SRC 8222-R2-CA

First Year Load Impact Study of Southern California Gas Company's 1994 Home Energy Fitness Program

Study ID 708

Prepared for

Ms. Melissa Cuaycong
Southern California Gas Company
555 W. Fifth Street
ML-25D1
Los Angeles, CA 90013-1011

Prepared by

Andrew A. Goett
Jamie Howell
Synergic Resources Corporation
2855 Telegraph Avenue, Suite 410
Berkeley, CA 94705

February 27, 1996

TABLE OF CONTENTS

TABLE OF CONTENTS

Chapter 1
EXECUTIVE SUMMARY

Chapter 1 EXECUTIVE SUMMARY

This report presents the results of the impact evaluation of Southern California Gas Company's 1994 Home Energy Fitness (HEF) Program. The HEF Program promotes the adoption of energy efficient measures and actions by providing informational audits to residential customers. The objective of the impact evaluation was to estimate the natural gas savings that resulted from the 1994 HEF Program audits. This evaluation was performed in compliance with the requirements of the Protocols for Verification of Demand-Side Management (DSM) Programs which govern the procedures that the California investor owned utilities must use in evaluating their programs.

The impact evaluation estimated the gross and net savings attributable to the 1994 HEF Program by examining the consumption patterns for a representative sample of participants and non-participants over a three year period spanning the receipt of the audits. Several regression models were estimated on these data, and they produced estimates of gross annual savings in the range of 57 to 88 therms per year. The estimates of net savings ranged from 44 to 54 therms per year. The net savings are the reductions in gas consumption from the program, after controlling for "naturally occurring" conservation and background environmental and economic trends.

Based on the performance of the different regression models that were estimated, one was selected as the preferred specification to calculate net program savings. This model produced a gross savings estimate of 88 therms and a net savings estimate of 44 therms. Given this model, the reliability of the estimates is fairly high. With a statistical confidence of 90%, one can state that the "true" net program savings are at least 26 therms and less than 62 therms per participant per year.

Table 1-1
Net Impact Estimates (Therms)

	Ex Ante Impact[1]	Ex Post Impact	Ex Post # of Audits[2]	Ex Post Program Impact	Realization Rate[3]
Do-It-Yourself Surveys (Audit)	24	44 (18,14)	39,681	1,745,964	183%

Notes:
1 Ex ante estimate filed in Advice No. 2267, February 1, 1994.
2 Number of surveys processed from January 1, 1994 through December 31, 1994.
3 Ex post percent of ex ante impact estimate.
4 The precision of the ex post impact estimate defined by the parenthetical values± the ex post impact estimate at the 90% and 80% confidence intervals, respectively.

The data preparation and analysis used to derive the estimates, as well as their documentation, adhered to the requirements of the Verification Protocols concerning load impact measurement. The statistical regression analysis is a Load Impact Regression Model as defined in the Protocols. The model is based on accepted empirical techniques for measuring the impacts of DSM programs, and it produces diagnostics that allow independent assessment of its performance. The specifications were based on sound behavioral and physical principles, and they included the major variables expected to influence gas consumption. Tests of important statistical issues that could arise in the estimation of model parameters were performed, and appropriate procedures were used to correct any significant problems that were identified.

The data met the requirements of the Verification Protocols concerning participant and non-participant sample sizes and coverage. The process of data acquisition and preparation is fully documented in this report and its appendices. All of the regression models that were estimated are presented in the report, with associated confidence statistics and related information.

Chapter 2
INTRODUCTION: IMPACT EVALUATION OVERVIEW

Chapter 2 INTRODUCTION: IMPACT EVALUATION OVERVIEW

IMPACT EVALUATION OBJECTIVE

This report presents the results of the impact evaluation of Southern California Gas (SoCal Gas) Company's 1994 Home Energy Fitness (HEF) Program and documents the data and analysis upon which the results are based. The HEF Program promotes the adoption of energy efficient measures and actions by providing informational audits to residential customers.

The objective of the impact evaluation was to estimate the natural gas savings that resulted from the 1994 HEF Program audits. The estimates include the gross changes in gas consumption from the program and the net changes, after accounting for the effects of actions that participants would have taken in the absence of the program.

The data collection and analysis used to obtain the impact results must conform to the requirements of the Protocols for the Verification of Demand-Side Management (DSM) Programs (Verification Protocols) governing the procedures used by the California investor owned utilities to estimate the impacts of their DSM programs. The Verification Protocols specify a series of requirements about data sources, analysis procedures, and reporting for impact evaluations. The key requirements that relate to this report are summarized in Tables 5, 6, 7, and C-11 of the protocols. Table 5 specifies the measurement methodology, the sample design and size, and the billing data requirements. Table 6 specifies the reporting requirements for the program impacts. Table 7 specifies the documentation requirements for data preparation and analysis. Table C-11 sets out additional requirements for the evaluation of Energy Management Services Programs, of which the HEF Program is an example.

DESCRIPTION OF THE HOME ENERGY FITNESS PROGRAM

SoCal Gas Company's 1994 HEF Program offered residential customers an analysis of their gas consumption and made recommendations as to how they could reduce their usage by adopting cost effective measures and practices. The analysis estimated a residential customer's annual natural gas use by end-use based on past consumption data and responses to a questionnaire that the participants completed on their gas using equipment characteristics and operation.

SoCal Gas marketed the HEF Program in 1994 by mass mailings. The marketing brochure invited SoCal Gas customers to complete an enclosed questionnaire and to receive a report by return mail. The marketing solicitation was not restricted to any one market segment. It was, however, in a relatively large number of cases, sent to residential customers in single family homes, with continuous gas service for over 10 years, and annual consumption of more than 640 therms in the colder regions of the service territory.

SoCal Gas also sent the offer to customers who had made bill inquiries. The offer letter and survey are reproduced in Appendix 2A.

In 1994, SoCal Gas sent out approximately 290,000 of the HEF brochures to its residential customers. Approximately 40,000 customers responded and received a personalized Home Energy Fitness Report. SoCal Gas treated these customers as HEF Program participants. This report included a personalized analysis of how much of the customer's annual gas bill was distributed by end-use and recommendations which potentially could save natural gas energy and reduce his/her gas bill. In addition, estimates of the cost savings associated with the recommendations were provided to the customer. An example of a typical report and a list of the types of recommendations are shown in Appendix 2A.

METHODOLOGY OVERVIEW

The estimated saving impacts of the HEF program were based on a statistical analysis of the gas consumption for a sample of participants and non-participants from January 1993 through November 1995. The statistical analysis consisted of the estimation of a series of regression models that relate consumption to variables representing program participation, weather, and gas equipment holdings, in addition to other residence and household characteristics. The specifications used in this analysis are referred to as Load Impact Regression Models (LIRM) in the Verification Protocols.

The estimated models are based on accepted empirical techniques for measuring program impacts, and the coefficients are estimated using techniques that account for the important statistical issues that arise in the estimation of regression model parameters. The estimation method accounted for both autocorrelation and heteroskedasticity which were found to be present in the data. In addition, other potential statistical problems were investigated. A detailed explanation of the model used to estimate program impacts appears in Chapter 4 of this report.

The data on which the regression models were estimated came from four sources. The first was the 1994 HEF program tracking database. This identified the participants and the dates on which the audit reports were sent to them. The second was SoCal Gas's billing system, from which consumption data were extracted for participants and a sample of non-participants covering the period from January 1993 through November 1995. The third was temperature data from the various weather stations in the SoCal Gas service territory. These were used to construct a heating degree day variable that was matched to the period of each gas bill. The last was a survey of a sample of participants and non-participants that gathered information on the respondents' demographic and dwelling characteristics.

The information from these sources was merged to construct a time series/cross sectional data set of consumption for a sample of participants and non-participants spanning the period from twelve months prior the program year through eleven months afterward (i.e. from January 1993 through November 1995). This data set permitted the comparisons of changes in gas consumption for participants and non-participants from before to after the program treatment, controlling for changes in weather, and other nonprogram effects.

In the course of the analysis, several different equation specifications were estimated. All of the models specified average daily gas use as the dependent variable and a core set of explanatory variables. These core explanatory variables included a binary variable for participation, heating degree days, several variables indicating equipment holdings, and an income variable. The various specifications, which entailed adding and redefining variables to the core set of independent variables, are presented and discussed in Chapter 4.

In the final model specification selected to estimate the net HEF program impacts, the program effects were represented by three distinct variables that capture different aspects of program participation. The first was a binary variable equal to one for participants in all billing periods, and zero for the non-participants. The second was a binary variable equal to one for all customers for billing periods from December 1994 through November 1995 and zero prior to December 1994. The third was a binary variable equal to one for program participants after they received the audit report, and zero for them before. The variable was zero for non-participants in all periods. The first variable captured any underlying differences in average consumption between participants and non-participants, after controlling for demographics and dwelling characteristics. The second variable captured any underlying change in gas consumption in the last year. This change is interpreted as the result of any "naturally occurring" conservation (i.e. free ridership effects), as well as the effects of background economic variables, on consumption. The last variable captured the average net change in gas consumption for participants from before to after receipt of the audit report, after netting out the naturally occurring conservation, and controlling for weather changes, and other effects.

ORGANIZATION OF REPORT

The remainder of this report in organized as follows. Chapter 3 describes, in detail, the data used in the analysis. Issues regarding the sampling procedure and connected data preparation are outlined. The associated Appendices contain a detailed accounting of specific information when appropriate.

The methodology underlying the statistical analysis used to estimate the energy savings impact of SoCal Gas Company's 1994 HEF Program are presented in Chapter 4. Additional potential and actual statistical problems are addressed and their solutions outlined when applicable.

The empirical results of the conditional demand analysis and the estimates of net and gross impacts on gas energy savings are reported in Chapter 4.

Chapter 3
SAMPLING AND DATA COLLECTION AND PREPARATION

Chapter 3 SAMPLING AND DATA COLLECTION AND PREPARATION

BILLING DATA EXTRACTION AND PREPARATION

The data used to estimate the HEF program impacts were drawn from the following four sources:

- The HEF Program Tracking Data Base;
- SoCal Gas Company's billing system;
- NOAA temperature readings at weather station in the SoCal Gas service territory; and
- a telephone survey of a sample 1994 HEF participants and non-participants.

Initially SoCal Gas provided a file of all 1994 HEF participants with continuous billing records from January 1993 through November 1995. This file contained information for 30,363 of the 39,861 customers categorized as 1994 participants. In addition, SoCal Gas provided the bills for a large random sample of non-participant residential customers with continuous service since 1993 who lived in single family residences. The length of service and dwelling type criteria matched those upon which SoCal had targeted its HEF marketing efforts. The size of this initial non-participant random sample was 44,914 customers.

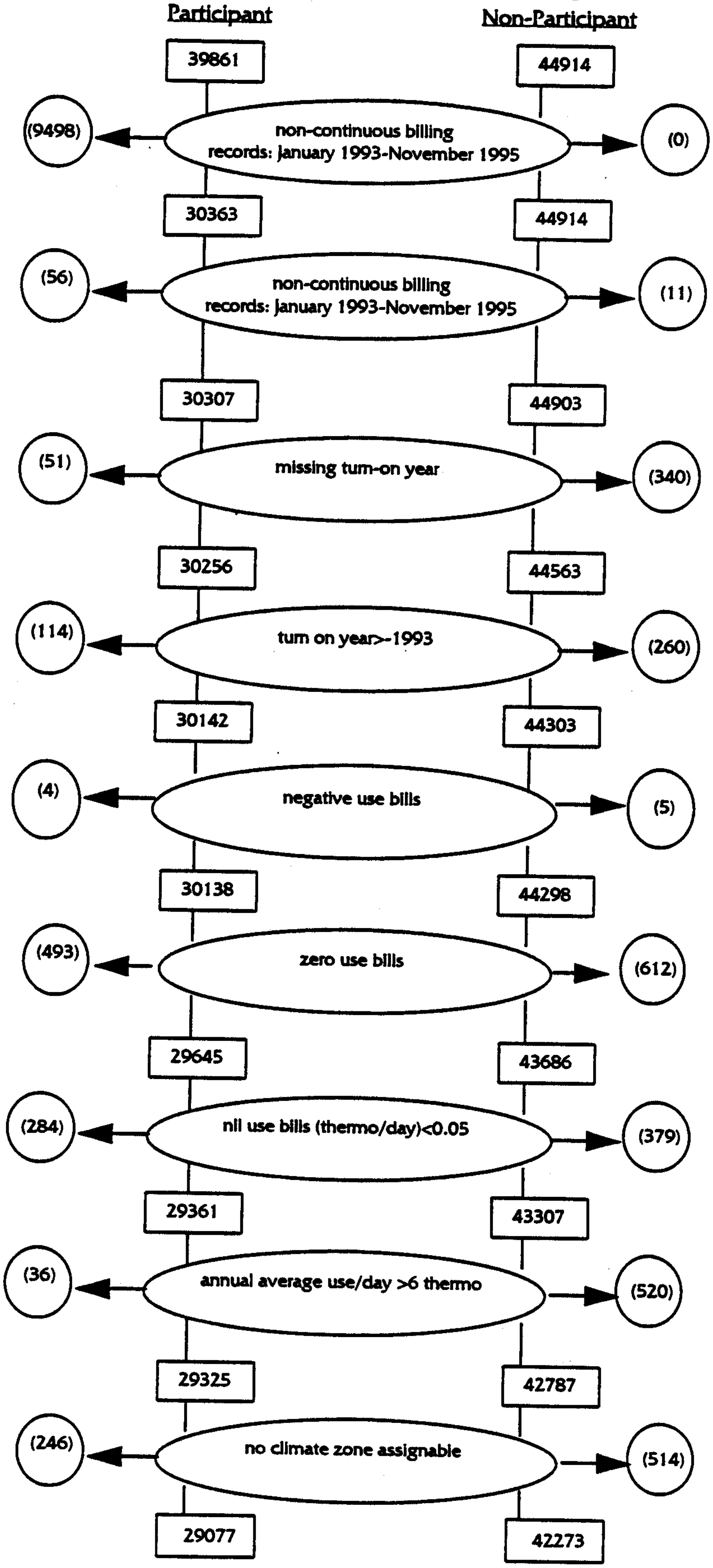
The billing data from these two files were reviewed for completeness and plausibility. Cases with missing information for key variables, negligible consumption, or average daily use greater than six therms were deleted form the data sets. The billing data provided by SoCal Gas Company was in very good condition in terms of completeness and plausibility of the values of consumption. It already had been screened to some extent by SoCal Gas -- at least to ensure that the full period from Jan./Feb. 1993 through June/July 1995 was represented. SRC's additional screening eliminated only a small number of cases from the billing records before the gross survey samples were drawn.

Some of these billing records had missing information which prevented assignment of sample selection categories (no turn-on year, no climate zone). Others were considered unsuitable for inclusion in the analysis because :

- there was a change in customer ID during the analysis period,
- the turn-on year listed was 1993 or later,
- there was at least one billing period of negligible use (implying a likely vacancy),
- annual average use (1993, 1994, or 1995) was much greater than would be expected for a single-family home.

After the screening process, 29077 participants and 42273 non-participants remained. None of these screening criteria affected the size of the available sample to any significant degree. Appendix 3A (Table 3A-1) contains a listing of how many customers were eliminated at each stage of this process. The numbers shown are those eliminated after all preceding criteria have been applied. Figure 3-1 diagrams the screening process.

FIGURE 3-1
Attrition of Cases in Inital Sample by Screening Criterion



PARTICIPANT SAMPLE

This section documents the process used to select the sample of participants that was surveyed for SoCal Gas Company's HEF Program evaluation project. The sample frame used to draw the sample was the full set of 29077 participating customers for which there was billing data and which survived the screening process described above.

The gross survey sample (i.e. cases sent to the survey implementation subcontractor) was drawn to match the distribution of cases in the sample frame with respect to climate zone, turn-on year, average 1993 use per day, and a seasonality factor (the ratio of average winter to average annual use in 1993). Climate zones were defined by assigning a climate zone category to each postal zip code recorded for each customer. The correspondence of climate zone to zip codes in contained in Appendix 3A (Table 3A-2).

First, categories were defined for each of the four stratification dimensions. These dimensions were Climate Zone, Turn-On Year, 1993 Average Daily Use, and 1993 Winter/Annual Average Use per Day. The distributions of the full sample of participants by these parameters is presented in Appendix 3B (Table 3B-1).

Next, a matrix of frequencies of the 288 possible combinations of these categories was created. A copy of this matrix, expressed as percentages, may be found in Appendix 3B (Table 3B-2).

A sample of 2790 participants was then drawn proportionally to this matrix (every nth occurrence of a specific attribute combination) with the exception that matrix cells corresponding to Climate Zone 1 were over sampled by a factor of about 5:1, and those of Climate Zone 2 were over sampled by about 2:1. The resulting climate zone distribution of the drawn sample is presented in Appendix 3B (Table 3B-3).

Including those selected for the pretest, a total of 2891 participants were sent to the survey implementation subcontractor, Northwest Research Group (NRG), as potential survey participants. The distribution of these 2891 customers across the 288 matrix cells defined above is shown in Appendix 3B(Table 3B-4). NRG was asked to ensure that the climate zone distribution of completed surveys was approximately equal to that of the gross selected sample.

NONPARTICIPANT SAMPLE

This section documents the process used to select the sample of non-participants to be surveyed for SoCal Gas Company's program evaluation project. The sample frame was the full set of 42273 comparison customers provided by SoCal Gas Company for which there was billing data and which survived the screening process described earlier.

The procedure used to select the non-participant sample was comparable to that used to select the participants. The non-participant survey sample was drawn to approximately match the distribution of the participant sample frame, except that Climate Zone 1 cells were over sampled by a factor of about 5:1 and Climate Zone 2 cells were over sampled by about 2:1.

Together with the pretest, a total of 2734 non-participants were sent to NRG. NRG was asked to ensure that the distribution of climate zones among completed surveys was close to that of the participant sample frame.

The selection matrix distribution of the full non-participant sample and of the selected non-participants sent to NRG may be found in Table 3C-1 of Appendix 3C. Tables 3C-2, 3C-3, and 3C-4 correspond to Table 3B-2, 3B-3, and 3B-4 respectively for non-participants. Aggregate distributions are listed there along with those of participants for purposes of comparison.

SURVEY ADMINISTRATION

The survey implementation subcontractor randomized the combined participant and non-participant gross samples and broke the merged file into subsets. Only after a minimum numbers of attempts were made for all cases in one subset was the next subset used. NRG used 201 cases in the survey pre-test. Of the remaining 5424 cases, NRG only attempted to contact 4,307 of them. 548 were not contacted because there was no valid telephone number available. The remainder were not needed to fill the completion quotas of 1000 participants and 1000 non-participants. Out of the 3008 qualified customers that NRG succeeded in contacting, almost 68% completed the full survey. The disposition of the 4307 cases is summarized in Appendix 2A (page 8 of the Field Service Report).

SURVEY RESULTS

The completed survey file returned by the survey implementation subcontractor contained responses for 1017 participants and 1019 non-participants. The means for key variables used in the regression analysis, broken down by participants and non-participants, are presented in Appendix 3Da and 3Db. Even though the survey included multiple questions addressing free-ridership it is apparent from the frequency calculations done for these variables that in the majority of cases (page 2Da-1 through page 2Da-34) theses variables corresponding to the survey questions associated with free-ridership contained no information. Thus, these data were not used in the analysis because in a preponderance of cases the information did not exist. Note that the these mean values may differ from those reported in Table 4-1 of Chapter 4 because Table 4-1 reports the mean values of the regression variables based on cases where there was no item nonresponse to any question that affected the variable's construction. The frequencies of all the responses to the survey questions are reported in Appendix 3Da and 3Db.

Table 3-1
Variable Definitions and Mean Values in the Full Survey Data

Variable Label	Variable Definitions	Mean Value	
	Independent Variable	Participants	Non-participants
INCOME	Household income in thousands	35.031	40.913
V46T65	Binary variable equal to 1 if dwelling was built between 1946 and 1965, 0 otherwise	0.489	0.456
V66T77	Binary variable equal to 1 if dwelling was built between 1966 and 1977, 0 otherwise	0.253	0.220
VPOST77	Binary variable equal to 1 if dwelling was built after 1977, 0 otherwise	0.117	0.153
AGEP65	number or residents of each premise who were 65 years or older	0.902	0.682
HHSIZE	Number of people living in the residence	2.460	2.715
DAYHOME	Binary variable equal to 1 if some residents are normally at home during the day, 0 otherwise	0.769	0.757
GASDRYR	Binary variable equal to 1 if a gas clothes dryer is used, 0 otherwise	0.628	0.674
GASWHT	Binary variable equal to 1 if a gas hot water heater is used, 0 otherwise	0.975	0.970
GASHEAT	Binary variable equal to 1 if gas is primary heating fuel, 0 otherwise	0.945	0.937
PILOT	Binary variable equal to 1 if stove has pilot light, 0 otherwise	0.329	0.387
GASCOOK	Binary variable equal to 1 if gas is primary cooking fuel, 0 otherwise	0.730	0.748
GASPOOL	Binary variable equal to 1 if gas is used to heat a pool, 0 otherwise	0.056	0.062
SBT	Binary variable equal to 1 if a set back thermostat is used, 0 otherwise	0.265	0.291
NEWHTSYS	Binary variable equal to 1 if new space heating equipment was installed, 0 otherwise	0.069	0.062
GOODINS	Binary variable equal to 1 if premise reported "good insulation", 0 otherwise	0.653	0.573
SQFT	Area in square feet of each premise	1646.04	1777.5

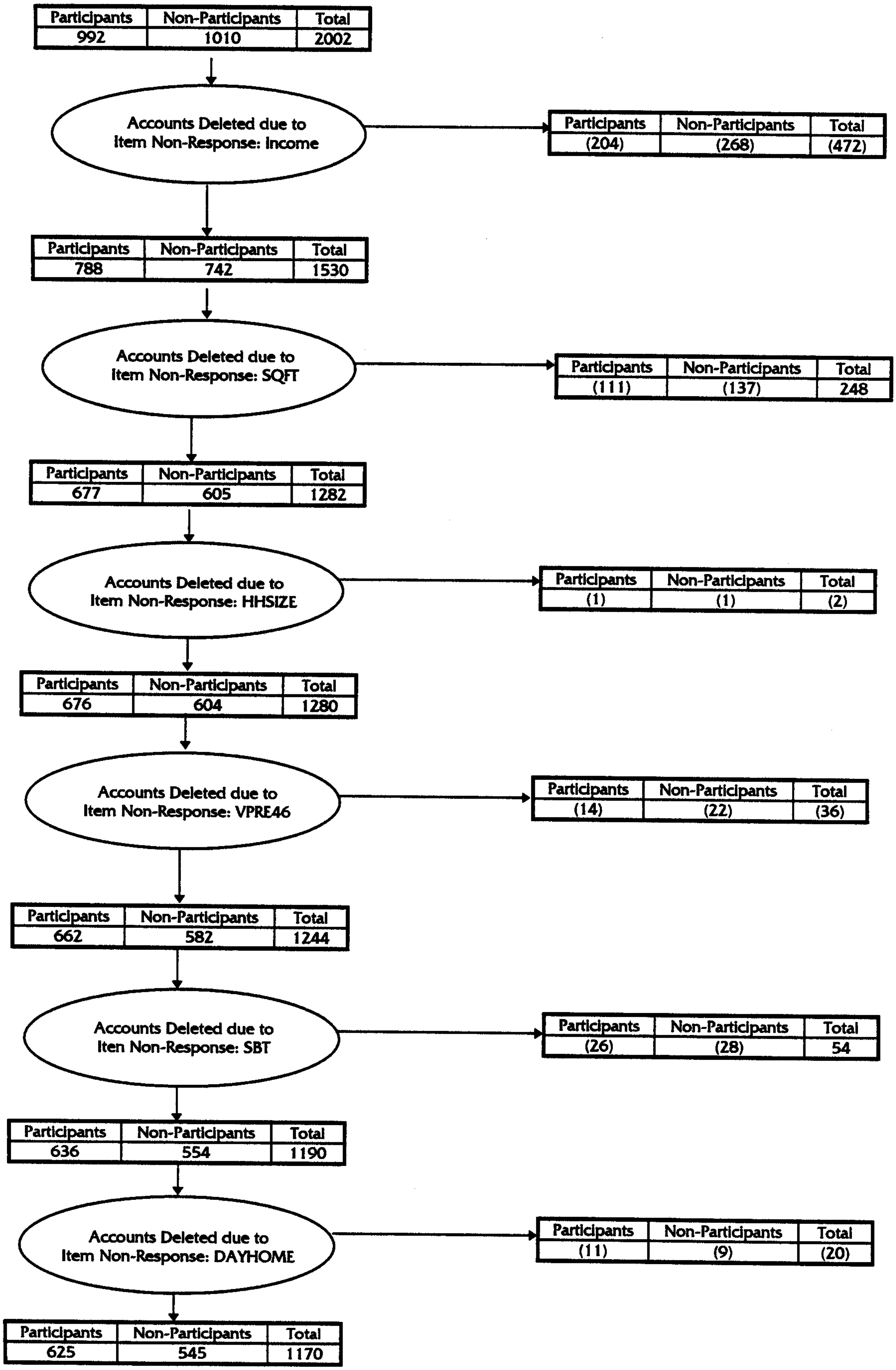
THE ANALYSIS DATA BASE

The data base used in the analysis consisted of a cross section of customers each with a monthly time series including its gas consumption and contemporaneous weather information. In order to organize the data into a usable form for the regression analysis, a total of 34 customers with less than the full 35 months of billing data were deleted from the original program data base. Out of these 34 cases, 25 were participants, and 9 were non-participants. Thus, prior to any attrition in the number of observations due to item nonresponse to the survey, the data base contained survey, billing, and weather variables on 2002 household accounts over 35 consecutive and contemporaneous months corresponding to January 1993 through November 1995 (70070 observations). This represented 1010 non-participant accounts and 992 participant accounts. Frequency distributions of this balanced data set by participation and climate zone may be found in Appendix 3D (Tables 3D-1 and 3D-2).

As will be discussed more fully in the next chapter, the regression analysis required a time series, cross sectional data set with the same number of consecutive time series on each account. In addition, when missing values, due to item nonresponse, for a independent variable used in the analysis were identified, it was the case that the values were missing for the entire associated time series. Given the nature of the missing data, no obvious interpolation techniques appeared appropriate. Therefore it was necessary, when missing data on certain variables were encountered, to delete the complete time series associated with a customer.

Applying a non-missing data rule for variables with important theoretical impacts on gas usage reduced the number of observations available for the regression based analysis to 41650. These observations represented 1190 household accounts (554 non-participant accounts and 636 participant accounts). Thus, because of missing data on crucial variables, 812 accounts were deleted from the data set used in the analysis. Out of this total, 456 non-participant accounts were deleted and 356 participant accounts were deleted. A detailed frequency distribution of participation status by climate zone are included in Appendix 3D (Tables 3D-3 and 3D-4). In Appendix 3D, Table 3D-5 lists attrition due to missing values. The numbers reported are those eliminated after all preceding criteria have been applied. Figure 3-2 diagrams the deletion process.

FIGURE 3-2
Deletions Due to Item Non-Response
(Deletions to Balanced Data Base t=1,...35 i=1,...2002)



Chapter 4
METHODOLOGY AND RESULTS

Chapter 4 METHODOLOGY AND RESULTS

METHODOLOGY AND RESULTS

Regression Based Demand Analysis

Synergic Resources Corporation (SRC) performed a regression analysis on the merged billing, weather, and survey data to estimate the natural gas savings attributable to the 1994 HEF Program. The objective of the regression analysis was to estimate the net impact of the HEF Program, after controlling for other factors that influence natural gas consumption. In addition, the analysis quantified differences in participant and non-participant natural gas usage and differences in usage due to underlying economic and environmental trends.

The general form of the specification of the conditional demand relationship which was estimated was

$$Y_{it}=\Sigma\beta_k X_{kit} + \varepsilon_{it}$$

where
- i = 1,2,3,...,I(I=total number of accounts);
- t = 1,2.3,..,T(T=35=number of billing cycles) ;
- Y_{it} = average natural gas use per day in billing cycle t by customer i;
- X_{kit} = the k^{th} explanatory variable defined in Table 4-1;
- β_k = the k^{th} regression coefficient;
- ε_{it} = random component representing unobserved factors affecting consumption;

A linear specification was chosen because the relationship between total gas consumption and the ownership of different end-uses is arithmetically linear.

Model Estimation

The parameters of a basic model specification were estimated using statistical procedures that test and correct for the presence of serially correlated errors and heteroskedasticity when either is found. The tests indicated that the errors in the models were both serially correlated and heteroskedastic. A detailed discussion of the techniques used to test and correct for these problems is presented later in this chapter.

In all, ten model specifications were examined. The definitions for the variables used in the different models are presented in Table 4-1. Table 4-2 presents a summary of the different model specifications. If the variable was included in the model specification and had a positive sign, this is indicated by an X. If the X is enclosed in parentheses (i.e. (X)), the parameters estimate was negative. An asterisk (*) indicates that the parameter estimate for that variable was **not** statistically significant. The estimated models are presented in Tables 4-3 through 4-11 in this chapter.

The key variables that capture the effects of the HEF program as well as any "background" factors that are contemporaneous with the program impacts are PARTIC, POST, GOTHEF, SG, SP and GSASQ. PARTIC is defined as one for participants and zero for non-participants. It captures any underlying difference in gas consumption between the two groups, after controlling for other observable variables. POST, which enters only in Models 6 through 10, captures any underlying trend in gas usage in the population. It is defined as 1 for both participants and non-participants for all billing periods from December 1994 through November 1995 and zero prior to December 1994. PSA measures the interaction of POST and AHDD65 (heating degree days). The coefficient of POST (and PSA) represents the effects of economic and environmental trends on gas consumption, as well as the impacts of any "naturally" occurring conservation actions on the average daily gas usage which occurred during the last year. The parameter estimate should be interpreted as the reduction in the average amount of natural gas consumption from 1994 to 1995, after controlling for the other variables. Thus, the estimated coefficient of POST(and PSA) captures any "free ridership" effects in the program, as well as the impact of general economic trends.

GOTHEF is defined as one for participants in months after they received the HEF audit report, and zero in preceding months. SG measures the interaction of GOTHEF and AHDD65, and GSASQ measures the interaction SG and SQFT (square feet of interior space). These three variables are zero for all months for non-participants. When GOTHEF (and SG and GSASQ) enter the model specification without POST, it captures the gross change in gas consumption following the receipt of the audit report. When it enters the specification along with POST, it is interpreted as the net impact, after controlling for any naturally occurring conservation (picked up by POST).

No attempt was made in any of the models to estimate the program impacts by end-use category. There were two primary reasons for this. First, the program tracking did not record what measures and actions were recommended to each recipient of the audit report. While the survey asked respondents what, if any, actions they took based on the report, the response rate to this question was very low. As a result, there was no way to identify what end-uses could have been affected by the recommendations with any degree of reliability. Second, almost all of the customers in the sample had both space and water heating. If binary variables representing these end-uses interacting with program participation variables were included in the specification, we expected that the program effects would be confounded by the high level of collinearity among the end-use variables and the constant term in the model.

In the model specifications where the key program participation variable, GOTHEF, interacts with heating degree days (discussed below), one might argue that the effect of the program on space heating consumption is captured by this weather sensitive variables. However, water heating has a seasonal use pattern whose effect would be partially picked up by this variable. We reject the argument that the weather sensitive portion of any program impacts can be attributed exclusively to space heating.

In the first model specification that was estimated, daily gas usage was regressed on the following set of variables:

- gas equipment holdings (GASDRYR, GASCOOK, GASHEAT, GASWHT, GASPOOL),
- whether a new water heater or space heater was installed recently (PM_WH, NEWHTSYS),
- an indicator of the quality of home insulation (GOODINS),
- an indicator for the presence of a set back thermometer (SBT),
- indicators of the age of the residence, as well as the respondent's estimate of the square feet (V46T65, V66T77, VPOST77, SQFT),
- household income, number of household members, whether people are typically at home during the day, and how many people 65 and over lived in the residence (INCOME, HHSIZE, DAYHOME, AGE65P),
- average daily heating degree days (AHDD65), and
- program specific independent variables and interactions of these variables with heating degree days (PARTIC, GOTHEF, SP, SG).

The estimation results of the first model (Model 1) were generally reasonable. The coefficient of PARTIC is positive and significant, indicating that participants have higher consumption than non-participants, after controlling for the other factors. GOTHEF is negative and statistically significant. This is interpreted to mean that participants reduced gas consumption significantly after receiving the HEF audit report. The interaction term of GOTHEF with heating degree days is also negative and significant, indicating that the reductions are also weather sensitive. At the sample mean of heating degree days, this implies an average daily reduction in gas consumption of almost 0.2 therms.

Most, but not all, of the other parameter estimates are reasonable from the standpoint of interpretation and statistical significance. Average gas use is greater for older homes. Gas consumption is positively related to household income, presence of older people in the dwelling (AGE65P), heating degree days, household size, dwelling square feet, and the ownership of major gas appliances. It is negatively related to indicators of good insulation and a recently installed new water or space heater.

The values of the indicators representing gas cooking (negative) and set back thermostat (positive) are implausible, and the statistical significance of other variables is low (SP, GASWHT, PM_WH, GOODINS).

In an effort to correct some of these inconsistencies and to examine the effects of excluding other variables, the second model (Model 2) was modified in the following manner. The two weather/participation interaction variables were omitted. The nontemporal participation variable was also omitted. The variable which measured the number of people aged 65 and over was also dropped because of the possibility of collinearity with the income variable. The binary variable indicating a pilot light in a stove replaced the gas cooking binary variable.

Because the variable indicating gas water heat and a variable indicating if a recommended water heating related measure was installed were statistically insignificant in both the initial specifications, they were omitted it in the estimation of Model 3.

The variables which indicated the presence of a gas pilot light and the presence of "good insulation" continued to be statistically insignificant, and were omitted in the specification of Model 4. Model 5 contains Model 4 variables in addition to the square of the heating degree day variable and the square of the interior space variable. These were included in order to capture any nonlinearity in the effects of these factors. Both these new variables were statistically significant.

The Model 6 specification is identical to that of Model 5 with the reintroduction of the earlier nontemporal program participation variable PARTIC and the variable POST which is one for both participants and non-participants in the last twelve months of the sample period and zero otherwise. This last variable was introduced in order to isolate changes in gas usage which occurred independently of the program's influence. The square of the heating degree day variable was deleted because the heating load should be approximately linearly related to AHDD65 (based on the physical relation).

Model 7 eliminates the GASHEAT variable from the previous model because over 97% of respondents indicated they had gas heating. The variable is highly correlated with the constant term in the model, causing a potential collinearity problem. Moreover, all of the customers in the sample have natural gas service, and most of them use an above average amount of gas. Based on their gas consumption, there is reason to believe that even the 3% who say they do not have gas heating actually do use it. As expected, the effect of the elimination of the indicator of natural gas heat in Model 7 is almost completely absorbed in the constant term, and the parameters of the other important variables change little from their previous values.

In Model 8, the key variable that captures the net effect of the program, GOTHEF, was entered in the specification interactively with heating degree days (GOTHEF*AHDD65), as well as entering alone. While the results show that the savings depend on weather, the estimate of the simple variable, GOTHEF, falls and becomes insignificant.

The final two models that were estimated (Model 9 and Model 10) more fully incorporated weather sensitivity into the model specification. In Model 9, four new variables were included. PSA (POST*AHDD65) measured the interaction between POST and heating degree days. When it was included in the model along with POST, its coefficient was significant while the coefficient of POST was no longer significant. In Model 9 the overall effect of POST was approximately half of what it was in Model 6 or Model 7.

As in Model 8, the key variable that captures the net effect of the program, GOTHEF, was entered in the specification interactively with heating degree days times interior square footage (GOTHEF*AHDD65*SQFT), as well as entering alone. Both these variables were significant in the final two models. In addition to these two program participation variables, space and water heating interaction variables were included in Model 9. GWHH (GASWHT*HHSIZE) was not significant, while GHSAQ (GASHEAT*AHDD65*SQFT) was significant. While both GSASQ and GHSASQ were significant, the coefficient of POST which measures free ridership fell substantially.

Table 4-1

Variable Definitions and Mean Values in Regression Sample

Variable Label	Variable Definitions	Mean Value	
	Independent Variables	Participants	Nonparticipants
DAYUSE	Daily Gas Usage (Gas use for billing cycle divided by # of days in billing cycle)	1.648	1.686
PARTIC	Binary variable equal to 1 if the respondent was a program participant, 0 otherwise	1.000	0.000
GOTHEF	Binary variable equal to 1 for participants after the date of HEF package was sent, 0 before and for non-participants in all months	0.391	0.000
POST	Binary variable equal to 1 for participants for last 12 months of the time period, 0 otherwise	0.343	0.343
SP	AHDD65*PARTIC	5.012	0.000
SG	AHDD65*GOTHEF	2.118	0.000
GSASQ	AHDD65*GOTHEF*SQFT	3237.50	0.000
GHSASQ	AHDD65*GASHEAT*SQFT	7335.21	7866.05
PSA	AHDD65*POST	1.973	1.950
GWHH	GASWHT*HHSIZE	2.415	2.785
V46T65	Binary variable equal to 1 if dwelling was built between 1946 and 1965, 0 otherwise	0.489	0.435
V66T77	Binary variable equal to 1 if dwelling was built between 1966 and 1977, 0 otherwise	0.275	0.241
VPOST77	Binary variable equal to 1 if dwelling was built after 1977, 0 otherwise	0.121	0.175
INCOME	Annual income in $1000's	38.184	44.567
AGE65P	number or residents of each premise who are 65 years or older	0.865	0.610
AHDD65	Average daily heating degree days in billing cycle	5.012	4.987
AHDD65Q	AHDD65 SQUARED	64.265	63.958
HHSIZE	Number of people in the household	2.463	2.823
DAYHOME	Binary variable equal to 1 if some residents are normally at home during the day, 0 otherwise	0.758	0.747
GASDRYR	Binary variable equal to 1 if a gas clothes dryer is used, 0 otherwise	0.646	0.711
GASWHT	Binary variable equal to 1 if a gas hot water heater is used, 0 otherwise	0.981	0.987

Variable Label	Variable Definitions	Mean Value	
GASHEAT	Binary variable equal to 1 if gas is primary heating fuel, 0 otherwise	0.972	0.975
PILOT	Binary variable equal to 1 if stove has pilot light, 0 otherwise	0.330	0.360
GASCOOK	Binary variable equal to 1 if gas is primary cooking fuel, 0 otherwise	0.726	0.737
GASPOOL	Binary variable equal to 1 if gas is used to heat a pool, 0 otherwise	0.064	0.071
SBT	Binary variable equal to 1 if a set back thermostat is used, 0 otherwise	0.266	0.332
NEWHTSYS	Binary variable equal to 1 if new space heating equipment was installed, 0 otherwise	0.040	0.037
PM_WH	Binary variable equal to 1 if a recommended source heating related measure was installed, 0 otherwise	0.023	0.000
GOODINS	Binary variable equal to 1 if premise reported "good insulation", 0 otherwise	0.652	0.589
SQFT	Area in square feet of each premise	1549.61	1637.95
SQFTQ	SQFT SQUARED	2939833.18	3314947.53

Table 4-2

Model Specifications

Dependent Variable= DAYUSE	MODEL NUMBER									
Independent[1] Variable	1	2	3	4	5	6	7	8	9	10
CONSTANT	(X)	(X)	(X)	(X)	(X)	(X)	(X)*	(X)	X	X*
PARTIC	X					X	X	X	X	X
GOTHEF	(X)	(X)	(X)	(X)	(X)	(X)*	(X)	(X)*	(X)	(X)
POST						(X)	(X)	(X)	(X)*	(X)*
SP	(X)*									
SG	(X)							(X)		
V46T65	(X)	(X)	(X)	(X)	(X)	(X)	(X)	(X)	(X)	(X)
V66T77	(X)	(X)	(X)	(X)	(X)	(X)	(X)	(X)	(X)	(X)
VPOST77	(X)	(X)	(X)	(X)	(X)	(X)	(X)	(X)	(X)	(X)
INCOME	X	X	X	X	X	X	X	X	X	X
AGE65P	X									
AHDD65	X	X	X	X	X	X	X	X	X	X
AHDD65Q					(X)					
HHSIZE	X	X	X	X	X	X	X	X	X*	X
DAYHOME	X*	X	X	X	X	X	X	X	X	X
GASDRYR	X	X	X	X	X	X	X	X	X	X
GASCOOK	(X)									
GASHEAT	X	X	X	X	X	X				
GASWHT	X*	(X)*								
PILOT		X*	X*							
GASPOOL	X	X	X	X	X	X	X	X	X	X
SBT	X*	X	X	X*	X	X*	X	X	X*	X*
NEWHTSYS	(X)	(X)	(X)	X	(X)	(X)*	(X)	(X)*	(X)*	(X)*
PM_WH	X*	(X)*								
GOODINS	(X)*	(X)*	X*							
SQFT	X	X	X	X	X	X	X	X	X	X
SQFTQ					(X)	(X)	(X)	(X)	(X)	(X)
GSASQ									(X)	(X)
GWHH									(X)*	
GHSASQ									X	X
PSA									(X)*	

X indicates that the variable was included in the model specification and its parameter estimate was positive.

(X) indicates that the variable was included in the model specification and its parameter estimate was negative.

* indicates that the parameter estimate was not statistically significant at the 90% confidence level.

[1] Variable definitions are presented in Table 4-1.

Table 4-3

MODEL 1: Dependent Variable = DAYUSE

Variable	Parameter Estimate	Corrected Standard Error	T-Statistic
CONSTANT	-0.470	0.211	-2.226
PARTIC	0.120	0.032	3.797
GOTHEF	-0.087	0.026	-3.396
SP	-0.002	0.004	-0.556
SG	-0.022	0.004	-6.128
V46T65	-0.146	0.060	-2.426
V66T77	-0.228	0.062	-3.675
VPOST77	-0.469	0.068	-6.924
INCOME	0.004	0.001	4.746
AGE65P	0.148	0.024	6.084
AHDD65	0.180	0.002	73.081
HHSIZE	0.091	0.012	7.407
DAYHOME	0.020	0.040	0.496
GASDRYR	0.143	0.036	4.005
GASCOOK	-0.103	0.040	-2.612
GASHEAT	0.495	0.120	4.133
GASWHT	0.023	0.162	0.141
GASPOOL	0.599	0.084	7.142
SBT	0.048	0.037	1.314
SQFT	0.2E-3	0.3E-4	6.551
PM_WH	-0.001	0.087	-0.012
NEWHTSYS	-0.128	0.073	-1.752
GOODINS	-0.040	0.034	-1.181

Number of Observations = 39200
Number of Customers = 1120
R^2 =.543*

---

* $R^2 = 1 - \frac{\Sigma\Sigma(Y_{it} - \Sigma\hat{B}_j X_{jit})^2}{\Sigma\Sigma(Y_{it} - \bar{Y})^2}$

Table 4-4

MODEL 2: Dependent Variable = DAYUSE

Variable	Parameter Estimate	Corrected Standard Error	T-Statistic
CONSTANT	-0.329	0.217	-1.515
GOTHEF	-0.156	0.026	-5.977
V46T65	-0.105	0.062	-1.700
V66T77	-0.205	0.064	-3.180
VPOST77	-0.476	0.070	-6.771
INCOME	0.003	0.001	3.668
AHDD65	0.175	0.002	110.393
HHSIZE	0.065	0.012	5.378
DAYHOME	0.094	0.040	2.339
GASDRYR	0.093	0.035	2.623
PILOT	0.043	0.035	1.225
GASHEAT	0.506	0.125	4.041
GASWHT	-0.006	0.168	-0.034
GASPOOL	0.599	0.085	7.031
SBT	0.064	0.038	1.712
SQFT	0.2E-3	0.3E-04	6.865
NEWHTSY	-0.128	0.076	-1.700
PM_WH	-0.001	0.087	-0.008
GOODINS	-0.020	0.034	-0.598

Number of Observations = 40915
Number of Customers = 1111
R^2 =.538

Table 4-5

MODEL 3: Dependent Variable = DAYUSE

Variable	Parameter Estimate	Corrected Standard Error	T-Statistic
CONSTANT	-0.335	0.155	-2.154
GOTHEF	-0.156	0.026	-6.031
V46T65	-0.105	0.062	1.700
V66T77	-0.205	0.064	-3.182
VPOST77	-0.476	0.070	-6.759
INCOME	0.003	0.001	3.669
AHDD65	0.175	0.002	110.393
HHSIZE	0.065	0.012	5.378
DAYHOME	0.094	0.040	2.345
GASDRYR	0.093	0.035	2.627
PILOT	0.043	0.035	1.224
GASHEAT	0.505	0.125	4.041
GASPOOL	0.599	0.086	7.000
SBT	0.064	0.038	1.708
SQFT	0.2E-3	0.3E-4	6.865
NEWHTSY	-0.128	0.075	-1.701
GOODINS	-0.021	0.034	-0.600

Number of Observations = 38885
Number of Customers = 1111
R^2 =.538

Table 4-6

MODEL 4: Dependent Variable = DAYUSE

Variable	Parameter Estimate	Corrected Standard Error	T-Statistic
CONSTANT	-0.380	0.140	-2.72
GOTHEF	-0.162	0.026	-6.34
V46T65	-0.123	0.058	-2.101
V66T77	-0.217	0.060	-3.592
VPOST77	-0.502	0.067	-7.532
INCOME	0.003	0.001	3.997
AHDD65	0.174	0.002	114.382
HHSIZE	0.072	0.012	6.059
DAYHOME	0.091	0.039	2.343
GASDRYR	0.075	0.034	2.194
GASHEAT	0.547	0.114	4.818
GASPOOL	0.552	0.082	6.763
SBT	0.066	0.038	1.773
SQFT	0.2E-3	0.3E-4	7.276
NEWHTSYS	-0.140	0.071	-1.963

Number of Observations = 40915
Number of Customers = 1169
R^2 =.528

Table 4-7

MODEL 5: Dependent Variable = DAYUSE

Variable	Parameter Estimate	Corrected Standard Error	T-Statistic
CONSTANT	-0.602	0.125	-4.831
GOTHEF	-0.167	0.023	-7.116
V46T65	-0.129	0.051	-2.538
V66T77	-0.237	0.053	-4.510
VPOST77	-0.525	0.058	-9.027
INCOME	0.002	0.001	2.797
AHDD65	0.189	0.003	66.500
AHDD65Q	-0.001	0.1E-3	-5.687
HHSIZE	0.078	0.010	7.558
DAYHOME	0.096	0.034	2.847
GASDRYR	0.073	0.030	2.423
GASHEAT	0.516	0.099	5.212
GASPOOL	0.509	0.073	6.991
SBT	0.045	0.033	1.379
SQFT	0.5E-3	0.5E-4	10.488
SQFTQ	-0.4E-09	0.7E-10	-6.291
NEWHTSYS	-0.112	0.0634	-1.767

Number of Observations = 40915
Number of Customers = 1169
R^2 =.542

Table 4-8

MODEL 6: Dependent Variable = DAYUSE

Variable	Parameter Estimate	Corrected Standard Error	T-Statistic
CONSTANT	-0.662	0.143	-4.622
PARTIC	0.092	0.036	2.560
GOTHEF	-0.120	0.030	-4.030
POST	-0.119	0.027	-4.416
V46T65	-0.144	0.058	-2.49
V66T77	-0.245	0.060	-4.110
VPOST77	-0.536	0.066	-8.155
INCOME	0.002	0.001	2.688
AHDD65	0.173	0.002	115.370
HHSIZE	0.079	0.012	6.816
DAYHOME	0.093	0.038	2.45
GASDRYR	0.077	0.034	2.277
GASHEAT	0.549	0.113	4.857
GASPOOL	0.508	0.081	6.307
SBT	0.050	0.037	1.356
SQFT	0.5E-3	0.5E-4	9.700
SQFTQ	0.5E-10	0.8E-10	-5.731
NEWHTSYS	-0.090	0.071	-1.270

Number of Observations = 40915
Number of Customers = 1169
R^2 =.532

Table 4-9

MODEL 7: Dependent Variable = DAYUSE

Variable	Parameter Estimate	Corrected Standard Error	T-Statistic
CONSTANT	-0.122	0.086	-1.421
PARTIC	0.093	0.036	2.554
GOTHEF	-0.121	0.030	-4.049
POST	-0.119	0.027	-4.389
V46T65	-0.139	0.058	-2.388
V66T77	-0.250	0.060	-4.156
VPOST77	-0.524	0.066	-7.896
INCOME	0.002	0.001	2.602
AHDD65	0.173	0.002	113.951
HHSIZE	0.076	0.012	6.481
DAYHOME	0.099	0.038	2.580
GASDRYR	0.071	0.0340	2.101
GASPOOL	0.525	0.081	6.508
SBT	0.062	0.037	1.683
SQFT	0.5E-3	0..5E-4	9.681
SQFTQ	-0.5E-9	0.8E-10	-5.705
NEWHTSYS	-0.076	0.071	-1.070

Number of Observations = 40915
Number of Customers = 1169
R^2 =.529

Table 4-10

MODEL 8: Dependent Variable = DAYUSE

Variable	Parameter Estimate	Corrected Standard Error	T-Statistic
CONSTANT	-0.148	0.086	-1.718
PARTIC	0.099	0.036	2.731
GOTHEF	-0.037	0.030	-1.226
SG	-0.021	0.003	-6.549
POST	-0.083	0.027	-3.050
V46T65	-0.137	0.058	-2.345
V66T77	-0.247	0.060	-4.101
VPOST77	-0.524	0.066	-7.890
INCOME	0.002	0.001	2.574
AHDD65	0.177	0.002	101.653
HHSIZE	0.074	0.012	6.388
DAYHOME	0.099	0.039	2.570
GASDRYR	0.072	0.034	2.110
GASPOOL	0.524	0.081	6.485
SBT	0.061	0.0371	1.637
SQFT	0.5E-3	0.5E-6	9.644
SQFTQ	-0.5E-9	0.8E-10	-5.674
NEWHTSYS	-0.079	0.072	-1.101

Number of Observations = 40915
Number of Customers = 1169
R^2 =.532

Table 4-11

MODEL 9: Dependent Variable = DAYUSE

Variable	Parameter Estimate	Corrected Standard Error	T-Statistic
CONSTANT	0.144	0.056	2.599
PARTIC	0.103	0.035	2.960
GOTHEF	-0.074	0.032	-2.347
POST	-0.028	0.032	-0.867
V46T65	-0.133	0.056	-2.401
V66T77	-0.239	0.057	-4.179
VPOST77	-0.536	0.063	-8.500
INCOME	0.002	0.001	2.268
AHDD65	0.125	0.006	27.453
HHSIZE	0.024	0.067	0.360
DAYHOME	0.100	0.037	2.680
GASDRYR	0.074	0.033	2.221
GASPOOL	0.539	0.076	7.073
SBT	0.041	0.036	1.150
SQFT	0.3E-3	0.7E-4	4.879
SQFTQ	-0.4E-7	0.1E-7	-3.834
NEWHTSYS	-0.091	0.070	-1.294
GSASQ	-0.1E-4	0.3E-5	-3.869
GHSASQ	0.4E-4	0.3E-5	-11.527
PSA	-0.007	0.003	-2.339
GWHH	-0.040	0.065	-0.614

Number of Observations = 40915
Number of Customers = 1169
R^2 =.550

Table 4-12

MODEL 10: Dependent Variable = DAYUSE

Variable	Parameter Estimate	Corrected Standard Error	T-Statistic
CONSTANT	0.144	0.091	1.586
PARTIC	0.104	0.035	2.962
GOTHEF	-0.074	0.032	-2.344
POST	-0.028	0.032	-0.874
V46T65	-0.134	0.056	-2.411
V66T77	-0.241	0.058	-4.182
VPOST77	-0.538	0.064	-8.469
INCOME	0.002	0,001	2.279
AHDD65	0.125	0.005	27.464
HHSIZE	0.078	0.067	6.968
DAYHOME	0.099	0.011	2.665
GASDRYR	0.072	0.033	2.197
GASPOOL	0.542	0.077	7.065
SBT	0.042	0.036	1.179
SQFT	0.3E-3	0.7E-4	4.892
SQFTQ	-0.4E-7	0.1E-7	-3.842
NEWHTSYS	-0.090	0.070	-1.278
GSASQ	-0.1E-4	0.2E-5	-3.875
GHSASQ	0.4E-4	0.3E-5	11.519

Number of Observations = 40915
Number of Customers = 1169
R^2 =.510

STATISTICAL ISSUES

Correction Technique for First Order Autocorrelation

Based on the Durbin-Watson test, first order autocorrelation was found to be present in the data set used to estimate the equation specified above. For MODEL 7 the Durbin-Watson Statistic was .5391. Table 4-12 at the end of this section presents the Durbin-Watson Statistics for the ten models. The critical value at a 5% significance level for the lower limit (k=21 and n=200) is 1.55, indicating an autocorrelated error structure in the data. All model specifications yielded a numerically similar statistic. Although the parameter estimates continue to be unbiased in the presence of autocorrelation, the associated standard errors will be biased. Thus, any inference concerning the statistical significance of the coefficient is unreliable. In order to correct for the statistical problem connected with autocorrelation, the standard data transformation and estimation procedure was employed. The procedure used in the is outlined below.

The original equation was first estimated using Ordinary Lease Squares (OLS) on the time series cross sectional data set. The residuals from this estimated equation were then computed and an average autocorrelation coefficient(ρ) was computed (SAS does this in Proc Autoreg where the ρ is computed using the Yule-Walker method). The estimated ρ for MODEL 7 was -.731. All model specifications yielded a numerically similar statistic. These values are included in Table 4-13. Both the dependent and independent variables (including the constant term) were then transformed by the standard transformation ($Z_{it} - \rho Z_{it-1}$).

The transformed data set, having been purged of autocorrelation, was used to re-estimate the original equation by OLS and to test for heteroskedasticity.

Testing and Correction Technique for Heteroskedasticity

Evidence of some form of heteroskedasticity was tested for using a White Test. Under the White Test, a test statistic is computed using the R^2 from a regression of the squared errors from an estimated equation, on all the independent variables, their respective squares, and all cross terms. The actual test statistic is computed as (total number of observations) X (the R^2 referred to above) and is distributed as a Chi-Squared variable with degrees of freedom equal to the number of independent variables.

The computed value of the relevant χ^2 statistic under the null hypothesis of homoskedasticity was 1127.2081 for MODEL 7, which leads to the clear rejection of homoskedasticity in the transformed data. The high value of this test statistic was consistently reflected in all model specifications and is reported in Table 4-13.

In the presence of heteroskedasticity, OLS will yield unbiased parameter estimates. The standard errors however will be biased under OLS. The White Correction was used to correct the standard errors of the coefficients computed under OLS. The White correction factor is

$$S_0=(n)^{-1}\Sigma e_i^2 x_i x_i' \text{ and the estimated } Var(B)=n(X'X)^{-1}S_0(X'X)^{-1}$$

where n = total number of observations

The corrected standard errors are unbiased and consistent. They were used to compute the corrected t-statistics used for inferences concerning the statistical significance of the independent variables in the model. The t-statistics reported in Tables 4-3 through 4-12 were computed by dividing each coefficient estimate by its corresponding corrected standard error(square root of the coefficient's corrected variance).

Table 4-13

Statistical Summary of Estimated Models

	Number of Bills	R^2	Durbin-Watson	ρ	χ^2	df
MODEL 1	39200	0.543	0.540	-0.731	1307.915	276
MODEL 2	38885	0.538	0.538	-0.732	1027.629	190
MODEL 3	38885	0.538	0.538	-0.732	987.230	153
MODEL 4	40915	0.528	0.538	-0.731	955.384	120
MODEL 5	40915	0.542	0.602	-0.699	1281.187	153
MODEL 6	40915	0.532	0.540	-0.730	1142.106	171
MODEL 7	40915	0.529	0.539	-0.731	1127.208	153
MODEL 8	40915	0.532	0.539	-0.731	1178.372	171
MODEL 9	40915	0.550	0.538	-0.731	1386.933	231
MODEL 10	40915	0.510	0.538	-0.731	1355.389	210

Measurement Error

There was no reason to expect any abnormal measurement error in the data. Many of the variables used are binary and thus not prone to measurement error which stems from recording errors. The income and square foot measures are subject to measurement error but the direction of any potential biases are rather difficult to evaluate in a multivariate regression context. In general, income and square feet were recorded as interval values, thus lowering the probability of measurement error which stems from recording errors. Thus, there was no attempt to correct for measurement error in the analysis.

Collinearity

Generally, collinearity did not appear to be a problem in the data set used in the analysis, with the exception of GASHEAT, GASWHT and AGE65P. As mentioned previously, over 97% of all customers in the sample had gas space and water heating. Various alternative specifications were estimated with fairly consistent results concerning the important coefficient estimates. Most of the t-statistics (corrected) were highly significant and the estimated coefficients had their expected signs. This may be taken as evidence that collinearity did not affect the precision of the estimation procedure to any appreciable extent. AGE65P may have been collinear to some degree with INCOME and DAYHOME. It was therefore omitted from the final specification.

Item Non-Response and Missing Data

Most instances of missing sample data involved missing data on a variable for an entire time series. Thus attempting to replace the missing data by some regression based interpolation procedure was not feasible. Because of this, when a crucial variable's values were missing the entire time series representing a residence was deleted. Appendix 3D Table 3D-5 reports the attrition of the time series omitted from the analysis because of missing sample values as does Figure 3-2 .

The effect of deleting accounts appears to <u>lower</u> the estimate of net energy impacts. When a regression was estimated using a subset of variables for which there was no item nonresponse and almost the full set of data, the annual net impact was 50.28 therms. The estimated equation and relevant statistics are presented in Table 4-1 through Table 4-3 in Appendix 4. Because the estimate of the net impact was larger if the deleted and non-deleted accounts were used when compared to the estimate when only non-deleted accounts were used, the estimate of net energy impact computed under Model 7 may be biased downward, if in fact it was biased.

Treatment of Outliers

During the initial screening process of the data preparation stage of this evaluation, observations corresponding to billing data that was either implausible or extreme were omitted. A detailed discussion may be found in Chapter 2 of this report.

Estimates of Net and Gross Saving

The gross and net annual gas savings estimates based on the different models are presented in Tables 4-14 and 4-15. Table 4-15 reports the gross savings implied by the parameters in each model, along with the precision of these estimates at the 90% and 80% confidence levels. In the cases where the gross savings estimates depend on heating degree days (Model 1 and 10), the savings were calculated for the long term average heating degree days in the SoCal Gas service territory, weighted by the distribution of 1994 participants by climate zone. This distribution is shown in Appendix 3B, Table 3B-1. That weighted average (5.01) turns out to be very close to the mean for the sample used for model estimation. The confidence intervals are calculated based on the corrected standard errors reported in the regression models. The critical values for the t-statistic used to calculate the upper and lower bounds are 1.645 for the 90% confidence level and 1.282 for the 80% level.

Table 4-14

	Annual Gross Savings Estimates				
Model	Gross Annual Savings	90% Gross Lower Bound	90% Gross Upper Bound	80% Gross Lower Bound	80% Gross Upper Bound
1	72.15	53.35	90.94	57.50	86.79
2	57.10	41.39	72.82	44.85	69.35
3	57.11	41.53	72.68	44.97	69.25
4	59.18	43.83	74.52	47.22	71.13
5	60.97	46.88	75.06	49.99	71.95
6	87.39	63.24	111.54	68.57	106.21
7	87.60	63.37	111.83	68.72	106.48
8	82.41	56.22	108.59	62.00	102.81
9	78.02	47.00	109.04	53.84	102.20
10	78.10	47.07	109.14	53.92	102.29

Table 4-15

	Annual Net Savings Estimates				
Model	Net Annual Savings	90% Net Lower Bound	90% Net Upper Bound	80% Net Lower Bound	80% Net Upper Bound
6	43.84	25.94	61.73	29.89	57.78
7	44.14	26.21	62.08	30.17	58.12
8	52.16	31.68	72.64	36.20	68.12
9	54.39	32.12	76.66	37.03	71.75
10	54.39	32.12	76.66	37.04	71.75

The net annual savings estimates were calculated in a similar manner to the gross estimates. Only Models 6 through 10 include the variable POST which is meant to capture the "naturally occurring" conservation trend. As a result, these are the only models for which the savings estimates implied by the coefficient of GOTHEF should be interpreted as net savings.

In all cases, the estimates of annual savings are highly significant. The introduction of POST raises the estimate of the gross savings, as expected. The net savings, which are the key estimate of interest, rise significantly when they are expressed as a function of weather.

Calculation Methodology for Net and Gross Impacts

There were two broad classifications of model specifications that were estimated. These classifications are categorized by the manner in which program participation was modeled. In the first class of equations, program participation related variables were measured by PARTIC, GOTHEF, and POST. In the second category, in addition to PARTIC, GOTHEF, and POST, two interaction program participation variables were entered into the equation. These two variables were GSASQ (GOTHEF*AHDD65*SQFT) and PSA (POST*AHDD65). To simplify the exposition of the calculations associated with computation of the estimates for net and gross energy impacts, the two general specifications may be written as:

$$DAYUSE = B_0 + B_1*PARTIC + B_2*GOTHEF + B_3*POST + B_4*X \quad (4.10)$$

$$DAYUSE = C_0 + C_1*PARTIC + C_2*GOTHEF + C_3*POST + C_4*GSASQ + C_5*PSA + C_6*X \quad (4.20)$$

In the equations above the set of all non-program independent variables is represented by one independent variable, X.

In an equation of the form of (4.1) both gross and net impacts can be calculated. The estimate of the coefficient of GOTHEF measures decreases in energy consumption per day due to the program net of any decrease that naturally occurred over the program year and thus measures the daily net impact. The sum of the estimates of the coefficients of GOTHEF and POST measure the decreases in energy consumption per day due to the program in addition to any decrease that naturally occurred over the program year and measures the gross impact. This interpretation and illustrations of the computation of gross and net annual impacts that were calculated in this evaluation are presented below.

Derivation of Net Impact

The net impact per day is the change in DAYUSE after a participant received a HEF report. If a participant had not received a HEF report, GOTHEF=0, PARTIC=1, and $X=X_0$. Since the estimate of net impact (and gross impact) is for the program year, POST=1. Thus, the estimate of DAYUSE can be calculated by

$$DAYUSE^b = B_0 + B_1*1 + B_2*0 + B_3*1 + B_4*X_0 \quad (4.11)$$

In contrast, if this participant did receive a HEF report and all other factors remained the same, GOTHEF=1, POST=1, PARTIC=1, and $X=X_0$, then the estimate of DAYUSE would be computed as

$$DAYUSE^a = B_0 + B_1*1 + B_2*1 + B_3*1 + B_4*X_0 \quad (4.12)$$

The daily gross impact would then be calculated as $DAYUSE^b$ - $DAYUSE^a$ or

$$DAYUSE^b - DAYUSE^a = B_0 + B_1*1 + B_2*0 + B_3*1 + B_4*X_0 - (B_0 + B_1* + B_2*1 + B_3*0 + B_4*X_0) \quad (4.13)$$

$$= (B_0 - B_0) + (B_1*1 - B_1*1) + (B_2*0 - B_2*1) + (B_3*1 - B_3*1) + (B_4*X_0 - B_4*X_0)$$

$$= 0 + 0 - \mathbf{B_2} + 0 + 0$$

The annual estimated of the impact would be $-B_2$ times 365 or

$$\text{ANNUAL NET IMPACT} = -B_2*365. \quad (4.14)$$

For example, in MODEL 7 - B_2 =-(-.121) =.121 and the estimated annual net impact is 365*.120941= 44.14.

Net impacts can be estimated from the second type of equation (4.20) in an identical manner except that the coefficient of GSASF and PSA will now appear in the result of the derivation. The equation specification is now

$$DAYUSE = C_0 + C_1*PARTIC + C_2*GOTHEF + C_3*POST + C_4*GSASQ + C_5*PSA + C_6*X.$$

Using the average of heating degree days (5.035) and square feet (1545.64) for participants, PARTIC=1, GOTHEF=0, POST=1, GSASQ=0*5.035*1545.64=0, and PSA=1*5.035 equation 4.20 becomes

$$DAYUSE^b = C_0 + C_1*1 + C_2*0 + C_3*1 + C_4*0 + C_5*5.035 + C_6*X \quad (4.21)$$

If PARTIC=1, GOTHEF=1, POST=1, GSASQ=0*5.035*1545.64=7782.42, and PSA=1*5.035 equation 4.4 becomes

$$DAYUSE^a = C_0 + C_1*1 + C_2*1 + C_3*1 + C_4*7782.42 + C_5*5.035 + C_6*X \quad (4.22)$$

Again the daily gross impact would then be calculated as $DAYUSE^b$ - $DAYUSE^a$ or

$$DAYUSE^b - DAYUSE^a = C_0 + C_1*1 + C_2*0 + C_3*1 + C_4*0 + C_5*5.035 + C_6*X$$

$$-(C_0 + C_1*1 + C_2*1 + C_3*1 + C_4*7782.42 + C_5*5.035 + C_6*)$$

$$= (C_0 - C_0) + (C_1*1 - C_1*1) + (C_2*0 - C_2*1) + (C_3*1 - C_3*1) + (C_4*0 - C_4*7782.42) + (C_5*5.035 + C_5*5.035) + (C_6*X - C_6X)$$

$$= 0 + 0 - C_2 + 0 - C_4*7782.42 + 0 + 0$$

$$= -(C_2 + \mathbf{7782.42*C_4}) \quad (4.23)$$

The annual estimate of the impact under Model 9 would be - (C_2 + $7782.42*C_4$) times 365.Using the coefficient estimates from MODEL 9 the calculation is

$$\text{Annual NET Impact} = -(-.074 - .000009*7782.42)*365 = 54.39 \quad (4.24)$$

Derivation of Gross Impact

To calculate the gross impact the decrease in energy consumption, due to free ridership must be added to the net effect. The gross impact per day is the change in DAYUSE after a participant received a HEF report and instituted some conservation measures independently of the program. Using equation (4.10), if a participant had not yet received a HEF report and had not instituted any independent conservation measures, GOTHEF=0, POST=0, PARTIC=1, and $X=X_0$. Thus, the estimate of DAYUSE can be calculated by

$$DAYUSE^b = B_0 + B_1*1 + B_2*0 + B_3*0 + B_4*X_0 \quad (4.15)$$

In contrast, if this participant did receive a HEF report and did institute some conservation measures independently of the program and all other factors remained the same, GOTHEF=1, POST=1, PARTIC=1, and $X=X_0$, then the estimate of DAYUSE would be computed as

$$DAYUSE^a = B_0 + B_1*1 + B_2*1 + B_3*1 + B_4*X_0 \quad (4.16)$$

The daily gross impact would then be calculated as $DAYUSE^b$ - $DAYUSE^a$ or

$$DAYUSE^b - DAYUSE^a = B_0 + B_1*1 + B_2*0 + B_3*1 + B_4*X_0 - (B_0 + B_1*1 + B_2*1 + B_3*0 + B_4*X_0)$$

$$= (B_0 - B_0) + (B_1*1 - B_1*1) + (B_2*0 - B_2*1) + (B_3*0 - B_3*1) + (B_4*X_0 - B_4*X_0)$$

$$= 0 + 0 - B_2 - B_3 + 0$$

$$= -(B_2 + B_3) \quad (4.17)$$

The annual estimated of the gross impact would be - $(B_2 + B_3)$ times 365 or

$$ANNUAL\ GROSS\ IMPACT = -(B_2 + B_3)*365. \quad (4.18)$$

For example, in MODEL 7, $-(B_2 + B_3) = -(-.121 - .119) = .240$ and the estimated annual gross impact is 365*(.240)= 87.60.

Gross Impacts can be estimated from the second type of equation (4.20) in an identical manner except that the coefficient of GSASF and PSA will now appear in the result of the derivation. The equation specification is now

$$DAYUSE = C_0 + C_1*PARTIC + C_2*GOTHEF + C_3*POST + C_4*GSASQ + C_5*PSA + C_6*X.$$

Using the average of heating degree days (5.035) and square feet (1545.64) for participants, PARTIC=1, GOTHEF=0, POST=0 GSASQ=0*5.035*1545.64=0, and PSA=0*5.035=0 equation 4.20 becomes

$$DAYUSE^b = C_0 + C_1*1 + C_2*0 + C_3*0 + C_4*0 + C_5*0 + C_6*X \quad (4.25)$$

If PARTIC=1, GOTHEF=1, POST=1, GSASQ=0*5.035*1545.64=7782.42, and PSA=1*5.035 equation 4.4 becomes

$$DAYUSE^a = C_0 + C_1*1 + C_2*1 + C_3*1 + C_4*7782.42 + C_5*5.035 + C_6*X \quad (4.26)$$

Again the daily gross impact would then be calculated as $DAYUSE^b$ - $DAYUSE^a$ or

$$DAYUSE^b - DAYUSE^a = C_0 + C_1{*}1 + C_2{*}0 + C_3{*}0 + C_4{*}0 + C_5{*}0 + C_6{*}X$$

$$-(C_0 + C_1{*}1 + C_2{*}1 + C_3{*}1 + C_4{*}7782.42 + C_5{*}5.035 + C_6{*})$$

$$= (C_0 - C_0) + (C_1{*}1 - C_1{*}1) + (C_2{*}0 - C_2{*}1) + (C_3{*}0 - C_3{*}1) + (C_4{*}0 - C_4{*}7782.42) + (C_5{*}0 - C_5{*}5.035) + (C_6{*}X - C_6X)$$

$$= 0 + 0 - C_2 - C_3 + - C_4{*}7782.42 - C_5{*}5.035 + 0$$

$$= -(C_2 + C_3 + \mathbf{7782.42{*}C_4} + C_5{*}5.035) \quad (4.27)$$

The annual estimated of the impact would be $-(C_2 + C_3 + 7782.42{*}C_4 + C_5{*}5.035)$ times 365. Using the coefficient estimates from MODEL 9 the calculation is

$$\text{Gross Impact} = -(-.074 - .028 - .000009{*}\ 7782.42 - .007 * 5.03) * 365 = 78.02. \quad (4.28)$$

SELECTION OF "BEST" ESTIMATE OF NET PROGRAM IMPACT

For purposes of the claim for verified program savings, it is necessary to select one model and its implied estimate of net savings as the "best" or most reliable one. Our recommendation is that Model 7 be designated as that model. This model produces an estimate of gross and net savings of 88 and 44 therms, respectively. This recommendation is based on both statistical and plausibility considerations. These considerations are summarized here:

- **Overall Explanatory Performance of the Models.** All of the models are very comparable in terms of their overall power in explaining variations in natural gas consumption. While some may be superior in a strict statistical sense based on an F test of overall performance, this consideration alone is not sufficient for eliminating any one of the estimated models. The R-squared for Model 7 is .53, which is in the middle of the range of .51 to .55 obtained for all of the models.

- **Statistical Significance of Parameter Estimates.** Almost every model had one or more parameter whose estimate was not significant from a statistical standpoint. In general, this is not sufficient to reject any one specification. However, the models which had the most reliable estimates of the parameters directly affecting the estimate of net program impacts were generally considered superior. For Model 7, the estimates for these variables were all statistically significant. In contrast, the estimates for POST were statistically insignificant in Models 9 and 10. The estimate for GOTHEF was not significant in Model 8.

- **Plausibility of Parameter Estimates.** Some of the first models that were estimated were eliminated based on the plausibility of key parameter estimates. For example, the coefficient of the variable representing gas cooking is negative and statistically significant in Model 1. This is not consistent with the physical relation between the ownership of gas equipment and average gas use.

• **Plausibility of Implied Impact Estimates.** The implied estimates of net program savings range from 44 to 54 therms per year. This represents seven to nine percent reduction in annual consumption for participants. Among participants who said what measures they took in response to the recommendations, the most significant ones were replacements of water heaters and space heating systems. These types of measures could account for the savings in the order of magnitude implied by the models, but only if a large percent of the participants took them. The evaluation of the 1990 HEF Program obtained an estimate of net savings of 39 therms for water and space heating energy use. This represented an 8% reduction for the participant sample. Based on these considerations, we concluded that the estimate of savings at the lower end of the range is the most plausible. Model 7 produces the second lowest estimate of net savings among all of the models.

SUMMARY TABLE: Completed Load Impact Study (February, 1996)
Southern California Gas Company
(In fulfillment of Table 6 of the *Protocols and Procedures for the Verification of Costs, Benefits, and Shareholder Earnings from Demand-Side Management Programs*)

Study Title

First Year Load Impact Study of Southern California Gas Company's 1994 Home Energy Fitness Program

Study ID 708

Program/Program Year
Southern California Gas Company's 1994 Home Energy Fitness Program

Program Description
Chapter 1, page 1-1; Chapter 2, pages 2-1 to 2-2; Appendix 2A

SoCalGas' 1994 Home Energy Fitness Program promoted the adoption of energy efficient measures and actions by providing informational audits to residential customers.

1. Average Participant Group and Average Comparison Group Usage

TABLE S1: AVERAGE PARTICIPATION AND AVERAGE COMPARISON GROUP USAGE

AVERAGE DAILY GAS USAGE PER CUSTOMER	NON-PARTICIPANTS	PARTICIPANTS
Pre-Installation (1993) Usage	1.4157001	1.3381708
Impact year usage per designated unit.(1995)	1.567575	1.545855
% change in usage: 1993-1995	15.52%	10.73%

2. Average net and gross end use load impacts for the 1994 program year.

TABLE S2: AVERAGE NET AND GROSS END USE LOAD IMPACTS (ANNUAL TOTALS BASED ON 39,861 PARTICIPANTS)

IMPACT	ANNUAL IMPACT PER PARTICIPANT	IMPACT AS A % OF 1993 ANNUAL USAGE:PARTICIPANTS
Calculated With Point Estimates		
Gross	87.60	14.5
Net	44.14	7.31
Calculated at 90% upper limit		
Gross	111.83	18.51
Net	62.08	10.28
Calculated at 90% lower limit		
Gross	63.37	10.49
Net	26.21	4.34
Calculated at 80% upper limit		
Gross	106.48	17.63
Net	58.12	9.62
Calculated at 80% lower limit		
Gross	68.73	11.38
Net	30.17	4.99

3. Net to Gross Ratios

TABLE S3: NET-TO-GROSS RATIOS

	NET-TO-GROSS RATIO
Point Estimate	50.39%
Upper 90%	55.51%
Lower 90%	41.36%
Upper 80%	54.58%
Lower 80%	43.90%

4. Designated Unit Intermediate Data

Mean Values in the full survey data are shown in Table 3-1 of the report. Mean values in regression sample are shown in Table 4-1 of the report.

5. **Precision of Load Impact Estimates**
The precision of the load impact estimates at the 90% and 80% confidence levels are shown in Table S1.

6. **Measure Count Data**
Not applicable.

7. **Market Segment Data**

**TABLE S4: MARKET SEGMENT DATA
FREQUENCY BY PARTICIPATION STATUS AND
WEATHER ZONE AFTER MISSING VALUE
(no deletions due to item nonresponse on DAYHOME)**

NONPARTICIPANTS BY WEATHER ZONE							
WEATHER ZONE[1]	**1**	**2**	**3**	**4**	**5**	**6**	**Total**
Frequency	1,645	910	3,185	4,515	7,665	1,470	19,390
Percentage of Total	3.95	2.18	7.65	10.84	18.4	3.53	46.55
Percentage of Nonparticipants	8.48	4.69	16.43	23.29	39.5	7.58	100.00

PARTICIPANTS BY WEATHER ZONE							
WEATHER ZONE[1]	**1**	**2**	**3**	**4**	**5**	**6**	**Total**
Frequency	1,820	1,225	3,710	5,495	8,960	1,050	22,260
Percentage of Total	4.37	2.94	8.91	13.19	21.51	2.52	53.45
Percentage of Participants	8.18	5.5	16.67	24.69	40.25	4.72	100.00

Notes:
1. Weather Zones: 1-Mountain,2-Lower Desert, 3-Coastal Strip, 4-Upper Desert, 5-San Fernando/San Gabriel Valleys, 6-Los Angeles Basin

SUPPLEMENTAL SUMMARY INFORMATION:

(In fulfillment of Table 7 of the *Protocols and Procedures for the Verification of Costs, Benefits, and Shareholder Earnings from Demand-Side Management Programs*)

End Uses: *Chapter 4 page 4-2*
End uses covered:

> gas space heat, gas water heat, gas dryers gas ovens and stoves, gas pilot lights, gas pool water heaters. Usage was measured as aggregate usage of all gas operated equipment.

Methods and Models Uses: *Chapter 4*
The final model was a linear regression model. The dependent variable was average daily gas usage in therms per billing cycle. The independent variables were: a constant; three participation binary variables PARTIC, GOTHEF; and POST; three residence age binary variables V46t65, V66t77, Vpost77; household income INCOME; average heating degree days AHDD65; household size HHSIZE; a binary variable indicating the presence of household members at home during the day DAYHOME; five binary variables indicating the presence of a gas dyer, a gas heated pool, a set back thermometer, and a new heating system GASDRYR, GASPOOL, SBT, and NEWHTSYS; and, two interior square feet measures SQFT and SQFT*SQFT. All alternative specifications may be found in Chapter 4, pages 4-9 through 4.-18.

Participant Definition: *Chapter 2 pages 1-1 though 1-2 and Chapter 3*
Participants were defined as those customers who received a Home Energy Fitness Report. The comparison group was defined as non-participants or those customers who did not receive a Home Energy Fitness Report.

Analysis Sample Size: *Chapter 3 page 3-7 and Appendix 3-D*

TABLE S5: FREQUENCY OF CUSTOMERS AND BILLS BY PARTICIPANT STATUS (used in the regression analysis)

PARTICIPANT STATUS	FREQUENCY OF CUSTOMERS	FREQUENCY OF BILLS	PERCENT
No = 0	554	19,390	46.6
Yes = 1	636	22,260	53.4
Total	1,190	41,650	100.0

DATABASE MANAGEMENT

Relationships Between Data Elements and Data Sources: *Chapters 3 and 4, and Appendix 2-C*

The data sources were:

- The HEF Program Tracking Data Base;
- SoCal Gas Company's billing system;
- NOAA temperature readings at weather station in the SoCal Gas service territory; and
- a telephone survey of a sample 1994 HEF participants and non-participants.

Variable definitions(relationships) and sources are listed below.

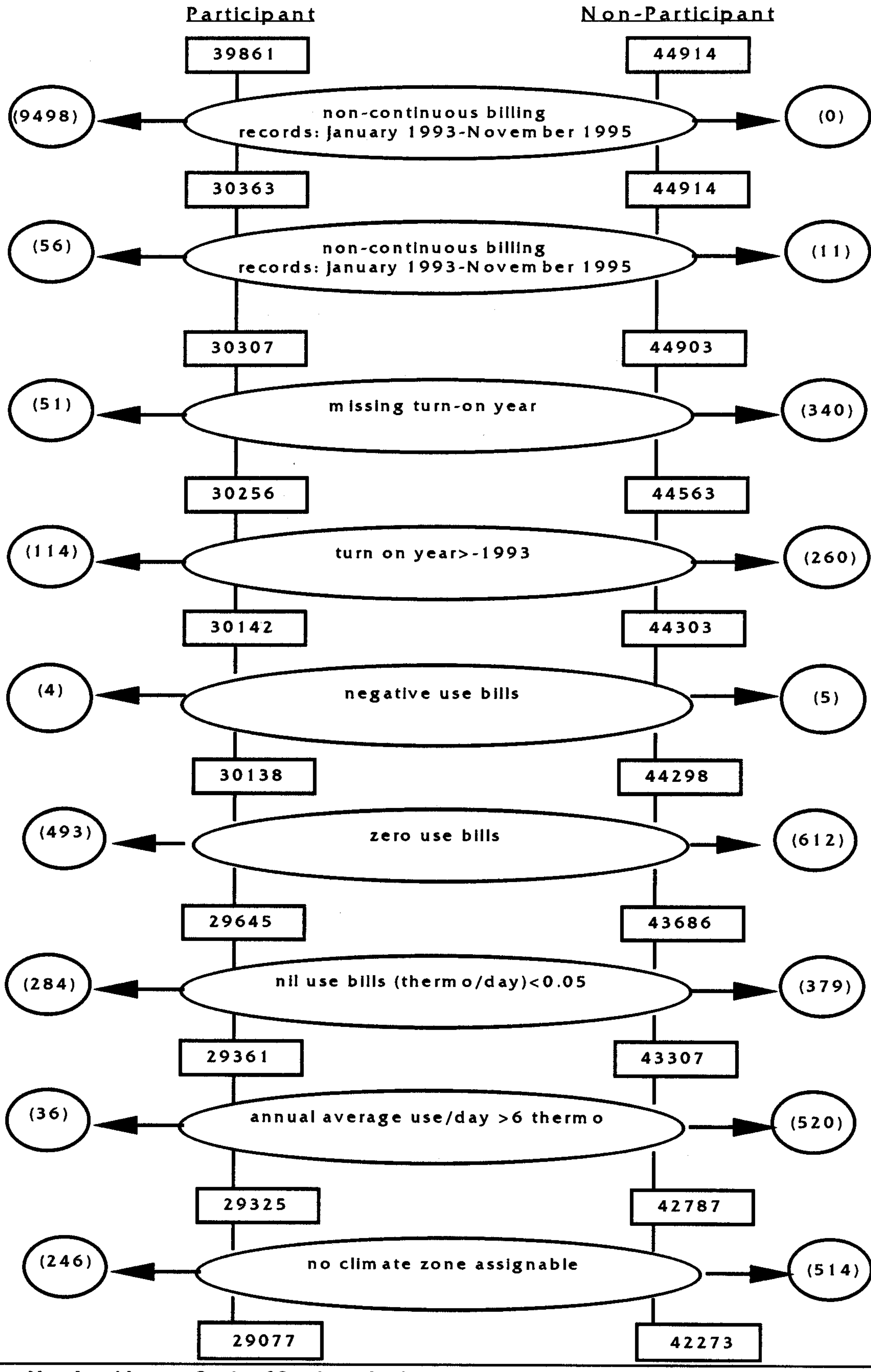
TABLE S6: VARIABLE DEFINITIONS AND SOURCES

VARIABLE LABEL	SOURCE	VARIABLE DEFINITIONS
		Independent Variables
DAYUSE	Billing Database	Daily Gas Usage (Gas use for billing cycle divided by # of days in billing cycle)
PARTIC	Program Database and Survey Database	Binary variable equal to 1 if the respondent was a program participant, 0 otherwise
GOTHEF	Program Database	Binary variable equal to 1 for participants after the date of HEF package was sent, 0 before and for non-participants in all months
POST		Binary variable equal to 1 for participants for last 12 months of the time period, 0 otherwise
SP	Program Database and Weather Database	AHDD65*PARTIC
SG	Program Database and Weather Database	AHDD65*GOTHEF
GSASQ	Program Database and Weather Database and Survey Database	AHDD65*GOTHEF*SQFT
GHSASQ	Program Database and Weather Database and Survey Database	AHDD65*GASHEAT*SQFT
PSA	Weather Database	AHDD65*POST
GWHH	Program Database and Survey Database	GASWHT*HHSIZE
V46T65	Survey Database	Binary variable equal to 1 if dwelling was built between 1946 and 1965, 0 otherwise
V66T77	Survey Database	Binary variable equal to 1 if dwelling was built between 1966 and 1977, 0 otherwise
VPOST77	Survey Database	Binary variable equal to 1 if dwelling was built after 1977, 0 otherwise
INCOME	Survey Database	Annual income in $1000's
AGE65P	Survey Database	number or residents of each premise who are 65 years or older
AHDD65	Whether Database	Average daily heating degree days in billing cycle
AHDD65Q	Survey database and Weather Database	AHDD65 SQUARED
HHSIZE	Survey Database	Number of people in the household
DAYHOME	Survey Database	Binary variable equal to 1 if some residents are normally at home during the day, 0 otherwise
GASDRYR	Survey Database	Binary variable equal to 1 if a gas clothes dryer is used, 0 otherwise

VARIABLE LABEL	SOURCE	VARIABLE DEFINITIONS
GASWHT	Survey Database	Binary variable equal to 1 if a gas hot water heater is used, 0 otherwise
GASHEAT	Survey Database	Binary variable equal to 1 if gas is primary heating fuel, 0 otherwise
PILOT	Survey Database	Binary variable equal to 1 if stove has pilot light, 0 otherwise
GASCOOK	Survey Database	Binary variable equal to 1 if gas is primary cooking fuel, 0 otherwise
GASPOOL	Survey Database	Binary variable equal to 1 if gas is used to heat a pool, 0 otherwise
SBT	Survey Database	Binary variable equal to 1 if a set back thermostat is used, 0 otherwise
NEWHTSYS	Survey Database	Binary variable equal to 1 if new space heating equipment was installed, 0 otherwise
PM_WH	Survey Database	Binary variable equal to 1 if a recommended source heating related measure was installed, 0 otherwise
GOODINS	Survey Database	Binary variable equal to 1 if premise reported "good insulation", 0 otherwise
SQFT	Survey Database	Area in square feet of each premise
SQFTQ		SQFT SQUARED

Data Attrition: Chapter 3 and Appendices 3A, 3B, 3C, and 3D.
The following table describe the data attrition process for the billing data and for the billing database.

Attrition of Cases in Inital Sample by Screening Criterion



D

Data Quality Checks: *Chapter 3*
The quality checks are reported in the discussion of attrition above. The merging of the billing, survey, and weather data was done by matching records by their premise identification number.

Data Collected But Not Used: *Chapter 3 and Appendix 3D page 3Da-1 through page 3Da-34*
Even though the survey included multiple questions addressing free-ridership it is apparent from the frequency calculations done for these variables that in the majority of cases these variables corresponding to the survey questions associated with free-ridership contained no information. Thus, these data were not used in the analysis because in a preponderance of cases the information did not exist.

SAMPLING

Sampling Procedures and Protocols: *Chapter 3 and its Appendices*
The data met the requirements of the Verification Protocols concerning participant and non-participant sample sizes and coverage. The process of data acquisition and preparation is fully documented in this report and its appendices.

Survey Information: *Chapter 3 and Appendices 2B, 3Da, and 3Db*
The survey instrument is provided in a appendix along with the sample disposition.

Statistical Descriptions of Variables: *Chapter 4*
There is a list of variable definition and their mean values in Chapter 4 (Table 4-1).

DATA SCREENING AND ANALYSIS

Outliers, Missing data and Weather Adjustment: *Chapter 3 and its Appendices*
Outliers were screened for in the sampling phase. Customers with missing data on crucial variables were deleted from the regression database. Variables were not weather adjusted.

Effects of Background Variables: *Chapter 4*
Background variables were explicitly incorporated into the analysis by including them in the regression model specification. A binary variable which was 1 for the last twelve months of the sample period (o otherwise) was included to measure the decrease in consumption that would have taken place without the program.

Attrition of Final Data set: Chapter 3

Participants	Non-Participants	Total
992	1010	2002

Accounts Deleted due to Item Non-Response: Income

Participants	Non-Participants	Total
(204)	(268)	(472)

Participants	Non-Participants	Total
788	742	1530

Accounts Deleted due to Item Non-Response: SQFT

Participants	Non-Participants	Total
(111)	(137)	248

Participants	Non-Participants	Total
677	605	1282

Accounts Deleted due to Item Non-Response: HHSIZE

Participants	Non-Participants	Total
(1)	(1)	(2)

Participants	Non-Participants	Total
676	604	1280

Accounts Deleted due to Item Non-Response: VPRE46

Participants	Non-Participants	Total
(14)	(22)	(36)

Participants	Non-Participants	Total
662	582	1244

Accounts Deleted due to Item Non-Response: SBT

Participants	Non-Participants	Total
(26)	(28)	54

Participants	Non-Participants	Total
636	554	1190

Accounts Deleted due to Item Non-Response: DAYHOME

Participants	Non-Participants	Total
(11)	(9)	(20)

Participants	Non-Participants	Total
625	545	1170

Regression Statistics: *Chapter 4 pages 4-9 through 4-18 and 4-20*
Regression statistics for the final model and all alternative specifications are included in Chapter 4.

Specification: *Chapter 4*
Initial Model *-page 4-3, 4-9*
Final Model *-page 4-4, 4-15*
Alternative Models *-Chapter 4*
Other Issues Addressed
Heterogeneity of customers by including customer specific variables, such as income and square feet as independent variables.
A binary variable was included to recognize and estimate changes that effected gas consumption during the last twelve months of the sample period. Its coefficient estimated free-ridership effects.
The calculation of net impact measurement *-pages 4-23 through 4-27*
Self-selection issues were not considered relevant in the context of this report.

Measurement Error: *Chapter 4, page 4-20*
Errors in measuring variables was not considered a problem.

Autocorrelation: *Chapter 4, page 4-19*
Autocorrelation was tested for and found to be present. The standard correction procedure was the applied to the data.

Heteroskedasticity: *Chapter 4, pages 4-19 and 4-20*
Heteroskedasticity was tested for, found to be present. The standard errors were corrected using the White procedure.

Collinearity: *Chapter 4, pages 4-21*
Collinearity was not a serious problem.

Influential Data Points
This was not a problem due the large number of observations in the data set and thus the small weight given to each observation by the OLS procedure

Missing Data
Discussed above.

Precision: *Chapter 4*
The standard errors for the estimates of gross and net impacts were calculated by adding the corrected variances of the relevant coefficients and then taking the square root to get the standard error.

DATA INTERPRETATION AND APPLICATION
The calculations of net and gross impacts may be found in chapter 4.