Final Report for the Evaluation of the California 2003 Home Energy Efficiency Survey Program and the Southern California Edison Local In-Home Energy Audit Program

Submitted to

Southern California Edison Company
Pacific Gas and Electric Company
San Diego Gas and Electric Company
Southern California Gas Company

Submitted by

Ridge & Associates
3022 Thompson Ave.
Alameda, CA 94501

In association with

KVD Research Consulting
and
StatWizards

May 16, 2006

Table of Contents

Tables

Figures

Appendices

1 Executive Summary

In this section, we describe the Statewide Home Energy Efficiency Survey (HEES) Program and the Southern California Edison Local In-Home Audit Program as they were implemented during program year (PY) 2003, covering such topics as the statewide Mail-In and Online Audits, program goals, customer eligibility, hard-to-reach customers, market barriers, languages addressed, funds expended, program outreach, and the definition of the program year.

1.1 Program Description

During PY 2003, the Pacific Gas and Electric Company (PG&E), Southern California Edison Company (SCE), Southern California Gas Company (SoCalGas), and San Diego Gas & Electric Company (SDG&E) offered the Home Energy Efficiency Survey (HEES) Program. The HEES Program is available statewide through both direct mail and the Internet and provides residential customers with valuable information to assist them with understanding, controlling and reducing energy use in their homes. In addition, SCE also offered its Local In-Home Audit Program. The primary market barriers addressed by the HEES Program are lack of consumer information and lack of high-efficiency products. The components of the HEES Program are briefly described in the following sections.

1.1.1 Mail-In Audit

Mail-In surveys were available in English and Spanish in the service territories of all four utilities; in Chinese in the PG&E, SoCalGas, and SCE service territories, and in Vietnamese in the SDG&E service territory. Mail-In surveys were distributed to customers via direct mail marketing efforts while other customers directly approached SCE to request an in-home audit. Participating customers were given a survey and materials explaining the value of the program. Once completed, the surveys were mailed back to the statewide mail-in survey vendor for processing. Completed surveys were then analyzed against the customer's actual energy usage, and a report representing actual energy usage in graph form was mailed to the customer. Reports included information on energy efficiency products and services, rebate programs, and other energy-related information to encourage adoption of energy efficiency measures identified through the energy survey.

1.1.2 Online Audit

For participants in the online version of the HEES, a simple log-on procedure allowed consumers to access the energy survey. Consumers input specific data regarding their energy use and received immediate results through an online report that provides an explanation of where energy dollars are spent. This easy-to-use tool provides customers with immediate short- and long-term changes they can make to become more energy efficient. The online surveys were available in English and Spanish in the service territories of all four utilities; offered in Chinese in PG&E, SoCalGas, and SCE service territories, and in Vietnamese in the SDG&E service territory. However, the online audits were interactive only in English. For other languages, web-posted versions (in PDF format) could be downloaded from the utility websites, completed, and returned to the utility for processing.

1.1.3 In-Home Audit Program

The In-Home Energy Survey Program is one of Southern California Edison Company's (SCE's) oldest energy efficiency programs. It has taken a variety of forms over the years, but the current design of the Program has been in use for the last decade. This current design is based on a two-page form that an energy auditor completes as s/he walks through the home and questions the customer about appliance usage patterns. A copy of the completed form and a list of recommendations are left with the customer. . Ideally, the auditor also discusses the recommendations with the customer and answers questions. When the audit is completed, the audit date and other basic information are recorded in a database and submitted to SCE.

While this program is called an in-home program, and most of the energy surveys are completed with an auditor who comes to the customer's home, participants have the option of doing the audit over the telephone. Thus, a small percentage of audits were conducted by telephone.

1.2 Program Period

For the most part, the program period for the Mail-In, Online, and In-Home Audits was defined as January 1, 2003 through December 31, 2003. However, PG&E defined its program period for achieving the Mail-In Audit goal as March 1, 2003 through December 31, 2003, while SCE defined theirs as January 1, 2003 through February 29, 2004. SCE explained that some of the participants in the first quarter of 2004 were responding to marketing efforts conducted in 2003.

1.3 Evaluation Goals

The evaluation of the PY 2003 Home Energy Efficiency Survey (HEES) Program and SCE's In-Home Audit Program has four primary objectives:

1. To verify whether the goals for completed HEES Mail-In and Online Audits and SCE's In-Home Audits were achieved.
2. To verify whether the hard-to-reach (HTR) outreach goals for the statewide HEES Mail-In Audit and SCE's In-Home Audit Programs were achieved.[1]
3. To improve the target marketing of the HEES Mail-In Audits and SCE's In-Home Audits.
4. To determine customer interest in other types of audit configurations and how this interest varies across customer groups.

1.4 Methods

1.4.1 Verification of Audit Participation Goals

To achieve the first objective, the program tracking databases for each utility were reviewed and the number of participants within the program period, as defined by each utility, was recorded. This was then compared to the goals established for each utility and to the completed audits claimed in their respective fourth quarter reports.

[1] For the Mail-In Audit, the goal was to send at least 50 percent of the direct mailers to HTR zip codes. For SCE's local In-Home Audit Program, the goal that at least 50 percent of those who completed the audit must be in HTR zip codes.

1.4.2 Verification of HTR Goals

To achieve the second objective, we analyzed those customers who received the direct mail solicitation and those customers who completed the SCE Local In-Home Audit with respect to the ZIP Codes defined by the California Public Utilities Commission (CPUC) as hard-to-reach (HTR).

1.4.3 Improvement of Target Marketing

The achievement of the third objective, involved an analysis of 510,228 residential households that received a mailed invitation to participate in the audit. The analysis, which used both logistic regression and classification and regression trees (CART), attempted to discriminate between those households that chose to participate in the Mail-In Audit and those that chose not to participate. In these models, the 2000 block-level U.S. Census data and weather zones were assigned to each of the 510,228 residential households. The models developed were then used to predict participation in the Mail-In Audit. Once these models were estimated, we illustrated how they could be used to predict which types of customers are most likely to participate in the future so as to better target recruitment efforts.

1.4.4 Assessment of Customer Interest in Other Audit Configurations

The fourth objective was addressed using data collected via an Internet survey of a random sample of 793 residential households in the service territories of the four utilities. These data were analyzed using the latent-class discrete choice approach. The objective of this analysis was to determine the level of interest in various types of home audits, each with a varying set of audit attributes, such as the length of time needed to complete the audit, level of detail in the audit report, and the availability of post-audit technical support. Customers fell into 8 segments based on their interests in certain audit configurations. As a part of this analysis, we also determined the characteristics of these customer segments so that utilities would be able to target market the various audit configuration. A simulation model was also developed so that program managers could test a large number of audit configurations and observe the estimated market shares so associated with each configuration.

1.5 Conclusions and Recommendations

1.5.1 Verification of Audit Participation and HTR Goals

First, all four utilities exceeded their goals for the *Mail-In Audit*. Table 1-1 presents these results.

Table 1-1
Verified Completed Mail-In Audits, by Utility
Fourth Quarter Report, by Target

Utility	Targets	Claimed	Verified	Verified as Percentage of Claimed	Verified as Percentage of Goal
PG&E	29,000	43,245	42,465	98%	146%
SCE	18,000	26,515	25,917	98%	144%
SCG	6,000	7,694	9,222	120%	154%
SDG&E	8,000	7,824	8,066	103%	101%
Total	61,000	85,278	85,670	100%	140%

Table 1-2 present the results for the Online Audit. PG&E and SCE exceeded their goals by 24 percent and 38 percent, respectively[2].

Table 1-2
Verified Completed Online Audits, by Utility Fourth Quarter Report, by Target

Utility	Goal	4th Quarter Report	Verified
PG&E	12,000	14,848	14,848
SCE	12,000	16,513	15,676

In addition, SCE had a target of 4,500 In-Home audits for 2003. In its 4th quarter report, it reported that it had completed 5,362. We were able to verify 99.7 percent, or 5,348, which is 18.8 percent greater than their original goal.

Finally, all four utilities exceeded their HTR goals for the mail-in audit and SCE exceeded its HTR goal for its local In-Home Audit Program.

1.5.2 Improvement of Target Marketing

The first analysis involved the 510,228 residential households that received a mailed invitation to participate in the audit. Because the discrete-choice analysis did a poor job of predicting participation, we attempted another approach involving ordinary least squares (OLS) regression. One of the more important findings of our OLS regression analysis was that the assumption that the higher the annual energy consumption the greater the likelihood a household will participate in the audit was not supported. The relationship is much more complex. To the extent that marketing of the Mail-In Audit by the utilities is based on this assumption, participation rates may be lower than they could be. However, while the OLS regression performed well, it was outperformed by the CART model, which not only did a good job of predicting participation for each utility but also produced results that were more *actionable* on the part of the utilities. Table 1-3 presents the prediction rates for each utility for the CART model.

[2] SoCal Gas and SDG&E did not have goals for their Online Audit.

Table 1-3
CART Prediction Rates, by Utility

Utility	Percentage of Participants Correctly
PG&E	67%
SCE	75%
SoCalGas	67%
SDG&E	74%

Next, we illustrated how these estimated CART models could be used to predict future participation in order to better target recruitment efforts. The estimated CART models, for PG&E, SCE, and SDG&E[3], were used to predict participation for a random sample of residential households for each of the three utilities. Finally, so that the utilities could more easily use these estimated models in the future, we translated them into SAS (Statistical Analysis System) code.

To empirically field-test the effectiveness of these CART models, each utility could select a random sample of 10,000 residential customers and attach block-level demographic variables[4] to each customer. Then, using the CART model (translated into SAS[5] code), each customer would be assigned a predicted probability of participating in the HEES Program. Only those customers with the higher probabilities of participating would receive the invitation to participate. One could compare the rates of participation between this group and the larger group who were targeted using the traditional approach. We note that, if the new marketing strategy is demonstrated to be superior to the traditional approach, utilities might not always reach out to customers with the highest predicted probabilities of participation. This is due to the fact that, while HTR goals are not required for the 2006-2008 funding cycle, utilities remain concerned with meeting the needs of the HTR customers.

1.5.3 Assessment of Customer Interest in Other Audit Configurations

The second analysis was based on the data collected from 793 households via the Internet. Each respondent was presented with various types of audits, which varied in terms of key attributes like the time to complete the audit and the availability of post-audit technical support. Based on their choices, we identified eight distinct customer segments and the audit characteristics most strongly associated with membership in each segment. Based on further analysis, customer characteristics of each segment were determined. Table 1-4 presents the segment names and the percentage of the 793 respondents within each segment or class.

[3] SoCalGas was unable to provide a random sample of their residential households.
[4] SCE could also attach variables that it had purchased from Acxiom. Other utilities could purchase the same Acxiom data if they wished. Axciom data is described more fully in Section 3.3.1.1.6.
[5] Statistical Analysis System.

Table 1-4
Segment Characterizations and Membership Proportions

Class #	Segment Characterization	Proportion
Class 1	Demanding but willing to pay	28.6%
Class 2	Meticulous	23.9%
Class 3	Subsidy required	10.4%
Class 4	Enthusiasts	9.1%
Class 5	Not interested	8.8%
Class 6	Personal attention	7.8%
Class 7	Hard to convince	6.4%
Class 8	Fast and thorough	5.0%

Based on the latent-class discrete choice analysis, a simulation model was developed that is applicable to any of the four utilities. This simulation model operates within an Excel framework and can be used to estimate expected participation resulting from a wide variety of audit configurations. Such simulations could assist HEES program managers in exploring new audit configurations that the different customer segments might find interesting.

Based on this analysis, a series of experiments could be conducted by the utilities to explore those audit configurations that, based on the results of the simulations, appear to produce the largest increases in market share. These experiments would involve modifying the traditional audits (Mail-In and Online Audits and SCE's In-Home Audit) and marketing them to those most likely to participate in these new types of audits. For example, some of the most popular audit features identified in the stated preference analysis were the availability of post-audit technical support and the provision of links to national and local vendors. For a random sample of 3,000 of those who receive the traditional mailer in a given mailing effort, one could insert an offer to provide, for example, technical support via a toll-free hot line after the audit results are delivered to the customer. Several expert residential auditors could be assigned to staff the hot line for a period of two months after the results are mailed to participants. The adoption rates of those who received the traditional mailer could be compared to those who received the offer of technical support. An analysis could be conducted to determine whether any benefits, in the form of higher adoption rates, exceed the additional costs of providing technical support.

1.5.4 Targeting the Hard-to-Reach

One of the recommendations made as part of the evaluation of the PY 2002 HEES (Ridge et al., 2004) was that targeting of HTR population should be done using block-level U.S. Census data rather than zip code-level[6]. A zip code is comprised of smaller units called tracts, block groups and blocks, zip codes manifest greater variation with respect to demographic characteristics than

[6] Zip codes were established by the United States Postal Service for efficient mail delivery. Because zip code boundaries follow the routes of mail carriers, they do not conform to boundaries of Governmental Units or to those of the Bureau's Statistical Units. In fact, zip code areas usually do not have clearly identifiable boundaries. They change periodically to meet postal requirements and they do not cover the total land area of the United States.

do the smaller units such as tracts and blocks. This makes targeting by zip code areas less refined than targeting by tracts, block groups, or blocks. Thus, we recommend that targeting the HTR population should be done based on demographic data at the Census block level. We make this recommendation even though HTR goals have not been established for the 2006-08 period, since utilities will likely continue to be concerned about serving the HTR population.

2 Introduction

In this section, we describe the Statewide Home Energy Efficiency Services (HEES) Program and the Southern California Edison Local In-Home Audit Program as they were implemented during program year (PY) 2003, covering such topics as the statewide Mail-In and Online Audits, program goals, customer eligibility, hard-to-reach customers, market barriers, languages addressed, funds expended, program outreach, and the definition of the program year.

2.1 Program Description

During PY 2003, Pacific Gas and Electric Company (PG&E), Southern California Edison Company (SCE), Southern California Gas Company (SoCalGas), and San Diego Gas & Electric Company (SDG&E) offered the Home Energy Efficiency Survey (HEES) Program[7]. The HEES Program is available statewide through both direct mail and the Internet and provides residential customers with valuable information to assist them with understanding, controlling and reducing energy use in their homes. In addition, SCE also offered its Local In-Home Audit Program. The primary market barriers addressed by the HEES Program are lack of consumer information and lack of high-efficiency products. The components of the HEES Program are briefly described in the following sections.

2.1.1 Mail-In Audit

Mail-In surveys were available in English and Spanish in service territories of all four utilities; in Chinese in PG&E, SoCalGas, and SCE service territories, and in Vietnamese in SDG&E's service territory. Mail-In surveys were distributed to consumers via direct mail marketing efforts, outreach channels such as county and regional fairs and other major events using the Energy Efficiency Mobile Education Unit, and through customer requests. Participating customers were given a survey and materials explaining the value of the program. Once completed, the surveys were mailed back to the statewide mail-in survey vendor for processing. Completed surveys were then analyzed against the customer's actual energy usage, and a report representing actual energy usage in graph form was mailed to the customer. Reports include information on energy efficiency products and services, rebate programs, and other energy-related information to encourage adoption of energy efficiency measures identified through the energy survey. For the Mail-In version of the HEES, consumers were selected from a database and sent a solicitation package. Customers who needed assistance with the survey or had additional questions could telephone the statewide vendor or their utility and have a trained energy specialist walk them through the HEES process. A goal was also established that utilities had to send at least 50 percent of the direct mail solicitation to households in HTR zip codes.

2.1.2 Online Audit

For participants in the online version of the HEES, a simple log-on procedure allowed consumers to access the energy survey. Consumers input specific data regarding their energy use and received immediate results through an online report that provides an explanation of where energy dollars are spent. This easy-to-use tool provides customers with immediate short- and long-term changes they can make to become more energy efficient. The online surveys were available in English and

[7] See Figure I-1 in Appendix G for a map depicting the service territories of these four utilities.

Spanish in service territories of all four utilities; in Chinese in the PG&E, SoCalGas, and SCE service territories, and in Vietnamese in the SDG&E service territory. However, the online audits were interactive only in English. For other languages, web-posted versions (in PDF format) could be downloaded from the utility websites, completed, and returned to the utility for processing.

2.1.3 In-Home Audit Program

The In-Home Energy Survey Program is one of Southern California Edison Company's (SCE's) oldest energy efficiency programs. It has taken a variety of forms over the years, but the current design of the Program has been in use for the last decade. This current design is based on a two-page form that an energy auditor completes as s/he walks through the home and questions the customer about appliance usage patterns. The completed form, which is left with the customer, provides a list of recommendations. Ideally, the auditor also discusses the recommendations with the customer and answers questions. When the audit is completed, the audit date and other basic information are recorded in a database and submitted to SCE.

While this program is called an in-home program, and most of the energy surveys are completed with an auditor who comes to the customer's home, participants have the option of doing the audit over the telephone. Thus, a small percentage of audits were conducted by telephone. A goal was also established that at least 50 percent of the households that *completed* the audit had to be in HTR zip codes.

2.2 Audit Participation Goals

Table 2-1 presents the targets for each utility for completed HEES audits (Mail-In and Online) and SCE's Local In-Home Audits. Note that neither SDG&E nor SoCalGas established goals for the Online Audit.

Table 2-1
Mail-in Audit Goals for PY 2003, by Utility

Utility	Mail-In Targets	Online Targets	SCE In-Home Target
PG&E	29,000	12,000	
SCE	18,000	12,000	4,500
SoCalGas	6,000	N/A	
SDG&E	8,000	N/A	
Total	61,000	24,000	

2.3 Hard-To-Reach Goals

During PY 2003, for the HEES Mail-In Audit and the SCE Local In-Home Audit, a special effort was made to contact hard-to-reach (HTR) customers. The utilities used the HTR definition developed by the California Public Utilities Commission (CPUC). This definition was based on the following five attributes:

1. Language: Primary spoken language is other than English

2. Income: Those customers who fall into the moderate income level
3. Housing Type: Multi-family and mobile home tenants
4. Geography: Residents of areas other than the San Francisco Bay Area, San Diego area, Los Angeles Basin, or Sacramento
5. Tenure: Renters

While the definitions of renter and spoken language other than English are fairly obvious, below we provide the specific definition "moderate income" used by the utilities that are perhaps less obvious.

Table 2-2 presents the operational definition of moderate income.

Table 2-2
Operational Definition of Moderate Income

	Moderate-Income	
Size of Family Unit	Lower Limit	Upper Limit
1	$ 12,525	$ 33,400
2	$ 16,875	$ 45,000
3	$ 21,225	$ 56,600
4	$ 25,575	$ 68,200
5	$ 29,925	$ 79,800
6	$ 34,275	$ 91,400
7	$ 38,625	$ 103,000
8	$ 42,975	$ 114,600
9	$ 47,325	$ 126,200
10	$ 51,675	$ 137,800
11	$ 56,025	$ 149,400
12	$ 60,375	$ 161,000
13	$ 64,725	$ 172,600

In Decision 02-03-056, the CPUC required that 50 percent of the *Mail-In Audit* targets be sent to HTR customers. Per the draft decision, a target was established that at least 50 percent of the utility direct mail solicitations would be sent to HTR customers, as defined by the CPUC. The total number of solicitations mailed by each utility in PY 2002 is presented in Table 2-3.

Table 2-3
Direct Mail Solicitations in PY 2002, by Utility

Utility	Direct Mail Solicitations
PG&E	225,226
SCE	200,000
SoCalGas	50,000
SDG&E	35,002
Total	510,228

For SCE's In-Home Audit, the HTR goal was framed in terms of *completed* audits rather than in terms of mailings. That is, at least 50 percent of the completed audits had to be in HTR zip codes. In their respective 4th Quarter Reports, all four utilities claimed to have met their HTR direct mail solicitation targets for the Mail-In Audit and SCE claimed that they had met their HTR goal for its In-Home Audit. To examine these claims, we conducted an analysis of the utility mailings for the Mail-In Audit and examined SCE's program tracking database for its In-Home Audit. The results are presented in Section 4.1.

2.4 Program Expenditures

The final PY 2003 expenditures for the HEES Program are presented in Table 2-4 for each utility.

Table 2-4
PY 2003 Expenditures for the HEES Program, by Utility

Utility	Mail-in Audit	Online Audit	In-Home Audit	Total
PG&E*	$719,538.85	$32,511.75		$1,044,650.60
SCE	$760,574.11	$388,924.53	$464,057.14	$1,613,555.78
SDG&E	$202,779.00	$86,906.00		$289,685.00
SoCalGas	$221,873.00	$95,089.00		$316,962.00

* The Mail-In Audit expenditures include costs paid to Kema-Xenergy, the contractor responsible for administering the Mail-In audit, but exclude internal implementation costs. Online Audit expenditures include costs paid to Nexus for implementing the PG&E online audit, but exclude internal implementation costs.

SCE spent the largest amount; SDG&E spent the least. In all, the four utilities combined to spend $3,264,853 in PY 2003.

2.5 Program Period

For the most part, the utilities defined their program year as January 1, 2003 through December 31, 2003. However, SCE defined its program period for achieving their Mail-in Audit goals as January 1, 2003 through February 29, 2004. For the In-Home Audit Program, SCE defined the program period as January 1, 2003 through March 31, 2004. SCE's argument for including participants in the first quarter of 2004 is that they were responding to SCE's marketing campaigns conducted in 2003. In order for a customer who completed the Mail-In or In-Home Audit during the first quarter of 2004 to be counted as a participant in SCE's PY 2003 Mail-In Audit or In-Home Audit Program, they had to have been solicited or sent the Mail-In survey sometime during 2003. Finally, PG&E defined its program period for achieving Mail-in Audit goals as March 1, 2003 through December 31, 2003. Table 2-5 presents these program period definitions.

Table 2-5
Program Period Definitions,
by Utility, by Audit Type

Utility	Mail-in Audit		Online Audit		In-Home Audit	
	Begin Date	End Date	Begin Date	End Date	Begin Date	End Date
PG&E	03/01/03	12/31/2003	1/1/2003	12/31/2003		
SCE	01/01/03	02/29/04*	1/1/2003	12/31/2003	1/1/2003	03/31/04*
SDG&E	01/01/03	12/31/2003	1/1/2003	12/31/2003		
SoCalGas	01/01/03	12/31/2003	1/1/2003	12/31/2003		

* Mail-In and In-Home Audits solicited in 2003 and processed/completed in 2004 were counted toward 2003 completed surveys.

2.6 Evaluation Goals

The evaluation of the PY 2002 HEES was completed just before this evaluation began. The HEES Advisory Group comprised of EM&V representatives from the four utilities felt that little new information could be gleaned from another similar study. Instead, they decided that the following four evaluation goals should be addressed:

1. To verify whether the goals for completed Mail-In, Online, and SCE In-Home Audits were achieved.
2. To verify whether the Mail-In and SCE In-Home Audit outreach goals for the HTR population were achieved.
3. To improve the target marketing of the HEES Mail-In and SCE In-Home Audits.
4. To determine customer interest in other types of audit configurations and how this interest varies across customer groups.

3 Methods

This section discusses the methods used to verify the audit goals and the HTR goals for each utility. In addition, this section discusses the methods for improving the target marketing of the mail-in audit for all four utilities and for assessing customer interest in a wide variety of audit configurations.

3.1 Verification of Audit Participation Goals

3.1.1 Data Collection

As one of the first tasks in this evaluation, we requested the PY 2003 program-tracking databases for the following four programs:

1. Statewide Mail-In Audit
2. PG&E Online Audit
3. SCE Online Audit
4. SCE Local In-Home Audit

We did not request the program tracking databases for SoCalGas or the SDG&E Online Audits since neither utility had established a goal for the Online Audit.

Each customer record in the statewide database for the Mail-In Audit and the databases for SCE's Online and Local In-Home Audits, contained the customer account number, date of audit, name, address, and zip code. The databases for the short version of the SCE Online Audit did not have an account number or a zip. The PG&E online audit database, maintained by Nexus, contains the date of completion, user ID, city and zip code.

3.1.2 Analysis

The verification process involved simply counting the number of records with a date on which the audit report was mailed that fell within the date range that defined the program period for PY 2003 (see Table 2-5). These counts were then compared to the goals established for each utility and to the completed audits claimed in their respective fourth quarter reports.

3.2 Verification of Hard-to-Reach Goals

First, we recognize that both PG&E and SCE achieved their HTR goals since they both mailed 100 percent of their direct mail solicitations to HTR zip codes. PG&E mailed all of its mailers to zip codes outside the nine Bay Area counties and outside the city of Sacramento. An examination of SCE's mailing strategies revealed that it mailed all of its mailers to HTR zip codes (rural, renter, and moderate income). Separate additional mailings were sent to Spanish-speaking or Chinese-speaking households. However, because it was not clear what strategies SDG&E or SoCal Gas used, we were forced to develop a different verification strategy for these two utilities.

Finally, so that we would have comparable results for all four utilities, we subjected SCE and PG&E to this same analysis.

3.2.1 Data Collection

To achieve the second evaluation objective for the Mail-In Audit, we obtained from Kema-Xenergy the zip codes for all customers who received the direct-mail solicitation to participate in the PY 2003 HEES. We also obtained a list of zip codes and the extent to which each had been defined in the Statewide Residential Needs Assessment Study (Reed, et al., 2001) as hard-to-reach. The verification of the HTR goal for SCE's local In-Home Audit Program relied on an extract from SCE's program tracking database.

3.2.2 Analysis

Our classification of the HTR zip codes relied on the Statewide Residential Needs Assessment Study (Reed, et al., 2001). To identify the HTR customers, 27 HTR segments were developed and assigned a score of 0 to 4 (except for "rural", which is a binary variable and was assigned a score of 1 if "rural" or 0 if not "rural") to reflect the extent to which each zip code could be characterized as HTR. The higher the score, the more hard-to-reach the segment was. Finally, each of these 27 segments was mapped into the five CPUC HTR criteria:

1. Language: Primary spoken language is other than English
2. Income: Those customers who fall into the moderate income level
3. Housing Type: Multi-family and mobile home tenants
4. Geography: Residents of areas other than the San Francisco Bay Area, San Diego area, Los Angeles Basin, or Sacramento
5. Tenure: Renters

Using the results of this study, we developed the most restrictive definition of HTR. For example, we assigned a particular zip code a *renter* score of 1 if it received a score of *4,* the highest score, on at least one of the 10 segments associated with renters. If it did not score a 4 on any of the 10 criteria, we assigned it a score of 0. (Note that the criterion *housing type* was combined with Tenure.) Using this approach, we assigned to each zip code a score of 1 or 0 for the remaining three HTR criteria for language, income, and geography. Each zip code could qualify as HTR by meeting *one or more* of the four criteria.

Using these definitions, we then matched the 510,228 customer zip codes that received direct mail solicitations to the zip codes identified as HTR. This analysis provides a kind of "sanity check" on the different methods used by utilities for reaching the HTR population.
For the SCE's In-Home Audit, we examined the extract prepared by SCE that identified each of the participants as HTR or not, using the same five HTR criteria listed above.

3.3 Improvement of Target Marketing

This effort focused on those customers who received the direct mailer and compares the characteristics of those who chose to participate versus those who did not with the aim of estimating models that predicted the choice to participate. Such an analysis has been referred to in the literature as a *revealed preference* study since it studies those customers who have

already revealed their choices to participate or not and seeks to understand their decisions. Going forward, the results of this modeling effort can be used to predict the audit participation behavior of all utility residential customers who have not as yet participated in the HEES.

The methods used for the revealed preference study are discussed below where we first describe all the data collected to support the goal of improving target marketing. After describing these data, we then discuss the role played by these data in the efforts to improve the target marketing of the HEES Program. We conclude with a description of the analytical techniques we used.

3.3.1 Data Collection

In this section, we discuss the various data collected to support the revealed preference study.

3.3.1.1 HEES Program Tracking Database for the Mail-In Audit

Each customer record in the statewide tracking databases for the Mail-In Audit, maintained by Kema-Xenergy, contains the customer account number, date of audit, name, address, and zip code.

3.3.1.2 Direct Mailer Files

These files containing the accounts, names, and addresses of those customers who received the direct mailer in 2003 were obtained from Kema-Xenergy through the four utilities. For the reader's convenience, Table 2-3 showing the number of direct mail records for each utility is duplicated below in Table 3-1.

Table 3-1
Direct Mail Solicitations in PY 2002, by Utility

Utility	Direct Mail Solicitations
PG&E	225,226
SCE	200,000
SoCalGas	50,000
SDG&E	35,002
Total	510,228

Kema-Xenergy was responsible for conducting the direct mailing on behalf of the utilities. A list of these direct mailing files is presented in Appendix E.

3.3.1.3 Utility Billing Records

For each customer in the direct mailer files, we obtained annual kWh and/or therm consumption for 2002 or 2003. For PG&E, while we obtained annual kWh consumption for 2003, we received no therm consumption data.

3.3.2 Geo-Code Data

Since the only information that we had for those customers in the direct-mailer files was a name, address, climate zone, and annual energy use, we needed to add demographic data from the 2000 U.S. Census at the Block Group level. The Census Block Group is an intervening level of Census geography that appears between Blocks (the smallest area of Census geography) and Tracts. As the name indicates, Block Groups are created by combining Blocks to form larger units[8]. Before we could merge the Census data with the customers in the direct mailer files, we needed to attach the Block Group ID to each customer record. Thus, for each record in the direct mailer files, we requested from the California State Department of Finance, the GIS code and Census Block Group identification number.

Having the Census Block Group ID allowed these direct mail files to be merged by Block Group ID with the Block Group-level census data, also purchased from the California State Department of Finance, which is described in the following section.

3.3.2.1 US 2000 Census Data

The following Block Group-level 2000 U.S Census variables were purchased from the California State Department of Finance:

- Number of each age
- Number of each sex
- Number of each race
- Number of each household composition configuration
- Number of each household size
- Number of family size
- Median household income
- Median family income
- Occupancy rate for owners
- Occupancy rate for renters
- Median home value
- Number of families below poverty line
- Number of individuals below poverty line
- Number of structures built in each year
- Number of dwelling units with each number of rooms
- Number of dwellings with each fuel type
- Median occupancy per room
- Number of units in structure
- Level of educational attainment
- Year householder moved into unit

[8] While the original plan called for obtaining Block level census data, the California Department of Finance was, at the time we were assembling the data, only able to provide data at the Block-Group level. We recommend that Block-level data, which are now available, be used for future target-marketing efforts.

3.3.2.2 ACXIOM Data

To drive the models, we planned to use U.S. Census data at the Block Group-level, which provides better resolution than data at the zip code level. However, we decided that it would likely be beneficial to use data at an even finer level of resolution. This is possible using a private company called Acxiom (www.acxiom.com) to get similar variables by individual household. Acxiom collects data from a variety of sources including credit applications, credit companies, warranty forms, etc., and can provide a substantial amount of data on any group of households and individuals.

We experienced mixed success in obtaining Acxiom data. SCE already subscribed to Axciom data and was able to merge those customers who received the direct mailer with their specific household data in the Acxiom files. However, SoCal Gas and SDG&E did not grant permission for us to provide Acxiom with the names and addresses of all customers who received the HEES mailers[9] in order to create a similar data file for our analysis as provided by SCE. By the time PG&E eventually granted permission to obtain these data, the decision had been made to increase the sample size from 600 to 800 households for the stated preference study. As a result, the funds were no longer available to purchase the Acxiom data for the PG&E customers who received the HEES mailer.

It is worth noting that the point of testing the Acxiom data is a very practical one. Is there an increase in predictive power due to Acxiom variables, which are more individually-based data than Census-based data that would be worth the cost of purchasing them? Many of the variables available through Acxiom are very similar to the census variables at a different level of aggregation. However, they aren't exact matches; but it is not important that they be exact matches to answer the practical question above.

SCE also had data from an Acxiom competitor, Claritas. These variables were also merged onto the SCE analysis files. However, they were never sufficiently predictive to enter the models. Thus, they are not described further.

3.3.2.3 Data Development and Integration

Once collected, these data were then assembled into four analysis datasets, one for each utility. Table 3-2 summarizes the available data for the revealed preference analysis.

[9] This involved giving Acxiom nothing but names and addresses, with no household information, or consumption data or any other data that they could use. Further, Acxiom was willing to execute a signed agreement that said they would not use any information we (the Evaluation Team) gave them, and that we would not use any information on the customers that they provided except for the purposes of this project. Given these conditions, we saw little risk in providing Acxiom with customer names and addresses in order to obtain data for the purposes of this analysis.

Table 3-2
Available Data for the Revealed Preference Analysis

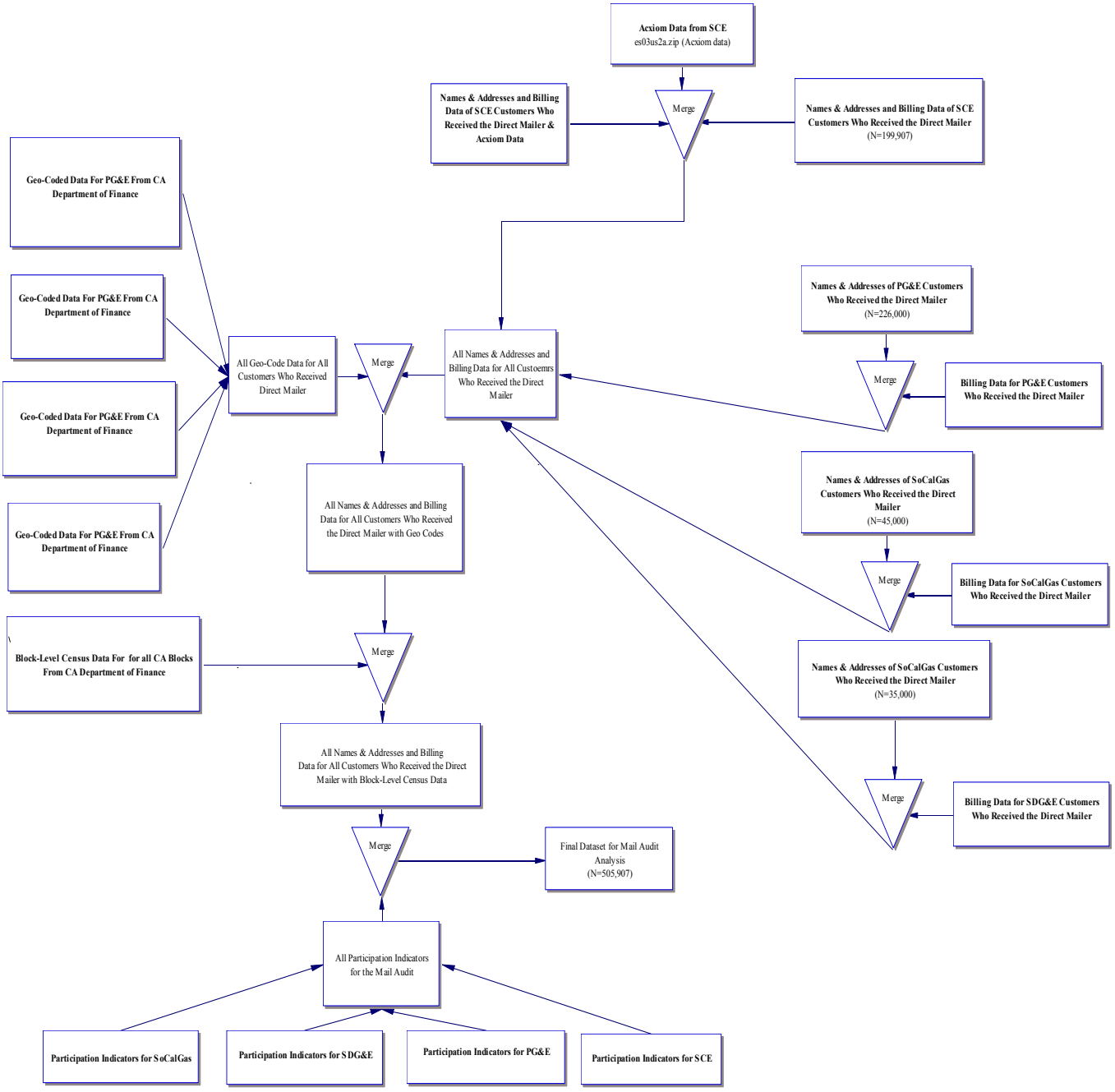
	PG&E	SCE	SCG	SDG&E
Number Mailed	225,226	200,000	50,000	35,002
Number& % with Billing Data	kWh: 166,499	kWh: 200000	kWh: NA	kWh: 27,736
	[74%]	[100%]		[79%]
	Therms: None Provided	Therms: NA	Therms: 50,000	Therms: 34,534
			[100%]	[99%]
Number & % with Geo-code Matches for Census Variables[10]	194,871 [87%]	181,390 [91%]	45,049 [90%]	31,302 [89%]
Number & % with both kWh and Census Variables	146,438 [65%]	181,390 [91%]	45,049 [90%]	24,925 [71%]
Number & % with Acxiom Data	NA	200,000 [100%]	NA	NA
Number & % Participants	40,676 [18%]	24,109 [12%]	7,696 [15%]	5,248 [15%]

First, the customers who received the direct mailer were merged by the Census Block Group ID with the Block Group-level Census data. The annual kWh and/or therm data were then merged with the customers who received the direct mailer. The resulting file was one in which each row represented one customer along with Block Group-level census variables and annual kWh. For SCE customers, this file was then merged with Axciom data. Figure 3-1 graphically illustrates this process.

[10] Even though the percentage of participant addresses that the California Department of Finance (DOF) was able to match to a Block Group ID was quite high, there was some concern that the non-matches were the result of some systematic process that introduced bias. There was a similar concern with the account number non-matches with utility billing files, the source of kWh data. With respect to non-matches with the DOF files, the DOF staff examined a random sample of the non-matches and discovered two basic reasons for non-matches. The first is that there were differences in spelling in the addresses in the two files. This appeared to be a random process. The second is that while some addresses in the files that we provided to DOF were apparently valid, hand checking revealed that there were no corresponding addresses in the DOF file that were even close to matching. Their conclusion was that these addresses were not part of the 2000 Census. How a household is not included in the Census could be that the house was built after the 2000 Census, in which case newer homes would be underrepresented.

However, any potential bias with respect to demographic variables or kWh consumption was mitigated since our best models of participation were estimated using CART. One of the features of the CART software is its handling of missing values. At each node, the software searches all variables and all splits for all of the variables for the most efficient variable. It chooses the most efficient for actual use. However, it also identifies any other variables that split the sample in a way similar to the chosen (primary) splitter. These variables are called "surrogates" and are actually used to complete a split when there are one or more missing values in the primary splitter. Thus, some missing values in a primary splitter need not stop the growth of the tree or even make assumptions about what the missing values would have been if they weren't missing. Actual, present values on a similarly-performing variable are used instead. See Section 3.3.3.1 for more details regarding CART.

Figure 3-1
Data Integration for the Revealed Preference Study



3.3.3 Analysis

Two types of analyses were originally envisioned for the revealed preference study: CART and discrete choice. Each of these analytical techniques is described in the context of the available data in the following sections.

3.3.3.1 CART

The goal of the analysis of data on all customers solicited for the program, with those actually participating flagged in the dataset, is to find the set of variables that most efficiently distinguishes those who participated from those who did not. Once we know what variables make this distinction, the same variables can be used in the future to predict who will and will not participate. Customers with these characteristics can then be targeted by marketing efforts. One very powerful method to determine the best variables is found in CART (Classification and Regression Tree analysis). Following is a description of that method, in general, followed by a short discussion of how this method was applied in this specific situation.

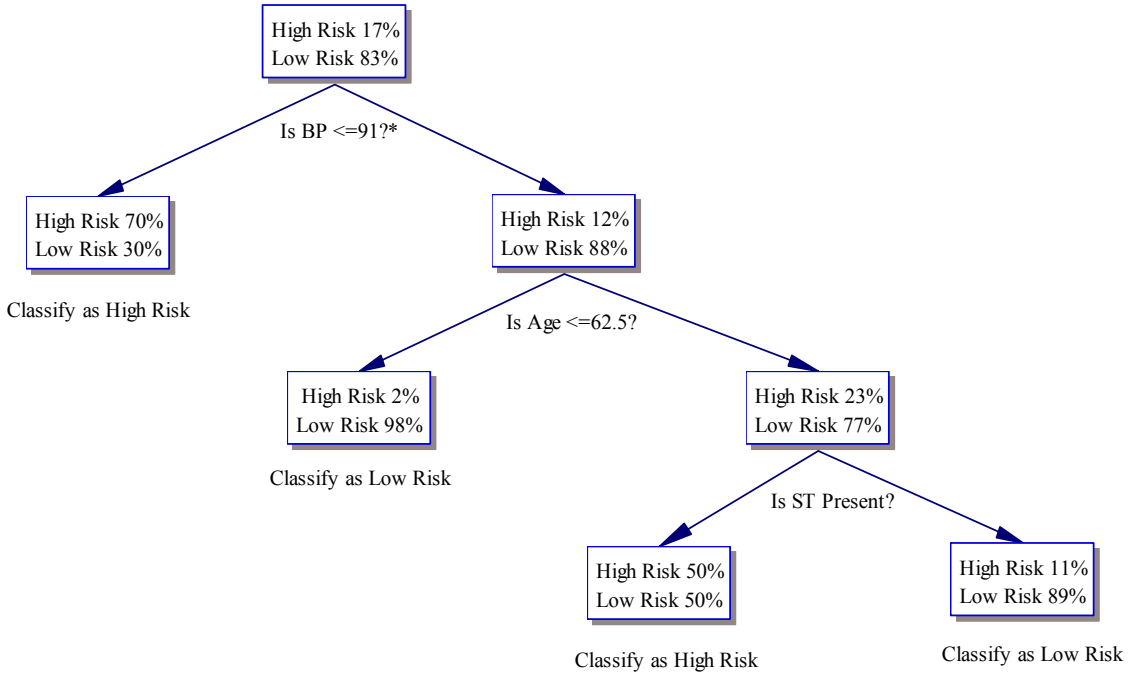
CART creates "trees" that show how the original group of customers who received solicitation letters are divided into participators and non-participators by successive group splits created by one of the variables in the dataset. These trees show the most efficient and the most stable splitters (Note: Think of splitters are predictor variables) and the resulting subgroups or clusters of customers who did and did not participate. CART algorithms can deal equally effectively with categorical and continuous variables as splitters CART is remarkably flexible in that it can consider the same variable for any number of different subgroups defined by other variables, allowing the same variable to behave differently for groups appearing in different parts of the tree. The software is also well designed to handle very large datasets, and to handle the very lopsided distributions such as we have in this dataset. An example will help illustrate this method.

The example given here is a medical one, and comes from a published study (Gilpin, Olshen, Henning and Ross, 1983). The data for the analysis came from the emergency room records of 215 patients who arrived having had a heart attack. The goal of the analysis was to use common data collected on such patients to predict who would quickly have a second heart attack. Thirty-seven patients had died within 30 days of admission, while the remaining 178 were surviving at that time. Thus, the grouping variable is survival. Since this analysis attempts to use actual patient histories to predict the risk status of future patients, the grouping variable will be called "Risk", with two groups: "High Risk" and "Low Risk". The medical data collected at the hospital are the potential splitters. The splitter variables available are: blood pressure, age, presence of tachycardia, and enzyme concentrations.

Figure 3-2 illustrates the visual result of this analysis. It appears as an upside down tree. It is also a more simple resolution than is usual. The top box (also called a node in CART terminology) begins with 17 percent high-risk patients and 83 percent low risk. CART has determined that a blood pressure reading above or below 91 is a critical splitter in classifying patients as either high or low risk. If the person's blood pressure is less than or equal to 91, s/he goes left, otherwise, right. The group represented by the left node cannot be further purified by more splitting with additional variables. It is a terminal node, containing 70

percent high-risk patients. The node to the right contains 88 percent low-risk patients, but this group can be further purified by the age variable. If the patient's age is less than or equal to 62.5, s/he is sent to the left at the next level. This group is now 98 percent low risk. The group represented by the node to the right at this level contains 77 percent low risk. This group can be further purified by the sinus tachycardia variable. The result of that split is seen at the next level with two terminal nodes. The high-risk node contains 50 percent high-risk patients, and the low-risk node contains 89 percent low-risk patients.

Figure 3-2
Binary Decision Tree



* "Yes" answers always go to the left.

With this basic understanding of the method, a few more important details can be added. First, past data mining/tree programs have suffered from validation problems, i.e., results tended to be sample-specific because the method capitalized on chance as much as stepwise regression does. CART has developed methods to address this problem. CART allows two approaches to validation. One method allows the user to specify two samples, one as a "learning sample" and the other as a "validation sample." The initial tree is grown in the learning sample, and it is corrected based on its application to the validation sample.

A second available approach (and the one used in this study) uses just one sample but divides the sample into 10ths. The tree and its nodes are developed on 9/10s of the sample provided, and then successively tested on nine other sets of 9/10 samples. This is accomplished by adding back the 1/10 originally held out and removing a different 1/10, over and over until all 10 versions of a 9/10 sample have been used. These procedures are referred to as cross-validation. This does not, however, refer to efforts to determine validity as the term is used in

the literature on measurement. Rather, it refers to a procedure that changes the sample enough each time to be sure the splits are not based on some entirely unique configuration of cases and variables that is not replicable.

One of the features of the CART software is its handling of missing values. At each node, the software searches all variables and all splits for all of the variables for the most efficient one. It chooses the most efficient for actual use. However, it also identifies any other variables that split the sample in a way similar to the chosen (primary) splitter. These variables are called "surrogates" and are actually used to complete a split when there are one or more missing values in the primary splitter. Thus, some missing values in a primary splitter need not stop the growth of the tree or even make assumptions about what the missing values would have been if they weren't missing. Actual, present values on a similarly-performing variable are used instead. The step-by-step splitting process, and the use of surrogates as supplemental splitters when there are missing values on the primary splitter, means that there is no downside to including many variables in the available pool of variables. Many of them may be highly correlated, but problems of multi-collinearity will not be evidenced. In fact, correlated predictors can be an advantage because they increase the pool of potential surrogates.

CART was applied in this setting to determine which sets of customer characteristics define one or more clusters of customers likely to participate in future programs, and one or more clusters highly unlikely to participate. There are some differences in type and availability of variables for each utility. There are also weather and other differences as well, which could affect how customers make these decisions. In addition, each utility targeted customers differently for the direct mailers that solicited participation. For all of these reasons, it made sense to develop utility-specific CART models.

All utilities provided variables for annual energy consumption and customer tariff, including a CARE (California Alternate Rates for Energy) flag. Also consistently available were the US Census variables described in Section 3.3.1.1.5. However, as noted above, SCE also provided variables from both Acxiom and Claritas. Most of these variables are demographic, but have the advantage of representing specific individuals and residences rather than Census Block Groups. Thus, more precision is possible. Beyond these variables, pre-defined customer segments and clusters were available. The presence of these variables, individualized demographics and marketing segments allow us to determine whether those variables provide a real advantage in targeting solicitation efforts compared to the less expensive Census variables.

3.3.3.2 Discrete Choice

Logistic regression is a frequent approach when the dependent variable is dichotomous; that is, takes on only two discrete values. Our revealed-preference data set met this criterion, as the domain of our dependent variable, audit participation, consists of the values 0 ("did not participate") and 1 ("participated"). When a dependent variable such as this takes on just two values, the ordinary least squares (OLS) formulation used for continuous dependent variables,

$$Y = a + \beta X + \varepsilon \qquad \text{Eq. 1}$$

suffers from three problems:

1. The error term e is heteroskedastic. That is, the variance of the independent variables differs for each value of the dependent variable.
2. The distribution of the error term e violates OLS assumptions, in that the term is not normally distributed because Y takes on only two values.
3. The predicted values of Y can lie outside the range {0…1}, creating absurd estimates for probabilities.

A transformation of the OLS formulation gives us the logistic regression model:

$$\ln\left(\frac{p}{1-p}\right) = a + \beta X + \varepsilon \qquad \text{Eq. 2}$$

or

$$\left(\frac{p}{1-p}\right) = e^{(a+\beta X+\varepsilon)} \qquad \text{Eq. 3}$$

where:

ln=	natural logarithm
p=	the probability that event Y occurs
$\frac{p}{1-p}$ =	the odds ratio
$\ln\left(\frac{p}{1-p}\right)$ =	the log odds ratio, or logit

The remaining components of the model are the same, except the OLS distributional assumptions about e are satisfied.

Solving for p, the estimated probability, gives us

$$p = \frac{1}{1+e^{-a-\beta X}} \qquad \text{Eq. 4}$$

so that

If $-a-\beta X = 0$ then p = 0.5

As $-a-\beta X$ gets large, p approaches 1

As $-a-\beta X$ gets small, p approaches 0.

3.3.3.2.1 Simulation

The results of the CART analysis estimated on the total direct-mail recipient group were used to apply to customers who could be targeted in future program years. In this simulation exercise, data from previously unsolicited customers were used to evaluate the estimated model in order to predict the likelihood of their future participation. For each utility, the estimated CART model was simulated using a random sample of residential customers. Table 3-3 presents the size of the random samples for each utility.

Table 3-3
Sample Size of Residential Customers for Simulation, by Utility

Utility	Sample Size
PG&E	3,000
SCE	5,000
SDG&E	5,000
SoCal Gas	None Provided
Total	13,000

Before the simulation could be conducted, the demographic variables used in the earlier CART analysis had to be attached to each customer. In order to do this, the addresses of these 13,000 customers were sent to the California State Department of Finance so that the Block Group ID could be attached. Having the Block Group ID allowed each customer to be merged with the Block Group-level demographic data from the U.S. Census.

3.4 Assessing Customer Interest in Other Types of Audit Configurations

In the research plan, one of the activities was to learn more about the various reasons why customers might choose to participate in an audit program, what features (audit configurations) are attractive to them, and what factors, such as demographic characteristics and attitudes of the customer, trigger decisions to participate. Understanding these issues could help program planners modify the design of their programs as well as better target market audits that may have any of these new features.

Data were collected via the Internet from a representative sample of residential customers who were asked to state their preference for different kinds of audits. Such an analysis is called a *stated preference* study, the methods for which are presented next.

3.4.1 Experimental Design

In the experimental design phase of the project we create product offerings that respondents will see in hypothetical stores[11]. During the survey, respondents are asked to choose between these offerings. In building the design, we adhere to two principal objectives. First, we want the product attributes to be completely uncorrelated with each other; second, we would like each level for every attribute to appear an equal number of times throughout the entire design. The statistical terms for these desirable characteristics are *orthogonality* and *balance,* respectively. The better the orthogonality and balance are, the more efficient the design.

Fortunately, the research community has assembled an extensive library of arrays that meet these criteria. A particular class of arrays having perfect orthogonality and balance is the set of orthogonal arrays, one of which was used to develop the experimental design for this study. Appendix B presents this array in detail.

3.4.2 Sample Design

The original plan was to gather these data by telephone interviews. We anticipated that we would randomly select residential customers and interview100 customers from each utility, totaling 400 interviews. However, the structure of the questionnaire became sufficiently complex that we grew concerned that the response rate would be unacceptably low. Note that response rates for both mail and telephone surveys have been decreasing steadily over the last 10 years with a corresponding increase in the non-response bias.

We therefore decided to implement the survey using the services of Infosurv (see www.infosurv.com), which has identified an existing panel of residential customers provided by e-Rewards (see *e-Rewards Panel Quality* in Appendix D). Also note that all panel establishment methodologies employed by e-Rewards fully comply with CASRO (Council of American Survey Research Organizations) guidelines (see www.casro.org), of which e-Rewards is a member organization. These guidelines cover the following topics:

- Problem Definition Guidelines
- Sample Design Guidelines
- Interview Design Guidelines
- Data Collection Guidelines
- Data Processing Guidelines
- Survey Reporting Guidelines

Through a contract with Blockbuster Video, e-Rewards relies on the Blockbuster member list as the source for panel recruitment. The existing panelists are compensated by e-Rewards for participating in no more than three consumer surveys a year. Panelists are invited to participate in the survey such that, as a group, they are representative of the residential households in each of the four utility service territories. The number of available e-Reward panelists who reside within each utility service territory is presented in Table 3-4.

[11] The concept of the virtual store was used in order to make the experience more familiar to the respondent.

Table 3-4
Available E-Reward Panelists, by Utility

Utility	Number of Available Panelists
PG&E	51,657
SCE	26,276
SoCalGas	47,870
SDG&E	12,049
Total	137,852

A random sample of these residential panelists was invited to log on to the designated website and take the survey, with a goal of completing 200 surveys per utility. This questionnaire was programmed to be taken over the Internet with the results automatically delivered to Infosurv, which cleaned and prepared the data for our analysis.

3.4.3 Data Collection Process

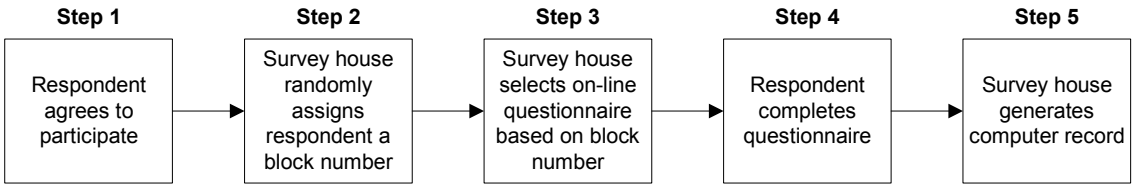
The Internet questionnaire is presented in Appendix A and was organized using the following topics:

- screening questions to determine eligibility
- customer choices regarding various types of audits
- attitudes about energy use
- awareness of energy efficiency
- past participation in programs
- demographic characteristics
- household energy use

After a panelist launched the survey, they were asked a series of screening questions. Panelists who indicated that their California residence was not their permanent residence and who indicated they were not served by one of the four utilities were terminated. The remainder of the questionnaire was organized using the following topics:

- customer choices regarding various types of audits
- attitudes about energy use
- awareness of energy efficiency
- past participation in programs
- demographic characteristics
- household energy use

Because of its importance, these next few paragraphs provide a more detailed illustration of the data collection process with respect to customer choices. We follow a hypothetical record from the point at which a hypothetical respondent agrees to complete the online survey through the collection of that response and its presentation in a dataset. An overview of the process appears below:



For purposes of illustration, we'll call our hypothetical respondent Jane Smith.

Step 1. Smith, a member of the survey house's standing panel agrees to participate in the survey. She is assigned an ID of 1 and logs on to the website containing the questionnaire.

Step 2. The survey house randomly assigns Smith a block number[12] from the set 1 through 6. To carry our illustration through we will assume she is assigned block 4.

Step 3. The survey house selects block 4's store visits[13] and uses them in the discrete-choice section of the questionnaire. A virtual store (see Table 3-5) presents the respondent with 8 types of audits from which to choose. A ninth option is "None of These." Each customer was guided through 9 stores[14]. When she gets to that part of the questionnaire, the first store she sees looks like the following:

Table 3-5
Example of First Store in Block 4

Store 1 **Block 4**

Directions: Please choose the audit package that you most prefer from among the following options, or specify "None of these" if you don't want any of them. Click the radio button corresponding to your choice.

	1	2	3	4	5	6	7	8	9
Mode of delivery:	Online	Online	Online	Mail	Mail	Online	In-home	Mail	None of these
Time required for you to complete audit:	45 min.	45 min.	15 min.	40 min.	60 min.	20 min.	15 min.	20 min.	
Additional services:					Post-audit technical support				
	Usage profile	Usage profile		Usage profile	Usage profile				
Links to vendors:	Local vendors	Local vendors	Local vendors	No links	National outlets	National outlets	Local vendors	No links	
Level of detail and accuracy in savings estimates:	Refined estimate of savings	Rough estimates of savings	Rough estimates of savings	Refined estimate of savings	No savings estimates	Refined estimate of savings	Refined estimate of savings	Rough estimates of savings	
Audit fee or incentive:	You pay $15	You pay $75	You pay $45	Free	You pay $75	You pay $75	We give you $30*	You pay $75	
Time between decision to participate and results:	3 weeks	3 weeks	1 week	2 weeks	4 weeks	2 weeks	3 weeks	4 weeks	

*in energy-saving equipment

Now from among these same 9 choices, select the one you LEAST prefer. Click the radio button corresponding to your choice.

In Table 3-6, we present what the second store would look like.

[12] To keep the number of stores presented to respondents at a manageable level, choice sets were divided into 6 blocks of 9 stores each. Every respondent was randomly assigned one of these 6 blocks and subsequently went through the 9 stores in that particular block.

[13] The use of a *virtual store* is simply a way to couch the choice experiment in terms familiar to the respondent.

[14] There was a total of 54 stores (6 blocks times 9 stores within each block).

Table 3-6
Example of Second Store in Block 4

Store 2 **Block 4**

Directions: Please choose the audit package that you most prefer from among the following options, or specify "None of these" if you don't want any of them.
Click the radio button corresponding to your choice.

	1	2	3	4	5	6	7	8	9
Mode of delivery:	In-home	Mail	Online	Mail	Online	Online	Mail	Online	None of these
Time required for you to complete audit:	30 min.	40 min.	50 min.	30 min.	45 min.	45 min.	60 min.	15 min.	
Additional services:	Post-audit technical support		Post-audit technical support	Post-audit technical support			Post-audit technical support		
		Usage profile	Usage profile		Usage profile	Usage profile	Usage profile		
Links to vendors:	Local vendors	National outlets	National outlets	No links	No links	No links	Local vendors	No links	
Level of detail and accuracy in savings estimates:	Rough estimates of savings	Rough estimates of savings	Rough estimates of savings	No savings estimates	No savings estimates	Rough estimates of savings	Rough estimates of savings	Refined estimate of savings	
Audit fee or incentive:	You pay $15	Free	We give you $30*	We give you $15*	You pay $75	You pay $15	We give you $15*	We give you $30*	
Time between decision to participate and results:	3 weeks	2 weeks	3 weeks	3 weeks	3 weeks	1 week	4 weeks	3 weeks	

*in energy-saving equipment

Now from among these same 9 choices, select the one you LEAST prefer.
Click the radio button corresponding to your choice.

The remaining seven stores would then be presented. The complete description of the audit options for all 54[15] combinations of 9 stores and 6 blocks is contained in *Final Experimental Design 2005-09-12 with Export.xls*, submitted as a part of this report[16].

Step 4. Smith completes the first part of the questionnaire, then moves on to the discrete-choice section. For purposes of illustration, assume that in store 1 her most favorite option was alternative 7, and her least favorite option was 9; then in store 2 her most favorite option was alternative 8, and her least favorite option was 5. Smith goes on to make other choices in the remaining 7 stores and then proceeds to complete the questionnaire. While Smith filled out the discrete-choice section, the survey's HTML code checked to see that she followed two basic rules:

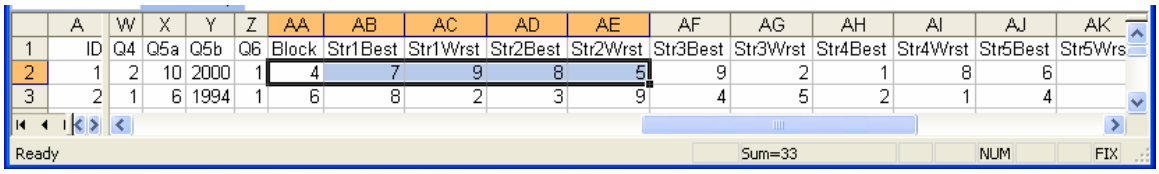
1. In each store, her choice did not equal her least favorite alternative.
2. She did not straight-line her answers by selecting the same numbered alternatives in every store.

The first rule forced a respondent not to violate the definition of "least favorite alternative" since, by definition, there can only be one. The second rule was designed to make sure that she was not simply checking, without any thought, the same options in every store, i.e., there had to be evidence that she was *thinking* about her choices. Had Smith violated either of these strictures, she would have been dropped from the survey.

Step 5. The survey house collects Smith's answers and adds them to the database of responses for the entire survey. The database is structured as a rectangular file with separate columns for each question and separate rows for each respondent. In the discrete-choice section one column contains the assigned block number. There are two columns of each of the nine stores: one for the first choice, and the other for the least preferred. For Smith's record in Row 2, entries might look like this (with the example responses we have used highlighted). A second hypothetical respondent's answers are recorded in Row 3. An ID of 2 is assigned to this second respondent (each respondent is assigned a unique ID).

[15] See Appendix B for a technical discussion of the rationale for using 6 blocks, with 9 stores within each.
[16] This Excel file is available upon request.

Table 3-7
Example of Database Structure



	A	W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG	AH	AI	AJ	AK
1	ID	Q4	Q5a	Q5b	Q6	Block	Str1Best	Str1Wrst	Str2Best	Str2Wrst	Str3Best	Str3Wrst	Str4Best	Str4Wrst	Str5Best	Str5Wrs
2	1	2	10	2000	1	4	7	9	8	5	9	2	1	8	6	
3	2	1	6	1994	1	6	8	2	3	9	4	5	2	1	4	

Ready Sum=33 NUM FIX

Subsequent processing by the consulting team converts this relatively simple format into a form used to estimate the model.

3.4.4 Sample Disposition

Table 3-8 shows the quotas and the final achieved sample for each utility.

Table 3-8
Quotas and Achieved Interviews
by Utility/Program Combination

Utility	Quotas	Completes: Electricity	Completes: Gas
PG&E	200	237	197
SCE	200	319	N/A
SDG&E	200	231	188
SoCal Gas	200	N/A	273
Total	800	787	658

Table 3-9 presents stages of the survey starting with the population of households and ending with those households that completed the survey.

Table 3-9
Survey Stages

Survey Stage	Panelists
Available Panelists	137,852
Received E-Mail	20,316
Clicked from E-Mail to Offer	3,180
Read Offer	1,188
Completed Survey	793

In all, of the available 137,852 e-Reward panelists, 20,316 (14.7 percent) were invited to participate in the survey. Of these, 3,180 (15.7 percent) clicked from the e-mail to the actual offer to participate in the survey. Of these, 1,188 (37.4 percent) actually read the offer and began the survey. Of these, 793 (66.8 percent) completed the survey. In the end, 3.9 percent of those who were invited to participate completed the survey (793/20,316). Given the complexity of the survey, this is a much higher response rate than if the same survey had

been administered by mail. The complexity of the survey made it impossible to conduct by telephone.

One goal of any survey is that the respondents resemble the population about which one wishes to generalize. To meet this goal, e-Rewards invited panelists to participate such that in aggregate they would resemble the demographic characteristics of those households in the service territories of the four California utilities (i.e., the e-Rewards invited panelists are representative of those households in the service territories of the four California utilities). While e-Rewards has relatively little control over who actually responds, the hope is that those who complete the survey resemble the population of California households about which we wish to generalize survey results.

Table 3-10, Table 3-11, Table 3-12, Table 3-13, and Table 3-14 compare three groups with respect to each demographic variable (i.e., owner vs. renter, age, ethnicity, educational attainment and household income).

Table 3-10
Owner vs. Renter Status, by Panelists Who Received the E-Mail Invitation, Those Who Completed the Survey, and All California Households

Housing Status	Received E-Mail	Completed Survey	California
Own	45%	70%	57%
Rent	30%	29%	43%
Other	6%	1%	0%
No Response	19%		
	100%	100%	100%

Table 3-11
Age, by Panelists Who Received the E-Mail Invitation, Those Who Completed the Survey, and All California Households

Age	Received E-Mail	Completed Survey	California
Under 25 years	9%	3%	14%
26 - 35 years	27%	29%	21%
36 - 45 years	23%	26%	23%
46 - 55 years	18%	22%	17%
56 - 65 years	9%	13%	12%
65+	3%	4%	13%
No Response	10%	3%	0%

Table 3-12
Ethnicity, by Panelists Who Received the E-Mail Invitation, Those Who Completed the Survey, and All California Households

Ethnicity	Received E-Mail	Completed Survey	California
African American/Black	2.5%	2.1%	6.3%
Asian American	7.8%	13.9%	10.2%
Caucasian/White	31.1%	53.7%	46.5%
Hispanic Origin	26.3%	20.5%	32.4%
Native American, Inuit or Aleut	0.4%	0.3%	0.3%
Other	4.3%	5.6%	4.4%
No Response	27.7%	3.9%	
Total	100%	100%	100%

Table 3-13
Educational Attainment, by Panelists Who Received the E-Mail Invitation, Those Who Completed the Survey, and All California Households

Educational Attainment	Received E-Mail	Completed Survey	California
Less then 9th grade	-	-	11%
9th grade, no diploma	1%	0%	12%
High school graduate (includes equivalency)	7%	5%	20%
Some college, no degree	24%	22%	30%
Bachelor's degree	34%	41%	17%
Graduate or professional degree	25%	31%	10%
No Response	9%	1%	-
	100%	100%	100%

Table 3-14
Household Income, by Panelists Who Received the E-Mail Invitation, Those Who Completed the Survey, and All California Households

Income	Received E-Mail	Completed Survey	California
Less Than $25,000	7.5%	3.7%	25.5%
$25,000 to $49,999	14.3%	13.7%	26.6%
$50,000 to $74,999	14.7%	20.2%	19.1%
$75,000 to $99,999	12.6%	18.6%	11.5%
$100,000 to $149,999	13.8%	22.5%	10.4%
$150,000 to $199,999	6.0%	7.5%	3.3%
$200,000 or more	5.5%	6.7%	3.6%
No Response	25.7%	7.2%	-
	100%	100.0%	100%

The first thing we note is that these tables are somewhat difficult to interpret since for some variables such as household income and ethnicity there is a fairly high rate of non-response

among the e-Rewards panelists. Only if one knows in which response categories for any given variable these non-respondents would fall can one reliably compare the distributions. However, based on the available responses, we make the following observations: Compared to the California residential population,

- owners are slightly under-represented among those who received the e-mail and slightly over-represented among those who completed the survey.
- the age distribution of both those who received the e-mail and those who completed the survey are very similar.
- Hispanics are somewhat under-represented among those who received the e-mail and among those who completed the survey.
- Caucasians are slightly under-represented among those who received the e-mail and slightly over-represented among those who completed the survey.
- the less educated are significantly under-represented among those who received the e-mail and among those who completed the survey.
- the more educated are significantly over-represented among those who received the e-mail and among those who completed the survey.
- the less wealthy are somewhat under-represented among those who received the e-mail and among those who completed the survey.
- the wealthier are somewhat over-represented among those who received the e-mail and among those who completed the survey.

In general, while the biggest difference is in educational attainment and income, those who completed the survey are moderately consistent with the California population of residential households. Decisions regarding the use of weights to correct for these discrepancies are discussed near the end of Section 4.4.1.1.

Infosurv has no way to determine why the 17,136 (20,316 – 3,180) households failed to click from the e-mail to the offer to participate in the survey. As with a mailed survey, one has no way to determine exactly why a household failed to return a survey it received. Perhaps many were on vacation or were away on business. Likewise, Infosurv can only speculate why 1,992 panelists (3,180 – 1,188) who clicked from the e-mail to the offer did not read the offer. While Infosurv did not capture the reasons why the 395 households (1,188 – 793) did not complete the survey, they expect that the primary reasons were that they were not customers of one of the four utilities, or were not permanent California residents (i.e., did not live at their California address at least nine months of the year).

However, we can, with available demographic data (age, gender, household income, ethnicity, age, and education) on these five groups, make comparisons to determine how different they are. Table 3-15 shows the four pairwise comparisons that were made among these five groups. For example, we compared those who received the e-mail to those who chose to click on the e-mail offer to determine whether the latter group remained reasonably representative of the former group. Other comparisons were made ending with the one that compared those who received the e-mail with those who completed the survey.

Table 3-15
Response Group Comparisons Made With Respect to Demographic Characteristics

Group	Response Groups	Pairwise Comparisons			
1	Received E-Mail	X			X
2	Clicked from E-Mail to Offer	X	X		
3	Read Offer		X	X	
4	Completed Survey			X	X

For each comparison, a Chi-square statistic was calculated. Because the magnitude of the chi-square statistic is related to the sample size, the large number of observations involved in each of the five comparisons meant that even small differences of little practical significance to any decision-maker were *statistically* significant. To provide a better sense of whether the various groups were different from one another, we focused on the magnitude of the association between the two variables assuming that the variable Response Group was the dependent variable and a given demographic variable was the independent variable[17]. For each of the four comparisons, Table 3-16 presents the Somers' d statistics for age, educational attainment, and household income, all ordinal variable. Table 3-16 also presents the lamda asymmetric statistic for gender, owner/renter, and ethnicity, all nominal variables.

Table 3-16
Association of Demographic Variables to Group Membership,

	Comparisons*			
Demographic Variable	Group 1 vs. Group 2	Group 2 vs. Group 3	Group 3 vs. Group 4	Group 1 vs. Group 4
Age	-0.0008	-0.0901	0.1189	0.0055
Educational Attainment	-0.0125	-0.0299	-0.0213	-0.009
Household Income	0.0033	-0.0375	-0.0305	-0.0082
Gender	0.0000	0.0000	0.0000	0.0000
Own/Rent	0.0000	0.0000	0.0211	0.0000
Ethnicity	0.0000	0.0339	0.0119	0.0000

*** Group 1 =** Received E-Mail **Group 2 =** Clicked on E-Mail
Group 3 = Read Offer **Group 4 =** Completed Survey

The results across all four comparisons suggest very little, if any, non-response bias, at least based on observed data. However, it is possible that these groups could be different in ways that we have not been able to observe.

In summary, there appears to be little non-response bias and the results of our analyses are reasonably generalizable to the residential households in the service territories of the four utilities.

[17] Both Somers'd and lamda asymmetric can be interpreted as the proportionate reduction in errors in predicting ranking on the dependent variable (Loether and McTavish, 1976). Somers' d ranges from -1.0 to +1.0 while lamda asymmetric ranges from 0.0 to +1.0.

3.4.5 Analysis

Two types of analysis were used for the stated preference study: CART and discrete choice. Each of these analytical techniques is described in the following sections in the context of the available data. We note here that understanding the following technical discussions is not necessary to understand the ultimate results of this analysis and their value. However, this technical discussion may be of interest to some.

3.4.5.1 Latent-Class Discrete Choice

The second major task of the project is to generate a latent-class discrete-choice (LCDC) model of program participation, which used the same dataset that was used for the CART analysis. This methodology combines the strengths of latent-class analysis and discrete-choice estimation in a single analytical framework. The discrete-choice component will help inform the relationship between independent variables such as program offerings, respondent characteristics such as geography and energy consumption, and the probability of program participation. The latent-class component tested for the existence of separate customer groups who respond to these variables in distinct ways.

Classification is based on a probability model. For each respondent, the technique calculates a probability for membership in each class, the probabilities summing to one for a given respondent. In so doing, the approach creates a profile of class membership across the sampled population.

To estimate probabilities of audit participation, we employed a LCDC methodology developed by Jay Magidson and Jeroen Vermunt (2003). For notation, i represents one respondent among the total number of respondents I. We presented each subject i with S choice sets consisting of K alternatives, where k is a particular alternative in choice set s. Each alternative k is described by a set of attributes A, where a is a single attribute. Let y_{is} represent the choice respondent i makes among the K alternatives in choice set s. More generally, let vectors $\mathbf{y}_i$, $\mathbf{z}_i^{att}$ and $\mathbf{z}_i^{cov}$ refer respectively to all responses, attributes and covariates for individual i. In this sense, attributes are characteristics of alternatives presented to subjects, and covariates are characteristics of the subjects themselves. Within this context, z_{ias}^{att} represents the attributes of a single alternative k evaluated by subject i in choice set s, and z_i^{cov} represents the set of R covariates describing subject i. We will also estimate the probability that each respondent falls into a latent class x, where x is an integer value $1 \le x \le C$, where C is the total number of latent classes.

For each latent class, a conditional logit model is estimated, using the form:

$$P(y_{is} = k \mid z_{is}^{att}) = \frac{e^{V_{k|z_{is}}}}{\sum_{k'=1}^{K} e^{V_{k'|z_{is}}}} \qquad \text{Eq. 5}$$

Where $V_{k|z_{is}}$ is the systematic component in the utility of alternative k for subject i in choice set s, and k' is an index for each alternative in K. V, sometimes called representative value, is a linear combination of part-worths and attributes, plus an error term ε that is assumed to have a Gumbel distribution.

$$V_{k|z_{is}} = \sum_{a=1}^{A} \beta_a^{att} z_{iask}^{att} + \varepsilon \qquad \text{Eq. 6}$$

For simplicity we will omit the error term below and focus on the systematic component of utility. Also note that this particular specification omits alternative-specific constants, though other specifications sometimes include them.

In a latent class (sometimes called finite mixture) model, individuals are assumed to belong to latent classes that differ with respect to one or more of the β parameters. The choice probabilities therefore depend on latent class membership *x*, and the logit model takes the form:

$$P(y_{is} = k \mid x, z_{is}^{att}) = \frac{e^{V_{k|x,z_{is}}}}{\sum_{k'=1}^{K} e^{V_{k'|x,z_{is}}}} \qquad \text{Eq. 7}$$

The term $V_{k|x,z_{is}}$ represents the systematic component of the utility of alternative *k* within choice set *s* for respondent *i*, who is a member of latent class *x*. The representative value equation therefore becomes:

$$V_{k|x,z_{is}} = \sum_{a=1}^{A} \beta_a^{att} z_{iask}^{att} + \varepsilon, \qquad \text{Eq. 8}$$

so the only difference between this version and the aggregate model is that the β parameters are class-specific.

The probability density associated with the LCDC model is:

$$P(y_i \mid z_i) = \sum_{x=1}^{C} P(x) \prod_{k=1}^{K} P(y_{is} \mid x, z_{is}^{att}). \qquad \text{Eq. 9}$$

Here, $P(x)$ is the unconditional probability of belonging to class *x*. It is also the size of class *x*. We will show that we can modify this probability so that it depends on an individual's covariates z_i^{cov}, so $P(x)$ is replaced by $P(x \mid z_i^{cov})$.

As the above equation implies, the y_{is} choices from each set of alternatives are assumed to be independent of each other given class membership. This is equivalent to the assumption of local independence common in latent class models. Responses are also assumed to be independent conditional on the value of the random coefficients.

3.4.5.1.1 Covariates

Our LCDC model includes covariates, which are used to predict class membership. With covariates, the model specification changes to:

$$P(y_i \mid z_i) = \sum_{x=1}^{C} P(x \mid z_i^{\text{cov}}) \prod_{k=1}^{K} P(y_{is} \mid x, z_{is}^{att}) . \qquad \text{Eq. 10}$$

Now we assume that class membership of individual *i* depends on a set of covariates z_i^{cov}. We accomplish this by specifying a multinomial logit model in which class membership is regressed on covariates:

$$P(x \mid z_i^{\text{cov}}) = \frac{e^{m_{x|z_i}}}{\sum_{x'=1}^{C} e^{m_{x'|z_i}}}, \qquad \text{Eq. 11}$$

where *m* is a linear combination of parameters and covariates:

$$m_{x|z_i} = \delta_{0x} + \sum_{r=1}^{R} \delta_{rx} z_{ir}^{\text{cov}} . \qquad \text{Eq. 12}$$

Here δ_{0x} represents the intercept or constant term corresponding to latent class *x*, and δ_{rx} represents the coefficient for the r^{th} covariate for class *x*.

Notice that the treatment of covariates differs from the one often employed in traditional logit specifications. For traditional logit models, covariates are often evaluated using specifications that involve interaction terms in which one or more covariates (*e.g.,* gender) are interacted with one or more attributes (*e.g.*, price). If the resulting coefficient(s) pass significance tests, one cannot reject the hypothesis that members of the covariate group express different utilities for the interacted attributes. Formally, a traditionally specified logit model with covariates looks like:

$$P(y_{is} = k \mid z_{is}) = \frac{e^{V_{k|z_{is}}}}{\sum_{k'=1}^{K} e^{V_{k'|z_{is}}}} \qquad \text{Eq. 13}$$

where

$$V_{k|z_{is}} = \sum_{a=1}^{A} \beta_a^{att} z_{iask}^{att} + \sum_{b=1}^{B} \beta_b^{\text{cov}} z_{ibsk}^{\text{cov}} + \sum_{c=1}^{C} \beta_c^{i} z_{icsk}^{att} z_{icsk}^{\text{cov}} + \varepsilon . \qquad \text{Eq. 14}$$

By contrast, LCDC models don't allow for interactions between covariates and attributes because of the separate and distinct role each plays in specification (3). In the LCDC specification, covariates are used to model the probabilities of membership in each latent class. The latent classes themselves are groups of respondents that share a common set of utilities expressed in their selections among alternatives with different attributes. The function that was served by using interactions between covariate and attribute terms in the traditional specification is served by latent classes in the LCDC specification.

3.4.5.1.2 Estimation

Latent-class models assume that a sample population consists of discrete segments, each of which is characterized by a separate logit model relating participation to a set of independent variables, and within each of which the IID assumption[18] holds. Not only are these models less restrictive than aggregate logit, they also can reveal insights into marketing strategy that aggregate models miss. In essence, they assume that individual tastes are homogeneous within classes but heterogeneous between classes. Marketing executives will recognize this as the underlying premise for the theory of market segmentation.

Latent-class choice models describe relationships between a number of elements such as program attributes, covariates that describe individual respondents, and segment membership. For each segment a logit model relating product attributes to program participation is estimated, while simultaneously calculating at the individual level probabilities of membership in each segment. Using covariate values, separate logit models are estimated concurrently to predict membership in each segment. This entire process is repeated for different segment counts, assuming that the total number of segments is 1, 2, 3 … n, and the "best" model is chosen based on the calculated Bayesian Interaction Criterion (BIC) for each iteration. Significance tests are applied to each parameter, as are Wald tests for equality of parameters across all segments.

Estimation results lend themselves to placement in a simulator that lets program managers vary program characteristics and gauge the impact on program participation.

We tested a number of attributes hypothesized to affect program participation. The following table lists these attributes, including the variable abbreviations used in the model summary later in this section and the description of attributes and levels provided to respondents. The table also lists interaction variables that were included in the model but because they were derived from manifested attributes were not shown to respondents.

Table 3-17
Variables Available for Latent-Class Discrete Choice Models

Category	Variable Mnemonic	Description provided to respondents
Inclination not to participate	None	None of these
Delivery mode	DlvrMode	Ways to provide information about your home and appliances and get recommendations for saving energy
	Mail	You complete a paper questionnaire about your home, appliances, & energy habits & mail it to your utility. You receive recommendations in the mail.
	Online	You complete an online questionnaire about your home, appliances, & energy

[18] Independent variables are assumed to be Independently and Identically Distributed (IID), such that the off-diagonal elements of the variance/covariance matrix are zero.

Category	Variable Mnemonic	Description provided to respondents
		habits & receive recommendations online.
	In-home	An expert comes to your home to record information about your home, appliances and energy habits. You will receive a printed report with recommendations.
Time required for you to complete audit	TmRqAdt	The amount of your time it takes for you to provide, for the audit, the information about your home, appliances, and energy habits.
Additional services		Services in addition to the recommendations you always get with an energy audit
Post-audit technical support	PsTcSppr	You can call your utility and talk to someone to help you understand the audit recommendations and how to implement them.
Usage profile	UsgPrfl	A graph that shows how much of your utility bill goes to each type appliance based on audit information. This requires that you provide your utility account number in the audit.
Links to vendors	Lnk2Vndr	The amount of help your utility will give you in finding retailers and contractors to install equipment recommended by the audit.
	No links	You must find your own retailers and contractors to install equipment recommended by the audit.
	National outlets	Audit recommendations would come with suggestions about which national retailers will usually carry the recommended equipment.
	Local vendors	Audit recommendations would come with specific information about reputable local retailers that carry the recommended equipment.
Level of detail & accuracy in savings estimates	LvlOfDtl	There can be a lot of variation in the accuracy and detail in the estimated savings that the audit provides with each recommendation.
	No savings estimates	You get general recommendations and tips for changing appliances and energy habits but without any estimates of kWh or therm savings.
	Rough estimates of savings	You get recommendations for changing appliances and energy habits along with estimated savings experienced by households similar to yours.

Category	Variable Mnemonic	Description provided to respondents
	Refined estimate of savings	You get much more accurate estimates of kWh or therm savings for each recommendation, based on your billed energy use. This requires that you provide your utility account number as part of the energy audit.
Audit fee or incentive	Fee (Range: We give you $30 to You pay $75)	You could be asked to pay for an energy audit, or the utility could offer you an incentive to complete one, or the audit could be offered free of cost or incentive.
Time between decision to participate & results	Tim2Rslt	The time between when you request an energy audit and when you receive the recommendations.

Interaction variables		*[The following were not shown to respondents]*
	Subsidy	Binary variable where 1 = Fee < 0 and 0 = Fee >= 0
	NatLnSub	Subsidy x National outlets
	OnLinSub	Subsidy x OnLine
	MailSub	Subsidy x Mail
	MailPsT	Mail x Post-audit technical support
	UsgPrLcl	Usage profile x Local vendors
	LclVnSub	Subsidy x Local vendors
	PsTcSUsg	Usage profile x Post-audit technical support
	MailUsg	Usage profile x Mail
	OnLinRgh	Rough estimate of savings x OnLine
	PsTcSRgh	Rough estimate of savings x Post-audit technical support
	PsTcSRef	Refined estimate of savings x Post-audit technical support
	OnLinFee	Fee x OnLine
	InHomFee	Fee x In-home

Parameters of the LCDC model are estimated using a combination of EM (expectation-maximization) and the Newton-Raphson variant of Maximum Likelihood (ML) estimation. Random points are chosen as seeds to begin the estimation process. EM is typically chosen to get the estimates close to a solution. Then the algorithm switches to ML to speed up convergence and provide estimates for parameter variances. Wald tests are conducted for parameter significance and equality across classes. Separate Wald statistics test the significance of covariates.

3.4.5.2 CART

The CART method was used in this segment of the study as well (see Section 3.4.1.1.1 for a description of the method). In the case of the stated preference study, the best method for

determining market segments, or classes, was the LDCD approach. The best use of CART for this section of the study was to use it to determine the customer characteristics associated with each class. Specifically, it was used to see what clusters of customer characteristics distinguished each class from the others. While LCDC is capable of completing this kind of task, in this particular case, CART did a better job. This is largely because of the ability of CART to find a different place for a given variable in different parts of the tree, i.e., interactions can be better and more efficiently captured in CART than in other methods. The data for this part of the study came entirely from the LCDC survey instrument.

4 Results

4.1 Verification of Completed Audits

4.1.1 Mail-In Audit

We reviewed the available program-tracking databases to determine the extent to which utilities met their respective targets for both Mail-In and Online Audits. Table 4-1 and Table 4-4 present the targets, the numbers reported in utilities' fourth-quarterly reports, and the number of audits that we were able to verify through our review of the program-tracking databases.

Table 4-1
Verified Completed Mail-in Audits,
by Utility Fourth Quarter Report, by Target

Utility	Targets	Fourth Quarter Report	Verified	Verified as Percent of Fourth Quarter Report	Verified as Percentage of Goal
PG&E	29,000	43,245	42,465	98%	146%
SCE	18,000	26,515	25,917	98%	144%
SCG	6,000	7,694	9,222	120%	154%
SDG&E	8,000	7,824	8,066	103%	101%
Total	61,000	85,278	85,670	100%	140%

Based on the verified results, all four utilities exceeded their Mail-In Audit goals. The average across all four utilities was 140 percent.

4.1.2 Mail-In Outreach Effort

The four utilities distributed 510,228 mailed audits to residential households. Table 4-2 presents the number mailed out, the date of the mailing, and the language in which the audit was written.

Mail-In Audit acceptance or "take" rates, defined as the number of verified Mail-In Audits divided by the number of audits mailed directly to customers, are presented in Table 4-3. Note that the verified audits in Table 4-3 are slightly less than those in Table 4-1 since some households chose to participate even though they did not receive a mailer.

Table 4-2
Direct Mailing, by Utility, Language, and Date

Utility	Number Mailed Out	Date of Mailing
PG&E Mailing #1	13,000 Chinese	3-Mar
	13,000 Spanish	
PG&E Mailing #2	50,000 English	3-Apr
PG&E Mailing #3	99,613 English	3-May
PG&E Mailing #4	49,613 English	3-Jul
Total PG&E	**225,226**	
SCE Mailing #1	50,000 English	3-Feb
SCE Mailing #2	50,000 English	13-Jun-03
SCE Mailing #3	25,000 Chinese	11-Jul-03
	25,000 Spanish	
SCE Mailing #4	50,000 English	25-Aug-03
Total SCE	**200,000**	
SoCal Gas Mailing #1	25,000 English	22-Oct-03
SoCal Gas Mailing #2	25,000 English	23-Oct-03
Total SoCal Gas	**50,000**	
SDG&E Mailing	35,002 English	16-Oct-03
Total SDG&E	**35,002**	
Grand Total	**510,228**	

Table 4-3
Acceptance Rates for the Mail Audit, by Utility

Utility	Direct Mail Pieces	Mailing-Related Verified Audits	Acceptance Rate
PG&E	225,226	40,676	18.1%
SCE	200,000	24,109	12.1%
SoCalGas	50,000	7,696	15.4%
SDG&E	35,002	5,248	15.0%
Overall	510,228	77,729	15.2%

4.1.3 Online Audit

Table 4-4 presents the Online Audit goals, the number claimed in the 4th quarter report and the number of verified Online Audits, by utility.

Table 4-4
Verified Completed Online Audits,
by Utility Fourth Quarter Report, by Target

Utility	Targets	Fourth Quarter Report	Verified	Verified as Percent of Fourth Quarter Report	Verified as Percent of Goal
PG&E	12,000	14,848	14,848	100%	124%
SCE	12,000	16,513	15,729	95%	131%
SCG	N/A	104	N/A	N/A	N/A
SDG&E	N/A	271	N/A	N/A	N/A

Regarding the Online Audit, both PG&E and SCE exceeded their goals by 24 percent and 31 percent, respectively. Recall that both SoCal Gas and SDG&E did not establish Online Audit goals.

4.1.4 In-Home Audit

SCE had a target of 4,500 In-Home audits for 2003. In its 4th quarter report, it reported that it had completed 5,362. We were able to verify 99.7 percent, or 5,348, which is 18.8 percent greater than their original goal.

4.2 Verification of the Hard-to-Reach Goals

Table 4-5 presents the results of the HTR analysis for all four utilities.

Table 4-5
Direct Mail Solicitations and HTR Achievement, by Utility

Utility	Direct Mail Solicitations	Percentage Mailed to Hard-To-Reach
PG&E	225,226	85.50%
SCE	200,000	86.20%
SoCalGas	50,000	84.90%
SDG&E	35,002	63.00%
Total	510,228	84.17%

Using the restrictive definition of an HTR zip code, all four utilities exceeded their goal of 50 percent.

With respect to SCE's In-Home Audit Program, our analysis of the extracts from the program tacking database revealed that 72.1 percent of the 5,362 completed audits were in HTR zip codes.

4.3 Improvement of Target Marketing

We began the analysis by estimating a number of logistic regression models. However, when the results of the logistic regression models proved disappointing, we attempted

another approach. Below we discuss the logistic regression results and then proceed to introduce the results of the alternative approach.

4.3.1 Logistic Regression Models

Although each utility model was estimated with different source data, the final specifications turned out to be remarkably similar, with cubic relationships between energy usage and program participation. In addition, the models shared some independent variables. Table 4-6 presents these results for each utility.

Table 4-6.
Parameter Estimates for Logistic Regression Models
(P-Values in Parentheses)

Parameter	PG&E	SCE	SCG	SDG&E
Intercept	-1.824 (0.000)	-0.997 (0.001)	-2.487 (0.000)	-1.248 (0.000)
Avg kWh (therms) consumed per day	1.707 (0.000)	-6.756 (0.000)	0.685 (0.002)	-0.085 (0.012)
Avg kWh (therms) consumed per day^2	-2.583 (0.000)	6.313 (0.000)	-0.222 (0.001)	0.006 (0.026)
Avg kWh (therms) consumed per day^3	0.276 (0.000)	-1.787 (0.000)	0.015 (0.007)	-0,000 (0.034)
Population % 65 and older*	0.891 (0.000)			
Population % female*			0.983 (0.011)	
Population % African-American*			1.060 (0.001)	
Population % Asian*			1.009 (0.018)	
Asian population < 1*			0.114 (0.002)	
Asian population = 3% - 8%*		-0.040 (0.014)		
Population % white*	0.461 (0.000)			
Hispanic population <= 8%*		0.975 (0.000)	0.661 (0.000)	
Hispanic population 8% - 17%*		1.029 (0.000)	0.714 (0.000)	
Hispanic population 17% - 37%*		0.916 (0.001)	0.664 (0.000)	0.092 (0.011)
Hispanic population > 37%*		0.934 (0.001)	0.508 (0.001)	
Median home value*†	-0.760 (0.000)			

Parameter	PG&E	SCE	SCG	SDG&E
Minimum occupancy*				-0.474 (0.000)
Percentage of homes built before 1950*				-0.494 (0.010)
Percentage of homes built before 1960*	-0.190 (0.026)	-0.128 (0.000)		0.439 (0.018)
Percentage of homes built before 1970*				-0.417 (0.000)
Percentage of homes built before 1980*				0.316 (0.004)
Percentage of homes headed by female only*		0.611 (0.000)		
Percentage of homes occupied by renters*		-0.488 (0.000)		-0.243 (0.015)
Percentage of homes occupied by renters: 11%-28%*			0.098 (0.006)	0.109 (0.010)
Percentage of homes occupied by renters: 28%-62%*		-0.165 (0.000)	0.153 (0.000)	
Percentage of homes heated by electricity		0.330 (0.000)	-0.437 (0.018)	
Mean family size*		-0.088 (0.000)	-0.449 (0.000)	
Percentage of households with 3 – 5 persons*			-1.272 (0.000)	
Mean household income*			0.376 (0.000)	
Median family income*		-0.000 (0.019)		
Climate zone 9		-0.091 (0.000)		
Climate zone 10		0.102 (0.000)		
Climate zone 13		0.325 (0.000)		
Climate zone 14		0.194 (0.000)		
Mode income: $0 - 27		-0.209 (0.000)		
Mode income: $28 – 36		-0.076 (0.009)		
Mode income: $43 - 48		0.049 (0.004)		
Model p-value	0.000	0.000	0.000	0.000
McFadden's Rho-Squared	0.009	0.017	0.008	0.004

*Data aggregated at Block level †Rescaled

While the fit of each of these models appears reasonable and they all pass significance tests, the Rho-squares are low. In addition, from Table 4-7, one can see that the prediction success proved disappointing for all four utilities.

Table 4-7.
Predicted Success Tables for Logit Models
(P-Values in Parentheses)

Measure	PG&E	SCE	SCG	SDG&E
Actual number of participants	4,002	22,216	6,994	3,755
Predicted number of participants	847	2,969	1,128	576
Actual number of non-participants	15,517	159,140	38,055	21,166
Percentage correct (sensitivity)	21.2%	13.4%	16.1%	15.3%

While we felt the sensitivity of these models was passable, we concluded that they were not sufficient to render the models useful in a commercial setting. We speculate that this circumstance arose from one or more causes:

1. Distributional assumptions for the error term were violated. Perhaps the assumption that the error term followed an extreme-value type 2, of Gumbel distribution was too presumptuous for this dataset.
2. Data measurement problems at the individual level made it difficult to obtain good estimates. Much of the independent data came from Block Group-level Census data, which are by their very nature aggregated to protect individual privacy. Associating aggregate (and in some cases spotty) data to individuals introduces a source of error that can reveal itself in model estimation.
3. Estimates suffered from known problems associated with the use of logistic regression in modeling rare-events data. King and Zeng (2001) have shown that for a variety of reasons, logistic regression produces distorted estimates when used to model rare-events data. The overall participation rates of 15% or less across our datasets do not strictly match the criteria of <5% to be designated rare-events, but they are low enough to raise the possibility of being influenced by the distorting effects.

The disappointment with this approach led us to explore an alternate model structure for our revealed-preference data.

4.3.2 Share Regression Models

To get around the presumed shortcomings of modeling revealed-preference data at the individual level, we decided to aggregate the data by energy consumption. We averaged both dependent and independent variables by energy consumption at 500 kWh and 50 therm increments, noting that averaging dummy-coded participation was the same as calculating the average participation rate for each energy consumption range. Once the data were transformed, we produced the following plots of audit participation rate by annual kWh consumption category. These plots are presented in Figure 4-1, Figure 4-2, Figure 4-3, and Figure 4-4.

Figure 4-1
PG&E Audit Participation Rate by Annual kWh Consumed

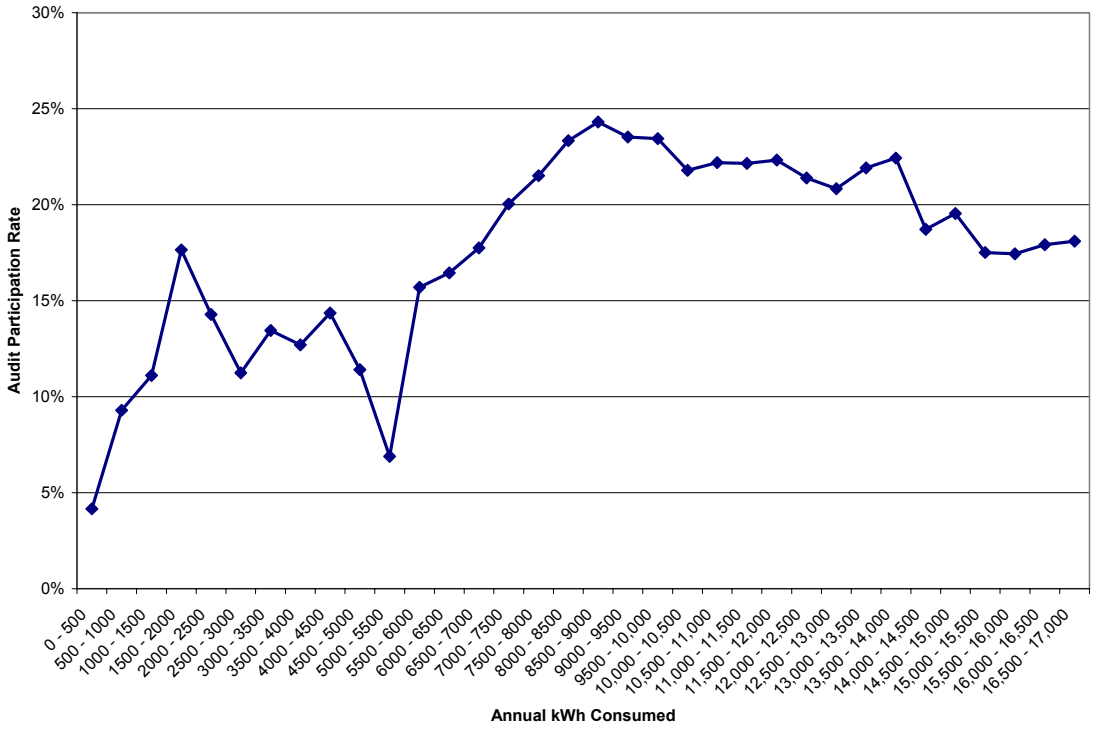


Figure 4-2
SCE Audit Participation Rate by Annual kWh Consumed

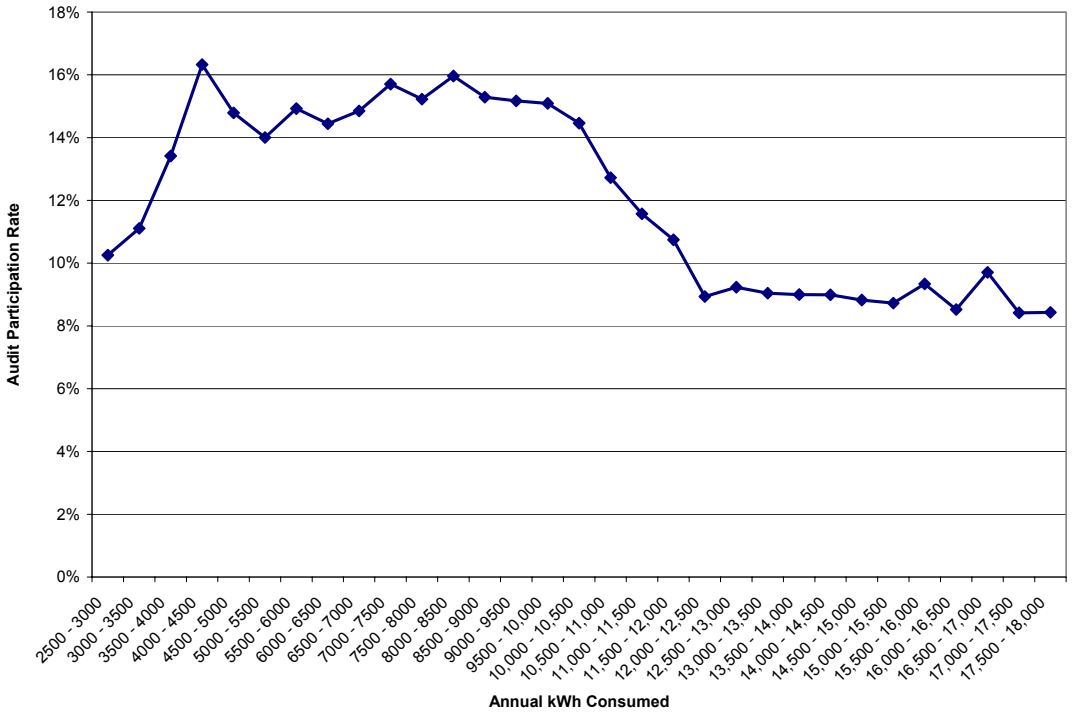


Figure 4-3
SCG Audit Participation Rate by Annual Therms Consumed

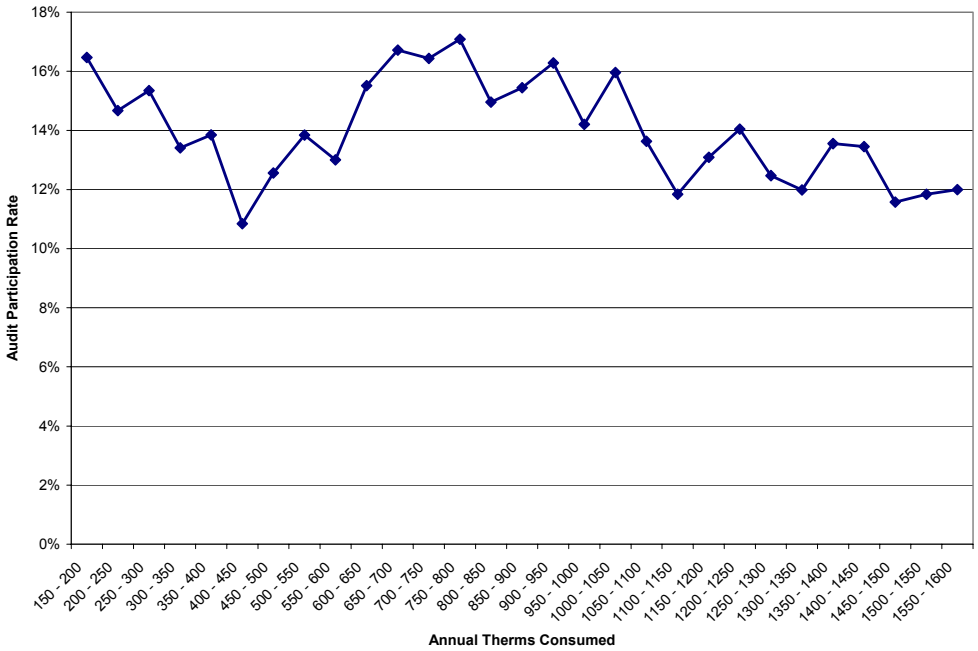
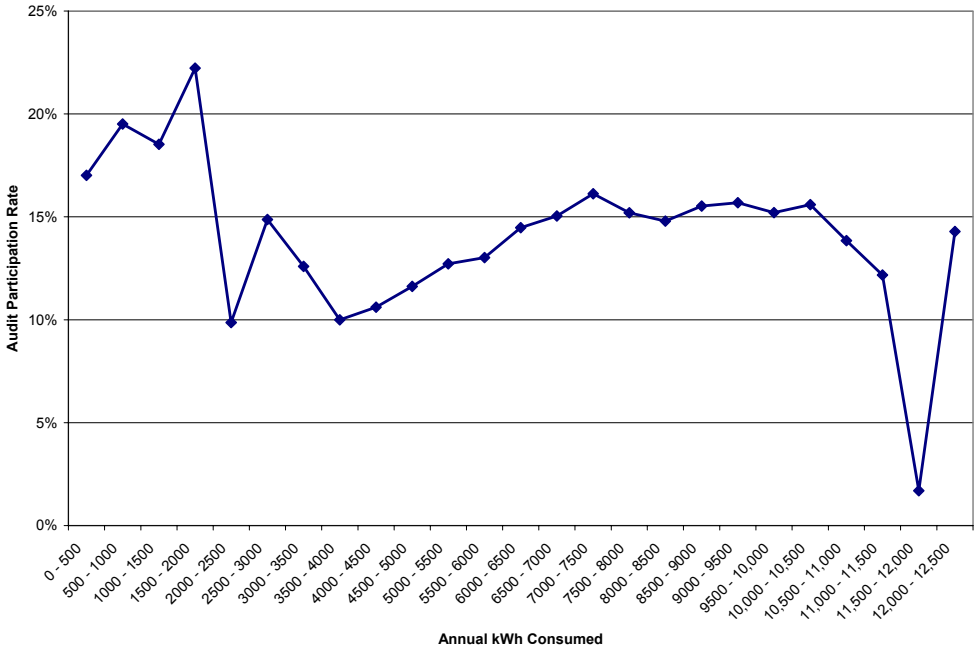


Figure 4-4
SDG&E Audit Participation Rate by Annual kWh Consumed



Notice that the kWh data for PG&E, SCE and SDG&E are distinct but share some common characteristics. Participation rates peak at relatively low levels of consumption, retreat as consumption rates increase, rise to a plateau at higher rates of consumption, then trail off at very high consumption rates. The magnitude and specific levels at which these changes occur vary among utilities, but the commonalities are unmistakable. It is not surprising, therefore, that models for the three electric utilities shared a common cubic form. Most important, these findings appear to contradict the common assumption among some utilities that the higher the energy use the greater the probability of participation. *That is, participation as a function of energy use appears to be a much more complex relationship*.

We used OLS to estimate the relationship between audit participation rate and a host of independent variables. This approach satisfied the concerns we had with logistic regression models, in that:

1. Distributional assumptions for the error term appeared to be satisfied. An assumed normal distribution for the error term is less heroic than an assumption of an extreme-value distribution.
2. Data measurement problems at the individual level were mitigated through aggregation. Distortions arising from matching Block Group-level aggregates to individuals are reduced when individual-level data are aggregated.
3. Estimates suffered from known problems associated with the use of logistic regression in modeling rare-events data. Converting rare-events to aggregate percentages eliminated this problem.

4.3.2.1 Results

Model fits were good for SCG and exceptional for the other three utilities. As occurred with the logistic regression attempts, our multiple regression models shared both specifications and certain independent variables. The following table summarizes the parameter estimates.

Table 4-8
Parameter Estimates for Regression Models on Aggregated Data
(P-Values in Parentheses)

Parameter	PG&E	SCE	SCG	SDG&E
Intercept	1.176 (0.000)	0.487 (0.001)	1.254 (0.000)	-8.988 (0.002)
Avg kWh (therms) consumed per day	0.237 (0.016)	0.876 (0.000)	0.096 (0.007)	-0.571 (0.000)
Avg kWh (therms) consumed per day^2	-0.268 (0.005)	-0.746 (0.000)	-0.050 (0.053)	1.199 (0.000)
Avg kWh consumed per day^3	0.074 (0.017)	0.197 (0.000)		-0.681 (0.000)
Median home value*	-1.146 (0.000)			
Mean household size*	-0.191 (0.000)			

Parameter	PG&E	SCE	SCG	SDG&E
Median household income*		-0.700 (0.000)		
Percentage of families living in poverty*	-0.921 (0.019)	-0.740 (0.000)		
Population % White*				1.954 (0.000)
Population % Asian*	0.556 (0.000)			2.918 (0.000)
Population % Hispanic*				2.219 (0.000)
Percentage of homes heated with electricity*	0.878 (0.000)	1.193 (0.000)		
Percentage of households headed by female only*		1.112 (0.000)		
Percentage of population graduating high school*		-1.977 (0.000)		-1.563 (0.030)
Percentage of population with some college*				-1.161 (0.020)
Percentage of population completing BA degree*	-1.225 (0.000)		-1.169 (0.039)	
Percentage of population completing graduate degree*			-4.019 (0.000)	-2.142 (0.012)
Percentage of single-family detached homes*				8.367 (0.004)
Percentage of single-family attached homes*				6.457 (0.019)
Percentage of multi-family homes with <5 units*				5.808 (0.011)
Percentage of multi-family homes with 5+ units*				9.215 (0.003)
Percentage of mobile homes*				8.158 (0.008)
Percentage of population in climate zone 13*		-0.447 (0.004)		
Percentage of population in climate zone 14*				1.155 (0.002)
Percentage of population in climate zone 15*			-0.249 (0.011)	
Percentage of population in climate zone 16*		-0.652 (0.001)		
Dummy variable for 5000-5500 kWh	-0.104 (0.000)			
R^2 (coefficient of determination)	0.976	0.984	0.754	0.993
R^2adj. (adjusted coefficient of determination)	0.966	0.976	0.686	0.982

*Census Block Group-level data

Adjusted R-squares (a statistic that can range from 0 to 1 and represents the percent of the variance explained) for the electric utilities were 0.96 or better, suggesting that 96 percent or more of the variance in average audit participation rates was explained by the model. For SCG, the comparable figure was 69 percent. A comparison of actual vs. fitted data shows how well the models track the non-linear portions of the participation curve, suggesting that the goodness of fit was not just in matching an overall trend.

Figure 4-5
PG&E Actual vs. Fitted Data

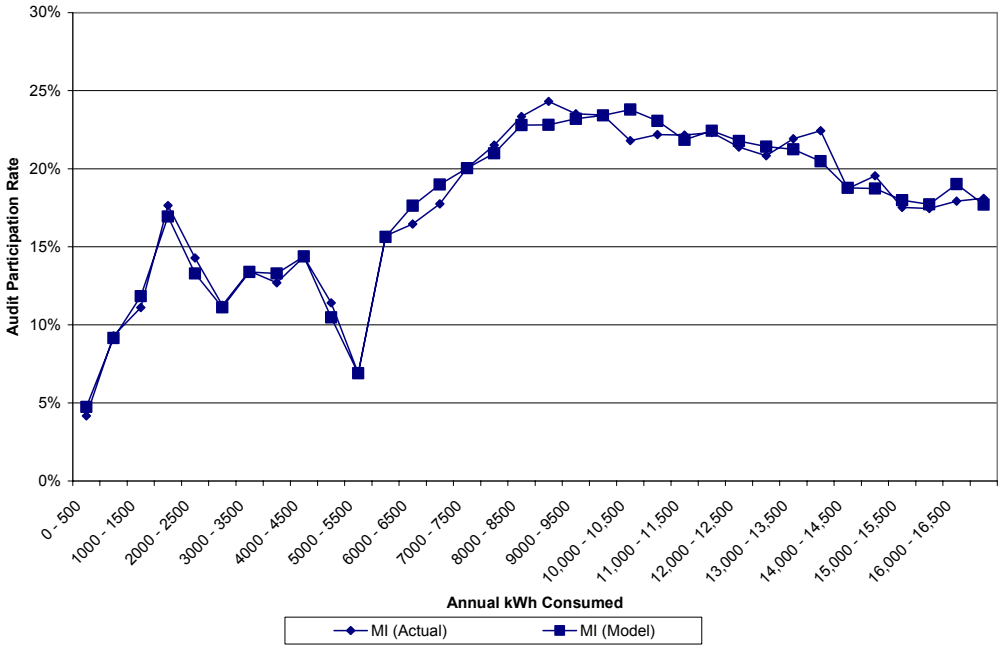


Figure 4-6
SCE Actual vs. Fitted Data.

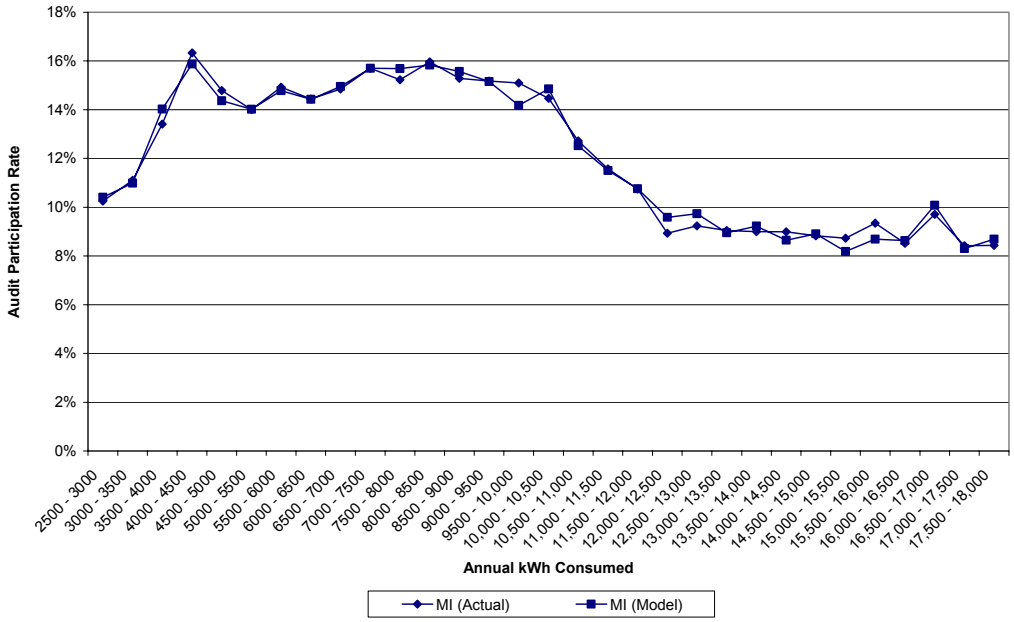


Figure 4-7
SCG Actual vs. Fitted Data.

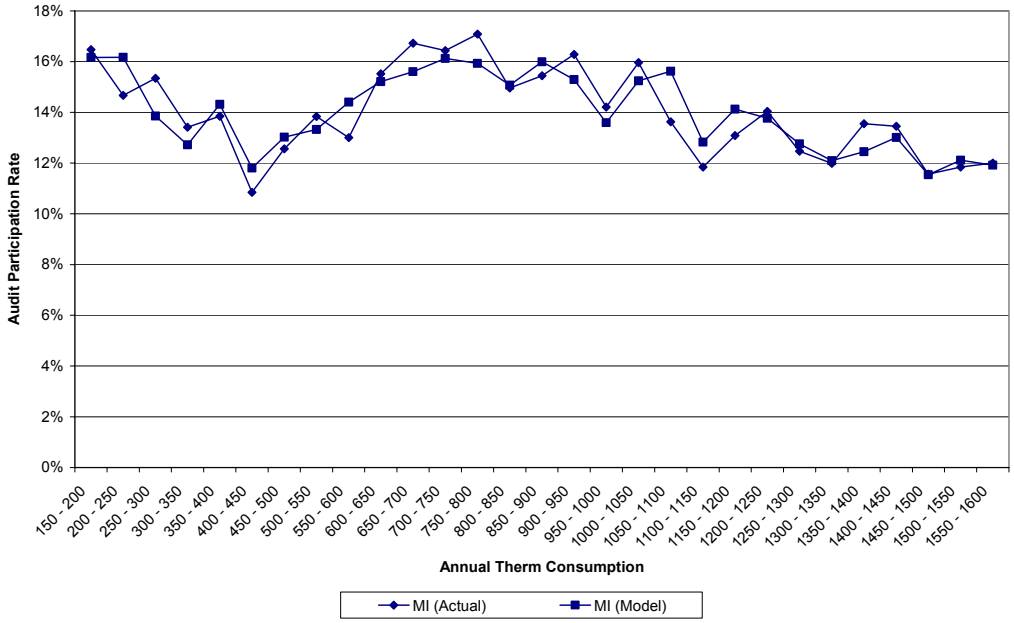
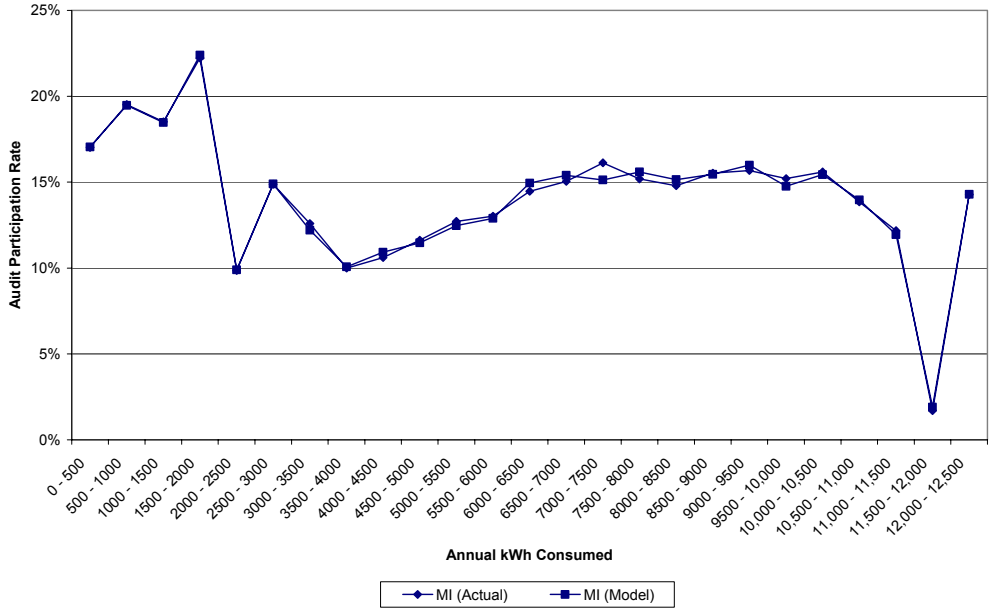


Figure 4-8
SDG&E actual vs. fitted data.



A check of standardized residuals revealed no obvious influential observations. On the contrary, the models do a particularly good job of capturing peaks and valleys, i.e., turning points, in the data. The model for therms did not enjoy the exceptional fit of the kWh models, but it was nevertheless quite acceptable.

4.3.2.2 Application

Because of the aggregations used to develop the model, care should be taken in applying it to external data. Our concern is based on the fact that a given Block Group-level demographic variable assigned to a given household will at times be far outside the range for that variable in the aggregated model. In such a situation, the prediction error can be quite large. With this concern in mind, we suggest following these steps in forecasting participation at the individual level:

1. Obtain an estimate of the individual's annual energy consumption in therms or kWh as appropriate.
2. Assign the individual to the appropriate energy use category.
3. Based on the location of the individual's residence, retrieve Block Group-level data for the Census-based independent variables.
4. Retrieve the climate-zone proportions for the energy-usage category from the Block retrieved in step 3.
5. Calculate the individual's projected program participation using the parameters from the regression model.

We also developed an Excel-based simulator for each utility to illustrate use of these models. However, due to our concerns about selecting input values for the variables for use in the simulator that, while quite reasonable, are nevertheless outside the range on which the original model was estimated, we have chosen not to rely on these models. While we have learned a great deal from these models (e.g., the relationship of annual energy use to participation and problems associated with relatively rare events), we have chosen to rely on the CART models for predicting the likelihood of participating. These CART models are discussed in the next section.

4.3.3 Revealed Preference CART Results

For the CART portion of the revealed preference component of the study, three major questions are addressed. First, the degree of predictive success achieved by CART, second, the significance of the Acxiom variables in predicting participation, and last, a description of the variables associated with the decision to participate or not.

Following this, we will present two other activities. The first is a procedure to cross-validate two of the model results. Cross-validation is possible for the PG&E and SCE trees because the models were developed on only 30,000 of more than 200,000 customers who received direct mail solicitations. Therefore, the models can be tested on additional portions of those direct mail recipients.

The final endeavor in this section was to apply the CART tree results to new, random samples of residential customers. That effort is described last.

Table 4-9 shows the degree of prediction that resulted from the CART models for each utility. In the left section of the table, one can see the results of running CART models using the Census and utility variables only. The percent correctly predicted among the participants varies between 67 percent and 74 percent, and for the non-participants, between 44 percent and 62 percent. These results could be stated another way: of all the actual participants, between 67 and 74 percent were classified in participant nodes by CART. Conversely, of all the known non-participants, between 44 and 66 percent were correctly classified in non-participant nodes. While it may be most important to predict participants correctly, it may also be important to determine who are highly unlikely to participate since marketing to them may be a waste of resources. Alternatively, unlikely participants could be a focus of more intense marketing efforts on the assumption that they are hard-to-reach customers.

The best rate of correct prediction for participants (74 percent) was found in the SDG&E territory, but it took 223 terminal nodes to achieve this. If we were trying to understand, conceptually, who these participants are, 223 terminal nodes would preclude that. That is just too much complexity to understand. However, if we are interested strictly in prediction, it doesn't matter how many nodes there are as long as we can use the tree to predict, mechanically, a new set of customers who have not been solicited.

The lowest prediction accuracy for participants is seen in the SCG and PG&E territories, at 67 percent, although it took only 8 and 6 nodes, respectively, to arrive at this, making them very efficient models. SCE had a good prediction rate at 71 percent for participants, and 50 percent for non-participants.

Table 4-9
Success of CART Trees in Predicting Participation by Utility

	Trees Excluding Acxiom Variables		Trees Including Acxiom Variables	
Utility	**% Correctly Predicted (Partic/NonPart)**	**# of Terminal Nodes**	**% Correctly Predicted (Partic/NonPart)**	**# of Terminal Nodes**
PG&E	67/43	6	n/a	n/a
SCE	71/50	14	75/50*	31
SCG	67/44	8	n/a	n/a
SDG&E	74/56	223	n/a	n/a

* Excludes Census variables of education, dwelling type, and years in dwelling
Note: Runs that included Acxiom variables as potential splitters also included Claritas cluster variables. However, these variables did not enter the resulting trees as splitters.
Note: the % correctly predicted numbers are not directly comparable to the sensitivities reported in Table 4-7. Such comparisons are not really practical as different methods produce different statistics.

It is worth noting that the tree results presented in Table 4-9 for PG&E is not the first tree grown. An earlier tree predicted 70 percent of the participants correctly, and 62 percent of the non-participants. However, it was a very complex tree at 226 terminal nodes. A tree of this complexity, even with 30,000 cases, is in danger of showing sample-specific results that cannot be replicated on other samples. Since the 30,000 cases that were the basis for the original tree did not exhaust the PG&E pool of customers, it was possible to apply the model to another sample of 30,000 to check the model's predictive accuracy (Section 4.3.3.2 describes this in detail). The fears about replicability were confirmed, so a third sample of 30,000 was used to grow another tree, hoping to generate a simpler model, and the tree reported here is the result of that effort. If it were possible to generate another model for SDG&E we would have done that, but there was not enough sample to do it. We do think that it is still worth using the SDG&E model, though, because based on the PG&E experience, even with reduced accuracy, improvements on current take-rates would probably occur.

4.3.3.1 Significance of Acxiom Variables

A test of the predictive importance of individualized demographic characteristics rather than Census Block Group-level demographics was conducted using Acxiom data. It should be said that both Acxiom and Claritas segment/cluster variables were available, but the Claritas variables never entered any tree. So, the comparison really just considers how much improvement can be made by the demographic variables that are measured differently, and more precisely, which is represented by the census Block Group-level of measurement of the Census data versus the individual level of measurement in the Acxiom data. This could only be tested on the SCE dataset as that is the only utility for which the latter variables were available.

A tree using the individual-household level variables produced correct prediction of 75 percent of the participants (compared to 71 percent using Block Group-level variables), although prediction of non-participants was not improved. CART required 31 terminal nodes to reach this level of prediction. This represents an increase in accuracy, but not a striking one.

The CART models, overall, did quite well in predicting participation and, to a lesser extent, non-participation. The issue of why the models do a less effective job of pinpointing non-participants is an interesting question. Our hypothesis is that some of the utilities have already done some work to understand who will or won't participate, and have tailored their mailings accordingly. CART (or any other modeling technique) would probably predict those customers' decisions better if the full range of customer interest had been represented in the samples. Perhaps it becomes difficult for a model to distinguish non-participants when a large proportion of them have already been eliminated from the pool. There would be value in an experiment to see how effectively CART models would predict among a random sample of customers.

4.3.3.2 Variables Predicting Participation

When using data mining techniques like CART, a distinction must be made between models that predict best and those whose content is understandable. If one is interested strictly in predicting individuals' behavior, and if the use of the model in that prediction is mechanized, the complexity of the tree isn't important. If understanding is a goal, then complexity is very important, as it is very difficult to make meaning of a tree with, e.g., 223 terminal nodes. In this study, it is probably most important to have predictive accuracy, but we want some understanding as well. Therefore, the approach taken here tries to balance those needs. For purposes of prediction, SAS code for the best tree in each utility will be provided upon request. A sample of SAS code for one of the most efficient trees is provided in Appendix H of this report.

Part of the CART output is a list of all variables available to the model, together with relative importance weights. The most important variable is given a weight of 100, and all of the other variables are given weights relative to that, based on their use and effectiveness in purifying the participant/non-participant groups. The importance weights are based on the improvements in predictive probabilities at each stage of the model. After each split, an improvement score is generated that is based on the change in node probabilities from one stage to the next. These improvement scores are tied to the splitters at each step, and a splitter's improvement scores are summed across the entire tree to determine its importance. The splitter with the highest total improvement score is assigned an importance weight of 100. All other splitters are given importance weights that reflect how much smaller their improvement scores are compared to the highest-importance splitter. Importance scores can be based on improvements tied only to primary splitters, or than can be based on primary and surrogate splitters' improvement scores.

The reader is reminded that at each stage the most efficient splitter is the primary splitter, the one that appears in the tree nodes. In addition, the less effective splitters (sometimes only slightly less effective) are recorded in the output, and are called surrogates. Surrogates are only those variables that act similarly to the primary splitters, and are used as back up to the primary splitters when they have missing values.

In this section, we present a series of tables that shows which variables are the important splitters, together with their relative importance weights. It is important to recognize that these tables don't inform us about which values of the variables predict participation or non-participation; the reader must study the trees to see that. This inability to show the form of the relation between splitter and the participation variable actually reflects an advantage of this method: it does not require a variable to behave in the same way for every subgroup of customers. Thus, its relation with participation can be different in different parts of the tree. Therefore, these tables can only show which variables it is important to measure for each customer if this method of targeting customers is to be used. Also note that to save space, variables with importance weights of less than 10 are not shown.

Importance weights in Table 4-10 are shown for the tree grown for PG&E, but there are two sets of variables. The left side of the table shows all of the splitter variables appearing

in the tree, and their importance to the model, whether they were primary splitters or surrogate splitters. The right side of the table shows those variables (a subset of the first group) that function as primary splitters. Recall that the surrogates are only used when there are missing values in the primary splitters. It may be helpful to think of this table as showing the core variables and their importance on the right, and the variables that can be substituted if the core variables are not available or have missing values are shown on the left.

For PG&E, the major predictors are the annual kWh, county, the percentage of renters in the Census Block Group, and mean occupancy rate, with kWh and county being, by far the most important. Most of the other variables come into play mainly as surrogates when one of the primary splitters is missing.

It may avoid confusion to explain at this point that both Percent White and Percent Non-White appear as splitters on the left side of the table, and one might wonder what the difference is between these two variables. The answer is that there is no real difference; both act essentially identically in opposite directions, of course. If one comes in as a primary splitter, the other will certainly be a surrogate for it, though not very helpful since if you have one you have the other. The only difference between the two could be in the rounding error that might come from the way the variables are stored digitally.

Table 4-10
PG&E CART Tree Variable Importances

Importance of Primary and Surrogate Variables		Importance Based on Primary Splitters Only	
Variable	**Relative Import**	**Variable**	**Relative Import**
Annual KWh	100.00	Annual KWh	100.00
County	47.19	County	73.18
% Renters	31.54	% Renters	48.91
Mean Occupancy	18.09	Mean Occupancy	28.05
% Couples	16.29		
% Apts 5 or more units	12.31		
% White	11.97		
% Non-white	11.97		
% Latino	11.65		
% Single Family Detached	11.45		
Mean Rooms in Households	10.70		

The best tree for SCE included Acxiom variables (see Table 4-11); specifically, it is the demographic variables from Acxiom, not the market segments that entered the trees. The most important Acxiom variables are ethnicity, occupation, marital status and age. Of course the utility variables of kWh and tariff are essential predictors as well. Interestingly, with the exception of household size (2-person) and income, census variables, at the Block Group level, act only as surrogates for when there is missing information on the Acxiom demographics.

Table 4-11
SCE CART Tree Variable Importances: Acxiom Variables Included

Importance of Primary and Surrogate Variables		Importance Based on Primary Splitters Only	
Variable	**Relative Import**	**Variable**	**Relative Import**
Ethnicity: Acxiom	100.00	Ethnicity: Acxiom	100.00
Total KWh	90.93	Total KWh	92.17
Occupation: Acxiom	86.07	Occupation: Acxiom	88.97
Religion: Acxiom	64.73	Tariff Class	36.74
Tariff Class	56.54	Marital Status: Acxiom	32.12
Marital Status: Acxiom	31.92	% 2-Person Households	12.35
% White	20.46	Age: Acxiom	12.25
Mean Family Size	19.09	Median Household Income	11.70
% Non-white	16.90		
% 2-Person Households	16.81		
Median Household Income	16.75		
Median Family Inc	16.57		
CARE Rate	15.50		
Title 24 Climate Zone	15.19		
Household Size: Acxiom	15.13		
HTR Zip: Rural	14.52		
% Population Under 18	13.98		
Age: Acxiom	12.48		
% 6+ Person Households	10.20		

The SCE tree without Acxiom variables (see Table 4-12) performed only a little worse than the Acxiom tree. KWh was the most important splitter, followed distantly by tariff. Ethnicity enters the picture at a distant third position. Note that in this table, % White is a primary splitter and % Non-white is a surrogate. Consequently, the % White variable receives a higher importance score because it came in as a primary splitter.

Table 4-12
SCE CART Tree Variable Importances: No Acxiom Variables

Importance of Primary and Surrogate Variables		Importance Based on Primary Splitters Only	
Variable	**Relative Import**	**Variable**	**Relative Import**
Total KWh	100.00	Total KWh	100.00
Tariff Class	69.01	Tariff Class	47.05
HTR Zip: Rural	24.92	% White	19.91
% White	21.44	% Completed grad degree	9.99
% Non-white	20.06	HTR Zip: Rural	9.87
CARE Rate	19.13		
Median Family Inc	18.25		
% Latino	16.20		
% Less than 9th grade Ed	15.22		
Mean Occupancy	14.37		
Median Household Income	11.57		
% Completing grad degree	10.81		
Median Home Value	10.47		
% 6+ Person Households	10.14		

The SCG tree represented in Table 4-13 shows the two top splitters as ethnicity (although they represent just one concept: White versus Non-White), but both of those variables act only as surrogates since they do not appear on the list of primary splitters. The most important splitters are related to occupancy density: 2-person households and mean occupancy rate. Rate or tariff is also quite important, as is climate zone, although less so. Education appears as a primary splitter, although at a much lower level than household size, rate class, and climate. Interestingly, although the density of gas-heated homes appears as a primary splitter, annual therm consumption does not appear at all, even as a surrogate.

Table 4-13
SCG CART Tree Variable Importances

Importance of Primary and Surrogate Variables		Importance Based on Primary Splitters Only	
Variable	**Relative Import**	**Variable**	**Relative Import**
% Non-white	100.00	% 2-Person Households	100.00
% White	100.00	Utility Rate Class	56.26
Mean Occupancy	97.46	Mean Occupancy	51.84
% 2-Person Households	91.75	Title 24 Climate Zone	30.62
Mean Family Size	80.29	% Homes Heated with Gas	25.25
% 3-5 Person Households	62.66	% Completed BA	18.13
Utility Rate Class	51.50	Median Home Value	17.17
CARE Rate	38.60		
% Some college	32.51		
% Latino	31.38		
% Less than 9th grade Ed	30.72		
Title 24 Climate Zone	28.03		
HTR Zip: Latino	24.18		
% Homes Heated with Gas	23.11		
% Completed BA	16.59		
Median Home Value	15.81		
Median Family Inc	10.47		

The tree for SDG&E (see Table 4-14) is very large (223 terminal nodes). However, there are variables that are clearly the most important in predicting participation. Specifically, kWh is by far the most important predictor, almost twice the importance of the next most important (therms), at least among the primary splitters. After kWh, percent renters in the Census Block Group and the percent of Census Block Group residents who have completed high school occupy approximately equal positions. All other variables serve only as surrogates.

Table 4-14
SDG&E CART Tree Variable Importances

Importance of Primary and Surrogate Variables		Importance Based on Primary Splitters Only	
Variable	**Relative Import**	**Variable**	**Relative Import**
KWh	100.00	KWh	100.00
Annual Therms	47.51	Annual Therms	45.60
KWh Category	45.86	% Renters	16.24
Mean Rooms in Households	25.62	% Completed some HS	11.85
Median Family Inc	21.48		
% Renters	19.31		
Median Household Income	17.69		
% Completed some HS	16.86		
Mean Household Size	14.64		
Mean Family Size	13.83		
% Family Households	13.70		
% 3-5 Person Households	13.52		
% 2-Person Households	12.99		
% Moved betw 1999-2000	12.31		
% 6+ Person Households	12.05		
% Population 65 or Over	11.85		
Median Home Value	11.82		
% Apts 5 or more units	11.73		
% Latino	11.28		
% Less than 9th grade Ed	11.16		
Mean Occupancy	10.78		
% Couples	10.64		
% White	10.26		

How each of the variables described here acted on the decision to participate for each of these utility territories can be seen in the trees, which are contained in Appendix F[19]. We believe that this approach is quite promising for use in targeting populations for program solicitation. This method is ideal for working with low-frequency events, and in this case, it was working under less than ideal conditions. It was working on customer pools that have already been filtered for what the utilities have considered to be variables that increase the possibility of participation. We have seen from other sections of this report that the assumptions about what is important in this prediction have not always been correct. For instance, kWh consumption works in a complex way, not even close to linear in its relation to the probability of participation. So, to the extent that the assumptions underlying selection for participation have been incorrect, some promising groups may have been eliminated from the pool. To the extent that the prior assumptions made were correct, the least likely participants were eliminated, and this may have made accurate prediction more difficult. In spite of this, the models have performed quite well.

[19] To aid in interpreting the trees with respect to Title 24 weather zones, we have provided a map of these 16 climate zones in Figure H-1 in Appendix H.

4.3.3.3 CART Cross-Validation

Two of the participating utilities (PG&E and SCE) sent out so many direct mailers that it was necessary to take a random sample of 30,000 mailers to submit to the CART modeling procedure, due to the number of variables, and the limits of the CART license. The trees for those utilities are based on those samples of 30,000 customers. The other two utilities mailed few enough to use all in the modeling effort. However, because of the need to sample for PG&E and SCE, it is possible to test the models for those utilities on separate samples from the same base. The accuracy check is possible because we know the actual participants and can compare them to the predictions made by the model, which was built on a different sample.

A number of questions could be asked in this situation, but perhaps the most important would be what the rate of correct predictions was, especially for those predicted to be participants. Table 4-15 addresses this question.

Table 4-15
Comparison of Actual Participants in Original Sample Version the Cross-Validation Sample

	% of Customers in Participant Nodes Who Participated	
Sample	**PG&E**	**SCE**
Original sample of 30,000	21%	16%
Second sample of 30,000	20%	15%
% Decrease in Accuracy	5%	6%

Of course, we would not expect to see exactly the same accuracy in a second sample that we saw in the original sample on which the predictive model was built. In spite of the cross-validation that takes place during model building, there will always be a certain portion of predictive accuracy that is sample-specific. This is the case here as well. In the original PG&E sample, 21 percent of the customers in participant nodes actually participated. In the second PG&E sample, 20 percent actually participated, a 5 percent reduction (1 percentage point reduction [21-20] divided by the original 21 percent=5 percent). The picture is similar for SCE where the original rate of correct prediction within participant nodes was 16 percent, with a rate of 15 percent for the second SCE sample, a reduction of 6 percent.

Looking at accuracy from the other side, we can ask what percentage of the actual participants was correctly classified in participant nodes (this is the view presented in the original analysis). From this perspective, the results are similar: 65 percent of the participants in the PG&E program were correctly classified in participant nodes. This compares to 67 percent in the original sample, and a reduction in predictive accuracy of 3 percent. For the SCE sample, the original correct classification of actual participants into participant nodes was 71 percent, and for the second sample, it was 67 percent, for a reduction of 6 percent.

It should be noted here, as it was in Section 4.3.3.1, that another PG&E tree was grown that was extremely complex. It correctly classified 70 percent of participants, but took 226 terminal nodes to do so. With a tree this complex, it is very likely to be somewhat sample-specific, and this was borne out by the accuracy check with a second sample of 30,000. The reduction in accuracy was about 30 percent. It should be said, however, that even with this reduction, the take-rate would have been substantially improved over the current methods of targeting. In fact, it would have been a similar performance to the final tree that we have described here.

4.3.3.4 An Illustrative CART Application

Given that CART models were generated to determine whether "take-rates" (i.e., the percent of customers who, after receiving the Mail-In audit direct mailer, chose to participate) of the direct mail audits might be improved, we thought it would be helpful to provide a way for utilities to experiment with increasing take-rates by using the results of the CART trees to mail to new groups of customers. For this purpose we requested random samples of 3,000-5,000 residential customers from each utility that could be "scored" by the revealed preference trees. Three of the utilities, PG&E, SCE, and SDG&E provided these samples and each was scored by its own tree generated by the original sample of 30,000 direct-mail recipients. In the results presented below, we will be comparing some results from the new random samples to the old original samples. In this section of the report, we will refer to the original samples of 30,000 from each utility as the "original samples." The new samples will be referred to as the "random samples."

Before discussing the results of the scoring, it may be helpful to review some basics about the original trees based on the *original samples* followed by a description of what the scoring procedure produces for *new samples* that are subjected to them. Recall that when a tree is grown based on a sample where we know who actually was and was not a participant, each customer is assigned to a terminal node, based on the customer's characteristics that predict participation according to the splitting rules of the tree. Each terminal node is characterized as either a participant node or a non-participant node. Every customer is assigned to one type or the other. Participant nodes are so named because the percentage of customers in that node that were actual participants is higher than the overall take-rate of the entire group. For example, if the overall take-rate for a sample is 12 percent, then a terminal node that has a percentage of participants of 15 percent will be categorized as a participant node. Naturally, this means that some *actual participants* will be placed in *non*-participant nodes, and vice versa.

When scoring a new sample, where actual participation is not known, each customer in the sample is assigned to the node governed by the original tree's splitting rules. (Note that the percentage of customers occupying a particular node, say, Terminal Node 5, could be greater or less than the percentage of customers occupying that same node in the original sample.) This is because the second sample may [and was in this case] selected differently than the original sample. After being assigned to his correct node, each customer is assigned a probability of participating. This probability is based on the take-rate of the analogous terminal node from the original tree. So, if the percentage of actual participants in the original tree for Terminal Node 5 was 25 percent, the probability of

participation for all customers in that node for the new sample will be assigned a probability of participation of 0.25.

Now we are in a position to consider Table 4-16. One of the interesting comparisons in this table is between the random sample and the original sample on the percent of sample customers classified in participant terminal nodes. For PG&E, this comparison is 23 percent (line D) versus 13 percent (line B). This comparison is likely a reflection of the difference in samples; the same algorithm classified 13 percent of sample customers in participant nodes for one (screened) sample (original sample), but 23 percent in a different (random) sample. The analogous comparison for SCE is 78 percent versus 8 percent and for SDG&E, 43 percent versus 11 percent. These differences imply that a random sample identifies more high-probability customers (i.e., customers who have a high probability of participation) than existing methods. However, it isn't clear how much these differences are a result of ineffective targeting, and how much are due to deliberate targeting of hard-to-reach customers who are, almost by definition, less likely to respond.

Table 4-16
Comparison of a Random Sample with the Original Sample CART Results

Line	Analysis	PG&E	SCE	SDG&E
	Original Sample			
A	Overall Take-Rate	18%	12%	15%
B	% Customers in "Participant" Nodes	13%	8%	11%
	Random Sample			
C	Predicted Take-Rate for Total Sample	13%	14%	13%
D	% Customers in "Participant" Nodes	23%	78%	43%
E	Predicted Take-Rate in Participant Nodes* (weighted by # in Nodes)	30%	16%	23%
F	Predicted Take-Rate in "Non-Part" Nodes (weighted by # in Nodes)	7%	7%	7%

*This rate is calculated by assigning all random sample cases to their appropriate nodes based on the original sample splitting rules, then assigning the probability of participating to the cases in each node, based on the percent participation in the analogous nodes from the original sample. Finally, these probabilities are averaged over all random sample cases that occupy participant nodes.

Another comparison of interest in this table is between the overall take-rate from the original sample (line A), and the predicted take-rates (recorded as probabilities) averaged over all customers in all participant nodes in the random sample (line C). Since, in the scoring procedure, each customer in the sample is assigned the probability of participation connected with the terminal node s/he occupies, averaging over all of those probabilities yields a self-weighted mean of the node probabilities. This average is affected both by the distribution of probabilities over the nodes, but also by the distribution of customers over nodes. For example, if customers are concentrated in nodes with a probability of 0.30, this will yield a higher overall predicted take-rate than if customers were concentrated in nodes with probabilities of 0.03.

So, for PG&E, the average, or overall predicted take-rate for the random sample, is 13 percent, lower than for the original sample (18 percent-line A)[20]. However, when we calculate the predicted take-rate for only those nodes that are categorized as participant nodes, the figure is 30 percent (line E). The difference between 30 percent and 18 percent is the potential improvement in take-rates based on these CART models on a random sample, using a strategy of mailing only to customers in participant nodes. The predicted improvement for SCE is only 16 percent versus 12 percent, and for SDG&E, 23 percent versus 15 percent.

Further, it should be noted that these potential improvements are based on the assumption that all participant nodes would be the basis for choosing direct mailers. If the utility were to be more selective in the nodes it used, take-rates could be improved further.

It will be quite clear to the reader at this point that original samples that we received for the initial revealed preference modeling was very far from representative of the utility territories' residential customer base. This is very well illustrated by the difference in how many customers occupied certain nodes in the trees. The issues with the mailed sample were well known, as were the characteristics of the residents of the territories. The possibility of weighting the sample was discussed extensively at an early stage. However, we believe that whatever gain there might have been in weighting the samples has been realized in the application of the tree results to the random samples. We now know the proportions of customers who would fall into each node because of this application. The development of splitting rules would not have been affected by weighting except to the degree that any small change to a sample can change results slightly.

In actually using these trees or others like them to create mailing lists, with anticipated participant group sizes, it will be important for program planners to look carefully at the trees and the contents of the nodes. A terminal node may be very rich in participants, but very sparse in the number of customers occupying the node. In other words, a very large percentage of a small number of customers could be involved. So, to the extent that program planners must meet targets of absolute numbers of participants, this will have to be folded into calculations.

4.4 Assessment of Customer Interest in Other Audit Configurations

In the case of the stated preference study, the best method for determining market segments, or classes, was the LCDC approach. The best use of CART for this part of the study is to use it to determine the customer characteristics associated with each class. In

[20] This seems unexpected since we have already shown that there is a higher concentration of customers in participant nodes in the random sample than there were in the original sample, thus the predicted participation rate should have gone up. However, the increased richness of the participant nodes has clearly been offset by a corresponding (and more) concentration of customers in the lowest probability non-participant nodes. Recall that there is a variety of probabilities associated with various non-participant nodes, and it is obvious that the lower probability nodes (in terms of participants) have received a disproportionate share of customers among non-participant nodes.

sections 4.4.1 and 4.4.2, we present the results of both analyses. Readers who are not interested in the following technical discussion of these results can skip to Section 4.4.1.2.

4.4.1 LCDC Results

4.4.1.1 Technical Discussion

Using the procedure described above, a model specification including all these variables was run over a range of classes, from 1 to 9. The Bayesian Interaction Criterion (BIC) statistic guided the selection of the optimum number of segments. The BIC balances the increase in number of parameters (N par) with the goodness of fit (reduction in log likelihood LL), in essence weighing fit against parsimony. For our model the BIC criterion reached a minimum at 8, so the 8-class model was selected for subsequent analysis. Key statistics for this model are presented in Table 4-17.

Table 4-17
Key Diagnostics for the LCDC Model

		LL	BIC(LL)	Npar	L^2	df	p-value	Class.Err.	$R^2(0)$	R^2
Model1	1-Class Choice	-15214.7	30609.6	27	20,177	766	1.2e-3674	0.000	0.017	0.013
Model2	2-Class Choice	-14341.9	29044.3	54	18,431	739	2.2e-3329	0.007	0.096	0.092
Model3	3-Class Choice	-13929.4	28399.5	81	17,606	712	2.9e-3176	0.010	0.124	0.121
Model4	4-Class Choice	-13608.4	27937.7	108	16,964	685	2.7e-3061	0.017	0.157	0.153
Model5	5-Class Choice	-13468.8	27838.8	135	16,685	658	5.7e-3022	0.070	0.170	0.166
Model6	6-Class Choice	-13316.9	27715.3	162	16,381	631	1.3e-2977	0.076	0.188	0.184
Model7	7-Class Choice	-13157.2	27576.0	189	16,061	604	4.9e-2930	0.075	0.201	0.197
Model8	**8-Class Choice**	**-13064.6**	**27571.2**	216	15,876	577	1.1e-2910	0.085	0.214	0.210
Model9	9-Class Choice	-12976.8	27575.7	243	15,701	550	2.0e-2893	0.087	0.225	0.221

Notice that the p-value for the model is close to zero, indicating the model passed an overall significance test, and the value for R^2 (not to be confused with the ordinary least squares R^2), is 0.2024, an acceptable value. Information on how this R^2 is calculated appears in a following section.

Figure 4-9 graphically illustrates the relationship between the BIC and the identification of the correct number of classes.

Figure 4-9
Relationship of BIC to Identification of Correct Number of Classes

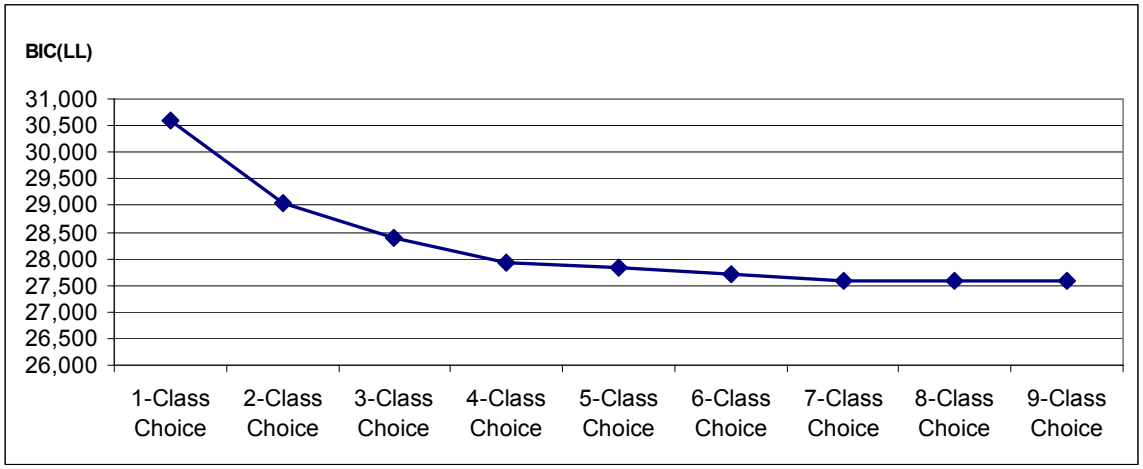


Table 4-18 displays parameter estimates along with significance tests for the 8-class model. The first column in this table lists the variables used in the model; lengthier descriptions for these variables appeared earlier in Table 3-17. The next eight columns contain parameter estimates for the conditional logit models associated with the eight classes (i.e., segments) in the model. It is important to note that the variables that indicate which utility serves these households turn out to be non-significant. As a result, there was no need to estimate separate utility models.

Table 4-18
Parameter Estimates and Significance Tests for Eight-Class Model

Model for Choices	Demanding, but willing to pay Class1	Meticulous Class2	Subsidy required Class3	Enthusiasts Class4	Not interested Class5	Personal attention Class6	Hard to convince Class7	Fast and thorough Class8	Overall
R^2	0.0299	0.0105	0.2677	0.3055	0.0155	0.1884	0.1594	0.2817	0.2098
$R^2(0)$	0.0541	0.0323	0.2939	0.332	0.8243	0.2281	0.1762	0.3274	0.2136

Attributes	Class1	Class2	Class3	Class4	Class5	Class6	Class7	Class8	Wald	p-value	Wald(=)	p-value	Mean	Std.Dev.
None														
	-2.1688	0.8578	0.2166	-5.2314	5.2649	1.7121	4.023	2.0136	174.5498	1.40E-33	167.8881	7.00E-33	0.0831	2.8129
DlvrMode														
Mail	-0.9952	0.0606	0.4793	1.0928	1.3759	-1.1961	-1.6212	-3.7612	155.432	6.60E-25	134.3277	9.50E-22	-0.3859	1.1931
Online	0.5712	0.0381	0.6811	-0.7387	-1.753	-0.5482	1.6949	1.8789					0.1813	0.901
In-home	0.424	-0.0987	-1.1604	-0.354	0.3771	1.7444	-0.0737	1.8823					0.2046	0.7693
TmRqAdt														
	-0.0188	0.0155	-0.0697	-0.063	-0.1806	0.0008	-0.0041	-0.0316	19.6275	0.012	15.7722	0.027	-0.0323	0.0538
PsTcSppr														
	0.568	0.4814	0.6253	-0.4974	3.0874	5.2521	-1.9696	1.803	99.7976	4.70E-18	86.2113	7.40E-16	0.9438	1.6517
UsgPrfl														
	0.5986	0.0374	1.4914	0.6731	4.7761	3.403	1.0574	3.9877	43.7384	6.40E-07	24.6828	0.00086	1.3486	1.5267
Lnk2Vndr														
No links	0.0162	-0.1412	0.1889	0.2083	-1.0239	-0.9704	-1.3167	-1.0741	149.0711	1.20E-23	136.2652	3.90E-22	-0.2943	0.51
National outlets	-0.1255	0.0252	-0.0787	-0.1466	0.5823	-1.5181	0.557	0.3139					-0.0681	0.4876
Local vendors	0.1093	0.1161	-0.1102	-0.0617	0.4416	2.4885	0.7597	0.7602					0.3624	0.6673
LvlOfDtl														
No savings estimates	-0.1067	-0.0236	-0.6041	1.0022	0.014	-1.2239	0.3727	-0.9262	116.9751	2.10E-17	113.5923	1.10E-17	-0.1245	0.5355
Rough estimates of savings	0.2014	0.0617	0.0034	-0.7212	-0.701	2.2061	-1.1994	1.0092					0.0925	0.7867
Refined estimate of savings	-0.0947	-0.0381	0.6007	-0.281	0.687	-0.9822	0.8266	-0.083					0.0321	0.4554
Fee														
	0.0168	-0.0056	-0.1255	-0.0326	-0.0145	0.0417	0.0388	0.0516	99.5702	5.20E-18	90.428	1.00E-16	-0.0055	0.0466
Tim2Rslt														
1 week	-0.5033	-0.1151	0.1495	0.7981	-0.0767	-0.1225	-1.3228	-1.0092	320.0667	1.50E-53	238.8342	7.10E-39	-0.2344	0.4956
2 weeks	0.2436	0.2691	-0.4847	0.875	0.5694	-0.4788	0.1508	-0.2718					0.1721	0.3879
3 weeks	0.2437	-0.1092	0.0166	-1.07	-0.7574	-0.1168	0.0308	-0.087					-0.1304	0.399
4 weeks	0.016	-0.0447	0.3186	-0.6031	0.2646	0.718	1.1411	1.368					0.1927	0.4803
NatLnSub														
	-1.0987	-0.4667	-0.2078	0.5987	0.1007	-4.2299	-1.1976	-16.258	80.007	4.90E-14	45.1329	1.30E-07	-1.6075	3.5508
OnLinSub														
	0.0917	-0.8822	-1.0077	-4.3891	2.0006	2.7449	1.2276	0.6888	87.204	1.70E-15	81.2071	7.80E-15	-0.1869	1.755
MailSub														
	2.0811	0.2036	-2.1903	-0.902	-0.0273	4.5146	3.1437	3.9582	99.7755	4.70E-18	58.8328	2.60E-10	1.0851	1.912
MailPsT														
	1.0063	0.3772	-0.7712	-0.1052	-0.4494	2.2655	0.6063	2.7264	72.0191	1.90E-12	30.7373	6.90E-05	0.6021	0.9165
UsgPrLcl														
	0.351	0.134	0.4558	0.8042	-0.4821	-0.9424	-3.9638	-5.7959	143.9513	3.60E-27	143.5134	9.30E-28	-0.4068	1.6531
LclVnSub														
	-0.8157	-0.4152	-0.4352	-3.4704	0.2126	-2.8978	-0.258	-1.0872	115.4341	2.90E-21	49.7852	1.60E-08	-0.9742	1.0621
PsTcSUsg														
	-0.4966	-0.2198	0.6454	1.1153	0.6965	-3.1138	0.1006	-2.9742	66.4636	2.50E-11	57.7995	4.10E-10	-0.3518	1.1671
MailUsg														
	-1.056	-1.056	-1.056	-1.056	-1.056	-1.056	-1.056	-1.056	101.1633	8.50E-24	0	.	-1.056	0
OnLinRgh														
	0.4971	0.9538	0.2286	2.6642	2.6732	0.9071	-8.1239	2.8412	140.5839	1.80E-26	75.3443	1.20E-13	0.5668	2.4348
PsTcSRgh														
	-0.2306	-0.5526	0.6453	2.6057	-1.7392	-6.2722	4.416	-1.3377	101.2339	2.40E-18	99.9547	1.10E-18	-0.3227	2.3089
PsTcSRef														
	-0.0551	-0.1677	0.978	3.9318	-1.2387	-1.8136	2.3249	-0.412	152.4985	5.90E-29	149.8838	4.30E-29	0.2816	1.4783
OnLinFee														
	-0.0288	-0.0084	0.0465	-0.0314	0.049	-0.0121	-0.0825	-0.1267	111.0163	2.40E-20	50.9207	9.50E-09	-0.0166	0.0419
InHomFee														
	-0.0275	0.0041	0.0447	0.0139	0.0332	-0.0379	-0.0243	-0.0668	78.4069	1.00E-13	54.0195	2.30E-09	-0.0059	0.0297

Model for Classes										
Intercept	Class1	Class2	Class3	Class4	Class5	Class6	Class7	Class8	Wald	p-value
	1.006	0.8288	-0.0094	-0.1345	-0.1766	-0.2874	-0.4932	-0.7337	196.333	6.90E-39

Turning to rows, the first few repeat names we assigned to each latent class. The next two rows are pseudo-R^2 measures: R^2 and $R^2(0)$. These terms measure reduction of error compared to baseline models, such that

$$R_k^2 = \frac{Error(baseline) - Error(\mathrm{mod}el)}{Error(baseline)} \qquad \text{Eq. 15}$$

where *k* indexes the two measures. The baseline for R^2 is an average-probability model; the baseline for $R^2(0)$ is a constants-only model.

The remaining rows display parameter values along with significance tests. Parameters for interaction variables in Table 4-18 begin after the variable "4 weeks." As a visual aid, the highest value for each parameter estimate in a row (i.e., across classes) is shaded orange; the lowest value, green; and any value above 0.1, yellow. Employing this scheme allows the distinctive nature of each segment to become apparent.

Columns 10 through 13 contain Wald significance tests. The first Wald / p-value combination tests whether the parameter set across classes equals 0, whereas the second pair (Wald(=) and p-value) tests the hypothesis that the true values of the parameters across all classes are equal. The low p values indicate that all coefficient estimates differ significantly from zero and from each other across classes. Columns 14 and 15 contain the means and standard deviations of coefficients in each row, weighted by class sizes as shown in Table 4-19.

Table 4-19
Weights by Class Size

Class #	Characterization	Percentage
Class 1	Demanding but willing to pay	28.6%
Class 2	Meticulous	23.9%
Class 3	Subsidy required	10.4%
Class 4	Enthusiasts	9.1%
Class 5	Not interested	8.8%
Class 6	Personal attention	7.8%
Class 7	Hard to convince	6.4%
Class 8	Fast and thorough	5.0%

Note that class proportions equal the marginal latent class probabilities for each segment.

Segment descriptions come from a careful analysis of the unique characteristics of each segment. A number of tables helped us develop these characterizations. We will discuss three of these: an importance table (Table 4-20), profile table (Table 4-21) and ProbMeans table (Table 4-22).

Importance as used here represents the maximum effect for attribute variables listed in Table 3-17 within each latent class. The following table displays a relative importance measure, defined as

$$releff_{xp} = \frac{\max eff_{xp}}{\sum_{p} \max eff_{xp}}, \qquad \text{Eq. 16}$$

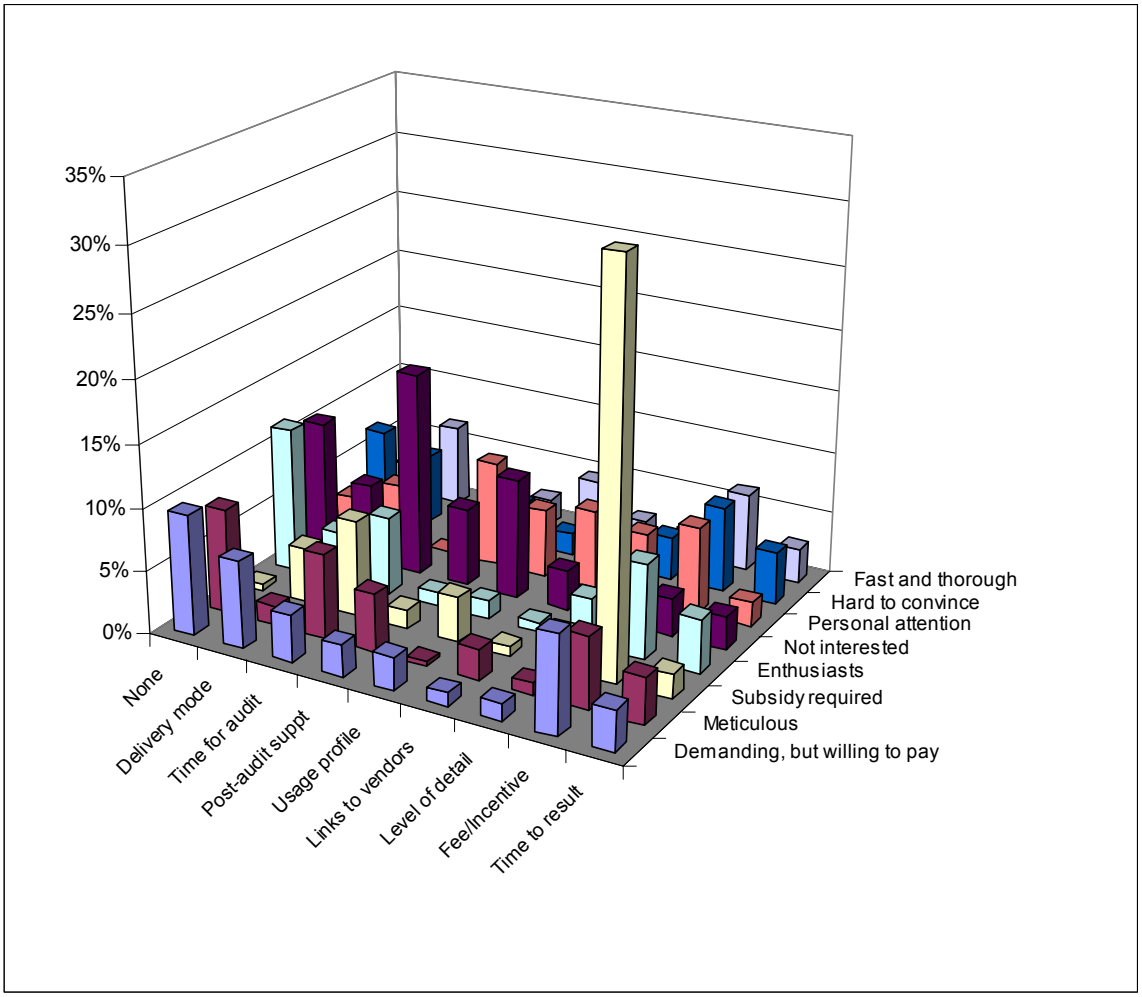
For each latent class x and attribute p. The *maximum effect* for attribute p is the difference in utility U_{max} - U_{min}, where U_{max} is the utility for the level that generates the maximum value for attribute A, and U_{min} is the utility for the level that generates the minimum value for attribute A. Table 4-20 presents the relative importance scores (values of $releff_{xp}$) for attribute by class. The higher the value the more important the attribute for a class.

Table 4-20
Relative Importance Table

Attribute Variable	Demanding, but Willing to Pay	Meticulous	Subsidy required	Enthusiasts	Not Interested	Personal Attention	Hard to Convince	Fast and Thorough
Participation	0.0964	0.0833	0.0053	0.1165	0.1063	0.0275	0.0684	0.0234
Delivery mode	0.0697	0.0155	0.0449	0.0408	0.0632	0.0471	0.0564	0.0657
Audit completion time	0.0377	0.0675	0.0765	0.0631	0.1641	0.0005	0.0031	0.0165
Post-audit tech suppt	0.0253	0.0467	0.0153	0.0111	0.0623	0.0842	0.0335	0.021
Usage profile	0.0266	0.0036	0.0364	0.015	0.0964	0.0546	0.018	0.0464
Links to vendors	0.0104	0.025	0.0073	0.0079	0.0324	0.0642	0.0353	0.0214
Level of detail	0.0137	0.0097	0.0294	0.0384	0.028	0.055	0.0344	0.0225
Fee/incentive	0.0784	0.0575	0.3214	0.0762	0.0307	0.0702	0.0693	0.0631
Time to results	0.0202	0.0146	0.0067	0.0511	0.0009	0.0119	0.0411	0.0371
NatLnSub	0.0557	0.0112	0	0.0125	0.009	0.0945	0.0186	0.2167
OnLinSub	0.0015	0.1019	0.022	0.0831	0.0472	0.0619	0.0166	0.008
MailSub	0.0981	0.025	0.0694	0.0722	0.0154	0.028	0.051	0.0334
MailPsT	0.0384	0.0337	0.0045	0.0122	0.0151	0.019	0.0039	0.0416
UsgPrLcl	0.0089	0.0216	0.0069	0.0197	0.0031	0.0056	0.0587	0.0733
LclVnSub	0.0276	0.0382	0.0119	0.1039	0.0062	0.0619	0.0074	0.0107
PsTcSUsg	0.0264	0.0236	0.0148	0.0092	0.0074	0.0116	0.0002	0.0348
MailUsg	0.057	0.1132	0.0335	0.0107	0.0006	0.013	0.0016	0.014
OnLinRgh	0.0281	0.0774	0.0058	0.0581	0.0475	0.0947	0.174	0.0343
PsTcSRgh	0.0068	0.0397	0.0181	0.0076	0.0497	0.1134	0.0321	0.0037
PsTcSRef	0.0052	0.0143	0.0214	0.0468	0.0472	0.0855	0.0159	0.0004
OnLinFee	0.1442	0.0463	0.1621	0.0085	0.0904	0.0295	0.1463	0.1562
InHomFee	0.1446	0.0751	0.1647	0.1065	0.0434	0.0118	0.0547	0.0525

A chart in Figure 4-10, based on this table, makes the relationships easier to see. For convenience, we show only principal variables, omitting the interaction terms.

Figure 4-10
Importance by Segment



A simple example illustrates how to read this chart. Notice that the most important attribute by far for Class 3, the "Subsidy Required" segment, is the Fee/incentive offered to respondents. This pronounced characteristic gave rise to this particular segment's name. In this respect, importance resembles its English connotation.

We now take up two related tables containing Profiles and ProbMeans, terms we will explain momentarily. The profile table (Table 4-21) on the next page displays a special kind of choice probability that varies only with respect to the attribute concerned. These values are calculated as follows. If a is a level of attribute p, where A_p is the total number of levels, and U is the utility associated with level a for latent class x, then the isolated choice probabilities for attribute p are

$$\hat{P}_p(a \mid x) = \frac{\exp(U_{a|xp})}{\sum_{a=1}^{A} \exp(U_{a|xp})} \qquad \text{Eq. 17}$$

For every attribute, taking "Time required to complete audit" on the next page as an example, the vertical probabilities associated with the levels *a* of attribute *p* within class *x* sum to 1. In those cases where the attribute takes on numeric values rather than discrete categories, the mean of the probabilities for that attribute is also displayed.

Color-coding helps interpret this table. For each *p* attribute within a class, the largest probability is colored orange; the smallest, green. Probabilities that lie within these extremes yet exceed 0.5 are colored yellow. We see, for example, that the conditional probability for a respondent in class 5 to participate in an audit is only 0.5 percent. That's why this was branded the "Not interested" segment. Conversely, respondents in class 4 ("Enthusiasts") demonstrated a participation probability of 99.5 percent.

The ProbMeans table (Table 4-22) that follows the Profile table resembles the latter table in interpretation, the only difference being that ProbMeans probabilities sum to 1 across classes rather than attributes. The calculation is

$$\hat{P}_p(x \mid a) = \frac{\hat{P}(x)\hat{P}_p(a \mid x)}{\sum_{x'=1}^{K} \hat{P}(x')\hat{P}_p(a \mid x')} \qquad \text{Eq. 18}$$

In this context, the value can be interpreted as the probability of being in class *x* given choice of attribute level *a* on attribute set *p*. The color-coding scheme is similar, with the largest probability in a row being colored orange, and the smallest, green. Other cells with probabilities greater than 0.5 are colored yellow.

Table 4-21
Profile Table

	Class1	Class2	Class3	Class4	Class5	Class6	Class7	Class8
Class Size	28.6%	23.9%	10.4%	9.1%	8.8%	7.8%	6.4%	5.0%
Attributes	**Demanding, but willing to pay**	**Meticulous**	**Subsidy required**	**Enthusiasts**	**Not interested**	**Personal attention**	**Hard to convince**	**Fast and thorough**
None								
Participation	0.897	0.298	0.446	0.995	0.005	0.153	0.018	0.118
No participation	0.103	0.702	0.554	0.005	0.995	0.847	0.982	0.882
Mean	0.103	0.702	0.554	0.005	0.995	0.847	0.982	0.882
DlvrMode - Delivery mode								
Mail	0.101	0.353	0.414	0.717	0.708	0.046	0.030	0.002
Online	0.483	0.345	0.506	0.115	0.031	0.088	0.829	0.498
In-home	0.417	0.301	0.080	0.169	0.261	0.867	0.141	0.500
TmRqAdt - Time required for you to complete audit								
15-20 min.	0.339	0.184	0.571	0.544	0.850	0.247	0.269	0.401
25 min.	0.147	0.103	0.167	0.168	0.099	0.124	0.130	0.158
30-40 min.	0.245	0.242	0.177	0.187	0.047	0.250	0.250	0.233
45 min.	0.101	0.141	0.041	0.048	0.003	0.126	0.120	0.084
50-60 min.	0.168	0.330	0.044	0.053	0.001	0.254	0.231	0.124
Mean	31.724	38.981	24.008	24.760	18.114	35.789	34.756	29.360
PsTcSppr - Post-audit technical support								
No	0.362	0.382	0.349	0.622	0.044	0.005	0.878	0.142
Yes	0.638	0.618	0.651	0.378	0.956	0.995	0.122	0.859
Mean	0.638	0.618	0.651	0.378	0.956	0.995	0.122	0.859
UsgPrfl - Usage profile								
No	0.355	0.491	0.184	0.338	0.008	0.032	0.258	0.018
Yes	0.645	0.509	0.816	0.662	0.992	0.968	0.742	0.982
Mean	0.645	0.509	0.816	0.662	0.992	0.968	0.742	0.982
Lnk2Vndr - Links to vendors								
No links	0.337	0.288	0.399	0.406	0.097	0.030	0.065	0.089
National outlets	0.293	0.340	0.305	0.285	0.483	0.017	0.421	0.356
Local vendors	0.370	0.372	0.296	0.310	0.420	0.953	0.515	0.556
LvlOfDtl - Level of detail in savings estimates								
No savings estimates	0.297	0.325	0.162	0.687	0.290	0.030	0.359	0.098
Rough estimates of savings	0.404	0.354	0.298	0.123	0.142	0.931	0.075	0.676
Refined estimate of savings	0.300	0.321	0.541	0.190	0.568	0.038	0.566	0.227
Fee - Audit fee or incentive								
We give you $30	0.065	0.211	0.848	0.427	0.283	0.009	0.011	0.003
We give you $15	0.083	0.194	0.129	0.262	0.228	0.016	0.020	0.007
Free, You pay $15	0.245	0.341	0.023	0.260	0.331	0.087	0.101	0.051
You pay $45	0.228	0.138	0.000	0.037	0.096	0.198	0.207	0.165
You pay $75	0.378	0.117	0.000	0.014	0.062	0.691	0.662	0.774
Mean	37.496	8.174	-27.334	-12.553	-0.756	61.039	59.235	65.727
Tim2Rslt - Time to results								
1 week	0.145	0.220	0.279	0.403	0.207	0.199	0.048	0.061
2 weeks	0.306	0.323	0.148	0.435	0.396	0.139	0.208	0.128
3 weeks	0.306	0.221	0.244	0.062	0.105	0.200	0.185	0.154
4 weeks	0.244	0.236	0.330	0.099	0.292	0.461	0.560	0.658

Table 4-22
ProbMeans Table

	Class1	Class2	Class3	Class4	Class5	Class6	Class7	Class8
Overall	28.6%	23.9%	10.4%	9.1%	8.8%	7.8%	6.4%	5.0%
Attributes	**Demanding, but willing to pay**	**Meticulous**	**Subsidy required**	**Enthusiasts**	**Not interested**	**Personal attention**	**Hard to convince**	**Fast and thorough**
None								
Participation	0.5296	0.1472	0.0954	0.1876	0.0009	0.0248	0.0023	0.0122
No participation	0.0568	0.3259	0.1112	0.0009	0.1689	0.1288	0.1216	0.0858
DlvrMode - Delivery mode								
Mail	0.0996	0.2924	0.148	0.2263	0.2144	0.0124	0.0066	0.0003
Online	0.3718	0.2229	0.1412	0.0283	0.0073	0.0185	0.1425	0.0674
In-home	0.3504	0.2122	0.0245	0.0453	0.0672	0.2	0.0265	0.0738
TmRqAdt - Time required for you to complete audit								
15 min.	0.233	0.098	0.159	0.132	0.244	0.044	0.040	0.050
20-25 min.	0.298	0.162	0.139	0.121	0.098	0.066	0.057	0.058
30 min.	0.328	0.229	0.105	0.096	0.030	0.084	0.070	0.058
40-45 min.	0.316	0.340	0.054	0.054	0.004	0.103	0.081	0.048
50 min.	0.292	0.405	0.034	0.035	0.001	0.110	0.084	0.040
60 min.	0.249	0.487	0.017	0.019	0.000	0.114	0.083	0.030
PsTcSppr - Post-audit technical support								
No	0.291	0.258	0.102	0.160	0.011	0.001	0.158	0.020
Yes	0.283	0.229	0.105	0.054	0.130	0.121	0.012	0.067
UsgPrfl - Usage profile								
No	0.350	0.406	0.066	0.107	0.003	0.009	0.057	0.003
Yes	0.260	0.172	0.119	0.085	0.122	0.107	0.067	0.069
Lnk2Vndr - Links to vendors								
No links	0.366	0.262	0.157	0.141	0.032	0.009	0.016	0.017
National outlets	0.269	0.262	0.102	0.084	0.136	0.004	0.086	0.057
Local vendors	0.248	0.209	0.072	0.066	0.086	0.175	0.077	0.065
LvlOfDtl - Level of detail in savings estimates								
No savings estimates	0.285	0.262	0.056	0.211	0.085	0.008	0.077	0.016
Rough estimates of savings	0.315	0.232	0.084	0.031	0.034	0.199	0.013	0.093
Refined estimate of savings	0.255	0.228	0.167	0.052	0.148	0.009	0.108	0.034
Fee - Audit fee or incentive								
We give you $30	0.083	0.227	0.395	0.176	0.112	0.003	0.003	0.001
We give you $15	0.183	0.355	0.103	0.184	0.153	0.010	0.010	0.003
Free	0.275	0.382	0.018	0.132	0.144	0.021	0.021	0.007
You pay $15	0.355	0.352	0.003	0.081	0.116	0.040	0.037	0.016
You pay $45	0.444	0.225	0.000	0.023	0.057	0.105	0.090	0.056
You pay $75	0.389	0.100	0.000	0.005	0.020	0.195	0.152	0.140
Tim2Rslt - Time to results								
1 week	0.207	0.264	0.145	0.185	0.091	0.078	0.015	0.015
2 weeks	0.307	0.271	0.054	0.140	0.122	0.038	0.047	0.023
3 weeks	0.405	0.246	0.117	0.026	0.043	0.073	0.055	0.036
4 weeks	0.232	0.188	0.114	0.030	0.085	0.121	0.119	0.110

We conclude with a brief discussion of our decision regarding the use of weights in the LCDC models to adjust for discrepancies presented in Section 3.4.4 with respect to some of the demographic variables. After careful consideration, we concluded that weighting during the estimation phase was not necessary since separate coefficients are calculated for each individual during the estimation process. These are the coefficients for the N classes times the probability that an individual is a member of each class. The process also determines which demographic characteristics are significant in determining class membership. Therefore, the proper place to weight the sample would be in the simulator, not during the estimation phase. However, our team agreed that because there were enough cases in each demographic group on which to base stable estimates of their effects, weighting observations in the LCDC models would not substantially changes the market shares. Note that even though there were enough cases within each demographic category to obtain stable parameter estimates, the demographic variables did not emerge as predictors in the models, making the weighting issue moot.

4.4.1.2 LCDC Summary

The result of the LCDC analysis has been the identification of eight distinct customer segments or classes who have been classified in terms of the particular audit configurations in which they appear to be interested. Up to this point we have hinted at why we characterized the segments the way we did. Now that we have presented, in the previous section (Technical Discussion), the tools used as references in this pursuit, we can present in Table 4-23 the label for each of the eight customer classes, the audit configurations in which they are interested upon which the label is based, and the analysis table sources, discussed in the previous section, upon which they were based.

Table 4-23
Segment Names, Description, and Source

Class	Label	Description	Source
1	Demanding but willing to pay	High probability of participation (89.7%), wants fast, non-mail audits, post-audit technical support, usage profiles and at least some estimate of savings. Willing to pay up to $75 for the audit and wait up to 2-3 weeks for results.	Profile table
2	Meticulous	Reluctant to participate (30%). Prefers the audit to be free and delivered by mail. Willing to devote up to 60 minutes to completing it, but wants results within one or two weeks. Wants post-audit technical support and a refined estimate of savings.	Profile table
3	Subsidy required	Fee/incentive has overwhelming influence (10.4%. Demands $30 subsidy.	Importance table
4	Enthusiasts	Very high probability of participation (98.5%). Respond well to subsidies. Prefer mail. Willing to spend up to 60 minutes to complete audit, but shun extras such as savings estimates, post-audit tech support or usage profiles.	Profile table

Class	Label	Description	Source
5	Not interested	Extremely low probability of participation (0.5%)	Profile table
6	Personal attention	Low probability of participation (15.3%) but high willingness to pay. Prefer subsidized in-home audits (86.7%) with refined estimates of savings and post-audit technical support (99.5%).	Profile table
7	Hard to convince	Very low probability of participation (2.8%)	Profile table
8	Fast and thorough	High willingness to pay (mean fee of $66), but wants a rapid, thorough audit including usage profiles, post-audit technical support and some estimates of savings. Willing to wait for results.	Profile table

These differing segment profiles can be quite useful to management. As an example, from a program manager's point of view classes 5 and 7 can be ignored, as their probabilities of participation are very small and difficult to influence through tested program mechanisms. At the other extreme, enthusiasts are virtually sold on participating, but their zeal can be dampened by the wrong choice of program attributes. Other groups can be swayed by particular combinations of program components.

4.4.2 LCDC-CART Analysis

The CART method was used in this segment of the study as well (see Section 3.3.1.2.1 for a description of the method). In the case of the stated preference study, the best method for determining market segments, or classes, was the LCDC approach. The best use of CART for this part of the study is to use it to determine the customer characteristics associated with each class. Specifically, it is used to see what clusters of customer characteristics distinguished each class from the others. While LCDC is capable of completing this kind of task, in this particular case, CART did a better job. This is largely because of the ability of CART to find a different place for a given variable in different parts of the tree, i.e., a variety of interactions can be better and more efficiently captured in CART than in other methods.

The data for this part of the study came primarily from the LCDC survey instrument, including reports from the respondent's most recent utility bill. This includes a set of attitudes and motivations about conservation. In addition, the respondents were asked for their zip codes, which allowed CEC Climate Zones to be appended to the analysis file. Not all respondents were able to provide information from their utility bills. Missing values for kWh were imputed using the SPSS Missing Values Analysis module, based on Expectation Maximization (EM)/regression methods.

One CART tree was grown for each of the eight LCDC classes, with a value of 1 assigned to the class under study, and a 0 otherwise. This procedure resulted in one to three terminal nodes that are rich in participants. The characteristics of these nodes help us to understand what customer characteristics are associated with each class. In this, as with the revealed preference analysis, choices must be made that balance the need to understand the customer segments with the need to predict. In this case, tree size does not

complicate the choice since the trees for the eight classes were small to moderate in size. The more important issue is the attitude questions, which help us understand the segments, but will not help in future customer targeting, since we will not have this information on non-surveyed customers. Because of these two somewhat conflicting goals, two sets of trees were grown; one that included attitude questions and one that didn't.

The CART software, internally, keeps growing a tree until groups can no longer be split in a way that results in improved accuracy. Typically, this last tree is not the one selected by CART for presentation because the tree would be too complex, and this complexity is not warranted by the increases in accuracy. The "decision" of which version of the tree to present is based on an algorithm that considers complexity and misclassification rate. However, the user can make a different choice if increased accuracy is desired or, alternatively, increased simplicity. When the "default" trees (those chosen by algorithm) that did not include attitudes were generated, in some cases there were a lot of terminal nodes (high complexity), and in others there was low complexity, and fairly low accuracy. In these cases, alternative (custom) trees are reported to give the reader an idea of what it is possible to achieve for each class. If IOUs wish to apply CART findings to target these segments, they can use this presentation to select the level of accuracy versus complexity they desire.

Table 4-24 provides an overview of the results of the CART analysis toward an understanding of the eight LCDC classes. The first two data columns show the level of predictive accuracy of the trees that include attitudes. The percentage of correct classifications achieved by the trees over the eight classes ranges from a low of 61 percent to a high of 98 percent for the prediction of class membership, and from 48 percent to 76 percent for non-membership. The range of tree sizes starts with a low of 3 terminal nodes to a high of 27. Overall, prediction of class membership is very good, even better than the revealed preference predictions.

Table 4-24
Prediction Success for CART Trees: Including Attitudes, Not Including Attitudes, Default & Modified

	Trees Including Attitude Questions		Default Trees--No Attitude Questions		Modified Trees--No Attitude Questions	
Class Description	% Correctly Predicted (Member/Not)	# of Terminal Nodes	% Correctly Predicted (Member/Not)	# of Terminal Nodes	% Correctly Predicted (Member/Not)	# of Terminal Nodes
1 Demanding but Willing to Pay	76/48	7	83/79	61	73/80 60/77	43 21
2 Meticulous	61/60	3	63/62	7		
3 Subsidy Required	93/80	27	68/57	5	66/76	11
4 Enthusiasts	92/74	21	100/87	51	93/72	28
5 Not Interested	59/76	4	99/75	30	99/67	24
6 Personal Attention	76/70	10	56/76	6	88/77	19
7 Hard to Convince	75/74	9	100/91	42	94/81	26
8 Fast & Thorough	98/59	8	85/57	3	100/71	15

The second set of columns reflects removing the attitude questions, and the resulting trees are labeled default trees. In some cases, removing attitudes increases accuracy, and in some cases it results in lower accuracy of classification, and this is accomplished with somewhat larger trees. The number of terminal nodes for this set of trees ranges from 3 to 61.

To experiment with pursuing more accuracy and/or more parsimony, different versions of the tree were explored. For most classes, one modified tree summary is presented. Two presented interesting choices for Class 1 and are shown in the appropriate row. In one case, prediction accuracy is reduced by 12 percent and the complexity of the tree by 18 percent. In the other case, accuracy is reduced by 23 percent, but the complexity is reduced by 64 percent. For Class 2, no satisfactory improvements were made, and the relevant cells of Table 4-24 are blank. These alternatives are interesting to consider, but the focus of the rest of this report will be on the trees that include attitude questions where they appear (attitudes do not always enter as splitters) because they can offer more insight into these market segments.

Now, we turn to a more detailed description of the characteristics of each class.

4.4.2.1 Class 1: Demanding but Willing to Pay

Sections 4.4.2.1 through 4.4.2.9 all refer to Table 4-25, which summarizes both the audit configurations/classes generated by the LCDC analysis and the customer characteristics associated with each class, which were produced by CART. This analysis helps us see what important customer characteristics lead to the classes choosing different configurations/attributes and, therefore, how these segments might be targeted or ignored.

The group of customers in Class 1, Demanding but Willing to Pay, comprising 29 percent of the sample, want fast, non-mail audits, post-audit technical support, savings estimates, and are willing to pay up to $75 for this service. They are also willing to wait for the results. Three terminal nodes are rich with these customers. All three groups or clusters tend to live in temperate climes (with some exceptions!). The characteristics that separate the clusters are home size and type as well as amount of gas used (different variables separate different subgroups). Customers in the first cluster generally live in a temperate climate and have larger homes, greater than 2,725 square feet. Customers in the second cluster come from the same climate zones but have smaller homes, yet higher gas usage (more than 295 therms). Customers in the third cluster live in the same climate zones, live in smaller attached or mobile homes, or apartments in large buildings (more than 5 units), and have lower gas usage than second cluster customers.

Table 4-25: Customer Characteristics of 8 Classes

Class Information	Audit Package Chosen	Terminal Node Summary Customer Description	Detailed Customer Characteristics
1 Demanding, but willing to pay 76/48* 29% of sample 10% of class chose none	1. Online delivery (In-home OK) 2. Time required to complete audit: 15-20 minutes (not 45 min) 3. Post-audit technical support 4. Usage profile 5. Links to local vendors 6. Provide rough estimates of savings 7. Charge customer $75 8. 2-3 weeks for audit results	6 Temperate Climate Zones (CZ) Larger homes	1. CZ: Arcata, Oakland, Sunnyvale, Santa Maria, L.A., S.D., El Toro, Pasadena, Sacramento, El Centro, Mt Shasta 2. KWh < 24k 3. SQFT > 2725
		4 Temperate CZ Smaller homes More gas used	1. CZ: Arcata, Oakland, Sunnyvale, Santa Maria, L.A., S.D., El Toro, Pasadena, Sacramento, El Centro, Mt Shasta 2. KWh < 24k 3. SQFT <= 2725 4. No rebate for Wx 5. Therms > 295
		3 Temperate CZ Smaller attached or mobile homes Less gas used	1. CZ: Arcata, Oakland, Sunnyvale, Santa Maria, L.A., S.D., El Toro, Pasadena, Sacramento, El Centro, Mt Shasta 2. KWh < 24k 3. SQFT <= 2725 4. No rebate for Wx 5. Therms < 295 6. Hometype: SF Att, Lg Apts, Mobile Homes

Table 4-25 (Continued)

Class Information	Audit Package Chosen	Terminal Node Summary Customer Description	Detailed Customer Characteristics
2 Meticulous 61/60* 24% of Sample 70% of class chose none	1. Mail delivery (Online OK) 2. 50-60 min to complete audit 3. Post-audit technical support 4. Usage profile 5. Links to local vendors 6. Provide rough estimates of savings 7. Free or charge customer $15 8. 2 weeks for audit results	3 High electric use	1. KWh > 8000
		2 Low-Mod electric use Familiar with energy efficiency programs	1. KWh < 8000 2. Recalled 6+ EE Progs
3 Subsidy Required 93/80* 10% of Sample 55% of class chose none	1. Online delivery (Mail OK) 2. Time required to complete audit: 15-20 minutes (not 45 min) 3. Post-audit technical support 4. Usage profile 5. No links to vendors 6. Provide refined estimates of savings 7. Most important, pay the customer $30 8. 4 weeks for audit results	24 Small, old homes Electric heat Temperate CZ	1. # Rooms >2 2. Elec Heating 3. Built before 1987 4. # People <6 5. SQFT >675 6. CZ: Santa Rosa, Oakland, Sunnyvale, S.D., El Toro, Pasadena, Riverside, Red Bluff
		11 Moderate size SF home No electric heat Low-mod gas use Temperate CZ Unaware of energy efficiency programs	1. # Rooms > 2 2. No Elec Heating 3. Therms < 245 4. CZ: Oakland, SD, El Toro, Pasadena, Riverside, Red Bluff, Sacto, El Centro 5. SQFT <= 2370 6. SF-Att, SF-Det, Oth 7. Don't recall turn-in prog 8. Born after 1946 9. Therms > 66

Table 4-25 (Continued)

Class Information	Audit Package Chosen	Terminal Node Summary Customer Description	Detailed Customer Characteristics
4 Enthusiasts 92/74* 9% of Sample 1.5% chose none	1. Mail delivery (definitely) 2. Time required to complete audit: up to 60 min 3. No post-audit technical support 4. No usage profile 5. No links to vendors 6. No estimates of savings 7. 2 weeks for audit results	14 Extreme CA Boomers Above average kWh Past energy efficiency program participation	1. CZ: Oakland, El Toro, Riverside, Sacto, Fresno, China Lk 2. KWh < 16687 3. Partic in < 3 EE progs 4. < 6 people in HH 5. Built after 1949 6. Elec: PG&E, SCE 7. Therms < 698 8. Born after 1949
5 Not Interested 59/76* 9% of Sample 99.5% chose none	1. Mail delivery 2. Time required to complete audit: 15-20 minutes (not 50-60 min) 3. Post-audit technical support 4. Usage profile 5. Links to national vendor outlets 6. Provide refined estimates of savings 7. Free to customer 8. 2 weeks for audit results	4 Mixed CZ	1. CZ: LA, Red Bluff, Sacto, Fresno
		1 Mixed CZ Non-conservation oriented	1. CZ: everything but above 2. Does not believe conservation helped in 2001 crisis 3. Does not believe scarce energy will be a problem
6 Personal Attention 76/70* 8% of Sample 85% chose none	1. In-home delivery (not mail) 2. Time required to complete audit: 50-60 minutes 3. Post-audit technical support 4. Usage profile 5. Links to local vendors 6. Provide rough estimates of savings 7. Willing to pay $75 8. 4 weeks for audit results	7 Temperate CZ Mod-High gas use Non-professional	1. CZ: Sunnyvale, LA, El Toro, Pasadena, Riverside 2. Therms < 1646 3. Homeage > 7.5 yrs 4. Gas heat 5. Educ < Prof Deg 6. Built after 1966

Table 4-25 (Continued)

Class Information	Audit Package Chosen	Terminal Node Summary Customer Description	Detailed Customer Characteristics
7 Hard to Convince 75/74* 6.4% of Sample 98% chose none	1. Online delivery 1. Time required to complete audit: 15-20 minutes (not 45 min) 2. No post-audit technical support 3. Usage profile 4. Links to local vendors 5. Provide refined estimates of savings 6. Willing to pay $75 7. 4 weeks for audit results	8 Young Small home	1. SF-Det, Apt-sm, Apt-lg 2. SQFT < 2600 3. 2+ 18-20 yr olds in HH
		3 Not 18-21 age Small home Philosophical conservationist	1. SF-Det, Apt-sm, Apt-lg 2. SQFT < 2600 3. Less than 2 18-20 in HH 4. Believe cons is worth it 5. Cons not econ necessity
		5 Not 18-21 age Small home Warm CZ Conservation minded Mod income	1. SF-Det, Apt-sm, Apt-lg 2. SQFT < 2600 3. Less than 2 18-20 in HH 4. Believe cons is worth it 5. Cons econ necessity 6. CZ: Oakland, LA, Pasadena, Riverside 7. Inc: $50-74k
8 Fast & Thorough 98/59* 5% of Sample 88% chose none	1. In-home (No mail) 2. Time required to complete audit: 15-20 minutes (not 45 min) 3. Post-audit technical support 4. Usage profile 5. Provide rough estimates of savings 6. Willing to pay $75 7. 4 weeks for audit results	6 Low income boomer Cold climate	1. CZ: Arcata, S Rosa, Oakland, Sunnyvale, SD, Riverside, Red Bluff, Sacto, Mt Shasta 2. Born after 1947 3. Built after 1951 4. Do not recall CARE rate 5. Inc: $25-34k

*% correctly classified as members/% correctly classified as not members

4.4.2.2 Class 2: Meticulous

This group of customers is somewhat reluctant to participate in any program configuration (only 30 percent made any choice; i.e., of the nine program configurations offered in the store, the respondent preferred none of them), and 24 percent of the surveyed sample falls into this group. For the most part, they prefer the free, mail audit, and are willing to put up to 60 minutes into completing the audit. They also want post-audit technical support and a rough estimate of savings. The CART tree predicted 61 percent of these members correctly (in only 3 terminal nodes), and these members fall mainly into two terminal nodes, representing two subgroups of these class members. The first is defined simply by high electricity use, over 8,000 kWh. Those who use less than this are characterized by familiarity with over six energy-efficiency programs, suggesting that they are already oriented to energy-efficiency. So, the group that demands this program configuration either is high in electricity consumption or is energy conscious.

4.4.2.3 Class 3: Require Subsidy

The members of this class of customers have only a moderate probability of participating at all (about 55 percent chose "none of the above"), but if they make a choice it involves getting a subsidy of $30. This is by far the most important consideration to this group, which contains 10 percent of the sample. The CART tree correctly classified 93 percent of these class members (in 27 terminal nodes), and two nodes were particularly rich in them. The first cluster is defined by old, small homes, electric heat, and most customers live in temperate climates. The other group also lives in moderate climates. In addition, these customers occupy moderate-sized homes, don't have electric heat, are low to moderate low gas users, and are probably not conservation minded.

4.4.2.4 Class 4: Enthusiasts

Only 9 percent of the surveyed customers are enthusiasts, so called because only 1 percent failed to choose an audit package. These customers are not demanding, as they require very few services. They prefer mail as a mode of delivery, but don't look for savings estimates, usage profiles, or technical support. The characteristics of these members are well associated with membership: CART accurately classified 92 percent of this group, and placed the bulk of them in one node. Enthusiasts are baby boomers, tend to live in extreme climates, can use a broad range of electricity quantities, up to almost 17,000 kWh, and have participated in past utility programs. The latter characteristic may account for their enthusiasm.

4.4.2.5 Class 5: Not Interested

About 9 percent of the sample falls into this cluster, 99.5 percent of whom chose no audit package as the preferred one. This group is not very well predicted, with CART accurately classifying only 59 percent of them. It is easier to predict who will *not* be in this group (76 percent of non-members were correctly classified). Most of the customers who were not interested fell into two groups: Those who live disproportionately in extreme climates, versus those who live in a wide range of climates but are very much not conservation minded. They do not believe that conservation helped in the 2001 energy crisis, and they do not believe that scarce energy will be a problem in the future.

4.4.2.6 Class 6: Personal Attention

This group involves 8 percent of the sample, and has a low probability of choosing in any program configuration. Those who may be willing to participate prefer subsidized, in-home audits with refined estimates and post-audit technical support. They also want links to local vendors, but are willing to pay up to $75 for this package. Seventy-six percent of the members of this group were correctly classified by the CART tree, using 10 terminal nodes. Most of this group was contained, however, in just one node. This node is defined by temperate weather, use of gas heat, moderate gas usage, and a non-professional level of education (i.e., a college degree or less).

4.4.2.7 Class 7: Hard to Convince

This is a low-probability group; only 1.8% of them chose a preferred audit package. This select group (6.4 percent of the sample) requires online delivery, a usage profile, refined estimates of savings, and links to local vendors. However, they are also willing to pay for the package. Three terminal nodes carried the bulk of these members. The first group is comprised of young people in small homes. The second is older, also in small homes, but defined by being philosophical conservationists, i.e., they believe conservation is worth it but that it is not an economic necessity for them. The third group is similar to the second, but lives in warm climates and has moderate income.

4.4.2.8 Class 8: Fast and Thorough

This group, encompassing only 5 percent of the sample, has quite a low probability of participating (about 12 percent). They don't want to put much time into the audit, they are willing to pay, but they want a lot in return: Post-audit technical support, a usage profile, and rough estimates of savings. They are willing to wait for the results, though. CART correctly predicted 98 percent of these members, who tend to be low-income baby boomers in cold climates.

4.4.2.9 Summary

In summary, there are several segments in this sample that have members willing to pay for a lot of services. These demanding segments or classes constitute a relatively small proportion of customers, but taken together could be worth pursuing. They are relatively easily identified by their characteristics. Those who are easy to convince are identifiable by their past program participation and tend to be fairly undemanding.

Six of the eight segments wish to have links to vendors who offer energy-efficient equipment. Five of the six segments requested local vendors, the other wanting national vendors. Members of 7 of the 8 segments expressed a desired to have some estimates of savings they would achieve by making recommended changes. None of the current HEES programs offer links to vendors, and some versions don't offer savings, or if they do, the information is not consistent across all measures.

Finally, the reader is reminded that the trees just described include customer attitudes toward energy efficiency, because those attitudes sometimes help us understand who is in the segments. However, additional trees were produced that do not depend on those questions

that cannot be asked of all customers. For 5 of the 8 classes, the additional trees equaled or surpassed the original trees in predictive accuracy in their default form. Of the remaining trees, modifications can be made that increase their accuracy to surpass the attitude-inclusive trees, as seen earlier in Table 4-24.

4.4.3 Application

Managers would do well to concentrate on segments for which program characteristics hold more sway. Because of the complexity of the model and the large number of interaction effects, it is hard to gauge the precise impact of changing program attributes without a simulator, which we provide as a deliverable with this study to the four utilities (*HEES Stated preference simulator 2006-01-13.xls (*see Appendix K for images of the three main worksheets). The Excel-based stated preference simulator (SPS) allows managers to assess the impacts of program structure on any combination of segments. Appendix K contains illustrations of the three key pages of the simulator. The first worksheet, the Main page, allows managers to simulate up to nine different types of audits, with each version of the audit varying one or more of the nine audit attributes. The second worksheet takes the estimate of market potential for a given alternative in the Main worksheet and plots an innovation diffusion curve using the Bass model (Mahajan and Peterson, 1985; Rogers, 1995). By choosing appropriate values of the two key parameters, α and β, you can simulate the adoption path of each alternative over time. α is the *coefficient of external influence* and represents the effect of external factors such as media communications on the rate of adoption. β is the *coefficient of internal influence and* represents the effect of prior adoption and word-of-mouth on new adoption. The third worksheet calculates and plots price versus potential revenue and price versus potential market share. Complete instructions for using the SPS are contained in a flash video, *Tutorial.exe*, that was also provided as a deliverable to the four utilities. The tutorial has three main sections: 1) Overview, 2) Generating Price Curves, and 3) Focusing on Segments. Only the first is of interest to users. Note that at the bottom of screen in the Overview section is a set of controls similar to those found on a standard VCR. The "buttons" are from left to right: Rewind, Back, Play, Pause, Forward, and Exit. These can be used to manage the learning experience.

5 Recommendations

5.1 Improvement of Target Marketing

There are a variety of experiments that could be carried out to determine whether modifications in the marketing and/or design of the audits (mail-in, online, and in-home) would result in greater participation and adoption rates. In this section, we present examples of the types of experiments that could be conducted to determine whether marketing to customers with the higher CART-estimated probabilities of participating results in higher take-rates as compared to the traditional approach. While various experiments are being conducted, the direct mailings can be carried out as they have been in the past several years.

One experiment could involve selecting a random sample of 3,000 residential customers and attaching Block Group-level demographic variables to each customer[21]. Then, using the revealed preference CART model (after translating into SAS code), identify those households that have higher CART-predicted probabilities of participating. These customers would be sent the direct mailer. One could compare the rates of participation between this group and the larger group who were targeted using the traditional approach.

5.2 Exploration of Customer Interest in Other Audit Configurations

To verify customer interest in various audit configurations, we recommend conducting a number of experiments. Using the HEES stated preference simulator, program managers can identify those configurations that they are interested in testing and the characteristics of those customers most interested in participating. One could test the new configuration(s) on a relatively small sample of very likely interested customers and compare their take rates and adoption rates among those who chose to participate with customers who received the standard mail-in audit.

For example, the availability of post-audit technical support is one of the most popular audit features identified in the stated preference analysis. For a random sample of 3,000 of those who receive the traditional mailer in a given mailing effort, one could insert an offer to provide technical support, via a toll-free hot line, after the audit results are delivered to the customer[22]. Several expert residential auditors (for example, such as those who conduct the local in-home/phone audits for SCE) could be assigned to staff the hot line for a period of two months after the results are mailed to participants. The adoption rates of those who received the traditional mailer can be compared to those who received the offer of technical support. Subsequently, an analysis could be conducted to determine whether any benefits, in the form of higher adoption rates (and thus higher savings), exceed the additional costs of providing technical support.

[21] The actual sample sizes will be based on the desired level of statistical power, the alpha, and an effect size of practical importance.

[22] The actual sample sizes will be based on the desired level of statistical power, the alpha, and an effect size of practical importance.

5.3 Improvement of Hard-To-Reach Targeting

One of the recommendations made as part of the evaluation of the PY 2002 HEES (Ridge et al., 2004) was that targeting of HTR population should be done using BlockGroup-level U.S. Census data rather than zip code-level[23]. A zip code is comprised of smaller units called tracts, block groups and blocks, zip codes manifest greater variation with respect to demographic characteristics than do the smaller units such as tracts and blocks. Thus, targeting by zip code areas is less refined than targeting by tracts, block groups, or blocks[24].

The 2004 report illustrated this problem by examining the demographic characteristics of participants who live in zip codes that have been defined as HTR (Moderate Income, Latino/Hispanic, or Renter). For example, a person who lives in a zip code defined as Moderate Income and whose self-reported income based on the telephone survey falls in the moderate income category is declared as being accurately targeted. On the other hand, a person who lives in a zip code defined as Moderate Income and whose self-reported income does not fall in the moderate income category is declared as being inaccurately targeted. Table 5-1 presents these results.

Table 5-1
Error Rates Using Zip Code Targeting

HTR Group	Error Rates
Moderate Income	69.5%
Latino/Hispanic	87.2%
Renter	86.4%

The report concluded that the error rates are all quite large and that targeting through the use of demographic data at the Census tract or block level would likely produce smaller errors, i.e., more efficient marketing.

We conclude by noting that even though there are no HTR goals for the four utilities in the 2006-08 period, we expect that utilities will continue to be concerned about targeting the HTR population. Thus, we recommend that targeting the HTR population should be done based on demographic data at the Census block level. The first step would be to define each Census block in a given utility's service territory as HTR or not using the utility HTR definitions. Next, each utility residential customer should be assigned their Census block Identification Number (ID). Using this block ID, the HTR classification of that block can be attached to the customer along with other demographic data. Finally, those customers living in HTR Census blocks can be mailed the survey with the expectation that there will be a

[23] Zip codes were established by the United States Postal Service for efficient mail delivery. Because zip code boundaries follow the routes of mail carriers, they do not conform to boundaries of Governmental Units or to those of the Bureau's Statistical Units. In fact, zip code areas usually do not have clearly identifiable boundaries. They change periodically to meet postal requirements and they do not cover the total land area of the United States.

[24] Note that Block-level Census data are now available from the California Department of Finance and could be used for better targeting of HTR customers.

greater likelihood that any given household in an HTR block will actually be a member of the HTR population. Of course, SCE could reduce its error rate even more by using Acxiom data. If other utilities purchased Acxiom data, they could also experience similar reductions in the error rate.

Appendix A

Internet Questionnaire for 2003 HEES Program Evaluation

Introduction

Infosurv in cooperation with Ridge & Associates is conducting a study sponsored by four California utilities, Pacific Gas & Electric, Southern California Edison, Southern California Gas, and San Diego Gas & Electric. We are conducting this study of residential customers to determine your preferences for various kinds of residential energy audits. An energy audit involves collecting information about your home and the equipment in it that uses electricity and natural gas. This information is used to make recommendations about how to reduce your use of electricity and natural gas.

There are many features of an energy audit that can vary and we would like to know your preferences for different combinations of these features. If you have any questions about this survey, please feel free to call one of the following numbers, depending on which of these four utilities serves you.

Pacific Gas & Electric	(800) 933-9555
Southern California Edison	(800) 736-4777.
Southern California Gas	(858) 636-6838
San Diego Gas & Electric	(858) 636-6838

Any information that you provide will be kept strictly confidential. Any information that we collect will only be reported in aggregate.

Please take a few minutes to complete this survey that will provide valuable information to these four utilities to improve their energy efficiency services to people like yourself.

Some Preliminary Questions

FOR THE FOLLOWING QUESTION AND FOR ALL QUESTIONS IN THIS SURVEY, PLEASE REFER TO YOUR PRIMARY RESIDENCE.

Q1. Is your primary residence in California?

1 ____ Yes **[CONTINUE]**
2 ____ No **[TERMINATE]**
77____Other **[TERMINATE]**
88____I refuse to answer **[TERMINATE]**
99____I don't know **[TERMINATE]**

Q2. Do you live at your primary residence at least 9 months out of the year?

1 ____ Yes **[CONTINUE]**
2 ____ No **[TERMINATE]**
77____Other **[TERMINATE]**
88____I refuse to answer **[TERMINATE]**
99____I don't know **[TERMINATE]**

Q3. Which of the following provides your *electricity*?

1 __ Pacific Gas and Electric **[CONTINUE]**
2 __ Southern California Edison **[CONTINUE]**
3 __ San Diego Gas and Electric **[CONTINUE]**
4 __ Southern California Gas Company **[CONTINUE]**
5 __ Other **[TERMINATE]**
6___I don't know **[TERMINATE]**

Q4. How is your electric bill paid?

1 I pay my own electric bill **[CONTINUE]**
2 It is included in my mortgage or rental payment **[TERMINATE]**
77____Other **[TERMINATE]**
88____I refuse to answer **[TERMINATE]**
99____I don't know **[TERMINATE]**

Q5. Which of the following provides your *natural gas*?

1 __ Pacific Gas and Electric
2 __ Southern California Gas
3 __ San Diego Gas and Electric
4 __ Other
5 __ I don't get natural gas
6___I don't know

Q6. When did you move to your primary address?

Month [INSERT DROPDOWN LIST]

____ ____ ____ ____ Year88____I refuse to answer
99____I don't know

Q7. Do you own or rent your home?

1 ___ Own
2 ___ Rent
77____Other
88____I refuse to answer
99____I don't know

Choice Section

In this next section of this questionnaire, you will be going through 9 different stores. Each store will have different energy audit packages for you to choose among. Imagine that you are making choices for your own family based on your own energy needs. You can select "None of these" if you don't see anything that you would value.

Each store is a new set of choices. In each store, you can choose up to one audit package or decide that you do not want any of the choices. Don't be concerned about what you chose in past stores or may choose in future stores. You should make your comparisons only within a store; you will not be able to go back and forth comparing between stores. At the bottom of each set of audit packages, you will then be asked to select the least attractive option in the set.

Once you get the feel of it, you should take about one minute per store. When you finish the last store, the survey will take you to some additional questions.

NOTE THAT "BACK" BUTTON WAS DISABLED DURING THE DISCREET-CHOICE QUESTIONS.

NOTE THAT THERE ARE 9 STORES PER BLOCK AND 8 AUDIT CONFIGURATIONS PER STORE PLUS "NONE OF THESE." THERE ARE 6 BLOCKS. EACH RESPONDENT WAS PRESENTED WITH 9 STORES IN ONE RANDOMLY SELECTED BLOCK. HERE WE SHOW ONLY THE 9 STORES ASSOCIATED WITH THE FIRST BLOCK. BECAUSE THERE ARE 6 BLOCK, THERE WERE 54 STORES REPRESENTED IN THIS CHOICE EXPERIMENT. THE OTHER 45 STORES IN THE OTHER 5 BLOCKS CAN BE PROVIDED UPON REQUEST. ALSO NOTE THAT THE FOLLOWING DEFINITIONS WERE PLACED BELOW EACH OF EACH STORE STORE SO THAT THE RESPONDENT COULD HAVE THEM READILY AVAIALBLE.

Energy audit	An energy audit involves collecting information about your home and the equipment in it that uses electricity and natural gas. This information is used to make recommendations about how to reduce your use of electricity and natural gas. There are many features of an energy audit that can vary. Below we describe some of these features and ways they can vary.		
Mode of delivery	Ways to provide information about your home and \appliances and get recommendations for saving energy	Mail	You complete a paper questionnaire about your home, appliances, & energy habits & mail it to your utility. You receive recommendations in the mail.
		Online	You complete an online questionnaire about your home, appliances, & energy habits & receive recommendations online.
		In-home	An expert comes to your home to record information about your home, appliances and energy habits. You will receive a printed report with recommendations.
Time required for you to complete audit	The amount of your time it takes for you to provide, for the audit, the information about your home, appliances, and energy habits.	Range: 15 - 60 min.	
Additional services	Services in addition to the recommendations you always get with an energy audit	Post-audit technical support	You can call your utility and talk to someone to help you understand the audit recommendations and how to implement them.
		Usage profile	A graph that shows how much of your utility bill goes to each type appliance based on audit information. This requires that you provide your utility account number in the audit.
Links to vendors	The amount of help your utility will give yo in finding retailers and contractors to install equipment recommended by the audit.	No links	You must find your own retailers and contractors to install equipment recommended by the audit.
		National outlets	Audit recommendations would come with suggestions about which national retailers will usually carry the recommended equipment.
		Local vendors	Audit recommendations would come with specific information about reputable local retailers that carry the recommended equipment.
Level of detail & accuracy in savings estimates	There can be a lot of variation in the accuracy and detail in the estimated savings that the audit provides with each recommendation.	No savings estimates	You get general recommendations and tips for changing appliances and energy habits but without any estimates of kWh or therm savings.
		Rough estimates of savings	You get recommendations for changing appliances and energy habits along with estimated savings experienced by households similar to yours.
		Refined estimate of savings	You get much more accurate estimates of kWh or therm savings for each recommendation, based on your billed energy use. This requires that you provide your utility account number as part of the energy
Audit fee or incentive	You could be asked to pay for an energy audit, or the utility could offer you an incentive to complete one, or the audit could be offered free of cost or incentive.	We give you $30*	
		We give you $15*	
		Free	
		You pay $15	
		You pay $45	
		You pay $75	
Time between decision to participate & results	The time between when you request an energy audit and when you receive the recommendations.	Range: 1 week - 4 weeks	

Store 1 Block 1

Directions: Please choose the audit package that you most prefer from among the following options, or specify "None of these" if you don't want any of them.
Click the radio button corresponding to your choice.

	1	2	3	4	5	6	7	8	9
Mode of delivery:	In-home	Mail	Online	Online	Mail	Mail	In-home	In-home	None of these
Time required for you to complete audit:	60 min.	20 min.	25 min.	30 min.	15 min.	40 min.	25 min.	25 min.	
Additional services:	Post-audit technical support Usage profile		Post-audit technical support	Post-audit technical support		Usage profile	Post-audit technical support	Post-audit technical support	
Links to vendors:	No links	Local vendors	Local vendors	Local vendors	National outlets	National outlets	Local vendors	National outlets	
Level of detail and accuracy in savings estimates:	Refined estimate of savings	Refined estimate of savings	Rough estimates of savings	Rough estimates of savings	No savings estimates	Refined estimate of savings	Rough estimates of savings	Refined estimate of savings	
Audit fee or incentive:	You pay $75	You pay $75	You pay $45	You pay $75	You pay $45	We give you $30*	Free	Free	
Time between decision to participate and results:	2 weeks	2 weeks	4 weeks	1 week	3 weeks	2 weeks	4 weeks	2 weeks	

*in energy-saving equipment

1	2	3	4	5	6	7	8	9

Now from among these same 9 choices, select the one you LEAST prefer.
Click the radio button corresponding to your choice.

Store 2 Block 1

Directions: Please choose the audit package that you most prefer from among the following options, or specify "None of these" if you don't want any of them.
Click the radio button corresponding to your choice.

	1	2	3	4	5	6	7	8	9
Mode of delivery:	Online	Mail	In-home	In-home	In-home	Mail	In-home	Mail	None of these
Time required for you to complete audit:	50 min.	20 min.	45 min.	30 min.	30 min.	20 min.	15 min.	40 min.	
Additional services:	Post-audit technical support Usage profile		Usage profile	Post-audit technical support	Post-audit technical support			Usage profile	
Links to vendors:	Local vendors	No links	No links	National outlets	No links	Local vendors	Local vendors	Local vendors	
Level of detail and accuracy in savings estimates:	No savings estimates	Rough estimates of savings	Refined estimate of savings	Refined estimate of savings	Refined estimate of savings	No savings estimates	Refined estimate of savings	Refined estimate of savings	
Audit fee or incentive:	We give you $30*	You pay $75	You pay $75	You pay $15	You pay $75	You pay $15	We give you $30*	You pay $45	
Time between decision to participate and results:	1 week	2 weeks	3 weeks	3 weeks	3 weeks	2 weeks	1 week	4 weeks	

*in energy-saving equipment

1	2	3	4	5	6	7	8	9

Now from among these same 9 choices, select the one you LEAST prefer.
Click the radio button corresponding to your choice.

Store 3 Block 1

Directions: Please choose the audit package that you most prefer from among the following options, or specify "None of these" if you don't want any of them.
Click the radio button corresponding to your choice.

	1	2	3	4	5	6	7	8	9
Mode of delivery:	Online	Online	In-home	Online	Online	Online	In-home	Mail	None of these
Time required for you to complete audit:	50 min.	40 min.	30 min.	30 min.	30 min.	50 min.	60 min.	25 min.	
Additional services:	Post-audit technical support Usage profile	Usage profile	Post-audit technical support	Post-audit technical support	Post-audit technical support	Post-audit technical support Usage profile	Post-audit technical support Usage profile	Post-audit technical support	
Links to vendors:	National outlets	Local vendors	No links	Local vendors	No links	National outlets	Local vendors	National outlets	
Level of detail and accuracy in savings estimates:	No savings estimates	Refined estimate of savings	Rough estimates of savings	Rough estimates of savings	Refined estimate of savings	No savings estimates	Refined estimate of savings	No savings estimates	
Audit fee or incentive:	Free	Free	We give you $15*	You pay $75	We give you $15*	Free	We give you $15*	You pay $45	
Time between decision to participate and results:	1 week	4 weeks	3 weeks	3 weeks	1 week	3 weeks	4 weeks	4 weeks	

*in energy-saving equipment

1	2	3	4	5	6	7	8	9
○	○	○	○	○	○	○	○	○

Now from among these same 9 choices, select the one you LEAST prefer.
Click the radio button corresponding to your choice.

Store 4 Block 1

Directions: Please choose the audit package that you most prefer from among the following options, or specify "None of these" if you don't want any of them.
Click the radio button corresponding to your choice.

	1	2	3	4	5	6	7	8	9
Mode of delivery:	Online	Online	In-home	Mail	In-home	Mail	Mail	In-home	None of these
Time required for you to complete audit:	30 min.	20 min.	20 min.	60 min.	30 min.	15 min.	50 min.	50 min.	
Additional services:	Post-audit technical support			Post-audit technical support Usage profile	Post-audit technical support		Post-audit technical support Usage profile	Post-audit technical support Usage profile	
Links to vendors:	National outlets	National outlets	Local vendors	National outlets	National outlets	Local vendors	National outlets	No links	
Level of detail and accuracy in savings estimates:	Rough estimates of savings	Rough estimates of savings	No savings estimates	Rough estimates of savings	Rough estimates of savings	Refined estimate of savings	Refined estimate of savings	Refined estimate of savings	
Audit fee or incentive:	We give you $15*	We give you $15*	You pay $75	You pay $15	You pay $75	You pay $45	We give you $30*	You pay $45	
Time between decision to participate and results:	1 week	4 weeks	4 weeks	2 weeks	1 week	3 weeks	3 weeks	3 weeks	

*in energy-saving equipment

1	2	3	4	5	6	7	8	9
○	○	○	○	○	○	○	○	○

Now from among these same 9 choices, select the one you LEAST prefer.
Click the radio button corresponding to your choice.

Store 5 Block 1

Directions: Please choose the audit package that you most prefer from among the following options, or specify "None of these" if you don't want any of them.
Click the radio button corresponding to your choice.

	1	2	3	4	5	6	7	8	9
Mode of delivery:	Mail	In-home	Mail	Mail	Online	Mail	Online	Online	None of these
Time required for you to complete audit:	50 min.	40 min.	40 min.	60 min.	50 min.	40 min.	45 min.	30 min.	
Additional services:	Post-audit technical support			Post-audit technical support	Post-audit technical support			Post-audit technical support	
Links to vendors:	Usage profile National outlets	Usage profile Local vendors	Usage profile National outlets	Usage profile No links	Usage profile Local vendors	Usage profile No links	Usage profile National outlets	 National outlets	
Level of detail and accuracy in savings estimates:	Rough estimates of savings	Refined estimate of savings	Rough estimates of savings	Refined estimate of savings	No savings estimates	No savings estimates	Rough estimates of savings	Refined estimate of savings	
Audit fee or incentive:	Free	We give you $30*	Free	You pay $15	We give you $30*	We give you $30*	We give you $15*	You pay $75	
Time between decision to participate and results:	1 week	2 weeks	4 weeks	4 weeks	3 weeks	4 weeks	1 week	1 week	

*in energy-saving equipment

1 2 3 4 5 6 7 8 9

Now from among these same 9 choices, select the one you LEAST prefer.
Click the radio button corresponding to your choice.

Store 6 Block 1

Directions: Please choose the audit package that you most prefer from among the following options, or specify "None of these" if you don't want any of them.
Click the radio button corresponding to your choice.

	1	2	3	4	5	6	7	8	9
Mode of delivery:	Online	In-home	Mail	Mail	Online	In-home	Mail	Online	None of these
Time required for you to complete audit:	60 min.	40 min.	45 min.	45 min.	40 min.	30 min.	15 min.	15 min.	
Additional services:	Post-audit technical support					Post-audit technical support			
Links to vendors:	Usage profile No links	Usage profile No links	Usage profile National outlets	Usage profile National outlets	Usage profile No links	 Local vendors	 No links	 National outlets	
Level of detail and accuracy in savings estimates:	No savings estimates	Refined estimate of savings	No savings estimates	No savings estimates	No savings estimates	Refined estimate of savings	Rough estimates of savings	Refined estimate of savings	
Audit fee or incentive:	You pay $75	You pay $45	You pay $75	You pay $75	You pay $45	We give you $15*	You pay $45	You pay $45	
Time between decision to participate and results:	2 weeks	2 weeks	3 weeks	1 week	4 weeks	3 weeks	3 weeks	1 week	

*in energy-saving equipment

1 2 3 4 5 6 7 8 9

Now from among these same 9 choices, select the one you LEAST prefer.
Click the radio button corresponding to your choice.

Store 7 Block 1

Directions: Please choose the audit package that you most prefer from among the following options, or specify "None of these" if you don't want any of them.
Click the radio button corresponding to your choice.

	1	2	3	4	5	6	7	8	9
Mode of delivery:	Online	Online	Online	Mail	Online	Mail	In-home	In-home	None of these
Time required for you to complete audit:	45 min.	60 min.	15 min.	15 min.	25 min.	20 min.	25 min.	20 min.	
Additional services:		Post-audit technical support			Post-audit technical support		Post-audit technical support		
	Usage profile	Usage profile							
Links to vendors:	National outlets	National outlets	No links	Local vendors	National outlets	National outlets	National outlets	No links	
Level of detail and accuracy in savings estimates:	No savings estimates	Rough estimates of savings	Rough estimates of savings	Refined estimate of savings	Rough estimates of savings	Refined estimate of savings	No savings estimates	No savings estimates	
Audit fee or incentive:	You pay $15	We give you $15*	Free	You pay $45	We give you $30*	We give you $15*	We give you $30*	You pay $15	
Time between decision to participate and results:	1 week	4 weeks	1 week	1 week	4 weeks	2 weeks	4 weeks	2 weeks	

*in energy-saving equipment

1	2	3	4	5	6	7	8	9

Now from among these same 9 choices, select the one you LEAST prefer.
Click the radio button corresponding to your choice.

Store 8 Block 1

Directions: Please choose the audit package that you most prefer from among the following options, or specify "None of these" if you don't want any of them.
Click the radio button corresponding to your choice.

	1	2	3	4	5	6	7	8	9
Mode of delivery:	In-home	Online	In-home	Online	Online	Mail	In-home	In-home	None of these
Time required for you to complete audit:	25 min.	25 min.	25 min.	30 min.	25 min.	20 min.	20 min.	45 min.	
Additional services:	Post-audit technical support	Post-audit technical support	Post-audit technical support	Post-audit technical support	Post-audit technical support				
								Usage profile	
Links to vendors:	No links	No links	Local vendors	National outlets	National outlets	National outlets	National outlets	National outlets	
Level of detail and accuracy in savings estimates:	Refined estimate of savings	No savings estimates	Refined estimate of savings	No savings estimates	No savings estimates	Refined estimate of savings	Rough estimates of savings	Refined estimate of savings	
Audit fee or incentive:	You pay $45	You pay $45	We give you $30*	You pay $15	Free	We give you $15*	You pay $75	You pay $15	
Time between decision to participate and results:	4 weeks	2 weeks	2 weeks	3 weeks	4 weeks	4 weeks	4 weeks	3 weeks	

*in energy-saving equipment

1	2	3	4	5	6	7	8	9

Now from among these same 9 choices, select the one you LEAST prefer.
Click the radio button corresponding to your choice.

Store 9 Block 1

Directions: Please choose the audit package that you most prefer from among the following options, or specify "None of these" if you don't want any of them.
Click the radio button corresponding to your choice.

	○ 1	○ 2	○ 3	○ 4	○ 5	○ 6	○ 7	○ 8	○ 9
Mode of delivery:	Mail	Online	In-home	Mail	Online	In-home	Mail	In-home	None of these
Time required for you to complete audit:	15 min.	40 min.	60 min.	50 min.	60 min.	15 min.	40 min.	40 min.	
Additional services:			Post-audit technical support	Post-audit technical support	Post-audit technical support				
		Usage profile	Usage profile	Usage profile	Usage profile		Usage profile	Usage profile	
Links to vendors:	Local vendors	National outlets	Local vendors	No links	No links	Local vendors	National outlets	Local vendors	
Level of detail and accuracy in savings estimates:	Rough estimates of savings	Refined estimate of savings	Rough estimates of savings	No savings estimates	No savings estimates	No savings estimates	No savings estimates	No savings estimates	
Audit fee or incentive:	We give you $30*	You pay $45	You pay $15	We give you $30*	You pay $75	You pay $45	You pay $45	You pay $45	
Time between decision to participate and results:	1 week	2 weeks	4 weeks	1 week	4 weeks	1 week	2 weeks	4 weeks	

*in energy-saving equipment

1 ○	2 ○	3 ○	4 ○	5 ○	6 ○	7 ○	8 ○	9 ○

Now from among these same 9 choices, select the one you LEAST prefer.
Click the radio button corresponding to your choice.

Your Attitudes about Energy Use

Q8. People have different opinions about energy efficiency and the availability of natural resources such as energy. Using a 10-point scale, with a "1" meaning you "Strongly Disagree" and a "10" meaning you "Strongly Agree," please show how much you disagree or agree with each of the following statements

		Strongly Disagree									Strongly Agree	Don't Know
		1	2	3	4	5	6	7	8	9	10	99
a.	My life is too busy to worry about making energy related improvements to my home.											
b.	Scarce energy supplies will be a major problem in the future											
c.	Instead of building new power plants, customers should use less electricity											
d.	It is possible to save energy without sacrificing comfort by being energy efficient											
e.	It is worth it to me for my household to use less energy in order to help preserve the environment											
f.	Conservation efforts helped reduce the effects of the energy crisis during the summer of 2001											
g.	Conserving energy in my home is an economic necessity											
h.	There is little I can to do reduce the amount of electricity that I am now using											

Your Awareness of Energy Efficiency

Q9. Have you ever seen or heard of ENERGY STAR?

1 ____ Yes → ENABLE Q10
2 ____ No **[SKIP TO Q11]**
88____I refuse to answer **[SKIP TO Q11]**
99____I don't know **[SKIP TO Q11]**

Q10. What is your understanding of what ENERGY STAR means? It means the appliance….(check all that apply)

1____ saves energy/uses less energy
2____ is less harmful to the environment, causes less pollution
3____ costs less to operate, saves money on electric bill
4____ comes with a rebate offer
5____ meets a government standard for energy efficiency
6____ other (please specify): _____

88___ I refuse to answer

Q11. Over the years, the electric utilities and others, including the State of California, have offered a variety of energy conservation programs such as energy surveys. They have also offered energy efficiency programs that have provided rebates for such items as energy efficient refrigerators and insulation. Not counting the program covered by this survey, are you aware of any other energy conservation or energy efficiency programs?

1 ____ Yes → ENABLE Q12-13
2 ____ No **[SKIP TO Q15]**
88____I refuse to answer **[SKIP TO Q15]**
99____I don't know **[SKIP TO Q15]**

Q12. Which of the following energy conservation programs do you recall? (check all that apply)

01 ____ Rebates
02 ____ Product Give-Away/Turn-In Event (CFLs, Torchieres)
03 ____ Refrigerator Turn-In/Re-Cycling
04 ____ Home Repair/Retrofit (Insulation, Windows, Etc.)
05 ____ Energy Efficient Mortgages
06 ____ Energy Survey/Audit Delivered On-Site
07 ____ Energy Survey/Audit Delivered Through the Mail
08 ____ Energy Survey/Audit Delivered Over the Telephone
09 ____ Energy Survey/Audit Delivered Via the Internet
10 ____ Energy Survey/Audit Delivered At the Time Of Sale
11 ____ New Construction
12. ____ Flex Your Power
13. ___ CARE
61 ____ Other 1 (please specify): _____
62 ____ Other 2 (please specify): _____
88____I refuse to answer
99____I don't know

Past Participation in Programs

Q13. If you have participated in any of the programs listed below, please choose the most recent year you were enrolled in each program.

01 ____ Rebates → ENABLE Q14
02 ____ Product Give-Away/Turn-In Event (CFLs, Torchieres)
03 ____ Refrigerator Turn-In/Re-Cycling
04 ____ Home Repair/Retrofit (Insulation, Windows, Etc.)
05 ____ Energy Efficient Mortgages
06 ____ Energy Survey/Audit Delivered On-Site
07 ____ Energy Survey/Audit Delivered Through the Mail
08 ____ Energy Survey/Audit Delivered Over the Telephone
09 ____ Energy Survey/Audit Delivered Via the Internet
10 ____ Energy Survey/Audit Delivered At the Time Of Sale
11 ____ New Construction
12. ___ Flex Your Power
13. ___ CARE
61 ____ Other 1 (please specify): _____
62 ____ Other 2 (please specify): _____

NOTE THAT FOR EACH CHOICE A DROPDOWN LIST OF THE PAST 25 YEARS WAS INSERTED.

Q14. In the prior question, you indicated that you participated in a rebate program. For which of the following did you receive a rebate in the year you indicated? Please check all that apply.

01__ Attic Insulation
02 __Central AC
03 __Central Heat Pump
04 __Efficient Water Heater
05 __ Gas Furnace
06 __ Evaporative Cooler
07 __ High Performance Windows
08 __ Programmable Thermostats
09 __ Room Air Conditioner
10 __ Refrigerator
11 __ Water Heater
12 __ Water Heater Pipe Insulation
13 __ Wall Insulation
14 __ Water-Saving Shower Heads
15 __ Whole House Fan
16 __ Swimming Pool Items (Timers/Heaters/Vacuum Cleaners, Etc.)
17 __ Clothes washer
18 Dishwasher
19 Oven
77 Other (please specify): _____
88____I refuse to answer
99____I don't know

Background Characteristics

Q15. What type of home do you live in?

1 ____ Single family **attached** home
2 ____ Single family **detached** home
3 ____ An apartment with **less** than 5 units
4 ____ An apartment with five or **more** units
5 ____ Mobile home
6. ___ Boat, RV, or van etc.
77 ____ Other (please specify): _____
88____I refuse to answer

Q16. In what year was your home built?

____ ____ ____ ____ YEAR **[SKIP TO Q18]**

88____I refuse to answer **[SKIP TO Q18]**
99____I don't know **[CONTINUE]**

Q17. Can you approximate what year your home was built?

1 ____ After 2002
2 ____ Between 1999 to 2002
2 ____ Between 1995 to 1998
3 ____ Between 1990 and 1994
4 ____ Between 1980 and 1989
5 ____ Between 1970 and 1979
6 ____ Between 1960 and 1969?
7 ____ Between 1950 and 1959
8 ____ Between 1940 and 1949
9 ____ Before 1940
88____I refuse to answer
99____I don't know

Q18. *Approximately*, how many square feet of living space do you now have?

_____ SQUARE FEET
88____I refuse to answer
99____I don't know

Q19. How many rooms are there in your house?

Rooms [INSERT DROPDOWN LIST 1-9, listing each number, and 10+ rooms]
88____I refuse to answer
99____I don't know

Q20. How many people live at this residence?

People [INSERT DROPDOWN BOX 1-9, listing each number, and 10+ people]
88____I refuse to answer
99____I don't know

Q21. In terms of the ages of the residents in your home,

a. how many are 17 years or younger? [INSERT DROPDOWN BOX: 1-9 and 10+ people]
b. how many are between 18 and 21? [INSERT DROPDOWN BOX: 1-9 and 10+ people]
c. how many are between 22 and 59? [INSERT DROPDOWN BOX: 1-9 and 10+ people]
d. how many are 60 or over? [INSERT DROPDOWN BOX: 1-9 and 10+ people]

88____I refuse to answer
99____I don't know

Q22. What is the approximate annual household income from all sources in 2004, before taxes?

Reminder: **This information will be kept confidential**.

01 ____ Under $10,000
02 ____ $10,000 to $14,999
03 ____ $15,000 to $24,999
04 ____ $25,000 to $34,999
05 ____ $35,000 to $49,999
06 ____ $50,000 to $74,999
07 ____ $75,000 to $99,999
08 ____ $100,000 to $49,999
09 ____ $150,000 to $199,999
10 ____ $200,000 or more
88____I refuse to answer
99____I don't know

Q23. What is the highest level of education you have completed?

1 ____ Some High School (no diploma)
2 ____ High School Graduate (include high school equivalency)
3 ____ Trade or Technical School
4 ____ Some College (no degree)
5 ____ College Graduate
6 ____ Master's degree
7 ____ Professional school degree
8 ____ Doctorate
88____I refuse to answer
99____I don't know

Q24. In what year were you born? _____

Q25. What type of fuel do you use to heat your house? (Check all that apply)

1. ____ Gas from utility company
2. ____ Bottled, tank or LP gas
3. ____ Electricity
4. ____ Fuel oil, kerosene, etc.
5. ____ Coal or coke
6. ____ Wood
7. ____ Solar energy
8. ____ Other fuel (please specify): _____
9. ____ No fuel used

88 ____ I don't know

Q26. Which of the following best describes your racial or ethnic background? Please check all that apply.

1 ____ Hispanic or Latino
2 ____ Caucasian
3 ____ Black or African American
4. ____ Asian
5 ____ American Indian
6 ____ Native Alaskan
7 ____ Native Hawaiian and other Pacific Islander
77 ____ Other (please specify): _____
88____ I refuse to answer
99____ I don't know

Household Energy Use

IF SOUTHERN CALIFORNIA EDISON CUSTOMER, ASK:
Q27. Please check your most recent electricity bill from your utility and provide your **service** account number. This number appears near the top of your bill.

Service Account Number: _____ → ENABLE Q29
88____I refuse to answer → ENABLE Q29
99____I don't know → ENABLE Q29

FOR ALL OTHERS, ASK:
Q28. Please check your most recent electricity bill from your utility and provide your account number. This number appears near the top of your bill.
Account Number: _____
88____I refuse to answer
99____I don't know

Q29. Please check your most recent electricity bill from your utility and provide the kWh used for the period for which you are being billed: _____ kWh.

Q29a. What date marks the beginning of the period of time your most recent electricity bill covers?
1 Day/Month/Year (dd/mm/yyyy): _____
88____I refuse to answer
99____I don't know

Q29b. What date marks the end of the period of time your most recent electricity bill covers?
1 Day/Month/Year (dd/mm/yyyy): _____
88____I refuse to answer
99____I don't know

NOTE THAT A SCRIPT WAS INSERTED TO FORCE DATES INTO THE STANDARD DD/MM/YYYY FORMAT.

Q30. Please check your most recent gas bill from your gas utility and provide your account number.

Account Number: _____
88____I refuse to answer
99____I don't know

Q31. Please check your most recent gas bill from your gas utility and provide the therms used for the period for which you are being billed: _____ therms

Q31a. What date marks the beginning of the period of time your most recent gas bill covers?
1 Day/Month/Year (dd/mm/yyyy): _____
88____I refuse to answer
99____I don't know

Q31b. What date marks the end of the period of time your most recent gas bill covers?
1 Day/Month/Year (dd/mm/yyyy): _____
88____I refuse to answer
99____I don't know

NOTE THAT A SCRIPT WAS INSERTED TO FORCE DATES INTO THE STANDARD DD/MM/YYYY FORMAT.

Household Location

Q32. In what ZIP Code (5-digits) is your home located?

ZIP Code: _____

End: Thank you very much for your time and your help.

Appendix B

Experimental Design

Constructing a design starts with an attribute list. For this survey we decided on nine attributes with varying numbers of levels, as shown below in Table B-1.

Table B-1. Attributes, Number of Levels, and the Decomposition

Attribute or Factor	Number of levels	Decomposition
Mode of delivery	3	3^1
Time required to complete audit	4	2^2
Custom results	2	2^1
Post-audit technical support	2	2^1
Usage profile for customer's residence	2	2^1
Links to vendors of recommended equipment	3	3^1
Provide expected savings for each recommendation	3	3^1
Audit fee or incentive	6	$2^1 \times 3^1$
Time between decision to participate and results	4	2^2
Overall design		$2^8 \times 3^4$

Following this attribute profile, we constructed the design from two orthogonal arrays: one 2^8 array with 16 rows (called runs) and one 3^4 array with 27 rows. The arrays were combined by replication, such that each row in the 2^8 array was replicated for each row in the 3^4 array, so the total number of rows in the combined array was 16 x 27 = 432. This approach preserved both the balance and orthogonality[25] of the overall design and provided a good compromise between design resolution and tractability.

A simpler example shows how this was done. Assume we wanted to test a set of four attributes, two with two levels and two with three levels, often called a $2^2 3^2$ design. We can build the matrix for our design from the 2^2 array. This matrix is illustrated in Table B-2.

Table B-2. Matrix for 2^2 Array

0	0
0	1
1	0
1	1

[25] When a linear model is fit with an orthogonal design, the parameter estimates are uncorrelated, which means each estimate is independent of the other terms in the model. More importantly, orthogonality usually implies that the coefficients will have minimum variance. For these reasons, orthogonal designs are usually quite good.

and the 3^2 array in Table B-3.

Table B-3. 3^2 Array

0	0
0	1
0	2
1	0
1	1
1	2
2	0
2	1
2	2

In Table B-4, we start by taking the first row in the 2^2 array and replicating it with the entire 3^2 array.

Table B-4. 2^2 Array Replicated with Entire 3^2 Array

0	0	0	0
0	0	0	1
0	0	0	2
0	0	1	0
0	0	1	1
0	0	1	2
0	0	2	0
0	0	2	1
0	0	2	2

Repeating this for the remaining rows in the 2^2 array gives us our full $2^2 3^2$ design, illustrated in Table B-5.

Table B-5. Full $2^2 3^3$ Design

0	0	0	0
0	0	0	1
0	0	0	2
0	0	1	0
0	0	1	1
0	0	1	2
0	0	2	0
0	0	2	1
0	0	2	2
0	1	0	0
0	1	0	1
0	1	0	2
0	1	1	0
0	1	1	1

0	1	1	2
0	1	2	0
0	1	2	1
0	1	2	2
1	0	0	0
1	0	0	1
1	0	0	2
1	0	1	0
1	0	1	1
1	0	1	2
1	0	2	0
1	0	2	1
1	0	2	2
1	1	0	0
1	1	0	1
1	1	0	2
1	1	1	0
1	1	1	1
1	1	1	2
1	1	2	0
1	1	2	1
1	1	2	2

Another desirable property of a design is *resolution*, expressed as an integer value. The resolution of a design tells how many effects, including interaction terms, are estimable. Assume we have a design with three attributes, A, B and C. Normally we want to be able to estimate a model that includes A, B and C as separate terms. In some cases, however, we may want a model to include interaction terms such as AB or BC. Because they are the product of two attributes, they are called two-way interaction terms. A variable that is the product of all three, ABC, would be a three-way interaction term. Correspondingly, a design that permits only main effects (i.e., A, B and C as separate terms) to be estimated is called a resolution III design. One that permits main effects plus two-way interaction terms is a resolution IV design. A resolution V design permits three-way interaction terms, and so forth.

Returning to the design for the study, we chose specific 2^8 and 3^4 orthogonal arrays for our design so that we could maintain a resolution of IV. This resolution was chosen because we wanted to have the ability to estimate *two-factor interactions, and a resolution IV design permits us to do tha*t. The resulting 432 runs were organized into 54 choice sets (stores) using the method outlined by Kuhfeld (1994). To keep the number of stores presented to respondents at a manageable level, choice sets were divided into 6 blocks of 9 stores each. Every respondent was randomly assigned one of these 6 blocks and subsequently went through the 9 stores in that particular block.

Appendix C

Internet Survey Frequencies

Q23. What is the highest level of education you have completed?

			Utility			Total
			PG&E	SCE	SDG&E	
Q23. What is the highest level of education you have completed?	Some high school	Count	2	1	2	5
		% within Utility	.8%	.3%	.9%	.6%
	High-school graduate	Count	10	12	10	32
		% within Utility	4.2%	3.7%	4.3%	4.0%
	Trade school	Count	5	12	7	24
		% within Utility	2.1%	3.7%	3.0%	3.0%
	Some college	Count	53	82	45	180
		% within Utility	22.4%	25.2%	19.5%	22.7%
	College graduate	Count	98	126	89	313
		% within Utility	41.4%	38.8%	38.5%	39.5%
	Master's degree	Count	47	63	44	154
		% within Utility	19.8%	19.4%	19.0%	19.4%
	Professional school	Count	13	11	11	35
		% within Utility	5.5%	3.4%	4.8%	4.4%
	Doctorate	Count	5	12	18	35
		% within Utility	2.1%	3.7%	7.8%	4.4%
	Refused	Count	4	5	4	13
		% within Utility	1.7%	1.5%	1.7%	1.6%
	Don't know	Count	0	1	1	2
		% within Utility	.0%	.3%	.4%	.3%
Total		Count	237	325	231	793
		% within Utility	100.0%	100.0%	100.0%	100.0%

Q23. What is the highest level of education you have completed?

			Utility		Total
			Not SCG	SCG	
Q23. What is the highest level of education you have completed?	Some high school	Count	4	1	5
		% within SCG	.8%	.4%	.6%
	High-school graduate	Count	23	9	32
		% within SCG	4.4%	3.3%	4.0%
	Trade school	Count	14	10	24
		% within SCG	2.7%	3.7%	3.0%
	Some college	Count	107	73	180
		% within SCG	20.6%	26.7%	22.7%
	College graduate	Count	213	100	313
		% within SCG	41.0%	36.6%	39.5%
	Master's degree	Count	99	55	154
		% within SCG	19.0%	20.1%	19.4%
	Professional school	Count	26	9	35
		% within SCG	5.0%	3.3%	4.4%
	Doctorate	Count	26	9	35
		% within SCG	5.0%	3.3%	4.4%
	Refused	Count	8	5	13
		% within SCG	1.5%	1.8%	1.6%
	Don't know	Count	0	2	2
		% within SCG	.0%	.7%	.3%
Total		Count	520	273	793
		% within SCG	100.0%	100.0%	100.0%

Q15. What type of home do you live in?

			Utility			Total
			PG&E	SCE	SDG&E	
Q15. What type of home do you live in?	Single family attached	Count	69	91	60	220
		% within Utility	29.1%	28.0%	26.0%	27.7%
	Single family detached	Count	93	154	126	373
		% within Utility	39.2%	47.4%	54.5%	47.0%
	Apartment <5 units	Count	20	21	7	48
		% within Utility	8.4%	6.5%	3.0%	6.1%
	Apartment 5+ units	Count	48	56	35	139
		% within Utility	20.3%	17.2%	15.2%	17.5%
	Mobile home	Count	1	0	2	3
		% within Utility	.4%	.0%	.9%	.4%
	Refused	Count	2	0	0	2
		% within Utility	.8%	.0%	.0%	.3%
	Other	Count	4	3	1	8
		% within Utility	1.7%	.9%	.4%	1.0%
Total		Count	237	325	231	793
		% within Utility	100.0%	100.0%	100.0%	100.0%

Q15. What type of home do you live in?

			Utility		Total
			Not SCG	SCG	
Q15. What type of home do you live in?	Single family attached	Count	143	77	220
		% within SCG	27.5%	28.2%	27.7%
	Single family detached	Count	234	139	373
		% within SCG	45.0%	50.9%	47.0%
	Apartment <5 units	Count	29	19	48
		% within SCG	5.6%	7.0%	6.1%
	Apartment 5+ units	Count	104	35	139
		% within SCG	20.0%	12.8%	17.5%
	Mobile home	Count	3	0	3
		% within SCG	.6%	.0%	.4%
	Refused	Count	2	0	2
		% within SCG	.4%	.0%	.3%
	Other	Count	5	3	8
		% within SCG	1.0%	1.1%	1.0%
Total		Count	520	273	793
		% within SCG	100.0%	100.0%	100.0%

Q22. What is the approximate annual household income from all sources in 2004, before taxes?

			Utility			Total
			PG&E	SCE	SDG&E	
Q22. What is the approximate annual household income from all sources in 2004, before taxes? Reminder: This information will be kept confidential.	Under $10,000	Count	1	5	0	6
		% within Utility	.4%	1.5%	.0%	.8%
	$10,000 to $14,999	Count	0	4	0	4
		% within Utility	.0%	1.2%	.0%	.5%
	$15,000 to $24,999	Count	8	9	4	21
		% within Utility	3.4%	2.8%	1.7%	2.6%
	$25,000 to $34,999	Count	11	14	13	38
		% within Utility	4.6%	4.3%	5.6%	4.8%
	$35,000 to $49,999	Count	18	27	22	67
		% within Utility	7.6%	8.3%	9.5%	8.4%
	$50,000 to $74,999	Count	25	63	35	123
		% within Utility	10.5%	19.4%	15.2%	15.5%
	$75,000 to $99,999	Count	46	52	39	137
		% within Utility	19.4%	16.0%	16.9%	17.3%
	$100,000 to $149,999	Count	59	63	44	166
		% within Utility	24.9%	19.4%	19.0%	20.9%
	$150,000 to $199,999	Count	20	24	15	59
		% within Utility	8.4%	7.4%	6.5%	7.4%
	$200,000 or more	Count	18	14	20	52
		% within Utility	7.6%	4.3%	8.7%	6.6%
	Refused	Count	29	43	35	107
		% within Utility	12.2%	13.2%	15.2%	13.5%
	Don't know	Count	2	7	4	13
		% within Utility	.8%	2.2%	1.7%	1.6%
Total		Count	237	325	231	793
		% within Utility	100.0%	100.0%	100.0%	100.0%

Q22. What is the approximate annual household income from all sources in 2004, before taxes?

			Utility		Total
			Not SCG	SCG	
Q22. What is the approximate annual household income from all sources in 2004, before taxes? Reminder: This information will be kept confidential.	Under $10,000	Count	2	4	6
		% within SCG	.4%	1.5%	.8%
	$10,000 to $14,999	Count	1	3	4
		% within SCG	.2%	1.1%	.5%
	$15,000 to $24,999	Count	14	7	21
		% within SCG	2.7%	2.6%	2.6%
	$25,000 to $34,999	Count	27	11	38
		% within SCG	5.2%	4.0%	4.8%
	$35,000 to $49,999	Count	47	20	67
		% within SCG	9.0%	7.3%	8.4%
	$50,000 to $74,999	Count	75	48	123
		% within SCG	14.4%	17.6%	15.5%
	$75,000 to $99,999	Count	91	46	137
		% within SCG	17.5%	16.8%	17.3%
	$100,000 to $149,999	Count	107	59	166
		% within SCG	20.6%	21.6%	20.9%
	$150,000 to $199,999	Count	40	19	59
		% within SCG	7.7%	7.0%	7.4%
	$200,000 or more	Count	38	14	52
		% within SCG	7.3%	5.1%	6.6%
	Refused	Count	70	37	107
		% within SCG	13.5%	13.6%	13.5%
	Don't know	Count	8	5	13
		% within SCG	1.5%	1.8%	1.6%
Total		Count	520	273	793
		% within SCG	100.0%	100.0%	100.0%

Q7 Do you own or rent your home?

			Utility			Total
			PG&E	SCE	SDG&E	
Q7 Do you own or rent your home?	Own	Count	160	221	175	556
		% within Utility	67.5%	68.0%	75.8%	70.1%
	Rent	Count	75	101	56	232
		% within Utility	31.6%	31.1%	24.2%	29.3%
	Other	Count	2	2	0	4
		% within Utility	.8%	.6%	.0%	.5%
	Refused	Count	0	1	0	1
		% within Utility	.0%	.3%	.0%	.1%
Total		Count	237	325	231	793
		% within Utility	100.0%	100.0%	100.0%	100.0%

Q7 Do you own or rent your home?

			Utility		Total
			Not SCG	SCG	
Q7 Do you own or rent your home?	Own	Count	360	196	556
		% within SCG	69.2%	71.8%	70.1%
	Rent	Count	158	74	232
		% within SCG	30.4%	27.1%	29.3%
	Other	Count	2	2	4
		% within SCG	.4%	.7%	.5%
	Refused	Count	0	1	1
		% within SCG	.0%	.4%	.1%
Total		Count	520	273	793
		% within SCG	100.0%	100.0%	100.0%

Q20. How many people live at this residence?

			Utility			Total
			PG&E	SCE	SDG&E	
Q20. How many people live at this residence?	1	Count	30	49	30	109
		% within Utility	12.7%	15.1%	13.0%	13.7%
	2	Count	92	114	85	291
		% within Utility	38.8%	35.1%	36.8%	36.7%
	3	Count	50	65	48	163
		% within Utility	21.1%	20.0%	20.8%	20.6%
	4	Count	40	52	46	138
		% within Utility	16.9%	16.0%	19.9%	17.4%
	5	Count	17	22	7	46
		% within Utility	7.2%	6.8%	3.0%	5.8%
	6	Count	4	11	8	23
		% within Utility	1.7%	3.4%	3.5%	2.9%
	7	Count	1	4	2	7
		% within Utility	.4%	1.2%	.9%	.9%
	8	Count	2	2	1	5
		% within Utility	.8%	.6%	.4%	.6%
	9	Count	0	0	1	1
		% within Utility	.0%	.0%	.4%	.1%
	10	Count	0	1	0	1
		% within Utility	.0%	.3%	.0%	.1%
	12	Count	1	5	3	9
		% within Utility	.4%	1.5%	1.3%	1.1%
Total		Count	237	325	231	793
		% within Utility	100.0%	100.0%	100.0%	100.0%

Q20. How many people live at this residence?

			Utility		Total
			Not SCG	SCG	
Q20. How many people live at this residence?	1	Count	64	45	109
		% within SCG	12.3%	16.5%	13.7%
	2	Count	197	94	291
		% within SCG	37.9%	34.4%	36.7%
	3	Count	111	52	163
		% within SCG	21.3%	19.0%	20.6%
	4	Count	94	44	138
		% within SCG	18.1%	16.1%	17.4%
	5	Count	25	21	46
		% within SCG	4.8%	7.7%	5.8%
	6	Count	16	7	23
		% within SCG	3.1%	2.6%	2.9%
	7	Count	4	3	7
		% within SCG	.8%	1.1%	.9%
	8	Count	3	2	5
		% within SCG	.6%	.7%	.6%
	9	Count	1	0	1
		% within SCG	.2%	.0%	.1%
	10	Count	0	1	1
		% within SCG	.0%	.4%	.1%
	12	Count	5	4	9
		% within SCG	1.0%	1.5%	1.1%
Total		Count	520	273	793
		% within SCG	100.0%	100.0%	100.0%

Q19. How many rooms are there in your house?

			Utility			Total
			PG&E	SCE	SDG&E	
Q19. How many rooms are there in your house?	1	Count	7	11	7	25
		% within Utility	3.0%	3.4%	3.0%	3.2%
	2	Count	21	33	18	72
		% within Utility	8.9%	10.2%	7.8%	9.1%
	3	Count	31	42	26	99
		% within Utility	13.1%	12.9%	11.3%	12.5%
	4	Count	36	54	27	117
		% within Utility	15.2%	16.6%	11.7%	14.8%
	5	Count	19	32	24	75
		% within Utility	8.0%	9.8%	10.4%	9.5%
	6	Count	21	40	32	93
		% within Utility	8.9%	12.3%	13.9%	11.7%
	7	Count	31	37	27	95
		% within Utility	13.1%	11.4%	11.7%	12.0%
	8	Count	27	29	19	75
		% within Utility	11.4%	8.9%	8.2%	9.5%
	9	Count	17	17	14	48
		% within Utility	7.2%	5.2%	6.1%	6.1%
	10+	Count	25	28	34	87
		% within Utility	10.5%	8.6%	14.7%	11.0%
	Don't know	Count	1	2	1	4
		% within Utility	.4%	.6%	.4%	.5%
	Refused	Count	1	0	2	3
		% within Utility	.4%	.0%	.9%	.4%
Total		Count	237	325	231	793
		% within Utility	100.0%	100.0%	100.0%	100.0%

Q19. How many rooms are there in your house?

			Utility		Total
			Not SCG	SCG	
Q19. How many rooms are there in your house?	1	Count	17	8	25
		% within SCG	3.3%	2.9%	3.2%
	2	Count	49	23	72
		% within SCG	9.4%	8.4%	9.1%
	3	Count	62	37	99
		% within SCG	11.9%	13.6%	12.5%
	4	Count	77	40	117
		% within SCG	14.8%	14.7%	14.8%
	5	Count	49	26	75
		% within SCG	9.4%	9.5%	9.5%
	6	Count	59	34	93
		% within SCG	11.3%	12.5%	11.7%
	7	Count	62	33	95
		% within SCG	11.9%	12.1%	12.0%
	8	Count	46	29	75
		% within SCG	8.8%	10.6%	9.5%
	9	Count	33	15	48
		% within SCG	6.3%	5.5%	6.1%
	10+	Count	60	27	87
		% within SCG	11.5%	9.9%	11.0%
	Don't know	Count	3	1	4
		% within SCG	.6%	.4%	.5%
	Refused	Count	3	0	3
		% within SCG	.6%	.0%	.4%
Total		Count	520	273	793
		% within SCG	100.0%	100.0%	100.0%

Q21a. How many are 17 years or younger?

			Utility			Total
			PG&E	SCE	SDG&E	
a. how many are 17 years or younger?	1	Count	155	198	141	494
		% within Utility	65.4%	60.9%	61.0%	62.3%
	2	Count	47	56	48	151
		% within Utility	19.8%	17.2%	20.8%	19.0%
	3	Count	25	43	30	98
		% within Utility	10.5%	13.2%	13.0%	12.4%
	4	Count	6	13	5	24
		% within Utility	2.5%	4.0%	2.2%	3.0%
	5	Count	0	7	3	10
		% within Utility	.0%	2.2%	1.3%	1.3%
	6	Count	3	2	1	6
		% within Utility	1.3%	.6%	.4%	.8%
	7	Count	0	1	0	1
		% within Utility	.0%	.3%	.0%	.1%
	12	Count	0	1	0	1
		% within Utility	.0%	.3%	.0%	.1%
	13	Count	1	4	3	8
		% within Utility	.4%	1.2%	1.3%	1.0%
Total		Count	237	325	231	793
		% within Utility	100.0%	100.0%	100.0%	100.0%

Q21a. How many are 17 years or younger?

			Utility		Total
			Not SCG	SCG	
a. how many are 17 years or younger?	1	Count	327	167	494
		% within SCG	62.9%	61.2%	62.3%
	2	Count	105	46	151
		% within SCG	20.2%	16.8%	19.0%
	3	Count	61	37	98
		% within SCG	11.7%	13.6%	12.4%
	4	Count	12	12	24
		% within SCG	2.3%	4.4%	3.0%
	5	Count	6	4	10
		% within SCG	1.2%	1.5%	1.3%
	6	Count	5	1	6
		% within SCG	1.0%	.4%	.8%
	7	Count	0	1	1
		% within SCG	.0%	.4%	.1%
	12	Count	0	1	1
		% within SCG	.0%	.4%	.1%
	13	Count	4	4	8
		% within SCG	.8%	1.5%	1.0%
Total		Count	520	273	793
		% within SCG	100.0%	100.0%	100.0%

Q21b. How many are between 18 and 21?Utility

			Utility			Total
			PG&E	SCE	SDG&E	
b. how many are between 18 and 21?	1	Count	192	294	202	688
		% within Utility	81.0%	90.5%	87.4%	86.8%
	2	Count	34	20	21	75
		% within Utility	14.3%	6.2%	9.1%	9.5%
	3	Count	8	5	5	18
		% within Utility	3.4%	1.5%	2.2%	2.3%
	4	Count	2	0	0	2
		% within Utility	.8%	.0%	.0%	.3%
	6	Count	0	1	0	1
		% within Utility	.0%	.3%	.0%	.1%
	12	Count	0	1	0	1
		% within Utility	.0%	.3%	.0%	.1%
	13	Count	1	4	3	8
		% within Utility	.4%	1.2%	1.3%	1.0%
Total		Count	237	325	231	793
		% within Utility	100.0%	100.0%	100.0%	100.0%

Q21b. How many are between 18 and 21?

			Utility		Total
			Not SCG	SCG	
b. how many are between 18 and 21?	1	Count	440	248	688
		% within SCG	84.6%	90.8%	86.8%
	2	Count	58	17	75
		% within SCG	11.2%	6.2%	9.5%
	3	Count	15	3	18
		% within SCG	2.9%	1.1%	2.3%
	4	Count	2	0	2
		% within SCG	.4%	.0%	.3%
	6	Count	1	0	1
		% within SCG	.2%	.0%	.1%
	12	Count	0	1	1
		% within SCG	.0%	.4%	.1%
	13	Count	4	4	8
		% within SCG	.8%	1.5%	1.0%
Total		Count	520	273	793
		% within SCG	100.0%	100.0%	100.0%

Q21c. How many are between 22 and 59?

			Utility			Total
			PG&E	SCE	SDG&E	
c. how many are between 22 and 59?	1	Count	15	15	23	53
		% within Utility	6.3%	4.6%	10.0%	6.7%
	2	Count	45	78	49	172
		% within Utility	19.0%	24.0%	21.2%	21.7%
	3	Count	151	189	133	473
		% within Utility	63.7%	58.2%	57.6%	59.6%
	4	Count	18	27	15	60
		% within Utility	7.6%	8.3%	6.5%	7.6%
	5	Count	6	7	6	19
		% within Utility	2.5%	2.2%	2.6%	2.4%
	6	Count	2	1	2	5
		% within Utility	.8%	.3%	.9%	.6%
	7	Count	0	2	0	2
		% within Utility	.0%	.6%	.0%	.3%
	8	Count	0	1	0	1
		% within Utility	.0%	.3%	.0%	.1%
	12	Count	0	2	0	2
		% within Utility	.0%	.6%	.0%	.3%
	13	Count	0	3	3	6
		% within Utility	.0%	.9%	1.3%	.8%
Total		Count	237	325	231	793
		% within Utility	100.0%	100.0%	100.0%	100.0%

Q21c. How many are between 22 and 59?

			Utility		Total
			Not SCG	SCG	
c. how many are between 22 and 59?	1	Count	38	15	53
		% within SCG	7.3%	5.5%	6.7%
	2	Count	100	72	172
		% within SCG	19.2%	26.4%	21.7%
	3	Count	324	149	473
		% within SCG	62.3%	54.6%	59.6%
	4	Count	37	23	60
		% within SCG	7.1%	8.4%	7.6%
	5	Count	14	5	19
		% within SCG	2.7%	1.8%	2.4%
	6	Count	4	1	5
		% within SCG	.8%	.4%	.6%
	7	Count	0	2	2
		% within SCG	.0%	.7%	.3%
	8	Count	0	1	1
		% within SCG	.0%	.4%	.1%
	12	Count	0	2	2
		% within SCG	.0%	.7%	.3%
	13	Count	3	3	6
		% within SCG	.6%	1.1%	.8%
Total		Count	520	273	793
		% within SCG	100.0%	100.0%	100.0%

Q21d. How many are 60 or over?

			Utility			Total
			PG&E	SCE	SDG&E	
d. how many are 60 or over?	1	Count	204	269	182	655
		% within Utility	86.1%	82.8%	78.8%	82.6%
	2	Count	17	35	25	77
		% within Utility	7.2%	10.8%	10.8%	9.7%
	3	Count	12	17	21	50
		% within Utility	5.1%	5.2%	9.1%	6.3%
	4	Count	1	0	0	1
		% within Utility	.4%	.0%	.0%	.1%
	8	Count	1	0	0	1
		% within Utility	.4%	.0%	.0%	.1%
	12	Count	0	1	0	1
		% within Utility	.0%	.3%	.0%	.1%
	13	Count	2	3	3	8
		% within Utility	.8%	.9%	1.3%	1.0%
Total		Count	237	325	231	793
		% within Utility	100.0%	100.0%	100.0%	100.0%

Q21d. How many are 60 or over?

			Utility		Total
			Not SCG	SCG	
d. how many are 60 or over?	1	Count	430	225	655
		% within SCG	82.7%	82.4%	82.6%
	2	Count	49	28	77
		% within SCG	9.4%	10.3%	9.7%
	3	Count	34	16	50
		% within SCG	6.5%	5.9%	6.3%
	4	Count	1	0	1
		% within SCG	.2%	.0%	.1%
	8	Count	1	0	1
		% within SCG	.2%	.0%	.1%
	12	Count	0	1	1
		% within SCG	.0%	.4%	.1%
	13	Count	5	3	8
		% within SCG	1.0%	1.1%	1.0%
Total		Count	520	273	793
		% within SCG	100.0%	100.0%	100.0%

Q26. Ethnicity

			Utility			Total
			PG&E	SCE	SDG&E	
Ethnicity of Respondent	Asian	Count	41	43	25	109
		% within Utility	24.0%	19.0%	13.7%	
	Black or African American	Count	4	13	3	20
		% within Utility	2.3%	5.8%	1.6%	
	Caucasian	Count	127	169	152	448
		% within Utility	74.3%	74.8%	83.5%	
	American Indian	Count	1	3	1	5
		% within Utility	.6%	1.3%	.5%	
	Pacific Islander	Count	2	4	4	10
		% within Utility	1.2%	1.8%	2.2%	
Total		Count	171	226	182	579

Percentages and totals are based on respondents.
a Dichotomy group tabulated at value 1.

Q26. Ethnicity

			Utility		Total
			Not SCG	SCG	
Ethnicity of Respondent	Asian	Count	75	34	109
		% within SCG	19.4%	17.7%	
	Black or African American	Count	13	7	20
		% within SCG	3.4%	3.6%	
	Caucasian	Count	298	150	448
		% within SCG	77.0%	78.1%	
	American Indian	Count	3	2	5
		% within SCG	.8%	1.0%	
	Pacific Islander	Count	7	3	10
		% within SCG	1.8%	1.6%	
Total		Count	387	192	579

Percentages and totals are based on respondents.
a Dichotomy group tabulated at value 1.

Q25. Heating Fuel

			Utility			Total
			PG&E	SCE	SDG&E	
Fuel Used for Heating	Gas from utility company	Count	189	265	172	626
		% within Utility	79.7%	81.5%	74.5%	
	Electricity	Count	75	100	81	256
		% within Utility	31.6%	30.8%	35.1%	
	Bottled, tank or LP gas	Count	6	4	4	14
		% within Utility	2.5%	1.2%	1.7%	
	Fuel oil, kerosene	Count	0	0	2	2
		% within Utility	.0%	.0%	.9%	
	Coal or coke	Count	1	0	0	1
		% within Utility	.4%	.0%	.0%	
	Solar energy	Count	1	3	3	7
		% within Utility	.4%	.9%	1.3%	
	Wood	Count	19	16	22	57
		% within Utility	8.0%	4.9%	9.5%	
	Other fuel	Count	0	8	8	16
		% within Utility	.0%	2.5%	3.5%	
	No fuel used	Count	2	3	6	11
		% within Utility	.8%	.9%	2.6%	
	Don't know	Count	0	2	0	2
		% within Utility	.0%	.6%	.0%	
Total		Count	237	325	231	793

Percentages and totals are based on respondents.
a Dichotomy group tabulated at value 1.

Q25. Heating Fuel

			Utility		Total
			Not SCG	SCG	
Fuel Used for Heating	Gas from utility company	Count	384	242	626
		% within SCG	73.8%	88.6%	
	Electricity	Count	190	66	256
		% within SCG	36.5%	24.2%	
	Bottled, tank or LP gas	Count	12	2	14
		% within SCG	2.3%	.7%	
	Fuel oil, kerosene	Count	2	0	2
		% within SCG	.4%	.0%	
	Coal or coke	Count	1	0	1
		% within SCG	.2%	.0%	
	Solar energy	Count	4	3	7
		% within SCG	.8%	1.1%	
	Wood	Count	43	14	57
		% within SCG	8.3%	5.1%	
	Other fuel	Count	12	4	16
		% within SCG	2.3%	1.5%	
	No fuel used	Count	9	2	11
		% within SCG	1.7%	.7%	
	Don't know	Count	1	1	2
		% within SCG	.2%	.4%	
Total		Count	520	273	793

Percentages and totals are based on respondents.
a Dichotomy group tabulated at value 1.

Q8a. My life is too busy to worry about making energy related improvements to my home.

			Utility			Total
			PG&E	SCE	SDG&E	
My life is too busy to worry about making energy related improvements to my home.	Strongly Disagree	Count	56	75	53	184
		% within Utility	23.6%	23.1%	22.9%	23.2%
	2	Count	17	35	25	77
		% within Utility	7.2%	10.8%	10.8%	9.7%
	3	Count	46	49	43	138
		% within Utility	19.4%	15.1%	18.6%	17.4%
	4	Count	27	41	24	92
		% within Utility	11.4%	12.6%	10.4%	11.6%
	5	Count	23	54	36	113
		% within Utility	9.7%	16.6%	15.6%	14.2%
	6	Count	27	14	20	61
		% within Utility	11.4%	4.3%	8.7%	7.7%
	7	Count	21	25	13	59
		% within Utility	8.9%	7.7%	5.6%	7.4%
	8	Count	6	8	8	22
		% within Utility	2.5%	2.5%	3.5%	2.8%
	9	Count	2	9	2	13
		% within Utility	.8%	2.8%	.9%	1.6%
	Strongly Agree	Count	11	14	5	30
		% within Utility	4.6%	4.3%	2.2%	3.8%
	Don't Know	Count	1	1	2	4
		% within Utility	.4%	.3%	.9%	.5%
Total		Count	237	325	231	793
		% within Utility	100.0%	100.0%	100.0%	100.0%

Q8a. My life is too busy to worry about making energy related improvements to my home.

			Utility		Total
			Not SCG	SCG	
My life is too busy to worry about making energy related improvements to my home.	Strongly Disagree	Count	122	62	184
		% within SCG	23.5%	22.7%	23.2%
	2	Count	49	28	77
		% within SCG	9.4%	10.3%	9.7%
	3	Count	93	45	138
		% within SCG	17.9%	16.5%	17.4%
	4	Count	61	31	92
		% within SCG	11.7%	11.4%	11.6%
	5	Count	66	47	113
		% within SCG	12.7%	17.2%	14.2%
	6	Count	49	12	61
		% within SCG	9.4%	4.4%	7.7%
	7	Count	37	22	59
		% within SCG	7.1%	8.1%	7.4%
	8	Count	16	6	22
		% within SCG	3.1%	2.2%	2.8%
	9	Count	4	9	13
		% within SCG	.8%	3.3%	1.6%
	Strongly Agree	Count	20	10	30
		% within SCG	3.8%	3.7%	3.8%
	Don't Know	Count	3	1	4
		% within SCG	.6%	.4%	.5%
Total		Count	520	273	793
		% within SCG	100.0%	100.0%	100.0%

Q8b. Scarce energy supplies will be a major problem in the future.

			Utility			Total
			PG&E	SCE	SDG&E	
Scarce energy supplies will be a major problem in the future	Strongly Disagree	Count	3	11	6	20
		% within Utility	1.3%	3.4%	2.6%	2.5%
	2	Count	0	3	3	6
		% within Utility	.0%	.9%	1.3%	.8%
	3	Count	3	5	6	14
		% within Utility	1.3%	1.5%	2.6%	1.8%
	4	Count	5	6	6	17
		% within Utility	2.1%	1.8%	2.6%	2.1%
	5	Count	14	29	13	56
		% within Utility	5.9%	8.9%	5.6%	7.1%
	6	Count	15	20	12	47
		% within Utility	6.3%	6.2%	5.2%	5.9%
	7	Count	32	32	25	89
		% within Utility	13.5%	9.8%	10.8%	11.2%
	8	Count	32	33	32	97
		% within Utility	13.5%	10.2%	13.9%	12.2%
	9	Count	31	41	33	105
		% within Utility	13.1%	12.6%	14.3%	13.2%
	Strongly Agree	Count	94	138	90	322
		% within Utility	39.7%	42.5%	39.0%	40.6%
	Don't Know	Count	8	7	5	20
		% within Utility	3.4%	2.2%	2.2%	2.5%
Total		Count	237	325	231	793
		% within Utility	100.0%	100.0%	100.0%	100.0%

Q8b. Scarce energy supplies will be a major problem in the future

			Utility		Total
			Not SCG	SCG	
Scarce energy supplies will be a major problem in the future	Strongly Disagree	Count	10	10	20
		% within SCG	1.9%	3.7%	2.5%
	2	Count	4	2	6
		% within SCG	.8%	.7%	.8%
	3	Count	8	6	14
		% within SCG	1.5%	2.2%	1.8%
	4	Count	12	5	17
		% within SCG	2.3%	1.8%	2.1%
	5	Count	29	27	56
		% within SCG	5.6%	9.9%	7.1%
	6	Count	32	15	47
		% within SCG	6.2%	5.5%	5.9%
	7	Count	59	30	89
		% within SCG	11.3%	11.0%	11.2%
	8	Count	70	27	97
		% within SCG	13.5%	9.9%	12.2%
	9	Count	68	37	105
		% within SCG	13.1%	13.6%	13.2%
	Strongly Agree	Count	214	108	322
		% within SCG	41.2%	39.6%	40.6%
	Don't Know	Count	14	6	20
		% within SCG	2.7%	2.2%	2.5%
Total		Count	520	273	793
		% within SCG	100.0%	100.0%	100.0%

Q8c. Instead of building new power plants, customers should use less electricity.

			Utility			Total
			PG&E	SCE	SDG&E	
Instead of building new power plants, customers should use less electricity	Strongly Disagree	Count	20	24	15	59
		% within Utility	8.4%	7.4%	6.5%	7.4%
	2	Count	6	13	13	32
		% within Utility	2.5%	4.0%	5.6%	4.0%
	3	Count	16	27	21	64
		% within Utility	6.8%	8.3%	9.1%	8.1%
	4	Count	19	20	26	65
		% within Utility	8.0%	6.2%	11.3%	8.2%
	5	Count	57	76	51	184
		% within Utility	24.1%	23.4%	22.1%	23.2%
	6	Count	27	43	33	103
		% within Utility	11.4%	13.2%	14.3%	13.0%
	7	Count	24	37	28	89
		% within Utility	10.1%	11.4%	12.1%	11.2%
	8	Count	26	33	17	76
		% within Utility	11.0%	10.2%	7.4%	9.6%
	9	Count	14	14	7	35
		% within Utility	5.9%	4.3%	3.0%	4.4%
	Strongly Agree	Count	22	30	18	70
		% within Utility	9.3%	9.2%	7.8%	8.8%
	Don't Know	Count	6	8	2	16
		% within Utility	2.5%	2.5%	.9%	2.0%
Total		Count	237	325	231	793
		% within Utility	100.0%	100.0%	100.0%	100.0%

Q8c. Instead of building new power plants, customers should use less electricity.

			Utility		Total
			Not SCG	SCG	
Instead of building new power plants, customers should use less electricity	Strongly Disagree	Count	37	22	59
		% within SCG	7.1%	8.1%	7.4%
	2	Count	18	14	32
		% within SCG	3.5%	5.1%	4.0%
	3	Count	42	22	64
		% within SCG	8.1%	8.1%	8.1%
	4	Count	47	18	65
		% within SCG	9.0%	6.6%	8.2%
	5	Count	123	61	184
		% within SCG	23.7%	22.3%	23.2%
	6	Count	69	34	103
		% within SCG	13.3%	12.5%	13.0%
	7	Count	58	31	89
		% within SCG	11.2%	11.4%	11.2%
	8	Count	47	29	76
		% within SCG	9.0%	10.6%	9.6%
	9	Count	24	11	35
		% within SCG	4.6%	4.0%	4.4%
	Strongly Agree	Count	47	23	70
		% within SCG	9.0%	8.4%	8.8%
	Don't Know	Count	8	8	16
		% within SCG	1.5%	2.9%	2.0%
Total		Count	520	273	793
		% within SCG	100.0%	100.0%	100.0%

Q8d. It is possible to save energy without sacrificing comfort by being energy efficient.

			Utility			Total
			PG&E	SCE	SDG&E	
It is possible to save energy without sacrificing comfort by being energy efficient	Strongly Disagree	Count	6	5	3	14
		% within Utility	2.5%	1.5%	1.3%	1.8%
	2	Count	1	4	1	6
		% within Utility	.4%	1.2%	.4%	.8%
	3	Count	5	10	1	16
		% within Utility	2.1%	3.1%	.4%	2.0%
	4	Count	6	9	3	18
		% within Utility	2.5%	2.8%	1.3%	2.3%
	5	Count	25	18	19	62
		% within Utility	10.5%	5.5%	8.2%	7.8%
	6	Count	14	20	20	54
		% within Utility	5.9%	6.2%	8.7%	6.8%
	7	Count	37	50	30	117
		% within Utility	15.6%	15.4%	13.0%	14.8%
	8	Count	47	56	57	160
		% within Utility	19.8%	17.2%	24.7%	20.2%
	9	Count	27	61	39	127
		% within Utility	11.4%	18.8%	16.9%	16.0%
	Strongly Agree	Count	65	87	57	209
		% within Utility	27.4%	26.8%	24.7%	26.4%
	Don't Know	Count	4	5	1	10
		% within Utility	1.7%	1.5%	.4%	1.3%
Total		Count	237	325	231	793
		% within Utility	100.0%	100.0%	100.0%	100.0%

Q8d. It is possible to save energy without sacrificing comfort by being energy efficient.

			Utility		Total
			Not SCG	SCG	
It is possible to save energy without sacrificing comfort by being energy efficient	Strongly Disagree	Count	12	2	14
		% within SCG	2.3%	.7%	1.8%
	2	Count	2	4	6
		% within SCG	.4%	1.5%	.8%
	3	Count	8	8	16
		% within SCG	1.5%	2.9%	2.0%
	4	Count	9	9	18
		% within SCG	1.7%	3.3%	2.3%
	5	Count	44	18	62
		% within SCG	8.5%	6.6%	7.8%
	6	Count	35	19	54
		% within SCG	6.7%	7.0%	6.8%
	7	Count	76	41	117
		% within SCG	14.6%	15.0%	14.8%
	8	Count	109	51	160
		% within SCG	21.0%	18.7%	20.2%
	9	Count	79	48	127
		% within SCG	15.2%	17.6%	16.0%
	Strongly Agree	Count	141	68	209
		% within SCG	27.1%	24.9%	26.4%
	Don't Know	Count	5	5	10
		% within SCG	1.0%	1.8%	1.3%
Total		Count	520	273	793
		% within SCG	100.0%	100.0%	100.0%

Q8e. It is worth it to me for my household to use less energy in order to help preserve the environment.

			Utility			Total
			PG&E	SCE	SDG&E	
It is worth it to me for my household to use less energy in order to help preserve the environment	Strongly Disagree	Count	4	6	7	17
		% within Utility	1.7%	1.8%	3.0%	2.1%
	2	Count	4	2	5	11
		% within Utility	1.7%	.6%	2.2%	1.4%
	3	Count	7	6	5	18
		% within Utility	3.0%	1.8%	2.2%	2.3%
	4	Count	7	12	7	26
		% within Utility	3.0%	3.7%	3.0%	3.3%
	5	Count	26	31	32	89
		% within Utility	11.0%	9.5%	13.9%	11.2%
	6	Count	22	33	26	81
		% within Utility	9.3%	10.2%	11.3%	10.2%
	7	Count	31	49	36	116
		% within Utility	13.1%	15.1%	15.6%	14.6%
	8	Count	40	58	33	131
		% within Utility	16.9%	17.8%	14.3%	16.5%
	9	Count	30	49	29	108
		% within Utility	12.7%	15.1%	12.6%	13.6%
	Strongly Agree	Count	65	76	50	191
		% within Utility	27.4%	23.4%	21.6%	24.1%
	Don't Know	Count	1	3	1	5
		% within Utility	.4%	.9%	.4%	.6%
Total		Count	237	325	231	793
		% within Utility	100.0%	100.0%	100.0%	100.0%

Q8e. It is worth it to me for my household to use less energy in order to help preserve the environment.

			Utility		Total
			Not SCG	SCG	
It is worth it to me for my household to use less energy in order to help preserve the environment	Strongly Disagree	Count	13	4	17
		% within SCG	2.5%	1.5%	2.1%
	2	Count	9	2	11
		% within SCG	1.7%	.7%	1.4%
	3	Count	12	6	18
		% within SCG	2.3%	2.2%	2.3%
	4	Count	17	9	26
		% within SCG	3.3%	3.3%	3.3%
	5	Count	58	31	89
		% within SCG	11.2%	11.4%	11.2%
	6	Count	49	32	81
		% within SCG	9.4%	11.7%	10.2%
	7	Count	78	38	116
		% within SCG	15.0%	13.9%	14.6%
	8	Count	84	47	131
		% within SCG	16.2%	17.2%	16.5%
	9	Count	65	43	108
		% within SCG	12.5%	15.8%	13.6%
	Strongly Agree	Count	133	58	191
		% within SCG	25.6%	21.2%	24.1%
	Don't Know	Count	2	3	5
		% within SCG	.4%	1.1%	.6%
Total		Count	520	273	793
		% within SCG	100.0%	100.0%	100.0%

Q8f. Conservation efforts helped reduce the effects of the energy crisis during the summer of 2001.

			Utility			Total
			PG&E	SCE	SDG&E	
Conservation efforts helped reduce the effects of the energy crisis during the summer of 2001	Strongly Disagree	Count	10	11	10	31
		% within Utility	4.2%	3.4%	4.3%	3.9%
	2	Count	1	9	2	12
		% within Utility	.4%	2.8%	.9%	1.5%
	3	Count	11	8	10	29
		% within Utility	4.6%	2.5%	4.3%	3.7%
	4	Count	7	8	9	24
		% within Utility	3.0%	2.5%	3.9%	3.0%
	5	Count	24	44	32	100
		% within Utility	10.1%	13.5%	13.9%	12.6%
	6	Count	29	31	27	87
		% within Utility	12.2%	9.5%	11.7%	11.0%
	7	Count	28	37	33	98
		% within Utility	11.8%	11.4%	14.3%	12.4%
	8	Count	30	39	27	96
		% within Utility	12.7%	12.0%	11.7%	12.1%
	9	Count	22	34	21	77
		% within Utility	9.3%	10.5%	9.1%	9.7%
	Strongly Agree	Count	43	52	27	122
		% within Utility	18.1%	16.0%	11.7%	15.4%
	Don't Know	Count	32	52	33	117
		% within Utility	13.5%	16.0%	14.3%	14.8%
Total		Count	237	325	231	793
		% within Utility	100.0%	100.0%	100.0%	100.0%

Q8f. Conservation efforts helped reduce the effects of the energy crisis during the summer of 2001.

			Utility		Total
			Not SCG	SCG	
Conservation efforts helped reduce the effects of the energy crisis during the summer of 2001	Strongly Disagree	Count	23	8	31
		% within SCG	4.4%	2.9%	3.9%
	2	Count	5	7	12
		% within SCG	1.0%	2.6%	1.5%
	3	Count	20	9	29
		% within SCG	3.8%	3.3%	3.7%
	4	Count	17	7	24
		% within SCG	3.3%	2.6%	3.0%
	5	Count	65	35	100
		% within SCG	12.5%	12.8%	12.6%
	6	Count	52	35	87
		% within SCG	10.0%	12.8%	11.0%
	7	Count	66	32	98
		% within SCG	12.7%	11.7%	12.4%
	8	Count	66	30	96
		% within SCG	12.7%	11.0%	12.1%
	9	Count	48	29	77
		% within SCG	9.2%	10.6%	9.7%
	Strongly Agree	Count	81	41	122
		% within SCG	15.6%	15.0%	15.4%
	Don't Know	Count	77	40	117
		% within SCG	14.8%	14.7%	14.8%
Total		Count	520	273	793
		% within SCG	100.0%	100.0%	100.0%

Q8g. Conserving energy in my home is an economic necessity.

			Utility			Total
			PG&E	SCE	SDG&E	
Conserving energy in my home is an economic necessity	Strongly Disagree	Count	11	10	11	32
		% within Utility	4.6%	3.1%	4.8%	4.0%
	2	Count	10	17	7	34
		% within Utility	4.2%	5.2%	3.0%	4.3%
	3	Count	10	14	17	41
		% within Utility	4.2%	4.3%	7.4%	5.2%
	4	Count	15	24	13	52
		% within Utility	6.3%	7.4%	5.6%	6.6%
	5	Count	31	33	30	94
		% within Utility	13.1%	10.2%	13.0%	11.9%
	6	Count	22	41	29	92
		% within Utility	9.3%	12.6%	12.6%	11.6%
	7	Count	34	38	38	110
		% within Utility	14.3%	11.7%	16.5%	13.9%
	8	Count	36	49	25	110
		% within Utility	15.2%	15.1%	10.8%	13.9%
	9	Count	22	37	25	84
		% within Utility	9.3%	11.4%	10.8%	10.6%
	Strongly Agree	Count	43	60	36	139
		% within Utility	18.1%	18.5%	15.6%	17.5%
	Don't Know	Count	3	2	0	5
		% within Utility	1.3%	.6%	.0%	.6%
Total		Count	237	325	231	793
		% within Utility	100.0%	100.0%	100.0%	100.0%

Q8g. Conserving energy in my home is an economic necessity.

			Utility		Total
			Not SCG	SCG	
Conserving energy in my home is an economic necessity	Strongly Disagree	Count	27	5	32
		% within SCG	5.2%	1.8%	4.0%
	2	Count	22	12	34
		% within SCG	4.2%	4.4%	4.3%
	3	Count	27	14	41
		% within SCG	5.2%	5.1%	5.2%
	4	Count	35	17	52
		% within SCG	6.7%	6.2%	6.6%
	5	Count	67	27	94
		% within SCG	12.9%	9.9%	11.9%
	6	Count	54	38	92
		% within SCG	10.4%	13.9%	11.6%
	7	Count	75	35	110
		% within SCG	14.4%	12.8%	13.9%
	8	Count	65	45	110
		% within SCG	12.5%	16.5%	13.9%
	9	Count	53	31	84
		% within SCG	10.2%	11.4%	10.6%
	Strongly Agree	Count	92	47	139
		% within SCG	17.7%	17.2%	17.5%
	Don't Know	Count	3	2	5
		% within SCG	.6%	.7%	.6%
Total		Count	520	273	793
		% within SCG	100.0%	100.0%	100.0%

Q8h. There is little I can to do reduce the amount of electricity that I am now using.

			Utility			Total
			PG&E	SCE	SDG&E	
There is little I can to do reduce the amount of electricity that I am now using	Strongly Disagree	Count	44	61	44	149
		% within Utility	18.6%	18.8%	19.0%	18.8%
	2	Count	25	36	30	91
		% within Utility	10.5%	11.1%	13.0%	11.5%
	3	Count	31	47	35	113
		% within Utility	13.1%	14.5%	15.2%	14.2%
	4	Count	28	53	24	105
		% within Utility	11.8%	16.3%	10.4%	13.2%
	5	Count	29	36	18	83
		% within Utility	12.2%	11.1%	7.8%	10.5%
	6	Count	21	23	23	67
		% within Utility	8.9%	7.1%	10.0%	8.4%
	7	Count	21	18	24	63
		% within Utility	8.9%	5.5%	10.4%	7.9%
	8	Count	14	11	12	37
		% within Utility	5.9%	3.4%	5.2%	4.7%
	9	Count	7	16	14	37
		% within Utility	3.0%	4.9%	6.1%	4.7%
	Strongly Agree	Count	15	13	5	33
		% within Utility	6.3%	4.0%	2.2%	4.2%
	Don't Know	Count	2	11	2	15
		% within Utility	.8%	3.4%	.9%	1.9%
Total		Count	237	325	231	793
		% within Utility	100.0%	100.0%	100.0%	100.0%

Q8h. There is little I can to do reduce the amount of electricity that I am now using.

			Utility		Total
			Not SCG	SCG	
There is little I can to do reduce the amount of electricity that I am now using	Strongly Disagree	Count	100	49	149
		% within SCG	19.2%	17.9%	18.8%
	2	Count	57	34	91
		% within SCG	11.0%	12.5%	11.5%
	3	Count	74	39	113
		% within SCG	14.2%	14.3%	14.2%
	4	Count	65	40	105
		% within SCG	12.5%	14.7%	13.2%
	5	Count	55	28	83
		% within SCG	10.6%	10.3%	10.5%
	6	Count	45	22	67
		% within SCG	8.7%	8.1%	8.4%
	7	Count	44	19	63
		% within SCG	8.5%	7.0%	7.9%
	8	Count	27	10	37
		% within SCG	5.2%	3.7%	4.7%
	9	Count	22	15	37
		% within SCG	4.2%	5.5%	4.7%
	Strongly Agree	Count	23	10	33
		% within SCG	4.4%	3.7%	4.2%
	Don't Know	Count	8	7	15
		% within SCG	1.5%	2.6%	1.9%
Total		Count	520	273	793
		% within SCG	100.0%	100.0%	100.0%

Q9. Have you ever seen or heard of ENERGY STAR?

			Utility			Total
			PG&E	SCE	SDG&E	
Q9. Have you ever seen or heard of ENERGY STAR?	Yes	Count	206	255	172	633
		% within Utility	86.9%	78.5%	74.5%	79.8%
	No	Count	26	64	51	141
		% within Utility	11.0%	19.7%	22.1%	17.8%
	Refused	Count	1	0	0	1
		% within Utility	.4%	.0%	.0%	.1%
	Don't know	Count	4	6	8	18
		% within Utility	1.7%	1.8%	3.5%	2.3%
Total		Count	237	325	231	793
		% within Utility	100.0%	100.0%	100.0%	100.0%

Q9. Have you ever seen or heard of ENERGY STAR?

			Utility		Total
			Not SCG	SCG	
Q9. Have you ever seen or heard of ENERGY STAR?	Yes	Count	418	215	633
		% within SCG	80.4%	78.8%	79.8%
	No	Count	88	53	141
		% within SCG	16.9%	19.4%	17.8%
	Refused	Count	1	0	1
		% within SCG	.2%	.0%	.1%
	Don't know	Count	13	5	18
		% within SCG	2.5%	1.8%	2.3%
Total		Count	520	273	793
		% within SCG	100.0%	100.0%	100.0%

Q10. What is your understanding of what Energy Star Means?

			Utility			Total
			PG&E	SCE	SDG&E	
Understandings of Energy Star Meaning	Q10-3. Costs less to operate.	Count	150	183	133	466
		% within Utility	72.8%	71.8%	78.7%	
	Q10-2. Is less harmful to the environment.	Count	83	74	55	212
		% within Utility	40.3%	29.0%	32.5%	
	Q10-5. Meets a government standard for energy efficiency.	Count	125	161	108	394
		% within Utility	60.7%	63.1%	63.9%	
	Q10-4. Comes with a rebate offer.	Count	73	83	57	213
		% within Utility	35.4%	32.5%	33.7%	
	Q10. What is your understanding of what ENERGY STAR means? Saves energy.	Count	189	234	157	580
		% within Utility	91.7%	91.8%	92.9%	
Total		Count	206	255	169	630

Percentages and totals are based on respondents.
a Dichotomy group tabulated at value 1.

Q10. What is your understanding of what Energy Star Means?

			Utility		Total
			Not SCG	SCG	
Understandings of Energy Star Meaning	Q10-3. Costs less to operate.	Count	309	157	466
		% within SCG	74.3%	73.4%	
	Q10-2. Is less harmful to the environment.	Count	147	65	212
		% within SCG	35.3%	30.4%	
	Q10-5. Meets a government standard for energy efficiency.	Count	259	135	394
		% within SCG	62.3%	63.1%	
	Q10-4. Comes with a rebate offer.	Count	139	74	213
		% within SCG	33.4%	34.6%	
	Q10. What is your understanding of what ENERGY STAR means? Saves energy.	Count	384	196	580
		% within SCG	92.3%	91.6%	
Total		Count	416	214	630

Percentages and totals are based on respondents.
a Dichotomy group tabulated at value 1.

Multiple Response Questions

Q12. Which of the following energy conservation programs do you recall?

			Utility			Total
			PG&E	SCE	SDG&E	
Programs Recalled	Audit thru the mail	Count	10	16	8	34
		% within Utility	9.7%	12.0%	9.5%	
	Audit by phone	Count	4	3	1	8
		% within Utility	3.9%	2.3%	1.2%	
	Audit at time of sale	Count	6	1	0	7
		% within Utility	5.8%	.8%	.0%	
	Audit on site	Count	19	15	11	45
		% within Utility	18.4%	11.3%	13.1%	
	Audit over the internet	Count	6	11	2	19
		% within Utility	5.8%	8.3%	2.4%	
	CARE rate	Count	27	33	9	69
		% within Utility	26.2%	24.8%	10.7%	
	Flex your power programs	Count	71	83	54	208
		% within Utility	68.9%	62.4%	64.3%	
	EE mortgages	Count	7	6	4	17
		% within Utility	6.8%	4.5%	4.8%	
	New construction programs	Count	26	17	11	54
		% within Utility	25.2%	12.8%	13.1%	
	Rebate programs	Count	87	108	69	264
		% within Utility	84.5%	81.2%	82.1%	
	Refrigerator turn-in	Count	57	109	56	222
		% within Utility	55.3%	82.0%	66.7%	
	Home repair-retrofit	Count	46	58	31	135
		% within Utility	44.7%	43.6%	36.9%	
	Product turn-in programs	Count	26	34	22	82
		% within Utility	25.2%	25.6%	26.2%	
Total		Count	103	133	84	320

Percentages and totals are based on respondents.
a Dichotomy group tabulated at value 1.

Q12. Which of the following energy conservation programs do you recall?

			Utility		Total
			Not SCG	SCG	
Programs Recalled	Audit thru the mail	Count	19	15	34
		% within SCG	9.2%	13.3%	
	Audit by phone	Count	4	4	8
		% within SCG	1.9%	3.5%	
	Audit at time of sale	Count	5	2	7
		% within SCG	2.4%	1.8%	
	Audit on site	Count	33	12	45
		% within SCG	15.9%	10.6%	
	Audit over the internet	Count	8	11	19
		% within SCG	3.9%	9.7%	
	CARE rate	Count	41	28	69
		% within SCG	19.8%	24.8%	
	Flex your power programs	Count	138	70	208
		% within SCG	66.7%	61.9%	
	EE mortgages	Count	11	6	17
		% within SCG	5.3%	5.3%	
	New construction programs	Count	39	15	54
		% within SCG	18.8%	13.3%	
	Rebate programs	Count	171	93	264
		% within SCG	82.6%	82.3%	
	Refrigerator turn-in	Count	130	92	222
		% within SCG	62.8%	81.4%	
	Home repair-retrofit	Count	79	56	135
		% within SCG	38.2%	49.6%	
	Product turn-in programs	Count	53	29	82
		% within SCG	25.6%	25.7%	
Total		Count	207	113	320

Percentages and totals are based on respondents.
a Dichotomy group tabulated at value 1.

Q13. In which programs have you participated in the past?

			Utility			Total
			PG&E	SCE	SDG&E	
Programs Participated In	Audit by mail	Count	5	8	3	16
		% within Utility	7.2%	8.4%	6.4%	
	Audit by phone'	Count	1	2	0	3
		% within Utility	1.4%	2.1%	.0%	
	Audit at time of sale	Count	1	3	0	4
		% within Utility	1.4%	3.2%	.0%	
	Audit on site	Count	7	10	5	22
		% within Utility	10.1%	10.5%	10.6%	
	Audit by internet	Count	5	5	0	10
		% within Utility	7.2%	5.3%	.0%	
	CARE rate	Count	9	13	2	24
		% within Utility	13.0%	13.7%	4.3%	
	Flex your power	Count	22	18	8	48
		% within Utility	31.9%	18.9%	17.0%	
	EE mortgages	Count	1	3	1	5
		% within Utility	1.4%	3.2%	2.1%	
	New construction programs	Count	7	5	4	16
		% within Utility	10.1%	5.3%	8.5%	
	Rebate	Count	54	65	35	154
		% within Utility	78.3%	68.4%	74.5%	
	Refrig recycling	Count	19	33	17	69
		% within Utility	27.5%	34.7%	36.2%	
	Home repair-retrofit	Count	12	19	12	43
		% within Utility	17.4%	20.0%	25.5%	
	Turn-in	Count	14	16	8	38
		% within Utility	20.3%	16.8%	17.0%	
Total		Count	69	95	47	211

Percentages and totals are based on respondents.
a Dichotomy group tabulated at value 1.

Q13. In which programs have you participated in the past?

			Utility		Total
			Not SCG	SCG	
Programs Participated In	Audit by mail	Count	7	9	16
		% within SCG	5.5%	10.7%	
	Audit by phone'	Count	1	2	3
		% within SCG	.8%	2.4%	
	Audit at time of sale	Count	1	3	4
		% within SCG	.8%	3.6%	
	Audit on site	Count	16	6	22
		% within SCG	12.6%	7.1%	
	Audit by internet	Count	5	5	10
		% within SCG	3.9%	6.0%	
	CARE rate	Count	13	11	24
		% within SCG	10.2%	13.1%	
	Flex your power	Count	33	15	48
		% within SCG	26.0%	17.9%	
	EE mortgages	Count	2	3	5
		% within SCG	1.6%	3.6%	
	New construction programs	Count	11	5	16
		% within SCG	8.7%	6.0%	
	Rebate	Count	95	59	154
		% within SCG	74.8%	70.2%	
	Refrig recycling	Count	38	31	69
		% within SCG	29.9%	36.9%	
	Home repair-retrofit	Count	26	17	43
		% within SCG	20.5%	20.2%	
	Turn-in	Count	25	13	38
		% within SCG	19.7%	15.5%	
Total		Count	127	84	211

Percentages and totals are based on respondents.
a Dichotomy group tabulated at value 1.

Q14. In the prior question, you indicated that you participated in a rebate program. For which of the following did you receive a rebate in the year you indicated? Please check all that apply.

			Utility			Total
			PG&E	SCE	SDG&E	
Measures Rebates Received For	Clothes washer	Count	19	14	8	41
		% within Utility	29.2%	16.9%	15.1%	
	Central ac	Count	7	17	6	30
		% within Utility	10.8%	20.5%	11.3%	
	Dishwasher	Count	13	8	8	29
		% within Utility	20.0%	9.6%	15.1%	
	water heater	Count	8	8	13	29
		% within Utility	12.3%	9.6%	24.5%	
	Evap cooler	Count	0	1	1	2
		% within Utility	.0%	1.2%	1.9%	
	Gas furnace	Count	3	4	3	10
		% within Utility	4.6%	4.8%	5.7%	
	Windows	Count	9	3	10	22
		% within Utility	13.8%	3.6%	18.9%	
	Whole house fan	Count	5	6	1	12
		% within Utility	7.7%	7.2%	1.9%	
	Heat pump	Count	3	1	0	4
		% within Utility	4.6%	1.2%	.0%	
	Insulation	Count	9	7	1	17
		% within Utility	13.8%	8.4%	1.9%	
	Oven	Count	3	5	3	11
		% within Utility	4.6%	6.0%	5.7%	
	Swimming pool items	Count	1	6	0	7
		% within Utility	1.5%	7.2%	.0%	
	Prog thermostats	Count	9	9	5	23
		% within Utility	13.8%	10.8%	9.4%	
	Refrigerator	Count	28	35	19	82
		% within Utility	43.1%	42.2%	35.8%	
	Room air conditioner	Count	4	2	1	7
		% within Utility	6.2%	2.4%	1.9%	
	Shower heads	Count	6	3	6	15
		% within Utility	9.2%	3.6%	11.3%	
	Wall insulation	Count	5	2	0	7
		% within Utility	7.7%	2.4%	.0%	
	Water heater	Count	5	9	7	21
		% within Utility	7.7%	10.8%	13.2%	
	Water heater pipe insul	Count	4	1	1	6
		% within Utility	6.2%	1.2%	1.9%	
Total		Count	65	83	53	201

Percentages and totals are based on respondents.
a Dichotomy group tabulated at value 1.

Q14. In the prior question, you indicated that you participated in a rebate program. For which of the following did you receive a rebate in the year you indicated? Please check all that apply.

			Utility		Total
			Not SCG	SCG	
Measures Rebates Received For	Clothes washer	Count	30	11	41
		% within SCG	23.3%	15.3%	
	Central ac	Count	15	15	30
		% within SCG	11.6%	20.8%	
	Dishwasher	Count	22	7	29
		% within SCG	17.1%	9.7%	
	water heater	Count	23	6	29
		% within SCG	17.8%	8.3%	
	Evap cooler	Count	2	0	2
		% within SCG	1.6%	.0%	
	Gas furnace	Count	6	4	10
		% within SCG	4.7%	5.6%	
	Windows	Count	19	3	22
		% within SCG	14.7%	4.2%	
	Whole house fan	Count	8	4	12
		% within SCG	6.2%	5.6%	
	Heat pump	Count	3	1	4
		% within SCG	2.3%	1.4%	
	Insulation	Count	10	7	17
		% within SCG	7.8%	9.7%	
	Oven	Count	6	5	11
		% within SCG	4.7%	6.9%	
	Swimming pool items	Count	1	6	7
		% within SCG	.8%	8.3%	
	Prog thermostats	Count	16	7	23
		% within SCG	12.4%	9.7%	
	Refrigerator	Count	52	30	82
		% within SCG	40.3%	41.7%	
	Room air conditioner	Count	5	2	7
		% within SCG	3.9%	2.8%	
	Shower heads	Count	12	3	15
		% within SCG	9.3%	4.2%	
	Wall insulation	Count	6	1	7
		% within SCG	4.7%	1.4%	
	Water heater	Count	11	10	21
		% within SCG	8.5%	13.9%	
	Water heater pipe insul	Count	5	1	6
		% within SCG	3.9%	1.4%	
Total		Count	129	72	201

Percentages and totals are based on respondents.
a Dichotomy group tabulated at value 1.

Interval Level Variables for PG&E

Descriptive Statistics(a)

	N	Minimum	Maximum	Mean	Std. Deviation
Q24. In what year were you born?	230	1867	1985	1962.80	13.846
Year home built with approximations inserted for missing values	220	1888	2005	1973.73	24.659
Annualized kWh	106	12	67579	8142.05	11353.762
Annualized therms	86	12	7320	548.70	1119.363
Valid N (listwise)	74				

a Utility = PG&E

Interval Level Variables for SCE

Descriptive Statistics(a)

	N	Minimum	Maximum	Mean	Std. Deviation
Q24. In what year were you born?	312	1931	1990	1963.94	11.626
Year home built with approximations inserted for missing values	304	1910	2005	1972.50	20.146
Annualized kWh	150	12	49836	8074.65	7566.417
Annualized therms	129	12	12024	422.51	1211.339
Valid N (listwise)	113				

a Utility = SCE

Interval Level Variables for SDG&E

Descriptive Statistics(a)

	N	Minimum	Maximum	Mean	Std. Deviation
Q24. In what year were you born?	225	1832	1985	1960.50	16.988
Year home built with approximations inserted for missing values	225	1906	2005	1977.08	17.844
Annualized kWh	114	60	60000	6183.95	7029.237
Annualized therms	94	3	6000	291.79	645.750
Valid N (listwise)	88				

a Utility = SDG&E

Interval Level Variables for SCG

Descriptive Statistics(a)

	N	Minimum	Maximum	Mean	Std. Deviation
Q24. In what year were you born?	261	1900	1990	1962.28	12.209
Year home built with approximations inserted for missing values	257	1910	2005	1972.51	20.724
Annualized kWh	127	12	49836	8041.39	7815.815
Annualized therms	118	12	12024	381.56	1123.902
Valid N (listwise)	100				

a SCG = SCG

Appendix D

e-Rewards Panel Quality

1. **e-Rewards built these panels from the ground up**
   e-Rewards has never acquired other panels that have been sourced with enrollment methodologies that diverge from (and that are sub-standard to) our highly structured enrollment process. e-Rewards does not "re-market" panels that are owned by others. e-Rewards only markets its own panels which represent fully owned and controlled assets.

   The panel counts that e-Rewards markets are accurate and reflective of the actual number of panel members enrolled into the e-Rewards panels. Counts are not inflated to include other panels that e-Rewards may have "access to" but that are not owned and maintained by e-Rewards. Counts are not inflated to include other *household members* who have not been enrolled and profiled into the panel.

2. **e-Rewards has adhered to a by-invitation-only panel enrollment approach which diminishes "self selection bias"**
   e-Rewards believes that panels that use *open* recruitment techniques and enrollment processes such as banner ads and co-registration sites are allowing large-scale self-selection biases into their panels and attracting "professional survey takers."

3. **e-Rewards uses both online and offline recruitment methods**
   In order to remain as balanced as possible, e-Rewards panel members have been invited through a mix of both online methods (e.g. solo e-mail invitations and other targeted online modes) and offline methods (e.g. physical post-card invitations, direct mail inserts, etc.). All panel establishment methodologies employed by e-Rewards fully comply with CASRO guidelines, of which e-Rewards is a member organization.

4. **e-Rewards' stringently weeds out "professional survey takers"**
   As mentioned above, the best way to eliminate professional survey takers is to not attract them in the first place. However, e-Rewards uses five additional methods to identify professional survey taking behavior within its panels and prevents professional survey takers from receiving future survey opportunities through the e-Rewards channel.
   - Obviously inconsistent profiling answers
   - Straight-lining answers
   - Answering surveys in too short a time frame
   - Client-reported complaints or observations
   - Blind pre-screening of study candidates

5. **e-Rewards panelist information is verified**
   e-Rewards collects a physical address on members who enrolls in the e-Rewards member panel. The physical addresses collected are verified against U.S. Postal information to verify address validity. e-Rewards members must enroll using a valid and unique e-mail address in order to receive market research survey opportunities. e-Rewards will not send e-mails to e-mail addresses that have *bounced back* or are no longer active.

6. **e-Rewards panelists are the most deeply profiled**
   During the panel enrollment process e-Rewards captures the most detailed possible demographic, psychographic, and geographic profile information from each panelist (over 300 segmentation variables). Over 90 percent of the questions asked during the enrollment process are "select only one answer" type questions which forces a respondent to choose the "best" answer that describes himself/herself.

7. **e-Rewards panelist activity is tracked for quality purposes**
   e-Rewards tracks the activity level of each panelist to ensure that inactive panelists are periodically phased out of receiving survey opportunities. Conversely, e-Rewards is able to monitor and prevent individual panelists from participating in an unhealthy amount of research studies. e-Rewards tracks each member's historical level of survey participation by topic area, and is able to prevent members from receiving invitations to participate in similar surveys until a proper amount of time has passed.

8. **e-Rewards' panels are the most normalized and representative**
   e-Rewards scientifically manages the demographic make-up of its panels using pre-recruitment targeting and a "by invitation only recruitment" method. e-Rewards is able to do this by working with its diverse set of sourcing partners up front to invite only the types of individuals that fit the current normalization needs of e-Rewards B2C or B2B panels. For example, the e-Rewards panel has been purposely constructed to be 56% male and 44% female. Accounting for the known female gender response bias, this allows e-Rewards to consistently achieve very close to a 50/50 gender split when it samples it panel naturally (e.g. without weighting or special targeting).

9. **e-Rewards panel profile information is fresher**
   e-Rewards utilizes its program participation rules and its proprietary Dynamic Profile Enrichment^{SM} capabilities to achieve industry leading levels of panel maintenance and data freshness. e-Rewards Dynamic Profile Enrichment^{SM} capability is an intelligent database algorithm that periodically presents panelists with opportunities to update certain aspects of their profile. Profiling questions are presented to panelists based a ranking that comprehends the amount of time since a profile question was last updated and the likelihood that the question's answers will have changed. This method is state-of-the-art in terms of panel profile refreshment and maintenance. e-Rewards uses Dynamic Profile Enrichment^{SM} to maintain the freshness of panel enrollment questions, as well as build deeper understanding of each segment of the e-Rewards' panels by introducing customized questions for future targeting.

10. **e-Rewards achieves industry leading member retention rates**
    e-Rewards has retained over 80% of its members since inception in 1999. This retention rate underscores e-Rewards commitment to its members in terms of quality communications, program design, and customer service.

11. e-Rewards achieves industry leading response rates

e-Rewards achieves average survey response rates of 15-25 percent for consumer surveys and 25-35 percent for B2B studies. When additional targeting or prescreening data is used for targeting, the response rates achieved typical increase from those cited above.

12. e-Rewards members are protected to provide honest answers

e-Rewards respects the privacy of each of its panel members and adheres to a highly ethical privacy policy that fosters maximum trust with panel members. Panelist trust equates to more honest responses to survey questions—even the most sensitive ones. e-Rewards research studies will report information only in aggregate, singular summary form, and will never reveal personally identifiable information unless it is expressly provided by the respondent themselves.

13. **e-Rewards' panelists are not over surveyed**

The average e-Rewards member qualifies and participates in less than 3 full surveys each year. Other sample vendors allow annual survey participation rates to exceed 12-24 surveys each year. e-Rewards asks each panel member about the maximum number of e-mails that they would like to receive from e-Rewards on a weekly basis. e-Rewards has system controls in place to enforce that each members' preferences are adhered to.

14. e-Rewards supports double-blind screening

e-Rewards advocates using a "blind screener" approach (e.g. a 2-step approach) to identify qualified respondents before they are invited to take a full survey instrument. That way every respondent receives the same amount of incentive during the screening process, and there are no detectable advantages (or disadvantages) to the respondent for answering one way or the other. In other words, there is no known incentive for providing the "correct" answers. e-Rewards suggests that respondents who are invited to take a full survey instrument are re-screened to ensure consistency in their answers toward qualifying for the study.

15. e-Rewards offers an established incentive currency

e-Rewards panel members earn e-Rewards currency (U.S.-dollar denominated) for the time they spend answering market research surveys and reacting to commercial e-mails. Members use their e-Rewards dollars to redeem valuable rewards which are provided by e-Rewards Sponsors and Program Partners.

Appendix E

Data Documentation

Stated Preference Analysis File

File Information

Analysis file of all survey respondents

Source		dc_final
Creation Date		19-FEB-2006 23:50:04
Label		Aggregated File
File Contents	Data Type	Case
Data Information	N of Cases	793
	N of Defined Variable Elements	241
	N of Named Variables	238
	Weight Variable	None
	Compressed	Yes

Revealed Preference Analysis Files

File Information

Analysis file of all PG&E customers receiving mailers

Source		pge_anal
Creation Date		26-FEB-2006 15:29:58
Label		None
File Contents	Data Type	Case
Data Information	N of Cases	225226
	N of Defined Variable Elements	93
	N of Named Variables	85
	Weight Variable	None
	Compressed	Yes

File Information

Analysis file of a random sample of 30,000 from pge_anal

Source		pge_anal_30k
Creation Date		23-FEB-2006 23:01:42
Label		None
File Contents	Data Type	Case
Data Information	N of Cases	30000
	N of Defined Variable Elements	91
	N of Named Variables	83
	Weight Variable	None
	Compressed	Yes

File Information

Analysis file of all SCE customers receiving mailers

Source		sce_anal
Creation Date		22-FEB-2006 17:11:30
Label		None
File Contents	Data Type	Case
Data Information	N of Cases	200000
	N of Defined Variable Elements	119
	N of Named Variables	113
	Weight Variable	None
	Compressed	Yes

File Information

Analysis file of a random sample of 30,000 from sce_anal

Source		scecen_ax_part_30k
Creation Date		22-FEB-2006 17:06:05
Label		None
File Contents	Data Type	Case
Data Information	N of Cases	30000
	N of Defined Variable Elements	115
	N of Named Variables	109
	Weight Variable	None
	Compressed	Yes

File Information

Analysis file of all SCG customers receiving mailers

Source		scg_anal
Creation Date		09-FEB-2006 10:37:50
Label		None
File Contents	Data Type	Case
Data Information	N of Cases	50000
	N of Defined Variable Elements	105
	N of Named Variables	93
	Weight Variable	None
	Compressed	Yes

File Information

Analysis file of a random sample of 30,000 from scg_anal

Source		scg_anal_30k
Creation Date		09-FEB-2006 10:39:22
Label		None
File Contents	Data Type	Case
Data Information	N of Cases	30000
	N of Defined Variable Elements	105
	N of Named Variables	93
	Weight Variable	None
	Compressed	Yes

File Information

Analysis file of all SDG&E customers receiving mailers

Source		sdge_anal
Creation Date		09-FEB-2006 10:44:52
Label		None
File Contents	Data Type	Case
Data Information	N of Cases	35002
	N of Defined Variable Elements	108
	N of Named Variables	96
	Weight Variable	None
	Compressed	Yes

Appendix F

Selected CART Trees for the Revealed Preference Models

Figure F-1
CART Tree Predicting Participation: PG&E

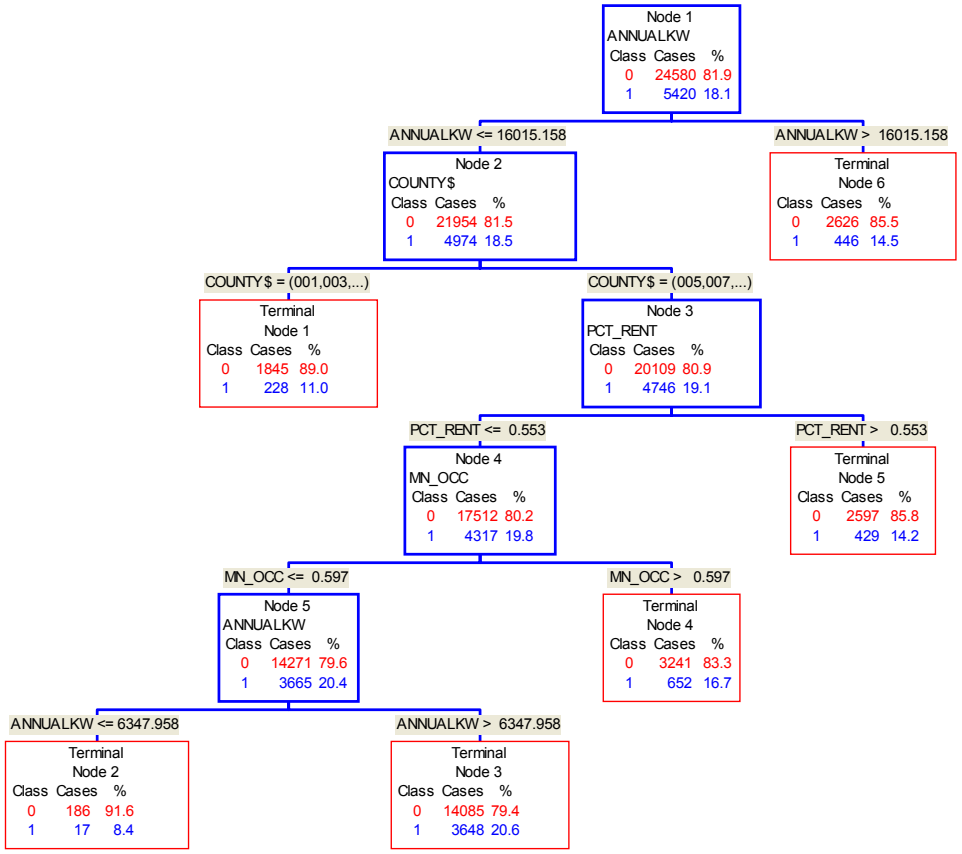


Figure F-2
CART Tree Predicting Participation: SCE

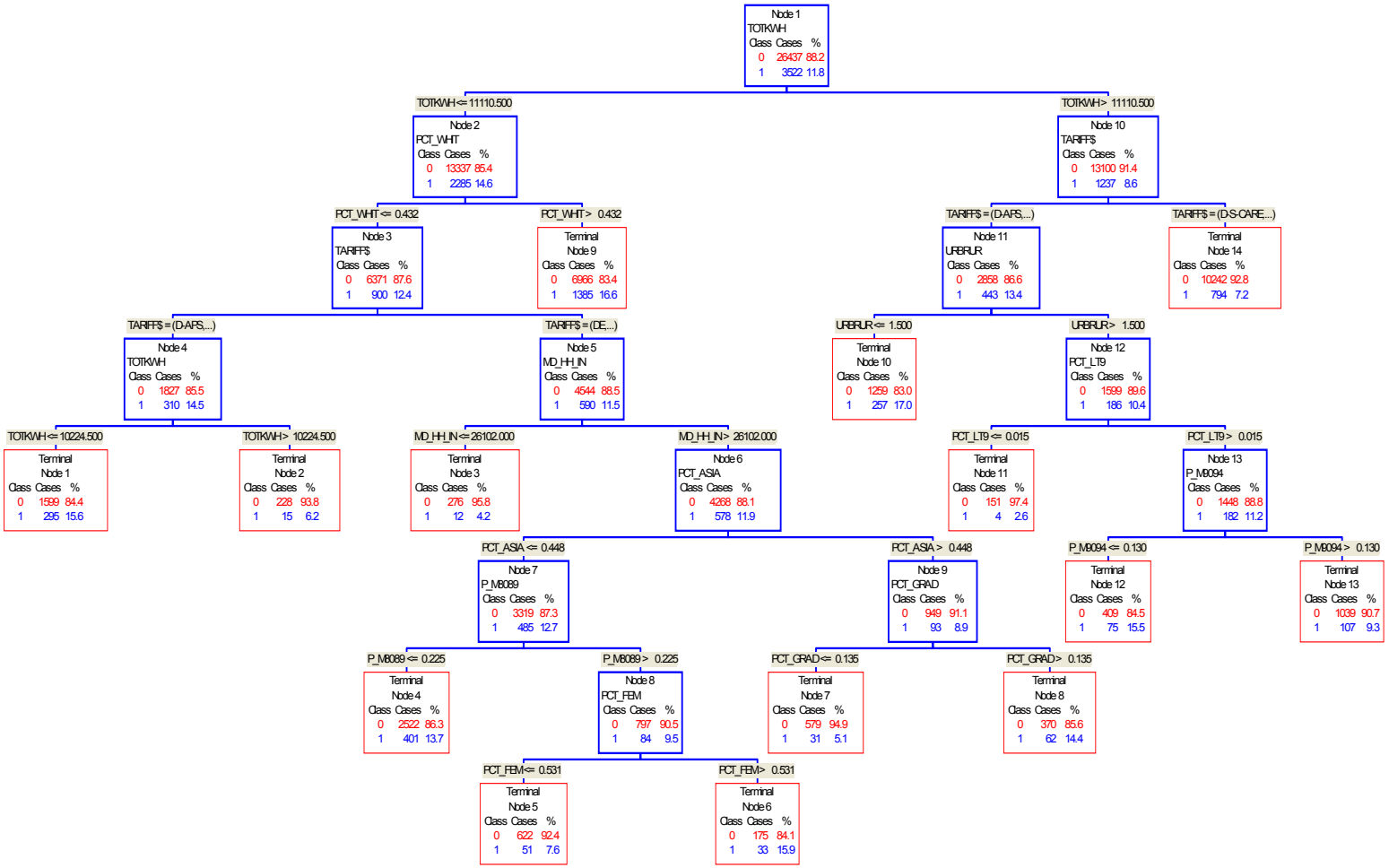


Figure F-3
CART Tree Predicting Participation: SCG

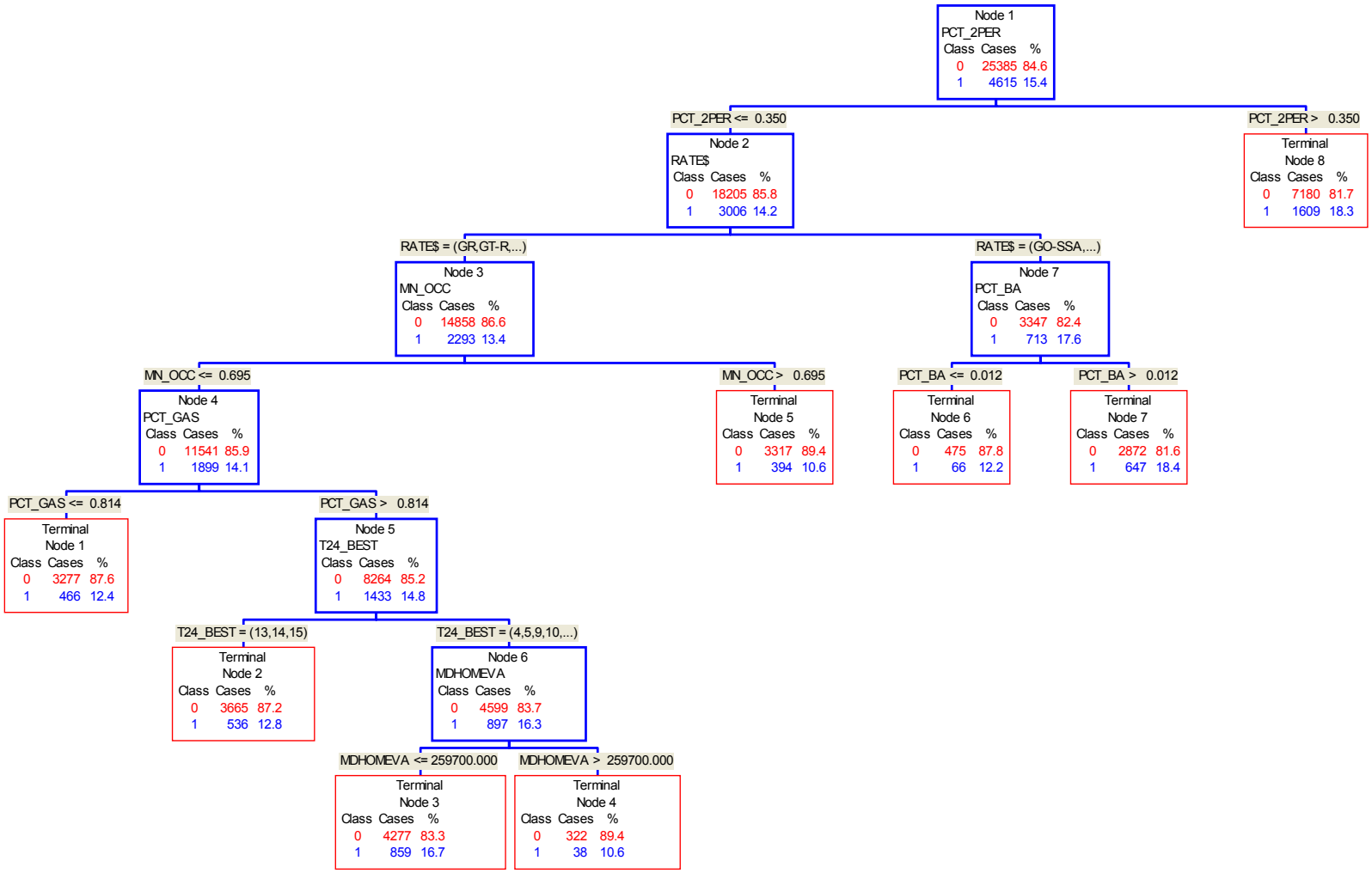
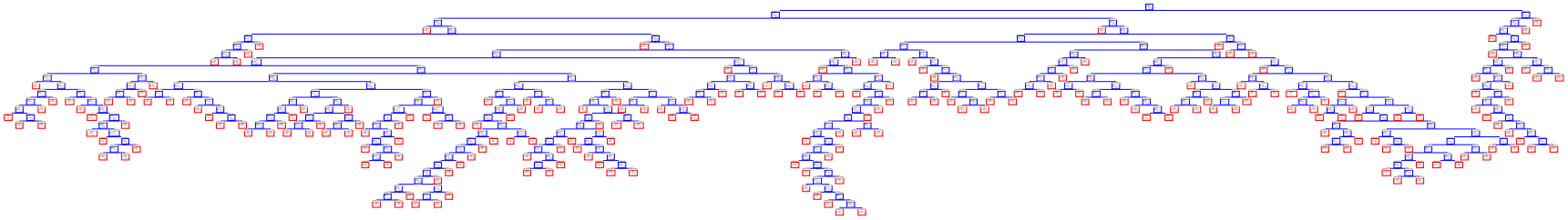


Figure F-4
CART Tree Predicting Participation: SDG&E



Note: See PDF version of this report (available at www.calmac.org) for a more easily readable tree.

Table F-1
Variable Meanings for Variables Acting as Splitters

CART Variable Name	Definition
AGE_NUM	Age:Acxiom
ANNUALKW	Annual KWh
ANNUALKWH	Annual KWh
AREA	Census Block Area
ASIANALO	HTR Zip: Asian
CARE$	CARE Rate
CAREFLAG	CARE Rate
COUNTY$	County
FORECAST	CEC Forecast Climate Zone
HISPLATI	HTR Zip: Latino
HSEHOLD	Household Size: Acxiom
KWH	KWh
KWH_CAT	KWh Category
LANGUAGE	Language: Acxiom
MARITAL$	Marital Status: Acxiom
MD_FAM_I	Median Family Inc
MD_HH_IN	Median Household Income
MDHOMEVA	Median Home Value
MEAN_HH	Mean Household Size
MN_FAMSZ	Mean Family Size
MN_OCC	Mean Occupancy

MN_ROOMS	Mean Rooms in Households
MODINC	HTR Zip: Moderate Income
OCCUPANT$	Occupation: Acxiom
P_M70-79	% moved to home betw 1970-79
P_M80-89	% moved to home 1980-89
P_M90-94	% moved to home betw 1990-94
P_M95-98	% moved to home betw 1995-98
P_M9900	% moved to home betw 1999-2000
P_MB4_69	% moved before 1969
PCT_17	% Population Under 18
PCT_1950	% Homes Built Before 1950
PCT_1960	% Homes Built Before 1960
PCT_1970	% Homes Built Before 1970
PCT_1980	% Homes Built Before 1980
PCT_1990	% Homes Built Before 1990
PCT_1PER	% 1-Person Households
PCT_2PER	% 2-Person Households
PCT_35PE	% Households with 3-5 Persons
Pct_5pl	% Apts 5 or more units
PCT_65	% Population 65 or Over
PCT_6PER	% 6+ Person Households
Pct_912	% completing some HS
PCT_AA	% African American
PCT_ASIA	% Asian
Pct_BA	% completing BA

PCT_COUP	% Couples
PCT_ELEC	% Heated with Electricity
PCT_F_ON	% Female Only Households
PCT_FAM	% Family Households
PCT_FEM	% Females in Population
PCT_GAS	% Homes Heated with Gas
Pct_Grad	% completing grad degree
PCT_HISP	% Latino
Pct_HS	% graduating HS
Pct_lt5	% Apt less than 5 units
Pct_LT9	% less than 9th grade Ed
PCT_M_ON	% Male Only Households
Pct_MH	% Mobile homes
Pct_NoBA	% some college
PCT_NONF	% Non-Family Households
PCT_NONW	% Non-white
PCT_NWHI	% Non-white
Pct_Ostr	% Other dwelling types
PCT_PO_A	% Families in Poverty
PCT_POV	% Population in Poverty
PCT_RENT	% Renters
Pct_SFA	% Single Family Attached
Pct_SFD	% Single Family Detached
PCT_WHIT	% White
POP2000	Population in Census Block

RATE$	Utility Rate Class
RELIGION$	Religion: Acxiom
RENTER	HTR Zip: Renter
ROLLUP_C$	Ethnicity: Acxiom
T24_BEST	Title 24 Climate Zone
TARIFF$	Tariff Class
THERMS	Annual Therms
TOTFAM	# Families in Census Block
TOTKWH	Total KWh
TOTPOP	Population of Census Block
URBRUR	HTR Zip: Rural

>

Appendix G

Selected CART Trees for the Stated Preference Models

Figure G-1
CART Tree of Customer Characteristics in Class 1: Demanding but Willing to Pay

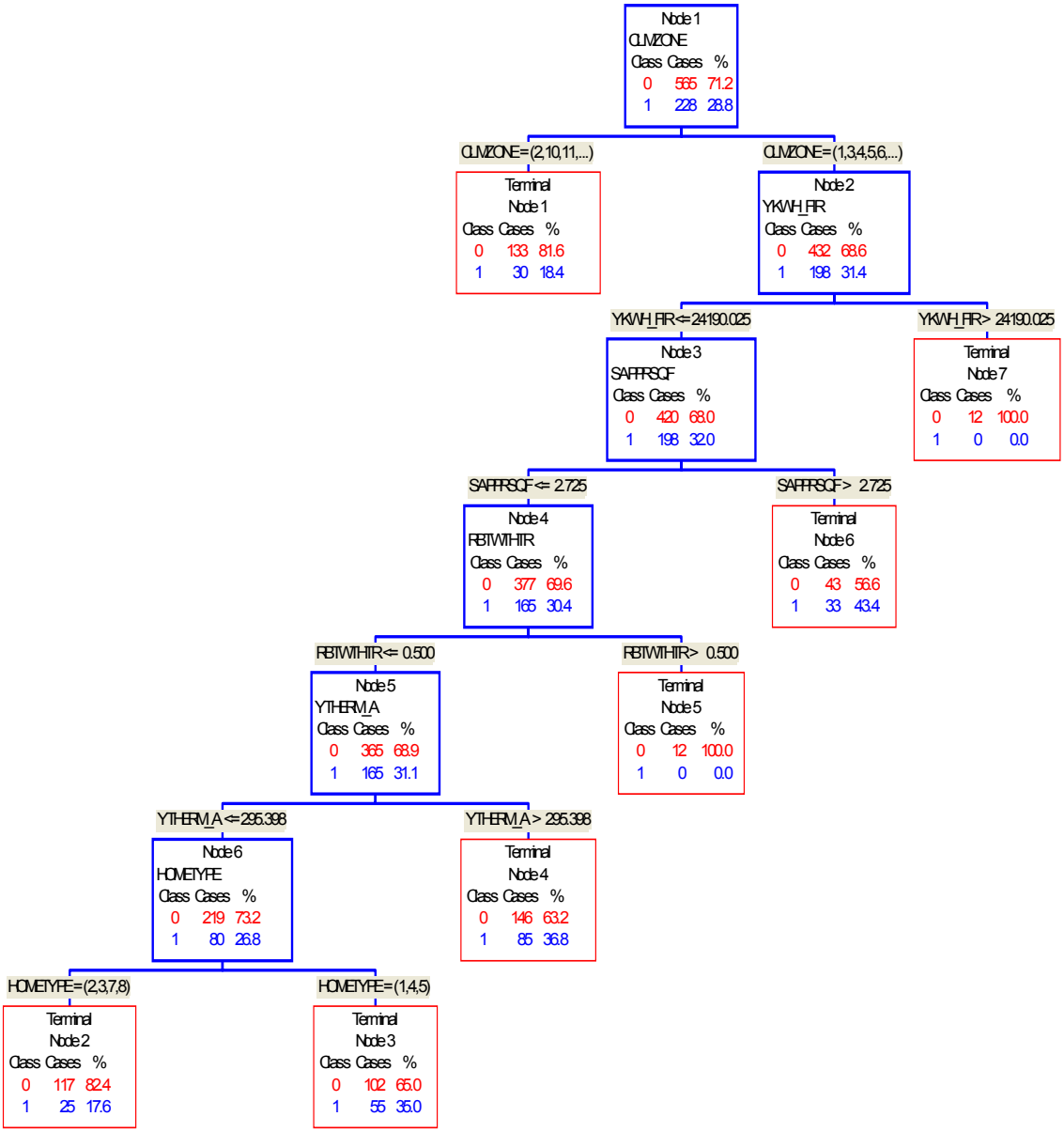


Figure G-2
CART Tree of Customer Characteristics in Class 2: Meticulous

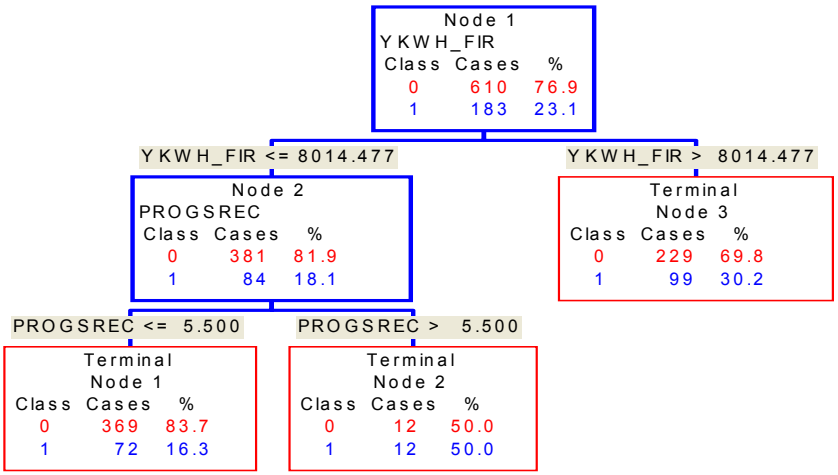


Figure G-3
CART Tree of Customer Characteristics in Class 3: Subsidy Required

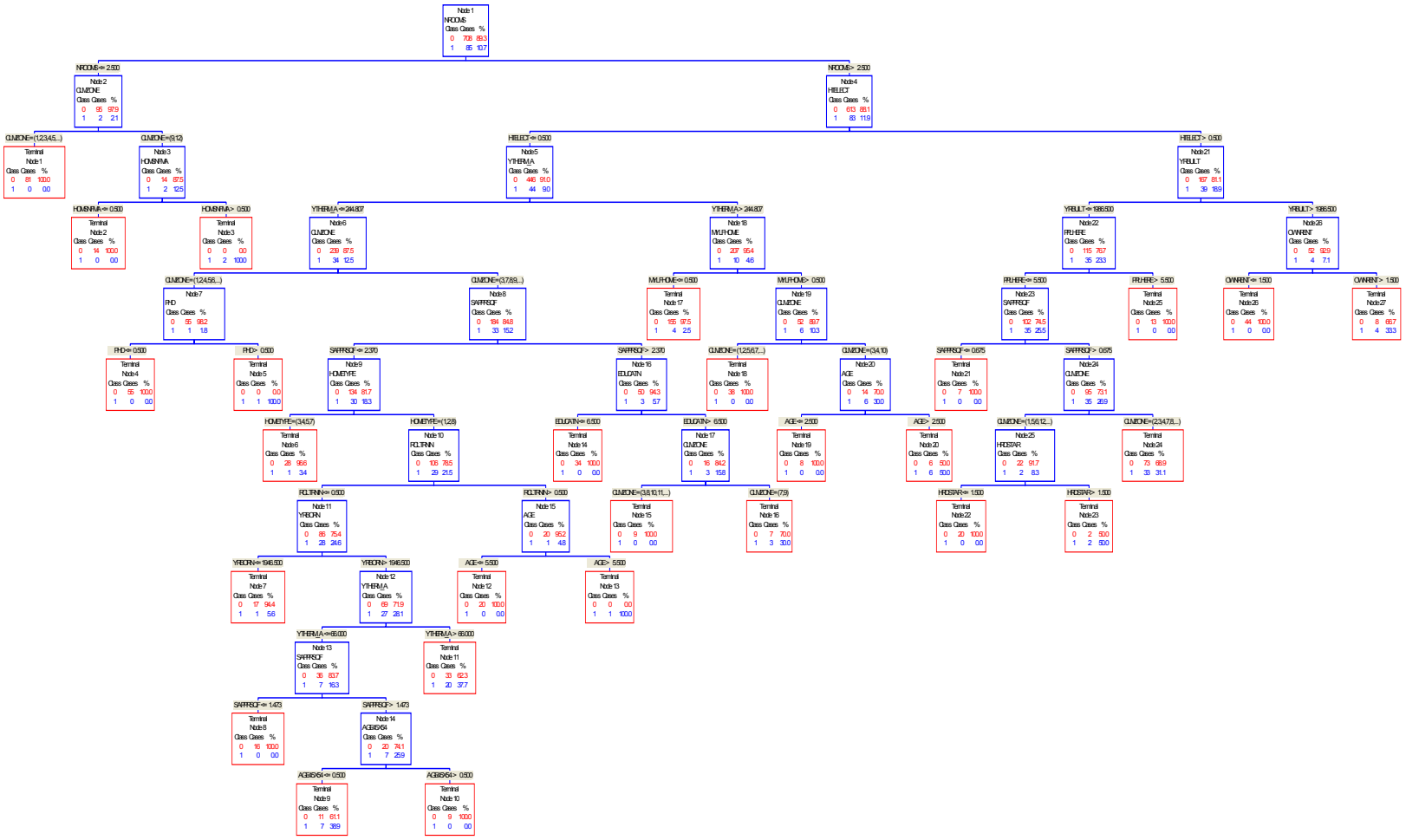


Figure G-4
CART Tree of Customer Characteristics in Class 4: Enthusiasts

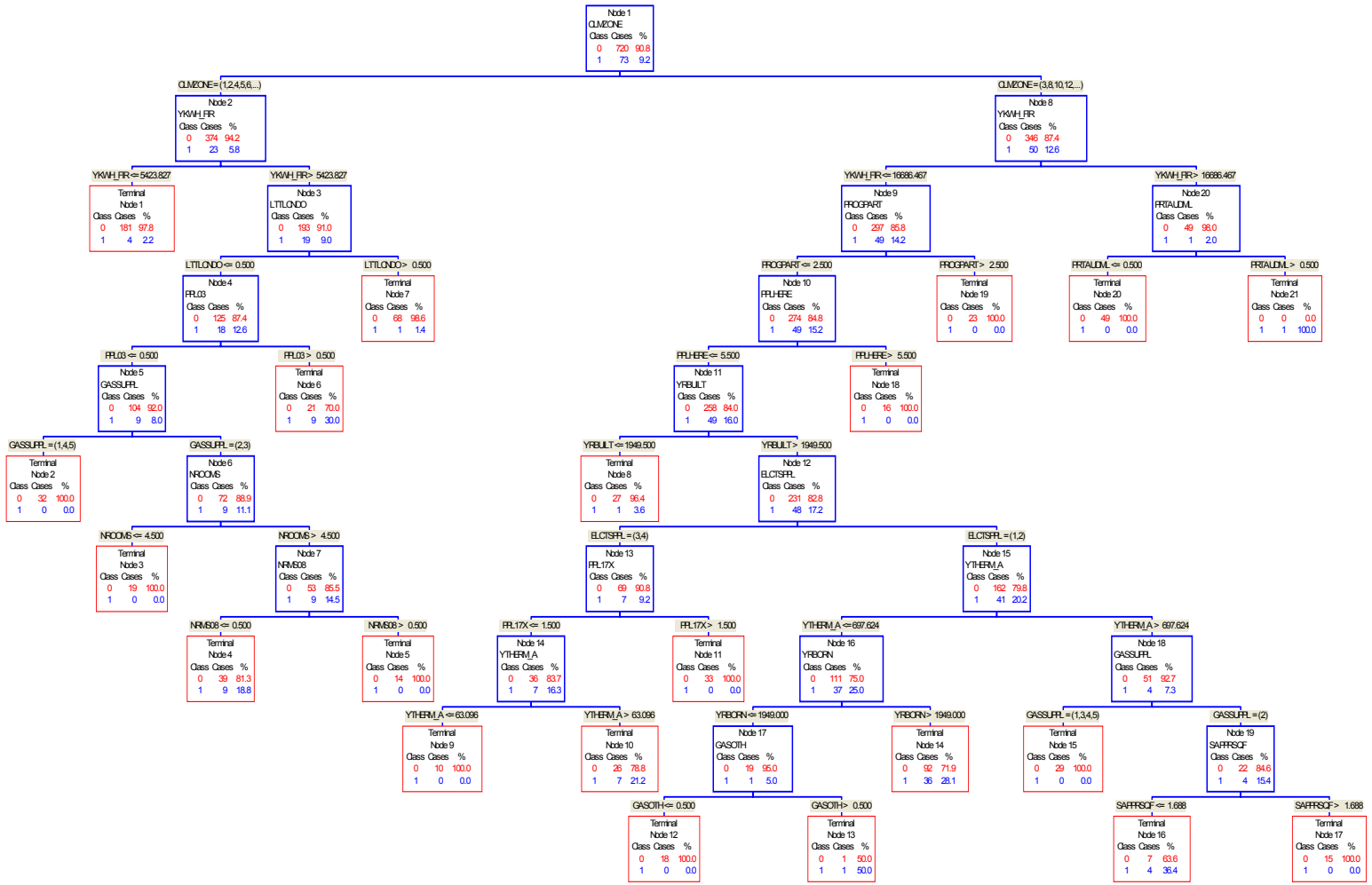


Figure G-5
CART Tree of Customer Characteristics in Class 5: Not Interested

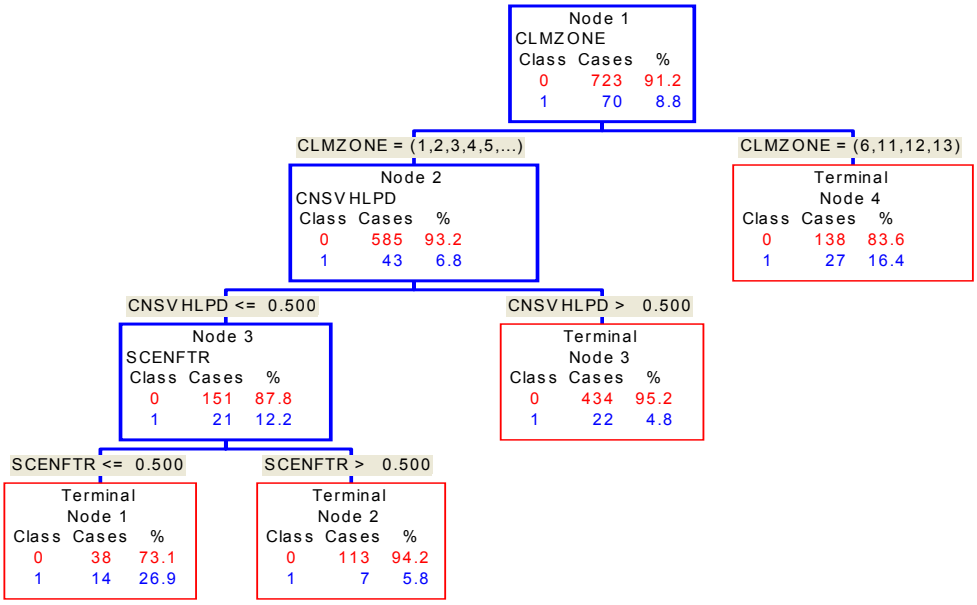


Figure G-6
CART Tree of Customer Characteristics in Class 6: Personal Attention

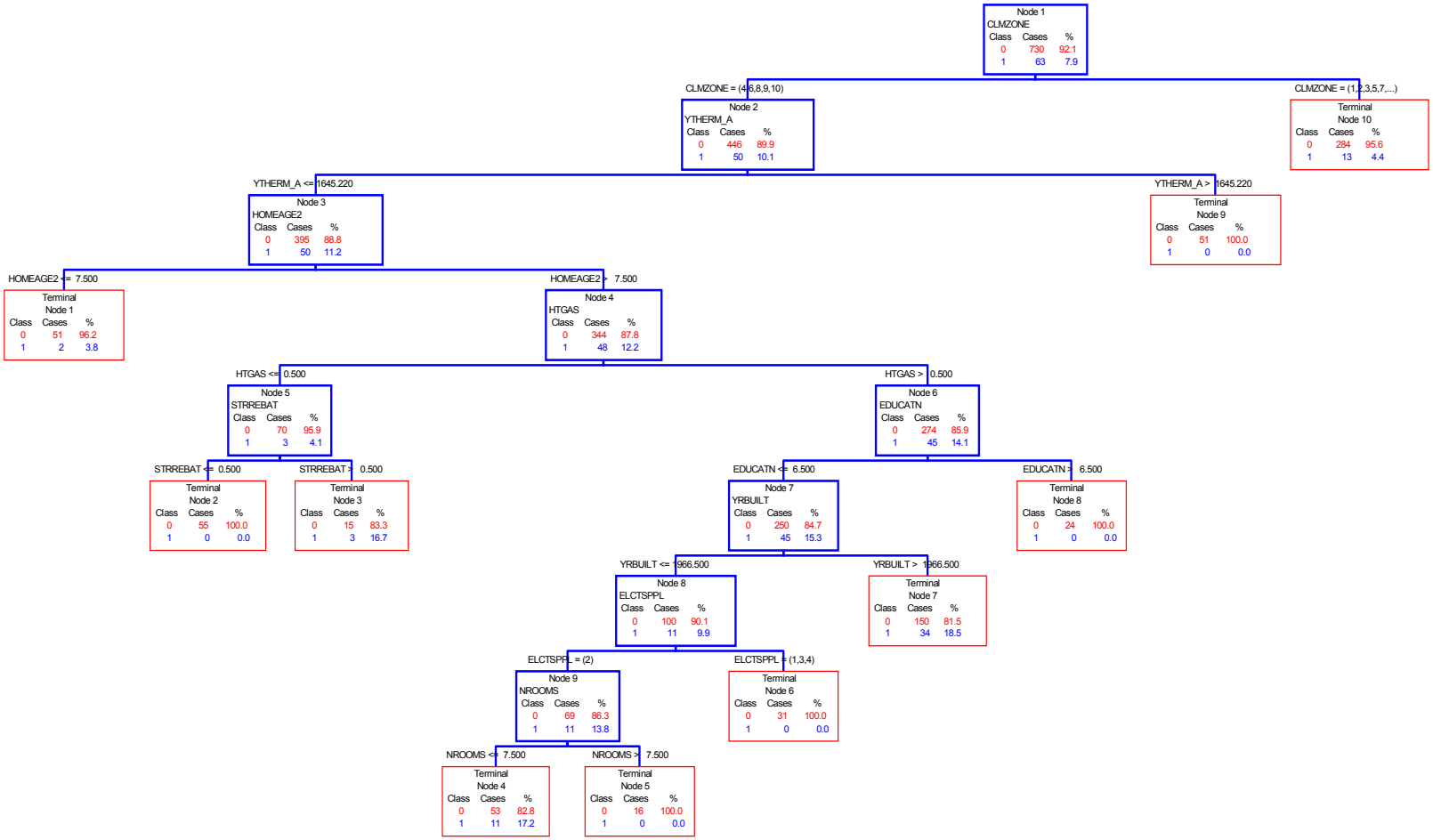


Figure G-7
CART Tree of Customer Characteristics in Class 7: Hard to Convince

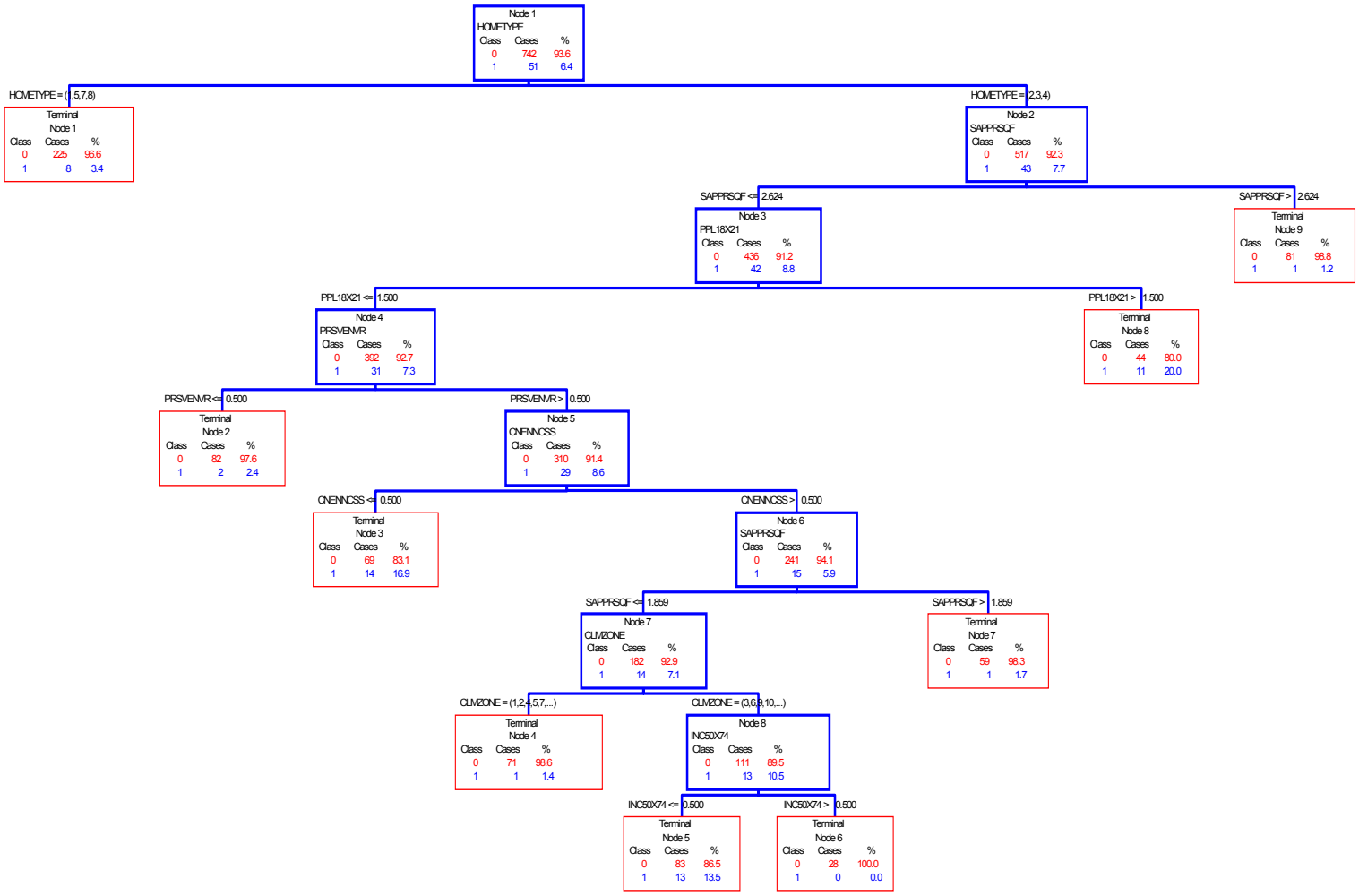
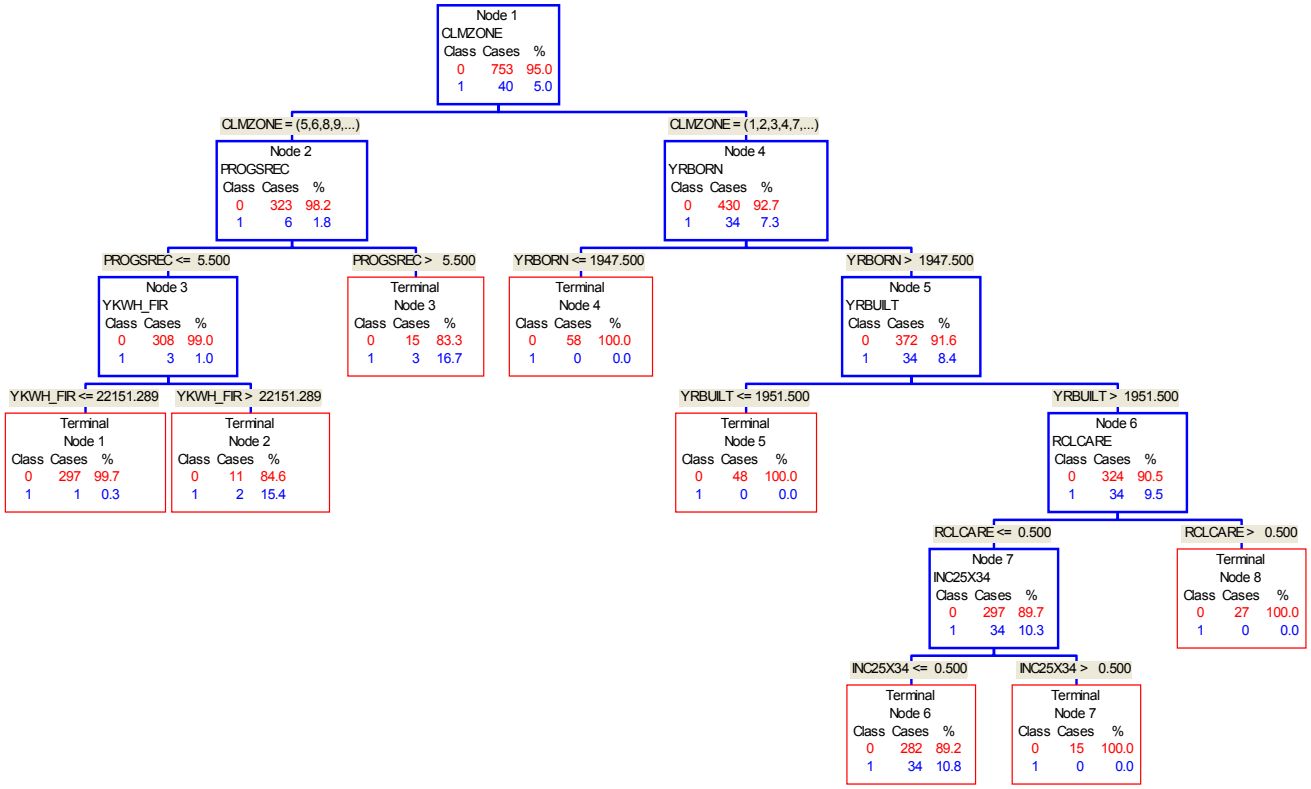


Figure G-8
CART Tree of Customer Characteristics in Class 8: Fast and Thorough



APPENDIX H

Illustrative SAS Code for Simplest CART Tree

```
libname CART 'INSERT FULL DIRECTORY PATH';

/*************************************************************
 * The following SAS-compatible code was automatically generated
 * by the TRANSLATE feature in the Salford Systems CART(tm)
 * program, version: 5.0.9.156
 *************************************************************/

/* Data Dictionary, Number Of Variables = 29 */
/*    Name = RATE, Type = categorical. */
/*    Name = CARE */
/*           Display Name = Indicates who is on the CARE rate */
/*           Type = categorical. */
/*    Name = T24_BEST, Type = categorical. */
/*    Name = PCT_65 */
/*           Display Name = Pct_65+ */
/*           Type = continuous. */
/*    Name = PCT_FEM, Type = continuous. */
/*    Name = PCT_WHIT, Type = continuous. */
/*    Name = PCT_ASIA, Type = continuous. */
/*    Name = PCT_NONW, Type = continuous. */
/*    Name = PCT_2PER, Type = continuous. */
/*    Name = PCT_35PE */
/*           Display Name = Pct_3-5Pers */
/*           Type = continuous. */
/*    Name = MN_FAMSZ, Type = continuous. */
/*    Name = MD_HH_IN, Type = continuous. */
/*    Name = MD_FAM_I, Type = continuous. */
/*    Name = PCT_PO_A, Type = continuous. */
/*    Name = MDHOMEVA, Type = continuous. */
/*    Name = PCT_GAS, Type = continuous. */
/*    Name = PCT_ELEC, Type = continuous. */
/*    Name = MN_OCC, Type = continuous. */
/*    Name = TOT_02 */
/*           Display Name = Annual Therms Consumed */
/*           Type = continuous. */
/*    Name = PCT_HISP, Type = continuous. */
/*    Name = MODINC, Type = continuous. */
/*    Name = HISPLATI */
/*           Display Name = HISPLATINO */
/*           Type = continuous. */
/*    Name = PCT_OSTR */
/*           Display Name = Percent Other structure */
/*           Type = continuous. */
/*    Name = PCT_LT9 */
/*           Display Name = Percent less than 9th grade */
/*           Type = continuous. */
/*    Name = PCT_NOBA */
/*           Display Name = Percent some college */
/*           Type = continuous. */
/*    Name = PCT_BA */
/*           Display Name = Percent completing BA */
/*           Type = continuous. */
/*    Name = PCT_GRAD */
/*           Display Name = Percent completing graduate degree */
```

```
/*           Type = continuous. */
/*    Name = P_M9900 */
/*           Display Name = Percent moved in 1999-2000 */
/*           Type = continuous. */
/*    Name = P_MB4_69 */
/*           Display Name = Percent moved in before 1969 */
/*           Type = continuous. */

data wk1;set CART.(INSERT FILE NAME);
rename PCT_2PERS=PCT_2PER;
rename PCT_35PERS=PCT_35PE;
rename PCT_WHITE= PCT_WHIT;
rename MD_FAM_INC= MD_FAM_I;
rename PCT_GAS_HT= PCT_GAS;
rename PCT_ELEC_HT= PCT_ELEC;
rename PCT_ASIAN= PCT_ASIA;
rename MD_HH_INC= MD_HH_IN;
rename MDHOMEVAL= MDHOMEVA;
rename PCT_POV_FAM=PCT_PO_A;
RUN;

dATA WK2;SET WK1;

MODELBEGIN:

/* CART version: 5.0.9.156 */
/* Tree: Tree_1 */
/* Timestamp: CART20060210162032000 */
/* Grove: C:\DOCUME~1\KATHER~1\LOCALS~1\Temp\s2vo57 */
/* CART Optimal tree, Complexity threshold = 0.00198795 */
/* Target variable: MI, integer discrete with 2 levels. */
/* N terminal nodes = 8, Depth = 7 */

%let target = predicted_response;
%let node = node;
%let prob = prob;

/* Correspondence between probabilities and     */
/* target class levels.  Probabilities are      */
/* based on weighted learn sample class counts. */
/*    &prob.1: 0 */
/*    &prob.2: 1 */

NODE1:
  if PCT_2PER gt .z then do;
    if PCT_2PER <= 0.350355 then goto NODE2;
    else goto TNODE8;
    end;
  else if MN_OCC gt .z then do;
    if MN_OCC <= 0.481585 then goto TNODE8;
    else goto NODE2;
    end;
  else if MN_FAMSZ gt .z then do;
    if MN_FAMSZ <= 3.12872 then goto TNODE8;
    else goto NODE2;
    end;
  else if PCT_35PE gt .z then do;
```

```
    if PCT_35PE <= 0.32818 then goto TNODE8;
    else goto NODE2;
    end;
  else if PCT_WHIT gt .z then do;
    if PCT_WHIT <= 0.792028 then goto NODE2;
    else goto TNODE8;
    end;
  else if PCT_NONW gt .z then do;
    if PCT_NONW <= 0.207972 then goto TNODE8;
    else goto NODE2;
    end;
  else goto NODE2;

NODE2:
  if RATE in ("GR","GT-R","GTO-SSA"
             ) then goto NODE3;
  else if RATE in ("GO-SSA","GO-SSB","GRL","GT-RL","GTO-SSB"
                  ) then goto NODE7;
  else if CARE in ("no"
                  ) then goto NODE3;
  else if CARE in ("yes"
                  ) then goto NODE7;
  else if PCT_NOBA gt .z then do;
    if PCT_NOBA <= 0.0692496 then goto NODE7;
    else goto NODE3;
    end;
  else if MD_FAM_I gt .z then do;
    if MD_FAM_I <= 16211.5 then goto NODE7;
    else goto NODE3;
    end;
  else if P_M9900 gt .z then do;
    if P_M9900 <= 0.00809717 then goto NODE7;
    else goto NODE3;
    end;
  else if TOT_02 gt .z then do;
    if TOT_02 <= 99.5 then goto NODE7;
    else goto NODE3;
    end;
  else goto NODE3;

NODE3:
  if MN_OCC gt .z then do;
    if MN_OCC <= 0.695203 then goto NODE4;
    else goto TNODE5;
    end;
  else if PCT_LT9 gt .z then do;
    if PCT_LT9 <= 0.197615 then goto NODE4;
    else goto TNODE5;
    end;
  else if PCT_WHIT gt .z then do;
    if PCT_WHIT <= 0.33532 then goto TNODE5;
    else goto NODE4;
    end;
  else if PCT_NONW gt .z then do;
    if PCT_NONW <= 0.66468 then goto NODE4;
    else goto TNODE5;
    end;
```

```
  else if PCT_HISP gt .z then do;
    if PCT_HISP <= 0.50295 then goto NODE4;
    else goto TNODE5;
    end;
  else if PCT_NOBA gt .z then do;
    if PCT_NOBA <= 0.241258 then goto TNODE5;
    else goto NODE4;
    end;
  else goto NODE4;

NODE4:
  if PCT_GAS gt .z then do;
    if PCT_GAS <= 0.813683 then goto TNODE1;
    else goto NODE5;
    end;
  else if PCT_ELEC gt .z then do;
    if PCT_ELEC <= 0.160852 then goto NODE5;
    else goto TNODE1;
    end;
  else if MD_FAM_I gt .z then do;
    if MD_FAM_I <= 71382.5 then goto NODE5;
    else goto TNODE1;
    end;
  else if PCT_OSTR gt .z then do;
    if PCT_OSTR <= 0.0212577 then goto NODE5;
    else goto TNODE1;
    end;
  else if PCT_ASIA gt .z then do;
    if PCT_ASIA <= 0.0690719 then goto NODE5;
    else goto TNODE1;
    end;
  else if MD_HH_IN gt .z then do;
    if MD_HH_IN <= 70425.5 then goto NODE5;
    else goto TNODE1;
    end;
  else goto NODE5;

TNODE1:
  &target = 0;
  &prob.1 =    0.8755009;
  &prob.2 =    0.1244991;
  &node = 1;
  goto MODELDONE;

NODE5:
  if T24_BEST in (13,14,15
                 ) then goto TNODE2;
  else if T24_BEST in (4,5,9,10,16
                      ) then goto NODE6;
  else if HISPLATI gt .z then do;
    if HISPLATI <= 3.5 then goto NODE6;
    else goto TNODE2;
    end;
  else if MDHOMEVA gt .z then do;
    if MDHOMEVA <= 131100 then goto TNODE2;
    else goto NODE6;
    end;
```

```
  else if PCT_NONW gt .z then do;
    if PCT_NONW <= 0.50036 then goto NODE6;
    else goto TNODE2;
    end;
  else if PCT_WHIT gt .z then do;
    if PCT_WHIT <= 0.49964 then goto TNODE2;
    else goto NODE6;
    end;
  else if MODINC gt .z then do;
    if MODINC <= 3.5 then goto NODE6;
    else goto TNODE2;
    end;
  else goto NODE6;

TNODE2:
  &target = 0;
  &prob.1 =    0.8724113;
  &prob.2 =    0.1275887;
  &node = 2;
  goto MODELDONE;

NODE6:
  if MDHOMEVA gt .z then do;
    if MDHOMEVA <= 259700 then goto TNODE3;
    else goto TNODE4;
    end;
  else if PCT_FEM gt .z then do;
    if PCT_FEM <= 0.580872 then goto TNODE3;
    else goto TNODE4;
    end;
  else if PCT_GRAD gt .z then do;
    if PCT_GRAD <= 0.137754 then goto TNODE3;
    else goto TNODE4;
    end;
  else if P_MB4_69 gt .z then do;
    if P_MB4_69 <= 0.226579 then goto TNODE3;
    else goto TNODE4;
    end;
  else if PCT_65 gt .z then do;
    if PCT_65 <= 0.239109 then goto TNODE3;
    else goto TNODE4;
    end;
  else if MD_FAM_I gt .z then do;
    if MD_FAM_I <= 91486.5 then goto TNODE3;
    else goto TNODE4;
    end;
  else goto TNODE3;

TNODE3:
  &target = 1;
  &prob.1 =    0.8327492;
  &prob.2 =    0.1672508;
  &node = 3;
  goto MODELDONE;

TNODE4:
  &target = 0;
```

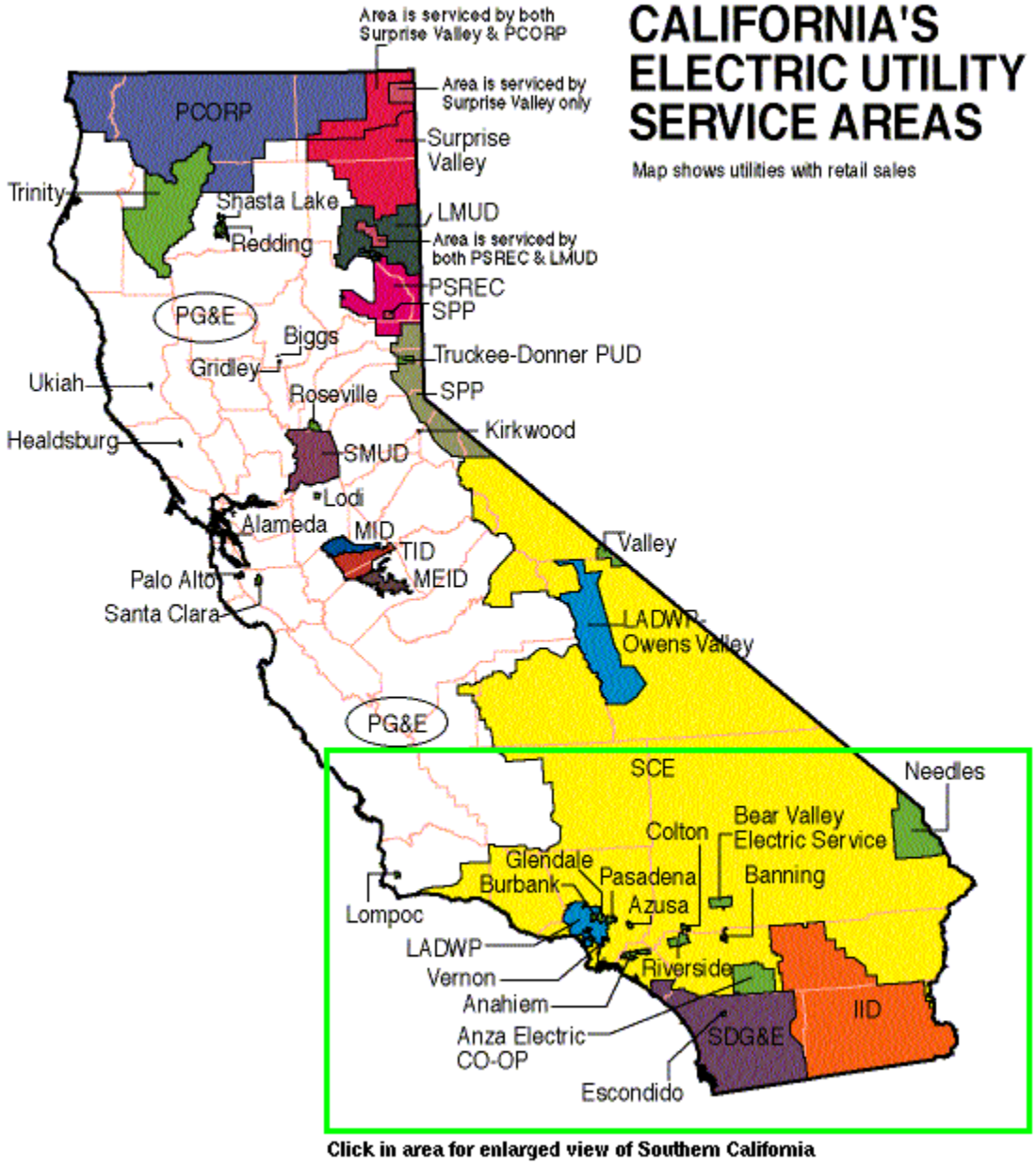
```
  &prob.1 =    0.8944444;
  &prob.2 =    0.1055556;
  &node = 4;
  goto MODELDONE;

TNODE5:
  &target = 0;
  &prob.1 =    0.8938292;
  &prob.2 =    0.1061708;
  &node = 5;
  goto MODELDONE;

NODE7:
  if PCT_BA gt .z then do;
    if PCT_BA <= 0.0118325 then goto TNODE6;
    else goto TNODE7;
    end;
  else if PCT_NOBA gt .z then do;
    if PCT_NOBA <= 0.075538 then goto TNODE6;
    else goto TNODE7;
    end;
  else if PCT_65 gt .z then do;
    if PCT_65 <= 0.023242 then goto TNODE6;
    else goto TNODE7;
    end;
  else if PCT_LT9 gt .z then do;
    if PCT_LT9 <= 0.534426 then goto TNODE7;
    else goto TNODE6;
    end;
  else if PCT_2PER gt .z then do;
    if PCT_2PER <= 0.0875339 then goto TNODE6;
    else goto TNODE7;
    end;
  else if PCT_PO_A gt .z then do;
    if PCT_PO_A <= 0.486124 then goto TNODE7;
    else goto TNODE6;
    end;
  else goto TNODE7;

TNODE6:
  &target = 0;
  &prob.1 =    0.8780037;
  &prob.2 =    0.1219963;
  &node = 6;
  goto MODELDONE;

TNODE7:
  &target = 1;
  &prob.1 =    0.8161409;
  &prob.2 =    0.1838591;
  &node = 7;
  goto MODELDONE;

TNODE8:
  &target = 1;
  &prob.1 =    0.8169303;
  &prob.2 =    0.1830697;
```

```
  &node = 8;
  goto MODELDONE;

MODELDONE:
return;

run;
```

Appendix I

Utility Service Territories

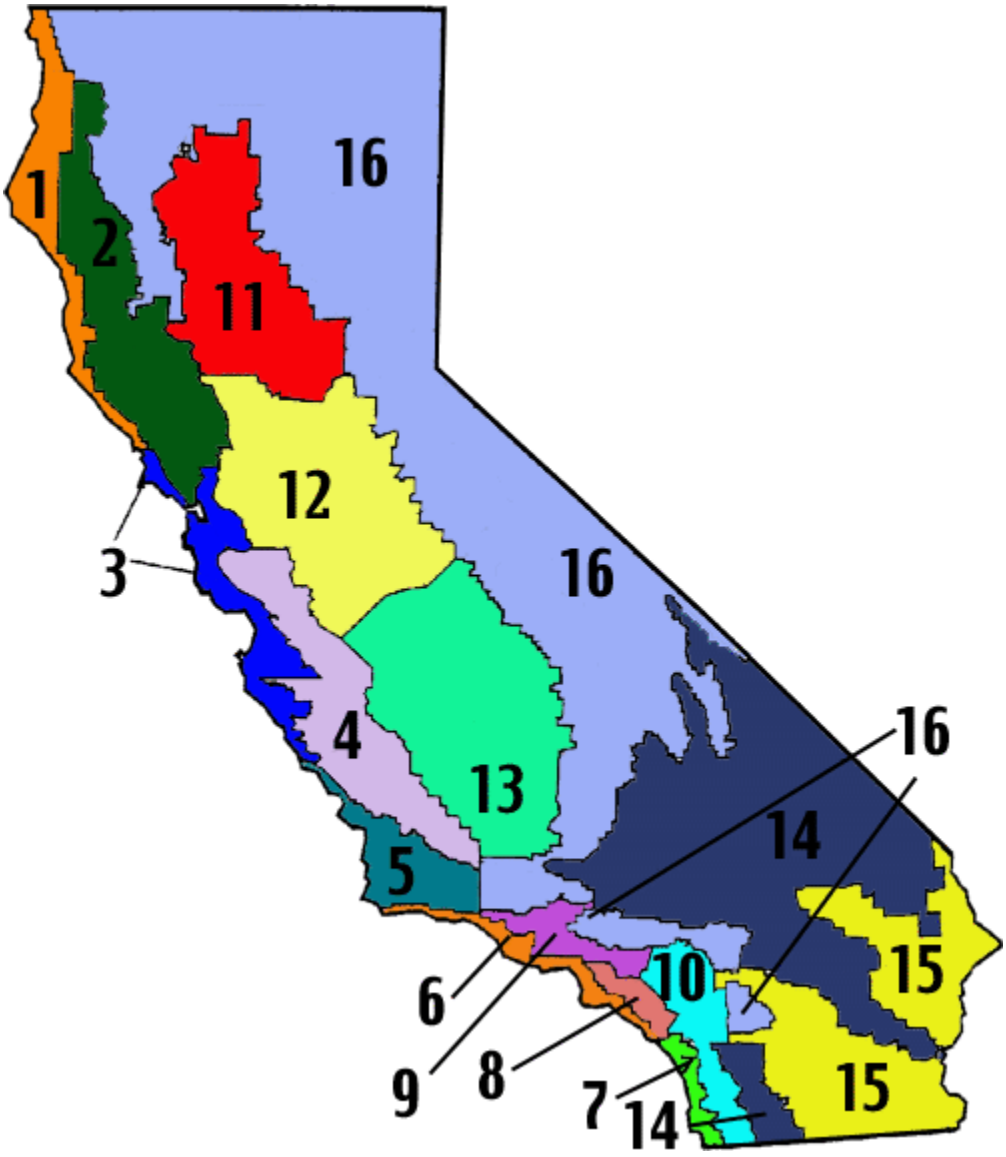
Figure I-1
Utility Service Territories



Appendix J

Title 24 Weather Zone Map

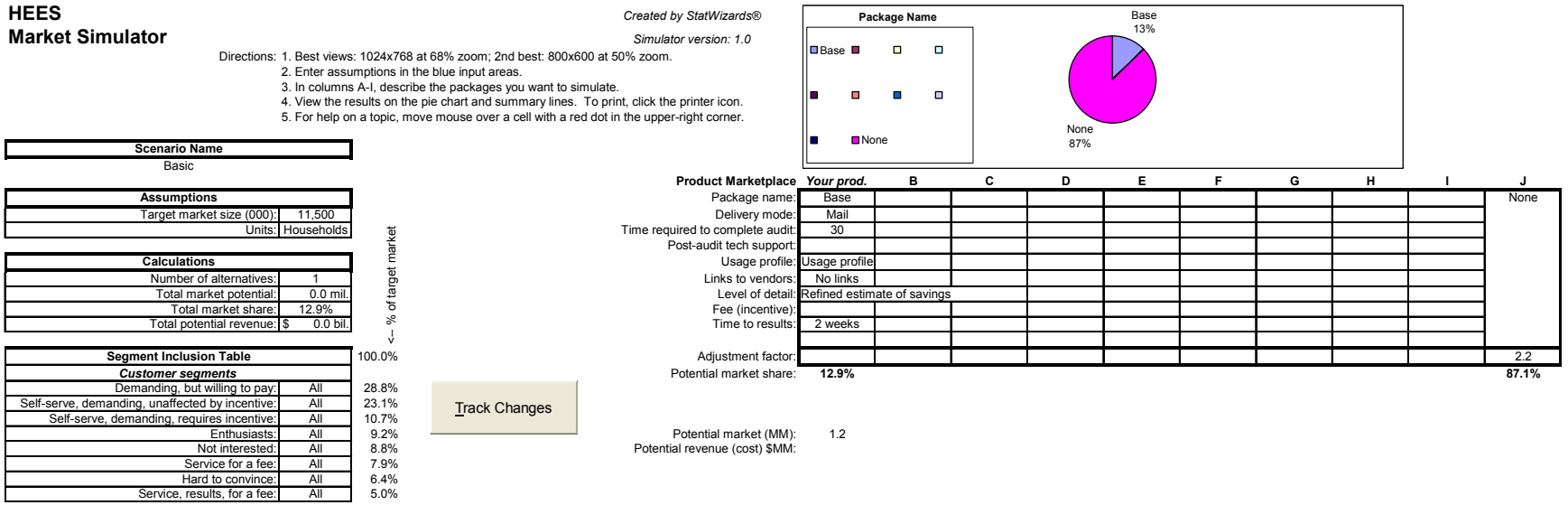
Figure J-1
Title 24 Weather Zone Map



Appendix K

Illustrations from Stated Preference Model

Figure K-1
Illustration of Main Page in the Simulator for the Stated Preference Model



HEES
Market Simulator

Created by StatWizards®

Simulator version: 1.0

Directions: 1. Best views: 1024x768 at 68% zoom; 2nd best: 800x600 at 50% zoom.
2. Enter assumptions in the blue input areas.
3. In columns A-I, describe the packages you want to simulate.
4. View the results on the pie chart and summary lines. To print, click the printer icon.
5. For help on a topic, move mouse over a cell with a red dot in the upper-right corner.

Scenario Name
Basic

Assumptions	
Target market size (000):	11,500
Units:	Households

Calculations	
Number of alternatives:	1
Total market potential:	0.0 mil.
Total market share:	12.9%
Total potential revenue:	$ 0.0 bil.

Segment Inclusion Table		<-- % of target market
		100.0%
Customer segments		
Demanding, but willing to pay:	All	28.8%
Self-serve, demanding, unaffected by incentive:	All	23.1%
Self-serve, demanding, requires incentive:	All	10.7%
Enthusiasts:	All	9.2%
Not interested:	All	8.8%
Service for a fee:	All	7.9%
Hard to convince:	All	6.4%
Service, results, for a fee:	All	5.0%

Track Changes

Product Marketplace	*Your prod.*	B	C	D	E	F	G	H	I	J
Package name:	Base									None
Delivery mode:	Mail									
Time required to complete audit:	30									
Post-audit tech support:										
Usage profile:	Usage profile									
Links to vendors:	No links									
Level of detail:	Refined estimate of savings									
Fee (incentive):										
Time to results:	2 weeks									
Adjustment factor:										2.2
Potential market share:	**12.9%**									**87.1%**

Potential market (MM): 1.2
Potential revenue (cost) $MM:

Figure K-2
Illustration of Diffusion Model in the Stated Preference Model

Bass Diffusion Model

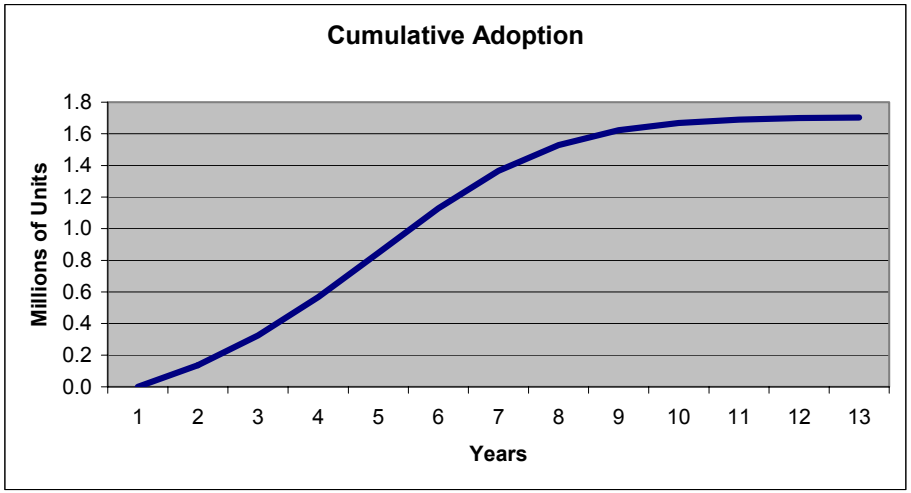
Directions: This worksheet takes the estimate of market potential for alternative A in the Main worksheet and plots an innovation diffusion curve using the Bass model. By choosing appropriate values of the parameters α and β, you can simulate the adoption path of alternative A over time.

α = 0.08 The *coefficient of external influence* , this parameter represents the effect of external factors such as media communications on the rate of adoption.

β = 0.5 The *coefficient of internal influence* , this coefficient represents the effect of prior adoption and word of mouth on new adoption.

Base Which package name (from Main sheet)?

Nbar = 1.7059 (potential market for alternative A from Main worksheet)



Years:	0	1	2	3	4	5	6	7	8	9	10	11	12
Diffusion factor*:	0.0	0.08	0.1904	0.3322419	0.4965912	0.6618581	0.8008104	0.8965021	0.951175	0.9783016	0.9906512	0.9960298	0.9983246
Adoption: 0	0.0	0.1	0.3	0.6	0.8	1.1	1.4	1.5	1.6	1.7	1.7	1.7	1.7
Unit sales:	0.0	0.1	0.2	0.2	0.3	0.3	0.2	0.2	0.1	0.0	0.0	0.0	0.0
Revenues:	$ -	$ 6.8	$ 9.4	$ 12.1	$ 14.0	$ 14.1	$ 11.9	$ 8.2	$ 4.7	$ 2.3	$ 1.1	$ 0.5	$ 0.2

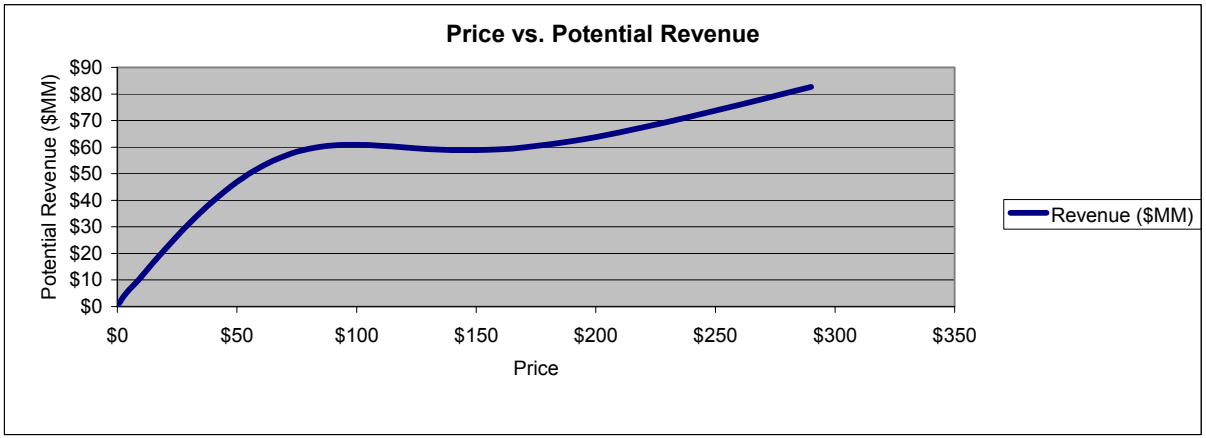
*Source: Easingwood, C.J., V. Mahajan, and E.W. Muller (1983) "A non-uniform influence innovation diffusion model of new product acceptance." *Marketing Science* 2: pp. 273-296. Discussed in Mahajan, V.,Robert Peterson, *Models for Innovation Diffusion* , 1985, Sage Publications, Newbury Park, California, 91320, pp. 32-34

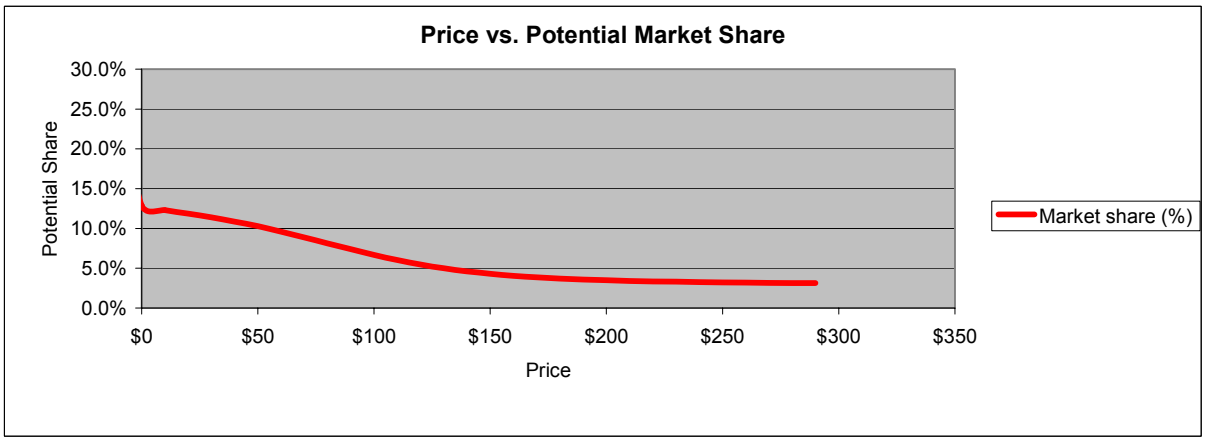
Figure K-3
Illustration of Price Curves in the Stated Preference Model

Directions: 1. Set starting price in B7 and increment in A7.
2. Press the button to generate price curves.
3. To restore your original scenario, press the button again.

Restore prior scenario

Price increment	Price	Revenue ($MM)	Market share (%)
$10	-$30	-$77.1	28.2%
	-$20	-$48.8	26.8%
	-$10	-$23.7	26.0%
	$0	$0.0	12.9%
	$10	$11.2	12.3%
	$20	$21.6	11.9%
	$30	$31.2	11.4%
	$40	$39.7	10.9%
	$50	$46.9	10.3%
	$60	$52.6	9.6%
	$70	$56.7	8.9%
	$80	$59.3	8.1%
	$90	$60.6	7.4%
	$100	$60.9	6.7%
	$110	$60.5	6.0%
	$120	$59.9	5.5%
	$130	$59.3	5.0%
	$140	$58.9	4.6%
	$150	$58.9	4.3%
	$160	$59.2	4.1%
	$170	$59.9	3.9%
	$180	$60.9	3.7%
	$190	$62.3	3.6%
	$200	$63.8	3.5%
	$210	$65.5	3.4%
	$220	$67.4	3.4%
	$230	$69.4	3.3%
	$240	$71.5	3.3%
	$250	$73.7	3.2%
	$260	$75.9	3.2%
	$270	$78.1	3.2%
	$280	$80.4	3.1%
	$290	$82.6	3.1%





Appendix L

References

Belsley, D.A., Kuh, E., and Welsch, R.E. *Regression Diagnostics*, New York: John Wiley & Sons, Inc., 1980.

Breiman, L., Friedman, J., Olshen, R., & Stone, C. (1984). Classification and Regression Trees. Pacific Grove: Wadsworth.

Gilpin, E; Olshen, R: Henning, H; and Ross, J., Jr. 1983. "Risk prediction after myocardial infarction: comparison of three multivariate methodologies." Cardiology, 70: 73-84.

Hosmer, David W. and Stanley Lemeshow. *Applied Logistic Regression*. New York: John Wiley & Sons, Inc., 1989.

King, Gary and Langche Zeng, "Logistic Regression in Rare Events Data" (2001), *Political Analysis*, vol. 9 (2), pp. 137-163.

Kuhfeld, Warren F., Randall D. Tobias and Mark Garratt. "Efficient Experimental Design with Marketing Research Applications." Journal of Marketing Research, November, 1994, pages 545-557.

Loether, Herman J. and Donald G. McTavish. *Descriptive and Inferential Statistics: An Introduction*. Boston, MA: Allyn and Bacon, Inc., 1976.

Magidson, Jay and Thomas Eagle & Jeroen Vermunt. (2003). New Developments in Latent Class Choice Models. *April 2003 Sawtooth Software Conference Proceedings*, 89-112.

Mahajan, Vijay, and Robert A. Peterson. (1985), *Models for innovation diffusion*. Newbury Park, CA: SAGE Publications.

Montgomery, Douglas C. (1997). *Design and Analysis of Experiments*. New York: John Wiley & Sons.

Ridge, Richard, Katherine Randazzo, and Quantum Consulting. "Final Report for the Evaluation of the California 2002 Home Energy Efficiency Survey Program." Submitted to the Southern California Edison Company, the Pacific Gas & Electric Company, the Southern California Gas Company, and the San Diego Gas & Electric Company, June 1, 2004.

Rogers, Everett M. (1995), *Diffusion of innovations*, New York: The Free Press.

Steinberg, D. & Colla, P. (1995). CART: Tree-Structured Non-Parametric Data Analysis. San Diego, CA: Salford Systems.