CAN WE RELY ON SELF-CONTROL?

Miriam L. Goldberg, XENERGY Inc., Madison, WI
Thomas S. Michelman, XENERGY Inc., Burlington, MA
Chris Ann Dickerson, Pacific Gas and Electric Company, San Francisco, CA

Introduction

Billing analysis for impact evaluation commonly incorporates survey data in regression models to control for non-program-related effects, and typically uses comparison groups as well. A limitation of survey-based regression models is that they are usually fit for only a small sample of the program participants. Some authors, however, have suggested the use of participant-only models. Others have emphasized the potential value of analysis based on entire participant populations, without survey data, rather than relying on small survey samples.

This paper compares billing analyses using survey data in three different ways. The first analysis uses the entire participant population, with no survey data, and a nonparticipant sample with limited survey data used for screening purposes only. We refer to this analysis method as the “screener survey model” . The second analysis uses participants only. The third incorporates survey variables on changes for a subset of the participants and nonparticipants. We refer to this as the “full-survey” model.

The program evaluated is a residential central air conditioning rebate program. The billing analyses all use similarly structured pooled time series/cross-sectional regression models. Such models have been discussed by Schiffman (1994), Megdal (1995) and Samiullah et al (1996). The comparison group is a pool of nonparticipants who had central air conditioning but did not replace or acquire a new system during the study period. The savings estimates from the regression are interpreted as gross savings relative to the prior equipment in place. Separate adjustments are made (1) to account for participants who added new systems rather than replacing existing equipment; (2) to compute gross savings relative to the standard-efficiency baseline; and (3) for free ridership. These adjustments use a combination of engineering analysis, regression analysis results, and proportions estimated from the participant survey.

The analysis methods are described in the next section. We then present the results of the analysis with the three different uses of survey data and the comparison group, and discuss the implications of these comparisons.

Methodology

Data Sources

The following data sources were used for each of the billing analyses.

Program Tracking Data**.** The tracking data included the customer identification number, type of measure installed, (packaged or split unit), tons of the installed units, and installation or program participation date and the program estimate of gross savings.

Billing Records**.** Billing records were matched to participants by identification number. The records for each customer included the beginning and ending of each meter reading period, number of days in the period, and amount consumed. The billing data used covered the period from January 1993 through October 1996.

Weather Data**.**. Each customer was assigned to one of PG&E’s 25 weather stations. The weather station assignment is part of the customer account number. Data taken from these weather stations were the daily temperatures for each day included in the billing analysis. In addition, we used the long-run average degree-days for each weather station, computed for the 12 year period from 1984 through 1995.

Customer Survey Data. A survey was conducted as part of the evaluation with participants in the Central Air Conditioner Program, as well as with a sample of nonparticipants. This survey was used both to support the billing analysis and for free rider estimation. Information collected on the survey included

- home ownership
- fuels used for end uses
- major changes that occurred over the study period and the dates of these changes

For the program participants, additional questions were asked regarding their participation. These questions were used to determine free ridership.

Participants were selected for the sample only if they had a minimum of 12 months of billing history prior to participation and nine months after participation. Nonparticipants were selected only if they had a minimum of 24 months of billing history. These are requirements of the CADMAC M&E Protocols (California Public Utilities Commission 1997) for inclusion in the analysis sample. A simple random sample of customers satisfying these criteria was selected for each surveyed group. Surveys were completed with a total of 214 participants and 1008 nonparticipants.

Initial Billing Analysis Model -- Screener Survey

The billing analysis approach was a pooled time-series/cross-sectional (TSXS) regression analysis to determine gross savings. That is, observations from all customers and all time periods in the analysis were combined into a single regression model. This regression was designed to estimate the gross effect on consumption of implementing the program measure. This "gross savings" actually included the effects of snapback, short-term measure persistence, and participant spillover. A separate adjustment for free ridership was made, based on survey results.

The comparison group included in the model was the set of all surveyed nonparticipants who had central air conditioning, and had not installed a new CAC system on their own over the time period included in the analysis. The model identifies the gross savings relative to the old system as the average change associated with participants' installation of the new system. Because nonparticipants who installed a new system are excluded from the model, there is no netting out of natural adoption. The nonparticipants do, however, control for other changes over the study period that are unrelated to the program but might have affected consumption.

The number of nonparticipants identified by the screener survey for inclusion in the model was much less than the number of participants available for the billing analysis. To adjust for this imbalance, and allow the nonparticipants to provide the desired control for exogenous changes, a weighted regression was used, with the nonparticipant observations weighted to correspond to roughly half the total. (Standard errors are still calculated recognizing the actual numbers of separate observations.)

Participants included in the initial model are all the participants for whom adequate billing records could be matched. This criterion provided a large pool of participants to include in the model, and allowed very good definition of the effect of installing the new system. The tradeoff was that survey data were not collected for most of these customers. Thus, the basic analysis used a large sample with limited information on each customer rather than a smaller sample with more detailed information on each customer. Effects of nonprogram changes are assumed to average out over time and over participants and nonparticipants included in the model. Variations on the model (a) using no comparison group and (b) using more detailed survey data are described below.

The terms included in the initial regression model are

- Customer-specific dummy variables (included implicitly, but not explicitly estimated by the model)
- Time-period dummy variables for each month in the analysis
- Heating degree-days, base 63°F (separate coefficients for nonparticipants, packaged system participants, and split system participants)
- Cooling degree-days, base 72°F (separate coefficients for nonparticipants, packaged system participants, and split system participants)
- Cooling degree-days interacted with tons of new equipment installed (for participants only, separate coefficients for packaged and split system participants)
- Time series participation dummy variable, interacted with cooling degree-days and the program estimate of savings.

The separate coefficients of degree-days for the different groups of customers allow for the possibility that these customers' response to temperature was different even prior to the installation of the new system. Thus, the average effect of installing each type of system is determined by the consumption change relative to that group's pre-installation pattern, not relative to the average pattern over all customers in the regression.

Likewise, the interaction of cooling degree-days with tons recognizes that homes with a higher projected cooling need, as reflected in the purchased tonnage, are likely to have higher consumption per degree-day. The tons in place prior to the installation of the new system is not necessarily the same as the new tons. Indeed, 64 percent of the participants reported that their new system had higher capacity than the old system. Nonetheless, the new tons installed is a useful indicator of the cooling load even in the pre-installation period.

Another reason to include the new tons as a predictor across all time periods is that the engineering savings estimate is proportional to tons. The incremental savings on a per-ton basis is most reliably determined if the baseline against which the increment is determined is also estimated on a per-ton basis. If the baseline usage is not scaled to tons but the savings effect included in the regression is, the coefficient of the savings term could be biased.

Because tons are known for participants but not for nonparticipants, the degree-day terms entered for these two groups take somewhat different forms. The nonparticipant degree-day coefficients have units of kWh/°F-day, while the participant coefficients have units of kWh/°F-day-ton. The two terms have similar effects in the regression.

__Initial TSXS model__. The form of the initial regression model fit is

$$\begin{aligned} Y_{jt} = \mu_j + \tau_t & \\ & + \beta_{HTg}\, HDD63_{jt} + \beta_{ACg}\, CDD72_{jt} \\ & + \beta_{TONg} CDD72_{jt} {*} TON_j \\ & + \gamma_g\, CDD72_{jt} {*} PST_{jt} {*} ENGKj \\ & + \varepsilon_{jt} \end{aligned}$$

where

g is a subscript indicating the participation type: packaged unit, split unit, or nonparticipant

Y_{jt} = consumption per day for customer *j* during time period *t*

$HDD63_{jt}$ = Heating degree-days per day base 63°F for customer *j*'s time period *t*

$CDD72_{jt}$ = Cooling degree-days per day base 72°F for customer *j*'s time period *t*

PST_{jt} = 0/1 dummy variable indicating that customer *j* implemented the program measure prior to time period *t*

TON_j = tons of the new unit installed through the program by participant j (zero for nonparticipants)

$ENGK_j$ = program engineering estimate of kWh savings for customer j (zero for nonparticipants)

ε_{jt} = residual error

In the pooled model, the terms μ_i are customer-specific intercepts. The terms τ_t are time trends. The coefficients β and γ are estimated by the regression. The dummy variables for participation PST_{jt} are zero for time periods *t* prior to customer *j*'s participation, and 1 thereafter.

The inclusion of the customer-specific and month-specific terms μ_j and τ_t is a first-order correction for the fact that observations for the same customer at different times or for the same time across customers are not all independent. Rather, some of the unexplained factors that make up the residuals, ε_{jt} will be similar across time periods t for a given customer j, and across customers j for a given time period. Excluding the customer - and time-specific effects would treat the model as if there were many more independent observations than there really are, with the result that the precision of the estimates would be exaggerated.

Some evaluation practitioners fit the pooled time series cross sectional models using participants only. The reasoning is that the exogenous changes are captured by those who have not yet participated in a given month. The limitation of this approach is that virtually all participants in a given year are "nonparticipants" during the first few months, and all are participants in the later months. As a result, any general (nonprogram) trends that made consumption different in the early months from that in the later months would be confounded with the participation effect. For this reason, a comparison group is included in the base model.

Since the model includes separate intercepts μ_j for each customer and separate heating and cooling slopes β_g for each group g, the effect of including nonparticipants in the model is not immediately apparent. The nonparticipants contribute to the estimation of the time-period fixed effects τ_t. Thus, for example, if there is an overall downward trend in consumption unrelated to the program, that downward trend would be captured in the time-period effects, rather than being erroneously associated with the implementation of savings measures.

For this program, the effect of the measure is expected to be temperature-related. To account for this relationship, the engineering estimate of savings ENGK is interacted with degree-days.

The index *t* indicates the month and year of the end date of the meter reading period. The dates used for the degree-day calculation are the reading dates specific to each customer. For example, for a customer *j* assigned to weather station 22 for a meter reading period *t* with begin date June 10, 1994 and end date July 8, 1994, cooling degree-days CDD_{jt} are computed using the daily temperatures from that weather station and that range of dates.

To estimate annual savings, the average annual value of each of the combination of terms interacted with the post-participation dummy variable is determined, and multiplied by the corresponding coefficient. The degree-day terms interacted with the post-participation dummy variables are calculated using long-run normal weather conditions, as used in the PG&E demand forecasting model. The average is computed across all customers in the tracking system. This approach satisfies the weather adjustment requirements of the CADMAC M&E Protocols (Tables C-1 and C-2).

Gross Savings Adjustments

As described above, the gross savings determined by the regression model is the savings relative to the prior condition. However, the gross savings as defined for the program are the savings relative to the baseline of standard efficiency equipment that would otherwise be installed. To determine the savings relative to the program baseline, two types of adjustments must be made. The first is to correct the regression gross savings for the inclusion of participants who added CAC systems where there was none before. The second is to apportion the total savings relative to old systems between (1) the savings moving from old efficiency to standard and (2) the savings moving from standard to program-eligible high efficiency.

<u>*Adjustment for CAC Participants Who Added CAC*</u>. The regression estimates the average change in consumption associated with acquisition of a new central air conditioning system. This average across all participants is the (weighted) average of the effect for replacers and the effect for adders.

For adders, the effect is an increase in consumption. Assuming the customer had no air conditioning before, the amount of this increase is the average UEC of a new efficient unit. For replacers, the effect is negative, with magnitude equal to the savings associated with changing from the old unit to the new one.

Thus, the estimated effect from the regression is

$$EFF_{REG} = a\ UEC_{NEW} - (1\text{-}a)\ SAV_{O}$$

where

a = fraction of participants who added CAC
UEC_{NEW} = average UEC of a new efficient unit
SAV_{O} = gross savings for replacement, relative to the old unit

We maintain the convention that an increase in consumption is a positive effect, but negative savings. Conversely, positive savings means a negative effect, or a decrease in consumption. That is, EFF_{REG} is a negative number, while SAV_{O} and UEC_{NEW} are positive numbers.

The same gross savings is assumed to apply to both replacement and added units. The base in either case is the standard-efficiency equipment that would otherwise have been installed. We assume that the rebate had no effect on the decision to replace or add a unit at all.

The UEC for a new unit is estimated by the UEC for old units, plus the incremental effect (savings) associated with replacing an old unit with a new one. That is

$$UEC_{NEW} = UEC_{OLD} - SAV_{O}$$

Thus,

$$\begin{aligned} EFF_{REG} &= a\ (UEC_{OLD} - SAV_{O}) - (1\text{-}a)\ SAV_{O} \\ &= -SAV_{O} + a\ UEC_{OLD}. \end{aligned}$$

Thus, we estimate the term of interest as

$$\begin{aligned} SAV_{O} &= -EFF_{REG} + a\ UEC_{OLD} \\ &= SAV_{REG} + a\ UEC_{OLD} \end{aligned}$$

where

$$SAV_{REG} = -\ EFF_{REG}$$

is the initial gross savings estimate from the regression.

Adjustment for Efficiency Base. The savings due to increasing the efficiency of a unit from $SEER_{LOW}$ to $SEER_{HI}$ can be calculated as the product of equivalent full-load hours of use, tons, and the difference in SEER, as follows:

$$SAV_{LOW\text{-}HI} = (\text{Hours})(\text{tons})C(1/SEER_{LOW} - 1/SEER_{HI})$$

where C is a conversion factor from tons to kWh. Thus, the total savings due to replacing old equipment with new high-efficiency equipment can be split between the savings increment for new standard equipment and the savings increment for moving above standard in proportion to the increments of 1/SEER. That is

$$SAV_{OLD\text{-}HI} = (\text{Hours})(\text{tons})C(1/SEER_{OLD} - 1/SEER_{HI})$$

and

$$SAV_{STD\text{-}HI} = (\text{Hours})(\text{tons})C(1/SEER_{STD}\ \text{-})\ 1/SEER_{HI}$$

so that

$$SAV_{STD-HI} = \frac{(1/SEER_{STD} - 1/SEER_{HI})}{(1/SEER_{OLD} - 1/SEER_{HI})} SAV_{OLD-HI}.$$

The standard-efficiency new-equipment baseline is specified by the program, as the 1993 Federal standard. The high-efficiency SEER actually installed is known from the program tracking data. The total gross savings from replacing old equipment with new high-efficiency equipment is determined from the regression analysis, with the adder adjustment described above. The final piece of information required to determine the gross savings relative to the program baseline is the SEER of the old equipment. This information is not known. Based on recent studies and practice, we assume that the stock efficiency of existing equipment is an average SEER of 8.8.

Discussion of Analysis Issues

Consistency of the Assumed Baseline. The free rider estimates are based on qualitative responses determining whether the customer would have purchased the efficient equipment or standard equipment in the absence of the program. The SEER associated with standard equipment is assumed to be the program baseline. However, what would actually have been sold as “standard” equipment is unknown, lacking market studies conducted during the program period.

Adjustment for Adders. The adder adjustment assumes that the UEC of existing equipment is accurately estimated by the same regression model used to isolate the effect of installing the new equipment. However, as shown in the Results section below, there is some seasonality evident in the estimated time-period effects τ_t. The direction and magnitude of these effects suggest that there is some cooling-related load captured in these effects rather than in the degree-day variables. If so, then the cooling UEC estimated by the degree-day terms may be understated.

On the other hand, the correction described may be overstated, because it assumes that all customers who reported adding a CAC unit where they didn’t have one before added a full-scale unit starting from nothing. In some cases, however, an adder customer may have previously had room air conditioning, or may have added CAC to a new addition rather than to the entire house. In either case, the amount of increase related to the addition of the new unit would be less than the average UEC.

Calculation of Standard Errors

The standard errors calculated for this analysis treat the adjustments for the adder proportion and for the program baseline efficiency as known constants. In fact, the proportion of adders is estimated from survey data; the error in this estimate should strictly be incorporated in the reported standard error. The error related to the baseline correction is not quantifiable statistically.

The free ridership rate is also estimated from the survey data, as the proportion of respondents who were classified as free riders based on their responses to this battery of questions. The standard error of the net savings estimate does take into account the uncertainty in this estimated proportion.

Other Modeling Approaches Attempted

In the course of this analysis, a number of other modeling approaches were explored. These attempts are described in brief below.

Other Efforts Attempted. Obtain the incremental savings for replacement through the program over and above the savings for basic replacement. This attempt was a model including both participants and nonparticipants, using customer-reported installation and dates. Meaningful results were not obtained, for the two reasons. First, we had too few nonparticipant replacers. Although there were about 40 nonparticipants who reported purchasing a new air conditioner, we had electric bills for only 17 of these customers. Second, customers did not seem able to report the installation dates accurately, even within one year, based on a comparison of survey-reported and tracking dates for participants.

- Separate the effects on consumption for replacers versus adders--customers who installed air conditioning equipment where there had not been any previously. These models gave unstable results, largely because of the limited number of adders in the participant and nonparticipant groups.
- Obtain the incremental savings per unit change in SEER. This model was fit across all participants in the tracking system, with one term for the base savings per ton associated with installation, and a separate term for the incremental savings per ton per SEER unit above the base level in the tracking system (SEER = 11). The intent was to develop a valid estimate of savings per ton per SEER from this model, and apply it to the entire SEER increment from the program base to the installed equipment. However, this model did not give meaningful results, probably because of the limited range of SEER above 11, particularly for packaged units.
- Allow separate degree-day coefficients for customers in different broad weather regions. This distinction was statistically significant, but did not substantially improve the quality of the estimates of interest.

Results of the Basic Analysis

Gross Savings Relative to Prior Conditions

Table 1 lists the variables used in the initial, screener survey regression model. Results of the regression and the specific form of the interaction terms included are shown in Table 2. The resulting gross savings estimates relative to prior conditions are shown in Table 3.

Table 1
Variables Included in the Pooled Regression Model

Variable	Description
HDD63	HDD/Day Base 63 (^{o}F-day/day)
CDD72	CDD/Day Base 72 (^{o}F-day/day)
NPART	Non-Participant Dummy
PSTCAC94	Time Series Participation Dummy
PACKAGE	Cross-Sectional Package CAC Participation Dummy
SPLIT	Cross-Sectional Split System CAC Participation Dummy
TON	New CAC Capacity (tons)
ENGK	Engineering Savings Estimate (kWh/year)

Table 2
Load Impact Regression Model

Parameter	Estimate	T	Pr > |T|	SE
Dec-92	-0.4642	-0.46	0.6465	1.0120
Feb-93	-1.1682	-5.40	0.0001	0.2165
Mar-93	-1.7554	-7.06	0.0001	0.2486
Apr-93	-2.1496	-8.23	0.0001	0.2610
May-93	-2.1167	-7.46	0.0001	0.2837
Jun-93	-0.4590	-1.53	0.1268	0.3006
Jul-93	1.3903	4.48	0.0001	0.3104
Aug-93	0.3587	1.17	0.2438	0.3077
Sep-93	-0.4938	-1.66	0.0964	0.2970
Oct-93	-1.0192	-3.59	0.0003	0.2841
Nov-93	-1.3398	-5.92	0.0001	0.2265
Dec-93	0.1766	0.86	0.3921	0.2064
Jan-94	-0.5687	-2.74	0.0061	0.2072
Feb-94	-1.4514	-6.76	0.0001	0.2147
Mar-94	-1.7579	-7.26	0.0001	0.2420
Apr-94	-1.9419	-7.37	0.0001	0.2636
May-94	-1.5564	-5.58	0.0001	0.2788
Jun-94	0.5545	1.82	0.068	0.3039
Jul-94	2.8851	9.15	0.0001	0.3153
Aug-94	2.0908	6.72	0.0001	0.3110
Sep-94	0.1683	0.56	0.5725	0.2983
Oct-94	-1.3162	-4.83	0.0001	0.2725
Nov-94	-1.2874	-6.16	0.0001	0.2091
Dec-94	0.1372	0.66	0.5119	0.2091
Jan-95	-0.1190	-0.56	0.5771	0.2134
Feb-95	-1.3433	-5.84	0.0001	0.2299
Mar-95	-1.6083	-6.98	0.0001	0.2303
Apr-95	-1.9078	-7.80	0.0001	0.2446
May-95	-1.5604	-5.75	0.0001	0.2712
Jun-95	0.7674	2.61	0.0091	0.2940
Jul-95	3.9001	12.55	0.0001	0.3108
Aug-95	3.3970	10.98	0.0001	0.3094
Sep-95	1.0966	3.67	0.0002	0.2990
Oct-95	-0.5768	-2.03	0.0422	0.2840
Nov-95	-0.2792	-1.07	0.2863	0.2618
Dec-95	0.5457	2.53	0.0115	0.2160
Jan-96	-0.3389	-1.62	0.1059	0.2096
Feb-96	-0.7631	-3.33	0.0009	0.2288
Mar-96	-1.4158	-5.92	0.0001	0.2391
Apr-96	-1.3339	-5.16	0.0001	0.2586
May-96	-0.1231	-0.43	0.6687	0.2877
Jun-96	2.4437	8.12	0.0001	0.3009
Jul-96	5.2596	16.56	0.0001	0.3176
Aug-96	5.0292	15.75	0.0001	0.3193
Sep-96	0.7051	2.37	0.0179	0.2979
Oct-96	-0.8101	-2.64	0.0082	0.3065
HDD63*NPART	0.3017	21.06	0.0001	0.0143
NPART*CDD72	1.5107	100.48	0.0001	0.0150
HDD63*PACKAGE	0.3529	22.13	0.0001	0.0159
CDD72*PACKAGE	0.9428	14.04	0.0001	0.0671
CDD72*PACKAGE*TON	0.3019	15.23	0.0001	0.0198
CDD72*PACKAGE*PSTCAC94*ENGK	-0.0019	-21.84	0.0001	0.0001
HDD63*SPLIT	0.3039	18.79	0.0001	0.0162
CDD72*SPLIT	0.4642	5.59	0.0001	0.0830
CDD72*TON*SPLIT	0.4169	17.55	0.0001	0.0237
CDD72*PSTCAC94*ENGK*SPLIT	-0.0017	-15.23	0.0001	0.0001

Dependent Variable:	**kWh/day**
Number of Customers:	**2,822**
Number of Observations:	**125,789**
R^2 =	**0.791**

Table 3
Unit Gross Savings Relative to Prior Conditions
from Direct Regression Results

Unit Type	Variable	Coefficient	T	SE	Variable Mean: Pooled regression data set	Variable Mean: Cross-Sectional Tracking Data (long-run normal)	Annual Savings per Unit
							kWh/year
Packaged	CDD72*PSTCAC94*ENGK*PACKAGE	-0.00191	-21.8	0.00010	443.18	558.60	389.8
Split	CDD72*PSTCAC94*ENGK*SPLIT	-0.00172	-15.2	0.00010	260.01	298.00	186.6

Gross Savings Adjusted for Adders

Table 4 shows the adder adjustment as described above. The results show a substantial understatement of the gross savings in the unadjusted regression estimate. One-eighth of the packaged unit participants and over one-fourth of the split unit participants added CAC systems where there had not been one previously. Correcting for the inclusion of these customers in the regression increases the packaged system gross savings by about one-third, and more then doubles the estimate for split systems.

Table 4
Unit Gross Savings Adjusted for Adders

	Direct Regression Savings Relative to Old	Cooling UEC, Old Eqt	Fraction of Adders	Adjusted Gross Savings Relative to Old
	(kWh/year)	(kWh/year)		(kWh/year)
A	B	C	D	E
Source:	Regression	Regression	Survey	B+D*C
Packaged	389.8	1566.2	0.125	586
Split	186.6	738.9	0.280	394

Gross Savings Adjusted to the Program Baseline

Table 5 shows the baseline adjustment as described above. Also shown in the table is the free rider rate. The free rider rate was determined by analysis of the survey data. Details on this analysis are given in the full report on this study (XENERGY Inc. 1997.)

Table 5
CAC Program Unit Gross Savings Adjusted for Program Baseline and Free Riders

	Adder-Adjusted Savings Relative to Old (kWh/year)	Unit Savings: Assumed Old SEER	Baseline SEER	Average Participant SEER	Baseline Adjustment	Savings Relative to Base (kWh/year)	Free Rider Rate	Net Savings (kWh/year)	Ex Ante Program Estimate
	A	B	C	D	E	F	G	H	I
Source:	Regression	Engineering	Federal Standard	Tracking	(1/C-1/D) / (1/B-1/D)	ExA	Surveys	(1-G)xF	Tracking
Packaged	586	8.8	9.7	11.6	0.62	363	0.12	319	234
Split	394	8.8	10.0	12.0	0.55	216	0.12	190	254
Program	470	8.8	9.9	11.8	0.58	274	0.12	241	246

The table shows that the overall electricity savings estimated by the evaluation are almost identical to the program planning estimates. However, the savings are higher than the planning estimates for packaged units, and lower than the planning estimates for split units.

The primary reason for the lower savings for split units and higher for packaged appears to be the location of the two types of units. The packages units are found in hotter climates, where usage and corresponding savings are somewhat higher than for a typical customer. By contrast, the split units were found in milder climates, where usage and savings were lower than for a typical customer.

Comparison of Results with Alternate Model Fits

Table 6 summarizes the regression results for the initial model, and for the two variations. For the participant-only model, the same terms are included as for the initial model (except for the nonparticipant terms). For the model requiring survey data for both participants and nonparticipants, the additional terms included are

- Change in number of occupants (0 prior to the change, equal to the number of additional people thereafter)
- Replaced windows (0 prior to replacement, 1 thereafter)
- Replaced windows interacted with cooling degree-days, separate coefficients for participants and nonparticipants
- Added insulation (0 prior to replacement, 1 thereafter)
- Added insulation interacted with cooling degree-days, separate coefficients for participants and nonparticipants

Table 6
Comparison of Regression Results

	Screener Survey		Participant-only		Full Survey	
	Co-efficient	t	Co-efficient	t	Co-efficient	t
RPLCW					-0.0285	-0.1
ADDINS					-1.267	-2.4
DPEOP					3.224	7.5
ADDINS*CDD72					0.0806	0.4
ADDINS*CDD72*CAC94					-0.156	-0.8
RPLCW*CDD72					0.616	3.8
RPLCW*CDD72*CAC94					-0.263	-1.5
HDD63*NPART	0.301	21.1			0.371	9.7
CDD72*NPART	1.510	100			1.541	38.0
HDD63*PACKAGE	0.352	22.1	0.438	28.7	0.304	7.4
CDD72*PACKAGE	0.942	14	0.991	19.8	0.221	1.4
CDD72*PACKAGE*TON	0.301	15.2	0.302	20.6	0.510	10.6
CDD7*PACK*PSTCA*ENGK	-0.00191	-21.8	-0.00196	-27.2	-0.00186	-8.6
HDD63*SPLIT	0.303	18.8	0.393	24.4	0.383	9.1
CDD72*SPLIT	0.464	5.59	0.569	9.1	0.0135	0.1
CDD72*TON*SPLIT	0.416	17.6	0.412	23.3	0.581	8.6
CDD7*PSTC*ENGK*SPLIT	-0.00172	-15.2	-0.00181	-19.2	-0.00279	-9.6
Number of cases	**observations**	**customers**	**observations**	**customers**	**observations**	**customers**
Participant n	116,797	2,618	116,797	2,618	8,063	192
Nonparticipant n	8,992	204	0	0	8,547	178
R^2	0.808		0.788		0.804	

Participant-Only Model

The results for the regression using participants only are quite similar to those for the initial, screener survey model. Both the magnitude of the coefficients and the estimated t-statistics are quite similar for each term. The R^2 for the model is also similar to that for the basic model.

These similarities indicates that there were no important time trends that the nonparticipant group was needed to control for. An unweighted version of the screener survey regression (not shown) also gave similar results. This similarity is expected, given that the presence of the nonparticipants does not appear to affect the results even with the weights.

Full Survey Model

For the full survey, there are more differences between the coefficients and those from the screener survey model. In addition, the t-statistics are smaller by about a factor of three. This reduction in t-statistic is consistent with what is expected based on the reduction in the sample size by a factor of about 8. The R^2 is slightly improved compared to the screener survey model. This result is expected, since additional terms have been added to the model.

The full survey model results show statistically significant increases in consumption associated with adding people, and statistically significant decreases associated with adding insulation. Somewhat curiously, replacing windows is associated with a statistically significant increase in consumption, possibly because of other expansion of the house associated with this measure. (For windows and insulation, the overall effect is the combined effect of the dummy term by itself and interacted with degree-days.) While the effects of these change variables appear to be important, they are not determined as accurately as the savings effects that are of primary interest. The difficulty with estimating these terms accurately stems in part from the smaller number of customers who had these changes, compared to the number who had the primary measure under study (Table 7).

Table 7
Average Values of Change Variables Included in the Model

Change Variable	Participants (n=340)		Nonparticipants (n=254)	
	Mean Value of Change Variable within the Analysis Time Frame	% with non-zero values	Mean Value of Change Variable within the Analysis Time Frame	% with non-zero values
Replace Windows	0.1676		0.1024	
Added Insulation	0.1559		0.065	
Change in # of People Living at Residence	-0.0191	15.9%	-0.0374	13.8%

An additional reason it may be difficult to estimate these effects with accuracy is that the timing of the changes is not reliably reported. As noted, participant reports of the timing of their CAC additions were often considerably at odds with the tracking system records. It is likely that similar inaccuracies would be associated with the timing of window replacements or insulation additions. Nonetheless, the goal of including the change terms in the model is not to determine these terms with great accuracy, but to incorporate sufficient information to avoid biases in the savings estimate due to their omission.

Comparison of Savings Estimates

Table 8 summarizes the savings estimates from the three models. Savings and standard errors for the participant-only model are very similar to those for the screener model. This similarity is expected, given the correspondence of the individual coefficients and their standard errors.

Table 8
Savings Estimates from Different Models

		Direct Regression Estimate of Savings	UEC of old equipment	Proportion of Adders	Savings Relative to Efficiency of Old Equipment	Baseline Adjustment	Savings Relative to Baseline Efficiency	Net-to-Gross	Net Savings
Screener Survey									
Packaged	Estimate	390	1566	0.13	586	0.62	363	0.88	319
	SE	18	19	0.00	20	0.00	12	0.02	13
Split	Estimate	187	739	0.28	394	0.55	216	0.88	190
	SE	12	13	0.00	15	0.00	8	0.02	8
Participants Only, No Survey									
Packaged	Estimate	400	1605	0.13	600	0.62	372	0.88	327
	SE	15	16	0.00	16	0.00	10	0.02	11
Split	Estimate	197	773	0.28	413	0.55	227	0.88	200
	SE	10	11	0.00	13	0.00	7	0.02	8
Full Survey									
Packaged	Estimate	380	1540	0.13	572	0.62	355	0.88	312
	SE	44	47	0.00	48	0.00	30	0.02	27
Split	Estimate	303	781	0.28	522	0.55	287	0.88	253
	SE	32	33	0.00	39	0.00	21	0.02	20

For the full-survey model, the packaged-unit savings estimates are again quite similar to those from the initial, screener survey model. The standard errors are larger, because of the smaller sample size. For the split units, however, there is a substantial difference between the savings estimates from the full-survey model and those from the screener-survey model.

These differences raise the question of which estimate is most reliable. The significant difference in split-system savings for the full-survey model compared to the initial model indicates that omitting the change terms that are available only for the surveyed sample leads to a biased savings estimate in this case. Thus, the tight precision indicated by the small standard error of the savings estimate for the models that use all the participants is deceiving.

This analysis has not attempted to develop the most complete and reliable model possible utilizing the available survey data. The full survey results shown are offered not as the best possible unbiased estimates, but as an indication of what the value of including survey data might be.

Conclusions

This analysis has explored the use of a large billing analysis sample, avoiding the need for survey data. The appeal of such an approach is

1. the high precision obtainable from the large sample size utilizing all participants with adequate billing histories;
2. the reduction in the complexity of the model development, and associated uncertainty regarding the “best” specification;
3. the potential for cost savings by eliminating the need for to conduct surveys.

The potential for cost savings from a survey-free billing analysis is actually somewhat limited. Usually, survey data are needed to address free ridership, unless either the program design is such that this is not an issue, or estimates are obtained from some other source. Development of free rider estimates relies at a minimum on participant survey responses, and often requires nonparticipant surveys as well for discrete choice analysis or self-selection corrections in billing analysis models that estimate net savings directly.

For the evaluation presented here, free ridership was estimated from participant surveys conducted specifically for the evaluation. The nonparticipant sample used for the billing analysis was drawn from a larger nonparticipants survey conducted to support this and other evaluations. The motivations for pursuing the survey-free analysis were the first two reasons above.

If we wanted to obtain the same level of precision from the survey-based analysis as was obtained from the survey-free analysis, survey sample sizes roughly eight times as large would have been required. This level of data collection would be impractical for most evaluations. Moreover, pursuing these larger sample sizes for the sake of improving the precision of the gross savings estimate would be a mis-allocation of evaluation resources. Even with the lower-precision, full-survey model for gross savings, the predominant source of uncertainty in the net savings estimate comes not from the error in the gross savings estimate, but from the uncertainty of the free rider estimate. The relative standard error of gross savings is about 8 percent, while the relative standard error of the free rider estimate is about 19 percent, resulting in an overall standard error of net savings of 20 percent. Even if the gross savings estimate were perfect, the net savings would still have a relative standard error of 19 percent.

Thus, rather than pursuing methods to refine the gross savings estimate, it would be more effective to expand the participant sample size to improve the free rider estimate. Taking into account both the free rider standard error and the regression error, doubling only the participant survey sample size would reduce the standard error of net savings to about 15 percent; quadrupling the participant survey sample size would cut the error roughly in half. If these sample size increases cannot be supported, the uncertainty in the free rider estimate remains the limiting factor, and improving the precision of the gross savings estimate will do little to improve the net savings precision.

In this analysis, the addition of the nonparticipants to the regression model using all participants, and no survey data, had little effect on the estimates. In this case, it might have been more cost-effective to conduct the surveys for a larger sample of participants, reducing the standard error of the free rider estimate, and relying only on participants, with surveys, in the billing analysis. The difficulty is that this approach does not include control for time trends. While such control appears not to have been needed in this case, that condition would be difficult to determined in advance. A broader issue in this regard is the need for further attention to how best to specify pooled regression models to control for exogenous time patterns.

There are contexts in which a survey-free analysis can be justified. Such cases would be where free ridership is not an issue, and there is a high degree of confidence that the nonparticipant group is comparable to the participant group in terms of the propensity to implement other changes that might affect consumption. For example, comparison groups composed of previous and future participants are sometimes appropriate.

More commonly, surveys are required to address free ridership, a comparison group is needed, and survey questions on key changes are needed from both participants and nonparticipants. The examples presented here show that the survey-free analysis can give results very close to those that would be obtained with the more complete survey data, or may give biased results. In this case, both results were obtained in the context of the same program. Without conducting the survey-based analysis, the extent of bias cannot be known. Thus, the apparent precision of the large-sample models using limited customer information should not be interpreted as high accuracy, if the potential for biases has not been accounted for.

References

California Public Utilities Commission 1997. *Protocols and Procedures for the Verification of Costs, Benefits, and Shareholder Earnings from Demand-Side Management Programs.* CA PUC Decision 93-05-063 Revised 1/97.

Megdal, L., E. Paquette, and J. Greer. 1995. "The Importance of Using Analysis of Covariance, Diagnostics, and Corrections within Billing Analysis for Large C&I Customers." *Proceedings of the 1995 International Energy Program Evaluation Conference*, pp. 433-439, Energy Program Evaluation: Uses, Methods, and Results, Chicago, IL.

Samiullah, S., M.L. Goldberg, E. Kademan, and Kenneth Train 1996. "Bells, Whistles, and Common Sense: Billing Analysis of a Residential HVAC Rebate Program." *Proceedings from the ACEEE 1996 Summer Study on Energy Efficiency in Buildings*, American Council for an Energy-Efficient Economy, Washington, D.C.

Schiffman, D.A. 1994. "A *Monte Carlo* Based Comparison of Techniques for Measuring the Energy Impacts of Demand-Side Management Programs." *Proceedings from the ACEEE 1994 Summer Study on Energy Efficiency in Buildings*, Vol. 7, pp. 7.213-7.222. American Council for an Energy-Efficient Economy, Washington, D.C.

XENERGY Inc. 1997. *1994 Residential Weatherization Retrofit Incentives and Appliance Efficiency Incentives Programs Impact Evaluation Final Report.* Prepared for Pacific Gas and Electric Company, San Francisco, CA.