Current Methods in Free Ridership and Spillover Policy and Estimation
FINAL

Prepared by

PWP, Inc.

Evergreen Economics

February
2017

Table of Contents

Executive Summary

This report presents the results of research conducted by PWP and Evergreen Economics (the PWP team) for Energy Trust of Oregon to assess policies and methods used to estimate free ridership and spillover of energy efficiency programs that are currently being used by organizations nationally. The basic goals of this research project were to:

- Determine what types of policies are currently used to determine free ridership and spillover.
- Determine what methods are being used by organizations around the country to estimate free ridership and spillover rates, including the strengths and weaknesses associated with each method, such as the costs of data collection and analysis, the accuracy and variance of estimates, the sample sizes needed, how or if the effects of long-term customer participation or program market presence are being incorporated into the estimates, etc.
- Develop actionable recommendations that will allow Energy Trust to improve its current free ridership and spillover methods and estimation process. An important consideration is that methods should have reasonable costs for the level of reliability provided relative to the overall budget and scope of the program.

To address the above goals, the evaluation team relied on Energy Trust input, on an extensive literature review, including review of Energy Trust materials, and eight telephone interviews with experts, each representing decades of experience conducting and refining net-to-gross (NTG) research, including the most recent advances in the state of the art. ES.1 Conclusions

Key findings reported throughout this report are summarized below.

- Perhaps the most significant result of our literature reviews and interviews is that there is no silver bullet; there are no breakthrough techniques in net to gross analysis that would allow Energy Trust to calculate net savings with greater confidence and at reasonable cost than the self-report approach it is currently using. Instead, the NTG landscape remains a patchwork of methods and policies where most jurisdictions use the same self-report technique that they and their consultants recognize as flawed.
  - Self-report may not be best practice, but it is standard practice, used by most jurisdictions for free ridership and participant spillover and sometimes for non-participant spillover.
  - Best practice is generally considered to be triangulation using a mix of methods, so that results from different methods can be used to validate one another and enhance confidence in the resulting NTG value.

- Precision for most methods, and particularly for self-report, typically refers to sampling precision only, and is a function of sample size and the number of segments or strata for which values are to be estimated. For free ridership, the relationship between sample size (and cost) and precision is straightforward, since all respondents will have undertaken the program-supported action. For non-participant spillover, it may be possible to estimate what percentage of the population replaced a certain type of equipment (i.e., with an average EUL of 12 years for air conditioners, roughly 8 percent would be replaced in any

given year), but it is extremely difficult to predict how many of those replacements would have involved program qualifying equipment, making it challenging to predict both the required sample size and cost for non-participant spillover studies.

- For both free ridership and spillover calculations, a number of states have guidelines to ensure that a consistent approach is used that mitigates the shortcomings of survey- and interview-based techniques. These guidelines, generally agreed to by all stakeholders, help ensure consensus regarding the net analysis approach and widespread acceptance of the results.

- In addition to the self-report approach, a variety of other techniques are available and proven, including econometric methods (such as billing regression and discrete choice models) and market effects or market transformation studies (including price response or elasticity models and market data analysis) and use of a standard practice baseline.
  - Both the literature and the experts we interviewed note that these techniques have their own limitations, including higher cost. However, they have significant value as a reality check/validation of self-report findings and as an added data point in the application of a triangulation strategy.
  - Many of these alternate approaches rely on the availability of a non-participant sample with enough households or businesses that have undertaken a targeted efficiency action without taking advantage of program incentives or services. Such a sample can be difficult and costly to identify, meaning that hundreds or even thousands of non-participants may need to be reached before enough customers are found who have taken the program-supported action without participating in the program.
  - For both self-report and alternate techniques, results often convey a false sense of precision, with NTG values calculated to the second decimal place. Few of the NTG techniques currently in use warrant this level of precision, especially when results are estimated values based on the counterfactual.

- Most experts say there are no hard and fast guidelines for the appropriate level of spending for net-to-gross analysis for various types and sizes of programs. Ideally, more budget might support the use of some alternate methods to support triangulation and improve estimates. If other methods are too costly, however, the best use of extra budget may be to increase the sample size for an existing approach rather than go for a more expensive technique. Increasing the sample size for the self-report approach might allow for separately calculated NTG values for specific measures or customer groups within a program, for example.

- For some types of programs (e.g., low income, hard-to-reach small business, new programs), stipulated NTG values of 1.0 or close to 1.0 are viewed as appropriate by the literature and by market experts. For other types of programs, stipulated values drawn from research in other states/service territories are generally considered acceptable as placeholder values until a program or region-specific value can be calculated. All the experts interviewed, however, say that generally this should be only a short-term

solution, and that spending on research to calculate NTG is a legitimate evaluation activity that should be built into overall evaluation costs.

- There is consensus that spillover is real and should be accounted for, and experts and the literature agree that NTG should not be calculated simply as one minus free ridership.
  - Participant spillover is estimated more often than non-participant spillover, but is generally of a smaller magnitude, rarely adding as much as 5% to the net-to-gross value. Estimated values for non-participant spillover vary widely, ranging from less than 5% to more than 1.0 (i.e., non-participant spillover savings exceed participant gross savings.)
  - While adders and deemed values are acceptable as a placeholder value to acknowledge spillover, region- and program-specific research should be conducted as soon as practicable to develop a more defensible spillover value.
  - The California 5% adder to account for non-participant spillover did not come from research or program evaluation results, but is instead intended as a conservative placeholder value. The filing submitted by the California IOUs in support of non-participant spillover cited numerous studies to justify a larger spillover value, but the CPUC Energy Division and its consultants disputed the relevance of many of those results, and the CPUC limited the amount of spillover to a 5% adder across the board until California-specific studies could be completed. The 5% is added to whatever NTG value is calculated, so that a calculated NTG of 0.6 would become 0.65, for example.

ES.2 RECOMMENDATIONS

Based upon suggestions from interviewed experts as well the literature review, we make the following recommendation to Energy Trust regarding its estimation of NTG values.

Considering the almost universal application of the self-report approach across the country, Energy Trust can continue its use of Fast Feedback surveys and the calculation of free ridership using a self-report algorithm. However, we suggest Energy Trust consider several adjustments or modifications to validate the results being obtained using this method.

- One option would be an alternate self-report approach using another instrument and algorithm, such as the basic format employed in California[1]. If an alternate technique was used for part of the survey population while the basic Fast Feedback instrument was used for the rest, the results could be compared.
- As one of the experts suggested, it might be appropriate to incorporate consistency checks into the existing Fast Feedback survey to address clear discrepancies in program influence/importance between the two free ridership components (stated intent /project

[1] http://www.energydataweb.com/cpucFiles/pdaDocs/910/Nonresidential%20NTGR%20Methods%202010-12%20101612.docx

change and program influence.) Currently, the free ridership algorithm accepts responses that the program resulted in no change in the project (FR=1) and that one or more elements of the program were extremely influential (FR=0) and generates a free rider score of 50%. We recommend that such contradictory responses lead to a clarifying question to resolve the inconsistency. For example: "You said that the program resulted in no change in the project and that your organization would have made the funds available anyway, suggesting that the project was not very influential. Yet you said that an Energy Trust funded study was extremely influential in the decision to incorporate energy efficient features in your project. Can you please clarify?"

- For specific measures where appropriate sample is likely to be available, it might be possible to try one of the quantitative techniques described in this report as an alternate estimate of free ridership. We recommend the discrete choice modeling approach over the less-vetted LCDC and SEM approaches.
- Consider adding a short battery of participant spillover questions; initially to determine whether there is evidence of any spillover and subsequently to try to quantify that spillover.
- For commercial and industrial customers with larger projects, start the survey with an open-ended question to get a "story" behind the project and the selection of energy efficient equipment.
- Because Energy Trust's self-report algorithm uses two components to score survey responses, consider including Cronbach's alpha test to test for internal consistency in responses.[2]
- Finally, while Fast Feedback already addresses the issue of timely contact with program participants, explore the possibility of methods that could reduce the time between the customer purchase decision and survey administration even further by collecting customer email information on rebate forms and leveraging web surveys, either via email, or in-store (for retail measures).

Other suggestions from the experts we interviewed focus on the net savings analysis framework and planning:

- Experts expressed the need to develop stakeholder consensus around net savings analysis and the methods used. To ensure transparency and stakeholder consensus, Energy Trust could develop a set of principles and guidelines that provide the theoretical background to net savings approaches adopted by Energy Trust, similar to that developed by NEEP.
- Develop a targeted approach to calculating NTG estimates, and spreading evaluations out over several years. First, categorize programs by their magnitude of impact and the

[2] Cronbach's alpha is a measure of internal consistency of a test or a scale. It is frequently used to measure reliability of data collection instruments such as a survey. Cronbach's alpha is expressed as a number between 0 and 1 and the value indicated the degree to which the components of a test measure the same concept or construct. In essence Cronbach's alpha tests the correlation of the components of the test with each other. The higher the level of correlation among the components, the more reliable the test is assumed to be.

potential for free ridership, spillover, and market effects. Then, based on these categorizations, Energy Trust could develop a priority of programs, focusing resources on the highest priority programs, sectors or measures first to develop calculated estimates of net savings components for these.

- As an example, a spillover analysis to supplement the free ridership information already being collected might be done for high-impact measures covering several residential end uses, such as linear LEDs for lighting and furnaces or ductless heat pumps for HVAC. The Fast Feedback survey could be used to capture participant spillover data, while a large scale non-participant survey could be used for non-participant spillover.
- For smaller budget programs or lower priority items, review data available from other regions to identify similar programs/markets, which would support the use of adders and deemed values in lieu of calculated values.

For non-participant spillover, Energy Trust has been using a 1% adder for all programs except Existing Buildings, where it applies a 7% non-participant spillover value based on the results of a previous survey of non-participants[3]. An update of that survey is planned, which presumably will lead to a new non-participant spillover value. As Energy Trust wrestles with the appropriate approach to spillover, we make the following recommendations based on input from the literature and expert opinion.

- Given the "political" difficulty of spending evaluation resources to survey customers or trade allies in comparison regions, Energy Trust should continue to focus on approaches that rely on data collected from its own customers and trade allies who serve those customers.
- Where appropriate, data from other regions might be useful. This might involve using values developed for a portion of another state rather than the entire state. For example, while California-wide climate data would be very different from Oregon for heating and cooling measures, the Pacific Gas and Electric territory in Northern California would be relatively similar.
- Consider use of Delphi panels for developing consensus around values for spillover and market effects, using outputs from other methods as inputs for the Delphi to achieve triangulation. This is a relatively inexpensive approach to confirming validity of results and potentially improving consensus and confidence in results.

The extent to which program presence and activity over the long-term influences customer behavior and attitudes does not seem to lend itself to easy measurement using survey methods, and we are unable to make a recommendation on how long-term program influence can be incorporated into current Energy Trust NTG estimates. However, it may be possible to use surveys or interviews to capture data on trade ally sales practices over time and link it to their length and intensity of involvement with Energy Trust programs.

---

[3] Personal communication from Energy Trust Evaluation Manager.

MEMO

Date: July 5, 2017

To: Board of Directors

From: Philipp Degens

Subject: Staff Response to the Current Methods in Free Ridership and Spillover Policy and Estimation Report

In 2016, Energy Trust undertook research to examine the state of methods for free ridership and spillover policy and estimation. The goal was to determine if there have been any methodological advances that Energy Trust should consider implementing. The research by PWP and Evergreen Economics did not reveal any compelling new methods for estimating free ridership that would warrant moving away from the current self-report method employed by Energy Trust through Fast Feedback surveys. However, Energy Trust is considering making some of the survey changes that were recommended in the report, including a question that captures investment "story" and participant spillover questions. These changes will be made for surveys fielded in 2018, as major changes to the survey instruments are only made on an annual basis.

In the case of spillover, the research did not find any cost-effective methods for estimating non-participant spillover. Energy Trust will consider using a deemed estimate of non-participant spillover that is consistent across programs once a number of California spillover studies are completed in 2017. These studies will help inform Energy Trust and provide a basis for developing reasonable deemed estimates of non-participant spillover. In the case of participant spillover, we will consider adding a question or two to the Fast Feedback surveys in 2018 to determine whether there is any compelling evidence of spillover.

Energy Trust staff will continue to brainstorm and discuss other possible ways to estimate free ridership and spillover. Of special interest is to determine if, or how, long-term programs and repeat participation affect respondent statements about free ridership and spillover.

1. Introduction

This report presents the results of research conducted by PWP and Evergreen Economics (the PWP team) for Energy Trust of Oregon to assess policies and methods used to estimate free ridership and spillover of energy efficiency programs that are currently being used by organizations nationally. Energy Trust is particularly interested in how long-term relationships with customers and the long-term presence of Energy Trust programs in markets might affect free ridership and spillover estimates, and how others are dealing with these issues.

Free-ridership and spillover along with other terms related to net savings are defined differently across the industry, as we will illustrate in this report. However, for to help the reader of this report we provide below key terms as defined by Energy Trust (except for non-participant spillover and market effects which are definitions provided by the evaluation team):[4]

- **Gross Savings:** Gross savings represent all savings from program participants, regardless of whether they are free riders. Energy Trust reports all savings in net terms, not gross terms, unless otherwise stated in the publication.
- **Net Savings:** Savings that are adjusted for evaluation factors of free riders, spillover and savings realization rates. Energy Trust reports all savings in net terms, not gross terms, unless otherwise stated in the publication.
- **Net-to-Gross:** Net-to-gross ratios are important in determining the actual energy savings attributable to a particular program, as distinct from energy efficiency occurring naturally (in the absence of a program). The net-to-gross ratio equals the net program load impact divided by the gross program load impact. This factor is applied to gross program savings to determine the program's net impact.
- **Free Rider:** This evaluation term describes energy efficiency program participants who would have taken the recommended actions on their own, even if the program did not exist. Process evaluations include participant survey questions, which lead to the quantification of the level of free rider impacts on programs that is applied as a discounting factor to Energy Trust reported results.
- **Spillover (Participant):** Additional measures that were implemented by the program participant for which the participant did not receive an incentive. They undertook the project on their own, influenced by prior program participation.
- **Spillover (Non-Participant):** Additional energy savings achieved when a non-participant implements energy efficiency measures or behavior as a result of the program's influence (for example, through exposure to the program) but did not participate in the program and therefore are not accounted for in program savings.

[4] Definitions taken from "Glossary of Terms Related to Energy Trust of Oregon's Work" provided as part of December, 2016 Energy Trust of Oregon Board Meeting Packet. https://www.energytrust.org/wp-content/uploads/2016/12/December_2016_Board_Packet.pdf, with the exception of Non-participant spillover and market effects.

- **Market Effects:** A change in the market structure or in behavior of participants in a market that leads to increased adoption of energy efficiency measure, services, or behavior and can be attributed to program market intervention.

Energy Trust currently tracks and reports annually on free ridership rates at a variety of levels. The most basic level is by fuel (i.e., gas and electric) and program (e.g., Existing Buildings, Production Efficiency). In some cases, Energy Trust tracks the free ridership rates at the program track (e.g., prescriptive, custom, lighting) or measure (e.g., heat pump, ceiling insulation) level. Tracking is done through Energy Trust's Fast Feedback survey and thus uses customer responses to determine free ridership (the self-report approach). This survey is fielded on a continuous basis with a random sample of select program participants that completed a project in the previous month for each strata of interest. The sample sizes are set to obtain 10% precision or better with 90% confidence on a quarterly basis, however, this is not achieved in all instances. The survey is short (5-7 minutes) and delivered via telephone. Each group (e.g., Existing Buildings lighting, Existing Homes windows) has its own survey instrument that obtains feedback on investment decisions (i.e., free ridership questions) as well as satisfaction with different aspects of the program, measure, and/or service. Approximately 2,000 customers are surveyed each year. Since Q2 2010, Fast Feedback surveys have been completed with participants from most major Energy Trust programs. The basic free ridership algorithm – i.e., the scoring and weighting of responses to calculate a free ridership score for each respondent – has been in use since 2010.[5]

Free ridership rates are not estimated for some customer classes, services and equipment categories by design; in these cases, a free ridership rate of zero (net to gross ratio of 1.0) is stipulated and reported.

For example, free ridership is not estimated for program offerings that have income restrictions. The reasoning is that households with limited means already face difficulties in investing in basic energy services or new equipment, much less the additional costs of services focused on energy efficiency and higher efficiency equipment.

In addition, Energy Trust does not estimate free ridership for some measures and services, for example, if the market does not already offer measures or services, or if product availability was achieved through Energy Trust efforts. In these cases, if Energy Trust had not been providing the measure or service there would not have been an energy efficient market alternative.

---

[5] Energy Trust Free Ridership Methodology, Energy Trust of Oregon. http://www.energytrust.org/wp-content/uploads/2016/12/Energy_Trust_Free_Ridership_Methods.pdf

Energy Trust has conducted limited research on spillover, which led to large and highly uncertain estimates for many products. Since Energy Trust has not found highly precise or reliable approaches to spillover evaluation, it has limited its investment in this area and thus claims very modest spillover, as detailed later in this report[6].

Energy Trust and the Northwest Energy Efficiency Alliance (NEEA), extensively employ market transformation-based estimates of savings for appropriate programs. These estimates are not part of the scope of this review, because Energy Trust's spillover rates are specific to the influence on the market of specific resource acquisition programs and measures. Market transformation programs are designed to influence overall market preferences and codes and standards through a more systematic effort to address market barriers to efficient products and services becoming the preferred product in markets. While the distinction can never be clear, it is important; spillover attempts only to measure influences of specific interventions.

Energy Trust reports Net Savings in public reporting of Energy Trust program results, as well as for future program design and program savings forecasting. As noted above, net savings includes adjustment of gross savings for free riders and spillover, as well as other factors such as transmission and distribution factors. The adjustments are made, retroactively, at the end of each program year. Energy Trust conduct a True Up process at the end of each program year to make changes to savings adjustment factors based on any new data or studies throughout the program year.

1.1 RESEARCH OBJECTIVES

The basic goals of this research project were to:

- Determine what types of policies are currently used in regard to free ridership and spillover.
- Determine what methods are being used by organizations around the country to estimate free ridership and spillover rates, as well as obtaining a clear overview of the strengths and weaknesses associated with each method, including the costs of gathering and estimating free ridership and spillover, the accuracy and variance of estimates, the sample sizes needed, how or if the effects of long-term customer participation or program market presence are being incorporated into the estimates, etc.
- Develop actionable recommendations that will allow Energy Trust to improve its current free ridership and spillover methods and estimation process. An important consideration is that methods should have reasonable costs for the level of reliability provided relative to the overall budget and scope of the program.

---

6 Personal communication from Energy Trust Evaluation Manager

2. Methodology

To address the above goals, the evaluation team relied on an extensive literature review including review of Energy Trust program materials, and telephone interviews with experts.[7] Each of these data sources is discussed below. In addition, the evaluation team consulted with Energy Trust staff for their invaluable input.

2.1 LITERATURE REVIEW

The literature review focused on current methods in free ridership and spillover policy and estimation. The goal of the literature review task was to find, review, and summarize existing information from secondary sources that shed light on the use of net savings estimation methods across the nation, including:

- Recently developed methods of net savings estimation
- Areas of consensus and disagreement on net savings estimation methods
- Regional use of net savings estimation methods
- Use of net savings estimation methods across sectors (residential, commercial, industrial)
- Net estimation methods used in other industries
- Adoption of prescribed versus calculated net savings approaches by state, and rationale for these approaches

The research team reviewed a wide range of literature, sourced from journal and conference papers, program evaluation reports and other industry studies. Results of the literature review are presented in Chapter 3. A full list of sources can be found as Appendix A to this report.

2.2 EXPERT INTERVIEWS

Based upon the results of the literature review, an interview guide was developed and refined using input from Energy Trust staff. It is included as Appendix B. Interviews were conducted with eight national experts on free ridership and spillover analysis, including program evaluators as well as consultants who provide key decision support to utilities and regulators.

Respondents were asked to comment on:

- Spillover definition, estimation methods, and values calculated
- Effect of long-term program presence on and methods for calculating:

[7] Energy Trust program materials reviewed included documentation on current free ridership methods used by Energy Trust and reviews of recent Fast Feedback reports.

    - Free ridership
    - Participant spillover
    - Non-participant spillover/market effects
- Free ridership and spillover policies
- Free ridership and spillover estimation methods
- Challenges of collecting the required data
- Accuracy of estimates
- Costs and benefits of specific methods
- Frequency of estimation
- Sampling requirements
- Level of estimation (sector, program, track, customer class, technology/service) as well as criteria for determining the level of estimation needed
- Trends in methodology and policy in different jurisdictions, regions, and nation-wide
- Relationship of new methods and trends to policy framework

Respondents were also asked to identify any additional free ridership and spillover documents and research reports, as well as any other experts that should be interviewed as part of this research project. The experts interviewed each represent decades of experience conducting and refining net-to-gross (NTG) research, including the most recent advances in the state of the art.

3. Literature Review

This chapter presents results of a literature review conducted as part of the review of current methods in free ridership and spillover policy and estimation. The goal of the literature review task was to find, review, and summarize existing information from secondary sources that shed light on the use of net savings estimation methods across the nation, focusing on:

- Recently developed methods of net savings estimation
- Areas of consensus and disagreement on net savings estimation methods
- Regional use of net savings estimation methods
- Use of net savings estimation methods across sectors (residential, commercial, industrial)
- Net estimation methods used in other industries
- Adoption of prescribed versus calculated net savings approaches by state and rationale for these approaches

We reviewed a wide range of literature, sourced from journal and conference papers, program evaluation reports and other industry studies. We also reviewed recent reports and professional papers to examine current trends in the industry. A full list of sources is in Appendix A.

In the remainder of this chapter, we first present high-level findings. Then, based on our review of the literature, we discuss both methods that are well established and those that are still developing. We conclude the literature review with proposed next steps for additional focused literature review and expert interviews.

3.1 KEY FINDINGS AND SUGGESTED AREAS OF FOCUS FOR CURRENT STUDY

The literature review began with a look at established methods to evaluate net savings, including:

- Billing analysis with randomized control trial (RCT)
- Billing analysis with quasi-experimental design
- Survey-based approaches
- Standard market practice analysis
- Market sales data analysis
- Top-down evaluations (macro consumption models)
- Structured expert judgment approaches
- Deemed or stipulated net-to-gross ratios
- Historical tracing or case study methods

In addition, we found information on the following new or revised approaches to net savings estimation:

- Discrete choice model / revealed preference model

- Latent class discrete choice (LCDC) model
- Price response model
- Equipment inventory tracking
- Revenue neutral sales model
- Real-time survey data

Not surprisingly, both established and new approaches have associated benefits and disadvantages, as well as suitability for specific application to the components of net savings, free ridership, spillover, and market effects. Moreover, while approaches vary significantly across states and sectors, the most commonly used approaches are:

- Survey-based approaches
- Billing analysis
- Standard market practice analysis

Efforts to calculate external spillover are generally complex, expensive and most effective for "widget"-type markets. In many jurisdictions, policies instead recognize spillover either by setting NTG to 1, so that spillover is assumed to offset free ridership, or by stipulating an "adder" to the calculated value of one minus free ridership; e.g. 5% in Hawaii and California.[8][9]

In sum, there is little consensus around methodological approaches. There are, however, other areas of consensus regarding the calculation and application of net savings. These include:

- NTG should not be just one minus free ridership, but should incorporate participant and non-participant spillover
- States should use multiple methods and triangulation to estimate free ridership and spillover
- Net analysis should be used to improve program design
- Deemed values are acceptable, but should be based on solid research and updated periodically
- Results should not be used retroactively to penalize program savings (while this was generally agreed to in theory, many jurisdictions do, in fact, apply NTG ratios retroactively)

---

[8] Decision Approving 2013-2014 Energy Efficiency Programs And Budgets. Decision 12-11-015 November 8, 2012.

[9] Net-to-gross Issues in Hawaii Energy Efficiency Programs: Challenges, Near-term Options, and a Longer-term Approach. Evergreen Economics. 2013.

3.2 METHODS FOR NET SAVINGS ESTIMATION

As noted in the Uniform Methods Project (UMP) chapter on estimating net savings[10] and the ACEEE 2014 national survey of net savings methods[11], and as becomes obvious after even a cursory review of the relevant literature, there is a wide range of approaches to the issue of calculating net savings. While there is consensus on some issues, there remains little overall consensus on which methodologies are most appropriate for determining net savings. In fact, there remains a lack of consensus on whether net savings is a metric that should be used, and even the definition of net savings is unsettled (ACEEE 2014; UMP 2014). This section provides an overview of the commonly used methods of net savings estimation, including estimating core components of net savings, free ridership, participant and non-participant spillover, and market effects, as well as new methods under development or fresh approaches to existing methods.

3.2.1 Established Methods

The UMP, ACEEE 2014, as well as other recent summaries of net savings methods provide a thorough summary of well-established methods. The following two tables present a summary of these established methods that draws primarily on these sources. Table 1 presents a description and applications for established methods and Table 2 presents benefits and concerns of each method.

10 *Estimating Net Savings: Common Practices*. Daniel M. Violette, Navigant; Pamela Rathbun, Tetra Tech. 2011.

11 *Examining the Net Savings Issue: A National Survey of State Policies and Practices in the Evaluation of Ratepayer-Funded Energy Efficiency Programs. ACEEE. 2014*

Table 1: Established Methods of Net Savings Estimation and Areas of Application
(FR = Free Rider; PSO = Participant Spillover; NPSO = Non-Participant Spillover; ME = Market Effects)

General Method	Description / Variations	FR	PSO	NPSO	ME
Billing Analysis with Random Control Trial (RCT)	The commonly accepted evaluation "gold standard", this approach uses a randomized control trial approach to estimate savings controlling for free ridership, spillover, and potentially market effects. Three common approaches: • *Simple post-period comparison*: Customers assigned randomly to treatment and control group. Simple comparison of pre-period and post-period consumption, difference assumed as net savings. • *Difference-in-differences (DiD): C*ompare the change in energy use for the two groups between the pre- and post-participation periods. • *Billing regression:* Typically a linear fixed effects regression model using customer energy bills and comparing pre- and post-installation periods. Controls for other factors.	Yes – but not able to separate FR and PSO		No	Possibly
Billing Analysis with Quasi-Experimental Design	Quasi-experimental designs are similar to RCTs, except that random assignment is not possible. Instead researchers develop a comparison group. • Regression Discontinuity Design: Comparison group is developed by matching participants from different program periods. • Random Encouragement Design (RED): A randomly selected group of participants is given greater encouragement to participate.	Yes – but not able to separate FR and PSO		No	Possibly
Survey Based Approaches (Self Report Approach or SRA)	This approach relies upon self-reported survey data where actors are asked counterfactual questions (e.g., "what would you have done if there had been no incentive?") to derive a subjective measure of free ridership and spillover. They can be separated into Standard and Enhanced approaches with enhanced approaches utilizing a longer battery of questions and more elaborate scoring algorithm. Surveys include: •Program Participant Surveys •Program Non-Participant Surveys: used to triangulate participant self-report responses and collect data for calculating nonparticipant spillover or market effects. •Market Actor Surveys - used to examine the effects of upstream influences and possibly market effects.	Yes	Yes	Yes	Yes
Standard Market Practice (Common Practice Baseline Approaches)	Identifies the standard market practice baseline for custom projects and specific measures where possible. It is primarily used to calculate a net-to-gross ratio for custom programs, such as those tailored to industrial customers. The standard practice baseline is often determined through a combination of customer surveys, contractor interviews, and an engineering review of specific projects. Using a statistically representative sample, an average net-to-gross ratio for the entire program can be calculated.	Yes	Yes	No	Yes

General Method	Description / Variations	FR	PSO	NPSO	ME
Market Sales Data Analysis	Can capture the total net effect of the program, including both free ridership and participant and nonparticipant "like" spillover. Common approaches include: • Cross-sectional comparison area method in which post-program data are compared with data from a non-program comparison area (or multiple comparison areas) for the same point in time. • Panel interview with participant and non-participant trade allies	Yes	Yes	Yes	Yes
Top-Down Evaluations (Macro-consumption Models)	Uses macroeconomic data on energy consumption in a model that relates changes in energy consumption to a measure of EE effort (usually expressed as EE expenditures). Top-down approaches have appeal because they directly address overall net savings.	Assesses overall change in market incorporating FR, SO, ME but not able to separate these factors			
Structured Expert Judgment Approaches	A panel of experts is asked to estimate baseline market share for a measure or behavior. In some cases, they are also asked to forecast market share with and without the program in place. Structured expert judgment processes use a variety of specific techniques to ensure that the panel of experts specify and take into account key known facts about the program, the technologies supported, and the development of other influences over time. Common method: Delphi Panel Method	Yes	Yes	Yes	Yes
Deemed or Stipulated Net-to-Gross Ratios	Applies pre-determined values and that do not rely on a calculation- based approaches. Deemed values are often based on previous NTG research that was conducted using at least one of the other methods described in this chapter	Yes	Yes	Yes	No
Historical Tracing or Case Study Methods	Involves reconstructing the events (such as the launch of a product or the passage of legislation) that led to the outcome of interest. An example of this is developing a "weight of evidence" conclusion about the specific influence a program had on the outcome.	Yes	Yes	Yes	Yes

Table 2: Established Methods of Net Savings Estimation Benefits and Concerns

General Method	Benefits	Concerns
Billing Analysis with Random Control Trial (RCT)	• Random assignment reduces and limits bias in estimates • Increases reliability and validity • Controls for free riders and participant spillover • Widely accepted in natural and social sciences as the gold standard of research designs	• Difficult to implement and needs to be planned as part of program implementation to allow for appropriate randomization of program participants and a control group • Bias can result if random assignment occurs among volunteers or if the program drop-out rate differs by key characteristics • Does not address nonparticipant spillover • Equity/ethical concerns about assigning some ratepayers to a control group and not allowing them to participate in the program for a period of time • Generally not applicable to programs that involve large investments in measures and services • Participants in some C&I programs may be unique with few control group candidates
Billing Analysis with Quasi-Experimental Design	• Limits bias if matched comparison group can be identified based on actions influencing energy use • Unlike RCT, can be applied after program implementation • Increases reliability and validity • Controls for free riders and participant spillover • Widely accepted in natural and social sciences when random assignment cannot be used	• May be difficult to identify a matched comparison group if there are unobservable variables that affect energy use • Does not address nonparticipant spillover • Some C&I programs may have unique participants and few control group candidates
Survey Based Approaches (Self Report Approach or SRA)	• Can provide useful information to support process and impact evaluations (for example, source of awareness, satisfaction, and demographics) • Flexible approach that allows the evaluator to tailor questions to the program design or implementation methods • Can yield estimates of free ridership and spillover without the need for a nonparticipant control group	• Potential biases related to respondents' giving "socially desirable" answers • Consumers' inability to know what they would have done in a hypothetical alternative situation, especially in program designs using multiple methods to influence behavior • The tendency of respondents to rationalize past choices • Potential arbitrariness of scoring methods based on evaluator judgment that translate responses into free rider estimates • Consumers may fail to recognize the influence the program may have had on other parties who influenced their decisions (for example, program may have influenced contractor practices, which in turn impacted the participant) • Participant surveys capture only a subset of market effects
Standard Market Practice (Common Practice Baseline Approaches)	• Can help to avoid double counting of free ridership in circumstances where gross impacts incorporate some net savings factors • Can be used in upstream and market transformation	• Self-selection bias is not addressed and methods for addressing self-selection are not readily apparent • Does not capture nonparticipant spillover • Common practice baselines for measures and technologies will change over time

General Method	Benefits	Concerns
	programs • Can be applied market-wide	and require updating • Determining average market practice has accuracy challenges • Approach has been applied in the Pacific Northwest, along with other net savings estimation methods, but is relatively new and evolving as a general net savings estimation method
Market Sales Data Analysis	• Can estimate the total net effect of a program • Uses information on actual consumer behavior • Addresses trends in an entire market • Most appropriate for programs that promote a large numbers of homogeneous measures and have substantial influence upstream	• There may be a low availability and quality of sales and shipment data in the area of interest and in an appropriate comparison area(s) • Data may be expensive to acquire and/or may have gaps that can be misleading • May be difficult to determine the appropriateness of a comparison area
Top-Down Evaluations (Macro-consumption Models)	• Estimates net effects of all programs cumulatively • No need to adjust for free ridership, spillover, or market effects at the aggregate level	• Methods are not fully developed at the state or regional levels • Relies on high-quality energy consumption data and on data regarding EE efforts within each cross-section analyzed • Cannot provide savings at the measure, technology, or program level • Does not provide information on how to improve program design and implementation processes
Structured Expert Judgment Approaches	• The resulting estimate is the independent, professional judgment of a group of technology and/or market experts • It is a useful approach for programs with diverse and complex end uses or practices • Useful tool for consolidating results from multiple methods to develop a consensus estimate • Panel members can provide levels of confidence and procedures using appropriate elicitation methods	• The approach relies on high-quality data to inform the panel, leading to reasonable estimates of net savings • Sampling-based calculations of confidence and precision are not available
Deemed or Stipulated Net-to-Gross Ratios	• This approach can reduce contentious after-implementation adjustments to estimated program savings because agreed-upon net savings factors are developed in advance of program implementation	• An incorrect estimate can be deemed • It is not based on program-specific information • The evaluator cannot assign sample-based statistical precision to the estimate •. Developing deemed savings net values at the measure and technology levels can be

General Method	Benefits	Concerns
	• Cost savings since a policy decision does not require costly estimation studies	time consuming and expensive • The process for developing deemed net savings can be contentious
Historical Tracing or Case Study Methods	• Draws from multiple information sources • Can be used at market level for upstream EE programs • Can be useful for making a persuasive case for attribution and provide evidence to support a statistically derived net savings estimate	• It can be difficult to translate the influence factors into estimates of impacts without additional modeling • The evaluator cannot calculate sample-based statistical confidence and precision levels for the estimate

* Source: UMP; ACEEE 2014; Iowa Energy-Efficiency Net-to-Gross Report;

3.2.2 New or Revised Approaches to Net Savings Estimation

The literature review also identified continuing research in net savings estimation methods that results in new approaches or fresh approaches to existing methods. In this section, we detail some of the more promising new approaches found in the literature.

3.2.2.1 Econometric Modeling

Aside from billing regression analysis, econometric modeling does not receive a lot of attention in the UMP or other reviews of net estimation approaches. However, econometric modeling is an area that shows promise and has been used to evaluate net savings in several jurisdictions and across several program types.

Discrete Choice Model / Revealed Preference Model

Discrete choice models have been used in net savings estimation across a variety of programs, although most were some time ago. Some illustrative examples include the California Upstream Lighting Program (2010) [12], 2006-08 impact evaluations covering the nonresidential high impact measures for lighting across multiple programs targeting small commercial customers,[13] [14] and a high efficiency gas furnace program conducted in 1997.[15] In their 2010 National Review of Best Practices and Issues in Attribution and Net to Gross, Skumatz and Vine suggest that discrete choice models should be considered as they "help to address issues of imperfect control groups, unobserved factors, etc. to allow improved estimates of attributable impacts."[16]

The goal of a discrete choice model is to model choices made by customers among a finite set of discrete alternatives, which can then be used to describe, or predict these choices. In a net savings application, the approach compares measure adoption decisions of program participants to those of non-participants to examine how customers make choices for energy efficiency measures, identifying key factors that influence these choices such as programmatic, demographic, attitudinal, economic, and other factors. The models can be used to assess the net impact of the program by estimating the number of eligible customers that change their behavior in response to a given situation such as a new program or action – to "separate program participants into true

---

[12] Final Evaluation Report: Upstream Lighting Program. Volume 1. Prepared by: KEMA, Inc. 2010.

[13] Small Commercial Contract Group Direct Impact Evaluation Report. Prepared by: Itron, Inc. 2010.

[14] Free to Choose? A Comparison of a Nested Logit Model with a Billing Regression Model and Self-Report Analysis in a Commercial Impact Evaluation. Grover et al. IEPEC. 2011.

[15] Freerider and Freedriver Effects from a High-Efficiency Gas Furnace Program. Ken Seiden, Ph.D, Quantec, Portland, OR

Helen Platis, Union Gas Limited, Toronto ON. 1999.

[16] A National Review of Best Practices and Issues in Attribution and Net-to-Gross: Results of the SERA/CIEE White Paper. Lisa Skumatz, Skumatz Economic Research Associates; Edward Vine, CIEE. 2010.

participants and free riders and to separate nonparticipants into true nonparticipants and free drivers" (Seiden and Platis 1999).

For example, for a basic lighting rebate program an evaluator would collect actual data on customer purchases of lighting equipment, including eligible efficient equipment, and ineligible equipment. The evaluator would also administer a survey to both customers who received a rebate (participants) and customers who did not receive a rebate (non-participants), some of whom would have still purchased efficient equipment, to gather information on factors like program awareness, demographic data and other factors. Using these data a discrete choice model is specified and applied, typically using either a logistic or probit model. The model coefficients can be used to calculate the probability that a customer would purchase efficient equipment with the program rebate, or who is not a free-rider, and without the program rebate, or who is a free-rider.

Discrete choice models can also derive elasticities, which represent the percent change in participation in response to a given change in the market, such as a new or altered program (Skumatz and Vine 2010). Another advantage of the discrete choice analysis is that it is based on *revealed preference* data that demonstrate how customers actually behave in real life situations -- in contrast to *stated preference* data developed thru the self-report approach, where customers state how they intend to act in a given situation, but there is no guarantee that they will actually behave this way in real life.

The primary disadvantage to the discrete choice method for net impact estimation is that it is data intensive and requires a significant amount of data on purchases made outside the efficiency program, which can add substantial cost to the evaluation process. The data intensive nature of the method can also limit how detailed the analysis can be, with narrow focuses on individual small programs or customer segments within programs potentially resulting in too little data.

Latent Class Discrete Choice (LCDC) Model

A subcategory of discrete choice models is the latent class discrete choice (LCDC) model. LCDC modeling is based on customers' stated preferences that helps determine the factors or attributes that influence customers' choices. This method estimates what the uptake rate would be for programs dependent on the attributes of the program the program planner wishes to consider (e.g. rebate levels, eligible measures), while allowing for the possibility that different types of customers will have different preferences.

This approach presents measure purchase choices to customers in choice sets or bundles of products to mimic the real-world environment that consumers face when making choices. Based on their responses customers are assigned to groups, or classes, with other like customers based on their choice. Out of each choice set, a discrete choice model is specified and applied to the data. A simulator is produced from the estimated models that allows program planners to assess what the purchase rate (probability of uptake) would be for programs with any bundle of attributes/values the planner wishes to consider.

For example, in a basic LED lighting rebate program, potential customers are surveyed and presented with multiple-choice sets of purchasing scenarios based on possible program designs (Table 3).

Table 3: Example LCDC Choice Sets

Choice Set 1	Choice Set 2	Choice Set 3
Longest Life LED; Cost $30; Rebate $10	Longest Life LED; Cost $25; Rebate $5	Medium Life LED; Cost $20; Rebate $5
Medium Life LED; Cost $20; Rebate $5	Longest Life LED; Cost $30; Rebate $0	Longest Life LED; Cost $30; Rebate $10
Medium Life LED; Cost $20; Rebate $10	Medium Life LED; Cost $20; Rebate $5	Shortest Life LED; Cost $15; Rebate $0
None of the Above	None of the Above	None of the Above

Similar to the discrete choice model, these choice data along with other data on customer characteristics are modeled using a logistic or probit regression specification. The results are used to calculate the probability of measure purchases based on the choice bundle attributes and other customer characteristics. The probabilities can then be used to determine net savings rates.

This approach has not been extensively used, with the only uses we identified being in an evaluation of the Southern California Edison (SCE) LED Ambient Lighting Program in 2012, by Opinion Dynamics, and a 2006 evaluation of the SCE Home Energy Audit Program (Ridge et al. 2006).[17] A paper by Opinion Dynamics staff in 2015 applied the modeling approach to a California financing program (Galhotra and Randazzo 2015).[18] Unlike the discrete choice model, which is based on consumer revealed preferences, the LCDC model is based on consumer stated preferences. The method involves customers completing surveys and, in the case of the LED lighting program, participating in mock shopping trials to segment customers into groups based on their product attribute preferences using probability modeling.

The LCDC in theory can account for free ridership as well as participant and non-participant spillover.

The benefits of the LCDC model are that it is a flexible approach that can be used to test a variety of specific program features that might influence free ridership or spillover, it can make predictions for an overall program as well as for defined customer segments of interest, and it obtains attribution answers based on comparison of realistic market conditions outside of the program. The disadvantages of the LCDC method are:

---

[17] The Southern California Edison (SCE) Advanced Light Emitting Diode (LED) Ambient Lighting Program Customer Preference And Market. Opinion Dynamics. 2012.

[18] *Show Me the Attribution: Recommended Methods for Estimating Impacts for Finance Programs*. Jeevika Galhotra, Opinion Dynamics Inc., Oakland, California Katherine Randazzo, Opinion Dynamics Inc., San Diego, California

- The LCDC model is a relatively new and somewhat untested method in the energy efficiency industry.
- The results are based on stated, not revealed, preferences, introducing potential bias similar to self-report approaches.
- The LCDC method is a high-cost method.

Price Response Model

Price response models, also known as elasticity studies, have been used on occasion to estimate program spillover, free rider effects, and market effects or market transformation. Notable instances of use include a 2001 study of market transformation indicators in the national energy efficient lighting market[19], a 2007 study of long-term market effects of energy efficiency policies in the residential, commercial and industrial sectors nationwide[20], and a 2013 study of Maine and California lighting programs.[21] In Maine, Cadmus estimated the elasticity of demand for CFL lighting products for two utilities operating in the state in 2012 and 2015. Cadmus used tracking data collected by the program implementer. The data were then passed through a pricing model in which the dependent variable is the quantity of bulbs sold per month and the independent variables are price, CFL characteristics, and a monthly dummy variable. The price response model predicts purchases at different price points, estimating increased sales due to program markdowns. Free ridership can be estimated by subtracting the total reduction in sales from the sales that would have occurred without the program. Spillover is estimated by predicting the impact of promotional pricing on sales, using the same model. The difference in sales between this hypothetical price scenario and the prices customers actually encounter represents net sales attributable to the program, or the program's lift. The lift comprises two components: (1) incentives provided by the program and (2) additional price effects related to retailer discounts. Cadmus claims that these additional price effects can be viewed as program spillover (Song et al. 2013). Free ridership in 2015 was calculated as 20.6 percent for standard CFLs using this method. The spillover effect was estimated at less than 1 percent.

3.2.2.2 Surveys with Saturation Study Participants

As part of the 2013-2014 California Residential Program Evaluation, Opinion Dynamics developed a new approach to estimating non-participant spillover by leveraging two recently completed equipment saturation studies, the California Lighting and Appliance Efficiency

---

[19] Economic Indicators of Market Transformation: Energy Efficient Lighting and EPA's Green Lights. 2001. Horowitz, M. The Energy Journal, Vol. 22, No. 4.

[20] Energy Efficiency and Changes in Energy Demand Behavior. Demand Research LLC. Marvin J. Horowitz. ACEEE 2008.

[21] *Estimating Spillover in Upstream Lighting Programs: Hard Data for an Elusive Number*. Louise Song, Joshua Keeling, Eric Rambo, Andrew Carollo, Jason Christianson. 2013.

Saturation Study (CLASS) and the Commercial Saturation Survey (CSS).[22] This approach estimates spillover by conducting surveys with households that completed the CLASS or businesses that completed the CSS surveys. This information in the CLASS or CSS data gives the evaluator the knowledge of whether the customer installed an eligible measure outside the program prior to conducting a survey. In the survey, customers are asked if the utility program influenced their decision to purchase the eligible equipment. Following the survey, the evaluator uses either engineering analysis or deemed values in the DEER database to quantify the spillover effect.

3.2.3 New Approaches to Established Methods

In addition to new (or relatively new) methods, there has also been work to refine and refresh existing methods.

Revenue Neutral Sales Model

The revenue neutral sales model is a new approach to market sales data analysis. Opinion Dynamics developed this approach as part of an evaluation of a lighting program in Delaware and it was reported as part of that evaluation as well as in an IEPEC conference paper.[23] The underlying assumption of the model is that retailers will avoid programs that are insufficient to stimulate enough additional sales to make up for lost revenues. When retailers participate in an upstream program such as the lighting program in question, and reduce the price of their bulbs, they typically experience a fall in gross revenue because the discount does not spur enough additional sales to cover the total value of the discount. While the retailers are reimbursed the discount, so that their profits are not affected, retailers still care about gross revenue because this can influence shareholders, and often store and corporate bonuses are tied to revenue. Through a reverse calculation, the program can determine the number of units sold prior to the promotion based on the assumption that the retailer would participate in the program only if the program sales kept the retailer revenue neutral. This can then be used as a baseline to measure the program-related sales against.

The advantage of this approach is that it is based on total retailer sales rather than a sample and likely avoids some of the pitfalls of other approaches, such as in-store intercepts and survey non-response. In general, the approach is also relatively low cost compared to other approaches.

---

[22] Spillover – Worth Crying Over: An Innovative Methodology for Quantifying Non-Participant Spillover in California. Chris Murphy, Itron; Dan Buch, CPUC; Vincent Greco & Mary Sutter, Opinion Dynamics. 2015

[23] *The Revenue Neutral Sales Model: A New Approach to Estimating Lighting Program Free Ridership.* Tami Buhr, Opinion Dynamics, Waltham, MA Stan Mertz, Applied Proactive Technologies, Springfield, MA

Disadvantages of the method are the requirements for extensive data that retailers may be reluctant to provide. In addition, this approach only estimates free ridership, not spillover.

Real-Time Survey Data

In their National Review of Best Practices and Issues in Attribution and Net-to-Gross, Skumatz and Vine note that there "have been several instances in which utilities have introduced NTG-surveys as part of the program participation documents and gather early feedback – near the point of actual decision-making – on the program's influence in adopting the measures" (Skumatz and Vine 2010).[24] In essence these are standard self-report surveys that are administered as close to the time-of-purchase as possible. This provides several benefits: it increases return rate and sample size (and eliminates the problem of finding participants after they have moved or after years of delay); provides on-going data and allows evaluation at virtually any point after the program is implemented to support continuous refinement of programs; significantly reduces the cost of surveying and evaluation; provides more accurate data if the point of feedback is close to decision-making (recall may be improved); and helps to sort out which programs had what degree of influence. This may be suited to education and behavioral programs as well as "widget" programs, but needs testing, as the approach has not been widely applied.

3.3 MARKET EFFECTS OUTSIDE SPILLOVER

While there is much discussion of market effects outside of spillover, there is little active work being done to develop methods to estimate these market effects. As we will note in a subsequent section, only two states, Massachusetts and Vermont, are actively pursuing estimation of market effects. In addition, one state incorporates market effects in a specific "adder" – more on this later. Despite this, we did identify some documents that address market effect estimation.

***Preaching to the Choir: Are Repeat Participants Free Riders?* Jeff Erickson, Summit Blue Consulting 2008. ACEEE Summer Study.**

This paper discusses the approaches of two utilities struggling with the question of repeat participants and free ridership. Specific questions asked in the paper are:

- How can program managers and evaluators accurately distinguish between repeat participants who would implement the target actions now without program assistance and those that genuinely need program assistance?
- How should the savings generated by repeat participants be counted?
- How much attribution should be given to prior program efforts and how should that be determined?

---

[24] Integrated, Real Time (IRT), on-going data collection for evaluation – benefits and comparative results. Gordon, Susan L. Skumatz, Lisa A. Susan L. Gordon Skumatz Economic Research Associates, Inc. USA. 2007.

Utility one entered into a study of their large commercial and industrial customers designed to identify effects of current and past interactions on participant behavior, with an understanding with their commission that "the study results would be used to improve the program, not to justify ignoring the issue of free ridership." The evaluator asked customers questions about the effect of current and past program involvement on their energy efficiency implementation decisions. This approach found that 56 percent of current year savings was program attributable, 8.5 percent was attributable to prior program experience and 35 percent was pure free ridership.

Utility two made no distinction between program years. In this case, program managers were concerned that respondents would not be able to distinguish effects across years. Utility two designed their survey questions to treat all program interactions over time as the program intervention, i.e. repeat customers were asked about the influence of all interactions not just the current year interactions, so that the utility could not tease out prior activity influence. This utility had funded various feasibility and audit studies prior to measure installation. Because of the approach taken, the utility was not able to understand the effect of the audit versus the measure installation, and in some cases where audits had occurred several years previously, customers may have internalized the lessons from the audit and attributed the actions they took to their own behavior rather than the audit, thus looking like free riders when they may not have been. The evaluation separated free rider savings into two buckets: savings from measures preceded by an audit and savings from measures not preceded by an audit. The result was that over half the free rider savings went to the audit bucket. This result indicates that if a customer is asked about measure installation only, and not the effect of the audit, those who received audits prior to measure installation could appear to be free riders, when in fact the information learned through the audit influenced their purchase decision. This further suggests that a self-report study asking people to report over a long time may not be able to accurately measure the complexities of programs that develop long-term relationships.

2012 Residential Heating, Water Heating, and Cooling Equipment Evaluation: Net-to-Gross, Market Effects, and Equipment Replacement Timing. Cadmus; Navigant. 2012.

The evaluation team estimated net market effects (NME) of two downstream HVAC programs, that provided incentives to customers for installation of high efficiency HVAC equipment, and provided contractors with education and training opportunities. To estimate NME the evaluation team compared the market structure and market actor behavior of the cooling, heating, and water heating markets between the program state, Massachusetts, and a comparison state, Pennsylvania. The evaluation team asked distributors in both states about changes in total equipment sales and program-qualifying product and non-qualifying product market shares, and further asked Massachusetts distributors about sales that occurred as a result of the two programs (and would have occurred in the absence of the programs). The data captured from surveys with distributors included equipment sales information, the change in sales over time, and distributor reports about how the equipment programs impacted these changes. The evaluators calculated the NME estimates by comparing Massachusetts and Pennsylvania distributors' self- reported estimates of what their market sales and efficiency shares from 2010-2012 would have been in the absence of the programs. However, the report notes that the small distributor survey sample sizes and the potentially "cumulative" nature of market impacts from the programs prior to 2010 mean the NME estimates play a qualitative, supporting role in the determination of Cool Smart and HEHE net savings.

The evaluators estimated NME values from a series of questions designed to capture the distributors' change in total sales due to the programs. For each end-use measure, the evaluator asked:

- What were your company's total sales over the last year?
- What were your company's sales (or the share of total sales) of high-efficiency equipment (as defined by program-qualifying thresholds)?
- How have these sales patterns changed relative to a hypothetical environment in which the programs did not exist?

Results were weighted by each distributor's reported total sales. Then the percent change in the market share of high-efficiency equipment, adjusted for the total change in the equipment sales' overall efficiency levels as a result of the programs, was multiplied by the annual appliance turnover for each measure. The result was the net increase in high-efficiency sales as a result of the programs. To obtain the NME, the team then divided the resulting net increase in high-efficiency equipment sales by the annual programs' participation for each measure in Massachusetts. In this case a value of one would indicate that the net market effect was equal to the program participation, meaning there are no broader market effects beyond the program. A value greater than 1 indicates evidence of broader market effects beyond the program. While the results were not statistically significant due to the small sample sizes, the results indicated that there is qualitative support for consequential market effects for high efficiency measures in excess of 1, and in some cases as high as 6.85, suggesting that spillover effects could be much larger than direct program impacts.

3.4 TRENDS

The process of reviewing the industry literature did not reveal any obvious new trends in approaches to estimating net savings. The most common approach remains self-report, despite continued criticism of the reliability of this approach. While there are no obvious trends, there are some regions and evaluators that are investigating new methods, with an apparent emphasis on econometric modeling such as discrete choice and price response modeling.

Free ridership remains the primary focus of most net savings estimation approaches, as evidenced by the prevalence of self-report methods, which are not always well suited to measuring spillover.

Spillover, both participant and non-participant, continues to present challenges, and many jurisdictions continue to either ignore these effects, or provide a stipulated value. California is a prime example. In 2012, the California Public Utility Commission (CPUC) opted to adopt a spillover "adder" of 5 percent in decision D.12-11-015 after negotiations with the Investor Owned Utilities (IOUs).[25] The IOUs initially proposed a 10 percent spillover effect adder based on past market effects studies from California, New York and other states. The CPUC and its consultants noted that very little research on spillover had been completed in California and were skeptical

[25] Decision Approving 2013-2014 Energy Efficiency Programs And Budgets. Decision 12-11-015 November 8, 2012

that a 10 percent value was accurate or appropriate. While the CPUC were skeptical of the 10 percent value, they expressed a belief that the concept of spillover is well proven and therefore found it appropriate to apply a portfolio level adjustment of 5 percent for 2013-2014, analogous to NTG ratios prior to completion of studies on program free ridership. They also committed to provide evaluation funding for spillover effects studies. The studies were due to be reported in Q2 and Q3, 2016 but were not publicly available as of January 2017.

Despite the inherent difficulty in estimating spillover, there are some promising approaches listed above. Specifically, econometric modeling such as the LCDC model and price response models, as well as more focused self-report approaches, may improve spillover estimation.

3.5 AREAS OF CONSENSUS

As noted previously, there is little consensus with regards to the best methods for estimating net savings and its principal components, free ridership, spillover and market effects. A 2012 panel of the Midwest Energy Efficiency Alliance explained, "the interplay of these methodologies is currently one of the most controversial issues with respect to EM&V both in the Midwest and nationally".[26] Despite the lack of consensus around methodologies, the 2014 ACEEE National Survey of State Polices and Practices noted that there are points of convergence in this conversation, specifically:

- It is desirable to have some evaluation focus on the issues of free ridership and spillover/market effects when designing and improving programs.
- Attempts to evaluate net savings impacts of programs should incorporate multiple methods and attempt to triangulate a reasonable savings estimate. There was general agreement that no single method of estimating net savings has yet emerged as the preferred approach.
- It is important to incorporate all three basic components (free riders, spillover and market effects) and it is not acceptable to quantify only free ridership.
- Issues of free ridership and spillover/market effects should be examined for program design and program improvement purposes, even if they are not used to adjust official reported energy savings.
- Simplified approaches to quantifying net savings (e.g., assigned NTG values derived from expert opinion, comparable regions, etc.) are acceptable in most situations, assuming they are carefully established and subject to periodic updating.
- For purposes of judging program administrator performance, when *a priori* net savings assumptions (e.g., NTG ratios) have been agreed to, evaluation results should generally only be applied prospectively, to adjust future energy savings calculations.

---

[26] Perspectives at the MEEA 2012 Annual Meeting of the Membership. Midwest Energy Efficiency Alliance April, 2013

3.6 REGIONAL USE OF NET SAVINGS ESTIMATION METHODS

As with the wide range of methods being used to estimate net savings, there is a wide disparity in how states approach net versus gross savings and lack of regional consensus on how to best estimate these values.

The ACEEE 2014 National Survey reported that there are 43 states that have evaluated energy efficiency programs. Of these 43 states, 15 reported net savings, 9 reported gross savings and 19 states reported both net and gross savings. Free ridership is used to adjust savings in 30 states (an additional 3 states plan to), and free drivers/spillover adjustments are made in 25 states (an additional 5 states plan to). Market effects were claimed to be accounted for in 13 states (an additional 2 states plan to), however the authors of the ACEEE survey reviewed their responses in light of accepted definitions of market effects and determined that only two states, Massachusetts and Vermont, actively pursue estimation of actual market effects.

The 2014 National Survey also reported that some states that apply a deemed 1.0 net-to-gross ratio consider themselves to be using net savings, while some consider that to be using gross savings. Once this is taken into account, of 42 states, 22 use net savings with calculated values derived from evaluation studies, four states use a net-to-gross ratio of 1.0, one state (New York) uses a deemed net value other than 1.0 (0.9), four states use both net and gross, and the remaining 11 states using gross savings only. One state appears to be missing in the National Survey report.

Using the 2014 ACEEE study as our guide, we took a deeper look at nine states including California, Hawaii, Illinois, Indiana, Iowa, Massachusetts, New York, Pennsylvania, Vermont and Wisconsin. The next section provides some highlights from each of these states.

3.6.1 Specific State Highlights

3.6.1.1 California

California is the state with the largest energy efficiency portfolio in the country. It is also a state that has one of the more problematic relationships with net-to-gross approaches as evidenced by discussions in decision D.12-11-015, including that of the spillover adjustment discussed previously. There is some disagreement across stakeholders as to the best approach to the issue of net savings estimation, with at least one IOU in favor of returning to a gross savings approach. That said, California does use calculated net savings values derived from program specific evaluations, although essentially only free ridership adjustments are calculated, with a prescribed spillover adjustment. The predominant approach is self-report with some econometric modeling. While the state uses a prescribed spillover adjustment, there is research being conducted presently, and due for publication in the near future, on spillover adjustment calculation methods.

3.6.1.2 Hawaii

In 2011, Evergreen, as the evaluators working for the Hawaii Public Utilities Commission, carefully reviewed the research and regulatory decisions coming out of California and a few other states, and concluded that spillover studies tended to be costly and of questionable value in

producing definitive estimates.[27] As a result, they recommended a pragmatic approach of applying a spillover adjustment of not more than 10% to the previously established NTG ratios. The Hawaii Commission subsequently proceeded with a 5% spillover credit, in line with California's approach (Dickerson 2013). Finally, the evaluators also recommended focusing attention on designing programs to maximize spillover and minimize free ridership rather than trying to precisely measure those variables (Evergreen Economics 2013).

3.6.1.3 Illinois

Illinois reports net savings in their program evaluations. Free ridership is assessed for each program, while spillover is included whenever possible and feasible in each NTG calculation. Evaluators are not required to always include spillover in NTG calculations due to the costs of spillover research, but they recognize that excluding spillover might unfairly reduce calculated savings. Evaluators are told to consider spillover, including logical reliance on deemed values and secondary research developed from evaluations of other Illinois programs and other jurisdictions, to estimate spillover. Portfolio-level spillover analysis is to be considered by each utility at least once every three years when it is feasible and considered viable by evaluation.

3.6.1.4 Indiana

Indiana is an interesting example of a state that is currently using a fairly basic approach to quantifying energy efficiency program savings, but is laying the groundwork for a sophisticated examination of market effects. Indiana has developed the Indiana Evaluation Framework (Indiana Statewide Core Evaluation Team 2012), which defines net savings as inclusive of free rider, spillover, and market effects adjusters. The 2012 Evaluation Framework detailed several approaches to free ridership and spillover, but did not include a market effects approach. In 2015, Indiana planned on developing a revised framework with would include methodologies for estimating market effects. We were unable to find this study, but believe it could be of interest to this research.

3.6.1.5 Iowa

Historically, Iowa has relied on a deemed NTG ratio of 1.0. In 2015, the Iowa Utility Board approved a report to investigate NTG policies. This was presented to the board with recommendations for NTG analysis for each program based on practical feasibility and cost effectiveness. The key recommendations were for Iowa to:

- Use deemed NTG value of 1.0 for programs with low net benefits and savings, and where programs are likely to have a NTG value close to 1.0.
- Conduct secondary research to determine deemed values other than 1.0 for programs where the costs of NTG research are not justified, but research shows a NTG value of 1.0 to be unlikely.

---

[27] Net-to-gross Issues in Hawaii Energy Efficiency Programs: Challenges, Near-term Options, and a Longer-term Approach. Evergreen Economics. 2013

- Conduct primary NTG research to estimate NTG values and/or common practice market baselines for key programs contributing large savings to the utility's demand-side management (DSM) portfolio.
- For programs warranting primary NTG research, market-based methods may be used as the primary research methodology, providing a comprehensive understanding of energy efficiency markets, facilitating development of common practice market baselines, and/or generating estimates of the free ridership and spillover components of NTG values.
- NTG research should be updated, at a minimum, every five years.

3.6.1.6 Massachusetts

Massachusetts is a national leader in net energy savings evaluation of energy efficiency programs, and has had a standardized approach to net savings estimation since 2003. Massachusetts has also pursued clarity on the net energy savings issues, commissioning several studies of free ridership and spillover that are among the most detailed and comprehensive reports in any state on net savings methodologies. The most recent reports aimed at updating the state's standardized approach include reports by Tetra Tech and the NMR Group that are comprehensive reviews of existing approaches, including best practices in market effects estimation.[28]

3.6.1.7 New York

New York is another leader in energy efficiency evaluation. New York may be most notable for its work on the issue of spillover. In the August 2013 update of the New York Evaluation Plan Guidance for Energy Efficiency Portfolio Standard (EEPS) Program Administrators, the New York State Department of Public Service (NYSDPS) established specific guidelines for quantifying spillover effects. The guidelines, among other things, include a list of questions evaluators need to answer before making critical decisions about spillover, a system to determine the amount of rigor to use to assess the amount of spillover, and several methods to calculate spillover. Despite these studies, New York decided in 2010 to adopt a deemed NTGR of 0.9 for all programs, and continue to use this presently, despite a note in the relevant ruling encouraging adjustment of the factor over time. The rationale for the continued use of the deemed value is unclear from the documentation.

3.6.1.8 Pennsylvania

The Pennsylvania Public Utility Commission (PA PUC) Phase II and Phase III Implementation Orders (2013; 2015) direct the Pennsylvania utilities to use NTG adjustments for program design and implementation, but not for compliance purposes or reporting and claiming savings. Their reasoning for this is twofold:

---

[28] Cross-Cutting (C&I) Free Ridership and Spillover Methodology Study Final Report. Tetra Tech, Inc; KEMA; NMR Group. 2011.

Cross-Cutting Net to Gross Methodology Study for Residential Programs - Suggested Approaches. NMR Group, Inc.; Tetra Tech. 2011.

- "if a NTG ratio of less than 1.0 is used, this will raise the acquisition cost per annual kilowatt-hour (kWh) …, which will result in a lower target, due to the 2% budget cap. The current targets include an assumed NTG ratio of 1.0."
- "calculations of NTG ratios are inexact at best. 'Free riders' are difficult and expensive to calculate, but even more difficult and costly to calculate is 'spillover'."

Despite this, the PUC has encouraged the evaluation of methods for estimating net savings to improve program design, including free ridership, participant and non-participant spillover and market effects.

The PA PUC has approved the following three approaches for assessing program effects on the market: self-report (free ridership and spillover), econometric (net savings), and market share analysis to determine market effects.

3.6.1.9 Wisconsin

Wisconsin uses both gross and net savings reporting. Wisconsin's net savings approach includes adjustment for both free ridership and spillover, and they follow a detailed evaluation protocol to determine estimates of these factors. Wisconsin primarily uses a self-report approach for most programs to estimate free ridership and spillover, but also uses an Econometric Demand Modeling approach for their lighting programs to estimate free riders only.[29]

3.6.2 Use of Net Savings Estimation Methods across Sectors

There is a wide dispersion of methods being used across the various energy efficiency sectors. In general, our review revealed that, ignoring deemed values, self-report approaches are the most commonly used across all sectors, followed by econometric methods and common practice methods.

3.7 NET ESTIMATION METHODS USED IN OTHER INDUSTRIES

Net estimation methods are not limited to the energy efficiency industry. Similar issues of net impact estimation are faced by other industries. The Uniform Methods Project chapter on net estimation provides the following examples.

- Healthcare: What would the health effects have been without an investment in water fluoridation?
- Tax subsidies for economic development: Would the project—or a variant of the project—have proceeded without a subsidy?
- Education subsidies: What would happen if school lunch programs were not subsidized or if low-interest loans for higher education were not offered?

---

[29] *Cadmus, August 28, 2013. Focus on Energy* Calendar Year 2012 Evaluation Report Appendix.

- Military expenditures: What would have happened without an investment in a specific military program or technology?

While we have not identified many specific sources to investigate methods in these other industries, this is something we suggest spending some time investigating in future research. For example, we found research investigating the effect of tax credits on the purchase of hybrid vehicles using sales data from different regions; some with, and some without the credit.

3.8 PRESCRIBED VERSUS CALCULATED NET SAVINGS APPROACHES

As noted in the UMP, deemed NTG ratios are predetermined values and do not rely on a calculation-based approach. The use of deemed values for net savings versus the use of researched (or calculated) values varies across states. Deemed values are often based on previous NTG research that was conducted – sometimes within the same jurisdiction, but often in other areas. Typically, NTG ratios are deemed when the expense of conducting NTG ratio analyses cannot be justified or when the uncertainty of the potential results is too great to warrant a study. The UMP quotes a recent review of 42 jurisdictions in the United States and Canada (which represented nearly all jurisdictions with ratepayer-funded energy efficiency programs) finding that only 14% use a deemed approach to NTG for C&I programs compared to 50% of the jurisdictions using an active research approach to developing estimates of net savings factors (Navigant 2013a). This is mirrored by the ACEEE 2014 National Review, which found that five of the 42 states they covered (12%) use deemed values, although they found that all but one of these (New York) use a value of 1.0.

4. Expert Interviews

4.1 INTRODUCTION

The expert interviews were designed to build upon the insights gained from the review of the literature summarized in Chapter 3. We wanted the interviews and review of other studies or papers made available by people we contacted to focus on the issues and techniques described above and summarized here:

- Are there further refinements or new techniques that are not described in the published literature?
- Similarly, is there other recent work that has been or is being done that has not yet been published? Several referred us to other sources or material that we had not had access to before.
- Because we included several experts who had been involved in the research and discussion that led to the 5% adder being used in California, we asked them how that value was selected, whether the 5% adder is likely to be revisited in the near future, and on what basis it might be modified.

The experts we chose to interview represent eight of the leading practitioners of net analysis in the country, and have been involved in the development and refinement of most of the major analytical techniques currently in use. In addition to this long-term involvement, specific recent areas of expertise include:

- Involved in several evaluations/forums of NTG methods for a regional organization
- Involved in CA spillover/market effects studies
- Involved in NY spillover studies, CA market effect/spillover adder
- Applied LCDC to net impacts of financing
- Author of 2008 paper on repeat participation
- Net evaluator for MA, CA, NY, and multiple other states, representing both regulators and utilities
- ACEEE net savings article author
- Directed net analysis in CA, IL, IA, PA, and MD

The interview guide was developed with input from Energy Trust, and is included as Appendix B. Broad topic headings in the guide form the basis for the presentation of interview findings in the remainder of this chapter.

4.2 DEFINITIONS

Across the experts there was consensus that a key step in developing net savings protocols is the need for common agreement or understanding across stakeholders. One topic in particular is the need to have a common agreement of the definitions of components of net to gross calculations. All of the experts interviewed agreed that while there is a common conceptual definition of net

savings, "changes in energy consumption or demand that are attributable to an energy efficiency program" (SEE Action 2012), there is no clear industry consensus of the definitions of all components of net savings.[30] A recent Northeast Energy Efficiency Partnership (NEEP) study noted that different jurisdictions translate the agreed-on conceptual definition into different operational definitions based on the types of impacts or components of net savings that apply to the specific program designs in their regions.[31] For example, regions with programs or organizations that have a mission to focus on market transformation efforts, such as the Northwest, will generally have a different definition of market transformation, or market effects, than regions without these organizations. The NEEP study further recommends that along with clear definitions of the components of net savings, jurisdictions should also agree on "an equation identifying the components or factors that are viewed as components of net savings, whether they are specifically estimated or not;" for example,

> Net Savings = Gross Savings – Free ridership + Spillover + Market Effects (not already captured by Spillover)

The interviewed experts provided varying degrees of insight on operational definitions, the details of which we present below, starting with the most comprehensive definitions.

Two experts referred to the recent NEEP study as the most up to date, comprehensive source of definitions available. The study draws primarily on the Energy Efficiency Program Impact Evaluation Guide (SEE Action 2012) and the Uniform Methods Project chapter on Estimating Net Savings: Common Practices (UMP 2014) as well as input from regional experts in the Northeast. These definitions are provided in Table 4 below.[32]

[30] SEE Action 2012. Schiller Consulting, Inc., "Energy Efficiency Program Impact Evaluation Guide," prepared for State & Local Energy Efficiency Action Network, December 2012.

[31] Gross Savings and Net Savings: Principles and Guidance. Regional Evaluation, Measurement, and Verification Forum. Prepared on behalf of the Northeast Energy Efficiency Partnerships, Inc. (NEEP). . April 2016.

[32] Ibid.

Table 4: NEEP Definitions of Free Ridership, Spillover and Market Effects

Free ridership	Spillover	Market Effects

Free ridership is the program savings attributable to free riders (program participants who would have implemented a program measure or practice in the absence of the program). There are three types of free riders: • **Total free riders:** Participants who would have completely replicated the program measure(s) or practice(s) on their own and at the same time in the absence of the program. • **Partial free riders:** Participants who would have partially replicated the program measure(s) or practice(s) by implementing a lesser quantity or lower efficiency level. • **Deferred free riders:** Participants who would have completely or partially replicated the program measure(s) or practice(s) at a time after the program timeframe.	Spillover refers to additional reductions in energy consumption or demand due to program influences beyond those directly associated with program participation. There are generally two types of spillover: • **Participant spillover:** Represents the additional energy savings achieved when a program participant— as a result of the program's influence — installs EE measures or practices outside the efficiency program after having participated. Participant spillover subcategories include: o Inside spillover: Occurs when participants take additional program-induced actions at the project site. o Outside spillover: Occurs when program participants initiate actions that reduce energy use at sites that are not participating in the program. o Like spillover: Refers to program induced actions participants make outside the program that are of the same type as those made through the program (at the project site or other sites). o Unlike spillover: Refers to EE actions participants make outside the program that are unlike program actions (at the project site or other sites) but influenced by the program. • **Nonparticipant spillover:** Represents the additional energy savings that are achieved when a nonparticipant implements EE measures or practices as a result of the program's influence (for example, through exposure to the program) but are not	Market effects refer to "a change in the structure of a market or the behavior of participants in a market that is reflective of an increase in the adoption of energy efficiency products, services, or practices and is causally related to market intervention(s)" (Eto et al. 1996). For example, programs can influence design professionals, vendors, and the market at large (through product availability, practices, and prices), as well as influence product or practice acceptance and customer expectations. All these influences may induce consumers to adopt EE measures or actions (Sebold et al. 2001). Some experts suggest that market effects can be viewed as spillover savings that reflect significant program-induced changes in the structure or functioning of energy efficiency. As a result, care is needed to ensure that market effects include only those elements that are not already included in the spillover term.

	accounted for in program savings.	

Another expert referred to a recent IEPEC webinar that delved into the conceptual relationships between spillover, market effects and market transformation and argued that each is a subset of the other[33]:

- Market effects = spillover savings that reflect meaningful changes in structure and functioning of energy efficiency markets
- Market transformation = market effects that are substantial and relatively lasting
- Market effects are a subset of spillover, and market transformation a subset of market effects. However, both market transformation and market effects have greater duration than other kinds of spillover, and thus may have greater cumulative magnitude.

In addition to the perspectives of these three experts, other nuances to operational definitions of net savings components were raised, as the following quotes show:

- "Market effects are things that have permanently changed and are unlikely to go back, whereas participant spillover could be less permanent, more behavior-oriented effects … but in the larger scheme of things, if you are looking at spillover you are looking at market effects. One key difference is that market effects is what participants do plus spillover."
- "If you do want to look at market effects, you have to do research on the market itself. For non-participant or participant spillover you will do your research on the audience of the programs."
- One "approach is using non-participant and participant spillover and considering these as market effects. That said, I think of market effects approaches as a step removed from spillover."

4.3 THE NET TO GROSS LANDSCAPE

4.3.1 Most Pressing Issues

It seems that NTG analysis using the self-report is very much like the weather in that everyone complains about it, but nobody has found a way to improve it. One of the issues consistently raised by the experts we interviewed was the almost exclusive reliance on the self-report

---

[33] The Estimation of Spillover: EM&V's Orphan Gets a Home. Prahl. R, Ridge. R, Hall. N, Saxonis, W. IEPEC Webinar. December 15, 2014

approach. All of them, however, said there were few good alternatives. "It's a terrible system, but it is the best system we have," is how one respondent described the problem. "We have been casting around for other ways, but for most programs they are not workable, so we keep coming back to self-report. We have used, and do use, billing analysis if the program design supports it; it's just not often the case that it makes sense. We have tested some other methods like discrete choice, but while the theory sounds plausible, the reality is, when seeing how they work on real data, the models are very unstable and models are not very believable." In addition, most respondents emphasized the importance of budget constraints. Alternate methods often rely on a relatively large sample of non-participants who have taken the actions targeted by the program in question, which makes these methods impractical for many smaller programs and very expensive for others.

Several other experts agreed that alternative methods might be underutilized due to their cost and the difficulty of obtaining the necessary data to make those approaches practicable. This, they say, argues for devoting more attention to how the self-report approach is used, since careful design and implementation can address some of this standard approach's shortcomings. One expert said "the biggest issue is failure to do self-report with sufficient rigor." A problem with the way self-report is usually implemented is that "you can't straight out ask, if there was no program, what would this customer have done? You have to understand the story behind that decision, gather all those facts, and make sure everything hangs together in a thorough and structured way. That means you have to have consistency checks and make sure you are not getting gamed and the question is not taken out of context." While this often requires a longer interview and inquiry, particularly for large projects, this respondent noted that many of the best practices associated with self-report can be incorporated into net analysis for residential and smaller commercial projects.

Another expert pointed out that less than a half dozen jurisdictions (CA, IL, NY, WI, MA) have explicit guidelines for the use of self-report, so that technique's design and implementation vary from jurisdiction to jurisdiction. Evaluators using self-report in areas without specific guidelines should take care to follow best practices in terms of timing, using multiple questions to achieve triangulation within the self-report approach, employing consistency checks and creating a cohesive story that sheds light on the decision to select efficient equipment. Several experts noted that different scoring algorithms should be investigated and tested to ensure stable, consistent results.

In addition to concerns regarding over-reliance on the self-report approach, several experts mentioned the difficulty of estimating spillover as a pressing issue, while emphasizing that "if you are going to try to estimate net savings you need to look at both sides of the coin – free ridership and spillover. It is inappropriate to look at one side of that and not the other." And while there is general agreement that self-report is the best – if imperfect – option for free ridership in most cases, it is more difficult to find similar consensus on techniques for estimating spillover. Citing her experience with a recent regional initiative, one respondent said that "our motivation was: could we design something new and could we get people across the region to come to a consensus and think the same way -- and the answer to both was no. A top-down approach was the closest we could come to consensus, but that was still not satisfactory because you can't disaggregate and there is a lot of noise still there. And then on the midstream program, efforts that involve sales

data are desirable, but getting sales data is very hard. And participants often don't know they have participated, so you can't get free ridership or spillover without sales data."

This difficulty of reaching agreement on appropriate methods for both free ridership and spillover was also cited as a concern by another respondent, who argued that the specific technique is less important than reaching agreement with all stakeholders on the approaches to be used. "What is most important is that the key parties need to come to an agreement on this and resolve uncertainty regarding methods. In terms of how to do that I am not wed to methods but I prefer the principle of triangulation," referring to the use of different methods for a single program to get different perspectives and ultimately a more realistic determination of net impacts.

Finally, an overarching issue specifically mentioned by a single respondent, but underlying the comments from all the experts interviewed, is the fundamental challenge of estimating detailed NTG values. "There are other fields of social program evaluation, and every other field worries about attribution, but not in the same way we do -- in the sense that we really want a quantitative number, a highly precise, highly granular number in many cases. Other fields have more modest goals, through accumulation of knowledge about these free ridership and spillover issues. The objective we have been forced into over the years is unrealistic. The average policy maker has been made to believe that we can come up with a specific value -- which doesn't reflect the reality of the issue. Essentially, our quantitative numbers are not realistic."

4.3.2 Role of Delphi Panels

One method that has been used in net analysis, usually to supplement and integrate the results of other methods, is the Delphi panel, where a group of experts provides input as to what the market overall (and non-participant purchase decisions in particular) would look like in the absence of the program. The process typically involves two iterations, where all participants review the same data and provide initial input, and then use the results of that first round to refine their estimate in the second round. The expectation in a successful panel is that estimates will converge to a credible value in the second iterations.

All the experts we interviewed were familiar with the technique, and several had used it or participated as panel members, while one was suspicious about the validity of this approach:

- "I am suspicious that they are kind of taking a bunch of people and working to a mean that doesn't really reflect what is actually happening. Again, if it was on a really targeted basis and if you had people with strong expertise then it could work, but you have to have people who really know what they are talking about."

Respondents who had participated in or used Delphi panels reported mixed results:

- "I've been in five to seven of these, but my experience has been variable. In theory, this method should work for any issue that is complex and hard to quantify and where you can bring expertise to bear and where people have different forms of expertise. It worked well sometimes in Massachusetts, while another study failed completely."
- "I've seen it applied to NTG for big programs with multiple methods. Delphi is well suited for that."
- "I was not involved in the actual procedures and conduct, but I was looking at the methodology and results. The evaluators were pretty confident they were seeing a

consensus, and they thought it worked pretty well. In California, it also worked well. Having a second wave is good; it doesn't always change people's opinion, but there was some movement with people changing their results slightly based on the first wave."

Caveats around the use of Delphi panels include the need for supporting data to inform the panelists' decision and the selection of a group with diverse perspectives but extensive knowledge of the market and the technology being studied, without having a direct conflict of interest. One respondent said her team had planned to use a Delphi panel to investigate spillover in the multi-family housing market, using 200 interviews with owners/operators of buildings identified as similar, using Dodge data. They abandoned the Delphi, however, because they found minimal spillover activity among non-participants.

Other comments regarding the requirements of a successful Delphi panel included: "For a Delphi to be convincing, you need to see people of different backgrounds converge to a consensus; if you see no changing of opinion and you see clustering, it is hard to have confidence in the results. The whole purpose is that people with different experiences will come to some degree of consensus."

It is the process of reaching convergence and consensus that marks one of the most significant challenges for the Delphi approach, since it can result in a final value that is a mean of widely divergent individual estimates. "I think nobody is happy with taking the average, and yet that is what they have often chosen to do," said one expert with extensive knowledge of this method. Another respondent described a failure of the approach as follows: "When it was used for consumer electronics …, they found very little convergence and net to gross in the low 40s, high 30s. It basically did not work."

Most respondents felt, however, that on balance, this technique could still play a useful role:

- "At the end of the day, it is informed opinion, and could be susceptible to challenge by parties who don't like the outcome. Which is why you need buy-in ahead of time. If you have that buy-in, it is also well suited for triangulation as a secondary method or reality check."
- "If you know a parameter is not zero and you don't want to blow your budget on measuring something that you either don't really know or you don't have the resources to measure, you are better off putting in something that is not zero. And if you have access to relevant information and expert judgment and this isn't going to break the bank, Delphi can provide more info about market trends than an adder could, so this is a valid approach. There is also an element of transparency in a Delphi where you can track back through the process to see where the number came from."

4.3.3 Use of deemed/stipulated values and adders

In addition to often being used for small programs and those targeting low income or hard-to-reach customers, deemed NTG values are, according to the experts interviewed, useful as a placeholder for new or dramatically changed programs. "Rather than not do programs or wait

years, have a good faith multi-party process that sets up TRMs and [use deemed values as] the best estimate of NTG until periodic review 2-3 years later, looking at key assumptions."[34]

While there are still jurisdictions around the country that use deemed values over a longer time frame, our respondents agreed that this becomes less justifiable after programs have been in place for several years, with the exception of the above noted low income and hard-to-reach programs. They particularly emphasized that for larger programs, it is necessary to do program-specific research rather than a deemed value. One respondent explained that "It is easy to get insights into these issues, so for a large state with large programs, I don't see the rationale for guessing." Once a value is established, however, most respondents felt that value could be deemed for several years going forward.

Another respondent noted that a mix of deemed values and calculated NTG can be applied in a single jurisdiction for different types of programs. California, for example, has deemed values in the DEER database[35] for certain programs and measures, but still estimates NTG for non-residential custom and upstream programs.

The use of adders is a related issue. One of the most high profile of these is the 5% adder used by California to account for spillover. Several of our expert respondents were involved in this process. One of them explained that the California evaluation framework formerly did not acknowledge spillover. "For a long time California thought NTG was one minus free ridership, and didn't account for spillover. That policy changed and they decided they need to think about spillover, and the Energy Division of the CPUC invited the IOUs to give their opinion about the magnitude of spillover."

The result was a review of spillover estimates conducted by another one of our experts, who reviewed 45 spillover studies and said that "it was pretty Wild West," with values and even approaches "all over the place" and including the use of both adders and multiplicative factors[36]. The review concluded that spillover was real and could be significant, but would most appropriately be reported as ranges. However, as an interim value, the IOUs recommended a 10% adder to the NTG ratio to account for nonparticipant spillover.

This recommendation "ultimately went up the chain and someone at [the CPUC's] Energy Division decided to use an adder of 5% to acknowledge the existence of spillover until a more detailed study could be done." These follow up studies – one residential, one non-residential – are now wrapping up and will be used to determine whether the 5% adder is changed, kept or scrapped for another approach. As the individual involved in the process noted, "You can imagine a variety of approaches; you could come up with program-specific values and at the other end you could have a global adder. Between those extremes, you might say that some get studies and some

---

[34] TRM refers to technical resource manual.

[35] Database for Energy Efficient Resources, CPUC.

[36] https://www.socalgas.com/regulatory/documents/A-12-07-003/SCG%20Appendix%20I%20-%20Spillover.pdf

get adders depending on the size of the program, or you might say that non-participant spillover is a global issue so you don't want to study it separately for different programs, so non-participant spillover gets a global adder and participant spillover gets a program-specific adder – there is a continuum of approaches and all sorts of points in between." The goal of the current study – which has a $1.1 million budget for the non-residential component alone – is to determine an optimal approach in light of the actual findings.

4.4 NEW TECHNIQUES AND APPROACHES

Net savings estimates can be determined using many different approaches as described in the literature review chapter of this report. Selecting the appropriate analysis method or methods is dependent on several factors including the type of program being evaluated, the research questions to be answered by a net savings study, the level of rigor required and the available budget. We asked our interview subjects to elaborate on any new techniques or approaches for estimating net savings or components they were aware of or that they had recently been involved with, including any insights on advantages, disadvantages, as well as examples, references and potential applicability for Energy Trust.

4.4.1 Econometric Modeling Approaches

Discrete Choice Model / Revealed Preference Model

Discrete choice analysis uses econometric modeling techniques to compare the measure adoption decisions of participants to those of a comparison group to examine how customers make choices for energy efficient technologies, and identify key factors that influence these choices, including demographic, attitudinal, economic, programmatic, and other factors. The models can be used to assess the net impact of the program by estimating the number of eligible customers that change their behavior in response to a given situation such as a new program or action.

Four of the seven experts provided opinions on the use of these models. Among these experts, while they all agreed in theory that they could be useful methods, they also noted similar flaws. Firstly, that these methods are data intensive, requiring careful design and potentially expensive data collection efforts. The second concern was that discrete choice and even stated preference modeling can suffer from a lack of transparency and are difficult to explain to stakeholders without statistical backgrounds. This makes the results difficult to defend at times. One expert explained that they have tested various choice modeling approaches and while the theory sounds plausible, the reality is when seeing how they work on real data, the models are unstable, often changing results significantly with small changes to the model. This instability resulted in difficulty believing and defending the results. However, another expert stated that while evaluators and clients are sometimes nervous about stated or revealed preference models, these methods should not make people any more nervous than self-report methods because at least with preference models, you have the socially desirable choice masked from the respondent to some degree by the structure of the approach. With revealed preference models the evaluator can also calibrate the results based on what you know about customer's actual decisions "so you don't get too far afield from reality".

The interviewees were not aware of many instances where these methods were applied recently. According to the interviewees, discrete choice models were used to estimate net savings for the 2006-2008 and 2013-2014 California Upstream Lighting Program (DNV-GL, Cadmus), and are being considered for evaluation of net savings in several finance pilot programs in California (Opinion Dynamics).

Latent Class Discrete Choice (LCDC) Model

A subcategory of discrete choice models are latent class discrete choice (LCDC) models. As previously noted in the literature review, this approach has not been extensively used, with the only use we identified being in an evaluation of the Southern California Edison (SCE) LED Ambient Lighting Program in 2012 by Opinion Dynamics, and a 2006 evaluation of the SCE Home Energy Audit Program (Ridge et al. 2006).[37] According to one interviewee, this method is being considered for use in several finance pilot programs in California.

This interviewee concurred with the findings from the literature review that LCDC is a potentially viable method for net savings estimation. A benefit of this approach is that it is a flexible approach that could be used to test a variety of influences on free ridership or spillover. In addition, the LCDC method also reveals various choice patterns and does not assume that all customers have the same preferences, which was perceived to be a drawback of other choice modeling approaches. However, the LCDC approach is relatively untested and high cost. The approach is also reliant on a valid comparison group, which can be difficult to identify.

Structural Equation Modeling

Structural Equation Models (SEM) are a group of multivariate analysis techniques, including factor analysis, and path analysis, that can be used to test structural relationships between variables, for example, causal relationships between variables. These models are typically used in social sciences to analyze and quantify relationships between observable variables and latent (unobservable) variables. A classic example is that of intelligence, a psychological variable that cannot be measured directly, which can be inferred through the use of observable test data from intelligence tests where there is a theoretical relationship between test responses and intelligence. An energy efficiency example may be that free-ridership is a latent variable but can be explained or measured by other observable data. Simplistically, the steps involved in structural equation modeling are:

- The evaluator states, based on some underlying theory, the way they assume the variables of interest are inter-related, usually with the use of a *path diagram*. For example, an evaluator may assume that free-ridership for an energy efficiency measure is caused by impact of product cost, program awareness, the customer's socio-economic status and the customer's concern for the environment.

---

[37] *The Southern California Edison (SCE) Advanced Light Emitting Diode (LED) Ambient Lighting Program Customer Preference And Market.* Opinion Dynamics. 2012.

- The evaluator develops a set of internal rules about the relationships between variables and the implications of the relationships between variables. For example product impact of product cost and socio-economic status could be defined as highly correlated, and customer's concern for the environment and socio-economic status as weakly correlated.
- The evaluator then conducts primary data collection, which in the energy efficiency example would likely be a customer survey, and product cost data.
- The evaluator then tests, through multivariate analysis techniques, such as factor analysis or linear regression modeling, whether the variance and covariance of the variables fit the assumed model of them.
- On the basis of this information, the researcher decides whether the model seems like a good fit to the data. If the model is a good fit then the correlations between the observable variables and the underlying theory of the model can be used to estimate free-ridership based on attributes of the program and customers.

One expert interviewed explained that while this method has not been used extensively, it does hold some promise for use in net savings analysis or modeling in other situations. The only example in the efficiency realm the expert could point to was a 2009 IEPEC paper that used SEM to identify the impact of a social marketing campaign for energy efficiency, which included teasing out program effects from other potential influences.[38] Advantages noted by this expert were that SEM allows a researcher to estimate direct and indirect effects immediately from the model, and allows researchers to correlate errors of predictors and model the correlation so the researchers do not have to assume that variables are uncorrelated. Another advantage mentioned was the transparency of the approach and the ability of the researcher to communicate the structure and logic of the model easily through intuitive diagrams. The drawbacks of the modeling approach, according to this expert, are that it can be time consuming and expensive, and that structural models are untested in the net savings arena. Additionally, there is skepticism in modeling literature about the validity of SEM as they rely highly on accurately theorizing the relationships between variables and may falsely attribute a causal relationship where there is none.

4.4.2 Market Effects or Market Transformation Studies

We interviewed one expert who was particularly interested in market effects studies. This individual provided some overall thoughts about market effects studies in general, as well as insights into specific methods, presented below:

- There is a lot of consensus about market effects study design.

[38] Using Structural Equation Modeling (SEM) to Identify, Tease Out, and Quantify a Marketing Program's Influence on Energy Efficiency Intentions and Behaviors. Dougherty et al. Opinion Dynamics. IEPEC 2009. https://www.iepec.org/conf-docs/papers/2009PapersTOC/papers/070.pdf

- Good market effects studies are theory based, have a prospective view, and are conducted in the field from the start of a program. If a program is designed to generate or is likely to generate market effects, then at the outset the program, evaluators should establish leading and lagging indicators and conduct ongoing field studies to assess the consistency of what happened relative to original theory.
- Programs should distinguish between spillover and market effects, by definition. Spillover can happen with any program and we should be consistently looking for it, but market effects will occur with certain programs and should generally be by design. So market effects studies should be focused on programs with strong expectations of market effects.
- Top-down methods can help determine both spillover and market effects, however, in regions with a lot of programs, evaluators have to be aware of multiple areas of market effects causality across programs.

Price Response or Elasticity Models

The same expert that discussed market effects studies in general also discussed price elasticity or price response models. This expert stated that to their knowledge, the only firm doing these at present is Cadmus. They believed that these methods were useful in determining the impact of price changes on purchasing decisions and could be used to determine the impact of free ridership and spillover. However, this expert emphasized that it is important to understand that these methods are in essence price analyses and are not useful for understanding the impact of other non-price related methods of program intervention.

Market Data Analysis

While market data analyses methods are not necessarily new or innovative, three experts had comments regarding these methods, so we include these comments here. The term "market data analysis" can cover a variety of methods, but all three experts we spoke with discussed market data analysis as methods in which the evaluator compares the market share of efficiency products between a baseline, typically a non-program region, and actual market share in the program region. The evaluator collects actual data, as well as conducts interviews with key market actors to gather information about the market share of efficiency products.

One of these three experts was in favor of using market data analysis because this method gathers actual data and produces objective values to show how the market is changing. This method can also be fairly inexpensive. This expert acknowledged that there is always a challenge in finding an appropriate baseline region with customers that have not been influenced by a program, but believed there are still regions that are suitable for comparison.

The remaining two experts were generally not in favor of market data analysis methods. These experts cited two key areas of concern. Firstly, the assumption that efficiency programs have not influenced a baseline region is flawed given the ubiquitous coverage of energy efficiency programs and resulting difficulty in finding a region that is similar in climate and other factors but does not have efficiency programs. The second concern was that the method assumes that the entire difference between the baseline and program regions is attributable to the program effects, essentially "assuming away the problem so it was as if free ridership or spillover doesn't even factor in to the market."

Common or Standard Practice Baseline

A similar method discussed by four experts was the common or standard practice baseline method. This method identifies the standard market practice through some combination of customer surveys, in-depth interviews with market actors, or engineering reviews, and compares the actual efficiency installation with the standard practice to determine savings. In the development of the NEEP guidelines referenced above, this method, and the accompanying Uniform Methods Project chapter, generated a lot of debate. In general, this was not considered a viable approach for determining net savings component values to the group developing the NEEP guidelines because, similar to market data analysis, they believe that the baseline has net savings behavior built into it and "in some ways you are baking in the net savings components or are assuming them away". They were also concerned that the method risks double counting or biasing net savings estimates or impact estimates and ratios unless you make some kind of adjustment. Finally, each of these experts noted that there is no mechanism or component by which to determine spillover or market effects. These experts did identify some benefits:

- The method relies on actual market or field data so could be considered a step up from self-report surveys (one expert).
- The approach can be relatively cheap depending on the availability of market data. However, one expert noted that this approach should be applied on a measure-by-measure basis, so for programs with multiple measures this could become expensive.

4.4.3 Real-Time Survey Data

Lastly, while not technically a method in and of itself, three experts suggested the idea of surveying customers more quickly, both participants and non-participants. Each of these experts suggested that leveraging technology to conduct as close to real time surveying as possible through web surveys, or in-store surveys would help mitigate issues of incorrect recall among customers. Each of these experts warned that this would still be unlikely to mitigate any self-response or social desirability bias.

4.5 METHODOLOGY AND APPROPRIATE USE OF METHODS

4.5.1 Self-report use and issues

Interview responses regarding methodology and the appropriate use of methods generally echoed many of the findings mentioned in the literature review and in the previous discussion of major issues. As noted above, all the experts believe that the self-report approach continues to be standard practice despite the many shortcomings acknowledged even by its proponents. Several mentioned that the self-report approach still tends to be the most cost-effective option and, again as noted previously, some of its disadvantages can be mitigated by following the protocols that have been drawn up for multiple states.

As one of our respondents noted, "I am in the camp that self-reports are OK; yes, there are limitations and there are a number of problems in terms of reliability. There are ways to reduce those limitations, like don't take a year to survey people, get the survey respondent earlier. So self-reports are good; the things that are important are a larger sample size, multiple methods are

good, so don't just rely on self-report, that was why we have always looked at market data analysis as a good complimentary method one can use to confirm or not confirm what people say…. it is better than saying you are just going to stay with gross and not look at net because you don't think self-report is valid. We know there are free riders and we know there is spillover, so why don't we study it?"

Another expert who also backs the use of the self-report approach said that it is important to recognize that this works better in some situations than other. "Self-report works well if it is clear you can get to the person who is making the decision and that is not always easy, particularly in commercial and industrial. It works well if the decision is a major one and not a minor one. For example, with regards to light bulb purchasing, it used to be that LEDs were pretty expensive so you could have some confidence that people were recalling their thinking, but now as they are getting cheaper and less of a major purchase, it is harder to be confident that they are recalling their motivations accurately. One of the challenges is trying to get to the participant as quickly as possible so they haven't forgotten."

4.5.2 Rules of thumb for cost, applicability of methods to sectors, technologies, program types

We also asked our experts whether they think there are rules of thumb regarding program size, type or budget and the applicability of specific net analysis techniques. The consensus was "not really." Several of our respondents have participated in research into this topic, and pointed out that the potential budget required for a free ridership or spillover analysis depends not only on the method selected, but also on the measure, the sector, the program budget and a wealth of other variables. One respondent had worked on a tool that tried to estimate the cost of various net analysis methods for different programs, but had not been able to come up with definitive answers that could serve as a rule of thumb.

Several experts saw opportunity in reduced costs associated with online rather than telephone surveys, and also noted the ability of online surveys to provide a means of reaching participants almost immediately after the purchase decision. One explained that "for one residential appliance rebate program we put a link on the online rebate form that takes them to our online survey. There are other things, but also evaluation questions. It is easy, cheap and gets as close to real time as possible." He added that collecting email addresses during the participation process might support the use of a "semi-automated procedure to send surveys immediately, so you could get data cheaply and quickly."

A related question – whether there is a rule of thumb to choose multiple methods with smaller sample sizes or a single method with a larger sample – also could not be answered definitively. If there is a rule of thumb mentioned by several experts, it would be that multiple methods are better than a single method, recognizing that budget constraints can make that impractical.

4.5.3 Cross-jurisdiction applicability of results

There was some disagreement about whether values from one region could be applied to another. At one extreme, a respondent said, "I think if you are going to come up with a NTG value you should do an evaluation. The counterargument is the disadvantage that you have to take money; but evaluation takes money and this is part of the evaluation exercise. A lot of things are region- and local-specific, so if you use a NTG value from California as the stipulated value it may have

very little applicability in the Northwest. Regions have their own history, maturity of programs, awareness of energy efficiency, prices etc. so I am very reluctant to use a stipulated value from one part of the country for another part of the country."

In contrast, another respondent suggested that it might be possible to use values from other states, if it is done carefully and with an eye to the similarities and differences of the regions. "If it is truly a resource issue then I would first recommend that they figure out at a global level -- not just for NTG but also for gross savings -- where they can look to other states and see if there are results that can be transferred. Sometimes states don't spend enough time exploring whether they can use research results from other states." A third expert said, "I think that is fine, as long as they take into account climate, program design, etc."

Finally, as with deemed or stipulated values, the use of NTG values from other areas was viewed as "a stop gap measure until you can establish your own numbers." Over the longer term, "you can't just borrow other people's numbers; you have to do the research, even if it is expensive."

4.5.4 How to measure market effects and spillover

In addition to lack of clarity on specifically how to define market effects and spillover, there were no "new and improved" methods of measuring either of these. Again, the consensus is that the self-report approach is the most widely used, particularly for participant spillover. However, one expert said that "in our experience, the participant spillover rate is pretty low, so we may be making or building methods for rare occurrences."

Self-report can also be used for non-participant spillover, but two respondents pointed out the large sample sizes needed to encompass a significant number of non-participants who installed energy efficient equipment.

4.5.5 Tracking long-term program influence

None of the experts interviewed were aware of successful attempts to measure long-term program influence, whether through self-report or other methods. One respondent emphasized that use of the self-report in particular was not appropriate for long-term influence, explaining that "the recall problem is exacerbated in this situation, and for good market transformation, much is not perceived by the customer and this is the point. Customers aren't going to say the reason I did X was because there have been ten years of programs training installers and working with manufacturers to make more efficient products, etc. I haven't seen anyone try to tackle this."

Another expert pointed out the inherent challenge of getting simultaneous participant input on both the impact of a program on their recent decision and on longer-term influence. "When you do a free ridership survey, you should get this to a participant as soon as possible, which is the opposite of trying to measure long term effects when you want to get it to them later….Soon after the purchase they might tell you, 'hey I got $50 and I'm happy and I love this new dishwasher and I am going to do all this other energy efficiency stuff,' but six months later they will not really remember and they are back to normal life."

Several mentioned that long-term program influence overlaps with market effects, which has been attempted to be measured using several approaches. "There are things we have done like shelf space devoted to efficient products and [we can] see how this grows over time. I am not sure of the comprehensive list of market transformation indicators, but this would be one of them – how

retailers sell the product." Another expert explained that "we look for behavior changes as a result of long-term program effects -- like incorporating energy efficiency into procurement, having an energy manager, having an energy policy -- that suggest you are changing the behavior. Other than that, it's very hard to quantify long-term influence."

One respondent provided context for this issue by noting that it is difficult to assess a program's effect on the market if this has not been considered and addressed when the program is first launched. "What is important is to decide ahead of time the focus of the program and if there is a real attempt to change the market over time. If there is, then you need to adopt the consensus market approach to market transformation studies and that is early involvement, theory-based evaluation, specification of leading and lagging indicators, staying in the market for years -- and that is hard and expensive so it is important to separate what is and isn't a market transformation initiative.

4.5.6 Studies of variation in NTG component over program life

None of the experts interviewed were familiar with studies that explicitly examined changes in NTG ratios over the life of a program. One respondents noted that there is a public spreadsheet showing NTG values by year for programs in Illinois, but the focus of this effort is not to link changes in NTG to changes in the program life cycle, but to determine whether a consistent NTG over the years would make it unnecessary to track NTG every year.

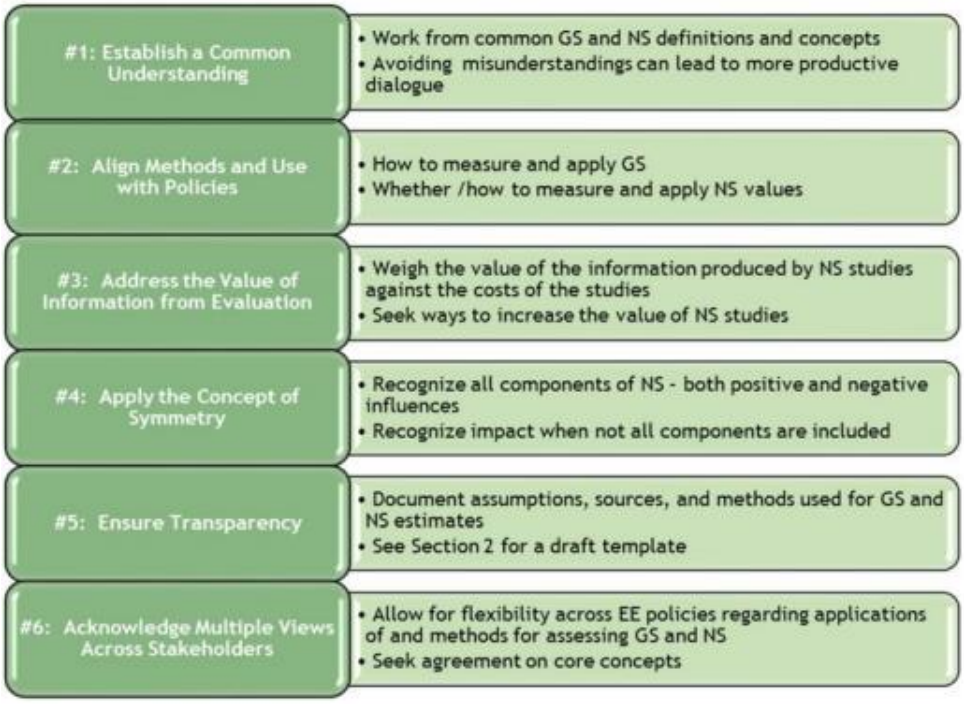
4.6 RELEVANCE TO ENERGY TRUST

We asked the experts to provide us with their recommendations for Energy Trust as the best way to estimate spillover or settle on a spillover value and evaluating the level of long-term market effects of multi-year programs, and several offered their suggestions. We present a summary of these recommendations below, by expert:

4.6.1 Expert 1

Firstly, this expert referred to the recent NEEP guidelines for a general overview of developing a net savings framework, and specifically noted the six guiding principles identified as a good starting point as depicted in the graphic below.

Figure 1: NEEP Figure 1-1. Six Guiding Principles for Considering a Net Savings Framework



This expert then suggested that Energy Trust should consider a targeted approach to developing estimates, and spreading evaluations out over several years. First, categorize programs by their magnitude of impact and the role of behavioral elements in them. Based on these categorizations, Energy Trust should develop a priority of programs and issues, and using these priorities, focus resources on the highest priority issues first and develop calculated estimates of net savings components for these. For smaller budget programs or lower priority items, adders and assumptions are a very reasonable way to go in lieu of calculated values.

Regarding methods, this expert advised multiple method approaches such as self-report and Delphi or self-report and market data analysis to try and triangulate on values. For programs very focused on market change or long term effects, this expert suggested blending in tracking elements as part of program implementation so data are collected to develop forecasts or possibly information that could be included in a Delphi process.

4.6.2 Expert 2

Expert two suggested that with a limited budget, Energy Trust should develop a multi-year plan to focus resources on high priority issues first and lower priority issues later. Alternatively, this expert suggested Energy Trust could also adopt a multi-stage effort across all programs where net savings are an issue, which could be designed as follows:

- Year 1: Self-report survey of participants and non-participants to determine spillover
- Year 2: Survey key market actors to determine spillover and market effects

- Year 3: Conduct a Delphi Panel using the results from the self-report surveys and key market actor interviews to come to a consensus on the most appropriate values for spillover, market effects or other components of net savings and feed into the Delphi.

This could also be thought of as stages rather than years, but could be over three years if budget was constrained. Doing this in progression and having the Delphi last ensures better information for the experts participating in the panel. This expert believed this approach would be robust in terms of being applicable to multiple technologies and sectors, in contrast to other methods that might only be applicable to certain programs or sectors.

4.6.3 Expert 3

This expert did not feel comfortable recommending specific methods or approaches, but did emphasize that multiple method approaches are more defensible and provide more flexibility to evaluators. This expert also explained that of key importance is obtaining a consensus among stakeholders around the validity of the chosen methods, such as developing a white paper or set of guidelines that are agreed upon by stakeholders such as regulators or utility staff.

4.6.4 Expert 4

Expert four also suggested a targeted, multi-stage effort to make the most of the limited budget resources of Energy Trust. This expert suggested picking high priority programs or measures where you think you will find spillover, long-term market effects, or other net savings component values and go at those areas with high rigor. This expert was of the opinion that it is very important to ultimately have calculated values that are accurate and defensible rather than just assigning a specified value for each component

4.6.5 Expert 5

Expert five also suggested a prioritized approach, but also leveraging other evaluations in similar states or with similar programs. This expert first recommended that Energy Trust determine at a global level, not just for net savings but also for gross savings and across all programs, where they can look to other states and see if there are results that can be transferred. Then use these values, where appropriate until adequate resources can be devoted to rigorous evaluation. For evaluating long term market change, this expert explained that they believe it is critical to have early involvement in a program and develop a rigorous, theory-based approach with ongoing market data tracking and pre-defined leading and lagging indicators of market change that will be measured and estimated.

4.6.6 Other experts

The three other experts did not have specific recommendations, but said that for any organization with a limited budget, the self-report approach was probably the most cost-effective, if it was properly implemented. One expert mentioned following one of the various sets of protocols (e.g., from California, Massachusetts and elsewhere). Another emphasized getting a more complete understanding of the equipment selection decision, even for smaller purchases or projects. As quoted previously, this interviewee said, “you have to understand the story behind that decision, gather all those facts, and make sure everything hangs together in a thorough and structured way. That means you have to have consistency checks and make sure you are not getting gamed and the question is not taken out of context.”

4. Conclusions and Recommendations

4.1 CONCLUSIONS

Key findings reported throughout this report are summarized below.

- Perhaps the most significant result of our literature reviews and interviews is that there is no silver bullet; there are no breakthrough techniques in net to gross analysis that would allow Energy Trust to calculate net savings with greater confidence and at reasonable cost than the self-report approach it is currently using. Instead, the NTG landscape remains a patchwork of methods and policies where most jurisdictions use the same self-report technique that they and their consultants recognize as flawed.
  - Self-report may not be best practice, but it is standard practice, used by most jurisdictions for free ridership and participant spillover and sometimes for non-participant spillover.
  - Best practice is generally considered to be triangulation using a mix of methods, so that results from different methods can be used to validate one another and enhance confidence in the resulting NTG value.

- Precision for most methods, and particularly for self-report, typically refers to sampling precision only, and is a function of sample size and the number of segments or strata for which values are to be estimated. For free ridership, the relationship between sample size (and cost) and precision is straightforward, since all respondents will have undertaken the program-supported action. For non-participant spillover, it may be possible to estimate what percentage of the population replaced a certain type of equipment (i.e., with an average EUL of 12 years for air conditioners, roughly 8 percent would be replaced in any given year), but it is extremely difficult to predict how many of those replacements would have involved program qualifying equipment, making it challenging to predict both the required sample size and cost for non-participant spillover studies.

- For both free ridership and spillover calculations, a number of states have guidelines to ensure that a consistent approach is used that mitigates the shortcomings of survey- and interview-based techniques. These guidelines, generally agreed to by all stakeholders, help ensure consensus regarding the net analysis approach and widespread acceptance of the results.

- In addition to the self-report approach, a variety of other techniques are available and proven, including econometric methods (such as billing regression and discrete choice models) and market effects or market transformation studies (including price response or elasticity models and market data analysis) and use of a standard practice baseline.
  - Both the literature and the experts we interviewed note that these techniques have their own limitations, including higher cost. However, they have significant value as a reality check/validation of self-report findings and as an added data point in the application of a triangulation strategy.
  - Many of these alternate approaches rely on the availability of a non-participant sample with enough households or businesses that have undertaken a targeted efficiency

action without taking advantage of program incentives or services. Such a sample can be difficult and costly to identify, meaning that hundreds or even thousands of non-participants may need to be reached before enough customers are found who have taken the program-supported action without participating in the program.
  - For both self-report and alternate techniques, results often convey a false sense of precision, with NTG values calculated to the second decimal place. Few of the NTG techniques currently in use warrant this level of precision, especially when results are estimated values based on the counterfactual.

- Most experts say there are no hard and fast guidelines for the appropriate level of spending for net-to-gross analysis for various types and sizes of programs. Ideally, more budget might support the use of some alternate methods to support triangulation and improve estimates. If other methods are too costly, however, the best use of extra budget may be to increase the sample size for an existing approach rather than go for a more expensive technique. Increasing the sample size for the self-report approach might allow for separately calculated NTG values for specific measures or customer groups within a program, for example.

- For some types of programs (e.g., low income, hard-to-reach small business, new programs), stipulated NTG values of 1.0 or close to 1.0 are viewed as appropriate by the literature and by market experts. For other types of programs, stipulated values drawn from research in other states/service territories are generally considered acceptable as placeholder values until a program or region-specific value can be calculated. All the experts interviewed, however, say that generally this should be only a short-term solution, and that spending on research to calculate NTG is a legitimate evaluation activity that should be built into overall evaluation costs.

- There is consensus that spillover is real and should be accounted for, and experts and the literature agree that NTG should not be calculated simply as one minus free ridership.
  - Participant spillover is estimated more often than non-participant spillover, but is generally of a smaller magnitude, rarely adding as much as 5% to the net-to-gross value. Estimated values for non-participant spillover vary widely, ranging from less than 5% to more than 1.0 (i.e., non-participant spillover savings exceed participant gross savings.)
  - While adders and deemed values are acceptable as a placeholder value to acknowledge spillover, region- and program-specific research should be conducted as soon as practicable to develop a more defensible spillover value.
  - The California 5% adder to account for non-participant spillover did not come from research or program evaluation results, but is instead intended as a conservative placeholder value. The filing submitted by the California IOUs in support of non-participant spillover cited numerous studies to justify a larger spillover value, but the CPUC Energy Division and its consultants disputed the relevance of many of those results, and the CPUC limited the amount of spillover to a 5% adder across the board until California-specific studies could be completed. The 5% is added to whatever NTG value is calculated, so that a calculated NTG of 0.6 would become 0.65, for example.

4.2 RECOMMENDATIONS

Based upon suggestions from interviewed experts as well the literature review, we make the following recommendation to Energy Trust regarding its estimation of NTG values.

Considering the almost universal application of the self-report approach across the country, Energy Trust can continue its use of Fast Feedback surveys and the calculation of free ridership using a self-report algorithm. However, we suggest Energy Trust consider several adjustments or modifications to validate the results being obtained using this method.

- One option would be an alternate self-report approach using another instrument and algorithm, such as the basic format employed in California[39]. If an alternate technique was used for part of the survey population while the basic Fast Feedback instrument was used for the rest, the results could be compared.
- As one of the experts suggested, it might be appropriate to incorporate consistency checks into the existing Fast Feedback survey to address clear discrepancies in program influence/importance between the two free ridership components (stated intent /project change and program influence.) Currently, the free ridership algorithm accepts responses that the program resulted in no change in the project (FR=1) and that one or more elements of the program were extremely influential (FR=0) and generates a free rider score of 50%. We recommend that such contradictory responses lead to a clarifying question to resolve the inconsistency. For example: "You said that the program resulted in no change in the project and that your organization would have made the funds available anyway, suggesting that the project was not very influential. Yet you said that an Energy Trust funded study was extremely influential in the decision to incorporate energy efficient features in your project. Can you please clarify?"
- For specific measures where appropriate sample is likely to be available, it might be possible to try one of the quantitative techniques described in this report as an alternate estimate of free ridership. We recommend the discrete choice modeling approach over the less-vetted LCDC and SEM approaches.
- Consider adding a short battery of participant spillover questions; initially to determine whether there is evidence of any spillover and subsequently to try to quantify that spillover.
- For commercial and industrial customers with larger projects, start the survey with an open-ended question to get a "story" behind the project and the selection of energy efficient equipment.

[39] http://www.energydataweb.com/cpucFiles/pdaDocs/910/Nonresidential%20NTGR%20Methods%202010-12%20101612.docx

- Because Energy Trust's self-report algorithm uses two components to score survey responses, consider including Cronbach's alpha test to test for internal consistency in responses.[40]
- Finally, while Fast Feedback already addresses the issue of timely contact with program participants, explore the possibility of methods that could reduce the time between the customer purchase decision and survey administration even further by collecting customer email information on rebate forms and leveraging web surveys, either via email, or in-store (for retail measures).

Other suggestions from the experts we interviewed focus on the net savings analysis framework and planning:

- Experts expressed the need to develop stakeholder consensus around net savings analysis and the methods used. To ensure transparency and stakeholder consensus, Energy Trust could develop a set of principles and guidelines that provide the theoretical background to net savings approaches adopted by Energy Trust, similar to that developed by NEEP.
- Develop a targeted approach to calculating NTG estimates, and spreading evaluations out over several years. First, categorize programs by their magnitude of impact and the potential for free ridership, spillover, and market effects. Then, based on these categorizations, Energy Trust could develop a priority of programs, focusing resources on the highest priority programs, sectors or measures first to develop calculated estimates of net savings components for these.
- As an example, a spillover analysis to supplement the free ridership information already being collected might be done for high-impact measures covering several residential end uses, such as linear LEDs for lighting and furnaces or ductless heat pumps for HVAC. The Fast Feedback survey could be used to capture participant spillover data, while a large scale non-participant survey could be used for non-participant spillover.
- For smaller budget programs or lower priority items, review data available from other regions to identify similar programs/markets, which would support the use of adders and deemed values in lieu of calculated values.

For non-participant spillover, Energy Trust has been using a 1% adder for all programs except Existing Buildings, where it applies a 7% non-participant spillover value based on the results of a previous survey of non-participants[41]. An update of that survey is planned, which presumably will lead to a new non-participant spillover value. As Energy Trust wrestles with the appropriate

[40] Cronbach's alpha is a measure of internal consistency of a test or a scale. It is frequently used to measure reliability of data collection instruments such as a survey. Cronbach's alpha is expressed as a number between 0 and 1 and the value indicated the degree to which the components of a test measure the same concept or construct. In essence Cronbach's alpha tests the correlation of the components of the test with each other. The higher the level of correlation among the components, the more reliable the test is assumed to be.

[41] Personal communication from Energy Trust Evaluation Manager.

approach to spillover, we make the following recommendations based on input from the literature and expert opinion.

- Given the "political" difficulty of spending evaluation resources to survey customers or trade allies in comparison regions, Energy Trust should continue to focus on approaches that rely on data collected from its own customers and trade allies who serve those customers.
- Where appropriate, data from other regions might be useful. This might involve using values developed for a portion of another state rather than the entire state. For example, while California-wide climate data would be very different from Oregon for heating and cooling measures, the Pacific Gas and Electric territory in Northern California would be relatively similar.
- Consider use of Delphi panels for developing consensus around values for spillover and market effects, using outputs from other methods as inputs for the Delphi to achieve triangulation. This is a relatively inexpensive approach to confirming validity of results and potentially improving consensus and confidence in results.

The extent to which program presence and activity over the long-term influences customer behavior and attitudes does not seem to lend itself to easy measurement using survey methods, and we are unable to make a recommendation on how long-term program influence can be incorporated into current Energy Trust NTG estimates. However, it may be possible to use surveys or interviews to capture data on trade ally sales practices over time and link it to their length and intensity of involvement with Energy Trust programs.