# Replicability of Commercial Lighting Program Impact Evaluations

**Allan A. Zebedee, University of California at San Diego**
**Marvin J. Horowitz and Andy Parece, XENERGY, Inc.**
**Dinesh Bhagani, Northeast Utilities**

Northeast Utilities Energy Saver Lighting Rebate (ESLR) Program is a commercial, industrial and multifamily building lighting program that has been the subject of four different evaluations over the past two years that have estimated net program-related annual energy savings. The findings of these studies have been relatively consistent suggesting, among other things, that the methods employed in these studies are replicable across years. These results tend to reinforce the idea that impact evaluation techniques and methods can become increasingly standardized and cost-effective for most commercial and industrial program. The focus of this paper is on describing the findings of these studies and the technical details of the research designs, data collection activities and statistical methods.

For the program years 1992 and 1993 both billing analyses and verification studies were completed for the ESLR Program. The billing analyses research designs followed the energy impact model method in which the variation in the annual changes in participant whole account energy consumption from the pre-installation to the post-installation period was explained as a function of program participation and customer characteristics. To control for non-program related changes in annual energy consumption a comparison group of nonparticipating customers were included in the research design. The verification studies followed the net-to-gross impact evaluation method in which on-site inspections, lighting loggers and self-reported free ridership data were used to enhance and revise customer specific engineering estimates of annual energy savings.

## INTRODUCTION

This paper compares the results and techniques used in four different impact evaluations conducted over the past two years of Northeast Utilities Energy Saver Lighting Rebate (ESLR) Program. The ESLR Program is designed to encourage commercial, industrial and multifamily customers to replace existing lighting with energy-efficient alternatives. The qualifying energy-efficient lighting alternatives include:

- T8 fluorescent systems and hybrid or electronic ballasts
- compact fluorescent and high intensity discharge lighting
- specular reflectors and parabolic fixtures
- occupancy sensors and exit sign fixtures.

The program provides cash rebates or account credits for the purchase and installation of these energy-efficient lighting products. The eligible market includes a range of customers from very small commercial customers to very large industrial plants. Since inception of the program in 1986, Northeast Utilities has received and processed more than 22,000 rebate applications.

The objective of the evaluations was to assess net program energy savings attributable to the ESLR Program. In deriving net program energy savings two distinctly different methodologies were used — a billing analysis and verification analysis. The billing analysis compares metered usage data prior to and after the installation of program measures. The analysis of program participants is complemented by a parallel analysis of a random sample of nonparticipating customers. Nonparticipants are matched to program participants by facility type, energy consumption level and state. In addition to the billing data, customer survey data are used to control for variations in customer characteristics and building features.

The verification analysis compares the energy use behavior of program participants and nonparticipants based on data collected from on-site inspections. The on-site inspections verify that program measures are installed properly in the reported building locations and that the measures have remained in place. Also, a representative sample of different types of lighting equipment in different types of building spaces was monitored so that the hours of use of the program measures could be verified. Participant and nonparticipant samples were drawn from those customers who were previously interviewed by telephone to support the billing analysis. The key objective of the nonparticipant on-site visits was to assess potential program spillover. So as not to give an unfair advantage to one evaluation technique over another, the sample sizes of both the billing analyses and verification analyze were designed so that the expected statistical precision of the program-related energy impacts for each study was $\pm 20$ percent at the 90 percent confidence level.

The second section of this paper describes the technical details of the statistical methods used for the billing analysis and the third section describes the verification analysis. The fourth section of this paper presents the 1992 and 1993 results of the billing and verification analyses and the fifth and final section offers observations and conclusions.

## BILLING ANALYSIS DATA PREPARATION AND METHODOLOGY

The primary purpose of the billing analysis is to estimate the reduction in energy consumption that have resulted from participation in the ESLR Program. To do so, the first steps in the billing analysis research plan follow a framework in which a customer s annualized and weather/seasonally-adjusted consumption, calculated by aggregation and analysis of monthly or periodic billing data, is compared before and after the installation of program sponsored energy efficiency measures. To strengthen the research design, the energy consumption of a comparison group of nonparticipating customers is also analyzed.

To control for the wide range of annual energy consumption among commercial customers, and to control for the heterogeneity of building types, the program study groups were stratified by energy consumption levels and building types. An addition, an attempt was made to match the participant and nonparticipant samples by the distribution by state of the program population.

Before billing data can be used to study impacts, the data must be screened for reasonableness and accuracy. In general, where the sources of raw data irregularities are easily found and corrected, the observation remains in the sample. However, if an irregularity casts doubt on the reliability of the data for an individual customer, or results in the loss of more than one meter reading, the observation is dropped from the study group. In addition, the participant and nonparticipants samples were screened for participation in other program during the study period.

Prior to statistically analyzing the billing data for program participants and nonparticipants, a stratification plan was developed to reduce to manageable size the great variance that is found in energy consumption in commercial and industrial sector facilities. For this evaluation, the annual level of energy consumption in the pre-installation period was used as the primary stratification variable. Once the strata are determined, the associated levels of energy consumption that form the minimum and maximum boundaries of each strata were found. To stratify the nonparticipant sample, these same minimum and maximum values were applied.

In addition to stratifying by annual energy consumption in the hope of minimizing the variance in the estimate of energy savings, sub-strata were also developed based on 12 building types. However, customers could not be placed in these sub-strata until after the participant survey was completed. Thus, the nonparticipant survey was not fielded until after the participant survey was completed and frequencies of building types within each strata were available. The purpose of this level of stratification was to assure that the sample of nonparticipants contained roughly the same mix of building types as the participant group. In keeping with the experimental design, this matching helps assure that qualitative differences among the observations are not overlooked. Finally, to assure that the nonparticipant sample matched the geographic location of the participant sample, strata were developed by state. This resulted in a total of 72 sampling cells for the sampling plan, i.e. three energy use levels by 12 building types by two states. Not all of these cells had sample points, and others had very small samples.

After screening and stratifying, the billing data, weather-adjusted and seasonally-adjusted annual consumption is estimated for each of the study participants and nonparticipants. Weather adjustment and annualization is necessary to ensure that the comparisons of electricity use between years, and between participant and nonparticipant groups, span the same number of days and the same weather conditions. In addition, weather-adjustment requires forecasting annual electricity consumption using long-run, average temperatures. This allows the results of the analysis to represent the savings expected to be achieved over a longer time period, one that is closer to the lives of the installed measures than the one or two years that comprise the study s post-installation period.

It is important to note that weather-adjustment may not be appropriate for all buildings. To empirically determine which buildings' energy use should be weather-adjusted, energy use was systematically analyzed with two separate models, the first designed to detect ''quarterly'' energy use sensitivity, and the second designed to detect ''temperature'' sensitivity. For each building in the sample, two models are estimated using pre-installation billing data. The first model is:

Q Model: $KWH_t = \alpha_1 + \alpha_2(Q1) + \alpha_3(Q2) + \alpha_4(Q3) + \varepsilon_t$

where:

$KWH_t$ = raw energy consumption in period t;

$Q1$ = an indicator variable for the first quarter of the year (as close as possible as the billing dates can come to January, February and March);

$Q2$ = an indicator variable for the second quarter of the year;

$Q3$ = an indicator variable for the third quarter of the year; and,

$\varepsilon_t$ = a random error term.

In this ordinary least squares (OLS) model, $\alpha_1$ is a constant (the regression intercept) that is interpretable as monthly energy use for the fourth quarter; and $\alpha_2$, $\alpha_3$ and $\alpha_4$ are coefficients providing estimates of monthly use in each of the first three quarters of the year. The model error term is the variation in kWh per period that remains unaccounted for by the model's estimates. This model is designed to indicate a farm's sensitivity to economic or calendar year cycles that are largely independent of outdoor temperature patterns because calendar year quarters are not identical to weather seasons.

The functional form of the second, temperature-related model that was estimated for each participant and nonparticipant is:

T Model: $KWH_t = \beta_1 + \beta_2(HDH_t^{65}) + \beta_3(CDH_t^{72}) + \varepsilon_t$

where:

$HDH_t^{65}$ = total heating degree hours calculated at a reference temperature of 65 degrees Fahrenheit in period t; and,

$CDH_t^{72}$ = total cooling degree hours calculated at a reference temperature of 72 degrees Fahrenheit in period t.

In this OLS model, $\beta_1$ is a constant (the regression intercept) that is interpretable as non-weather sensitive load per period; and $\beta_2$ and $\beta_3$ are coefficients providing estimates of how a marginal change in temperature affects electricity consumption.

After estimating these models, the statistical significance for each coefficient is calculated. Variables are selected based on whether or not their t-values are greater than 1.282 (the 80 percent level of confidence). Those variables that pass the threshold t-value screen are combined into a new model, called the *quarterly-temperature* (QT) model. For example, if $a_2$ and $b_2$ were significant at the threshold level, then the QT model for this building is:

QT Model: $KWH_t = \chi_1 + \chi_2(Q1) + \chi_3(HDH_t^{65}) + \varepsilon_t$

Once the QT model is settled on for a given account for the pre-installation period, no further model search is performed for the post-installation period billing data. Weather-adjusted and/or annualized consumption is forecast for each customer, for each period, based on their individual models. Where weather-adjustment is called for, annual consumption in the pre- and post-installation ( *PRE_KWH* and *POST_KWH*) periods for each customer is estimated using long-run heating and/or cooling hours per day ($LRHDH^{65}$ and $LRCDH^{72}$). If none of the temperature-related coefficients are significant, the billing data are annualized by pro-rating the aggregated monthly data to 365 days.

The formula for calculating weather-adjusted and seasonally-adjusted annual consumption in the pre-installation period for customer i where both Q1 and HDH are involved is:

$$PRE\_KWH_i = \phi_1 * 365 + \phi_2 * 90 + \phi_3(LRHDH^{65} * 365)$$

Similarly the formula for calculating post-installation, weather-adjusted and seasonally-adjusted annual consumption for customer i is:

$$POST\_KWH_i = \varphi_1 * 365 + \varphi_2 * 90 + \varphi_3(LRHDH^{65} * 365)$$

Once the customer annual, weather-adjusted energy consumption are prepared, the energy impact model can be specified and estimated. The aim of the energy impact model is to use a multiple linear regression model to control for key variables that differentially affect the changes in electricity consumption for participants and nonparticipants. This model is cross-sectional and is used to derive an empirical energy savings ''realization rate'' that represents *the fraction of the program tracking system estimate of energy savings that can be confirmed by the change in energy bills.* As the regression coefficient that represents this rate incorporates information on the change in nonparticipant energy bills, the realization rate is more than an estimate of gross participant realized savings — it is akin to an estimate of program-related or net realized savings.

The dependent variable of the energy impact model is post-installation energy consumption. Using pre-installation energy consumption as one of the independent variables effectively converts the model into a *change model* wherein the coefficients of all the remaining independent variables are interpretable as *the marginal change from the pre-installation to the post-installation period in annual total energy use per participant.*

The independent variable of interest in the energy impact model is a continuous variable that is calculated in the course of program implementation. It is an engineering estimate of the expected changes in energy use from installing specific energy efficiency lighting measures in a customer's facility. This variable, referred to as *IMPACT*, takes a value of zero for nonparticipants, since the program has not had any direct affect on the changes in lighting in these facilities. As the objective of the energy impact model is to compare the change in customer bills with the engineering estimate of the energy savings associated with the installation of program-subsidized energy efficiency lighting equipment at participant sites the model is designed to provide an estimate of how a marginal (1 kWh) change in expected savings based upon the program tracking database is related to an actual change in energy consumption. A realization rate of 1.00 means that a 1 kWh change in tracking system savings is related to 1 kWh in program-related energy savings; a rate of between 0.00 and 1.00 implies that tracking system estimates overstate program savings; and a realization rate of greater than 1.00 implies that tracking system estimates understate program savings. As such, the realization rate provides information about how well the program tracking system does in estimating energy savings. The general specification of the model is:

$$POST\_KWH_i = \beta_1 + \beta_2(PRE\_KWH_i) + \beta_3(X_i) + \beta_4(D\_FACILITY_i) + \beta_5(SSC_i) + \beta_6(IMPACT_i) + \varepsilon_i$$

where:

| | | |
|---|---|---|
| $X_i$ | = | a vector of variables related to customer characteristics affecting changes in energy use for customer i; |
| $D\_FACILITY_i$ | = | a vector of variables related to changes to a facility that can affect changes in energy use for customer i; |
| $IMPACT_i$ | = | the estimate of gross annual energy savings for measures installed by the program for customer i; for nonparticipant customer i this variable takes a value of 0; |
| $SSC_i$ | = | participation self-selection correction term for customer i; and, |
| $\varepsilon_t$ | = | model error term. |

Of special interest in this model is the self-selection correction term. This term is estimated for each customer in the energy impact model by way of a discrete choice (logit) model. Using telephone survey data that provides information on customer characteristics, the discrete choice model provides an estimate of the probability of participation. After converting these probabilities into the appropriate functional form, the self-selection term is included in the energy impact model to control for the unobserved propensity of certain customers to participate in the program. This term is necessary to ensure that the energy impact model is unbiased; if the propensity is related to changes in energy use and is not included in the energy impact model the model will suffer from omitted variable bias.

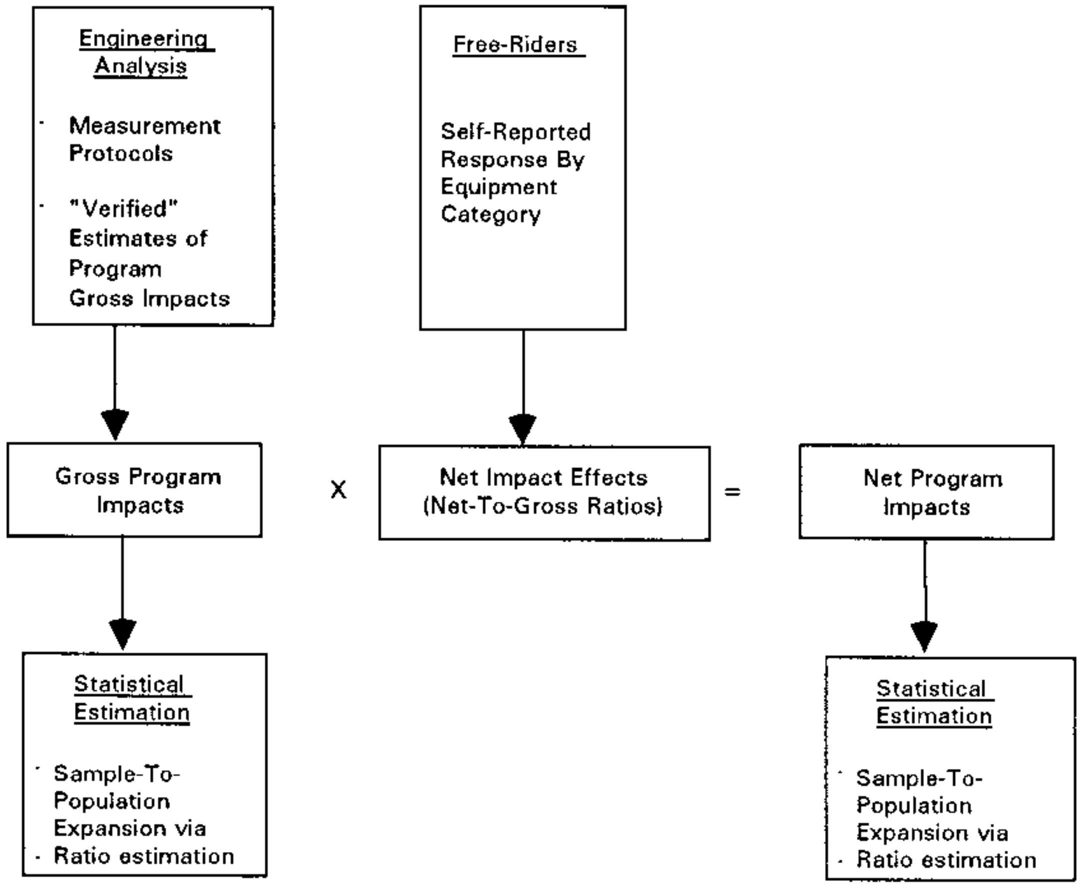
## VERIFICATION ANALYSIS DATA COLLECTION AND METHODOLOGY

The verification analysis is comprised of two interrelated tasks. As illustrated in Figure 1, improved *gross* program impacts are derived through expansion of measurement data for a sample of equipment installed through the program. The methodology for developing estimates of improved gross measure savings is based on engineering algorithms supplemented by data from on-site inspections and short-term monitoring. Second, estimates of free-ridership are applied to the estimates of improved gross savings to arrive at *net* program impacts which are then expanded to the program population level.

To provide accurate and unbiased estimates of energy savings, the components of the savings algorithms were observed and measured for a large sample of program-rebated equipment. On-site surveys were conducted for a sample of program participants — during the visits, operating hours monitoring equipment was installed on a sample of lighting fixtures. Based on the on-site inspections the behavioral assumptions and measure retention assumptions embedded in the engineering algorithms were revised. In addition to improving the estimates of gross savings, the on site inspections also provided valuable information on the reasons for discrepancies between expected and net savings. For example, the on site data collection effort was used to quantify the effects of such factors as:

- equipment disposition (e.g. operating, failed, or in stock)
- lighting profile (average daily hours of use, by month)
- space conditioning (heated or cooled spaces)
- baseline efficiency assumptions (for appliances and equipment).

***Figure 1.*** *Summary of the Verification Analysis Methodology*




In total, 52 of the 1992 program participants and 66 of the 1993 program participant sites were selected for inspection. For each inspected site, all measures covered under the program were inspected, and the operating characteristics listed above were observed and recorded. Monitoring of operating hours was performed using lighting loggers for all inspected sites except those for which exit signs were the only measure rebated. These detailed measurement and observed data were used to re-estimate the savings attributable to the program thus accounting for many factors not available through the program tracking data alone.

Ratio estimation was employed for expanding the sample data for computing program-wide savings. The study samples were drawn separately for each program year and were stratified by whether participants indicated that they were freeriders with respect to any measure installed through the program. This stratification was possible due to the nesting of the verification samples with the telephone survey samples used in the billing analysis.

The verified or improved gross savings estimate, when stated relative to the initial gross savings estimate for all surveyed sites in the tracking system, results in a *realization rate.* The realization rate is applied to the total program tracking estimates to calculate improved gross program savings at the population level.

Lastly, estimates of the improved gross program impacts derived through the engineering analysis are adjusted to reflect free ridership, that is, measure installations that would have occurred had program assistance not been not provided. Program free ridership was estimated by asking program participants a series of questions regarding what they would have installed had the program not existed, along with the timing and extent of these planned actions for each relevant measure category at a particular site. These data were applied, by customer and equipment category, to obtain an estimate to the net impact of the program for each sampled customer. The results are applied to the improved gross impacts of the sample respondents separately by measure category and then weighted to the program level to derive total net impacts for the program population.

## STUDY FINDINGS

As two years of post-installation billing data are available for the 1992 participants, the energy impact model is specified as a partial-panel model The term *partial* is used because for some customers only one post-installation period is available. This approach, which assumes that the underlying behavioral parameters of the model are relatively constant across years, provides estimates that are less volatile than the estimates that might result from individual models for different years. This is because by incorporating added information, the panel billing analysis coefficient estimates are less prone to be affected by non-program related factors that have temporary influences on changes in energy use. Using this specification, the energy impact model for 1992 program year participants represents the empirically confirmed program realization rate over a two year post-installation period.

The variables contained in the energy impact model are briefly described in Table 1. As described above, the dependent variable of the energy impact model is post-installation energy consumption. With pre-installation energy consumption and a selectivity correction term as independent variables there are 11 independent variables in the energy impact model. Five of the 11 independent variables in the study are related to changes in building conditions or operating schedules over the study period and a sixth represents the presence of supplemental heat in the building. Another explanatory variable is a dummy variable representing spillover, that is, nonparticipants who reportedly were influenced by their knowledge of the ESLR Program to invest in energy efficient lighting.

Table 2 displays the estimated findings for the energy impact models. The major finding of the 1992 program years model is that the estimated net program realization rate is 71.0 percent. The relative precision of this estimate is ±35 percent at the 90 percent confidence level, implying that if the true program realization rate was known there is a 90 percent probability it would be between 46 percent and 96 percent. Some of the other findings of the model are that building remodeling is statistically significantly related to increased energy consumption and that a supplemental electric heat is significantly related to a decrease in energy consumption. The coefficients of the remaining independent variables in the energy impact model are not statistically significant. However all variables incorporated into the model are of a reasonable magnitude and conform with expectations in terms of the direction of their correlation with the change in energy consumption.

***Table 1.*** *Independent Variables in the Energy Impact Model*

| Label | Definition |
|---|---|
| PRE_KWH | Dependent variable representing pre-installation weather-adjusted annual energy consumption |
| REMODEL | Indicator variable (0/1)* for remodeling times building square feet |
| D_ACTIVE | Indicator variable (0/1)* for change in building activities times building square feet |
| D_SQFT | Change in building square feet |
| D_COOL | Variable representing changes in cooling system times building square feet |
| D_VAC | Variable representing changes in building vacancy rate |
| D_OPP | Change in annual hours of operation |
| SUP_ELEC | Indicator variable (0/1)* for supplemental electric heating times building square feet |
| NP_INF | Indicator variable (0/1)* if nonparticipant was influenced by ESLR to install energy efficient lighting |
| SSC | Selectivity correction term |
| IMPACT | Engineering estimate from NU program tracking system of a participant's expected savings for all measures |

*A coding of (0/1) indicates that it is an indicator variable with a value of 1 representing an affirmative response.

The major finding of the 1993 program year model is that the estimated net program realization rate is 92 percent. The relative precision of this estimate is ±44 percent at the 90 percent confidence level, implying that if the true program realization rate was known there is a 90 percent probability it would be between 52 percent and 133 percent. Other findings of the model are that a change in building s function

*Table 2. Energy Impact Model: Dependent Variable = POST_KWH*

| Independent Variable | 1992 Program Year Coefficient | 1992 Program Year t-statistic | 1993 Program Year Coefficient | 1993 Program Year t-statistic |
|---|---|---|---|---|
| INTERCEPT | 5,192.37 | 1.31 | − 1,253.70 | 0.37 |
| PRE_KWH | 1.04 | 241.54 | 1.04 | 237.12 |
| REMODEL | 0.65 | 3.10 | 0.015 | 0.14 |
| D_ACTIVE | − 0.27 | 0.67 | 1.73 | 3.24 |
| D_SQFT | 0.44 | 0.17 | − 0.53 | 0.18 |
| D_VAC | − 3.65 | 1.09 | − 0.47 | 1.02 |
| D_COOL | 0.83 | 0.90 | − 12.28 | 1.48 |
| D_OPP | 39.34 | 1.63 | 20.70 | 1.78 |
| SUP_ELEC | − 1.20 | 3.86 | − 0.51 | 1.52 |
| NP_INF | − 195,214 | 3.99 | − 219,619 | 5.10 |
| SSC | 7,708.31 | 2.30 | − 100.99 | 0.04 |
| IMPACT | − 0.71 | 4.64 | − 0.92 | 3.74 |
| n | 781 | | 451 | |
| Dep. Variable Mean | 258,427 | | 220,890 | |
| Root Mean Sq. Error | 69,039 | | 43,024 | |
| Adjusted R-Square | 0.99 | | 0.99 | |

is statistically significantly related to increased energy consumption while nonparticipants who were influenced by the program to install energy efficient lighting on average decreased consumption. The coefficients of the remaining independent variables in the energy impact model are not statistically significant. However all variables incorporated into the model are of a reasonable magnitude and conform with expectations in terms of the direction of their correlation with the change in energy consumption.

The estimated program annual energy savings impacts from the verification analysis for the 1992 and 1993 program years are summarized in Table 3. These impacts were calculated based on an engineering analysis and statistical expansion of data collected through the on-site inspections of participant facilities and self-reported free rider information. In this analysis, the difference between the improved gross impacts and the net impacts is the inclusion of free rider estimates on a measure-by-measure level. Table 4 summarizes the estimated net energy realization rates by program year and methodology.

As the full set of findings indicate, the net realization rates are in reasonable conformance with each other, the only surprise being the billing analysis realization rate for the 1993 program year. This atypically high rate is likely due to abnormally high annual energy use levels in 1992, the pre-installation year for the research design for the 1993 program year. At this point, all that is known about 1992 that makes it different from the other years was that it sustained an much colder winter season, and a much colder summer season, than the average year. However, further

**Table 3.** *ESLR Verification Analysis Impact Estimates*

| Program Year | N | Improved Gross Impacts Realization Rate | Improved Gross Impacts Relative Precision | Net Impacts Realization Rate | Net Impacts Relative Precision |
|---|---|---|---|---|---|
| 1992 | 959 | 88.1% | 21% | 78.6% | 23% |
| 1993 | 744 | 100.8% | 18% | 70.3% | 31% |

**Table 4.** *ESLR Estimated Energy Impacts by Year and Methodology*

| Program Year | Billing Analysis Realization Rate | Billing Analysis Relative Precision | Verification Analysis Realization Rate | Verification Analysis Relative Precision |
|---|---|---|---|---|
| 1992 | 71% | 35% | 79% | 23% |
| 1993 | 92% | 44% | 70% | 31% |

investigation is needed to determine why this realization rate is as high as it is. In any event, at the 95 percent probability level there are no statistically significant differences between any of the four realization rates.

## CONCLUSION

There are a number of observations and conclusions that can be drawn from these four studies. Each pair of studies uses methods that come at the problem of estimating net, program-related savings from entirely different perspectives. To derive net program savings the billing analysis approach must estimate the change in energy use due to the program measures by successfully extracting it from the change in whole account energy use. On the other hand, the verification analysis relies on modifying the basic engineering equations of gross savings for each installed measure by focusing on major physical or behavioral parameters whose initial values may be biased. At its most comprehensive, the verification analysis approach would estimate net energy savings by reexamining and improving on each and every component of energy savings.

Unfortunately, because perfect studies are unattainable using either approach it is very difficult to pinpoint the flaws or shortcomings in one study by referring to the findings of the other. Rather, lessons must be learned through inference. With respect to the four studies, what may be inferred is that the two techniques appear to yield relatively consistent results across program years and also appear to yield relatively consistent results between each other. Assuming the program delivery, the program measures and the basic participant population has not changed very much across years, it would seem that both these techniques are either doing a good job in estimating program impacts, or, are systematically biased in ways that we have yet to understand. Of added interest, a similar billing analysis of the 1991 ESLR program and two verification studies of much smaller scope than the present ones, one of the 1990 and one of the 1991 program year, also arrived at similar point estimates of net realization rates. All in all, it is our belief that these studies are doing a good job of providing reliable estimates of program savings.

Using standardized approaches to program evaluation thus appears, at least in this case, to yield beneficial results. The approaches are proven to be replicable and the impact estimates themselves are shown to be consistent with follow-up studies and with complementary studies. In addition, multiple use of the same methods minimizes the amount of resources that go into reinventing the wheel and then having to explain its working to a puzzled audience.