Facing Uncertainty from Within and Without

An Impact Evaluation of the California Low Income Energy Efficiency Program

Kathryn Parlin, West Hill Energy and Computing, Chelsea, VT
Al Bartsch, West Hill Energy and Computing, Chelsea, VT
Scott Pigg, Energy Center of Wisconsin, Madison, WI

ABSTRACT

This paper provides an overview of methods for, and key results from, an impact evaluation of the 2002 California Low Income Energy Efficiency (LIEE) program. Impact estimates were based on statistical analysis of monthly utility billing data, which indicate that overall, the program results in savings of about 7 percent of annual electric use per household and 2 percent for gas. A variety of program- and data-related issues as well as external factors create significant uncertainty in assessing measure-level savings, despite the large study groups available for the analyses. These issues point to the importance of supporting research results with independent sources and alternative analysis, as well as providing sufficient lead time to allow for a variety of impact evaluation methods to be considered.

Introduction

This project was designed to evaluate the impacts of the California Low Income Energy Efficiency ("LIEE") program for program year 2002. As the project unfolded, a number of challenges arose, demonstrating the difficulties of constructing a study that would produce reliable and defensible estimates for this program. The nature of the obstacles can roughly be divided into two categories: internal factors stemming from the characteristics of the program and external factors outside the control of the managers.

The structure of the program has a direct effect on the delivery of services, from the eligibility requirements and measures installed through to the delivery mechanism and data collection. Management decisions also have an impact on key elements of the program, such as the accessibility and scope of the tracking information. In the LIEE program, numerous contractors install a wide variety of measures and data collection needs are relatively basic. These features of the program limited our options and increased the uncertainty associated with the results.

Ideally, billing analysis should be based on typical participants during a typical year. In this case, our analysis period spanned from 2001 through the end of 2003, encompassing the volatility in the energy market caused by the 2001 California Energy Crisis. In addition, the time frame of our study was coincident with the transition period related to changes in the program structure instituted during 2001.

The impact evaluation was initiated at the beginning of 2004 and designed to assess the savings for program year 2002. The fact that we were brought in after the completion of the 2002 program year and that the program tracking data included little detail about the pre-installation conditions severely constrained the choice of methods.

Given these circumstances, our research plan incorporated a number of complementary strategies. The final report contained these components:

- overview of evaluation methods and discussion of the specific strategies that could be applied to the LIEE program as currently implemented
- billing analysis to develop disaggregated measure-specific estimates (to the extent possible), using alternative methods to give a more realistic estimate of the precision of the results
- review of independent studies to provide a range of estimated savings for the major measures
- analysis of the consumption trends within the LIEE billing history
- summary of available research regarding the impacts of the 2001 Energy Crisis on consumption patterns

Our objective was to provide an accessible report that placed our results in context of the expected range of estimated savings and overall trends in energy use during the period. This approach gave us the information needed to assess the validity, and make recommendations for the application, of the results.

Background

The LIEE program is delivered throughout the state of California by the major gas and electric utilities. The participating utilities include Pacific Gas & Electric (PG&E), Southern California Edison (SCE), San Diego Gas & Electric (SDG&E) and Southern California Gas (SoCalGas). The program is designed to help low income households conserve energy (both electricity and natural gas), thus lowering monthly energy costs and reducing the financial burden of energy bills. All services are provided free of charge to participating households. In PY 2002, the program reached over 118,000 households with electric measures, and improved gas efficiency in more than 70,000 households.

The goal of the program is to install all feasible energy efficiency measures in qualifying low-income households. The number of CFL bulbs installed per home is capped at five. Although the program was designed to be comprehensive, the penetration of major measures, such as attic insulation, was low due to the characteristics of the housing stock.

Overall, the services and measures offered through the participating utilities are equivalent and consist of energy education and the installation of energy savings measures. The program installs a wide range of energy savings measures associated with air conditioning, lighting, refrigeration, water heating and space heating. Beginning in PY 2001, the program also began offering “rapid deployment measures” in response to the California Energy Crisis. These measures include a variety of equipment replacement, maintenance and repair options targeted at cooling systems and water and space heating. The measures offered through the program vary somewhat depending on service territory and climate zone.

Sources of Uncertainty

The combination of constraints created by the internal and external factors limited our options and largely determined the methods and strategies to be used for the PY2002 impact evaluations. The internal influences were created by the structure and management of the program, and include data constraints, the number and variety of eligible measures, and variations across contractors and utilities. Circumstances beyond the control of the program manager, such as overall trends in energy use and policy changes, also had an effect on the evaluation strategies and results.

There were a number of aspects of the LIEE program itself that complicated the analysis, as discussed below.

- The utilities requested measure-level savings by utility, housing type and measure. Over twenty measures were offered by the four participating utilities to low-income customers in three housing types, many of which would typically be expected to result in small savings.
- Services are provided by a number of distinct contractors in each utility's territory and, even with standardized program protocols, there are likely to be differences in installation practices and applicability of measures that may affect savings.
- The program tracking systems were not an entirely reliable source of some valuable data to inform the billing analysis, such as the fuel type for space and water heating.
- Some measures, such as furnace and evaporative cooler repair and replacement, restore functionality to an appliance that was previously inoperable, with the result that energy use increases rather than decreases following participation in the program.
- Measurement error is an inherent part of billing analysis, due to the fact that overall changes in household consumption are used as a proxy to estimate the savings from efficiency programs although variations in the billing data may be caused by many factors external to the program (TecMarket Works, 2004).

External factors contributing to the difficulties in obtaining measure estimates include the following:

- The analysis period extended from 2001 through 2003, and energy consumption was during this period affected by the 2001 California Energy Crisis.
- On the policy level, changes were made to the program in 2001, including the addition of "Rapid Deployment" measures offered to both current participants and participants receiving services in PY2001, and implementation changes may still have been felt during PY2002.
- Our team was brought in to conduct the impact evaluation in 2004, well after the completion of the program year to be evaluated.

The sources of uncertainty, effects on the analysis and results are summarized in Table 1.

Implications

These uncertainties have ramifications on the types of analysis that can be conducted and the results of the analyses. Initially, we debated the relative merits of calibrated engineering approach and billing analysis, but the lack of detailed information regarding pre-installation conditions severely reduced the feasibility of any engineering-based approach. Participant surveys were considered but ultimately rejected, since they depend on the recollection of occupants who generally lack the necessary technical expertise and may not have perfect recall of events occurring two years prior to the survey. Given the large sample sizes (over 48,000 households with complete billing history for electric and 43,000 for gas), we decided to proceed with a billing analysis.

Table 1. Summary of Sources of Uncertainty and Effects

Sources of Uncertainty	Effect on Analysis	Impact on Results
Program Complexity • Many measures (> 20) • Some measures installed in almost all homes • Some measures typically installed together • 3 housing types • 4 utilities • Many contractors	• Increases random variations in savings • Creates many variables, some with small savings • Introduces collinearity among measures installed concurrently	• Increases variances and obscures treatment effects • Difficult to disentangle measure-level savings • May result in estimators of the incorrect sign
Data Restrictions • Unreliability of water and space heating fuel designations • Unknown existing condition for equipment replacements	• Introduces random error into the explanatory variables	• Exerts downward bias on estimators
Measurement Error • Billing data not a direct measurement of energy savings • Observations are not independent within homes	• Increases random variation (noise) in the response variable • Causes autocorrelation among the observations within each home	• Noise increases variances and obscures treatment effects • Precision tends to be overstated in auto-correlated data sets
Consumption Trends • Dip in average consumption levels during pre-installation period due to Energy Crisis	• Low pre-installation use obscures savings • Comparison group may not entirely account for effects	• Exerts downward bias on estimators
Structural Changes to LIEE • Addition of rapid deployment measures • Return visits to previous participants	• Increase noise in the savings estimates • May introduce systematic error for some groups of participants	• Increases variances • May create bias of unknown impact

Although the data are readily available, billing analysis has its own set of limitations. This type of analysis is most useful for estimating savings at the household level and for measures that save a substantial proportion of the total household consumption. However, for this evaluation we were seeking to develop measure-level estimates with savings that are small in comparison to consumption levels. Specific issues relating to the billing analysis include the following:

- Insufficient data due to minimal tracking requirements and variations in measurement across utilities and contractors introduce random error into the explanatory variables, making it more difficult to estimate treatment effects and creating a downward bias on the savings estimates (Ridge, 1997).
- The program covers many measures with some packages of measures installed concurrently, resulting in many variables and increasing the possibility of collinearity among measures installed as a group (Belsley, 1980).
- Monthly billing data for a given household tend to be correlated; while this effect will not produce biased estimators, the variances tend to be understated using ordinary least squares ("OLS") regression that assumes independent observations (Belsley, 1980).
- For equipment replacements, lack of information regarding the condition of the equipment prior to the installation prevented us from identifying the homes in which the savings may reasonably be expected to be found.
- External trends in energy use indicate depressed use during the pre-installation period; while the comparison group helps to mitigate the biasing effect on savings estimates, the period was unusual enough to create concern that residual effects may still have an impact on the results.
- Variability in installation methods and the identification of eligible applications is likely to be higher during the transitional period associated with the addition of the "rapid deployment measures."

Solutions

While billing analysis seemed to be the only viable option for producing program-specific savings estimates, we were concerned that a regression on the billing data may produce improbably precise and possibly biased results given the numerous issues with the data. Under these circumstances, we were uncomfortable relying solely on one source to establish the savings and developed a multi-pronged strategy for the research, including the following components:

- a cross-check of regression model results against simple pre/post estimates
- modifications to standard OLS regression techniques used in the billing analysis
- a literature search to determine ranges of savings for the measures found in other impact evaluations and in engineering simulations, providing a reality check for the results of the billing analysis
- additional research to place our results in the context of the larger trends of energy use over the analysis period

In combination, these components provided a more comprehensive view of the program activity, highlighted particular problem areas and suggested methods for improving the quality of future evaluations.

Billing Analysis. We conducted a regression analysis on the billing data, defining the measures or groups of measures that may reasonably be expected to be identified through this type of approach, and compared overall results to a simple pre/post analysis at the household level. These comparisons were also made for the relatively large number of households that received either a refrigerator replacement or lighting replacements as a single measure.

The billing data available for impact evaluations such as this one are both cross-sectional across participating households and time-series in nature. As with prior impact evaluations of the program, we fitted regression models to monthly consumption data. Since our interest lay in measuring the change in consumption associated with participation in program, we used a fixed-effects analysis that removed overall differences in consumption from house to house.

The large number of households that participate in the program each year provide for large datasets for analysis. Even after attrition for missing or insufficient billing histories and other reasons, approximately 48,000 households that received electric measures and 44,000 households that received gas measures were available for the study.

Our analysis differed from prior impact evaluations of the program in two key ways. First, we were concerned that confidence intervals from standard calculations for fixed effects models might be optimistic due to ignoring the degree to which monthly consumption data are correlated within households. To address this issue, we used an empirical bootstrap approach to resample households in the data set repeatedly and recalculate the results (Efron and Tibshirani, 1993). This approach assumes that the data in hand are a reasonable approximation of the population, and that while households are independent from one another, the monthly consumption data within households are not. Second, we employed a comparison group to help remove general trends in consumption from the analysis. The comparison group was made up of households that participated in the program after the post-treatment analysis period for the 2002 participants. We matched the treatment and comparison groups on utility, housing type, and pre-participation energy use.

The relationships among the variables were reviewed for collinearity. Some groups of measures were bundled and the measure-level savings were disaggregated proportionally based on the deemed savings from the DEER report (Xenergy, 2001). In other cases, alternative models were defined to assess the impacts of collinear sets of variables.

Review of External Data Sources. The literature search provided us with a range of savings found in other studies for the major measures installed through the LIEE program. In our report, we compared the PY2002 results to the savings estimates to the range of savings estimates for the three previous LIEE evaluations, as well as those reported in other studies (West Hill Energy and Computing, 2005). We also discussed key inputs into the savings estimates and strategies for improving the reliability of estimates for the different measures on an on-going basis. Some examples of these issues and approaches are provided in Table 2.

Table 2. Example of Measure-level Issues and Approaches

Measure	Issues	Approach
Lighting	▪ Reduction in watts ▪ Retention ▪ Hours of use	▪ Record delta watts and location ▪ Follow up survey for retention ▪ Selective logging
Refrigerators	▪ Major driver of program savings	▪ Regression results are stable and consistent
Evaporative Coolers	▪ Proper use of equipment ▪ Interaction w/ AC	▪ Record condition of existing equipment in the tracking system ▪ Survey of participants to clarify usage patterns
Furnace Replacement/Repair	▪ Replacement of non-functioning units results in increased usage	▪ Record functionality, vintage and condition of furnace unit being repaired or replaced

Research into Consumption Trends. During 2001, shortages of electricity were common, resulting in rolling black outs in some areas, price increases, frequent public service announcements requesting California residents to reduce use, and offers by the utilities and directly from the governor to discount bills for measurable reductions in usage. After review of the billing data, we suspected that the impacts of the 2001 Energy Crisis could be affecting our ability to estimate savings. To investigate this issue further, we analyzed the billing data for general trends and also conducted research to identify independent studies that assessed the impact of the Crisis on electric and gas usage.

Results

As may be expected from the previous discussion, our success at teasing out savings from the billing analysis was highly variable. In general, the billing data provide reasonable and stable estimates of the program average per-home savings in electricity and natural gas. At the measure level, regression results were reasonably precise for measures with substantial savings and a high frequency of installation, such as refrigerators. For many smaller and less frequently installed measures, the precision was low and the estimators were not particularly reliable. The discussion below focuses on the household and measure savings, the estimates of statistical precision, and the research into overall energy consumption trends during the period.

Household and Measure Savings. Table 3 shows the household savings as estimated from the regression and the simple/pre post analysis incorporating a comparison group adjustment. Savings are also shown for two measures (refrigerators and CFLs) where these were the only installed measures (in addition to energy education) in a significant number of homes. Overall electricity and gas savings amount to about 7 percent and 2 percent of pre-participation consumption, respectively.

Variations between the regression and pre/post methods are likely due to differences in the structure of the analyses. While the regression analysis can more flexibly account for weather and other non-program effects than the simple pre/post approach, the two sets of estimates are reasonably consistent with one another.

Table 3. Comparison of Household Savings and Selected Measures

	Regression	Simple Pre/Post
Average Electricity Savings (kWh/year)		
Overall	323	355
Refrigerators	701	702[a]
Lighting (per bulb)	26	34[a]
Average Gas Savings (therms/year)	7.9	4.1

[a]Based on analysis of households that received only this measure.

For discussion purposes, we selected the two measures above that dominate the program impacts: refrigerator replacement and replacement of incandescent light bulbs with compact fluorescents (CFLs). These two measures also neatly illustrate the range in ability to extract meaningful estimates of measure-level savings from the available data.

Refrigerator replacement has a major impact on electricity use that is readily detectable in customer billing data, as indicated by the robust estimator derived from the regression analysis. The savings estimated from the regression and pre/post analyses are highly consistent, and the refrigerator savings are stable in all of the alternative regression models. These savings also fit well within the range of savings suggested by other research. The criterion for refrigerator replacement is that the existing unit must be more than 10 years old, and the LIEE savings falls neatly within a range of savings based on the difference in rated energy consumption between a 15- to 20-year old, 18 cubic foot refrigerator and a new ENERGY STAR labeled unit.

We also noted that the program average electric savings per home has gone up in lockstep with an increase in the installation rate for this measure, from 8 percent in PY 2000 to 30 percent in PY 2002. This further indicates that refrigerator replacement is a major driver of electricity savings from the program.

Table 4. Reported Refrigerator Replacement Savings

	Savings (Annual kWh)[a]
PY2002 LIEE evaluation	665 – 700
PY2001 LIEE evaluation	665 – 795
PY2000 LIEE evaluation	645 – 712
Estimated savings range for 15-20 year old models[b]	630 – 851

[a]Ranges for LIEE evaluations represent range in point estimates across utilities and housing types.
[b]Based on analysis of difference in shipment-weighted average rated energy use for 15- and 20-year old, 18 cubic foot, top freezer models compared to new ENERGY STAR labeled unit

In contrast, savings from lighting tend to be small and much less certain. Our estimates of CFL savings are reasonably consistent with previous evaluations, but we also found that these savings are somewhat unstable in that the estimates vary under alternative formulations of the model. The high installation rate of the measure makes it difficult to statistically disentangle CFL savings from other near-ubiquitous program effects, such as potential impacts from energy education efforts.

Table 5. Reported CFL Savings

	Savings (Annual kWh/bulb)[a]
PY2002 LIEE evaluation	21 – 43
PY2001 LIEE evaluation	16 – 24
PY2000 LIEE evaluation	22 – 29
Other Studies[b]	34 – 63
[a]Ranges for LIEE evaluations represent range in point estimates across utilities and housing types. [b] (Brown, 1994, Dahlhoff, 2003, Quantec, 2002, Xenergy, 1997)	

Lighting savings from all three previous LIEE evaluations (which employed similar methods) are also on the low end of reported values from other programs around the country. However, until some corroborative evidence emerges for why the savings from the LIEE program should be lower than those found in other evaluations, we are hesitant to conclude that the billing-analysis estimates are an accurate reflection of this pervasive, but low-impact measure.

We were unable to estimate savings from the billing analysis for some weather-sensitive measures, including all electric space heating-related measures. The gas model showed substantially smaller savings for many space and water heating measures than found in previous years. A number of the equipment replacement measures showed an increase in use, possibly related to participants whose existing equipment was nonfunctional prior to the installation of the efficient equipment. Savings from hot water conservation tended to be imprecise and, in some cases, the estimators were not statistically significant. Generally, these estimates were lower than found in previous LIEE impact evaluations and external data sources. This result could be related to the random error introduced into the explanatory variables due to uncertainty surrounding the fuel type of the water heating system in each home.

Bootstrapping Results. The results of the bootstrapping analysis confirmed that modeling uncertainties are greater than standard calculations would suggest: the bootstrap standard errors for key model coefficients related to savings from various measures ranged from about 1.75 to 3.5 times those obtained from the SAS GLM procedure, and averaged about 2.4 times larger. Table 6 compares the OLS and bootstrap standard errors for a few selected measures for one of the utilities (SCE). In some cases, the results of the OLS regression indicated the savings were statistically significant, but standard errors from the bootstrapping were large enough to negate that conclusion.

Table 6. Comparison of Bootstrap and Regression Standard Errors

	Regression Coefficient[1]	OLS Standard Error	Bootstrap Standard Error	Percent Increase
Refrigeration	666	11.7	21.0	80%
Lighting	21	3.1	6.0	92%
DHW Package	261	33.7	71.1	111%
Evap Cooler Maintenance	24	6.7	24.1	257%
A/C Replacement	145	13.8	43.3	215%

Consumption Trends. Our supporting research suggests that the billing analysis was complicated by the voluntary energy conservation efforts of California consumers during and after the 2001 Energy

1 The value of the regression coefficient is not always equal to the annualized kWh savings for the measure, due to the inclusion of weather-dependent terms in the model.

Crisis. An initial review of the trends in the billing analysis shows reduced use in 2001, slowly climbing back up through the analysis period. This trend is corroborated by independent sources demonstrating that average energy use (both gas and electricity) during the 2001 pre-period was significantly lower than in the previous years and began a rebound in 2002. (Bartholomew, 2002, Marks, 2003, Lutzenhiser, 2002, Ridge, 2004).

Although the comparison group should in theory eliminate this confounding influence, we remain concerned that residual uncontrolled effects may compromise our ability to estimate savings from the billing analysis, particularly for weather-sensitive measures. While we matched the two groups on utility, housing type (for SCE, SDG&E and SoCalGas) and approximate pre-participation energy consumption, we were not able to control for differences in climate region and end uses, fueling concerns that the comparison group may not be a good proxy for the treatment group.

Conclusions

This evaluation was confounded by numerous effects from within the program and from outside influences. While estimates of average overall savings per home appear to be reasonable, unstable estimators tend to be more the rule than the exception at the measure level. For this reason, we advise caution in any attempt to use the measure-level results to recalibrate the cost effectiveness of individual measures implemented by the program.

In this regard, the information gleaned from the external data sources was critical to the interpretation of the regression results and to the decisions regarding the appropriate application of these results. Some of the causes of uncertainty associated with the program itself can be mitigated through improved program tracking. Future volatility in the energy market is unpredictable, although one hopes the gyrations of the 2001 California Energy Crisis will not be repeated in the near future. Measuring energy impacts remains an inexact science and significant sources of error are inherent in all of the available methods.

The results of this evaluation point to the importance of supporting research results with independent sources and alternative analysis, as well as providing sufficient lead time to allow for a variety of methods to be considered. Engineering methods may indeed provide better estimates for some small measures, with sufficient opportunity to collect the necessary pre-installation data. Regardless of the selected methods, an evaluation plan based on transparency, the clear identification of sources of uncertainty, a realistic reflection of actual precision of estimates and placing results in context should guide the next round of impact evaluation and inform the set of strategies used to improve the savings estimates.

References

Bartholomew, E., R.Van Buskirk, and C. Marnay. 2002. *Conservation in California During the Summer of 2001*. Ernest Orlando Lawrence Berkeley National Laboratory.

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

Brown, Marilyn and Mihlmeister, P., 1994. *Summary of California DSM Impact Evaluation Studies*. Oak Ridge National Laboratory/Aspen Systems Corp.

California Public Utilities Commission. 2001. *California Statewide LIEE Policy and Procedures Manual.*

Dalhoff and Associates. 2003. *Report On Impacts And Costs Of The Iowa Low-Income Weatherization Program -- Calendar Year 2002*. Iowa Statewide Low-Income Collaborative.

Efron, Bradley, and Robert J. Tibshirani. 1993. *An Introduction to the Bootstrap*, Chapman and Hall, New York.

Lutzenhiser, L. 2002. *An Exploratory Analysis of Residential Electricity Conservation Survey and Billing Data: Southern California Edison*. Washington State University.

Kema-Xenergy. 2003. *Impact Evaluation of the 2001 Statewide Low-Income Energy Efficiency (LIEE) Program*. Southern California Edison Company, Southern California Gas Company, San Diego Gas & Electric Company, Pacific Gas and Electric Company.

Marks, M. 2003. California Energy Commission Workshop on Natural Gas Supply, Price and Infrastructure Assessment: Demand Presentation. http://www.energy.ca.gov/naturalgas/documents/2003-01-24_workshop/

Quantec, LLC. 2002. *Impact Evaluation of the 2001 Appliance Management Program*. National Grid USA.

Ridge & Associates. 2004. *Final Report for the Evaluation of the California 2002 Home Energy Efficiency Survey Program.* Southern California Edison Company.

Ridge, Richard. 1997. "Errors In Variables: A Close Encounter of the Third Kind." 1997 International Energy Program Evaluation Conference, Chicago, IL.

TecMarket Works, Megdal & Associates, Architectural Energy Corporation, RLW Analytics, Resource Insight, B & B Resources, Ken Keating and Associates, Ed Vine and Associates, American Council for an Energy Efficient Economy, Ralph Prahl and Associates, Innovologie. 2004. *The California Evaluation Framework.* Project Number: K2033910. Prepared for the California Public Utilities Commission and the Project Advisory Group. Draft, February, 2004.

West Hill Energy and Computing, Inc., et. al. 2005. *Impact Evaluation of the 2002 California Low Income Energy Efficiency Program.* Prepared for Southern California Edison, Pacific Gas and Electric, San Diego Gas and Electric and Southern California Gas. 2005. Final draft report available at http://www/liob.org, under "Public Meetings."

Xenergy, Inc. 1997. *Impact Evaluation of the 1995 Residential Direct Assistance Program*. Southern California Edison.

Xenergy, Inc. 2002. *Volume 1: Impact Evaluation of the 2000 Statewide Low-Income Energy Efficiency (LIEE) Program.*

Xenergy, Inc.2001. *2001 DEER Update Study.*