appendix D. A brief summary is: the data are divided into groups according to the distribution of the abscissa; within groups, the median and first and third quartiles of the ordinate are calculated. Then the statistics for the ordinate are plotted by groups and centered over the group medians along the abscissa. Outlying points are noted to give a sense of the variability of the data, and group statistics are linked by line segments to give a sense of the structure in the data.

In viewing Figure 6, a tendency is observed for gas consumption to rise smoothly from the lower end of the degree-day scale to the 6,500 to 7,500 degree-day range. The rate of rise is about 100 x 10^6 Btu per 5,800 degree-days or about 17 cubic feet per degree-day where 1 cubic foot is valued at 1,020 Btu's. Over this entire range, the outlying values occur only on the upper end of the consumption scale. Within groups, outlying values as large as two and three times the group median are not uncommon.

Starting in the range 7,000-7,500 degree-days, the association between gas consumption and degree-days seems to change markedly. At this point, the group medians level off, tending to remain constant at around 175 x 10^6 Btu even at 10,200 degree-days. Unfortunately, we have no data in the range from 8,000 to 9,700 degree-days and it is therefore difficult to write with certainty of a change in trend.

There are several hypotheses explaining the drop in the upper end of the energy-consumption-per-degree-days curve observed in Figures 1 and 6:

i. Sampling variability: the households located at 9,800 and 10,200 may be lower than "expected" relative to the other households simply because of random sampling variability. In this case, the presumed change in trend is probably a feature of the sampled households but not of the population of all households.

ii. Systematic differences: the change in trend may go beyond sampling variability so that the population of households at the upper end of the degree-day range is indeed different than the balance of the population. In this case, the change in trend is indicative of a real difference which might be due to:

--Better home insulation or construction than the general population.

--Acclimation to colder climate; households in northern climates may have adjusted better to cooler interior temperatures than their southern counterparts.

--More conscientious energy conservation: Households in northern climates may be more conservative in their use of energy resources. Higher natural gas prices in the Northern climates may contribute to this.

--Greater reliance on secondary heating systems: Households in northern climates may make greater use of secondary heating systems, such as fireplaces or portable heaters, than their southern neighbors. These secondary uses are not accounted for in Figures 1 and 6 unless the secondary and main fuels are both natural gas.

iii. Problems with the heating degree-day measure: heating degree-days are a convenient summary of the coldness of a day. However, they may not be a satisfactory measure of the demand for heat in colder climates. Alternatively, the degree-day base used here, 65 degrees Fahrenheit may not be the best choice. A lower degree-day base would tend to squeeze the plots, a higher base expand them; depending on how much the plots are changed, they could become more linear.

Turning to Figure 7, we observe a rather different relationship for electricity consumption versus heating degree-days. The Figure starts off with a sharp drop from a center of about 87.5×10^6 Btu down to one of about 25×10^6 Btu and a rise back to its starting point all in the range from 300 to 1,200 degree-days. We believe this rapid variation is due either to systematic differences independent of the use of electricity for space-heating and cooling, or simply random variation.

After this initial burst of variation, the data in Figures 3 and 7 tend to settle down, rising slowly but steadily from about 75×10^6 Btu at the 2,800 degree-day mark to about 125×10^6 Btu at 7,800 degree-days, a rate of 50×10^6 Btu over a 5,000 degree-day range. In equivalent cubic feet of gas, this rate is 9.8 cubic feet per degree-day.

In summary, there are hypotheses which can account for the structure of Figures 1 and 6 and Figures 2 and 7. If the leveling off in Figure 6 is indicative of systematic differences, then it may be possible to account for them in the subsequent analysis. We return to this point in Sections 4 and 5.

The rather distinct differences between Figures 1 and 6 and 2 and 7 indicate that the heating uses of electricity and natural gas are very different phenomena which should be treated in separate models. This is the approach we have taken.

Section 4. Regression Results

The principle aim of this project is to develop linear models to explain the variability in household electricity and natural gas consumption. The models being developed are based on the theory outlined in Section 2 using selected household variables (see Table 3), transformations of these variables, and interaction terms. Natural gas consumption was modeled separately from electricity consumption as suggested by the graphs in Section 3. A single model was used within each fuel type.

Variables Used in the Models

From the large number of variables in the NIECS data, only a few dozen were considered for use in modeling. Of these, several were discarded because of known or suspected problems with the data. This section discusses the candidate variables and why some were used and others weren't. The variables are organized into groups: demographic characteristics, measures of size and insulation, measures of potential air conditioning usage, indicators of fuels used for various purposes, degree-day variables, and appliance indicators.

- Demographic Characteristics: The specific variables used were: number of individuals in the household (NHSLDMEM), age of respondent (NAGE01), and household income (KINCOME).

- Measures of size and insulation: Descriptive housing unit variables relating to the size of the house and limited measures of insulation formed the second group.

  Respondents were asked the size of their home and about the presence and amount of insulation. Unfortunately, nonresponse to these items was so large that we considered the data to be unusable for this analysis. In subsequent surveys, we plan to measure the square footage of the heated and unheated sections of individual housing units.

  Surrogate variables were used to estimate the size of the housing unit. These variables were the number of rooms (NROOMS), number of doors and windows (NDRSAWS), number of bathrooms (NBATHRMS) and the total number of rooms, doors, and windows (NTOTAL). NROOMS can be obtained directly from the public use file. NDRSAWS, NBATHRMS and NTOTAL are transformations of variables listed in Table C1 (Appendix C); equations for these transformed variables are found in Table C2.

  In computing NBATHRMS, the variables NCOMBATH and NHAFBATH are used. For some houses that do have complete plumbing, one or both of the last two variables are missing. For these households we assumed that NCOMBATH and NHAFBATH are actually zero.

We could only reliably obtain indirect measures of the insulation properties of the sampled houses. One measure is the number of storm doors and windows, NSDRSAWS. This term measures not only the effect of using storm doors and windows but also the propensity of houses with storm doors and windows to be better-insulated in other ways. For example, we expect a positive correlation between the number of storm windows and the amount of insulation. The other indirect measure is KYHSBREC. This variable codes the age of the house on an ordinal scale. The oldest houses are code 1 and the newest houses are code 7.

We are currently investigating the possibility that variables describing energy conservation activities will provide an additional indirect measure of insulation.

- Measures of Potential Air Conditioning Usage: the next group of variables gives the number of rooms that can be air conditioned by various types of equipment. The resulting variables are NRMELCAC, NRMELRAC, and NRMGASAC. These variables represent the number of rooms that can be air conditioned by electric central units, electric room units, and gas central units, respectively. The logic used to compute these variables is described in Table C2 (Appendix C). One of the variables used in the logic is NROOMAC. This variable is missing for some of the households that stated that they had central air conditioning. For these households, we set NROOMAC equal to NROOMS.

- Fuels by End-Use: The fourth group contains variables that describe the fuel used for water and space-heating. The variables are HELWHT, HGASWHT, HELMHT, HGASWHT, HSBNMELH, HSBNMGSH, HELHTPUM and KMHTEQRC. These variables are all transformations of variables that are listed in Table C1.

The air conditioning terms in the models always involve cooling degree-days as well as the number of rooms air conditioned by different types of equipment. The variable KINCOME is also interacted with the above variables to represent the lifestyle of the households. The terms used for modeling the consumption of electricity to air conditioning are NCOOLDD x NRMELCAC, NCOOLDD x NRMELRAC, NCOOLDD x NRMELCAC x KINCOME and NCOOLDD x NRMELRAC x KINCOME.

Fewer people use natural gas as fuel for air conditioning purposes. As a result, only the single term, NCOOLDD x NRMGASAC was significant in the natural gas consumption model.

- Degree-Day Variables: The fifth group gives the heating and cooling degree-days. The variables used were NHEATDD and NCOOLDD; both are based at 65 degrees Fahrenheit. These are the degree-days derived from 46-year averages as described in Appendix C. A new public use file will be issued shortly providing the approximate degree-days experienced by the households for several bases.

A preliminary analysis indicates that models using heating degree-days based at 60 degrees Fahrenheit and cooling degree-days based at 70 degrees Fahrenheit result in a slightly higher R^2 than models with degree-days based at 65 degrees Fahrenheit. This preliminary analysis did not investigate all possible bases. Due to the incompleteness of the analysis, we decided to use the base 65 degrees Fahrenheit degree-days that were obtained from the 46-year averages for this report.

Appliance Indicator: The last group of variables describes the major appliances that are contained in the household. The variables are NELCKDV, NELFRIG, HELDISHW, HELCLSDY, HSPFDFRZ, HAUTOWASH and NGASNDX. All of these except NELCKDV and NELFRIG are listed in Table C1 (Appendix C). These two additional variables are defined and described in Table C2.

In defining NELFRIG, three extra variables are defined first. These variables are HELFRIG1, HELFRIG2, and HELFRIG3. The details of these three extra variables are found in Table C2. The last variable (HELFRIG3) is an indicator variable that denotes if the third refrigerator in the household is fueled by electricity. The public use files does not list the fuel for the third refrigerator. Hence, if the first two refrigerators were electric, we assumed that the third one was also electric. Fortunately, all of the households in the sample that had three refrigerators also had only electric refrigerators for the first two that were listed.

All of the variables listed above except NGASNDX correspond to major electrical appliances. All of these appliances are commonly found in households. Hence, the model can attach an electric consumption figure to each of these appliances separately. Unfortunately, the corresponding gas appliances are rare and the noise in the gas consumption model obscures their contributions. Hence, the gas appliances cannot be treated individually using our data set. Instead, we are using the variable NGASNDX, an index for gas appliances. It is a combination of the effects for all of the gas appliances contained in the household, exclusive of water- and space-heating appliances.

- Interaction Terms: Various multiplicative combinations, or interactions, of the items mentioned above were also tried in the models. Interaction terms are suggested explicitly in Equation 5, Section 2, where housing unit characteristics are interacted with degree-days.

The interaction terms can be broken up into three groups, water-heating, space-heating, and air conditioning terms. Within each group, the terms can be divided into two subgroups, one for the electric consumption model and one for the natural gas consumption model.

Several interaction terms were tried when modeling the electricity consumption due to water-heating. The two terms that we found significant were HELWHT x NHSLDMEM and HELWHT x HELDISHW. For natural gas consumption, we only used the term HGASWHT x NHSLDMEM. The coefficient for the term HGASWHT x HELDISHW was positive, but it was only marginally significant.

These terms seem reasonable in that each additional household member is likely to increase the usage of hot water and some electric dishwashers tend to use a large amount of hot water. Also, some models require that the hot water heater be set at a higher temperature than if they were not present.

Turning to space-heating interactions, we note that the space-heating component of total usage is potentially much more complex than the water-heating component. In particular, the amount of fuel consumed for space-heating involves not only the fuel used, but also the heating degree-days, size of house, type of equipment, indirect insulation measures and indirect measures of the lifestyle of the household.

For natural gas consumption, there is no space-heating component unless the main or secondary heating fuel is natural gas. Hence, all of the other factors involved with space-heating need to be interacted with HGASMHT or HSBNMGSH. The type of equipment use for main heating is represented by the term KMHTEQRC, which might be thought of as system overhead, or start-up energy requirements. The size of the house comes in through the terms, HGASMHT x NHEATDD x NBATHRMS, HGASMHT x NHEATDD x NROOMS, and HGASMHT x NHEATDD x NDRSAWS. Note that the variables measuring the size of the house are always interacted with HGASMHT x NHEATDD. The lifestyle of the household is represented by the term HGASMHT x NHEATDD x NAGE01.

Finally, the terms HGASMHT x NHEATDD x KYHSBREC and HGASMHT x NHEATDD x NSDRSAWS indirectly represent the insulation properties of the dwelling. The component for secondary gas space-heating is not as easily broken up as the component for main gas space-heating. As a result, the only term used for modeling the secondary gas space-heating is the dummy variable HSBNMGSH.

The interaction terms used in the analysis of electricity were HELMHT x NHEATDD x NAGE01, HELMHT x NHEATDD x NBATHRMS, HELMHT x NHEATDD x NDRSAWS, HELMHT x NHEATDD x NSDRSAWS, and HELMHT x NHEATDD x HELHTPM.

Fewer terms were significant in the analysis of electricity than were significant in the analysis of gas consumption; fewer people heat with electricity than heat with gas. All of the terms listed except the last correspond to terms used in the gas analysis. The exceptional term represents to some degree the effect of using an electric heat pump.

Fitting by Iterative, Weighted Least Squares

Equations of the form described in Section 2 were fit to the data using iterative, weighted least squares and an outlier rejection procedure. The general linear model procedure of the SAS statistical package was used throughout. This section describes the techniques used to model the data.

Iteration: In the first step of each analysis, parameters were fit to the data using ordinary least squares. For each succeeding step up to a total of seven, new parameters were fit using a weighted least squares procedure with the weights equal to the reciprocal of the estimated consumption from the previous step. The weights reflect the observation that there is a larger variance in energy usage among households with high usage than among those with low usage. (See Figures 8 and 9.)

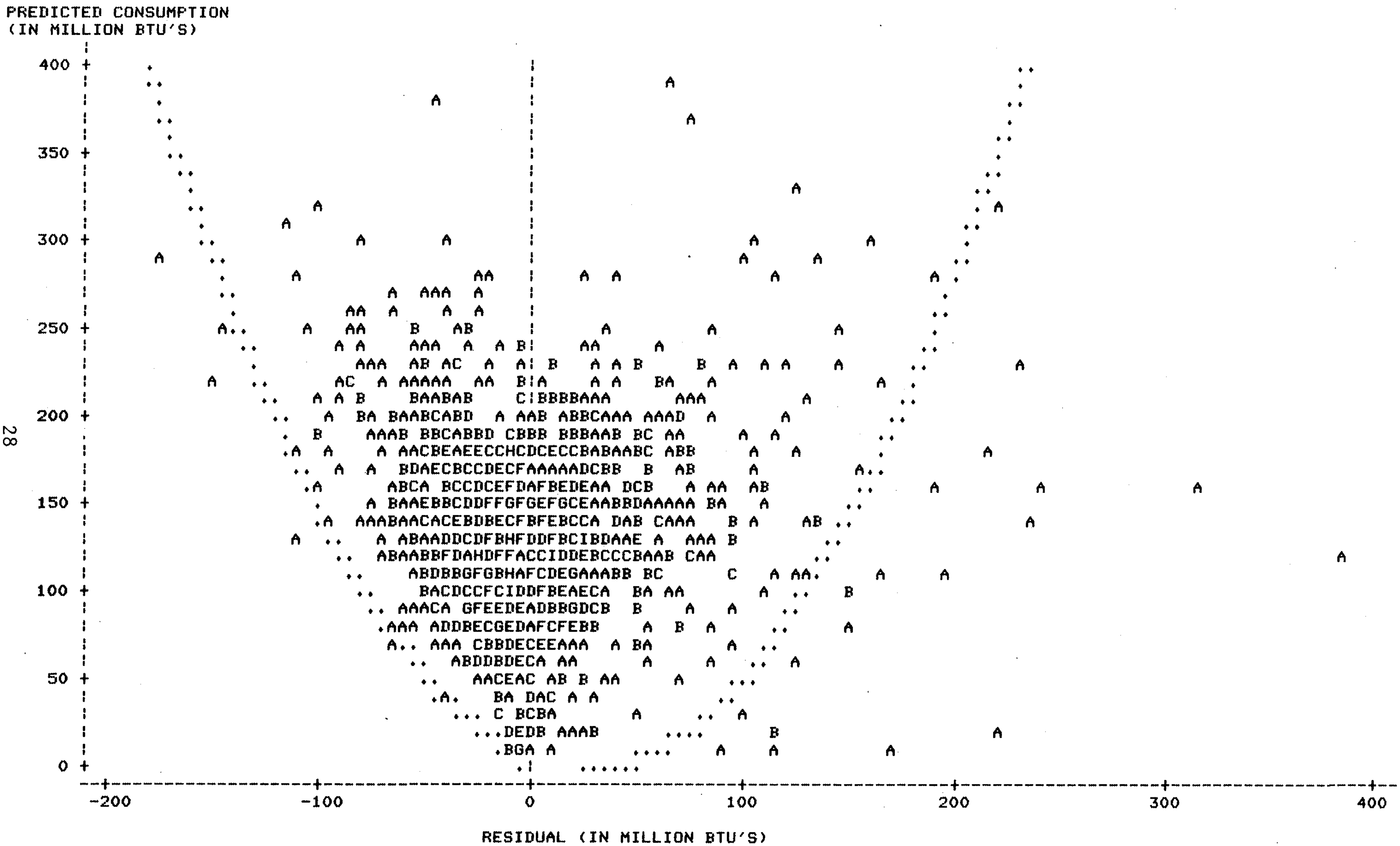
At the third iteration, we began systematically eliminating outliers before refitting the models. The process of elimination and refitting was continued until further outliers were removed, or until the seventh iteration was reached, as mentioned above.

The plots of residuals versus predicted consumption, Figures 8 and 9, revealed two outlier problems. One was the presence of outlying values, values far from the main body of residuals; and the other was the asymmetry of the residuals. Our method of dealing with the first problem was to eliminate values outside an interval determined by the standard error of the regression. Our method for dealing with skewness was an asymmetric interval based on a square root transformation of actual and predicted consumption after fitting the model. Methods for dealing with skewness based on transforming the dependent variables, energy consumption, and using it in the regression were rejected due to a desire to preserve the linear structure of the model. Using the logarithm of the energy consumption in fitting the model would result in a multiplicative model.

Let SRESID be the difference between the square root of the energy consumption and the square root of the predicted energy consumption. The values of SRESID were more normally distributed than the values of the residuals. A standard deviation was calculated based on SRESID using the equation:

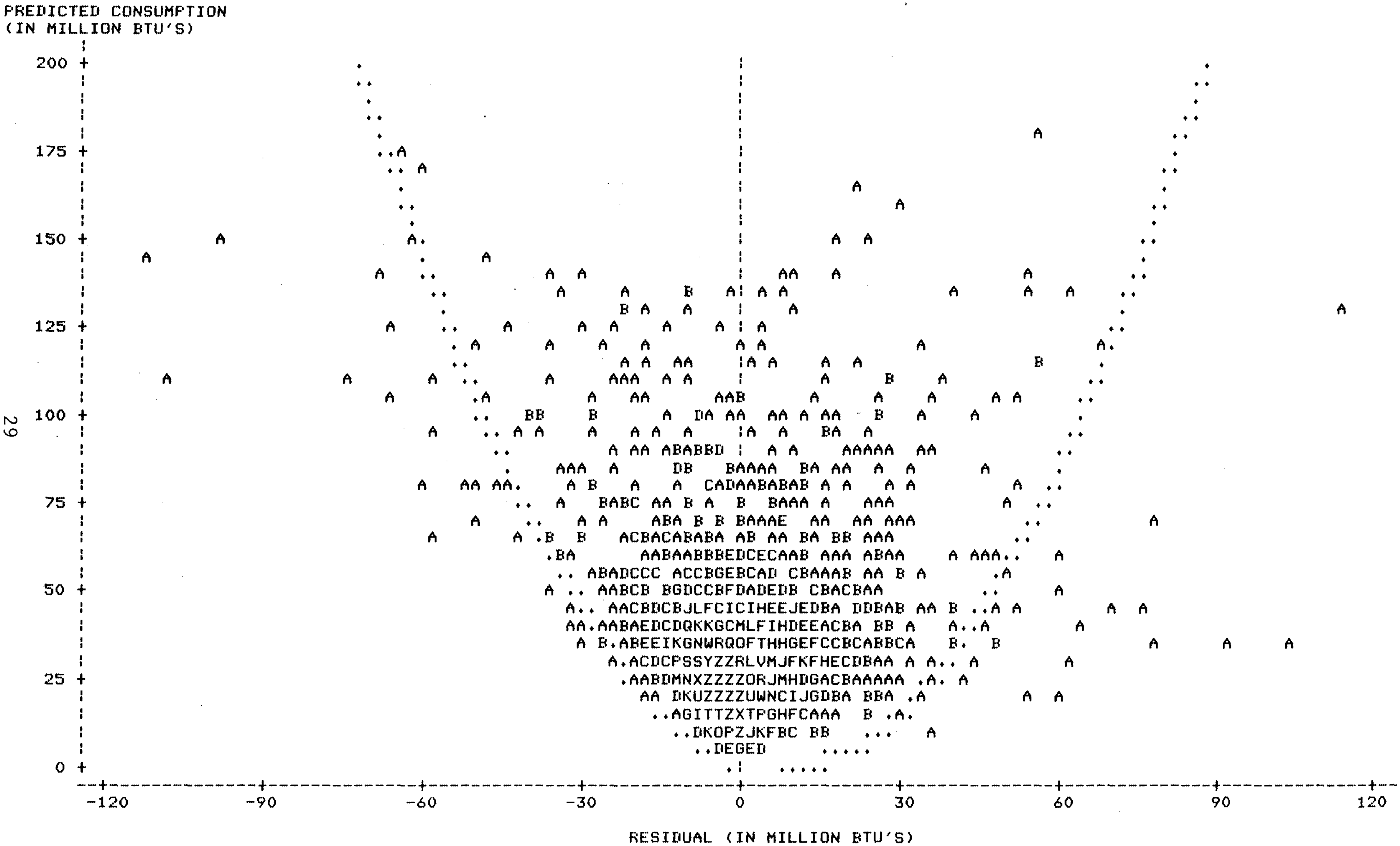
$$\text{Standard Deviation} = [\ \Sigma\ (\text{SRESID} - \overline{\text{SRESID}})^2/(\text{N}-\text{K})\ ]^{1/2}$$

FIGURE 8. OUTLIER DETECTION FOR NATURAL GAS MODEL: PREDICTED CONSUMPTION BY RESIDUAL.



LEGEND: A = 1 OBSERVATION, B = 2 OBSERVATIONS, ETC.
LIMITS FOR OUTLIERS SYMBOL USED IS '.'
ZERO AXIS SYMBOL USED IS '¦'

FIGURE 9. OUTLIER DETECTION FOR ELECTRICITY MODEL: PREDICTED CONSUMPTION BY RESIDUAL.



LEGEND: A = 1 OBSERVATION, B = 2 OBSERVATIONS, ETC.
LIMITS FOR OUTLIERS SYMBOL USED IS '.'
ZERO AXIS SYMBOL USED IS '¦'

where N is the number of observations, K is the number of parameters estimated in the linear model and $\overline{\text{SRESID}}$ is the mean value of SRESID. Any household whose value of SRESID was more than three such standard deviations from zero was rejected as an outlier. When outliers were deleted, the value of the standard deviations was decreased at the next iteration. Hence, additional households were deleted after each step.

In Figures 8 and 9, the curves represented by the dots show the final outlier detection regions. The curves are not symmetrical about zero and the distance from the zero vertical axis increases as YHAT increases.

Consumption Models

The estimated natural gas consumption model that was obtained in the final step is summarized in Tables 1 and 2. In construction of the model, 28 households were deleted as being outliers and were not used. The ID numbers, NSQIDDOE, of these outliers are listed in Table 3.

Five of the deleted households listed electricity as their main heating fuel, but they consumed large amounts of natural gas. This suggests that the main heating fuel is actually natural gas and not electricity. For most of the outliers, there is no obvious reason why the model does not fit. One exception is a single-person-household in a large house. This suggests that some of the rooms in the house may have been left unheated during the winter. Another exception is a household that used natural gas to heat a swimming pool. This fact came from a remark placed on the questionnaire by the interviewer. The RECS survey will ask all households if they heat swimming pools.

If the outliers are not removed, the intercept term in the natural gas consumption model is greatly increased. On the other hand, the effect on the other terms is only moderate.

The estimated electricity consumption model that was obtained in the final step is summarized in Tables 4 and 5. In constructing the models, 51 households were detected as outliers and were not used in constructing the model. The ID numbers (NSQIDDOE) of these households are listed in Table 6.

Four households are listed in Tables 3 and 6. Three of these appears to have the main heating fuel incorrectly listed. The other one consumed abnormally large amounts of both natural gas and electricity.

If the outliers had not been deleted when we constructed the electricity consumption model, the results would only be slightly different.

TABLE 1. ANALYSIS OF VARIANCE TABLE FOR NATURAL GAS MODEL (USING NCNGYRB AS THE DEPENDENT VARIABLE).

SOURCE	DEGREES OF FREEDOM	WEIGHTED SUM OF SQUARES	MEAN SQUARE	F-VALUE	ATTAINED LEVEL OF SIGNIFICANCE	WEIGHTED R-SQUARE
MODEL	13	44995013.38210071	3461154.87554621	290.05	0.0001	.752079
ERROR	1243	14832524.60716271	11932.84360995			
CORRECTED TOTAL	1256	59827537.98926342				

TABLE 2. WEIGHTED LEAST SQUARE SOLUTION FOR NATURAL GAS MODEL.

PARAMETER	ESTIMATE	T-STATISTIC FOR HO: PARAMETER=0	ATTAINED 2-SIDED LEVEL OF SIGNIFICANCE	STD ERROR OF ESTIMATE
INTERCEPT	7885.86117328 *	2.97	0.0030	2653.97724959
NGASNDX	478.57038686	3.28	0.0011	145.88367773
NHSLDMEM*HGASWHT	8183.12797801	12.97	0.0001	631.08680052
HSBNMGSH	39380.14733876	5.26	0.0001	7488.38074856
KMHTEQRC				
RADIATORS OR HOT WATER PIPES (1)	35976.62454035 *	6.10	0.0001	5899.72507242
CENTRAL FORCED AIR (2)	21177.30479874 *	5.03	0.0001	4206.54981219
OTHER (3)	7946.83589671 *	2.14	0.0323	3709.19030002
NO GAS MAIN HEATING (4)	0.00000000 *	**	**	**
NHEATDD*HGASMHT*NBATHRMS	4.35890180	7.91	0.0001	0.55073749
NHEATDD*HGASMHT*NAGE01	0.05823754	5.01	0.0001	0.01163191
NHEATDD*HGASMHT*KYHSBREC	-1.12679365	-8.12	0.0001	0.13882925
NHEATDD*HGASMHT*NROOMS	1.14812215	6.45	0.0001	0.17801440
NHEATDD*HGASMHT*NDRSAWS	0.36930515	7.11	0.0001	0.05192832
NHEATDD*HGASMHT*NSDRSAWS	-0.19222837	-4.62	0.0001	0.04156752
NCOOLDD*NRMGASAC	3.97906768	4.55	0.0001	0.87531020

* THE SOLUTION IS NOT UNIQUE BECAUSE THE DESIGN MATRIX IS SINGULAR.
** NO TEST IS GIVEN BECAUSE THE ESTIMATE WAS ARBITRARILY SET EQUAL TO ZERO.

Table 3. Households that were Determined to be Outliers for Natural Gas Consumption Models

NSQIDDOE	NCNGYRB	Residual	Remarks
1,065	541,313	226,768	
1,206	460,915	235,734	
1,208	129,336	104,742	
1,631	68,158	-143,958	
1,885	64,888	-110,471	
1,908	238,242	204,255	
2,275	502,923	379,765	
2,299	250,941	146,080	
2,302	191,954	124,853	
2,315	140,315	106,119	Listed electricity as main heating fuel.
2,357	253,221	144,288	
2,377	121,396	101,515	Listed electricity as main heating fuel.
2,416	348,149	190,599	
2,594	24,830	-115,094	
2,611	228,293	137,951	
2,721	178,054	149,304	Listed electricity as main heating fuel.
2,827	371,738	217,693	
2,895	6,890	-63,698	
2,933	104,077	81,423	Listed electricity as main heating fuel
3,262	305,750	184,224	
3,405	481,234	307,609	
3,414	395,757	197,073	
3,681	279,671	164,465	
3,972	48,398	-118,288	
4,187	404,907	226,933	Heats a swimming pool with natural gas.
4,473	130,936	96,173	Listed electricity as main heating fuel.
4,692	109,437	-164,902	
4,693	105,766	-140,375	Single retired person in a large house.

TABLE 4. ANALYSIS OF VARIANCE TABLE FOR ELECTRICITY MODEL (USING NCELYRB AS THE DEPENDENT VARIABLE).

SOURCE	DEGREES OF FREEDOM	WEIGHTED SUM OF SQUARES	MEAN SQUARE	F VALUE	ATTAINED LEVEL OF SIGNIFICANCE	WEIGHTED R-SQUARE
MODEL	22	25057741.26255367	1138988.23920699	318.01	0.0001	.771422
ERROR	2073	7424794.29083030	3581.66632457			
CORRECTED TOTAL	2095	32482535.55338398				

TABLE 5. WEIGHTED LEAST SQUARE SOLUTION FOR ELECTRICITY MODEL.

PARAMETER	ESTIMATE	T-STATISTIC FOR HO: PARAMETER=0	ATTAINED 2-SIDED LEVEL OF SIGNIFICANCE	STD ERROR OF ESTIMATE
INTERCEPT	1838.90033348	1.32	0.1874	1394.29804463
NELCKDV	1542.91063357	6.06	0.0001	254.42416974
KINCOME	249.49782225	2.47	0.0135	100.89977992
NAGE01	-46.19145874	-3.01	0.0026	15.32376895
NHSLDMEM	1800.26754569	10.21	0.0001	176.30579562
NHSLDMEM*HELWHT	3567.83543487	16.11	0.0001	221.41854047
HELWHT*HELDISHW	5010.04258170	4.08	0.0001	1227.15715352
NELFRIG	3176.97311550	5.45	0.0001	582.64498142
HELDISHW	3246.81149591	5.14	0.0001	631.57851582
HELCLSDY	3827.05159181	7.31	0.0001	523.81387074
HSPFDFRZ	4456.46196110	9.69	0.0001	460.08932694
HAUTOWSH	2441.14974480	4.22	0.0001	577.80748929
NTOTAL	108.33553652	3.57	0.0004	30.34494681
HSBNMELH	3282.23184877	4.23	0.0001	776.69185130
NHEATDD*HELMHT*HELHTPUM	-2.56940122	-2.72	0.0066	0.94493510
NHEATDD*HELMHT*NAGE01	0.04376684	4.27	0.0001	0.01025377
NHEATDD*HELMHT*NBATHRMS	2.16693210	5.19	0.0001	0.41768602
NHEATDD*HELMHT*NSDRSAWS	-0.11337950	-3.04	0.0024	0.03727163
NHEATDD*HELMHT*NDRSAWS	0.31150668	6.37	0.0001	0.04890592
NCOOLDD*NRMELRAC	0.69793901	4.04	0.0001	0.17282322
NCOOLDD*NRMELRAC*KINCOME	0.07027372	2.63	0.0087	0.02675765
NCOOLDD*NRMELCAC*KINCOME	0.12566062	5.81	0.0001	0.02162087
NCOOLDD*NRMELCAC	0.41258501	2.36	0.0184	0.17482890

Table 6. Households that were Determined to be Outliers for Electricity Consumption Model.

NSQIDDOE	NCELYRB	Residual	Remarks
1,063	82,707	47,082	
1,065	98,518	51,170	Natural gas consumption also high.
1,268	2,835	-31,040	
1,435	115,414	69,746	
1,559	85,358	-60,716	
1,583	7,510	-56,678	
1,700	10,680	-26,382	
1,770	138,428	104,481	
1,774	19,486	-51,069	
1,796	13,713	-34,123	
1,810	103,141	61,656	
1,913	3,583	-20,202	
1,954	1,692	-15,978	
2,273	90,418	47,084	
2,377	50,569	-49,744	Probably heats with natural gas.
2,417	28,596	-47,940	Probably heats with natural gas.
2,482	146,876	76,718	
2,493	109,525	59,047	
2,513	84,659	45,949	
2,520	82,291	58,491	
2,522	50,569	-89,558	
2,560	37,157	-42,787	
2,685	126,688	90,256	
2,695	74,419	53,162	
2,697	119,672	74,126	
2,721	34,042	-71,395	Probably heats with natural gas.
2,933	34,857	-103,679	Probably heats with natural gas.

Table 6. Households that were Determined to be Outliers for Electricity Consumption Model (Continued).

NSQIDDOE	NCELYRB	Residual	Remarks
3,029	22,304	-42,735	
3,040	72,113	-64,952	
3,042	56,687	-64,083	
3,209	34,680	-43,597	
3,390	826	-105,043	
3,404	103,640	49,911	
3,585	4,593	-22,713	
3,717	35,294	-50,226	
3,728	17,807	-48,319	
3,731	75,770	43,985	
3,758	7,435	-34,627	
3,780	28,190	-50,786	
3,794	90,500	61,269	
3,800	52,224	33,897	
3,809	12,362	-31,461	
3,887	48,805	35,279	
3,899	109,935	76,779	
3,968	67,998	42,327	
4,064	119,440	58,603	
4,134	86,549	44,729	
4,159	10,492	-26,514	
4,603	39,961	-55,691	
4,713	12,512	-29,070	
4,823	246,909	127,238	

Section 5. Discussion

This section contains a discussion on the results of the analysis summarized in Tables 1, 2, 4, and 5. Highlights include interesting features of the models, what the models suggest about the relative contribution of different factors to household energy consumption, and some of the pitfalls encountered in interpreting individual coefficients. The discussion is organized according to the variable groupings used in Section 4.

The t statistics listed in Tables 2 and 5 indicate the significance of the marginal contributions of the associated term in the model. Thus, the t statistic associated with the term NGASNDX in Table 2 indicates whether the addition of this term is significant given that all of the other terms in the model are already present.

The order of appearance of the terms in a model is not significant. That is, there is no relationship between a term's position in its list and its relative contribution to the multiple R-square for the model. The current models were arrived at by a trial-and-error approach with attention to what terms proved significant when other terms were already in the model.

We caution the reader again that the estimates presented here are preliminary and derived from only part of the NIECS data. In addition, there are the usual problems of interpreting the coefficients of a regression model, and they are discussed elsewhere in this report. For now, the results are presented as a description of the variation in energy consumption among households.

Tables 7 and 8 summarize, respectively, the modeled natural gas, and the modeled electricity consumption for a hypothetical household's water-heating, space-heating, and air conditioning.

The modeled results for the two hypothetical households should be taken as a package and not as a sum of well-defined individual contributions. Some of the independent variables are highly correlated. Hence, the effect of a variable, including those not used in the model, could easily be represented in the model by the contribution of another variable.

Table 7. Modeled Natural Gas Consumption for Hypothetical Household

Independent Variable	Hypothetical Value	Contribution to Model
Intercept	-	7,886
NGASNDX	17	8,136
NHSLDMEM x HGASWHT	4 x 1	32,733
HSBNMGSH	0	0
KMHTEQRC	2	21,177
NHEATDD x HGASMHT x NBATHRMS	5,700 x 1 x 2	49,691
NHEATDD x HGASMHT x NAGE01	5,700 x 1 x 40	13,278
NHEATDD x HGASMHT x KYHSBREC	5,700 x 1 x 5	-32,114
NHEATDD x HGASMHT x NROOMS	5,700 x 1 x 7	45,810
NHEATDD x HGASMHT x NDRSAWS	5,700 x 1 x 18	37,891
NHEATDD x HGASMHT x NSDRSAWS	5,700 x 1 x 12	-13,148
NCOOLDD x NRMGASAC	1,200 x 7	33,424
Total		204,764

Table 8. Modeled Electricity Consumption for Hypothetical Household

Independent Variable	Hypothetical Value	Contribution to Model
Intercept	-	1,839
NELCKDV	2	3,086
KINCOME	8	1,996
NAGE01	40	- 1,848
NHSLDMEM	4	7,201
NHSLDMEM x HELWHT	4 x 1	14,271
HELDISHW x HELWHT	1 x 1	5,010
NELFRIG	2	6,354
HELDISHW	1	3,247
HELCLSDY	1	3,827
HSPFDFRZ	1	4,456
HAUTOWASH	1	2,441
NTOTAL	25	2,708
HSBNMELH	0	0
NHEATDD x HELMHT x HELHTPUM	5,700 x 1 x 1	-14,646
NHEATDD x HELMHT x NAGE01	5,700 x 1 x 40	9,979
NHEATDD x HELMHT x NBATHRMS	5,700 x 1 x 2	24,703
NHEATDD x HELMHT x NSDRSAWS	5,700 x 1 x 12	- 7,755
NHEATDD x HELMHT x NDRSAWS	5,700 x 1 x 18	31,961
NCOOLDD x NRMELRAC	1,200 x 0	0
NCOOLDD x NRMELRAC x KINCOME	1,200 x 0 x 8	0
NCOOLDD x NRMRLCAC x KINCOME	1,200 x 7 x 8	8,444
NCOOLDD x NRMELCAC	1,200 x 7	3,466
Total		110,740

Strategy vs. Tactics

As outlined in Section 4, our strategy was to develop an overall intercept equivalent to a linear combination of household and appliance data and a coefficient of degree-days equivalent to a linear combination of housing unit characteristics. In practice, our strategy was modified to accommodate the data, dropping housing unit characteristics into the intercept, for example. One effect of not strictly following the strategy outlined above is that the electricity and natural gas models are not simply comparable. There is no term for term correspondence between the models. This is another factor which contributes to the difficulty of comparing the effects of different factors on the use of the two energy forms.

Demographic Characteristics

The variables NAGE01 and NHSLDMEM appear in both models, but in different forms, and the variable KINCOME appears only in the electricity model.

The age variable[1] enters with a negative coefficient in the electricity model and with a positive coefficient interacted with degree-days in the natural gas model. The positive coefficient in the gas model suggests increased demand for heating with increasing age of the respondent. On the other hand, the negative coefficient in the other model suggests a decreased demand for electricity, other than that associated with the heating (degree-days) components. The two indications are not inconsistent: households comprising older people may operate at lower levels of activity, thereby requiring less electricity on the average, while demanding higher interior temperatures than the average.

Both models indicate that the larger the household (NHSLDMEM), the greater the energy consumption. In addition, both suggest that the number of household members has a significant impact on the amount of fuel used for hot water-heating.

The income variable, KINCOME, enters with a positive coefficient in the electricity model. In the current gas model, the addition of this variable is not significant; it may be significant in other versions of the model.

[1]Note that NAGE01 is the age of the respondent, not necessarily that of the head of household.

Care needs to be taken in interpreting the income variable in any formulation of the model. The income data available in the NIECS are inherently only ordinal data. Respondents were asked what income range they were in, not their actual income. Thus, one cannot say from analyses such as these that if income changes by a stated amount, then electricity changes by a proportional amount. Rather, one can only say that if income increases, an increase in electricity consumption is indicated. This limitation is inherent in the data.

Measures of Size and Insulation and Heating Components

In both models, the measures of size and insulation data are concentrated in the heating component of the models, the interaction terms which include degree-days. The only exception to this is NTOTAL, which counts the total number of doors, windows, and rooms, and which stands alone in the electricity model.

All of the size of house coefficients are positive, suggesting that energy consumption increases with the size of the dwelling unit, as was expected. The fact that these terms appear as interactions with heating degree-days indicates this importance in determining the heating component of energy demand, and is consistent with the experiments and modeling strategy described in Section 4.

The importance of the number of bathrooms as a measure of size is interesting. In both models, it is the variable which, when interacted with degree-days, has the largest positive coefficient, in some cases, by orders of magnitude. Its importance likely derives from being a proxy for several other measures of the size of the housing unit or status of the household: square footage, income of the household, quality of housing and living habits of the household members.

The uncertainty of just what NBATHRMS may be a proxy for suggests the care that must be taken in interpreting this coefficient. For example, the conversion of storage space to a bathroom will not automatically result in adding 4,360 Btu per degree-day to the gas consumption of gas-heated homes. Nor should it even result in increased gas consumption, unless, perhaps, the addition signals other changes that result in higher energy consumption.

Note the differences between the values of NBATHRMS across the two models. The ratio of the coefficient in the gas model to that in the electricity model is on the order of 2 to 1. The analogous ratios for the other coefficients in the heating component are also large in the gas model, the ratios being on the order 1.5 to 1. These ratios recall the discussions of Figures 1 and 6, and 2 and 7 in Section 3.

The differences between NDRSAWS and NSDRSAWS within models suggests a relationship between energy consumption and size of dwelling, and the energy conserving effects of the presence of storm doors and windows.

Once again, the interpretation must be made with caution. The variable NSDRSAWS, the total number of storm doors and windows, likely reflects the fact that houses with storm windows tend to be better-insulated in other ways than those that do not have storm windows; installing storm windows on all windows in a house will not necessarily reduce energy consumption due to heating by the amount indicated in the models.

Households which heat with electricity and operate an electric heat pump may enjoy a very large reduction in their space-heating demand, at least as indicated by the coefficient associated with HELHTPUM interacted with NHEATDD. Out of the 2,095 households used in the electricity model, only approximately 40 have heat pumps, yet this factor was significant.

Appliances

Important differences in the consumption of electricity and gas are indicated by the composition of the intercepts. In the electricity models, individual appliances are accounted for, whereas in the gas model, the use of appliances is summarized in NGASNDX. On the other hand, important differences due to differences in heating equipment are indicated in the gas model.

Care should be taken in interpreting the coefficient of individual appliances in models like the electricity model. For example, the 4,456 MBtu that the model adds for having a separate food freezer may not all be consumed by the freezer. Households that have freezers may cook more or tend to have more electrical appliances generally than households that do have freezers. Therefore, multiplying the coefficient for separate food freezers by the number of freezers in residential use may give a biased estimate of the energy consumed by freezers.

The Heating Component and Heating Equipment

The natural gas model suggests there are large differences in consumption unrelated to degree-days but associated with differences in heating equipment. The large values of these baseload or intercept terms may be a result of the curve in the overall natural gas consumption with respect to degree-days.

It should be noted that because of the way the model was fit only certain linear combinations of the intercept term and the four coefficients for KMHTEQRC can be estimated. In particular, the intercept plus any of the four coefficients can be estimated, and the differences between pairs of equipment coefficients can be estimated.

Thus, we estimate that households which use natural gas, but not as their main heat source, (Category 4 of KMHTEQRC) have an annual baseload of INTERCEPT + 0 = 7,886 MBtu. Continuing by estimating differences, we see that Category 3 of KMHTEQRC (which includes all types of gas heating systems except forced-air, Category 2, and hot water-radiator systems, Category 1) increases a household's annual baseload by approximately

7,950 MBtu annually compared to a household which uses gas but not as its main heating fuel (Category 4). Similarly, forced-air systems have a baseload consumption which is 21,177 - 7,947 = 13,230 MBtu greater than other systems, and hot water systems have a baseload 35,977 - 21,177 = 14,800 MBtu greater than forced-air systems.

The model suggests that baseload consumption can be very substantial in relation to the heating component. For example, suppose a household has 2 bathrooms, the age of the head of household is 40, the house was built during 1960 (KYHSBREC=6), and the house has 6 rooms, and a dozen doors and windows, 6 of which are insulated. Then

(4.36 NBATHRMS + 0.06 NAGE01 1.13 KYHSBREC + 1.15 NROOMS + 0.37 NDRSAWS - 0.19 NSDRSAWS) x NHEATDD

=(4.362 + 0.0640 1.13 + 1.156 + 0.3712 - 0.196) x NHEATDD

=(14.5 MBtu/degree-day) x NHEATDD

=14.2 cu. ft./degree-day x NHEATDD

If the house is in a 5,000 heating degree-day zone, then the heating component is estimated to be approximately 71,100 cubic feet or 72.5 x 10^6 Btu annually. For a household with a hot-water system, the baseload is (7.886 + 35,976) MBtu = 43.9 x 10^6 Btu which is 61 percent of the heating component, (43.9/72.5) = 61 percent or 38 percent of the total heating load which is (43.9 + 72.5) = 116.4 x 10^6 Btu.

Originally, we interacted the equipment variable, KMHTEQRC, with degree-days on the grounds that the overhead associated with various systems should vary with the severity of winters. For example, households located in a 4,000 to 5,000 degree-day region were expected to have lower system baseloads than households in an 8,000 to 9,000 degree-day regions because winters degree-day are less severe. However, the interaction term did not show any significant differences between the various types of heating equipment.

The Cooling Component of Electricity

In discussing Figure 5, we noted an apparent lack of association between electricity consumption and cooling degree-days among households that do not use electricity for heating. However, in the electricity model interaction, terms which include the number of rooms air conditioned, income, and cooling degree-days, appear. This suggests that a component of variation in electricity consumption due to cooling can be observed when other characteristics are accounted for, such as a household's appliance inventory.

Residuals

Figures 10 and 11 show the residuals from the gas and electricity models, respectively, plotted against degree-days. Figure 12 presents another look at the data in Figure 10. Figure 13 is a plot of the residuals from the electricity model against the residuals from the gas model for those households which use both fuels.

Figures 10 and 12 repeat the patterns displayed in Figures 1 and 6: the residuals from the gas model rise gently from just below zero to slightly above it at about 6,100 degree-days, at which point they once again fall below zero. This is most easily seen in Figure 12.

It is apparent from these figures that the gas model does not account for the hypothesized change in trend observed in Figure 6 and discussed in Section 3. Since the present model includes proxy measures of size and insulation, the indication is that these variables are not sufficiently powerful to account for the differences between households in the coldest regions and the rest of the sample, assuming that the differences are due to differences in insulation. Therefore, real, unmeasured differences in insulation cannot be ruled out as the explanation. Neither can the other hypotheses (acclimation, use of secondary heating sources, conservation behaviors) advanced in Section 3 be ruled out.

Figure 11 shows the residuals from the electricity model versus degree-days. The plot indicates a reasonably good fit to electricity consumption across the range of degree-days. The residuals are centered at zero over the whole range, and the variability in the residuals is uniform.

Figure 13, the electricity residuals versus gas residuals, is a commentary on the "wasters are wasters" while "savers are savers" hypothesis. Under this hypothesis, households which are below-average in the consumption of one fuel should be below-average in the consumption of the other and those who are above-average consumers of one should be above-average consumers of the other. Under this hypothesis, we would expect the residuals in Figure 13 to exhibit a positive correlation. In fact, the diagram and correlation coefficient of .04 reveal a negligible association.

Concluding Remarks

In this paper, we have presented some preliminary investigations of the variability in energy use among single-family households. Several important indications appeared which we believe should be accounted for in further analysis:

- The differences between electricity and gas consumption; Because of the large differences between the heating components for these fuels, one should be wary of combining the data into a single, "total fuel consumption" model.

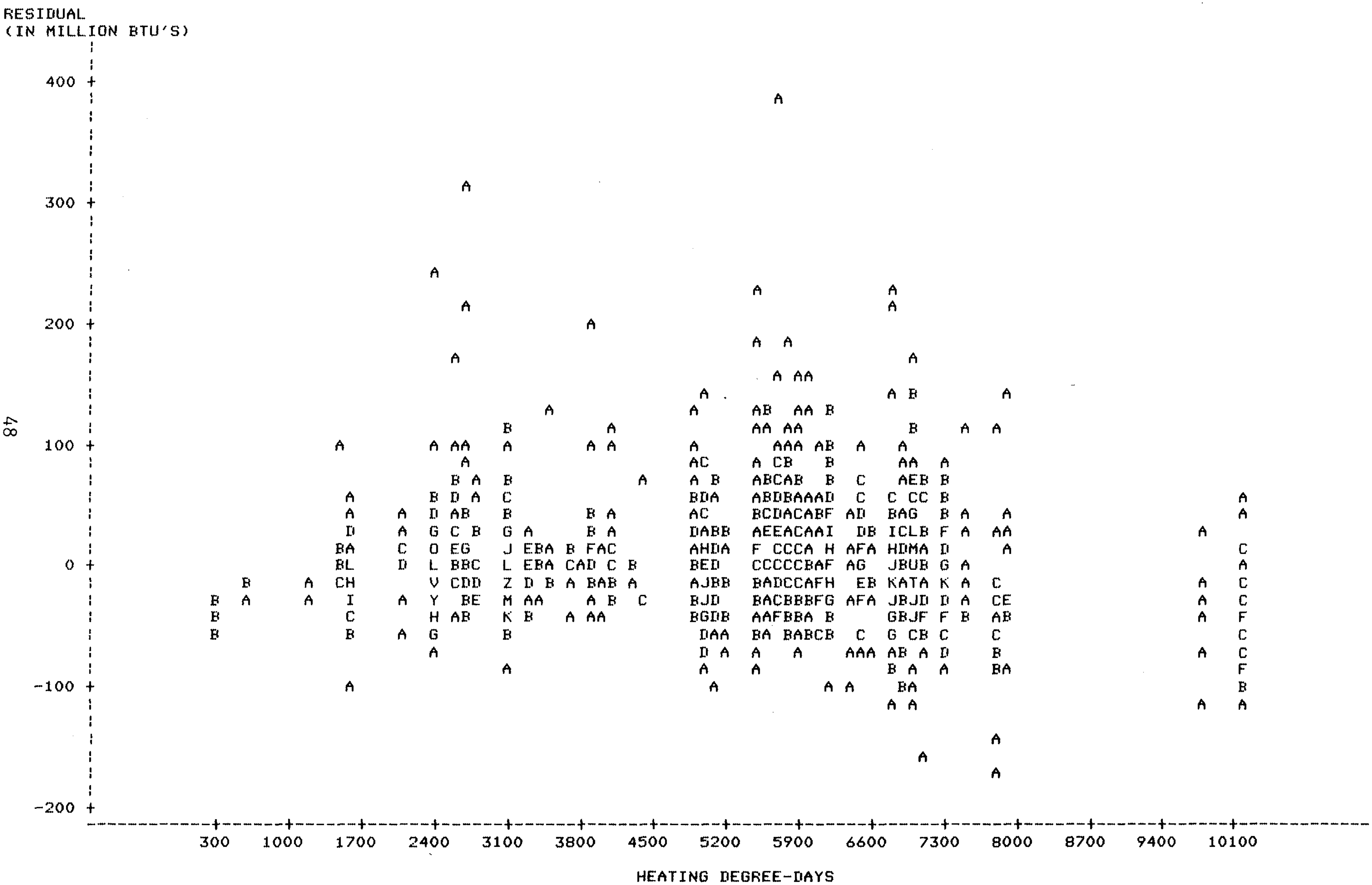
- The deviation of houses in the coldest regions from the regression average. More work needs to be done to uncover the systematic differences, if any, between the households in the coldest region from the rest of the population.

- The large differences in baseload requirements among different types of heating systems.

- The importance of the number of bathrooms as a predictor of energy consumption.

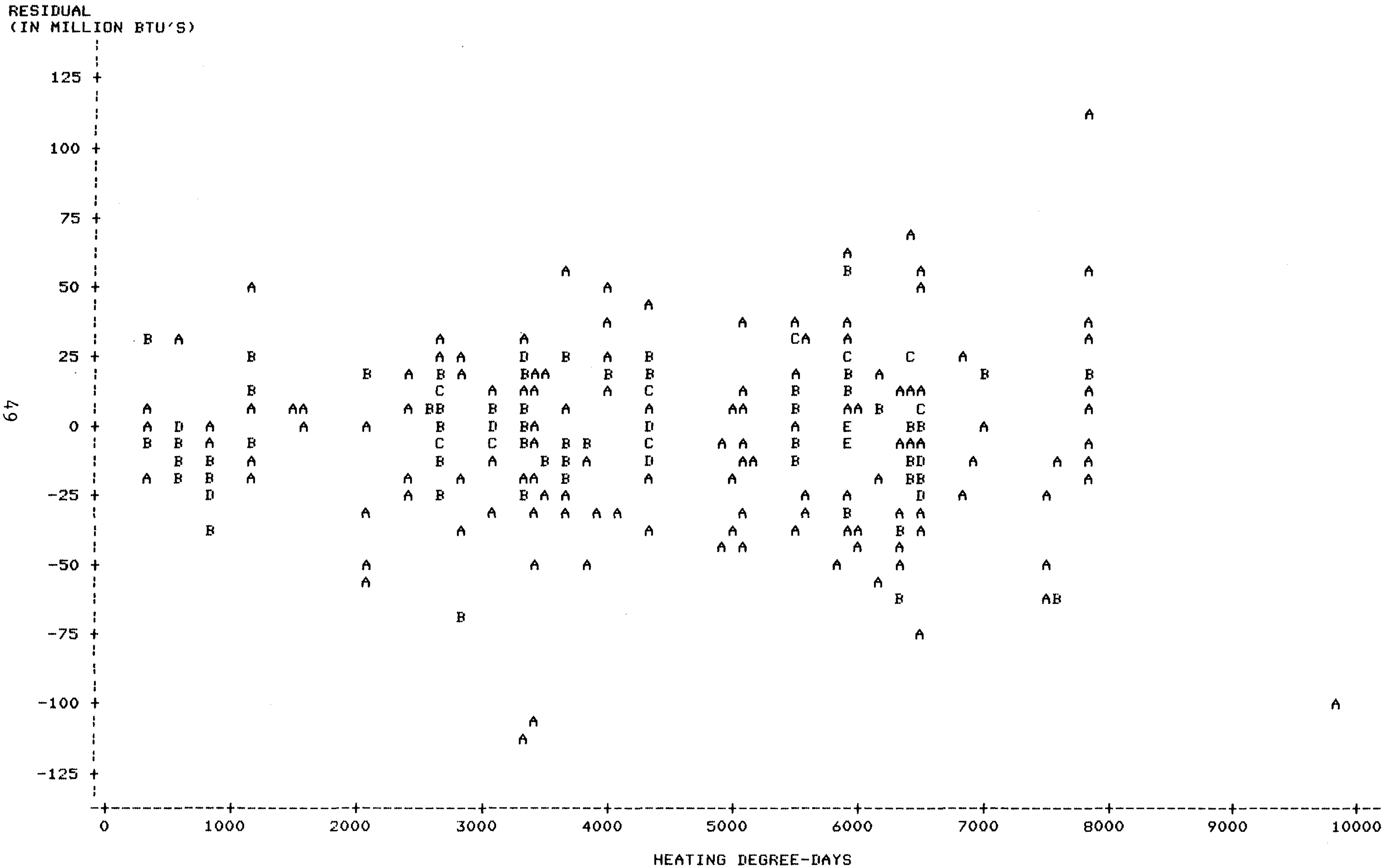
In subsequent analyses, we will extend the models to the full data set, estimate the sampling errors of regression coefficients, and attempt to estimate average heating and baseload components by characteristics of households.

FIGURE 10. RESIDUAL FOR NATURAL GAS MODEL BY HEATING DEGREE-DAYS FOR HOUSEHOLDS THAT USE NATURAL GAS FOR THE MAIN SPACE-HEATING FUEL.



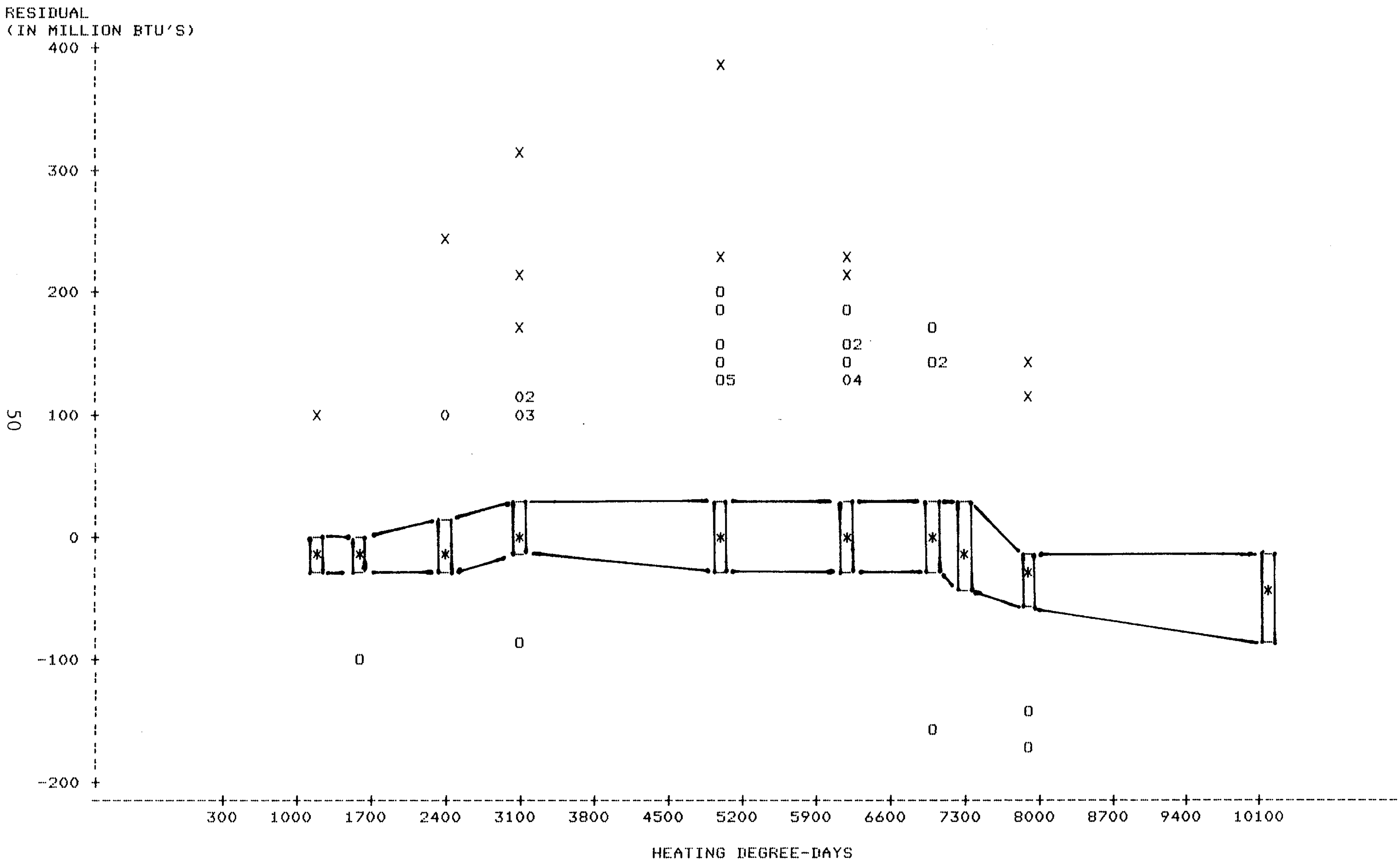
LEGEND: A = 1 OBSERVATION, B = 2 OBSERVATION, ETC.

FIGURE 11. RESIDUAL FOR ELECTRICITY MODEL BY HEATING DEGREE-DAYS FOR HOUSEHOLDS THAT USE ELECTRICITY FOR THE MAIN SPACE-HEATING FUEL.



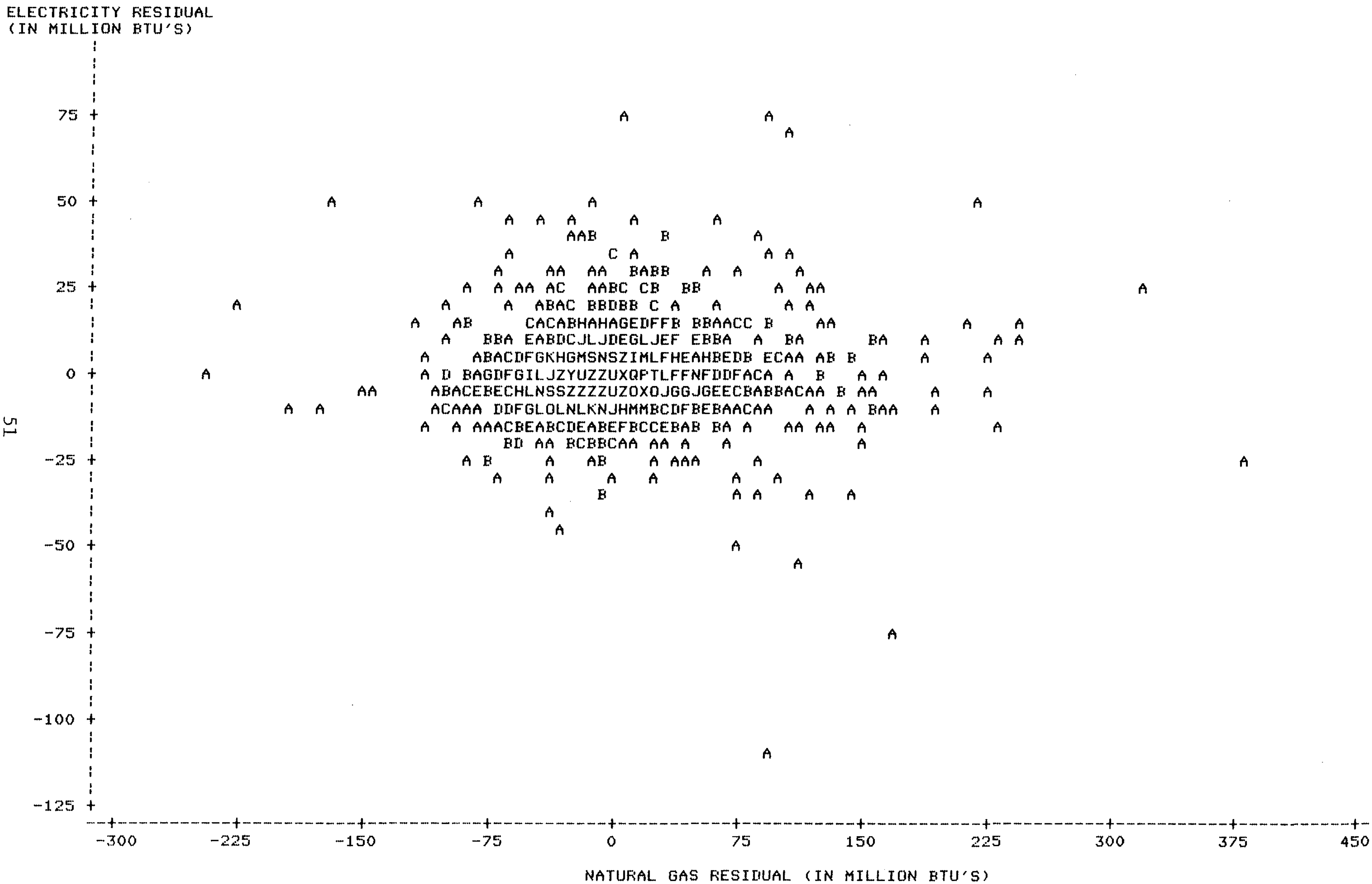
LEGEND: A = 1 OBSERVATION, B = 2 OBSERVATIONS, ETC.

FIGURE 12. WANDERING BOX PLOT OF RESIDUAL FOR NATURAL GAS MODEL BY HEATING DEGREE-DAYS FOR HOUSEHOLDS THAT USE NATURAL GAS FOR THE MAIN SPACE-HEATING FUEL.



LEGEND: X = EXTREME OUTLIER
0 = BORDERLINE OUTLIER
* = MEDIAN.

FIGURE 13. RESIDUAL FOR ELECTRICITY MODEL BY RESIDUAL FOR NATURAL GAS MODEL.



LEGEND: A = 1 OBSERVATION, B = 2 OBSERVATIONS, ETC.

APPENDIXES

Appendix A

SAMPLING PLAN FOR THE NIECS

Households used in the NIECS were selected from the universe of households according to a plan whereby each household had roughly the same probability of being selected.

Definitions of the universe of households used for the survey were generally the same as those used for U.S. Census Bureau surveys except that households in Alaska and Hawaii and those located on military installations were not included.

A number of steps were carried out in the selection. Briefly they were:

- The approximately 3,000 counties and independent cities in the United States (exclusive of Alaska and Hawaii) were grouped into 1,140 primary sampling units (PSU's).

- The 1,140 PSU's were next divided into 103 groups or strata; each stratum consisted of PSU's as much alike as possible in terms of geographic region, community type, and socio-economic characteristics.

- One PSU was selected for each of the 103 strata; these selected PSU's were the primary areas for the survey.

- Within selected PSU's, a number of subsampling steps were used to select specific clusters of housing units for the survey. These subsampling steps made use of 1970 Census data for small units such as block groups and enumeration districts, supplementary data to identify areas with substantial new residential construction since 1970, and field visits to make rough counts of housing units as well as detailed housing unit listings.

- These subsampling steps resulted in the selection of 456 ultimate sampling units (USU's). Each USU is a cluster of households, averaging approximately 10 per cluster.

After the households were selected and interviewed, they were assigned weights that reflected the nonresponse rate in their cluster and the approximate probability that they were selected. The weights were not used in the analysis presented in this publication.

Appendix B

ENERGY CONSUMPTION DATA, MISSING DATA, AND DEGREE-DAY DATA

A preliminary public use version of the NIECS data is available in machine-readable form from the National Technical Information Service (Reference 8).

There are two important features of the public use file. One is the manner in which the energy consumption data is reported, and the other is that degree-day information is included.

Table B1 summarizes the type of energy consumption data obtained from the utilities. The table also summarizes the situation by varying degrees of missing data. The category "11 or more months" corresponds to the combination of "complete" and "nearly complete" mentioned in Reference 5. The category "less than 5 months" corresponds to "missing" in the reference.

The reason why data are missing vary with the extent to which they are missing. In the case of completely missing data, the reasons range from outright noncooperation from households or utilities, to failure to find a household in a utility's records. For the other missing data situations, the most frequent cause is changing ownership of households: if ownership or occupancy of a housing unit changed hands during the period March 1978 to April 1979, we were able to obtain data for the time period when the house was occupied by the person signing the authorization.

Another reason for missing data is inefficient record-keeping by utilities. Some utilities do not keep 12 months of billing data on-line, or readily accessible, at any given time.

Table B1. Energy Consumption Records and Missing Data for Survey Households Using Electricity and Utility Gas

	Electricity: Number of households	Percent	Utility Gas: Number of households	Percent
Total Households Using Fuel	4080	100.0	2577	100.0
Data Received From Fuel Supplier	3509	86.0	1947	75.5
11 Months or More	3023	74.1	1754	68.0
5-10 Months	340	8.3	124	4.8
Less Than 5 Months	146	3.6	69	2.7
Household Pays Directly To Supplier -- No Data Available	334	8.2	270	10.5
Household Not Identified In Company Records	128	3.1	110	4.3
Company Refused to Participate	0	-	5	0.2
Company Unknown Or Not Located	0	-	0	-
Authorization Form Not Signed	206	5.1	155	6.0
Fuel Used Included In Rent Or Paid In Other Way	237	5.8	360	14.0

Source: NIECS: Report On Methodology, Part 1. Household and Utility Company Surveys, Response Analysis Corporation, Princeton, New Jersey; February 1981, Section 5.

The heating degree-day data and cooling degree-day data are annual figures, adjusted for the 1978 to 1979 weather, and rounded to the nearest 100 degree-days. The annual degree-days, prior to adjustment, were the long term averages published by the National Oceanic and Atmospheric Administration (NOAA). Adjustments were made to these long term averages on a regional basis (9 Census Regions). For heating degree-days, the adjustment was based on the 1978 to 1979 heating season. For cooling degree-days, the adjustment was based on the 1978 cooling season. The procedure used to obtain the adjustment factors is as follows:

The country was partitioned into 344 divisions as defined by NOAA, where each division is a geographic area within which climatic conditions are relatively homogeneous. The divisions generally follow county boundaries, the principal exceptions occurring in certain coastal and mountain areas. The division within which a household resides was then determined, and average annual heating and cooling degree-day figures were computed by averaging over all NOAA weather stations within the division and over the years from 1930 to 1975. The mathematical form for a household's unadjusted average then is:

$$HDD_h = \sum_k HDD_{ik}/46$$

where:

HDD_h is the 46-year average heating degree-day figure for $household_h$

i denotes the NOAA division in which the household resides

k denotes the years from 1930 to 1975, inclusive, and

HDD_{ik} is the number of heating degree-days reported in the i^{th} division for the k^{th} year.

The cooling degree-day average, CDD_h is defined analogously. Note that if two households, h and 1, are in the same NOAA division, then $HDD_h = HDD_1$; the same would be true for cooling degree-days. Thus, the variability in the degree-day within a NOAA division is zero.

The average values HDD_h were then adjusted on a regional basis for departures from the 46-year norm using the ratios in Table B1. The ratios were derived as follows:

$$Q_r = HDD_{sr}/HDD_{ar}$$

where r is the r^{th} census region

HDD_{sr} is the average heating degree-day figure for the survey year, April 1978 to March 1979, averaged over all NOAA Divisions in the r^{th}.

and HDD_{ar} is the 40-year average ove¨ all NOAA divisions in the r^{th} region.

Table B2. Ratio Adjustments

Census Region	78-79 HDD Adjustment Factor	1978 CDD Adjustment Factor
New England	1.020	.893
Middle Atlantic	1.044	.896
East North Central	1.112	.945
West North Central	1.151	1.026
South Atlantic	1.044	.992
East South Central	1.103	1.000
West South Central	1.174	1.036
Mountain	1.103	.984
Pacific	1.049	1.195

Appendix C
NIECS PUBLIC USE FILE VARIABLES USED IN FINAL CONSUMPTION MODEL

Table C1 defines the household variables selected for the final analysis of natural gas consumption and the final analysis of electricity consumption. The variable names and values are documented in the NIECS public use file. Table C2 defines the variables that are transformations of variables listed in Table C1.

Table C1. Household Variables Selected for Analysis of Natural Gas Consumption and Electricity Consumption

Variable Name	Variable Description
NCNGYRB	Annual Consumption of Natural Gas in Thousands of Btu's
NCELYRB	Annual Consumption of Electricity in Thousands of Btu's
NHEATDD	Number of Heating Degree-Days Based on 40-year Average
NCOOLDD	Number of Cooling Degree-Days based on 40-year Average
NAGE01	Age of Respondent
KINCOME	Code for Household Income Level
NHSLDMEM	Number of Household Members
KYHSBREC	Code for Year House Built
NROOMS	Number of Rooms in House
NCOMBATH	Number of Complete Bathrooms
NHAFBATH	Number of Half Bathrooms
NGASNDX	Gas Appliance Index
HELDISHW	Indicator Variable for Electric Dish Washer
HAUTOWSH	Indicator Variable for Automatic Washing Machine
HELCLSDY	Indicator Variable for Electric Clothes Dryer
HELOVEN	Indicator Variable for Electric Oven
HELRANGE	Indicator Variable for Electric Range
HSPFDFRZ	Indicator Variable for Separate Food Freezer
NREFRIG	Number of Refrigerators
KREFRIG1	Code for First Refrigerator Fuel
KREFRIG2	Code for Second Refrigerator Fuel
NDOORS1	Number of Outside Doors and Sliding Glass Doors
NSDOORS	Number of Storm Doors
NSWINSGD	Number of Storm Sliding Glass Doors
NUMWINDS	Number of Windows
NSTRMWIN	Number of Storm Windows
KMHEATEQ	Code for Main Heating Equipment
KFLMHEAT	Code for Main Heating System Fuel
KFLSHEAT	Code for Secondary Heating System Fuel
HELHTPUM	Indicator Variable for Electric Heat Pump
KWHEATFL	Code for Water-Heating Fuel
HROOMAC	Indicator Variable for Electric Room Air Conditioners
HCENTAC	Indicator Variable for Central Air Conditioning
KFLCNAC	Code for Central Air Conditioning Fuel
NROOMAC	Number of Rooms Air Conditioned

Table C2. Transformed Household Variables Selected for Analysis of Natural Gas Consumption and Electricity Consumption.

Variable	Definition		Description
NSDRSAWS	NSDOORS + NSWINSGD + NSTRMWIN		Number of storm doors and windows.
NDRSAWS	NDOORS1 + NUMWINDS		Number of doors and windows.
NTOTAL	NROOMS + NDOORS1 + NUMWINDS		Overall measure of the size of the house.
NBATHRMS	NCOMBATH + 1/2 (NHAFBATH)		Number of bathrooms.
NRMELCAC	0	if HCENTAC = 0	Number of rooms potentially air conditioned by electric central units.
	0	if HCENTAC = 1 and KFLCNAC = 1	
	NROOMAC-HROOMAC	if HCENTAC = 1 and KFLCNAC = 2	
NRMELRAC	0	if HROOMAC = 0	Number of rooms potentially air conditioned by electric room units.
	1	if HROOMAC = 1 and HCENTAC = 1	
	NROOMAC	if HROOMAC = 1 and HCENTAC = 0	
NRMGASAC	0	if HCENTAC = 0	Number of rooms potentially air conditioned by natural gas central units.
	0	if HCENTAC = 1 and KFLCNAC = 2	
	NROOMAC-HROOMAC	if HCENTAC = 1 and KFLCNAC = 1	

Table C2. Transformed Household Variables Selected for Analysis of Natural Gas Consumption and Electricity Consumption (Continued).

Variable	Definition	Description
HELWHT	1 if KWHEATFL = 5 0 otherwise	Indicator variable for electric water heater.
HGASWHT	1 if KWHEATFL = 1 0 otherwise	Indicator variable for natural gas water heat.
HELMHT	1 if KFLMHEAT = 5 0 otherwise	Indicator variable for electric main heating.
HGASMHT	1 if KFLMHEAT = 1 0 otherwise	Indicator variable for natural gas main heating.
HSBNMELH	1 if KFLSHEAT = 5 and KFLMHEAT ≠ 5 0 otherwise	Indicator variable for electric secondary but not main space-heating.
HSBNMGSH	1 if KFLSHEAT = 1 and KFLMHEAT ≠ 1 0	Indicator variable for natural gas secondary but not main space-heating.
HELHTPUM	1 if KMHEATEQ = 4 0 otherwise	Indicator variable for electric heat pump.
KMHTEQRC	1 if KFLMHEAT = 1 and KMHEATEQ = 1 or 2 2 if KFLMHEAT = 1 and KMHEATEQ = 3 3 if KFLMHEAT = 1 and KMHEATEQ ≠ 1,2, and 3 4 if KFLMHEAT ≠ 1	Class variable for gas main heating equipment. 1 - System where heat is distributed by water 2 - Central forced-air 3 - Other (usually space heater) 4 - Nongas main heating

Table C2. Transformed Household Variables Selected for Analysis of Natural Gas Consumption and Electricity Consumption (Continued).

Variable		Description
NELCKDV	HELOVEN + HELRANGE	Number of major electric cooking appliances.
HELFRIG1	1 if KREFRFL1 = 1 0 otherwise	Indicator variable for electric refrigerator.
HELFRIG2	1 if KREFRFL2 = 1 0 otherwise	Indicator variable for electric refrigerator.
HELFRIG3	1 if NREFRIG = 3, KREFRFL1 = 1 0 otherwise	Indicator variable for electric refrigerator.
NELFRIG	HELFRIG 1 + HELFRIG 2 + HELFRIG 3	Number of electric refrigerators.

Appendix D.

WANDERING BOXPLOTS

Figures 6, 7, and 12 are variations of the graphical technique known as wandering boxplots or wandering schematics, discussed by John Tukey in his book, "Exploratory Data Analysis". To construct our version,

- Compute the 3rd, 6th, 12th, 25th, and 50th percentiles of the abscissa, and their symmetric counterparts the 97th, 94th, 88th, and 75th percentiles.
- Divide the data into groups according to the values of the abscissa using the computed percentiles as dividing points.
- Within each group, calculate the quantiles and the median of the ordinate. These points are the ends and center, respectively, of the boxes.
- Calculate the midspread for each group, midspread = upper quantile - lower quantile.
- Flag all value between one and one-half and two midspreads of the upper quantile with an 'o', and all values beyond two midspreads of the upper quantile with an '*'. Perform the similar operation with respect to the lower quantile.
- Within each group, plot the box and the flagged values of the ordinate centered over the median of the abscissa for the group.

*U.S. GOVERNMENT PRINTING OFFICE : 1981 0-341-068/1152

REFERENCES

1. Residential Energy Consumption Survey: Characteristics of the Housing Stock and Households, 1978, Report No. DOE/EIA-0207/2, February 1980.

2. Residential Energy Consumption Survey: Conservation, Report No. DOE/EIA-0207/3, February 1980.

3. Residential Energy Consumption Survey: Consumption and Expenditures, April 1978 Through March 1979, Report No. DOE/EIA-0207/1, December 1979.

4. Single Family Households: Fuel Inventories and Expenditures: National Interim Energy Consumption Survey, Report No. DOE/EIA-0207/1, December 1979.

5. Woteki, T. H. "Imputing Residential Energy Consumption Data for the National Interim Energy Consumption Survey: A Progress Report", February 1980, Office of the Consumption Data System, Energy Information Administration, U.S. Department of Energy.

6. "NIECS Imputations and Adjustments for Missing Data" Memorandum from Reuben Cohen to Lynda Carlson, August 1980. Response Analysis Corporation, Princeton, New Jersey.

7. Energy and Buildings, 1, April 1978.

8. Shrader, T. F., "A Two Parameter Model for Assessing the Determinants of Residential Space-Heating", Princeton University Center for Environmental Studies, Report No. 69, June 1978, Princeton, New Jersey.

9. Mayer, L. S. and Benjamin, Y., "Modeling Residential Natural Gas Demand as a Function of the Coldness of the Month", Energy and Buildings, 1, April 1978.

10. Woteki, T. H. "The Princeton Omnibus Experiment: Some Effects of Retrofits on Space-Heating Requirements, "Princeton University Center for Environmental Studies, Report No. 43, December 1976, Princeton, New Jersey.

11. Fels, M. F. and Woteki, T. H. "Studies of Energy Consumption; An Exploratory Analysis of a Utility's Monthly Billing Data", Energy, 5, 1979.

Other Reports Produced by the Office of the Consumption Data System

Residential Energy Consumption Survey: Conservation, February 1980, DOE/EIA-0207/3, GPO Stock No. 061-003-00087-8, $6.00.

Single-Family Households: Fuel Inventories and Expenditures: National Interim Energy Consumption Survey, December 1979, DOE/EIA-0207/1, GPO Stock No. 061-003-00075-4, $1.75.

Residential Energy Consumption Survey: Characteristics of the Housing Stock and Households, 1978, February 1980, DOE/EIA-0207/2, GPO Stock No. 061-003-00093-2, $4.25.

Residential Energy Consumption Survey: Consumption and Expenditures, April 1978 through March 1979, July 1980, DOE/EIA-0207/5, GPO Stock No. 061-003-00131-9, $6.50.

Residential Energy Consumption Survey: Consumption Patterns of Household Vehicles, June to August 1979, June 1980, DOE/EIA-0207/4, GPO Stock No. 061-003-00156-4, $3.75.

Nonresidential Buildings Energy Consumption Survey: Building Characteristics, March 1981, DOE/EIA-0246, GPO Stock No. 061-003-00171-8, $5.50.

Residential Energy Consumption Survey: 1979-1980 Consumption and Expenditures, Part I: National Data (including Conservation), April 1981, DOE/EIA-0262/1, GPO Stock No. 061-003-00191-2, $5.00.

Residential Energy Consumption Survey: 1978-1980 Consumption and Expenditures, Part II: Regional Data, May 1981, DOE/EIA-0262/2, GPO Stock No. 061-003-00189-1, $5.00.

Copies of the above reports are available from:

Superintendent of Documents
U.S. Government Printing Office
Washington, D.C. 20402

National Interim Energy Consumption Survey: Household Interview File, DOE/DF-81/001 (Magnetic Tape), $125.00 available from the National Technical Information Service - Computer Products Division, 5285 Port Royal Road, Springfield, Virginia, Accession No. PB-81-108714, $125.00.

U.S. Department of Energy
Energy Information Administration
National Energy Information Center, EI-20
Forrestal Building
Washington, D.C. 20585