How Much Do We Know About Savings Attributable to a Program?

Stefanie Wayland, Opinion Dynamics, Oakland, CA
Olivia Patterson, Opinion Dynamics, Oakland, CA
Dr. Katherine Randazzo, Opinion Dynamics, Carlsbad, CA

ABSTRACT

As evaluators, we are often asked, “What is overall program savings?” a great follow-up question is, “How is that savings distributed across participants and measures?” The first question is about a single number. The second question requires answering how different parts of the program affect participants and/or how specific participants respond to the program. This paper looks at statistical and causal modeling approaches to answering these questions.

For energy impact evaluation, a primary research question is, “How much did the program cause participants to save energy and reduce demand?” This is the fundamental question for attribution of savings to a program, and getting to a useful answer can take more than just a fixed-effects model. This question of whether a program caused a change in energy usage moves us away from standard statistical analysis into causal analysis. In energy evaluation terms, we want to estimate net impacts.

Researchers specializing in causal modeling have made amazing gains in methods for assessing causality in just the last few years. We are working to bring some of these methods into energy evaluation to use recent advances and to take advantage of the huge increase in information that is now available with AMI data.

In this paper, we compare traditional methods for impact evaluation such as linear fixed effects models with several newer methods: multilevel modeling, Bayesian methods, and Bayesian additive regression trees (BART) to illustrate what we get from our current methods and how we can overcome those shortfalls to usefully attribute savings due to programs.

Different programs can require substantially different approaches to attribute savings. We have used multilevel modeling in a home energy report program to assess savings for individual participants, and are using Bayesian additive regression trees for a home energy report program. We discuss standard methods for attribution in these programs and compare results to show where the newer methods can yield substantially more information about how savings are distributed in the participant population.

Those who want to understand evaluation results will gain insight into why evaluators choose the models that we do, what impact results mean, and how emerging methods can aid decision making.

Introduction

As energy program evaluators we try to answer, “How much did the program cause participants to save energy and reduce demand?” For this paper, we are looking at this question from the standpoint of statistical models based on energy consumption data. The question of whether a program caused a change in energy usage moves us away from standard statistical analysis into causal analysis.

We perform impact evaluations to help inform decisions about programs. Is the program cost-effectively delivering demand and/or energy reductions? Should you call a demand response event tomorrow, and if so, how much load reduction is available? Should the program continue next year? Where does the program fit within a portfolio? Should you allocate more or fewer resources to the program? Should you modify the program? Should you bid the program capacity into your ISO?

All these decisions are made under uncertainty, since we cannot know for certain what will happen next year, or even tomorrow. We cannot know for certain what we should do. Part of evaluation is to assess uncertainty, and to quantify it, where possible, allowing informed decisions based on our work. The more information we have about when, how and how much a program affects customers, the better the decisions we can make. One of the most interesting sources of uncertainty is in the variety of program impacts on different customers. Are all of the participants realizing the same savings? Should we modify the program delivery for some participants where we expect lower than average savings?

In this paper, we highlight some newer methods such as Bayesian modeling, linear multilevel modeling, and Bayesian additive regression tree (BART) models to illustrate where our current methods fall short and how we can overcome those shortfalls to best attribute savings to programs. How we attribute program impacts can help to inform the most useful range of program-related decisions.

Different programs can require substantially different approaches to attribute savings. We have used multilevel modeling in a home energy report program to assess savings for individual participants, and are using Bayesian additive regression trees to estimate individual savings in a home energy report program. We discuss standard methods for attribution in these programs and compare results to show where the newer methods outperform.

Discussion

We give a brief explanation of attribution, causal data analysis, and propose some models to work within the framework of energy program evaluation. In order to get at attributable program impact, we use causal data analysis, which is a rapidly evolving field. Recent advances in the field have allowed much more granular program impact results, with some approaches allowing us to estimate impacts at measure or participant levels rather than the program level.

Causal Analysis

Causal analysis is applying techniques for quantifying the effects of causes. In energy evaluation, we are usually interested in what we call net effects, and causal analysis uses statistical modeling to make those estimates. In this paper, and most often in energy program evaluation, we are applying the approach to causal modeling formalized by Rubin (Rubin 1974), which is distinct from other schools of thought (e.g. Pearl 2009) on the subject. When using Rubin's approach to causal analysis, we are interested in measuring the difference between what actually happened and what would have happened without the program. Understanding Rubin's simple model (i.e. program impact is the difference between what happened and what would have happened) helps us to check assumptions before believing results. This is one way to approach the central problem of causal analysis – but we cannot know for certain what would have happened. We use causal models and experiments to estimate what would have happened, and the attributable impact of the program is the difference between what actually happened and the modeled estimate of what would have happened without the program.

How does causal modeling help to inform decision making?

Approaches to Attribution/Causal Modeling

These are some of the approaches we have used for calculating attributable program impacts. One of the main traditional econometric approaches is fixed or random effects modeling, which yields attributable average savings estimates for a program. The newer methods we outline are multilevel linear models which give individual participant savings estimates, Bayesian additive regression trees which also give individual participant savings estimates with some nicer properties than multilevel linear models, and Bayesian SAE models which give measure specific savings estimates.

Econometric Fixed and Random Effects Models

Linear fixed effects models are used for many consumption-data based impact analyses. These are traditional linear regression models, with the addition of customer-specific intercept terms. The intercept terms account for factors that do not vary before and after participation, such as square footage, appliance stock, habitual behaviors that remain unchanged from pre- to post-participation periods, household size, and any other factors that do not vary over the longer term. Such consistent factors are represented in account-specific constant terms in a fixed-effects model. The models also usually account for differences in weather and pre-program energy use between participants through inclusion of terms representing them in the equation.

These models can yield overall average attributable impacts for a program when they are used with appropriate comparison or control groups. It is also possible to explicitly generate attributable impacts that interact with other variables by directly including these interactions in the model. The results are not effective for measuring customer specific impact, since it is only possible to include a very few interactions between treatment and other variables.

Multilevel Linear Models

Multilevel linear models are a generalization of linear models that simultaneously include both group-level information and customer-level information (or even more explicitly defined levels). These models can estimate how much of the variation in savings is accounted for at each “level” of customer data. For example, a basic multilevel model that looked at the customer, neighborhood, and substation levels would indicate how much savings differed at each of those levels. This gives us considerably more information about where to attribute savings from a program.

Bayesian Additive Regression Trees (BART)

Bayesian additive regression trees (BART) is a Bayesian non-parametric model that can take many input variables and automatically build a model that selects the important variables and interactions of those variables. Bayesian models are a set of probabilistic methods to update beliefs based on evidence from data, while non-parametric methods use the data to build probability structure. These models can be used as (among other things) a non-linear extension of multilevel-linear models, which can include information at many levels (customer, group, measure, etc.) without having to explicitly define a single functional representation for the estimating equation.

Bayesian SAE Models

These statistically adjust engineering models use a Bayesian approach to combine engineering results with consumption data. For technical reasons, the Bayesian approach overcomes many of the shortfalls in standard SAE models, where the assumption basically decided the outcomes more than the data.

Practical Applications

Behavior Programs with Multilevel Modeling

Multilevel modeling can demonstrate the diversity in participant responses to a residential HER program. This variation in responses suggests that there is considerable scope for these types of programs to increase savings by moving away from a one-size-fits-all approach to home energy reports and studying the characteristics of high, medium, and low savers in order to target program offerings more effectively. For example, a utility might recruit new cohorts mostly made up of customers similar to current medium and high savers, or stop sending reports to negative savers.

Figure 1 below illustrates the differences between the results than can be estimated by a typical fixed effects model (left panel) and a multilevel model (right panel) using hypothetical data. The X-axis represents number of days in an HER program, and the Y-axis represents total program savings. The thick dashed lines represent average program savings, and the solid lines represent the savings of individual customers. It is easy to see that the overall estimates of program savings, which take all of the participants into account, are quite similar. However, the two types of models make very different assumptions about underlying customer behavior. In the fixed effects model, customer-level differences do play a role in the form of the different intercepts. However, each hypothetical customer gains the same amount of savings from being in the program for an additional day (i.e., the slopes of the lines are the same), and one would assume that maximum program savings would be achieved by adding as many customers as possible. In the multilevel model, customers respond to treatment in different ways. Customer C achieves huge savings, Customer A resembles the overall average, Customer C barely saves anything, and Customer D actually has negative savings. Overall program savings would be increased by attracting more customers like Customers A and C and changing or perhaps even ending reports for customers like Customer B and Customer D.

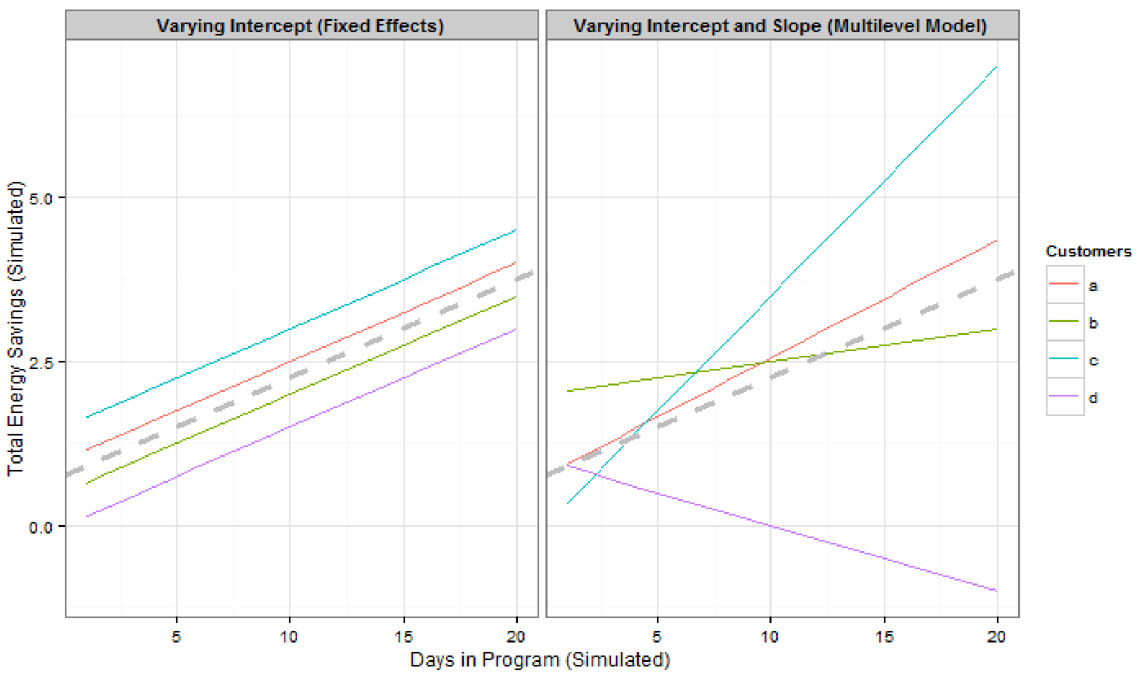

Figure 1. Comparison of Fixed Effects and Multilevel Models.

Numerous process evaluations and interviews with program participants and implementers suggest that customers respond differently to HERs, and that the real world looks more like the right-hand than the left-hand panel of Figure 1. It is also possible to estimate savings levels for individual households by running individual regression models for each participant. However, the multilevel modeling approach provides several important advantages over individual regression in establishing individual household savings levels. First, multilevel modeling statistically controls for weather differences between pre- and post-periods for an individual household as well as across households. In contrast, individual models solely control for weather differences between pre- and post-periods for an individual household. Second, multilevel modeling allows for modeling the influence of variables that do not change over time that apply to customers and for generating appropriate standard errors and statistical tests. Third, results from multilevel regression models adjust individual savings estimates based on control group usage during the treatment period, so the savings estimates are much closer to actual savings than results from individual regressions. Finally, information is shared across customers in multilevel models, so the unexplained variance in individual savings across participants is much lower when we make estimates using a multilevel model.

Another major advantage to multilevel modeling is its ability to estimate how much of the variation in savings is accounted for at each "level" of customer data. For example, a basic multilevel model that looked at the customer, neighborhood, and substation levels would indicate how much savings differed at each of those levels. This information would be extremely helpful in program design, as it would help implementers better understand whether to target different interventions at specific households or entire neighborhoods.

In short, multilevel modeling provides a useful compromise between estimating aggregate results but ignoring individual differences (traditional fixed effects models) and estimating individual results but overlooking program-level dynamics (individual regression models). One drawback of this type of multilevel modeling is the assumption that the individual participant savings estimates are normally distributed, which can mask informative patterns that are not directly included in the model.

We used multilevel modeling to examine savings for the most recent year of an HER program. The household-specific savings estimates obtained by the multilevel model showed not only that customers varied significantly in terms of the amount of energy they saved, but also that approximately 40% of customers (both gas and electric) actually had negative savings associated with the program. This is an important finding in its own right, as it suggests that residential HER programs could boost savings by enrolling more customers likely to benefit highly from the program and modifying their approach to negative savers.

We split the participants into five savings groups based on their results from the multilevel model to explore the characteristics of high versus low savers. We created separate groups for gas and electric customers, so the same customer could be a high saver for gas and a low saver for electricity. Figures 2 and 3 below show the distribution of individual electric and gas savings elements with colors that indicate savings groups.

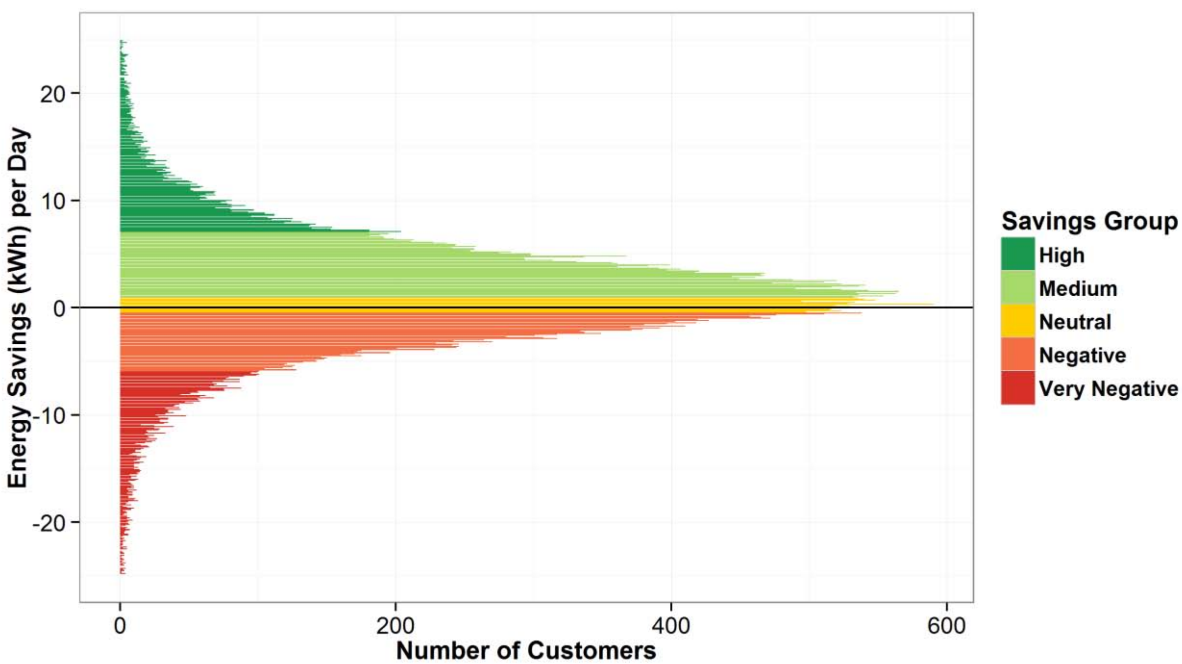

Figure 2. Distribution of individual electric savings estimates. Group cut-offs are as follows: High > 7 kWh; Medium > 1 & $\leq$ 7 kWh; Neutral > -0.5 & $\leq$1 kWh; Negative > - 6 & $\leq$ -0.5 kWh; Very Negative < -6 kWh

The variation in individual savings estimates is striking. High electricity savers save an average of 12.33 kWh per day, but some save over 20 kWh per day. Gas participants might save or increase usage by over 1.5 therms per day. It is important to realize that differences in savings might not only be a hostile reaction to the reports. For example, a participant might have a baby, start working from home, or expand a house. We discuss the implications of this later in the paper. For now, it is sufficient to note that removing these negative savers from the program would increase overall program savings.

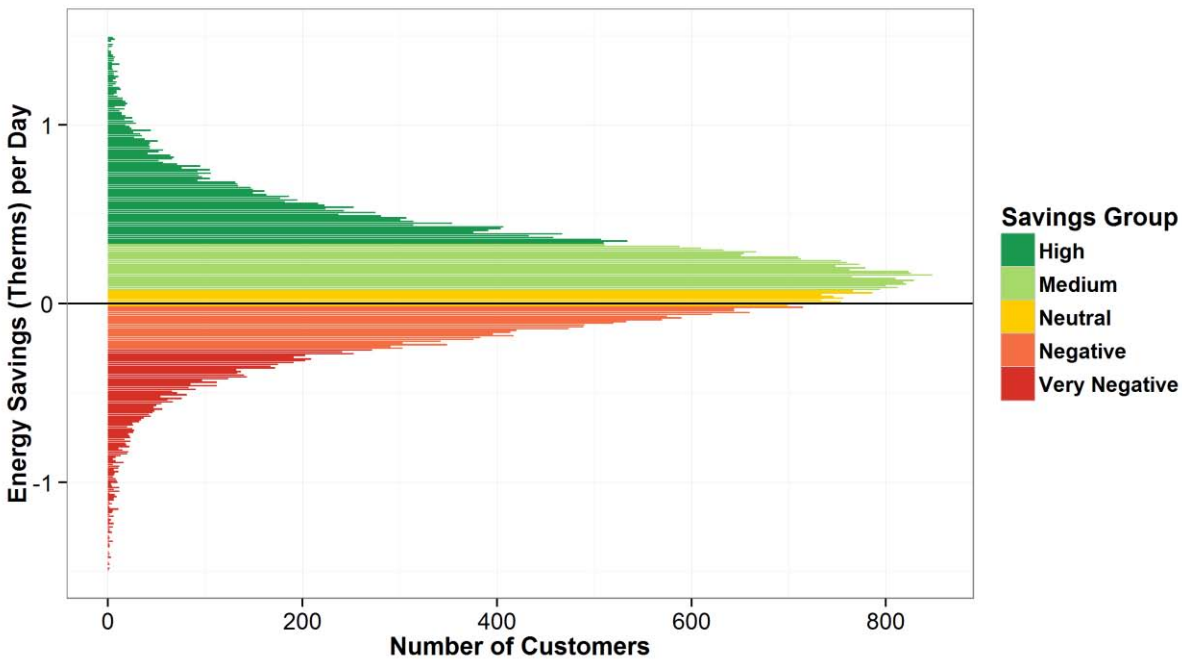

Figure 3. Distribution of individual gas savings estimates. Group cut-offs are as follows: High > 0.33 therms; Medium > 0.08 & $\leq$ 0.33 therms; Neutral > - 0.02 & $\leq$ 0.08 therms; Negative >-0.25 & $\leq$ -0.02 therms; Very Negative < - 0.25 therms

To put these numbers in context, overall savings for each cohort in the evaluation report range from 0.006 to 0.04 therms per day for gas customers and 0.15 to 0.6 kWh per day for electric customers, using the traditional weather-adjusted fixed effects regression models. Notice that these values fall in the highest density portions of the individual savings distributions. The cohort-level averages produced by traditional models are accurate, but hide a great deal of household-level variation.

BART for Individual Savings Prediction in Behavior Programs

A direct extension of the multi-level modeling approach above is BART. BART is a multi-level modeling approach that simultaneously incorporates participant and program-level information into a model for estimating energy savings. This is a non-linear model we can use for estimating individual participant savings, while relaxing the assumption of an approximately normal distribution of savings in the population. BART can also require much less modeling effort, since it is possible to mostly automate variable selection, and the tree structures automatically incorporate variable interactions. BART can be very effective at modeling heterogenous treatment effects.

We are currently evaluating a HER program to maximize program cost-effectiveness through analytical insights – in particular, to identify customers who contribute energy savings, and to develop strategies to support enhancing savings or decreasing costs through revisions to program design and implementation. These data-driven insights can reveal opportunities to revise existing report content or frequency, or support of future eligible customers for program inclusion. These are the same points that we discussed above, with the addition of substantial flexibility in the model structure.

Starting with consumption data for the treatment and control groups in the HER program, we use BART to estimate the savings for individual customers. This allows us to quantify the total number of negative, neutral and positive savers within the participant population, and classify each customer into their savings category. This approach allows us to accurately describe the distribution of savings across households, allowing a data-driven grouping of the participants into negative, neutral and positive savings groups.

SAE Analysis with Bayesian Modeling

A key challenge faced by evaluators and the energy efficiency industry is bridging the gap between engineering and statistical analysis for energy savings impact estimates. In many cases, these results do not always align in terms of absolute or relative savings.

Historically, evaluation protocols have suggested conducting a statistically adjusted engineering (SAE) analysis which combines some of the engineering and simulation analyses with a statistical analysis using consumption data. Revised engineering estimates can serve as inputs to an SAE analysis that may be more useful for decision making than the current *ex ante* values, providing a more holistic energy savings impact result.

SAE models make use of engineering knowledge about what customers installed in or out of the program. Specifically, instead of inserting a dummy variable for participation (1 for participation, and 0 for customers and time periods with no participation), or a set of dummies for multiple installations, engineering or deemed estimates of savings anticipated from the installation are entered for each customer and time period where it applies. This method has the advantage that it may fit the data better than the dummy versions, and therefore improve precision. The coefficients that are associated with these variables represent a realization rate for each measure installed for which there is a separate estimate in the model. In other words, if the engineering estimate for installation of weatherization is 20 kWh per day per customer, and the regression coefficient for weatherization was 0.9, this would mean that the statistical estimate of savings was 18 kWh, or 90% of the engineering estimate.

However, SAE models have consistently failed to produce results that go beyond scaling engineering findings to statistical analysis results. In addition, engineering estimates are generally stated in annual terms, while the consumption data is measured monthly, or with AMI data at hourly, or even sub-hourly intervals. This means that the engineering estimates must be distributed across all the time periods of the analysis. How this distribution is implemented has strong effects on the model results. Thus, if errors are made in the distribution of savings over time, estimates can be biased due to this measurement error.

We develop an SAE modeling approach based on Bayesian random effects regression analysis to estimate energy impacts at a measure level for all participants. This analysis starts with standard engineering savings estimates, but incorporates an additional step by inserting measure-specific engineering estimates into a regression analysis to identify distributional and confidence assumptions by pilot offering. Confidence assumptions about engineering savings estimates are developed through engineering expertise around the merits of the algorithms, assumptions and measures reviewed.

We use the engineering estimates as prior information in a Bayesian random effects model regression analysis with a matched comparison group to estimate energy savings associated with each measure. The Bayesian model is a random effects model that brings the engineering results into the statistical analysis, much like a standard SAE model. Unlike a standard SAE model, the results do not depend entirely on the choice of how the annual engineering savings is distributed through the months, days and hours of the year. Instead, the way that Bayesian modeling works is that it takes the distributions of engineering savings as a starting point, an educated guess that is then updated with information from the consumption data collected from participants and comparison customers.

At first, this may sound like the standard SAE approach, but the way that the distribution of engineering energy savings applies to the final answer is fundamentally different. In the standard SAE approach, the decision on how to distribute the engineering energy savings determines the distribution of savings in the final answer. The only way that the consumption data influences the result is on the overall scale, not the distribution. Bayesian models take the distribution of engineering savings as a starting point that is then updated with the actual distribution of consumption savings (based on differences in consumption between pre- and post-periods and treatment and comparison customers). The final answer is then a true combination of engineering-based savings and statistical consumption analysis-based savings.

When the engineering and consumption based results are fully integrated in this way, it is much easier to make decisions about a program. In contrast to most evaluation results, where disparate results from the two approaches require substantial personal judgement on which savings estimate to apply, we can clearly explain to what extent the two numbers align, and choose an optimal middle ground.

Implications

Attributable results are most useful if they are structured to help inform key decisions, and there is an expanding set of tools that make it possible to inform a much broader range of decisions than has been historically possible with standard evaluation approaches. We can learn much more about what is going on with a program when we have savings estimates at a measure or individual participant level. Then we can examine groupings of measures and participants to help improve a program based on model results.

These models can apply to a broad range of programs, but are potentially the most useful when programs are underperforming or may have significant improvements available. It may not be worthwhile to apply these methods when the program is well-established and serving the needs of the utility and implementer. Pilots or emerging programs are places where these methods will probably be most useful.

Conclusion

There are many new approaches that overcome some of the limitations of established approaches, but sometimes at a cost of more work for evaluators, more required data, and the application of different sets of assumptions that must be understood by both those doing the modeling and those making decisions based on the results. Sometimes the traditional approaches yield all the information necessary for making simple decisions about a program, but we find that often the additional information from one or more of these emerging techniques can be highly valuable for much more nuanced program-related decisions. We’ve also only covered one of several schools of thought on causal analysis. There is a huge amount of value available in attributable results based on good analysis whether that is a simple, traditional approach or one of these emerging ones.

References

Bleich, Justin; Kapelner, Adam; George, Edward I.; Jensen, Shane T. Variable selection for BART: An application to gene regulation. Ann. Appl. Stat. 8 (2014), no. 3, 1750--1781. doi:10.1214/14-AOAS755.

Chipman, H.A., George, E.L. and McCulloch, R.E. BART: Bayesian Additive Regressive Trees. The Annals of Applied Statistics, 4(1):266–298, 2010.

Pearl, J., Causality: Models, Reasoning, and Inference. Cambridge University Press, New York. 2nd edition, 2009.

Rubin, D. “Estimating causal effects of treatments in randomized and nonrandomized studies.” Journal of Educational Psychology, Vol. 56, 5, (1974) 488-701.