Scaling-Up Energy Efficiency Programs: The Measurement Challenge

Joe Loper, Steve Capanna, Rodney Sobin and Tom Simchak

Alliance to Save Energy, 2010

Acknowledgments

Many thanks to Karen Palmer (Resources for the Future), Steven Kromer, Marvin Horowitz, Peter Fox-Penner (Brattle Group), Elizabeth Titus (Northeast Energy Efficiency Partnerships), Nigel Jollands and Lisa Ryan (International Energy Agency), Mark Hopkins and I-Chun Hsiao (United Nations Foundation), Carl Blumstein (California Institute for Energy and Environment), Nikolaas Dietsch and Jeffrey Brown (Environmental Protection Agency), Carla Frisch (Department of Energy), and Mike Rufo (Itron) for their comments and suggestions. Alliance staff also provided helpful comments and other assistance include Brian Castelli, Lowell Ungar, Jeff Harris, Selin Devranoglu, Sally Larsen and Kateri Callahan. This report is made possible with support from U.S. Department of Energy, Pennsylvania Department of Environmental Protection, and Kresge Foundation. Of course, the authors are solely responsible for any opinions, errors or omissions.

Table of Contents

Abbreviations and Acronyms

ACEEE	American Council for an Energy Efficient Economy
AEE	Association of Energy Engineers
ARRA	American Recovery and Reinvestment Act
ASHRAE	American Society of Heating, Refrigeration and Air-conditioning Engineers
BAU	Business as Usual
CAFE	Corporate Average Fuel Economy
CALMAC	California Measurement Advisory Council
CDM	Clean Development Mechanism
CEE	Consortium for Energy Efficiency
CEO	Chief Executive Officer
CFL	Compact Fluorescent Light bulb
CMVP	Certified Measurement and Verification Professional
CPUC	California Public Utilities Commission
CO_2	Carbon Dioxide
CRS	Center for Resource Solutions
DOE	Department of Energy
DPUC	[Connecticut] Department of Public Utility Control
DSM	Demand-Side Management
EE	Energy Efficiency
EIA	Energy Information Administration
EERS	Energy Efficiency Resource Standard
eGRID	Emissions & Generation Resource Integrated Database
EMV	Evaluation, Measurement and Verification
EPA	Environmental Protection Agency
ESC	Energy Savings Credits
ESCO	Energy Service Company
ESPC	Energy Savings Performance Contract
EUL	Estimated Useful Life
EVO	Efficiency Valuation Organization
FEMP	Federal Energy Management Program
GHG	Greenhouse Gas
HVAC	Heating, Ventilation and Air Conditioning
IEPEC	International Energy Program Evaluation Conference
IPMVP	International Performance, Measurement and Verification Protocol
ISO	Independent System Operator
kWh	Kilowatt-hour
MWh	Megawatt-hour
M&V	Measurement and Verification
NAESB	North American Energy Standards Board
NAPEE	National Action Plan for Energy Efficiency
NEEP	Northeast Energy Efficiency Partnerships
NYSERDA	New York State Energy Research and Development Agency
PUC	Public Utility Commission
REC	Renewable Energy Credit
PJM	Pennsylvania - [New] Jersey - Maryland [Interconnection]
RUL	Remaining Useful Life
TRM	Technical Resource Manual
VER	Verified Emission Reduction
W	Watt

Introduction and Executive Summary

Everyone is talking about energy efficiency (EE). Congress and the President included roughly $20 billion in the economic stimulus program for EE deployment programs and investments in federal facilities. Electric company CEOs call EE the nation's "first fuel." Electric utility regulators are directing more than $3 billion annually to be spent for EE programs and that figure is expected to increase to $4.5 billion. A growing number of states are implementing energy efficiency resource standards (EERS).

Meanwhile, in a attempt to drive deeper investments in electricity end-use efficiency, at least a few states have implemented strong performance-based incentives to reward utilities for investing in customers' EE improvements, including (most notably) a substantial utility shareholder incentive in California. In addition, most proposals for a federal cap-and-trade regime to regulate carbon emissions include provisions that would allow regulated entities to claim credit for reducing carbon or other greenhouse gas emissions that are not covered under the cap. A portion of the carbon allowances that they are required to submit under the cap could be "offset" by reducing those uncapped emissions.[1]

Increased public funding and performance-based compensation schemes have drawn greater attention to how EE program savings are measured. Despite decades of experience and an impressive body of literature, estimating EE program energy savings is still plagued by disagreements over methods and appropriate levels of rigor as well as perennial data shortcomings. Evaluation of EE program impacts – as with evaluation of other types of programs – is as much art as science and evaluation results can vary widely depending on different evaluators' professional judgment and the assumptions they use.

Performance-based incentives are proving to be a mixed blessing in California so far. On the one hand, the deliberations over EE program evaluation, measurement and verification (EMV) results have been costly – about $100 million has been spent on EMV in California over the last three years to evaluate a $3 billion program budget. Moreover, the deliberations have been rancorous as hundreds of millions of dollars in utility compensation can hinge on the EMV results. On the other hand, they are incentivizing California's utilities to pursue increased investment in demand side EE and have garnered manifold improvements in EMV data, methodologies and, perhaps most important, the oversight process. By creating an adversarial process in which various interests argue over how much compensation should go to California's stakeholders, a spotlight has been shined on programmatic EMV.

Likewise, the Clean Development Mechanism (CDM), which awards credits to carbon-reduction projects in less developed countries under the Kyoto Protocol, has struggled for over a decade to develop reliable, but implementable, methods for evaluating energy efficiency projects and programs.[2] Whether the CDM Board has been successful in this balancing act depends on who you ask, and when, but the role of programmatic EE remains small in the CDM.

[1] Note that reductions in emissions that are covered by an emissions cap would not be eligible to receive carbon credits.
[2] The CDM was created as part of the Kyoto Protocol to enable participating developed countries to meet part of their carbon reductions commitments through investment in carbon reduction projects in developing countries. For

Credible savings measurement is a prerequisite for participation of EE projects and programs in carbon offset (credit) schemes such as the CDM, performance-based incentives to encourage utility investments in customer efficiency, for implementation of a meaningful national energy EERS, and generally assessing the cost-effectiveness of EE programs. It remains to be seen whether improved EE EMV will mean more programs, fewer programs, or different programs, but credible savings measurement will be a vital part of a sustained scale-up of energy efficiency programs.

In *The Measurement Challenge* section that follows, this report provides an overview of issues associated with measuring EE program impacts, including how some challenges – for instance, attribution of savings – is becoming more difficult as more and increasingly diverse projects, programs and policies are implemented.

Right-Sizing Certainty provides perspective on issues of measurement accuracy, precision, bias, and uncertainty. These EMV challenges are not unique to EE, but exist in many other fields of public policy importance, such as education, health, safety and security, poverty alleviation, defense, and economic policy. The needed level of certainty for EE programs can vary depending on program intents. So, for example, certainty required for load forecasting may differ from that needed for EERS compliance or carbon credit trading. Some types of evaluation and measurement biases may be unknowable and unmeasureable, so the quest for ideal methods and data will be frustrated.

The *Battling for Certainty* section argues for the adversarial process in which stakeholders deliberate over EMV. Critical questioning and debate can shine a light on EMV assumptions, methods, data, and analyses to overcome biases and enhance the quality and credibility of EMV and of EE programs. While the California process is not pretty at times, it is uncovering problems with EMV and beginning to address them. EMV problems are not new but are now being faced. In response to the argument that the California shareholder incentive should be eliminated because of acrimony over EMV, we ask why the EMV that is good enough for determining cost-effectiveness and portfolio decisions is not good enough for compensation. Or, why do people accept different levels of EMV rigor and certainty for allocating resources *ex ante* through utility planning and associated rate adjustments in contrast to awarding *ex post* compensation?

Toward a National EMV Protocol discusses the difficulties of developing such a protocol. The diversity of project and program types and characteristics make the development of common protocols very challenging. However, if a national EERS and carbon offsets for EE programs are enacted then consistent EMV protocols are absolutely critical. The Northeast Energy Efficiency Partnerships (NEEP) is exploring possibilities for a regional protocol, having started with an effort to harmonize EMV terminology. Other organizations, such as the North American Energy Standards Board (NAESB), the National Action Plan for Energy Efficiency (NAPEE), and the Efficiency Valuation Organization (EVO) also have relevant initiatives.

---

discussion of the CDM Board's efforts to credit EE programs, see Christiana Figueres and Michael Philips, *Scaling Up Demand-Side Energy Efficiency Improvements Through Programmatic CDM*, Energy Sector Management Assistance Program and Carbon Finance Unit, World Bank, 2007.

The report closes with *Conclusion and Recommendations*. The section underscores the value of adversarial processes where stakeholders can question and debate and, it is hoped, make EMV and EE programs more transparent and credible. It also reinforces that while increasing evaluation precision is an important objective, reducing systematic bias is ultimately more important to the credibility and reliability of evaluation results.

To enhance the credibility and reliability of EE program savings measurement, federal and state governments, along with various stakeholders, should: 1) Institute processes for EMV design and review that incite transparent and thorough debate over EMV methods, data and assumptions; 2) Improve EMV methods, data and assumptions; 3) increase consistency of methods and assumptions between regions and program types, 4) assure evaluation professional competency and integrity and 5) manage stakeholder expectations.

The Measurement Challenge

At the heart of the measurement challenges is a counterfactual – what would have happened if not for the program? We never actually know how much energy was saved as a result of the program activity. While we can estimate savings, we can never fully know whether the estimates are correct or the direction of bias. Thus we cannot recalibrate estimation methods and assumptions based on "actual" savings. This is not to say we cannot usefully and adequately estimate savings – just that there is considerable uncertainty and thus room for reasonable people to argue.

A full discussion of measurement issues could – and does – fill rooms full of reports and conference halls full of people. The International Energy Program Evaluation Conference has drawn hundreds of evaluation experts every other year since the early 1980s.[3] The California Measurement Advisory Council maintains a database of California evaluation studies and the Consortium for Energy Efficiency maintains a database of studies from other states.[4]

In this section, we provide an overview of some key issues that emerge when trying to measure EE program savings.

Gross versus Net Impacts

The steps involved in EE program impact evaluation generally include estimation and reporting of gross and net savings to, respectively, verify that savings actually occurred (are "real") and would not have occurred otherwise (are "additional"). NAPEE defines gross energy impacts as:[5]

> *…the change in energy consumption and/or demand that results directly from program-related actions taken by energy consumers that are exposed to the program, regardless of the extent or nature of program influence on these actions. This is the physical change in energy use after taking into account factors beyond the customer or sponsor's control (for example weather).*

Net impacts are gross energy impacts "that are attributable to the program" which is involves making at least several adjustment to gross energy impacts: [6]

- Free riders – Subtracting "the portion of energy savings that participants would have achieved in the absence of the program through their own initiatives and expenditures."

[3] Traditionally held in the US in odd numbered years, the IEPEC will hold its first conference in Europe in Paris in 2010.
[4] The CALMAC database can be accessed at http://www.calmac.org. The CEE database is available at http://www.cee1.org.
[5] National Action Plan for Energy Efficiency (NAPEE), *Model Energy Efficiency Program Impact Evaluation Guide*, Prepared by Steven R. Schiller, Schiller Consulting, November 2007, pp. 26-27, http://www.epa.gov/RDEE/documents/evaluation_guide.pdf.
[6] *Ibid*.

- Spillover – Adding energy savings resulting from actions taken by 1) people who did not participate in the program but were influenced by it, and 2) participants who took additional actions as a result of the program – e.g., as a result of participating in an air conditioner rebate program, they were influenced to purchase an more efficient clothes washer. Note that spillover can also be negative; for example, if participants do not like the CFLs that have been given to them for some reason, they could project that dislike onto other energy savings measures.
- Rebound effect – Some program-induced efficiency gains may be offset by increased demand for energy services. For example, a CFL may be left on more than the incandescent bulb it replaced. Rebound effects must be subtracted.

Net impacts may be more or less than gross impacts, depending on whether spillover exceeds free ridership and rebound effects.

Measurement and verification (M&V) is generally associated with estimating gross savings from individual, or a collection of, projects or measures. Evaluation is associated with estimation of net savings from a collection of projects or measures – i.e., programs. We use the terms EMV and evaluation interchangeably – M&V is a subset of both.

Measurement & Verification of Gross Savings

M&V is intended to establish whether measure or project savings actually occurred (i.e., are "real"). M&V elements include whether the measure or project was implemented, what the actual operating efficiency is (as opposed to the claimed or nameplate efficiency), if the measures are operating properly, the number of hours the measure is used and when those hours occur. M&V also involves development of project or measure BAU levels, from which adjustments are made to account for deviations from expected weather, hours of use and maintenance practices and the type of equipment that was replaced or upgraded.

Among the more difficult M&V challenges are:

- *Savings Persistence* – Few measures save energy by themselves. Their performance depends on whether they are installed correctly and how they are maintained and operated. Savings persistence can be affected by other factors as well, including (for example): equipment breakage or malfunctioning, equipment decommissioning, factory closing, or demolition of building in which the project was installed.
- *Useful life* – Measuring energy savings from a project requires comparing energy consumed during the estimated useful life (EUL) of the new equipment or measure being implemented to the energy consumed during the remaining useful life (RUL) of the equipment being replaced plus the energy consumed by the equipment that would have replaced the original equipment at the end of its life.[7] Estimating the useful life of various measures can be costly

[7] For example, if a light fixture being replaced has an expected useful life (EUL) of 10 years and is 6 years old, the RUL is 4 years. After 4 years, the light fixture would be replaced by a new fixture. Let us say this baseline replacement fixture also has a EUL of 10 years, and it will consume 200 kwh per year compared to the current

and time-consuming (perhaps even the life of the equipment) and, in some cases, by the time the data are available they are longer relevant.[8]

- *Behavior change* – Many EE measures can affect how and how much consumers operate energy-using equipment. "Smart meters" are intended, in part, to allow immediate feedback to consumers about their energy consumption and costs in hope that increased awareness will drive increased investment in EE or EE behavioral changes (say, turning off unused equipment or running full dishwasher loads). In the other direction, rebound effects may occur as a result of EE measures. These are just two of a multitude of examples of how behavior can complicate and increase the uncertainty associated with savings measurement.
- *Interactive effects* – Many EE measures will affect other energy-using systems. Replacing incandescent lamps with CFLs, for example, can increase heating needs and reduce air conditioning loads. Installing a more energy efficient heating, ventilation and air conditioning (HVAC) system will reduce the energy saving associated with an energy management control system. Better insulation or windows can reduce HVAC loads. Occupancy sensors will lower savings attributable to new light fixtures. Understanding and modeling these relationships in a single building presents significant engineering challenges. Accounting for interactive effects for many projects or within or between programs is even more difficult.

The Role of the International Performance Measurement and Verification Protocol

Methods and techniques for project M&V are often perceived to be straightforward with widely accepted best practices guidance in place, including most notably the International Performance Measurement and Verification Protocol (IPMVP). The IPMVP provides extremely useful discussion of the many dimensions of M&V, including: normalizing savings (for weather, occupancy, etc), interactive effects (increased heating load resulting from reduced lighting), directional causality, data sources, missing data, data preparation, meter accuracy, statistical bias, rounding of numbers, independent verification, measurement boundary and retrofit isolation, appropriateness of assumptions, building simulation and more.[9]

But while the IPMVP is extremely useful in defining terms and presenting frameworks for designing and implementing project level evaluations, it does not eliminate the professional judgment involved in M&V. The IPMVP is expressly designed to allow flexibility in creating M&V plans, while "adhering to the principles of: accuracy, completeness, conservativeness, consistency, relevance and transparency."

---

fixture which consumes 300 kwh/year. If the new more efficient light fixture consumes 150kwh per year and has a EUL of 10 years, then the savings (all other things equal) would be 900 kwh (4x300 + 6x200)-10x150 = 900).

[8] See Michael Rufo, "Evaluation and Performance Incentives: Seeking Paths to (Relatively) Peaceful Coexistence," International Energy Program Evaluation Conference, 2009, p. 1036.

[9] See Efficiency Valuation Organization, *International Performance Measurement and Verification Protocol: Concepts and Options for Determining Energy and Water Savings Volume 1*, EVO 10000-1.2007, April 2007 and John Cowan and Steve Sain, *Measurement and Verification Fundamentals: the International Performance Measurement and Verification Protocol for Energy Managers and Emission Traders*, Study Manual for Association of Engineers Certified Energy Manager Training prepared by Efficiency Evaluation Organization, October 2008. While IPMVP is the best known M&V guidance, there are others, for example: American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE), *Guideline 14-2002 - Measurement of Energy and Demand Savings*, 2002; and Federal Energy Management Program (US FEMP), *M&V Guidelines, Measurement and Verification for Federal Energy Projects*, Version 2.2, DOE/GO-102000-0960, 2000.

The IPMVP was designed primarily to support the use of energy services performance contracts (ESPCs), in which an Energy Services Company (ESCO) makes energy efficiency improvements in a client's facility and is paid out of energy bill savings. Key to the success of ESPCs is an adversarial relationship between the client and the ESCO, with each party looking out for its own interests. As long as the two have symmetrical knowledge and power, then the model can work well for both parties. By telling clients what questions to ask and providing a framework for evaluating the ESCO's answers, the IPMVP has helped to legitimize the ESPC concept. But even with the IPMVP, the judgment of independent EMV experts is still needed. U.S. federal agencies, which use ESPCs for federal buildings and are the largest clients of ESCOs, rely on DOE-supported National Laboratories to review EMV plans and ensure that ESCO proposals adequately answer the many questions provided in the IPMVP.

The IPMVP, like other M&V guidance documents, is seldom prescriptive. It tells the M&V professional what questions need to be answered and what issues to take into account in answering them. But it generally does not provide the answers. And it never provides the data.

Evaluation of Net Savings

NAPEE defines evaluation as: [10]

> *The performance of studies and activities aimed at determining the effects of a program; any of a wide range of assessment activities associated with understanding or documenting program performance, assessing program or program-related markets and market operations; any of a wide range of evaluative efforts including assessing program-induced changes in energy efficiency markets, levels of demand or energy savings, and program cost-effectiveness.*

Once the M&V is complete and savings have been determined to be real, evaluation of savings is largely an exercise in savings attribution. Attributing savings to specific programs (and projects) has always been one of the greatest and most contentious challenges faced by EE evaluators. EE programs are intended to induce energy savings that would not have happened otherwise. Adjusting for free riders and spillover requires the development of BAU projection for the program – what would have happened without the program? Many factors can determine whether a project or measure would have been implemented without the program, including the state of the economy, energy prices, capital turnover, technology adoption, and, very important of late, other programs and policies.

It may be very difficult or even impossible to attribute specific EE measure implementation and energy savings to certain types of programs, such as public information and education. Yet such programs still may be very important and even a prerequisite to the success of other more measurable EE programs. For instance, utility bill inserts, public service advertising, or elementary school energy education programs could prime customers for participation in coupon, rebate, or home energy audit programs.[11]

---

[10] NAPEE, 2007, p. B2.

[11] The difficulties of measuring and attributing savings from behavioral programs are discussed and argued about in a Connecticut DPUC docket, 05-07-19RE02, with Efficiency 2.0 and others supporting behavioral programs and the

In attempting to parse out the various influences on BAU investments in energy efficiency, evaluation employs many tools associated with the social sciences including surveys, sampling and econometric "bill" analysis.

Underscoring the importance of estimating net savings, only about one-quarter of US energy savings that occurred between 1973 and 2003 (the latest year available) were induced by policies and programs.[12] The remaining three-quarters of savings were BAU savings, resulting from energy price changes, natural turnover of capital and technology development.

The notion that program savings should be additional is not itself so controversial. The controversy occurs in how that additionality is determined or tested. The test that is used can have enormous impacts on estimates of program savings and costs. Furthermore, the value of EE program-related carbon offsets or energy savings credits depends on their relative scarcity. If credits are issued to anyone who reduces energy use, their value will be greatly diminished, and thus, the willingness of project developers to come forward with new projects will also be reduced. And to the extent that carbon offsets or energy savings credits offer a significant path for compliance with a carbon cap or EERS, respectively, a weak additionality test also could undermine the effective stringency of those regulations.

Many additionality tests have been discussed and argued by various stakeholders over the last several decades, including:[13]

- *Legal, Regulatory, or Institutional Test* – Did the program or project energy savings exceed the level required by any official policies, regulations, guidance, or industry standards?
- *Technology or Common Practices Test* – Did the program or project involve a technology that is not "business as usual?"
- *Investment or Financial Test* –Without the program's support, would the project have a lower than acceptable rate of return?
- *Market Barriers Test* – Did the program help overcome otherwise prohibitive barriers to implementation (e.g., information barrier, split incentives)?
- *Timing Test* – Was the program or project initiated after a certain date (*e.g.* the date the EERS was instituted)?

---

utility United Illuminating Company and the Office of Consumer Counsel among the most vocal opponents to creating ESCs from behavioral programs.

[12] Based on Alliance estimates of 43 quad savings between 1973 and 2003 – equal to half of the reduction in energy intensity. Kenneth Gillingham, Richard Newell and Karen Palmer, *Retrospective Examination of Demand-Side Energy Efficiency Policies*, Resources for the Future, 2004, p. 3, estimated savings collectively of 4 quads from utility demand side management programs, building codes and appliance standards. The National Research Council, *Effectiveness and Impact of Corporate Average Fuel Economy Standards*, 2002, p.3, estimated 2.8 million barrels per day savings from CAFE standards in place at that time, which equals about 6 quads per year. Of course, all the estimates used for these calculations are subject to same challenges discussed in this report. As such, they are suggestive, not conclusive.

[13] This list is adapted from Mark Trexler, Derik Broekhoff and Laura Kosloff, "A Statistically-Driven Approach to Offset-Based GHG Additionality Determinations," *Sustainable Development Law and Policy*, Volume VI, Issue 2, Winter 2006, pp.30-40.

- *Performance Benchmark Test* – Did the program or project achieve an energy consumption level that is lower than a predetermined benchmark level of energy consumption for the particular technology or practice?
- *Project In, Project Out Test* – Was energy consumption lower after the program or project was implemented than before?

Other tests that are commonly proposed include:

- *Utopian* – Anyone reducing energy is "doing the right thing" and should get credit.
- *Bureaucratic* – If a company or household is willing to process the paperwork, it should be considered additional.
- *Expedient* – Since we cannot figure out how to deal with the complexities involved in determining additionality, we should assume net-to-gross ratio of "one" and move on.

All of the additionality tests are either conceptually or practically challenging. For example, the market barriers test may be more conceptually compelling than the financial test, since a project can be economical but still not get implemented. But the market barriers test is probably more difficult to implement since market barriers are difficult to measure or even know whether they exist. The timing test is relatively easy to implement, but conceptually is not so compelling, since the date of project initiation only tells us the project *could* have been induced by the program, not that it was.

Care must be taken that a test does not simply become a circular argument for making all projects additional. For example, it is sometimes surmised that if a financially viable project has not been implemented, there must be market barriers. By this standard, any financially viable projects would be considered additional if it has not been implemented. Similarly, the utopian, bureaucratic and expedient tests, while easily implemented, would allow any project or program to be counted.

Spillover

Additionality tests focus on netting out free riders. But as noted above, savings also occur outside the program scope when non-participants implement measures due to exposure to participants or when, due to the program, participants implement other types of EE measures that were not within the scope of the program. Spillover can be an intentional part of program design, such as a CFL giveaway program used to get consumers interested in EE and allow program administrators to introduce them to other energy savings opportunities.[14] Inadvertent peer to peer communication can also have a profound effect on consumer behavior and investment decisions, whether we know it is occurring or not.

Over the longer term, programs can transform both private and public "markets" for EE. Programs can build critical mass in the marketplace and, especially where there are economies of scale associated with manufacture and sale of those products, could help reach a tipping point

[14] Beyond spillover effects, in Connecticut, at least, some observers are concerned that CFL giveaways also imply to consumers that the products are not cost-effective or market-viable on their own merits. Office of Consumer Counsel, 'Brief of the Office of Consumer Counsel, May 18, 2009, Docket 09-02-18.'

where the EE products are mainstreamed and costs fall to a point where they effectively compete with standard products without ratepayer or government subsidies. Once the more efficient products make it into the mainstream, government policies such as standards can begin to push less efficient products out of the market through regulation.

For example, numerous utility and government programs have provided rebates and awareness of CFL savings opportunity. Over the last few years, large retailers – e.g., Lowes, WalMart, and Home Depot – started to market and give significant shelf space to CFLs. Once CFLs were mainstreamed, Congress was able to enact a new standard that will effectively phase out incandescent general purpose lamps starting in 2012.[15] Without this history of CFL programs, leading to improved quality, reduced costs and growing public acceptance of CFLs, it is not clear that Congress would have passed a general purpose incandescent bulb phase-out.

While CFL programs have surely helped transform lighting markets, attributing the impact to any particular program(s) is difficult given literally hundreds of programs of various types run by numerous administrators at various places and times over a long period. It is also difficult to estimate the aggregate effect of past programs – there are plenty of examples of new technology becoming mainstream without government support and plenty of examples of technology not becoming mainstream with government support.

In sum, estimating spillover, especially market effects, may be even more difficult than measuring free riders. At least with free riders, evaluators can know who the program participants are and then try to discern whether those participants would have implemented the project or measure without the program assistance – they can know the population of possible free riders. Free riders can be like searching for a needle in a haystack when you do not know whether the needle is there. With non-participant spillover, the evaluator does not even know which haystack to search in.

Across states, treatment of free riders and spillover is diverse. Some count only free riders. This method increases the likelihood (all other things equal) that savings estimates will not overstate actual savings. Some count only spillover. Some states try to count both free riders and spill over. Some states take the most expeditious approach, basically assuming free riders and spillover will cancel each other out and counting neither.[16]

Attribution is getting even more complicated

The difficulties of attributing savings have been compounded over the last couple years as the nation faced an historic downturn in the economy, wild swings in energy prices and revolutionary information and controls technology advances. These factors can shift BAU energy use in unpredictable ways. For example, as the economy has slowed, businesses and households have postponed purchases of new cars, appliances and homes, thus slowing

---

[15] The Energy Independence and Security Act of 2007establishes a performance standard for general purpose lamps that cannot be met by conventional incandescent filament light bulbs.
[16] These differences were highlighted in Julie Michals and Elizabeth Titus, *The Need for and Approaches to Developing Common Protocols to Measure, Verify and Report Energy Efficiency Savings in the Northeast*, Northeast Energy Efficiency Partnerships, 2006, p.19.

efficiency gains that would have occurred “naturally.”[17] On the flip side, energy price volatility has caused people to be more concerned with energy consumption than a few years ago. At least some of these changes could persist even if prices stay low since people have been reminded that energy prices will not necessarily stay low and the severe economic downturn has reminded them the economic growth will not always be high.

In addition, significant changes in BAU energy use have come from increases in federal and state spending on EE programs and the plethora of new EE policies. As a result of more aggressive energy programs and policies, the boundaries between programs have become increasingly blurred. Customers, as well as equipment and services suppliers, may be affected by multiple programs and policies. Thus it is becoming even harder to determine the impacts a particular program is having. Dealing with the additionality question was hard enough when there was relatively little overlap between programs. It is even more unmanageable now.[18]

The Energy Independence and Security Act of 2007, perhaps the single most significant energy efficiency legislation for the last three decades, created several important new efficiency standards, including the phase out of general purpose incandescent lamps. CFL programs dominate most EE program portfolios, with more than half of ratepayer funded program savings coming from CFL programs.[19] As CFLs become the default option for consumers (the new BAU), the potential programmatic savings from CFL programs will be lower. Some states and utilities have begun discounting savings from CFL programs to account for increased market saturation, but there is significant ongoing debate about the remaining potential savings from CFL programs.[20]

The federal economic stimulus bill in March 2009 provides in excess of $20 billion for various technical assistance and financial incentives for energy efficiency. Other major sources of funding includes the Northeast’s Regional Greenhouse Gas Inventory, which to date has generated more than $400 million in revenues, much of which is dedicated to EE programs.[21] Many of these state and local programs will target the same customers and measures as existing utility programs.

---

[17] For example, as a result of the current recession, the average car on the road in Washington, DC – an area only mildly affected by the recession – is now 6 months older than it was in 2005, which is creating concerns about local air quality in the region. Katherine Shaver, “Older Cars Fouling Region’s Air Quality: Agency Study Links Recession, Pollution,” *The Washington Post*, July 4, 2009, http://www.washingtonpost.com/wp-dyn/content/article/2009/07/03/AR2009070302451.html?sid=ST2009070400257.

[18] For an excellent discussion of program overlap in California, see Doug Mahone and Wim Bos, “A New Energy Efficiency Portfolio - Developing an Expanded Evaluation Approach,” *International Energy Performance and Evaluation Conference Proceedings*, August 2009.

[19] According to Department of Energy, *Big Results Bigger Potential: CFL Market Profile*, March 2009, p.1, http://www.energystar.gov/ia/products/downloads/CFL_Market_Profile.pdf, nationally, more than 60 percent of program savings have come from CFLs and some utilities are getting as much as 97 percent of program savings from CFLs.

[20] For an interesting overview of data related to CFL market saturation see Stephen Bickel, “Maximizing Energy Savings with CFLs: Don’t Bench Your Superstar,” American Council for an Energy Efficient Economy, Market Transformation Conference, 2009, http://www.aceee.org/conf/09ee/pdfs/(5B_Bickel)EER2009.pdf.

[21] Nathanial Gronewold, “Climate: Prices take a sharp dip in fifth RGGI auction,” *Greenwire*, September 11, 2009, http://www.eenews.net/public/Greenwire/2009/09/11/5.

In sum, complete attribution of savings to certain EE programs requires netting out savings that are the result of other government policies, changes in economic growth or energy prices and other outside factors while adding in the savings that result from policies that were made possible by the programs. When the world is changing slowly, this requires a herculean effort by very knowledgeable evaluators using some heroic assumptions. When the world is changing rapidly, it may be impossible to attribute savings with any real certainty.

Deemed Values

Savings from some common measures or the assumptions used to calculate savings may be deemed rather than actually being measured. Many program evaluations rely on deemed (ex ante) assumptions to some degree, but they can be especially useful for common and homogeneous measures – such as residential lighting, appliances and HVAC equipment. Individually, these measures may save little energy, but in aggregate they comprise a large percentage of the savings claimed by energy-savings programs.[22]

Deeming savings assumptions is considerably cheaper and simpler from an administrative standpoint than relying on direct measurement of savings, especially when there are many measures of a similar type. But it is important to note that ex ante savings estimates do not reduce the level of effort necessary to accurately measure or verify savings. Deeming savings simply moves the difficult and often contentious decisions to before the program implementation rather than after.[23]

In systems in which evaluation results are used to determine the level of compensation given to program administrators, the choice between deemed and metered savings affects the amount of risk the administrators assume. Ex ante savings assumptions reduce the program administrator's risk of non-performance. Whether or not a given measure or program actually saves energy no longer matters (at least until the deemed savings are reevaluated). Compensation based on ex post savings involves greater risk to the administrator of non-performance.[24]

Deemed savings assumptions can vary widely.[25] As Table 3 shows, a 15W CFL installed in Vermont would be deemed to save 332.4 kWh over its lifetime, but only 149.3 kWh in Connecticut and even less in California at 118.76 kWh. Cumulative lifetime savings from the same bulb installed in the same house in the same type of room will be 180 percent higher in Vermont than in California and twice as high as Connecticut.

---

[22] Different programs rely on deemed savings to varying extents. New York, for instance, uses deemed savings values for only 20 percent of its portfolio, with the majority of savings measured using actual run hours and operating conditions. Other states, like Pennsylvania, rely much more heavily on deemed savings estimates. Sources: Cherie Gregoire, Impact Assessment Program Manager, Energy Analysis, NYSERDA, March 26, 2009; Pennsylvania Public Utility Commission, Implementation Order, Docket No. M-2008-2069887, Public Meeting held January 15, 2009.
[23] Rufo, 2009, p. 1032.
[24] *Ibid*, p. 1032-1033.
[25] Michals & Titus, 2006, p. 8. Michals and Titus flagged this issue and identified assumptions and methods in which TRMS in the Northeast varied.

Table 3: 15W CFL installed in a living room, with a rated lifetime of 10,000 hours	CT	PA	VT	CA	Energy Star	CT before 10/1/09[26]	MA[27]
Wattage of old bulb	51	varies depending on individual measures	63.7	52.95	60	51	TBD
How the equivalent was calculated	3.4 x CFL wattage	Measured	CFL W + 48.7	3.53 x CFL wattage	4 x CFL wattage	3.4 x CFL wattage	
Change in Wattage	36	varies depending on individual measures	48.7	37.95	45	36	
Hours/day	2.96	3	3.4	2.18	3	2.96	
Gross Watts saved per day (Change in Wattage x Hours/day)	106.56	Varies depending on individual measures	165.58	82.73	135	106.56	
Gross Kilowatts saved per year (Gross Watts saved per day x 365/1000)	38.89	Varies depending on individual measures	60.44	30.20	49.28	38.89	
Annual Gross Gas Savings (Therms)	not given	not given	not given	-0.99	not given	not given	
Life in years	5.25	6.4	6.4	6.6	9.13	6	
Life in Hours (Hours/day * 365 * Life in years)	5672.1	7008	7942.4	5251.62	9997.35	6482.4	
Lifetime Gross Electricity Savings (kWh)	204.2	varies depending on individual measures	387.5	197.93	450.1	233.4	
In-Service Rate	70%	84%	73%	not given	not given	70%	
Free ridership	9.2%	NTG = 1	6%	not given	not given	9.2%	
Spillover	15%		25%	not given	not given	15%	
Net Realization Rate	74%	84%	86%	60%	not given	74%	
Annual Net Electricity Savings (kWh)	28.4	varies depending on individual measures	71.1	17.99	49.3	28.43	
Annual Net Gas Savings (Therms)	not given	not given	not given	-0.59	not given	not given	
Lifetime Gross Electricity Savings (kWh)	204.2	varies depending on individual measures	387.5	197.93	450.1	233.4	
Lifetime Net Electricity Savings (kWh)	149.3	varies depending on individual measures	332.4	118.76	450.1	170.6	
Lifetime Gross Gas Savings (Therms)	not given	not given	not given	-6.52	not given	not given	
Lifetime Net Gas Savings (Therms)	not given	not given	not given	-3.91	not given	not given	
Lifetime Net Energy Savings (kWh including natural gas converted to kWh)	149.3	varies depending on individual measures	332.4	82.81	450.1	170.6	

[26] There is an extra Connecticut column because the state recently updated its CFL estimates in order to reflect the looming federal incandescent phase out. The only change is the lifetime in years that a CFL is now expected to last, a number which will be repeatedly updated in the next few years. In general, many Connecticut shareholders consider common CFL savings dubious not only because of the incandescent phase out but because CFL prices are now similar to incandescent lamps. Office of Consumer Counsel, 2009, p4.
[27] Massachusetts is in the process of developing a TRM but has yet to publish one.

The different savings estimates stem from differences in a variety of assumptions used in the calculation of gross savings (including assumptions about the wattage of the incandescent light bulb being replaced, the number of hours per day the CFL is in use, the number of years the CFL will last before burning out and whether or not interactive effects of the installation on non-lighting energy consumption is considered[28]) and of net savings (including the in-service rate, the free ridership rate and the spillover rate). Many of the assumptions should be different, however, to reflect differences in CFL market saturation, for example. Some of the differences in assumptions are difficult to explain – e.g., the change in wattage or life of bulbs.[29]

Table 3 illustrates how complex savings calculations for even the simplest type of measure can be – a residential CFL is perhaps the simplest and most commonly used energy-efficiency measure. Estimating the savings from a commercial lighting or HVAC retrofit project is far more complicated. Deemed savings assumptions do not eliminate the necessity to study saturation rates or understand the interactions between lighting and HVAC loads, but it does allow the costs of those studies to be spread across more measures.

[28] CFLs give off less waste heat than incandescent bulbs. As a result, a building's cooling load is reduced with the installation of the efficient bulb and its heating load is increased.
[29] For additional discussion on these CFL comparisons, see Appendix B.

Right-Sizing Certainty

Estimating programmatic savings will always involve some degree of uncertainty. The counterfactual nature of energy efficiency savings measurement ensures that we can never know for sure how much energy a program saved. Still, additional certainty can be had at a price. The question is how much certainty do we need and how much are we willing to pay for it?

This chapter discusses definitions of certainty, how certainty is measured (or not), the price of certainty and the cost of uncertainty, and provides thoughts on balancing certainty and costs.

Types and Sources of Uncertainty

The level of uncertainty is a function of the various sources of error that can be found in any estimation process, including:[30]

- *Measurement errors* that can result from inaccurate meters or errors in recording data.
- *Collection errors* that can arise from non-representative sampling, refusal by some in the sample to participate in a survey, biased responses or interpretation of responses, poor questionnaire design, failure to take into account behavioral factors and other sources of bias.
- *Modeling error* that may come from incorrect use of models, incorrect specification of relationships between variables, improperly included or excluded information or data, incorrect assumptions (stipulations) and other modeling deficiencies.

These sources of error combine in complex ways that can reinforce or offset one another.[31]

Errors can be random (statistical) or systematic (biased). Random errors are present in any measurement and are caused by unpredictable differences in the readings of a measurement apparatus, misinterpretation of the data and incorrect (but unbiased) responses to surveys. With random errors, the average of many repeated measurements would be close to the actual value, although the error associated with a single observation could be high. Reducing random error increases the precision of the savings estimate.

Systematic errors are biases where the average of multiple measurements differs significantly from the actual value. Systematic errors are not subject to chance, but rather are the result of evaluation "decisions and procedures" including the way that data are measured, collected and modeled.[32] Systematic error reduces the accuracy, as opposed to the precision, of savings estimates.

[30] NAPEE, 2007, pp. D1-D2.
[31] Cowan & Sain, 2008, p. 51.
[32] NAPEE, 2007, pp. D1-D2.

Increased precision is not much help if the estimate is precisely wrong. Reducing sampling error by increasing sampling size could be an illusion altogether if there are major systematic non-random errors embedded in the data.[33]

Measuring and Correcting Errors

If repeated measurements are available, random error can be estimated using common statistical methods.[34] Random error can be reduced by increased sampling, or otherwise increasing the size of the data set. The level of random error can often be statistically estimated and reported, allowing the balancing of additional evaluation costs with the desire for greater precision. Acceptable levels of random error for measuring devices are suggested and sometimes prescribed in various project evaluation guidance documents, including, for example, the IPMVP, which says savings should be at least twice the standard error.[35]

Addressing systematic error (bias) is far more challenging than addressing random error. Whereas random error can be discovered and estimated with repeated measurements, which will approach the true value, systematic error cannot since repeated experiments err in the same direction.[36] Increasing sample size – in attempt to increase precision – could amplify the error and skew the estimate further.

An analogy used to explain the difference between random and systematic error is shooting at a target. If shots center on the bull's eye but are scattered, that would represent random error – repeated shots, even if none actually hit the bull's eye, would reveal the location of the bull's eye. Systematic error, by contrast, would not center on the bull's eye. The amount of variation could be small (precise), even with every shot hitting the same exact spot, but the shots may be nowhere near the bull's eye. Knowing this error was occurring would be very difficult if one did not know where the bull's eye was to begin with.[37]

Moreover, systematic error is insidious. We often do not know it exists, and if we know it exists, we often cannot quantify it.[38] If we could quantify it, we could eliminate it or adjust for it. Any guidance for acceptable levels of systematic error would be simple – the acceptable level of known systematic error is, in effect, zero.

Unfortunately, there is no way to recalibrate the EMV methods or assumptions to make them more accurate since, due to the counterfactual, the actual savings are never really known, even many years after the fact. A challenge with EE EMV is that not only are future savings uncertain, so are past savings.

---

[33] Personal correspondence with Steve Kromer.
[34] For excellent discussions of systematic and random error (and uncertainty in general), see John R. Taylor, *An Introduction to Error Analysis, the Study of Uncertainties in Physical Measurements, 2nd Edition*, University Science Books, 1997, pp. 94-97.
[35] Cowan & Sain, 2008, p.52.
[36] Taylor, 1997, pp. 94-97.
[37] *Ibid*.
[38] One could know, for example, that self selection bias exists in a survey and even know the direction of the bias, but not know the extent or how much it affects the estimates.

Despite the futility of ever knowing whether EE savings estimates are really correct, devoting time and resources to reducing uncertainty of savings estimates through greater understanding of the programs and their impacts can increase the level of confidence that biases have been eliminated and actual and estimated savings are at least in the same ballpark.

Perhaps the best that can be done is to triangulate estimates – i.e., use multiple, independent evaluation methods for given projects and programs. For instance, a project or program may be assessed by both direct measurement and user surveys. If different methodologies deliver corroborating results, then there should be more confidence in the accuracy of the results.[39] One would still need to be careful of the possibility of the different methods possibly sharing a common source of systematic bias (say, the same assumptions of the relationship between two variables or the same deemed savings assumptions).

The Price of Certainty

But increased evaluation certainty has costs:

- Increased certainty can be expensive. Nationally, EMV costs exceeded $100 million last year, comprising about 4% of total program budgets. Over the last 3 years, California spent roughly 8 percent of program budgets on EMV.[40] Of course, money spent for evaluation is not available for the programs themselves.
- Increased certainty takes time. It can take years to collect data for an intensive impact evaluation, especially if the measures being evaluated are long-lived and have unpredictable persistence. If a program evaluation takes too much time, its usefulness in terms of providing feedback to guide changes in design or administration of programs, or for determining performance-based compensation levels, could be diminished.[41] By the same token, this should not be an excuse for shoddy evaluations – if the savings estimates are wrong, their usefulness in guiding program development will be diminished.
- Increased certainty can lead to lost opportunity. Savings from some types of programs are less easily quantifiable than others. For example, savings from consumer awareness programs are more difficult to measure than direct install programs. The difficulty measuring savings from these programs does not mean that savings are not being realized; in fact, the savings could be great even if they are hard to quantify or attribute to the program. If program portfolio or compensation decisions are based on measured savings, the portfolio will emphasize easily measurable programs.

[39] The California Public Utility Commission's evaluators' protocols refer to this approach as "triangulation," meaning "comparing the results from two or more different data gathering or measurement techniques on the same problem to derive a "best" estimate from the analysis of the comparison. California Public Utilities Commission, *California Energy Efficiency Evaluation Protocols: Technical, Methodological, and Reporting Requirements for Evaluation Professionals*, prepared by The TecMarket Works Team, April 2006, p. 243.
[40] In 2008, $2.01 billion was spent nationally on electric energy EE programs in 26 states plus about $384 million in load management programs. EMV budgets for these programs amounted to $102.4 million, or about four percent of total EE and load management program cost. Consortium for Energy Efficiency, "CEE 2008 Annual Industry Report," 2008, Table 4, http://www.cee1.org/ee-pe/2008/us_electric.php.
[41] Nick Hall and Patrick McCarthy, "Portfolio Evaluation Versus Program Evaluation," *International Energy Program Evaluation Conference*, 2009, p. 480.

Balancing Certainty and Costs

Conceptually, balancing costs and certainty is simple. Increased savings uncertainty represents a cost that should be subtracted from the net benefits of alternative programs or projects. For instance, a utility or grid operator uncertain about the level of EE being delivered may need to procure additional energy supply as insurance against possible EE shortfalls. If the level of certainty and costs associated with increasing certainty are known, the guidance would be to buy more certainty until the marginal benefit of the increased certainty equals the marginal cost of the increased certainty. Unfortunately, as discussed above, we never really know how much systematic uncertainty there is or how much uncertainty is reduced through more evaluation.

To date, there has been minimal use of quantitative uncertainty analysis related to M&V for projects.[42] Adjusting for risk at the program level is even more difficult. Attempts to develop quantitative performance risk models, including efforts by Enron in the 1990s to develop actuarial tables of EE project risks, have so far been thwarted by insufficient project (and program) data.[43] A large portfolio of project (and program) data, using standardized EMV on standardized projects, is required to develop a satisfactorily robust analytic tool.
Efforts to generate even small-group consensus on various prescriptive approaches for discounting savings estimates to reflect measurement uncertainty have also been frustrated. The best and most rigorous discussion of how to assess and address uncertainty that we have seen was produced in 2002 by the IPMVP Adjustments Committee, which developed and analyzed five options for assessing and discounting for project uncertainty:[44]

- Option 1 would apply different discount rates to the four different IPMVP measurement and verification options.
- Option 2 would require estimation and reporting of confidence intervals for different savings estimates.
- Option 3 would conduct a qualitative assessment of uncertainty and rating on a scale of 1 to 10.
- Option 4 would ask a range of question relating to the data collection and analysis to assess whether acceptable practices are being employed.
- Option 5 would discount savings depending on the extent to which savings are measured.

Ultimately, none of the options were able to garner consensus within the IPMVP Adjustments Committee. And as the authors acknowledged, even if there had been agreement, guidance

[42] Paul Matthew, Eric Koehling and Satish Kumar, "Use of Quantitative Uncertainty Analysis to Support M&V Decisions in ESPCs," *AEE Journal Energy Engineering*, v. 103, n. 2, 2006, p. 25.
[43] For discussion of how financial risk analysis techniques might be applied to energy efficiency projects, see Evan Mills, Steve Kromer, Gary Weiss and Paul Mathew, "From Volatility to Value: Analyzing and Managing Financial and Performance Risk in Energy Saving Projects," *Energy Policy* 34, 2006, pp. 188-199. For discussion of Enron's attempts to create actuarial models for EE projects, see Paul Mathew, J. Stephen Kromer, Osman Sezgen and Steven Meyers, "Actuarial Pricing of Energy Efficiency Projects: Lessons Foul and Fair," *Energy Policy*, 33, 2005, pp. 1319-1328.
[44] See Ed Vine, Gregory Kats, Jayant Sathaye and Hemant Joshi, "International Greenhouse Gas Trading Programs: A Discussion of Measurement and Accounting Issues," Energy Policy 31, 2003, pp. 211-224, which summarizes and discusses approaches developed by the DOE IPMVP Adjustments Committee in 2000.

related to defining the many source of uncertainty and how to calculate standard deviations and confidence intervals would still be needed.

Even more challenging, the options dealt only with direct energy savings from specific measures or projects and did not address more uncertainty associated with measurement of free riders, spillover, or other net-to-gross issues. We are unable to find rigorous discussion of, much less agreement on, addressing uncertainty related to net-to-gross calculations anywhere in the literature.

In the absence of more formal methodologies, a leading expert in the evaluation field suggests some factors to consider in balancing evaluation certainty and costs:[45]

- Size of the program in dollars
- Size of the program's projected energy impacts
- Newness of the program or the delivery approaches
- Newness of the measures covered
- Standard operations of the market and market sector being served
- Measures included in the program
- Marketing, enrollment and delivery approaches for the program and the program's measures
- Participants' energy use conditions and factors affecting the variation in energy use (weather, operations, timing, etc.)
- Estimated risk factors associated with the measure and the program-specific use conditions
- Stability of the savings expected
- Rigor of past program evaluations and confidence in the results
- Changes to the market, the program or the measures
- BAU conditions associated with the measures and their use conditions
- Market operations changes that impact measure selection and use decisions and options
- Expected accuracy and comprehensiveness of program-influenced load shape distributions
- Expected accuracy and comprehensiveness of program-influenced deemed savings metrics
- Cost of the various evaluation approaches
- Needs for special studies to support policy decisions
- Needs for special studies to support program or portfolio construction
- Past program management performance or impact-related program operational concerns
- Familiarity with the program administrator and service contactors
- Experience of the service provider
- Potential for waste, fraud or abuse

Of course, these leave a lot to professional judgment. There is little guidance here in terms of what a more rigorous evaluation would actually entail or whether the emphasis would be on

[45] Hall & McCarthy, 2009, pp. 478-479.

precision, accuracy, both, or neither. Nor do we know what defines a large program, presumably meriting more rigorous evaluation, or a small program – $1 million, $100 million? This is by no means intended as a criticism, but rather an observation that for this type of guidance to be useful to policy makers it would need far greater definitional clarity and professional judgment to prioritize the importance of various factors in various contexts.

Keeping it in Perspective

EE evaluation is not uniquely challenged. Most government activities intended for societal benefit are not measurable to a large extent. Do we, and will we ever, really know the benefits (or even costs) of military actions, public education, health and welfare programs, federal insurance for nuclear generators, tax incentives for oil and gas production? Estimates of impacts of many of these programs can contain far greater uncertainty than EE programs.[46] There is a perennial struggle to rigorously evaluate the impacts of government programs, but the final evaluation of these programs usually occurs in Congressional hearings and Sunday morning talk shows.

The uncertainty and lack of rigor in estimating the impacts of government programs is constantly lamented but accepted as being unavoidable. Few people suggest, at least successfully, that government should be closed down due to our inability to measure the impacts of government actions. For example, few would argue that a police department should be shut down because "EMV" does not allow good estimation of the department's productivity (how many instances and what severity of crime deterred or auto accidents avoided in addition to arrests and accident responses) or cost-benefit ratio. In general, more uncertainty is acceptable in estimates of savings from government programs than estimates of savings used for other purposes.

By contrast, the most common justification for ratepayer funded EE programs is that they are a cost-effective alternative to investing in supply-side resources. Though it is often made in a very public environment, if the objective is cost-effective provision of energy services to ratepayers it should be treated as a private investment decision. The required level of certainty in determining whether the EE programs are cost-effective could reasonably be higher than if a broader societal objective is the driver. In the latter case, the programs are competing with other programs benefiting society, many of which have highly uncertain (though important) benefits.

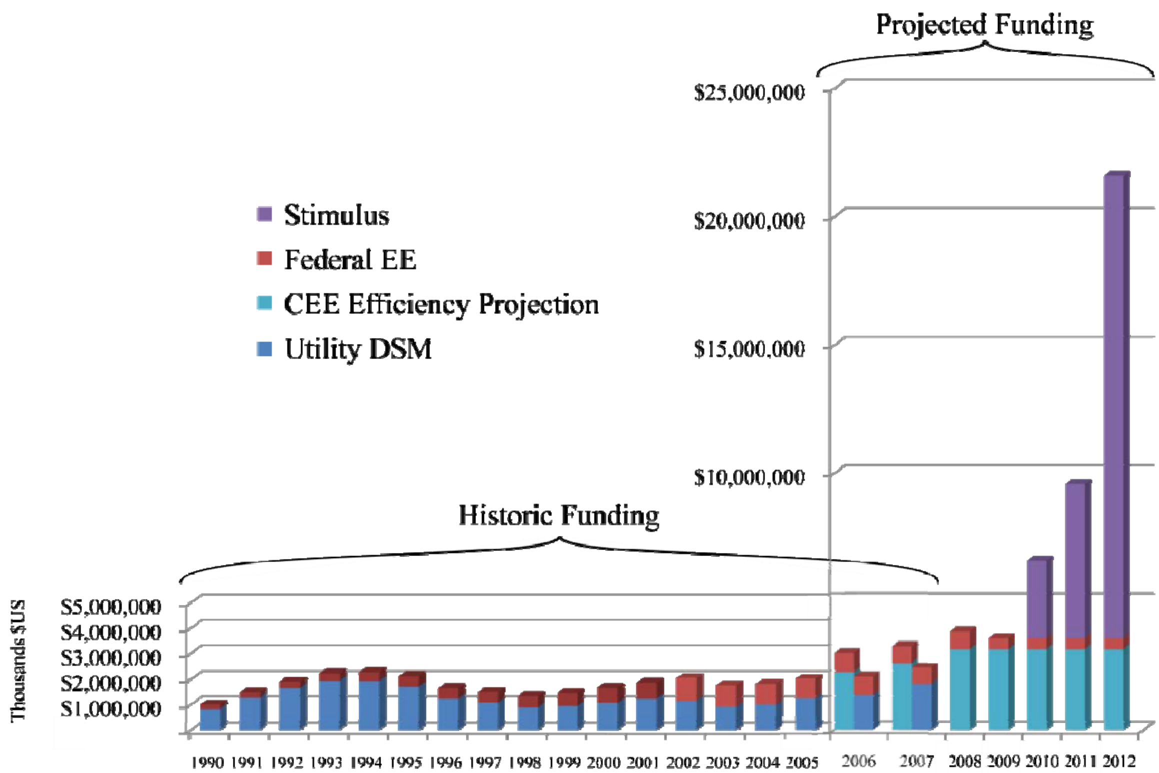
Required certainty also could be higher as program spending, and related savings claims, increase. As shown in figure 2, pending from the federal economic stimulus, along with growing state and utility budgets for energy efficiency programs, has increased dramatically. And the national funding masks wide variations among states – Massachusetts, for example has proposed to spend $1 billion over the next 3 years, equal to roughly 12 percent of total electric revenues. Increased funding has increased scrutiny of EE program effectiveness and costs. It is also making it harder for utilities and regulators to ignore programmatic savings in planning for future capacity needs.

---

[46] Author's observation based on participation in courses offered by The Evaluators' Institute in Chicago, IL, April 22-23, 2009. Participants from a wide range of disciplines including health care, education, museums, environment and more shared many experiences that were quite similar to those faced by energy efficiency program evaluators.

To the extent program savings are included in load forecasts, under or overestimating those savings could have a material effect on generation and transmission capacity requirements. For example, if utility planners expect 20 percent of demand to be met through EE programs in 2020 and only half is actually realized, most of their reserve margin will have evaporated. This would increase the likelihood of power supply disruptions. Perhaps even more important, supply shortages could fuel "emergency" construction of inefficient gas/oil "peakers" with heat rates twice as high as base load plants.[47]

Figure 2: Government and Utility Funding for Energy Efficiency[48]



Source: Alliance to Save Energy based on data from EIA, Electric Power Annual with data for 2007, January 2009 and CEE, "CEE 2008 Annual Industry Report," http://www.cee1.org/ee-pe/2008/us_combo.php.

Moreover, utility program administrators have historically been allowed to recover their program costs, with perhaps some additional bonus or penalties based on performance. To entice utilities to more aggressively pursue EE savings opportunities, some states are experimenting with innovative performance incentive mechanisms, including California utility shareholder incentive and Duke Energy's Save-a-Watt program, variants of which have also been approved in Indiana

[47] Personal correspondence, David Bellman, November 19, 2009.
[48] Federal spending includes only the DOE Office of Energy Efficiency and Renewable Energy budget. It does not include spending by the US Agency for International Development and EPA's ENERGY STAR program, for example. State government funding is not included except for 2006-2008. Federal energy services performance contracts and utility energy services agreements also are not included.

and North Carolina.[49] These mechanisms have greatly increased the stakes of saving estimates in those states. In California alone, hundreds of billions of dollars in utility incentives are riding on decisions relating to how energy savings are measured.[50]

The stated objective of EE advocates for decades has been to place EE on a level playing field with energy supply. Competing EE with power supply may require either that the certainty associated with EE program savings measurement be on par with power supply measurement or that methods for adjusting EE savings to account for that uncertainty be developed.

To be sure, supply side investment decisions involve considerable uncertainty as well. Cost overruns associated with power plant construction are commonplace. Projected power delivery may be widely different than actual due to delays in plant completion, unplanned outages, changes in dispatch due to changes in fuel prices and a host of other factors. Projecting future load requirements, which are affected by economic growth, energy prices and weather and a host of other factors, is also an uncertain enterprise.

But while there may be significant uncertainty associated with projecting power system costs and requirements, we eventually know with a high degree of certainty how much power was delivered and its cost. Thus the projections used for power system planning can be periodically recalibrated based on historical data. As noted above, the data to allow recalibration of EE program savings (i.e., discounting) or the redesign of the evaluation methodologies and assumptions are not available.

Some ratepayer funded EE programs are justified on the basis of larger social objectives, such as reducing carbon emissions, improving local air quality, or promoting economic development. If the program justification is a social objective then a secondary question could be asked in deciding who should pay for it – are rate ratepayers responsible for the problem being addressed? For example, if the objective is to reduce carbon emissions, ratepayers (as electricity consumers) are responsible for the emissions associated with electric they consume. Ratepayer funded programs to reduce or mitigate carbon emissions would seem appropriate. Billing ratepayers for the programs internalizes the externality cost of carbon emissions into electric rates.[51]

Whether programs are paid for by ratepayers or taxpayers at large, when EE programs are intended to achieve broader societal objectives such as carbon emissions abatement it becomes less clear whether the programs should be compared to government programs in general (e.g., military, education, etc) or alternative carbon abatement measures (e.g., generation fuel switching, carbon capture and sequestration, etc). This is really a question of the scope of the portfolio being considered. The portfolio could be limited to a single utility's energy efficiency

[49] Peter Cappers, Charles Goldman, Michele Chait, George Edgar, Jeff Schlegel and Wayne Shirley, *Financial Analysis of Incentive Mechanisms to Promote Energy Efficiency: Case Study of a Prototypical Southwest Utility*, Lawrence Berkeley National Laboratory, Environmental Energy Technologies Division, March 2009, p.xiii.
[50] Rufo, 2009, p. 1031.
[51] Making ratepayers pay for an EE program justified on the basis of economic development is less clear cut, since electricity consumption does not make them beneficiaries of the economic development program nor does electric consumption impose costs related to economic development. At any rate, most states EERS requirements contain a jumble of justifications, including cost effectiveness, carbon emissions, local air quality, as well as economic development.

programs, the range of carbon abatement options available to the utility, the range of abatement options available nationally, the entire portfolio of government energy programs or the entire portfolio of government programs. None seems clearly more conceptually compelling than the others.

When the primary objective of a program has little to do with EE, it is particularly hard to decide a reference for comparison. A topical example is the American Reinvestment and Recovery Act of 2009, which provides $25 billion-plus in federal funding for energy efficiency programs over a 3-year period 2009-2011.[52] While these funds are an enormous boost to energy efficiency efforts in the US, the first objective was to stimulate the economy and create jobs. The objective was to quickly inject money into the economy. Achieving energy savings was secondary to this objective and the rapid roll-out of these programs will involve many state and local governments and others with relatively little experience or capacity to design or administer these programs. While every effort should be taken to ensure these funds are spent effectively, both the expectations of energy savings and expectations of savings certainty need to be managed. And the short lead time for these programs precludes development and implementation of an evaluation system that even comes close to meeting the standards generally expected of a ratepayer funded program evaluation.

The Debates Rage On

There are many schools of thought on how to move forward given all the uncertainty and other challenges with EE EMV (see Table 4). Of course, the opinions and recommendations are usually more nuanced, but the table gives an idea of the range of opinions and recommendations.

Evaluation challenges are well documented and have been a subject of many hot debates and conference proceedings for decades, including a biennial conference that is dedicated solely to the topic.[53] Unfortunately, little consensus has emerged or appears on the horizon. While the authors of this report would call themselves "No-Choicers," the prevailing rule of thumb for evaluation spending and rigor could come down to "gut checks" and evaluating until stakeholders are content that estimates are "close enough." In some cases, that may mean fewer EE programs, in others cases more, and in still others it will not make a difference. The only thing we can say for certain is that who is doing the gut check and which stakeholders are in the room will be critically important.

[52] Could be as much as $75b for EE (defined loosely to include smart grid, federal rehab money, etc) over 3-plus years, but more likely on the order of $25b. Alliance to Save Energy, "Stimulus Resources: The American Recovery and Reinvestment Act of 2009 (ARRA)," 2009, http://ase.org/content/article/detail/5461.

[53] International Energy Performance and Evaluation Conferences have been held every other August (in odd numbered years) since the mid 1980s.

Table 4: EMV Schools of Thought

Type	Opinion	Recommendation
Blissfuls	No awareness EE measurement is an issue	NA
Real Politikians	EE EMV is not perfect, but neither is evaluation of other activities	Do not set higher standard for EE
Economists	Bottom-up EMV is unreliable and cannot be improved	Use econometrics
Optimists	EMV is not perfect but getting better	Do more of it
Pessimists	EMV is unreliable and cannot be improved	Limit EE as resource
Deemers	EMV is unreliable and cannot be improved	Ignore the EMV
Differentiators	EMV is OK for program feedback and cost effectiveness, but not performance compensation	Limit impact assessment
No choicers	EMV is unreliable, but it is necessary to scale up EE, so there is no choice but to improve	EMV is a necessary cost – do more

Battling for Credibility

Uncertainty reinforces predispositions about whether EE programs are achieving claimed savings. Such predispositions may be the product of experience or hearsay, or general attitude toward life – e.g., believers vs. skeptics, optimists vs. pessimists. While EMV methods and data will continue to improve and while there may be ways to enhance the credibility of evaluation professionals as a community, at least some of the key sources of uncertainty (e.g., measurement of free riders and spillover) are unlikely to be resolved any time in the near future, if ever.

Even with better data, methods and commitment to fairness and objectivity by evaluators (i.e., "professionalism") at least for now, there are two choices – ignore the uncertainty associated with EE evaluation or rely more heavily on stakeholder participation in the evaluation process to minimize bias and perceptions of bias. The first option is increasingly untenable as funding for EE programs and EERS, carbon offsets and other performance-based mechanisms raise the stakes for evaluation results.

Credibility from Contention

Stakeholders may intentionally or unintentionally introduce bias through assumptions and methods that increase or lower savings estimates. For example, project developers seeking carbon offsets or program administrators compensated on the basis of estimated energy savings may propose assumptions and methods that will overestimate savings. Other stakeholders might have vested interests in seeing underestimation of savings.

These biases can be reduced to some extent with more and better information (to check the veracity of stakeholder assumptions and claims), but even the best information can be ignored or misinterpreted if there are no checks and balances. Ultimately, mitigation of actual or perceived bias may best be achieved through an "adversarial" process in which stakeholders interests counter one another.

The most contentious processes for deciding how EE program savings are measured could be the best at eliminating actual and perceived bias, especially if those processes are complemented by decision structures that value evidence and reason. The California shareholder incentive system engages a wide range of stakeholders in highly informed and transparent deliberations over how savings are measured. Hundreds of millions of dollars hinge on the results the evaluation results.

Some observers and participants argue that the EE evaluation sciences are insufficient for making these high stakes performance-based determinations. For example, a CPUC Energy Division White Paper states:[54]

> "The implementation of [the shareholder incentive mechanism] has revealed fundamental flaws which lead Energy Division to propose that the EM&V process, at least as it is currently designed and administered, cannot serve as a tool to simultaneously determine incentive awards

[54] California Public Utilities Commission, "Proposed Energy Efficiency Risk-Reward Incentive Mechanism," Energy Division White Paper, April, 2009, p. 7.

> or penalties and produce accurate estimates of energy savings without protracted disputes concerning the magnitude of specific values or the fairness of allowing those values to be updated and applied retroactively."

Similarly, in his analysis of the California shareholder incentive, Carl Blumstein concludes:[55]

> To change this direction we need first to recognize that program evaluation cannot precisely and accurately determine the counterfactual question, what would have happened in the absence of a utility's energy-efficiency programs? There will always be substantial uncertainty. Next, we need to reduce the stakes. Given the current state of knowledge, the decision to tie all of the incentive [sic] to program outcomes is misguided... This suggests that most of the utilities' compensation should be cost plus and only a relatively small share should be tied to the performance of the energy-efficiency programs that they administer.

It is worth asking, however, if evaluation is not sufficient to determine compensation, why is it sufficient to tell us what programs to fund? Why should regulators (or the constituents they represent) need a higher level of certainty for performance-based compensation than for traditional portfolio decision making? If they are willing to discount savings estimates for traditional portfolio decisions (say, require a benefit-to-cost ratio greater than two), why could they not simply impose similar discounts for performance-based compensation (say, multiply savings claims by 0.5)? Why should performance-based compensation raise the bar for EE evaluation?

EMV for performance compensation could be identical to EMV used for portfolio analysis. Portfolio decisions are based on ex ante savings estimates which are used to justify program spending decisions and program administrator cost-recovery. If a performance-based compensation scheme was solely based on ex ante savings estimates, then "cost-based" compensation may be no different than the "performance-based" compensation. If the ex ante estimates overstate actual savings (and cost-effectiveness), then they do so regardless of the basis for compensation.

Most performance-based compensation schemes involve ex post savings measurement, however. Basing compensation on ex post savings estimates does not eliminate uncertainty or bias. As noted above, it does, however, transfer the risk of non-performance to the program administrator from the ratepayer, giving the administrator a greater stake in the results of the ex post savings analysis.[56] The administrator does not necessarily care more than it did before whether the results are accurate or precise. The administrator's interest is increasing its compensation. Increased compensation comes from higher savings estimates, not more certain savings estimates. By the same token, ratepayers may benefit from understating savings in the ex post analysis since the greater the savings estimate, the more compensation they have to pay.

[55] Carl Blumstein, "Program Evaluation and Incentives for Administrators of Energy-Efficiency Programs: Can Evaluation Solve the Principal/Agent Problem?" *European Council for an Energy-Efficient Economy Summer Study*, Center for the Study of Energy Markets, March 2009, pp. 9-10.
[56] Rufo, 2009, p. 1030.

In other words, performance-based compensation creates an adversarial process in which stakeholders with differing financial interests will have incentive to question assumptions, data and methodologies that may lead to results counter to those interests. This then can lead to checks and balances that may work to counter high levels of under-or overstatement of savings.

Without performance-based compensation, there may be less incentive for regulators and program administrators to question savings estimates after the portfolio decision is made since both benefit more from program success than failure. Neither regulators nor administrators want to tell ratepayers that the programs they paid for not work as stated. While litigating performance-based compensation can create conflict and discomfort (at least in California), the EMV becomes more transparent in the process and the savings estimates are probably greatly improved as a result of having a spotlight shined on them. Suggesting that less certainty would be okay if not for the compensation spotlight would basically be suggesting that less certainty is permissible if there is no spotlight.

Increased Compensation, Increased EMV

Performance-based compensation has been used on a limited basis for decades. What have changed are the greatly increased scale of efficiency expenditures and the increased share of benefits that can accrue to administrators.[57] To encourage utilities to pursue deeper energy savings, regulators have begun providing strong inducements to utilities. These include shared net benefits based on ex ante savings assumptions, cost capitalization and decoupling mechanisms, as well as performance targets. In most cases, the utilities face little risk of not recovering their costs. They gain greater up-side potential with little down-side risk in order to motivate them to place energy efficiency on a level playing field with supply side resources and to aggressively pursue ratepayer efficiency improvements.

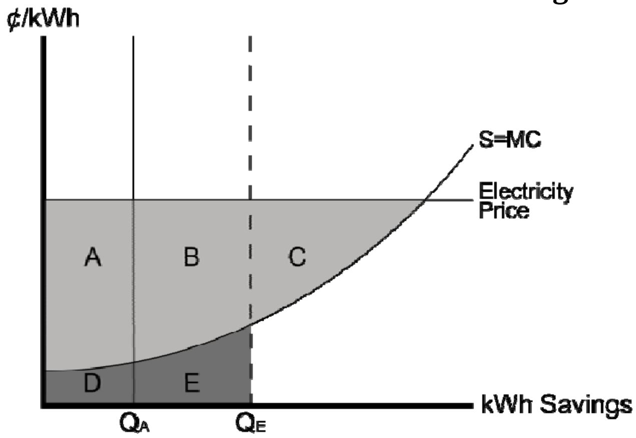
Under traditional cost-plus compensation, virtually all of the economic benefits of the EE investment accrue to ratepayers.[58] In some cases, utilities are not even compensated for lost profits due to the program-induced reductions in demand for electricity. It is hard to imagine a better deal for ratepayers as a class, as long as the claimed savings are realized and the marginal benefit exceeds the marginal cost billed by the utility to the ratepayer. Any change in the compensation mechanism is going to transfer some of the economic benefit from ratepayers to the utility.

When all of the net economic benefit accrues to ratepayers, the acceptable margin of error between estimated and actual savings is greater than if a share of the net benefit is paid to the administrator. Figure 3 illustrates the effect of overestimating savings on ratepayers' collective welfare.[59] If actual savings equals estimated savings (Q_E), the ratepayer benefit equals A+B. The area in C represents unrealized economic benefit due to not achieving all cost effective potential. If the actual savings (Q_A) are less than estimated savings (Q_E), then the ratepayer

[57] *Ibid*, p. 1031.
[58] One could argue that under cost-compensation ratepayers have had it too good relative to typical markets in which the net benefits are shared between consumers and producers.
[59] The arguments here are not affected by whether the marginal benefit curve represents benefits as determined by a ratepayer test, total resource test or societal test. For simplicity sake, we refer to the beneficiaries as "ratepayers."

benefit equals A. In either case, the ratepayers' payment to the administrator is represented by D+E. As long as the area in A equals or exceeds the area in D+E, the ratepayers are still better off than if there were no programs.

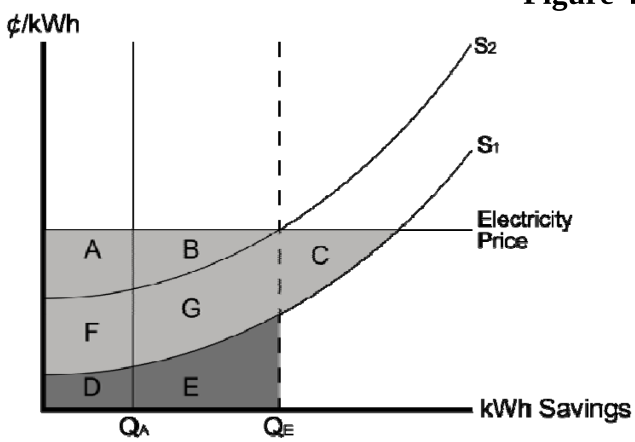
Figure 3



If actual savings = estimated savings = Q_E, then the net benefit to ratepayers = A+B, their costs equal D+E and C=unrealized net benefit opportunity. If actual savings equal Q_A, then ratepayers' net benefit is only A even though they are still paying D+E.

As shown in figure 4, an increase in administrator compensation will shift the supply curve up (S_2) and increase the compensation level by F+G.[60] Recall that the administrator's compensation (now represented by D+E+F+G) will be paid to the administrator regardless of the actual savings (Q_A). Further, the area in A is smaller due to the increased compensation represented in F. In other words the increased administrator compensation crowds out the ratepayer net benefit. For ratepayers to be indifferent between the previous and new compensation levels, the difference between the actual (Q_A) and estimated savings (Q_E) will need to be smaller either through increased actual savings or reduced estimated savings; graphically, Q_A needs to shift to the right so that the area in A increases by the area in F+G or Q_E needs to shift to the left so that the area in B decreases by the area in F+G.

Figure 4



For ratepayers to be indifferent between the previous and new compensation levels (represented by shift of S_1 to S_2, the difference between the actual (Q_A) and estimated savings (Q_E) will need to be smaller either through increased actual savings or reduced estimated savings; graphically, Q_A needs to shift to the right so that the area in A increases by the area in F+G or Q_E needs to shift to the left so that the area in B decreases by the area in F+G.

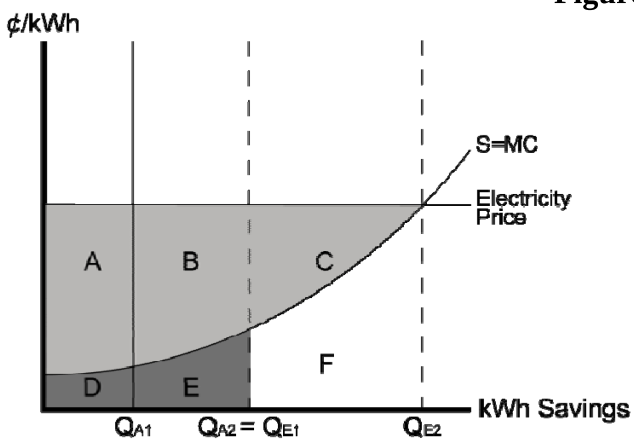
[60] Note that the parallel shift here would coincide with a cost-plus compensation increase. The new cost curve could look different for other types of compensation schemes. A simple performance-based compensation in which the utility program administrator gets a fixed price per kWh saved would be represented by a flat MC curve, while the cost curve for the California shareholder incentive would be tiered. The shape of the new cost curve is not important for this discussion. The important point is that it is above the original cost curve.

In other words, ratepayers are justified in demanding greater EMV certainty along with any increase in compensation given to administrators.

Deeper Savings, Deeper EMV

Deeper savings also justifies greater EMV certainty. The further up the supply curve one goes, the more certainty that should be required in the savings estimates since there will be a lower margin of error between positive and negative ratepayer net benefits. Ratepayers are justified in demanding greater EMV certainty for large-scale programs than small-scale programs.

Figure 5



A large program has a smaller margin of error than a small program – in other words, for a large program, a smaller percentage difference between actual and estimated savings could result in a negative net benefit to ratepayers. Graphically, (A+B) / (D+E+F) is less than A / (D+E).

In figure 5, estimated and actual savings from the large-scale program are represented by Q_{E2} and Q_{A2}, respectively. Estimated and actual savings from the small-scale program are represented by Q_{E1} and Q_{A1}, respectively.[61] Net benefits to ratepayers under the large program are represented by A+B and the compensation is represented by D+E+F. Under the small program, the ratepayer net benefit equals A and the payment compensation equals D+E. Given a downward sloping MB curve and upward sloping MC curve, the large program has a smaller margin of error than the small program – in other words, for the large program, a smaller percentage difference between actual and estimated savings could result in a negative net benefit to ratepayers. Graphically, (A+B) / (D+E+F) is less than A / (D+E).

This suggests that pursuit of deeper savings requires greater EMV certainty to leave ratepayers indifferent. It is reasonable for ratepayers to expect greater EMV certainty as deeper savings are pursued. This is the case no matter how administrators are compensated.[62]

Rigorous EMV is the Answer, Not the Problem

In sum, whether administrators are compensated based on cost or performance, the deeper the savings and the higher the compensation, the less margin for error and the more certainty is

[61] QA2 does not need to equal QE1. We have just portrayed it this way to limit clutter in the graph.
[62] To be clear, we are talking about uncertainty associated with savings estimates (whether the program *did* succeed or fail, not the uncertainty associated with whether a program *will* succeed or fail).

required in the savings estimates. It is not only the performance-based compensation mechanism that increases the need for improved EMV; it is the higher compensation and increased scale of the programs.

This could have several important implications. If the performance-based compensation mechanism is *not* itself driving the need for more EMV certainty, then:

> 1) Abandonment of EMV for performance-based compensation reflects poorly on EMV used for portfolio analysis, especially if abandonment of performance-based incentives is seen as a retreat from improving portfolio analysis EMV;
>
> 2) Abandonment of a performance-based incentive will not eliminate the need for greater EMV certainty if the real reason for more EMV certainty is increased program scale or increased utility compensation; and
>
> 3) Abandonment of a performance-based incentive could cause abandonment of some benefits such as increased transparency and credibility of EMV.

Performance compensation has created a more transparent and adversarial EMV process. But greater scale and greater compensation have created the *need* for greater EMV certainty. Backing away from performance based compensation will not eliminate the need for more rigorous EMV.

Ensuring cost effective efficiency improvements will require tradeoffs between: 1) scaling back programs, 2) reduced utility compensation, and 3) increased EMV certainty. At least for now, the first option is not on the table as governments concerned about reducing carbon emissions push the envelope on programmatic energy savings. The emphasis will need to be placed on improving EMV and reducing the cost of program delivery. Performance-based compensation mechanisms have the potential to both increase the transparency of EMV and foster competition for delivery of energy savings.

Symmetrical Knowledge and Power is Key

An adversarial process requires vested interests that are willing and able to identify problems they see in the estimation methods. The accuracy of energy savings evaluations used for ESPCs hinges on an adversarial process between ESCOs and their clients. The accuracy of EE program savings estimates depends on involvement of all the relevant stakeholders including electric utilities, regulators, ratepayers and environmental advocates. The credibility of carbon offsets depends on a strong and independent CDM-type board to push back on project developers seeking easy credits.

To be successful, an adversarial process also requires symmetric access to knowledge and data among the adversaries. Most successful energy savings evaluations are conducted in an adversarial process – even evaluations of single projects. Most ESPCs are entered into with large clients, such as federal agencies, schools, or hospitals, that have the resources necessary (with help from the IPMVP and other public resources) to identify potential systematic errors in the evaluation of savings and negotiate terms and conditions that either reduce the risk of those errors or discount for them.

Care must be taken to ensure that the adversarial process does not itself introduce systematic error. For example, if the only stakeholders in the "adversarial process" are stakeholders that do not want the programs to succeed, then savings estimates will be underestimated. In some cases, the stakeholders will appear organically – utilities, utility regulators, large ratepayers and environmental organizations in a state may effectively challenge one another's claims of savings from EE programs. At least the hope is that with countervailing interests, such an adversarial process can result in negotiated savings estimates that, while maybe not equal to actual savings, provide a reasonable basis for compensating the utility and balances other stakeholder interests.

In other cases, however, countervailing interests may need to be introduced from outside. For example, there may be relatively little interest of in-state stakeholders in ensuring compliance with a national EERS. Federal oversight by DOE or EPA may be required to ensure that estimates of energy savings are not exaggerated.

No matter the cause of the evaluation uncertainty, some degree of adversarial process is probably necessary to overcome it. This is an important point, because it suggests that evaluation protocols for addressing uncertainty may be more productively focused on the evaluation decision making process than the measurement methods or assumptions.

Creating a Community of Professionals

Evaluation does and will continue to rely heavily on the judgment of professional evaluators. As with other professions – engineers, lawyer, doctors, teachers, etc. – the evaluation community benefits from a professional association of peers that can accredit evaluators and monitor the quality of their work. EVO, in partnership with the Association of Energy Engineers, offers the Certified Measurement and Verification Professional (CMVP) program to raise the professional standards and practices measurement and verification practitioners and to recognize professionals who have demonstrated "a high level of competence and ethical fitness."[63]

The focus of the CMVP training and examination is on fundamental understanding of the IPMVP. Thus, it does not include any detailed discussion of engineering, statistics, metering, or survey design, nor does it cover the host of issues related to program evaluation or net impacts.[64] Nevertheless, the CMVP can serve as the foundation for certification covering these other important evaluation topics.

Triangulation of methods – or evaluations by different evaluators – could go a long way to increase awareness of possible sources and directions of bias associated with different methods

[63] See Steven Meyers and Steve Kromer, "Measurement and verification strategies for energy savings certificates: meeting the challenges of an uncertain world," *Energy Efficiency*, 2008, p. 313-321.
[64] See Cowan & Sain, 2008.

and data.[65] This is not unlike other professions – for example, medical professions where patients often obtain a second opinion. Accreditation of evaluation professionals does not provide a guarantee that their opinions will be correct, but it will ensure that the presence of a pool of experts from which administrators can draw evaluators.

[65] At the same time, triangulation increases awareness of the vagaries of savings estimation, which can lead to greater awareness of the uncertainty among stakeholders and thus reinforce perceptions of uncertainty – if not perceptions of bias, at least imprecision.

Toward a National EMV Protocol?

There are significant economies of scale associated with evaluating projects and programs, yet no single EMV formula can fully capture the multitude of program dimensions. Some projects lend themselves to standardization more than others – e.g., hotel lighting projects, constant load motor applications and packaged unit air conditioning upgrades in retail chain facilities. If enough of these projects are of the same type are being implemented at the same time, it may be feasible to develop common EMV procedures. For other types of projects, such as industrial process chiller retrofits, it is difficult to develop standardized EMV.[66]

Likewise, programs vary in their uniformity and ease of evaluation. It is easier, though not trivial, to evaluate a CFL direct install program than, for example, a small residential loan program in which many types of EE measures may be implemented. Unit evaluation costs can be lowered if spread over larger programs or savings – e.g., bundling projects, using deemed savings, or sampling rather than evaluating every project. Efforts to increase the number and diversity of parties delivering the programs and/or projects can require evaluation protocols to be developed for a wider variety of program and project types and thus could raise evaluation costs further.

Current Protocol Development

Numerous protocols for estimating savings from EE programs and projects have been developed over the last two decades. In simplest form, protocols can consist of a couple of pages of reporting requirements for government funded EE programs relying heavily on "intermediate indicators" such as number of workshop participants. The most complex protocols can easily be several hundred pages and reference hundreds more in additional documents. The California Energy Efficiency Evaluation Protocols provide guidance for project M&V and at least five types of program evaluation – program impact, process, market effects, emerging technologies and codes and standards – at two various levels of rigor. The California Protocols include guidance on a long list of topics, including:[67]

- acceptable methods;
- acceptable data sources including stipulated savings assumptions;
- key metrics;
- reporting contents;
- verification of measure installation;
- verification of commissioning;
- estimating data measurement error;
- survey design;
- engineering model design;
- target audiences;
- contacting customers;

[66] Mathew, et al., 2005, pp. 1319-1328, gives an excellent run down of Enron's efforts to "commoditize" EE projects.
[67] California Public Utilities Commission, 2006.

- confidentiality issues;
- required evaluator skills;
- responsibilities of various actors;
- sampling precision;
- measure persistence; and
- mid-stream changes in evaluation methods.

Many states have adopted their own less-ambitious protocols –TRMs or "cookbooks" – usually to be used to determine EERS compliance. Most state protocols provide deemed savings assumptions for common measures along with more general guidance for evaluating custom projects and programs. Some reference the International Performance Measurement and Verification Protocol (IPMVP) for project level savings estimates, but some do not.[68]

NAPEE developed a far less detailed but instructive guide to program impact evaluation.[69] ISO-New England and the PJM have both developed standards for estimating savings from EE projects and programs that are bid into their wholesale electric capacity markets.[70]

Additional efforts to develop regional and/or national protocols for estimating savings from EE programs are being pursued by a variety of organizations, including the NEEP, NAESB, NAPEE, EVO and the Center for Resource Solutions (CRS).[71] These efforts are generally in early stages of development.

NEEP, in a 2006 report outlining the variations in EMV methods, data, assumptions and reporting, was one of the first to flag the need for coordinating the various state protocols: "Fundamental to successfully treating EE as a resource is to create a common currency for savings that is credible and transparent."[72] NEEP subsequently established the Evaluation, Measurement and Verification Forum to pursue development of protocols by and for states in the Northeast and Mid-Atlantic, all of which, except for Pennsylvania, have joined this effort.

The NEEP EMV Forum has developed and is in the process of developing a number of products to further regional EMV cooperation, including:

- EMV glossary (March 2009)
- Comparison of different methods and assumptions (expected late 2009)
- Recommended common assumptions and reporting (expected early 2010)
- Load shape data (expected in 2010)

[68] For example, we can find no references to IPMVP in TRMs of Connecticut, Pennsylvania and Vermont.
[69] NAPEE, 2007, p. ES-4.
[70] Cheryl Jenkins, Blair Hamilton and Chris Neme, Vermont Energy Investment Corporation, "Playing with the Big Boys: Energy Efficiency as a Resource in the ISO New England Forward Capacity Market," 2008 ACEEE Summer Study on Energy Efficiency in Buildings, 2008, pp. 5-157 - 5-170.
[71] NEEP, EM&V Forum, accessed October 2009, http://neep.org/emv-forum. NAESB, "Demand Side Management and Energy Efficiency," accessed October 2009, http://www.naesb.org/dsm-ee.asp. NAPEE, 2007. Dr. Jan Hamrin, Dr. Edward Vine and Amber Sharick, *The Potential for Energy Savings Certificates (ESC) as a Major Tool in Greenhouse Gas Reduction Programs*, Center for Resource Solutions, 2007.
[72] Michals & Titus, 2006, p. 6.

- Estimates of commercial and institutional lighting measure persistence and life (expected early 2010)
- Net savings issues (expected early 2010)
- Impacts of codes and standards on programs (expected early 2010).

NAESB develops and promotes standards for the electric and gas industries through an American National Standards Institute-certified, consensus-based process. It is looking to build on recently completed standards for M&V of demand response programs and has created EE working groups to develop standards for measuring savings from EE programs and projects. The exact scope of this initiative – e.g., whether it will include program evaluation or just project M&V – is still being discussed, but NAESB's ability to develop standards for demand response programs bodes well for development of standards for EE EMV.

NAPEE, a joint effort of the US DOE and US EPA, developed a "National Energy Efficiency Program Impact Evaluation Guide" in 2007 (which we have cited in this report numerous times). The Guide describes EMV issues and program evaluation structures and model approaches. It considers estimation of energy and demand savings as well as emissions reductions. NAPEE created an EMV Technical Working Group in early 2009 to consider the need for new EMV methods, inventory the range of EMV programs and protocols and set priorities for NAPEE in the EMV arena, including investigating the feasibility and need for national EMV protocols.

The Feasibility of a National Protocol

Many evaluation experts express grave reservations about the creation of national evaluation protocols beyond the general content found in, for example, the NAPEE model EMV Guidance. It is easy to be skeptical about the feasibility of creating a national evaluation protocol and its longer term viability once it was created. Major challenges include:

- *Program heterogeneity* – A wide variety of program approaches (direct install, rebates, technical assistance, awareness, etc) are used by a wide range of people and organizations with various management capabilities (work effort, enthusiasm, charisma) targeting different measures (e.g., electric motors, appliances, lights) at different stages of development (e.g., emerging, mature) for different types of participants (low income renters, middle income home owners, small retailers, large manufacturers, early adopters, etc) at different times in different climates in different jurisdictions with different policies (codes, standards) for different purposes (e.g., compensation, compliance, cost-effectiveness).[73] A national protocol will need to contain many different protocols to be useful.
- *Contentious and difficult issues* – The more vexing issues around EMV– e.g., how to define and test for additionality, how much emphasis to place on measurable savings, etc. – are not new. Many smart people have argued these issues for decades. What has changed that will allow their resolution now?

[73] For discussion of these various program dimensions, see Nick Hall and Patrick McCarthy, "Portfolio Evaluation Versus Program Evaluation: Is There a Balance?," *International Energy Program Evaluation Conference*, 2009.

- *The power of the status quo* – Protocols attempt to undo existing practices, products and services. They disenfranchise some business models. Participants will defend existing prerogatives.
- *Many stakeholders* – Different stakeholders have different perspectives, interests, knowledge and levels of influence. To be widely accepted and credible, protocols will have to appease these various stakeholders equally.
- *EMV costs money* – The willingness of regulators and other stakeholders to devote resources to EMV varies widely. There is generally a direct correspondence between rigor and funding.
- *Updating* – Protocols must be continually updated to reflect changing BAU conditions, new data, improved methods, changes in program design and changing technologies.

Possible Benefits of a National Protocol

On the other hand, a number of benefits could be realized if national energy savings protocols can be developed:

- *More policy and program options* – The viability of some energy efficiency policies and programs depends on a national EMV. For example, if there is no home-grown desire to increase EE program savings, weak EMV could be some state's preferred path to compliance with a federal EERS. Even if there is no national EERS, states can use loose EMV standards to appear to be taking aggressive action when they are not.
- *Increased credibility* – At least over time, the credibility of EE as a major energy resource, on par with supply side resources, could be enhanced by a public process to develop national protocols.
- *Diversified delivery of EE savings* – It is harder and more costly for national ESCOs and other entities to participate in EE programs if forced to learn different protocols in every state.[74]
- *Lower transactions costs* – Substantial planning, administration, and other transactions costs can be avoided if individual states do not have to create their own cookbooks, conduct useful life and persistence studies, debate free riders, etc.

Creation of a national EMV protocol is clearly a challenging task. Moreover, a standardized national EMV methodology could also introduce a strong bias toward a standardized national program designs; some would argue that we need more innovation and experimentation, not less.[75]

But if it is not feasible and or not desirable to develop consensus protocols, we need to know it. We would then need to lower expectations about our ability to place EE on a level playing field with supply side energy resources and to revisit some of the energy efficiency policies being considered, including a national EERS, performance-based compensation and programmatic CDM for energy efficiency.

[74] Don Gilligan at Alliance to Save Energy Workshop on ESCs and RECs, February 14, 2008.
[75] Personal correspondence, Jeff Harris, November 19, 2009.

Conclusion and Recommendations

The public and policymakers are increasingly looking to EE programs as a frontline response to climate change and other concerns. Increased attention and funding for EE programs is bringing greater attention to how EE program savings are measured. If EE programs are going to play a major role in improving the nation's energy efficiency, the savings measurement challenges outlined in the article will need to be addressed.

Most of the program evaluation issues are not new. They have been discussed extensively for decades and are familiar to a large cadre of evaluation experts. There is little consensus on many of the EMV issues and the most productive path forward is not clear. But it is quite clear that the measurement issues cannot be cast aside if EE programs are to be a major component of a national and global energy and climate strategy. If it is not feasible to develop consensus on savings measurement, we need to know it. We might then need to revisit some of the energy efficiency policies being considered and perhaps even the scale of programs.

Credible savings measurement is a prerequisite for participation of EE projects and programs in carbon offset (credit) schemes such as the CDM, performance-based incentives to encourage utility investments in customer efficiency, a meaningful national EERS, and generally assessing the cost-effectiveness of EE programs. At the same time, by creating vested interests, performance-based mechanisms – e.g., carbon offset programs, the California shareholder incentive, Duke Energy's Save-a-Watt, Connecticut's energy saving credit program – can shine a spotlight on programmatic EMV and help ensure that we get the greatest energy savings for our money.

It remains to be seen whether improved EE EMV will mean more programs, fewer programs, or different programs, but credible savings measurement will be a vital part of a sustained scale-up of energy efficiency programs. While increasing evaluation precision is an important objective, the credibility and reliability of evaluation results.ultimately depends on our ability to reduce systematic bias in the estimates.

To enhance the credibility and reliability of EE program savings measurement, federal and state governments, along with various stakeholders, should: 1) Institute processes for EMV design and review that incite transparent and thorough debate over EMV methods, data and assumptions; 2) Improve EMV methods, data and assumptions; 3) increase consistency of methods and assumptions between regions and program types, 4) assure evaluation professional competency and integrity and 5) manage stakeholder expectations. For each of these broad objectives, we recommend the following actions be taken by federal and state governments, foundations and stakeholders.

1) Institute processes for EMV design and review that incite transparent and thorough debate over EMV methods, data and assumptions.

- Encourage participation of a wide variety of stakeholders in EMV protocol development and review of EMV results. Ensure diverse stakeholders have sufficient information, knowledge, power and interest to effectively engage in the evaluation design and review.

- Recognize that risk-reward mechanisms based on ex post savings may be necessary to create diverse and competing interests that will promote sufficiently rigorous savings estimates, that increased contention and EMV cost may be necessary to achieve reliable savings estimates.
- Encourage or require program evaluators to use multiple evaluation methodologies (triangulation) to reduce systematic error, including top-down and bottom-up approaches, econometric modeling and/or surveys.

2) Improve EMV methods and assumptions.

- Compile deemed savings assumptions and methodologies found in various TRMs with a view toward overcoming inconsistencies and developing regional or national TRMs.
- Support central compilation of data and information, including lessons learned from California and other major evaluation initiatives around the country. The Consortium for Energy Efficiency (CEE) and California Measurement Advisory Council (CALMAC) websites provide ready-made platforms for compiling this information.
- Build on the many existing data sets. For example, California's Database for Energy Efficiency Resources has been extensively detailed and updated over the last year and these efforts will continue. The Northeast Energy Efficiency Partnerships is compiling data on, among other things, load shapes and measure persistence for the Northeast.
- Encourage or require utilities and evaluators to share data, as well as the costs of obtaining that data, to achieve economies of scale in data collection and analysis.
- Support conferences and other types of venues to facilitate exchange of information and discourse among evaluation professionals and other stakeholders.
- Be prepared to continually update data and methods to take advantage of changing circumstances and latest best practices.

3) Increase consistency of methods and assumptions between regions and program types.

- Develop national and/or regional protocols to harmonize evaluation definitions, methods, assumptions and reporting.
- Develop different protocols for different evaluation purposes – e.g., utility planning, national EERS, bidding EE into capacity markets, carbon offsets, etc. – and for different types of programs.
- Build on existing EMV protocols and initiatives, including, for example, the California Energy Efficiency Evaluation Protocols (2006) and the California Evaluation Framework (2004). Products and decisions made in the NEEP EMV Forum and NAPEE work group should feed into the NAESB Energy Efficiency Working Groups.

4) Assure the competency and integrity of evaluation professionals.

- Enhance the capacity of national professional association(s), such as the Energy Valuation Organization (EVO), to monitor the professional integrity and competence of evaluation professionals and provide professional accreditation.
- Rely heavily on a national professional association(s) to help craft a national protocol(s) and code of conduct for evaluation professionals.

- Support programs to provide training of new EMV professionals and update and enhance the skills of existing evaluation professionals. Provide training that clarifies different roles of evaluators to foster adversarial evaluation oversight processes – analogous to attorneys being trained and bound to look out for their clients' interests, for example.

5) Manage stakeholder expectations.

- Provide clear and reasonable expectations about EMV certainty regarding various types of programs and evaluation purposes.
- Avoid arbitrary limits on EMV spending. Some types of EE programs require more evaluation than others to achieve stakeholder confidence and credibility. EMV should be viewed as an integral part of EE program budgets, as much as rebates or any other program cost, rather than a tax.
- Do not overemphasize the conflict between process and impact evaluation. If the impact estimates are wrong, the feedback to programs will be wrong too. Impact evaluation is unavoidable.

Bibliography

Alliance to Save Energy, "Stimulus Resources: The American Recovery and Reinvestment Act of 2009 (ARRA)," 2009, http://ase.org/content/article/detail/5461.

American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE), *Guideline 14-2002 - Measurement of Energy and Demand Savings*, 2002.

Association of Energy Engineers, Certified Energy Manager Training. Prepared by Efficiency Evaluation Organization, October 2008.

Stephen Bickel, "Maximizing Energy Savings with CFLs: Don't Bench Your Superstar," American Council for an Energy Efficient Economy, Market Transformation Conference, 2009, http://www.aceee.org/conf/09ee/pdfs/(5B_Bickel)EER2009.pdf.

Carl Blumstein, "Program Evaluation and Incentives for Administrators of Energy-Efficiency Programs: Can Evaluation Solve the Principal/Agent Problem?" *European Council for an Energy-Efficient Economy Summer Study*, Center for the Study of Energy Markets, March 2009.

California Public Utilities Commission, *California Energy Efficiency Evaluation Protocols: Technical, Methodological, and Reporting Requirements for Evaluation Professionals*, prepared by The TecMarket Works Team, April 2006.

California Public Utilities Commission, "Proposed Energy Efficiency Risk-Reward Incentive Mechanism," Energy Division White Paper, April, 2009.

CALMAC, accessed October 2009, http://www.calmac.org.

Peter Cappers, Charles Goldman, Michele Chait, George Edgar, Jeff Schlegel and Wayne Shirley, *Financial Analysis of Incentive Mechanisms to Promote Energy Efficiency: Case Study of a Prototypical Southwest Utility*, Lawrence Berkeley National Laboratory, Environmental Energy Technologies Division, March 2009.

The Climate Registry, *The General Reporting Protocol for the Voluntary Reporting Program, Version 1*, May 2008, pp. 97-108.

Consortium for Energy Efficiency, accessed October 2009, http://www.cee1.org.

Connecticut Department of Utility Control Docket No. D05-07-19RE02, 2009.

Consortium for Energy Efficiency, "CEE 2008 Annual Industry Report," 2008, http://www.cee1.org/ee-pe/2008/us_combo.php.

John Cowan and Steve Sain, *Measurement and Verification Fundamentals: the International Performance Measurement and Verification Protocol for Energy Managers and Emission Traders*, Study Manual for Association of Engineers Certified Energy Manager Training prepared by Efficiency Evaluation Organization, October 2008

Department of Energy, *Big Results Bigger Potential: CFL Market Profile*, March 2009, p.1, http://www.energystar.gov/ia/products/downloads/CFL_Market_Profile.pdf.

Efficiency Valuation Organization, *International Performance Measurement and Verification Protocol: Concepts and Options for Determining Energy and Water Savings Volume 1*, EVO 10000-1.2007, April 2007.

Environmental Protection Agency, "eGRID FAQ," http://www.epa.gov/RDEE/energy-resources/egrid/faq.html.

Federal Energy Management Program (US FEMP), *M&V Guidelines, Measurement and Verification for Federal Energy Projects*, Version 2.2, DOE/GO-102000-0960, 2000.

Christiana Figueres and Michael Philips, *Scaling Up Demand-Side Energy Efficiency Improvements Through Programmatic CDM*, Energy Sector Management Assistance Program and Carbon Finance Unit, World Bank, 2007.

Kenneth Gillingham, Richard Newell and Karen Palmer, *Retrospective Examination of Demand-Side Energy Efficiency Policies*, Resources for the Future, 2004.

Cherie Gregoire, "Impact Assessment Program Manager, Energy Analysis," NYSERDA, March 26, 2009.

Nathanial Gronewold, "Climate: Prices take a sharp dip in fifth RGGI auction," *Greenwire*, September 11, 2009, http://www.eenews.net/public/Greenwire/2009/09/11/5.

Nick Hall and Patrick McCarthy, "Portfolio Evaluation versus Program Evaluation," *International Energy Program Evaluation Conference*, 2009.

Dr. Jan Hamrin, Dr. Edward Vine and Amber Sharick, *The Potential for Energy Savings Certificates (ESC) as a Major Tool in Greenhouse Gas Reduction Programs*, Center for Resource Solutions, 2007.

Debra Jacobson and Colin High, "Fact Sheet: Assuring Accurate Estimates of the Avoided Emissions Benefits of Energy Efficiency and Renewable Energy Technologies," [2009], http://www.mwcog.org/uploads/committee-documents/a15cWlxY20090402113537.pdf.

Cheryl Jenkins, Blair Hamilton and Chris Neme, Vermont Energy Investment Corporation, "Playing with the Big Boys: Energy Efficiency as a Resource in the ISO New England Forward Capacity Market," 2008 ACEEE Summer Study on Energy Efficiency in Buildings, 2008.

Doug Mahone and Wim Bos, "A New Energy Efficiency Portfolio - Developing an Expanded Evaluation Approach," International Energy Performance and Evaluation Conference, August 2009.

Paul Mathew, Eric Koehling and Satish Kumar, "Use of Quantitative Uncertainty Analysis to Support M&V Decisions in ESPCs," *AEE Journal Energy Engineering*, v. 103, n. 2, 2006.

Paul Mathew, J. Stephen Kromer, Osman Sezgen and Steven Meyers, "Actuarial Pricing of Energy Efficiency Projects: Lessons Foul and Fair," *Energy Policy*, 33, 2005.

Steven Meyers and Steve Kromer, "Measurement and verification strategies for energy savings certificates: meeting the challenges of an uncertain world," *Energy Efficiency*, 2008.

Julie Michals and Elizabeth Titus, *The Need for and Approaches to Developing Common Protocols to Measure, Verify and Report Energy Efficiency Savings in the Northeast*, Northeast Energy Efficiency Partnerships, 2006.

Evan Mills, Steve Kromer, Gary Weiss and Paul Mathew, "From Volatility to Value: Analyzing and Managing Financial and Performance Risk in Energy Saving Projects," *Energy Policy* 34, 2006.

National Action Plan for Energy Efficiency (NAPEE), *Model Energy Efficiency Program Impact Evaluation Guide*, prepared by Steven R. Schiller, Schiller Consulting, November 2007, http://www.epa.gov/RDEE/documents/evaluation_guide.pdf.

North American Energy Standards Board (NAESB), "Demand Side Management and Energy Efficiency," accessed October 2009, http://www.naesb.org/dsm-ee.asp.

The National Research Council, *Effectiveness and Impact of Corporate Average Fuel Economy Standards*, 2002.

Northeast Energy Efficiency Partnerships EM&V Forum, accessed October 2009, http://neep.org/emv-forum.

Office of Consumer Counsel [Connecticut], "Brief of the Office of Consumer Counsel, May 18, 2009, Docket 09-02-18."

Pennsylvania Public Utility Commission, Implementation order, Docket M-2008-2069887, January 15, 2009, http://www.puc.state.pa.us/electric/pdf/Act129/EEC_Implementation_Order.pdf.

Michael Rufo, "Evaluation and Performance Incentives: Seeking Paths to (Relatively) Peaceful Coexistence," International Energy Program Evaluation Conference, 2009.

Katherine Shaver, "Older Cars Fouling Region's Air Quality: Agency Study Links Recession, Pollution," *The Washington Post*, July 4, 2009, http://www.washingtonpost.com/wp-dyn/content/article/2009/07/03/AR2009070302451.html?sid=ST2009070400257.

John R. Taylor, *An Introduction to Error Analysis, the Study of Uncertainties in Physical Measurements, 2nd Edition*, University Science Books, 1997.

Mark Trexler, Derik Broekhoff and Laura Kosloff, "A Statistically-Driven Approach to Offset-Based GHG Additionality Determinations," *Sustainable Development Law and Policy*, Volume VI, Issue 2, Winter 2006.

United Illuminating Company in United Reporters, Transcript of the Hearing of Docket 09-02-18, 4-28-09.

United Illuminating Company Response to EL-15 in D05-07-19RE02, July 7, 2009.

Ed Vine, Gregory Kats, Jayant Sathaye and Hemant Joshi, "International Greenhouse Gas Trading Programs: A Discussion of Measurement and Accounting Issues," Energy Policy 31, 2003.

Appendix A

Converting Energy Savings to Emissions Reductions

Enhancing the energy efficiency of electricity use can deliver benefits in the form of reduced stress to electricity grids and the deferral or avoidance of new investment in power generation, transmission and distribution. However, much of the impetus for improving energy efficiency is to avoid or reduce environmental impacts of electricity production and use, particularly greenhouse gas (GHG) emissions.

The environmental impacts of electrical power systems are numerous and include such things as impacts on land, water and biota of coal mining and natural gas extraction; disposition of nuclear wastes; management of ash and scrubber sludge; adverse effects on local wildlife from hydro-electric dams; water consumption by power plants; and emissions of the powerful GHG sulfur hexafluoride (SF_6) from electrical switchgear, among others. However, the main focus of various energy efficiency and renewable energy (EERE) policies is to reduce or avoid carbon dioxide (CO_2) emissions. In addition, since hundreds of counties in the U.S. have failed to attain the national air quality standards for ozone and other conventional pollutants, states have increasingly sought to utilize EERE measures to reduce emissions of ozone precursors, such as nitrogen oxides (NO_x).

The uncertainties and complications of EMV of energy savings are amplified when translating such savings to emission reductions. Even if one has an agreed upon figure for MWh saved by a particular program or set of measures, the emission reductions vary among different regions of the country because of variations in the mix of generating units (different percentages of coal, oil, natural gas, nuclear and renewable energy units). In addition, the emission reductions vary according to the season and time of day. Electric utilities rely on numerous generation units (whether owned by the utility or by third-party generators) and attempt to dispatch those units in the most cost-effective manner feasible (and also in consideration of technical matters such as maintenance schedules and transmission constraints) as customer power demand changes over time, both hourly and seasonally.

The emissions of these units can vary greatly. Nuclear, wind and hydroelectric units have essentially zero CO_2 emissions while coal-fired units are major CO_2 emitters. Natural gas units, often used to meet intermediate and peak loads, emit on average roughly half the CO_2 as coal-fired units per kWh produced, though their efficiency may vary by configuration (for instance, simple-cycle as compared to combined-cycle natural gas units). Oil-fired steam units, diesel, waste-to-energy and various biomass-fired generators also enter the mix. Calculation of NO_x reductions can be even more complicated than for CO_2 because NO_x emission rates depend on complex design and operational characteristics of generating units (e.g., burner configurations, air flow and temperature, fuel parameters, pollution abatement equipment).

Moreover, the amount of emission reductions estimated will vary substantially depending upon the choice of methodology used for such calculations. The EPA-developed Emissions and Generation Resource Integrated Database (eGRID) is frequently used for estimating emission reductions resulting from EE and RE measures because it is the only widely available public database. The eGRID database contains data on emission factors for 26 power markets in the United States calculated in two different ways: (1) the system average methodology, which is based on the average generation mix (including all fossil fuel-fired units, nuclear units, hydroelectric units, and other units); and (2) the non-baseload methodology, which is based on the average emission rates of only those generating units that have variable generation that follows the regional load. This includes generation at units that are usually classified as intermediate and peak load plants.[76] The current General Reporting Protocol for the Climate Registry, the GHG registry used by 40 states, relies on the eGRID system average methodology.[77]

Does it matter whether emissions reductions are estimated using system average emissions rates as reported in eGRID rather than using the eGRID nonbaseload methodology or even more precise methodologies based on marginal emissions at each hour of the year? Jacobson and High assessed emission reductions for both CO_2 and NO_x for five EERE technologies (high-efficiency commercial air conditioning, high-efficiency commercial lighting, LED traffic light retrofits, wind energy and photovoltaic electricity) hypothetically applied in the PJM Interconnection and Upstate New York power markets. According to their analysis, in the two power markets reviewed eGRID system average data results in 70 to 120 percent understatement of emissions savings relative to using a more precise time-matched marginal (TMM) methodology that is based on marginal emissions at each hour of the year.[78] The complete results of their analysis as well as comprehensive recommendations will be contained in an article that has been accepted for publication by the Journal of Energy and Environmental Law of The George Washington University Law School in the spring of 2010.

NAPEE notes that there are "medium effort calculation approaches" that lie between using broad grid averages and more precise power dispatch modeling.[79] The TMM methodology helps reduce the high expense of power dispatch modeling through the following steps: [80]

1. Estimating the hourly electric power generation for every fossil fuel-fired unit in a specific power market area based on: (a) the EPA's database of hourly information on CO_2 emissions from Continuous Emission Monitors (CEMs); and (b) the average CO_2 emission rate (lbs/MWh) for each unit. The CO_2 emission rate is used to estimate the generation at each hour.
2. Identifying the marginal fossil fuel-fired units at each hour by using the hourly generation to identify units that follow the total load at each hour. Based on this information, the TMM method estimates average emission rates of the marginal units based on their incremental contribution to the load.
3. Using data provided in publicly available databases or industry data to compile a profile of the energy savings or energy generation on an hourly basis over the 8760 hours of the

[76] Debra Jacobson and Colin High, "Fact Sheet: Assuring Accurate Estimates of the Avoided Emissions Benefits of Energy Efficiency and Renewable Energy Technologies," [2009], http://www.mwcog.org/uploads/committee-documents/a15cWlxY20090402113537.pdf.
[77] The Climate Registry, *The General Reporting Protocol for the Voluntary Reporting Program, Version 1,* May 2008, pp. 97-108.
[78] Jacobson & High, [2009]
[79] NAPEE, 2007, pp. 6-5 to 6-6. See Section 6 "Calculating Avoided Air Emissions" for more on this topic.
[80] Debra Jacobson, personal communication, November 22, 2009.

years. This profile is prepared for a particular technology (e.g., wind power, high-efficiency commercial air conditioning) and for a particular region.

4. The "time-matching" occurs when the load profile is matched for a specific technology on an hourly basis against the marginal emissions profile for the same hour.
5. The total hourly, monthly and annual emissions are used to produce summary charts, including hourly avoided emission rates.
6. The avoided emission rates can be used to produce a calculator in Microsoft Excel format, and this calculator provides total avoided emissions from annual generation or savings using either project-specific profiles or default regional profiles. This calculator also can be used to estimate avoided emissions on a monthly or seasonal basis.

So, while the state of EMV continues to challenge our ability to accurately and economically estimate energy savings due to EE measures and programs, parallel problems and tradeoffs between cost and accuracy affect the ability to translate energy savings into emission reductions. However, according to Jacobson and High, a good first step would be to use the eGRID non-baseload methodology instead of the eGRID system average methodology to calculate emission reductions from EERE.[81]

[81] Debra Jacobson, personal communication, November 22, 2009.

Appendix B

Comparison of State Technical Resource Manuals Treatment of Savings from Selected Measures

Table 7: Example Compact Fluorescent Lamp

15W CFL installed in a living room, with a rated lifetime of 10,000 hours (per Energy Star calculator)

State	CT	PA	VT	California	Energy Star	CT as of 10/1/09
Hours/day	2.96	3	3.4	2.18 [1]	3	2.96
Wattage of old bulb	51	varies depending on individual measures	63.7	52.95	60	51
Annual Gross Electricity Savings (kWh)	38.9	varies depending on individual measures	60.5	29.99 [2]	49.3	38.9
Annual Gross Gas Savings (Therms)	not given	not given	not given	-0.99	not given	not given
In-Service Rate	70%	84%	73%	not given	100%	0.7
Free-ridership	9%	NTG = 1	6%	not given	0%	0.1
Spillover	15%		25%	not given	0%	0.2
Net Realization rate	73%	84%	86%	60%	100%	0.7
Annual Net Electricity Savings (kWh)	28.43	varies depending on individual measures	51.9	17.99	49.3	28.4
Annual Net Gas Savings (Therms)	not given	not given	not given	-0.59	not given	not given
Life in years	6	6.4	6.4	6.6	9.13	5.25
Lifetime Gross Electricity Savings (kWh)	233.4	varies depending on individual measures	387.5	197.93	450.1	204.2
Lifetime Net Electricity Savings (kWh)	170.6	varies depending on individual measures	332.4	118.76	450.1	149.3
Lifetime Gross Gas Savings (Therms)	not given	not given	not given	-6.52	not given	not given
Lifetime Net Gas Savings (Therms)	not given	not given	not given	-3.91	not given	not given
Lifetime Net Energy Savings (kWh including NG converted to kWh)	170.6	varies depending on individual measures	332.4	82.81	450.1	149.3

[1] CA in Oakland, using the average existing home, with gas furnace and central AC
[2] CA takes into account interactive effects with both heating and cooling

Table 8: Example Compact Fluorescent Lamp

15W CFL installed in a bathroom, with a rated lifetime of 10,000 hours (per Energy Star calculator)

State	CT	PA	VT	California	Energy Star	CT as of 10/1/09
Hours/day	0.65	3	3.4	2.18 [1]	3	0.65
Wattage of old bulb	51	varies depending on individual measures	63.7	52.95	60	51
Annual Gross Electricity Savings (kWh)	12.1	varies depending on individual measures	60.5	29.99 [2]	49.3	12.1
Annual Gross Gas Savings (Therms)	not given	not given	not given	-0.99	not given	not given
In-Service Rate	70%	84%	73% (already incorporated in gross energy savings)	not given	100%	0.7
Free-ridership	9%	NTG = 1	6%	not given	0%	0.1
Spillover	15%		25%	not given	0%	0.2
Net Realization rate	73%	84%	86%	60%	100%	0.7
Annual Net Electricity Savings (kWh)	8.84	varies depending on individual measures	71.1	17.99	49.3	8.8
Annual Net Gas Savings (Therms)	not given	not given	not given	-0.59	not given	not given
Life in years	6	6.4	6.4	6.6	9.13	5.25
Lifetime Gross Electricity Savings (kWh)	72.6	varies depending on individual measures	387.5	197.93	450.1	63.5
Lifetime Net Electricity Savings (kWh)	53.1	varies depending on individual measures	455.3	118.76	450.1	46.4
Lifetime Gross Gas Savings (Therms)	not given	not given	not given	-6.52	not given	not given
Lifetime Net Gas Savings (Therms)	not given	not given	not given	-3.91	not given	not given
Lifetime Net Energy Savings (kWh including NG converted to kWh)	53.1	varies depending on individual measures	455.3	82.81	450.1	46.4

[1] CA in Oakland, using the average existing home, with gas furnace and central AC
[2] CA takes into account interactive effects with both heating and cooling

Table 9: Example Refrigerator

Energy Star auto-defrost refrigerator with top mount freezer and no door ice, 14.75 ft^3 fresh volume, 6.76 ft^3 freezer volume, so adjusted volume of 25.77 ft^3 (fresh volume + 1.63*freezer volume)

State	CT	PA	VT	CA	Energy Star
AV	25.8	25.8	25.8	20-25 ft^3 (TV)[1]	25.8
Baseline efficiency (kWh/year)	449.3	not given	528.5	697.0	528.5
Assumed Efficient model efficiency (kWh/year)	422.8	not given	422.8	452.0	422.8
Hours/year	not given	not given	5000.0	not given	not given
Gross savings (kWh/year)	26.4	80.0	117.1	161.5 [2]	105.7
Gross savings (therms/year)	not given	not given	not given	-7.3	not given
Free-ridership	not given	not given	0.3	not given	not given
Spillovers	not given	not given	0.3	not given	not given
Realization rate	1.0	1.0	1.0	0.6	not given
Net savings (kWh/year)	26.4	80.0	117.1	92.0	105.7
Net savings (therms/year)	not given	not given	not given	-4.1	not given
Lifetime (yrs)	13.0	13.0	17.0	14.0	12.0
Lifetime gross kWh savings	343.6	1040.0	1990.7	2260.7	1268.5
Lifetime net kWh savings	343.6	1040.0	1990.7	1288.6	1268.5
Lifetime gross therm savings	not given	not given	not given	-101.6	not given
Lifetime net therm saving	not given	not given	not given	-57.9	not given
Lifetime Net Energy Savings (kWh including NG converted to kWh)	not given	not given	not given	756.5	not given

[1] CA in Oakland w/gas furnace and central air
[2] CA takes into account interactive effects with both heating and cooling

Table 10: Example Dishwasher

Residential Energy Star dishwasher with electric water heating installed at the end of the previous dishwasher's life

State	CT	PA	VT	California[1]		Energy Star[2]
Assumed EF of efficient equipment	0.68	not given	0.58	0.58		0.68
Baseline EF	0.65	not given	0.46	0.46		0.46
kWh per year savings	15	137	113.3	97	72	96
Therms per year savings	0	not given	0	4	3	2.9
Gallons water per year savings	108	not given	134.6	not given		258
Washes per year	not given	not given	264	215	160	215
Life of measure (yrs)	11	11	13	11		not given
Total gross kWh savings	165	1507	1472.9	1067	792	not given
Total gross water savings (gallons)	1815	not given	19147.7	not given		not given
Net Realization Rate	100%	100%	100%	41%		not given
Total net kWh savings	165	1507	13	437.47	324.72	not given
Total net water savings	1815	not given	19147.7	not given		not given

[1] The two columns represent different assumptions about dishwasher cycles.
[2] Energy Star updated the dishwasher requirement on August 11, 2009.

Table 11: Example Room Air Conditioner

Residential Energy Star room AC, 12,000 Btu/hr

State	CT	PA	VT	CA	Energy Star
Assumed EER of efficient equipment	11.3	not given: ENERGY STAR	10.8	DEER only has measures for central AC	10.8
Baseline EER	9.8	not given	9.8		9.8
Btu/hr	10000	not given	10000		10000
hours in use per year [1]	272	737	375		1320
Gross kWh savings per year	36.84	70	37.4		125
Net Realization Rate	100%	100%	100%		
Free-ridership			33%		
Spillover			33%		
Net kWh savings per year	36.84	70	37.4		125
Lifetime (years)	12	10	13		9
Total net savings (kWh)	442.12	700	486.2		1125
Total net savings if 300 hrs per year	487.6	284.9	389.0		255.7

[1] PA assumes Pittsburgh; Energy Star assumes DC

Table 12: Example Commercial Lighting Retrofit

Commercial lighting retrofit, 20,000 s.f. office space, from ASHRAE 90.1-1999 compliant to 30% below ASHRAE 90.1-2007 without occupancy sensors, 3 lamps/fixture

State	CT	PA [1]	VT	CA [2]
Baseline lighting power density (W/ft2)	1.3	1.3	not given	not given
Baseline watts per fixture	not given	not given	110	115
Efficient LPD (W/ft2)	0.7	0.7	not given	not given
Efficient watts per fixture	not given	not given	86	78
Square footage	20000	20000	20000	10002.3
SF adjusted	n/a	n/a	n/a	20000
fixtures per s.f.	not given	not given	0.02	0.02
Total fixtures	not given	not given	332.23	366.30
Hours/year	3748	3640	3453.00	2594.00
Gross savings per fixture/year			86.09	96.03
Savings from lighting only (kWh/year)	44976	43680	28600.48	35174.36
Fraction of kWh savings that have to be removed by cooling	0.675	variable	not given	not given
Additional cooling savings (kWh/year)	12649.5	variable	not given	6113.55
Waste heat factor for energy	not given	not given	1.06	not given
Gross savings (kWh/year)	57,625.50	43,680.00	28,600.48	41,287.91
Space heating increase from reduced lighting load (MBTU/year for CT & VT, therms/year for CA)	35.53	not given	33.52	180.83
Avg gross realization rate/ISR	103%	100%	98%	not given
Avg free-ridership	15%	0%	21%	not given
Avg spillover	3%	0%	0%	not given
Avg net realization rate	89%	100%	77%	85%
Net savings (kWh/year)	51,540.82	43,680.00	22,049.06	35,094.73
Net heating increase (MBTU or Therms)	31.78	not given	25.84	153.71
Lifetime (yrs)	13	15	15	15
Net lifetime savings (MWh)	670.03	655.20	330.74	526.42
Net lifetime heating increase (MBTU or Therms)	413.13	not given	387.60	2305.62
Net total savings (heating converted to MWh)	669.99	655.20	330.70	314.51

[1] For PA, assumed 14 hrs/day, 5 days a week
[2] For CA, Oakland, 48in, T8 lamp assumed

Figure 13: Example Motor

FEMP-qualified Industrial motor for heating in an HVAC application in a 1-shift facility, ODP, 1800 RPM, 100 hp, 100% loaded 10% of the time, 80% loaded 40% of the time, 50% loaded 40% of the time, 30% loaded 10% of the time

State	CT	PA	VT	CA
Baseline efficiency	94.1%	94%	94.1%	not given
Efficient model efficiency	95.4%	95%	95.4%	not given
HP	100	100	100	100
Peak factor	0.8	0.75 [1]	0.75	not given
Hours/year [2]	2857	2380	2762	2820
Baseline kW	not given	not given	59.46	not given
Efficient model kW	not given	not given	58.65	not given
Annual gross kWh savings	2469.13	1928.33	2237.8404	2284.8
Avg gross realization/ISR	81%	100%	100%	84%
Avg free-ridership	36%	0%	7%	not given
Avg spillover	0.2%	0%	-18% [3]	not given
Net realization	52%	100%	76%	84%
Total net kWh savings/year	1286.15	1928.33	1699.237	1919.232
Lifetime (years)	17.5	20	20.00	15
Total net lifetime kWh savings	22507.67	38566.693	33984.74	28788.48

[1] No rated load factor given in the PA TRM and no value easily found - this may just be a reflection of the % loaded time

[2] for PA, based on PGH heating; for VT, based on miscellaneous HVAC heating pump

[3] for VT, this appears to be a negative spillover factor, but it is uncertain why this is so.

Figure 14: Example Industrial Non-HVAC Motor

Industrial motor for a non-HVAC application, ODP, 1800 RPM, 100 hp, 100% loaded 10% of the time, 80% loaded 40% of the time, 50% loaded 40% of the time, 30% loaded 10% of the time, 4000 hours per year total

State	CT	PA	VT	CA
Baseline efficiency	94.1%	94.1%	94.1%	not given
Efficient model efficiency	95.4%	95.4%	95.4%	not given
HP	100	100	100	100
Load factor		0.75	0.75	
Hours/year	4000	4000	4000	4000
Baseline kW	not given	not given	59.46	not given
Efficient model kW	not given	not given	58.65	not given
Hours at 100% or at 30%	400	not given	not given	not given
Hours at 80% or at 50%	1600	not given	not given	not given
kWh savings at 100%	432.12	not given	not given	not given
kWh savings at 80%	1382.78	not given	not given	not given
kWh savings at 50%	864.24	not given	not given	not given
kWh savings at 30%	129.64	not given	not given	not given

Total gross kWh savings/year	2808.78	3240.90	3240.90	3240.85
Avg gross realization/ISR	81%	100%	100%	84%
Avg free-ridership	36%	0%	7%	not given
Avg spillover	0.2%	0%	-18%	not given
Net realization	52%	100%	76%	84%
Total net kWh savings/year	1463.07 [1]	3240.90	2460.88	2722.31
Lifetime (years)	17.5	20	20	15
Total net lifetime kWh savings	25603.76	64817.97	49217.58	40834.72

[1] for CT, no life given for non-HVAC motors

Table 15: Example Commercial Chiller

500-ton Centrifugal Water-Cooled Chiller in an office

State	CT	PA	VT	CA [1]	FEMP
Baseline efficiency (COP)	not given	not given	6.1	not given	not given
Efficient model (COP)	not given	not given	varies	not given	not given
Baseline efficiency (kW/ton)	0.55	0.577	0.58	0.58	0.68
Efficient model (kW/ton)	0.53	AHRI compliant	varies	0.46	0.56
Tons	500	500	500	500	500
Baseline energy use (kWh/yr)	not given	not given	not given	not given	680000
Efficient energy use (kWh/yr)	not given	not given	not given	not given	560000
Cooling hours/yr	797	737 [2]	varies	not given	2000
Gross energy saving per ton	not given	not given	not given	196.62	not given
Gross Energy Savings (kWh/yr)	7970	varies	varies	98311	120000
Average Gross Realization rate	94%	not given	not given	not given	not given
Average free-ridership	16%	not given	4%	not given	not given
Average Spillover rate	2%	not given	0%	not given	not given
Net Realization Rate	81%	100%	96%	50%	not given
Net Energy Savings (kWh/yr)	6461.1	not given	not given	49155.5	120000
Years	20	20	25	20	23
Net Total energy savings (MWh)	129.22	varies	varies	983.11	2760

[1] for CA, based on Oakland
[2] for PA, based on Pittsburgh