Successful Techniques for Identifying, Measuring, And Attributing Causality in Efficiency and Transformation Programs

Lisa A. Skumatz, Skumatz Economic Research Associates, Inc.
Daniel Violette, Summit Blue Consulting, LLC
Rose A. Woods, Skumatz Economic Research Associates, Inc.

ABSTRACT

In the course of conducting recent evaluation and causality work on residential and commercial programs, the authors employed several useful steps to improve the attribution of impacts to program interventions. We focused on providing a burden of proof on par with the criteria for other public and private investments. As with many recent evaluations, the work included a theory based evaluation, but also allowed for goal free / alternative explanations for program progress. However, we undertook four key steps to improve the quality of the attribution estimates:

- We examined an array of causality modeling options available from the literature – some of which have not been used often in the energy field, but which show promise for application to measurement and assignment of program impacts.
- We used the distribution of the measure and intervention impacts, rather than relying on point estimates. This used much more of the information gathered during the data collection, and more fully reflects the range of impacts induced by the program.
- An enhanced method was used to address free riders, spillover, and free drivers – a method that allows for partial free riders and uses indicator methods to produce evidence on program-induced effects on spillover and free drivers.
- The attribution approach can also address the issue of uncertainty and risk in the attribution work, incorporating scenario analysis, decision tree, and other techniques to "bound" the effects, and options like a Bayesian approach are being explored to address "degree of belief" in the impacts. Thus, the attribution work is applying and examining several methods of addressing risk – an important component of using causal results.

These extra steps improved the reliability and robustness of the results of the causality analysis and provided a better foundation to guide program and investment decisions – an important goal of an evaluation. The paper highlights benefits and impact of these approaches.

Introduction

In association with recent projects to evaluate residential and commercial programs, the authors were tasked with estimating the market effects and their attribution to program efforts. The programs that were assessed included commercial new construction, ENERGY STAR® products and marketing, and ENERGY STAR® new homes among others.

As part of this work, we explored options for modeling and assigning impacts. We also determined to use more robust approaches to examining free riders / naturally occurring adoption, spillover / market effects, and attribution – relying not just on point estimates, but specifically incorporating distributions and uncertainties to provide a better foundation for guiding program-related decision making. The first section below addresses promising modeling alternatives; the later section describes the approach taken for attributing program effects.

Brief Summary of Measurement Options

There are a variety of techniques that have been used or can be adapted to causality and attribution work. This section provides a summary of several of these methods.

Granger Causality

This technique involves using time series data to estimate the effect of lagged variables of a series of data on the current (time period) of another variable. Specifically if we are interested in whether series X causes series Y, the Granger Causality method is used to test the joint significance of the coefficients for the lagged values of X when regressed on Y, controlling for (including) lagged values of Y. For example, to test whether increased advertising spending causes increased sales, we would set up the model as:

$$Sales(t)=b0+b1*Sales(t-1)+b2*Sales(t-2)+b3*Sales(t-3)+a1*Advertising(t-1)$$
$$+a2*Advertising(t-2)+a3*Advertising(t-3).$$

Then the joint hypothesis that a1 & a2 & a3 are zero is tested. Regression analysis can be run on the restricted model (a1, a2, a3 = 0) as well and then the sum of squared residuals for both regressions are compared to get our test statistic and then accept or reject at some level.

Treatment Effects and "Propensity Scoring"

This is an enhanced version of "control group" approaches. Treatment Effects involve estimating the difference in outcomes for a unit subject to a treatment and the outcome when (had) the same unit is (been) exposed to the control. Specifically, let Yi0 = outcome for unit i in the control group,

Yi1 = outcome for unit i in the treatment group.

The treatment effect is

$$\Delta T = E(Yi1/Ti=1) – E(Yi0/Ti=1) = E(Yi/Ti=1)-E(Yi0/Ti=0)$$

This is true if the outcomes are independent of the whether the unit is assigned to the treatment group or the control group. Rubin (1977) notes that if, for each unit, we observe a vector of covariates Xi and if the population treatment effect is identified for the treated, then the treatment effect is equal to the treatment effect conditional on covariates and on assignment to treatment. We can them estimate the treatment effect as:

$$\Delta T= E(Yi1/Xi,Ti=1) – E(Yi0/X,Ti=0)$$

The estimation is implemented by matching outcomes for treatment and comparison units that are similar in terms of their observable characteristics. This is called "matching on observables" estimation. Rosenbaum and Rubin (1983) suggest using the propensity score method (PSM) to match observations. The propensity score is the probability of receiving treatment conditional on covariates. One method of matching is the nearest neighbor matching.

Each treated unit (i) is associated with m comparison units which have the closest propensity scores. For many studies, m=1.

The propensity is estimated for the data and then the data are matched (1 to m) from the treatment and control group. Each observation in the treatment group has a match whose propensity score is closest.

In most of the literature, authors then compare the means of the variables of interest. For example, Nieswiadomy and Lee (2003) examine the means of water usage for recipients and non-recipients (matched using PSM) of a Xeriscape newsletter to evaluate the effectiveness of the newsletter.

"New Framework…"

This study represents a compilation of and in some cases adaptations / enhancements of a variety of techniques that have been used for some time in the literature. This study (Sebold, *et. al*. 2001) examined the methods for evaluating market effects and the market dynamics that occur during the duration of the programs and subsequent periods.

- Statistical methods to estimation adoption and net to gross. The report discusses methods that have been used for estimating adoption impacts and adjusting adoptions by participants for free ridership and free drivership. Approaches for estimating the impacts of intervention-induced adoptions on energy and demand are addressed, including: engineering approaches (simulations, regressions, and other statistical approaches), statistical analyses of billing, and hybrid approaches. The report highlights methods for estimating baselines (what would happen in the absence of the market intervention) using simple trends, s-curves, and other adoption patterns. Pre- and post- intervention surveys are discussed for assessing market changes and measure adoption.
- Methods to address dynamic effects. Methods to address dynamic baselines and market effects – including those that extend beyond the program period – are also addressed in the report. Time series forecasting techniques are used in estimating baselines. Evaluating the market effects (adoption) over time involves running simulations on awareness, willingness, and availability. The levels of awareness, willingness, and availability at time t are assumed to be functions of lagged values of awareness, other variables, and interventions at time t. The parameters of the model can be used directly to develop estimates of program-derived market effects, or the results can be refined using Delphi methods, or used as inputs to additional analyses.

Ordered Logit

The ordered logit approach is a qualitative choice model that can be used to estimate current and projected market shares, changes in market share due to alternative program interventions, or applied to other questions in program-related causality. The analysis uses information derived from the rankings of a set of options (usually supplied on cards) by survey respondents. The market actor / decision-maker respondents rank a set of technology alternatives that are described in terms of alternate values for a set of attributes, program interventions, or other considerations affecting technology choice. To maximize the analytical use of the data, the sets of cards / options are developed using orthogonal "Latin Square" or similar experimental designs. Parameters are estimated using maximum likelihood techniques, and the estimates are used to predict relative market shares – using baseline settings as well as alternate values for the technologies and interventions. Changes in market share that are due to the particular interventions can be analyzed to examine attributable impacts. Examples are provided in Skumatz and Weitzel (2001).

Delphi Technique

A Delphi technique uses an iterative judgment technique to develop quantitative estimates of parameters of interest. According to Martino (1993), Delphi approaches require / maintain three key features: iteration with controlled feedback; anonymity; and statistical representation of a group response. For example, applying Delphi techniques to obtain judgmental results for forecasting market penetration requires asking experts to answer questions regarding the adoption of technologies. Using a Delphi technique to develop estimates of market share, progress in market assessment indicators, or causality requires employing a panel of experts and surveying them on the key topics in interest. Then, the numerical results and supporting arguments are collected and summarized (e.g. using statistical summaries with means, quartiles, etc.). The results are then presented to the panel and they are re-surveyed, allowing them a chance to consider arguments from others and where in the distribution their initial estimates fall. Those outside of the middle range are encouraged to provide rationales for their information. The re-survey is then summarized as the results from the Delphi. The number of re-iterations can vary; however, Schuster (1985) found that most change occurred between the first and second rounds, and that a fifth round did not provide significant contributions. Delphi provides advantages over standard group voting including the fact that it provides a majority opinion and also a distribution or range of uncertainty.

Fishbein Causality

Ajzen and Fishbein (1980) model “reasoned actions” in explaining behavioral theories. According to the theory, intention to act has a direct effect on behavior, and can be predicted by attitude. The attitude is formed by subjective norms and belief, and the importance of those variables is decided by situational factors (advertising, marketing efforts, etc.). In short, the work looks at progress along a chain of awareness, intention to purchase / adopt behavior, and purchase / adopt behavior (with variations in intermediate step labels). Two areas of recent interest in the environmental field include social-based marketing and self-efficacy. Community-based social marketing (CBSM) emphasizes factors beyond economic best interests into elements of culture, social interactions, and human feelings and their role achieving modifications in (environmental) behavior. Self-efficacy argues that internal factors and underlying beliefs are strong components of behavioral change – specifically the perceived ability of an individual to effect change. This research has been most actively used in outreach and program design applications.

Hwang, Kim and Jeng (2000) attempt to examine causality and test a model with selected antecedents of responsible environmental behavior. They surveyed 523 visitors of a forest trail in Korea and asked questions regarding “knowledge about the issue”, “loss of control”, “attitude”, “personal responsibility”, and “intention to act”. They evaluate the influence of the different questions on “intention to act”. Woods and Skumatz (2004), Skumatz, et.al. (2003), Peters and Feldman (2001) and others evaluate the role of self-efficacy indicators in modifying behavior, including analyses of "green" energy conservation programs. There is also a growing literature in social-based marketing (McKenzie-Mohr 1995 and others) there is only limited analysis and evaluation of the work.

Meta Analysis

Essentially, Meta analysis refers to the analysis of analyses; that is, it encompasses the statistical analysis of a large collection of analysis results from individual studies for the purpose of integrating the findings. It is a rigorous alternative to the narrative discussion of research studies. Studies are collected, coded and interpreted using methods similar to those in primary analysis. Typically the steps for a Meta Analysis include: defining the Meta-Analysis, selecting the studies, identifying the different types of models, calculating summary effects, and interpreting results. Studies are selected based on a set of quality criteria, and a "score" is associated with each study. The analysis is based on the quality of the scores of the studies. While some studies have used fairly simplified approaches to combine the data, a literature exists outlining the appropriate statistical treatment of combined data from these studies, which improves the analytical robustness and reliability of the results.

Overall Issues / Approach

These causality approaches can be applied to analyzing / attributing observed changes in key program indicators. These might include: purchases / market share, changes in behavior, or other indicators from outreach, incentive, or other programs.[1] They can be applied to programs with clear participants / non-participants, or without directly-known participants. An array of advantages and drawbacks of each of the methods are listed below.

In general, causality can never really be proven. A thoughtful logical approach utilizing alternative techniques that demonstrate consistency with causality may prove to be the most persuasive in demonstrating the program's effectiveness. We view causality research as a series of steps, involving:

- noting whether / proving that the technology actually works;
- demonstrating causality in the most reliable method(s) possible, with multiple methods used to improve credibility, and
- providing added robustness to the results by reporting and analyzing confidence intervals, ranges, and distributions where possible. Point estimates by themselves are almost certainly incorrect; however, using ranges or confidence intervals provides more robust and reliable information about impacts, and can be used with greater confidence to guide decision-making about programs, budgets, impacts and other decisions.

[1] And of course, they are appropriate for residential or commercial programs.

Table 1. Advantages and Drawbacks of Several Causality Approaches

Method	Advantages	Drawbacks
Granger causality	• Simple straightforward analysis • More statistical than econometric, so doesn't impose economic assumptions about structure	• Traditionally used for macroeconomic analysis using many periods of time series data. Suitable data may not be available for programs without clear participant lists.
Treatment effects and Propensity Scoring	• Given information on participants and non-participants, scores can be matched and then differences can be evaluated between the variables of interest (saving, attitudes, awareness, etc.) to examine effectiveness of the program. • Rosenbaum and Rubin (1983) and others have used this method in the context of demonstrating causality. • Method has become popular in the economics literature as a method to control for bias, and may be especially helpful with difficult samples. There is a recent literature from which to draw information.	• Although it provides an enhancement on control group method, the treatment of differences in scores between "matched" pairs is not specific. • More difficult to apply if participants and non-participants aren't known a priori (e.g. outreach programs)
"New Framework …"	• Using simulations one can use a variety of assumptions to evaluate the effectiveness of programs under different conditions. • There is considerable literature using these analytical approaches, and they have been applied to a variety of program types.	• Models need to incorporate all other factors that might be affecting observed changes between baseline and estimated / projected outcomes in savings , etc. in order to establish "causality". Confi-dence in attributed results are limited by data and variables, quality, etc. • Getting good estimates requires having time series data on indicators and interventions, which may not always be available or of suitable quality.
Ordered Logit	• Does not require respondents to make a "purchase / no purchase" decision – just need to rank preferences. • Because it is able to use the information on all rankings, it can generate results even in the case of relatively small sample sizes.	• This is a simulation. It describes what may happen but not what did happen. Using this in conjunction with actual market share data can be helpful in assessing the effectiveness of programs.
Delphi Technique	• Does not require large numbers of respondents, and maximizes the use of data provided. • Expertise is well-used, and experts can often be relatively easily identified. • Provides some level of quantification of uncertainties around estimates. • Quick and inexpensive to estimate.	• Less quantitative / statistical than surveys.
Fishbein causality	• Provides alternative approach to looking at effectiveness of program by examining the behavior and beliefs of the participants. • Relatively easy to implement. • Literature and quantitative studies are growing.	• Most of the literature does not look at actual changed behavior but "intent to change behavior", so causality linkage may be somewhat weaker. However, intent to change may be a necessary condition for actual change and is used extensively in assessing advertising campaigns, etc.
Meta analysis	• Makes quantitative use of studies that have already been done. • Rather than simply doing literature review, the literature can be used empirically to support claims. • Provides additional information without collecting new data – minimizes data collection costs. • Statistical research explaining the properties and use of combined data is available.	• Only as good as the original studies. • Not original research. • Statistics associated with combining studies can be complicated.

The authors' recent market causality work employed these steps, and applied several of the causality approaches in attributing impacts to program interventions. The next section addresses the methods that were used for attributing portions of the impacts to the program interventions, and addressing issues of free riders and market effects.

Toward a More Robust Causality/Attribution Analysis

The primary objective of the authors' recent work efforts was to produce information useful in judging the investment in energy efficiency programs. In producing information useful for decision making, this work incorporated three themes:

1. Many factors important to assessments are, by their nature, uncertain and this uncertainty should be reflected in the information, results and conclusions of the assessment.
2. To address uncertainty present in assessing factors such as the net impacts of a program, range estimates are used that provide a lower bound that, given the information collected, can be viewed as almost certainly being exceeded; a best estimate that can be a component of a "most likely scenario"; and where appropriate an assessment of how large the effect might be; i.e., an estimate that can be used in a "most favorable" scenario assessment. Scenario assessment of program or project outcomes is commonly used to evaluate even large scale business investments (e.g., the building of a power plant), and in assessments of research and development portfolios conducted by private companies.
3. The presentation of information must be placed in the context of certain hypotheses and program theory. The nature of program evaluation makes absolute certainty around program impacts impossible. The modern formulation of the problem of causal analysis is based on the fundamental notion of a counterfactual. Every action (e.g., program participation) has two possible outcomes all else held fixed. The difference between the two is the causal effect of the event or action. Only one of the outcomes is observed in the data because an entity cannot do two things at the same time. The other unobserved outcome--the one not chosen--is called the counterfactual for that entity. A basic point in this literature is that the causal effect of the event on Y cannot be estimated without some type of minimal assumption or restriction, even in principle, because of the inherent unobservability of the counterfactual, and that such minimal assumptions and restrictions cannot be formally tested. This is true for any investment analysis examining any impact resulting from a policy, program or business decision. Consequently, they must be justified or rationalized on the basis of a priori argument, outside evidence, intuition, theory, or some other informal means.

Theory-Based Evaluation, and Characterization / Assessment / Attribution Steps

In the context of theory-based or hypothesis-driven program evaluation, a key component in evaluation is the verification of the validity of assumptions and hypotheses. The assumptions can be as straightforward as the assumed incremental savings associated with installing a high efficiency lamp, compared to standard efficiency options, and they can be complex when they address such things as the behavior of market actors (e.g., builder / contractor / buyer interactions; manufacturer / retailer / purchaser interactions, etc.).

A focused project theory and logic model construct can help identify pivot information or assumptions, i.e., what must be true for the project to achieve its target benefits. These pivot assumptions often focus on three factors: 1) baseline assumptions, 2) the performance of technology as it is applied in the field, and 3) the inter-related issue of attribution/causality. Baseline issues involve both market conditions at the start of the intervention/program, and how that market will change over time without the intervention. Technology performance will depend on field installation and operating characteristics.

Attribution and causality issues can often be viewed as aspects of selecting the correct baseline, i.e., the ways in which the market has changed in the event the project had not been offered and which changes can truly be attributed to program activities (Susser 1997, Huberman and Miles 1997, Oakley 1997).

From one perspective, this assessment can be viewed as an analysis of the investment in a program and the return on that investment as captured in the benefit-cost analysis. Attribution assessment often provides data used as inputs to benefit-cost analysis. There will be some elements that can be more easily quantified than others, but a thorough assessment will need to address the range of benefits produced by the program.

Attribution of Energy and Demand Savings

The general approach to assessing the energy savings attributable to the program of interest is based on a net-to-gross (NTG) multiplier. The NTG has two main components – 1) a net factor and 2) a market effects factor.

Defining the Net to Gross Multiplier

The net factor takes the calculated savings (termed gross savings) for the energy efficiency measures that were installed under the program and subtracts from these gross savings the energy savings that are due to actions that participants would have taken anyway, i.e., actions that were not induced by the program. Commonly termed the free-rider effect, this subtraction is meant to correct for energy efficiency measures that would have been installed at the project even if the project had not participated in the program.[2]

The market factor is designed to capture program effects and impacts that go beyond the measures installed through the program at the specific project sites. The overall net-to-gross multiplier is meant to capture these two attributes of the program – net impacts at participating projects and spillover impacts that result from the program but occur at other projects or otherwise are missed by the program's accounting for energy savings. As a result, overall net program impacts are based on the development of a net-to-gross (NTG) ratio. When the NTG ratio is multiplied times the estimated gross program impacts, as accounted for by savings from installed measures recorded in program records, the result is an estimate of program impacts that are attributable to the program (i.e., impacts that would not have occurred without the program). The basic equation is:

$$NTG = (\text{Net Factor}) \times (\text{Market Factor})$$

The net factor equals the attributed fraction of savings or (1 – free riders). As an example, if free riders are estimated at 0.25 (or 25%), the net factor is 0.75 (or 75%).

[2] Or for outreach programs, if the person had not been influenced by program advertisements or other interventions.

The market factor is a combination of program spillover factors that may influence actions to be taken outside of the program. This factor generally enhances or adds to the program's measured impacts.

$$\text{Market Factor} = [1 + \text{spillover}(1) + \text{spillover}(2) + \text{spillover}(3)]$$

Where the market multiplier is the sum of the spillover impacts across participant spillover (types 1 and 2) and non-participant spillover (type 3). Each type of spillover is defined below:

Spillover 1 = participant spillover that applies to "within program project" actions.
Spillover 2 = participant spillover that applies to other non- program projects.
Spillover 3 = non-participant spillover that occurs in non- program projects.

Spillover refers to a variety of indirect impacts stemming from an energy efficiency program. Spillover effects can occur through a variety of channels including:

1. Participant spillover within projects, where program participants undertake additional, but unaided, energy efficiency actions based on positive experience with the program. For example, homeowners or builders in a residential new construction program may be so encouraged by measure performance that they opt to install measures beyond program requirements or incentives.
2. Participant spillover on non-program project, where participants take unaided actions on projects outside the program. For example, the participating builder may learn design techniques that he then applies to other homes outside the program, achieving energy efficiency and savings in the market beyond program homes.
3. Non-participant spillover (sometimes called the free-driver effect) where a market actor may hear about a program measure from a customer who participated in the program, or through advertising and decide to pursue it on his or her own without participating in the program; as a result, program records will not include these impacts. An example may be a builder that has not participated in the residential new construction program, but in order to compete, adopts several program-recommended measures that are recognized in the marketplace, achieving additional energy savings in the market and moving the market forward.

There can also be "Other Market" Spillover which can occur through several pathways. For example, manufacturers may change the efficiency of their products, and/or retailers and wholesalers may change the composition of their inventories to reflect the demand for more efficient goods created through an energy efficiency program. Another example might be new building codes or appliance standards adopted in part due to the demonstration of technologies through an energy efficiency program.

The attribution concept applied here is to link program activities to impacts in the market, and to exclude actions that would have been taken anyway. The underlying concept is that only impacts "caused" by the program should be included in the final net program impacts estimate. Free riders reduce savings attributed to the program since they would have occurred anyway, but spillover adds to program savings as it refers to additional actions that produce energy savings.

Methods to Derive Net Factor Estimates

There are several approaches for estimating net savings of energy efficiency programs. The two basic methods are – 1) estimation approaches, including the differences of differences approach; and 2) self-report analyses.

In the differences of differences approach a non-participant control group is identified and the difference in energy use between participants and non-participants is used as the net impact of the program (Violette 1991). Other approaches were discussed in the first section of this paper; for example, this difference of differences "control group" approach may be refined using propensity scoring. Other approaches may be appropriate, depending on the program.

The self-report approach takes information directly from program participants and asks them what actions they would have taken in the absence of the program, and what actions they may have taken that comprise spillover due to the program. The self-report method was used in a recent evaluation in California to "better estimate participant free-ridership and spillover savings … the inclusion of both free-ridership and non-participant spillover savings at the measure level to produce 'comprehensive' net savings provides the more accurate measure of actual program savings (RLW Analytics 2003).[3] As mentioned earlier in the paper, absolute proof of attribution is unattainable. The modern formulation of the problem of attribution analysis is based on the fundamental notion of a counterfactual for the participant in an energy efficiency program; or, for other policy decisions, it can be individual, state, country, or other unit. Every individual has two possible outcomes all else held fixed. The difference between the two is the attribution effect of the program, event or action for that individual. Only one of the outcomes is observed in the data because an entity or individual cannot do two things at the same time. The other unobserved outcome--the one not chosen--is called the counterfactual for that individual.

A basic point in the literature on attribution is that the causal effect of the event on Y can not be proved without assumption or restriction, even in principle, because of the inherent unobservability of the counterfactual, and that any such minimal assumptions and/or restrictions that allow for estimations cannot be formally tested. Consequently, they must be justified or rationalized on the basis of a priori argument, outside evidence, intuition, theory, or some other informal means. This is partly the basis for the statement by Heckman that "there is no mechanical algorithm for producing a set of 'assumption free' facts or causal estimates based on those facts" (Heckman 2000). This in not only true for assessing attribution in the context of an energy efficiency program, but it is true for any policy assessment including those pertaining to education programs, health programs, fiscal policies and virtually any assessment of a business decision that purports to have a cause and effect. The information derived from the market effects and causality analyses must be considered in the context of information from the market characterization work, the market assessment showing the relationship between the program and market indicators and logic models and program theory. Taken together, these relationships and the self-report estimates discussed in this section can make a compelling case for attribution of the net-to-gross savings estimates for particular programs.

[3] Recent work (Borst 2003) on estimating net impacts for new construction programs has also favored the self-report method.

Estimating Program / Market Area Baseline, Free Riders, and Net Effects

Establishing a baseline level of energy efficiency is necessary to determine how much electricity is saved through energy efficiency measures. In examining a new construction program energy code can serve as a useful baseline. However, in assessing free ridership through surveys of participating owners, builders, and market actors, the analysis relies not on the provisions of the current energy code but rather on survey respondents' perceptions of standard practice at the time that their projects were underway. To gauge free ridership one must ascertain how much electricity the new construction program saved as a percentage of total savings from a given project. Since this attribution data is based on survey respondents' (e.g. builder's) professional estimates and opinions, and not on (measurable) hard data, the appropriate baseline for this assessment would be judged to be the one offered by the respondents themselves. Note that this applies only for determining attribution/free ridership. The energy code is an appropriate baseline against which to measure gross savings independent of the "net" effects of free ridership and spillover.

Estimating Free Riders as a Baseline Adjustment

Given that large pieces of the authors' recent work used the self-report technique to measure and attribute impacts, an extensive series of surveys were carried out. Free ridership was assessed through a series of questions asked of market actors that included program participants, non-participants, and program staff and implementers. These questions were designed both to elicit direct estimates of savings attributable to the program as well as to identify qualitatively whether the program's financial and technical incentives influenced energy efficiency decision making that might be important for estimating the market factor.

Participating home owner and home builder respondents were asked for their estimate of the electric energy savings (in terms of the percentage of total electric bill) as a result of all of the energy conservation measures installed at their home. They were also asked for their upper bound, lower bound, and best estimate of the electric bill savings that were attributable to the project's participation in the program. Program attribution was then calculated as the savings resulting from program participation divided by the total savings from all measures.

Non-participating home builders offer an outsiders' perspective on the influence of the program. To gauge their relative views of the potential for energy conservation, these firms were asked to estimate the average percent reduction in electric use when efficiency measures are installed with assistance from a program like the new home construction program in question. A detailed series of questions were asked to identify the fraction of savings for which the program could reasonably take credit.

Approach for Estimating Spillover

The basic approach for assessing spillover for each market actor was to ask a set of questions that comprised a three step approach:

- Step 1: Determine if the respondent believes that the effect exists at all. These are usually yes/no questions that ask, for example, whether the respondent believes that some participants may apply the knowledge gained through the program to other projects that are not program projects. Questions were asked related to extra measures that might have been installed beyond the project records (spillover 1), extra measures they might have installed in non-program projects (spillover 2), and non-participating respondents were asked about possible program influence on their practices (spillover 3).

- Step 2: Determine the extent of the effect in terms of the market share it might apply to. In the case of a residential new construction program, the question would address the number or share of program homes they installed extra measures in; how many non-program projects a market actor might apply knowledge gained from the program to over the evaluation period; or the number of homes to which non-participants applied program-derived knowledge or techniques.
- Step 3: Determine the size of the effect per project. As the final piece, respondents were asked about the savings "per spillover project", in a sense. For example, the question might be what percent reduction in electricity use might be achieved by those market actors that apply knowledge from residential new construction program projects to non-program projects.

The process of breaking the questions into incremental parts helps the respondent think through each step and, while difficult to address, it allows the respondent to provide their expert judgment as a participant in the new homes market. Bounds and ranges were asked as part of these question batteries, and we found respondents were often much more comfortable talking about ranges than point estimates, reinforcing the importance of moving beyond point estimates in attribution research.

These basic techniques have been adapted by the authors for a broad range of programs, including residential product outreach and product incentive programs; new homes and remodeling programs; commercial new construction; renewables, and other programs. Enhancements being examined that include scenario analysis and decision tree techniques to "bound" the effects, and options to use a Bayesian approach addressing "degree of belief" in the impacts. Thus, the attribution work is incorporating several methods of addressing risk – an important component of using causal results.

Summary and Implications

As part of recent projects to attribute market effects to interventions from dozens of different residential and commercial programs, the authors employed several useful steps to improve the robustness of the analysis. The focus was on using theory-based evaluation, but also worked toward providing a burden of proof on par with the criteria for other public and private investments. Four steps were undertaken for this effort.

- We examined an array of causality modeling options available from the literature.
- We used the distribution of the measure and intervention impacts, rather than relying on point estimates.
- We use an enhanced method to address free riders, spillover, and free drivers to allow for partial free riders and used indicator methods to provide evidence on program-induced effects on spillover and free drivers.

- We directly addressed the issue of uncertainty and risk in the attribution work using scenarios to "bound" the effects. Thus, the attribution work examined several methods of addressing risk – an important component of using causal results.

This approach used much more of the information gathered during the data collection, and more fully reflects the range of impacts induced by the program. Although causality can never be proven, these extra steps improved the reliability and robustness of the results of the causality analysis and provided a better foundation to guide program and investment decisions – one of the most important goals of an evaluation.

References

Ajzen and Fishbein. 1980. *Understanding Attitudes and Predicting Social Behavior*.

Borst, M. et al. 2003. "Contrasting Approaches to Estimating Program Net Savings in NRNC", *Proceedings of the 2003 Energy Program Evaluation Conference*, p. 251, Seattle WA.

Chao, Corradi, and Swanson. 2002. "An out of sample test for Granger causality", web document.

Granger and Newbold. 1986. *Forecasting Economic Time Series*.

Heckman, J. 2000. "Causal Parameters and Policy Analysis In Economics: A Twentieth Century Retrospective," *Quarterly Journal of Economics*, (2000), 45-97.

Huberman, A. and M. Miles. 1997. "Data Management and Analysis Methods," in *Collecting and Interpreting Qualitative Materials*, Denzin *et al.* eds., Sage Pubs., Thousand Oaks, CA.

Junsoo, Lee and Michael Niewsiadomy. 2003. "Comparing Self Selection Estimators to Propensity Scoring Methods: The Impact of a Xeriscape Newsletter on Residential Water Demand". Presented at Western Economic Association International Meetings, Denver, CO.

Martino, J.P. 1993. *Technological forecasting for decision making*, New York: McGraw-Hill.

McKenzie-Mohr, Doug, 1995. "Promoting a Sustainability Future: An Introduction to Community-Based Social Marketing", National Roundtable on the Environment and the Economy (NRTEE).

Michalopoulos, Charles, and Carolyn Hill. Unknown date. "Can Propensity Score Methods match the finding from a random assignment evaluation of mandatory welfare-to-work programs?" Web document.

NYSERDA (New York State Research and Development Authority). 2003. "Causality Analysis: Evidence linking the New York Energy $mart Program to Market Outcomes" in *Program Evaluation and Status Report to the System Benefits Advisory Group*, May 2003, p. 5-1 – 5-7, p 5-20 to 5-34, Albany, NY.

Oakley, A. 1997. "Confronting Causation," *Metascience,* **12** 149-152.

Peters, Jane, and Shel Feldman. 2001. "I Can Do It! The Role of Self-Efficacy in Motivating Changes in Attitudes and Behavior Relating to Energy Efficiency and Renewables", *Proceedings of the 2001 Energy Program Evaluation Conference*, Salt Lake City, Utah.

Peters, Jane, Ken Seiden, Sharon Baggett, and Lynn Morander. 1998. "Changing Consumer Attitudes to Energy", *Proceedings of the 1998 ACEEE Conference*, pp. 8.251-261. American Council for an Energy Efficient Economy, Washington DC.

RLW Analytics. 2003. "Final Report: Evaluation of the Savings by Design Program," prepared for California's Investor Owned Utilities, April 2003.

Rosenbaum and Rubin. 1983. "The Central Role of the propensity score in observational studies for causal effect", *Biometrika*.

Schuster, E.G. 1985. "The Delphi Method: Application to elk habitat quality. INT-353. Ogden, UT. USDA Forest Service, Intermountain Research Station.

Sebold, Fred, Alan Fields, Lisa A. Skumatz, Shel Feldman, Miriam Goldberg, Kenneth Keating, and Jane Peters. 2001. "A Framework for Planning and Assessing Publicly Funded Energy Efficiency", Report prepared for Pacific Gas and Electric Company, San Francisco, CA.

Skumatz, Lisa A., and David Weitzel. 2001. "Making The Most Of Your Data – Reliable Techniques For Estimating Baseline And Projected Market Shares From MT Interventions With Limited Observations", *Proceedings from the AESP Conference*, Ponte Vedra, Florida.

Skumatz, Lisa, David Weitzel, Rose A. Woods, and Blake Bear. 2003. "Buying Into Conservation – How much Does It Matter Whether Customers Believe They Can Effect Change?" *Proceedings of the EEDAL Conference*, Turin, Italy.

Susser, M. 1973. *Causal Thinking in Health Sciences: Concepts and Strategies,* Oxford University Press.

Violette, D. et al. 1991. *Impact Evaluation of Demand-Side Management Programs: A Guide to Current Practice*, Electric Power Research Institute (EPRI CU-7179).

Woods, Rose A., and Lisa A. Skumatz. 2004. "Self-Efficacy in Conservation – Relationships between Conserving Behavior and Beliefs in the Ability to Make A Difference", *Proceedings of the 2004 ACEEE Conference*, American Council for an Energy Efficiency Economy, Washington DC.

Yeon-Hyeon Hwang, Seong-Il Kim, and Jian-Min Jeng. 2000. "Examining the Causal Relationships among Selected Antecedents of Responsible Environmental Behavior", *Journal of Environmental Education*.