ASSESSING PERSISTENCE AND OTHER EVALUATION ISSUES

Daniel M. Violette, Navigant Consulting

Addressing other evaluation issues that have been raised in the context of energy-efficiency (EE) programs, this chapter focuses on methods used to address the persistence of energy savings, which is an important input to the benefit/cost analysis of EE programs and portfolios. In addition to discussing "persistence" (which refers to the stream of benefits over time from an EE measure or program), this chapter provides a summary treatment of these issues:

- Synergies across programs,
- Rebound,
- Dual baselines, and
- Errors in variables (the measurement and/or accuracy of input variables to the evaluation).

This first section of this chapter contains a definition of persistence and identifies issues in its evaluation. The state of the practice in persistence is addressed, examples taken from persistence studies are presented, and recommendations for addressing persistence are presented at the end of the section. The other evaluation issues are addressed in the second section of the chapter. Appendix A presents a matrix of persistence issues and methods by program type.

1 Persistence of Energy Savings

Understanding persistence is critical to making good decisions regarding EE investments, so this section outlines program evaluation methods that can be employed to assess persistence—the reliability of savings over time. EE program benefits are measured as the net present value (NPV) of a stream of benefits based on the energy and demand savings[1] achieved by the program. Depending on the mix of measures and their assumed lives, these benefits may extend to 15 years (or more) for some measures. As a result, assumptions about the persistence of savings over time influence the EE benefit-cost tests. Extrapolating savings beyond the evaluation period has often been based on engineering judgment, manufacturer specifications, and some empirical work (the factors used to develop projections of measure lifetimes and degradation).

The protocols developed under the Uniform Methods Project (UMP) in other chapters generally focus on estimating first-year savings. There is also some discussion, however, about estimating first- and second-year savings when more participants from a second program year are needed for the impact evaluation. These initial evaluations are often quite detailed, assessing both the

[1] This chapter focuses on estimating energy savings, but the persistence of reductions in demand may also be important for some measures and programs. Issues raised here may also be important for programs and policies focused on reducing demand during peak periods.

savings and the quality of the program in terms of installation, engineering calculations, and equipment selection (where on-site visits are used to validate initial "*claimed*" estimates).

1.1 Addressing Persistence

Persistence of savings encompasses both the retention and the performance degradation of measures. Together, these factors are used to estimate how the *claimed* persistence values used in program planning can be updated based on *evaluated* savings values.[2] Different jurisdictions define and treat the components of overall persistence differently. As a result, defining what is meant by overall persistence and addressing some of the subtle context issues are important to the discussion.

There are a number of subtle aspects to the context and definition of overall persistence. Consistent and practical definitions for use in developing estimates of the overall persistence of savings over time were developed for the Joint Massachusetts Utilities (Energy & Resource Solutions 2005).[3] In that study, overall persistence is divided into two components: (1) measure life, and (2) savings persistence.

Recognizing that definitions for the terms *persistence* and *realization of savings* are not nationally consistent, the definitions based on the Massachusetts framework and outlined below provide a structure that can be addressed by evaluation and verification methods. That is, these definitions use categories of effects and factors that can be quantified using evaluation methods. For example, it is difficult to estimate technical measure life based on on-site inspections, as there may be many reasons that a measure is no longer in place. Thus, technical measure life and other reasons for measure non-retention are combined in the definition "measure life," which is simply the time a measure can be expected to be in place and operable.

1.1.1 Definitions

The definitions of key terms used in this chapter are these.

1.1.1.1 Measure Life or Effective Useful Life (EUL)

This is the median number of years that a measure is in place and operational after installation. This definition implicitly includes equipment life and measure persistence (defined below), but not savings persistence.

- "Equipment life" is the number of years installed equipment will operate before it fails.

---

2 In this chapter and consistent with other chapters, *claimed* savings means the same as *ex ante* savings and *evaluated* savings is used instead of *ex post* savings. This note is to eliminate confusion for those more familiar with the use of "*ex ante*" (initial savings estimates) and "*ex post*" (evaluated savings) terminology in describing evaluation methods.

3 This study for the Joint Massachusetts Utilities' defines "measure life" as the median number of years that a measure is installed and operational. This definition implicitly includes equipment life and *measure* persistence. However, *savings* persistence is the percentage of change in expected savings due to changed operating hours, changed process operation, and/or degradation in equipment efficiency relative to the baseline efficiency option.

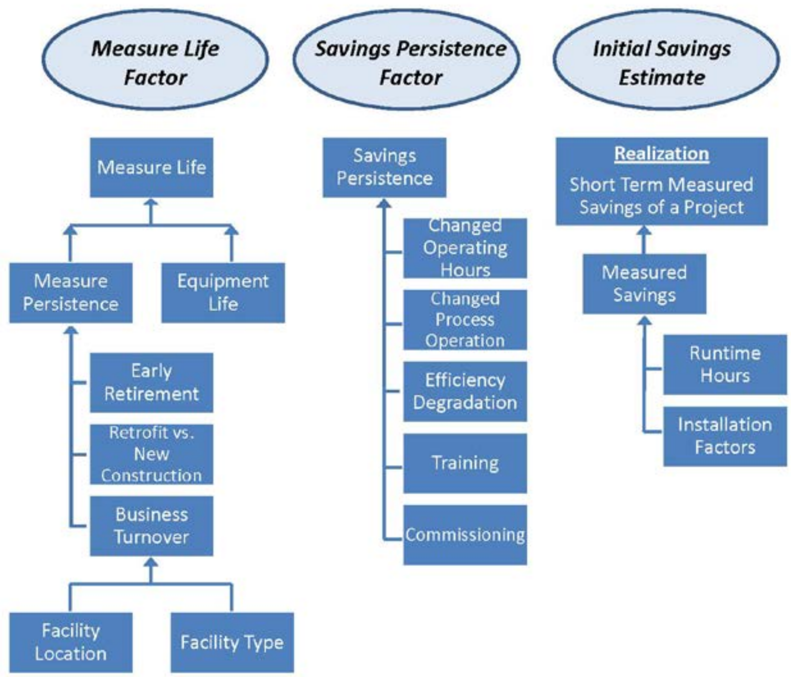
- “Measure persistence” takes into account business turnover, early retirement or failure of the installed equipment, and any other reason the measure would be removed or discontinued.

1.1.1.2 Savings Persistence

This is the percentage of change in expected savings due to changed operating hours, changed process operations, and/or the performance degradation of equipment efficiency relative to the baseline efficiency option. For example, an industrial plant that reduces operation from two shifts to one shift may then have a *savings* persistence factor of 50%, as only half of the projected energy savings would be realized. Also, improper operation of the equipment may negatively affect savings persistence, so training and commissioning could improve savings persistence. Finally, most equipment efficiency degrades over time, so annual energy savings may increase or decrease relative to the efficiency degradation of the baseline efficiency option.

Figure 1 illustrates how the two persistence factors are used to produce savings that are adjusted for persistence: Savings Adjusted for Persistence = (Measure Life Factor) x (Savings Persistence Factor) x (Initial Savings Estimate).

Figure 1. Relationship of Measure Life, Savings Persistence, and Initial Savings Estimates[4]



1.1.2 Factors for Selecting a Persistence Study

The following are several important factors to consider when selecting the type of study to examine energy savings persistence

1.1.2.1 Available Claimed Estimates of Persistence.

There are almost always initial *claimed* estimates of the assumed stream of savings for a program (based on current estimates of measure life and degradation). These estimates are used in the initial benefit/cost analyses conducted as part of program design or in the benefit/cost tests of initial program evaluations efforts. As a result, most studies of persistence test the initial *claimed* stream of savings against the *evaluated* results to check for significant differences.[5] The outcome is often presented as a realization rate (that is, the *evaluated* values divided by the initial *claimed* values), which is the year-by-year savings estimate used in benefit/cost studies.

---

[4] Source: *Adapted from Energy &Resource Solutions (2005).*

[5] Starting with a set of *claimed* savings allows for the use of evaluation methods that leverage these initial data through the use of ratio estimates and a "realization rate" framework.

1.1.2.2 Uncertainty in Claimed Estimates.

When deciding whether to conduct a new study of persistence—and the corresponding level of effort required—consider the confidence that the evaluator or decision-maker has in the *claimed* stream of savings values. If the uncertainty is perceived as being high *and* a sensitivity analysis shows that plausible revisions to persistence of energy savings substantively changes the results of benefit/cost tests, then a new study may be worthwhile. Such an undertaking regarding persistence may result in revisions to the current *claimed* estimates.

For example, measures that account for greater savings, have shorter measure life values, or may be subject to near-term degradation in savings are more important to evaluate, as they will have a greater impact on the resulting benefit/cost tests. However, changes in measure life that do not take effect until the 14th or 15th year of the measure may be discounted in the NPV calculation (discussed below). Thus, in terms of the effect on the benefit/cost calculation, the additional work needed to estimate these values may not be worthwhile.

1.1.2.3 Discounting Values of Energy Savings over the Life of the Measure.

The stream of program benefits over time is discounted, resulting in near-term savings estimates that have a larger impact on the NPV of benefits than the values further out in the future. For example, the effect of research on the measure life of a second refrigerator retirement that extends it from six years to eight years would be muted somewhat in the benefit/cost analysis, due to discounting. Specifically, the energy savings from this updated measure life of two additional years would be muted in its application by discounting the benefits for year 7 and year 8. The impact of discounting depends on the discount rate being used and the measure life.[6]

1.1.2.4 Differences in Baseline and EE Energy Streams of Benefits.

Energy savings calculations are based on the difference between the post EE state and the assumed baseline. If the baseline equipment has the same level of degradation in performance, then the energy savings factor due to degradation would be 1 *and* it would be appropriate to assume constant energy savings over the life of the EE measure.[7] In fact, if the relative persistence of savings is higher for the EE measures compared to a baseline consisting of standard measures, then energy savings not only persists, but can increase over time.

Those four factors are meant to address the following questions:

---

[6] For example, if a discount rate of 5% is used, the savings will be reduced by 0.78 multiplied by the energy savings at five years. At 10 years and a 5% discount rate, the new value would be 0.61 multiplied by the energy savings. At a discount rate of 7% for a 10-year period, the value would be 0.51 multiplied by the energy savings.

[7] The report from Peterson et al. (1999) is a good example of degradation being measured for both an efficient appliance offered by an EE program and standard equipment. This study showed that the high-efficiency coils start with and maintain a higher efficiency than standard efficiency coils. The slower degradation rate increases the life of the equipment, and the equipment uses less energy over its operational lifetime. Even though both high-efficiency units and standard units showed performance degradation over time, the lower rate of degradation in the high-efficiency units resulted in a recommended degradation factor exceeding 1.0 in most years. This factor increased from 1.0 to 1.08 over the 20-year expected life of the unit, indicating that savings not only persisted, but actually increased relative to the baseline over the assumed life of the equipment.

- If a persistence study is conducted, is there a reasonable likelihood that the new trend in energy savings over time would be substantively different from the assumptions used in the initial benefit/cost analyses?
- Would the NPV benefits of the program change with a new persistence factor, the discount rate being used, and the likely change in the baseline energy use level that may also be due to performance issues of the baseline equipment?

There may be good reasons to assess persistence, as many factors can influence the stream of energy savings over a three- to 10-year period. The most these common factors are listed in Table 1.

Table 2. Factors Influencing Persistence

Residential Sector Programs and Measures	Commercial and Industrial Sector Programs and Measures
1. Changes in ownership 2. Maintenance practices 3. Changes in equipment use 4. Behavioral changes 5. Occupancy changes 6. Inappropriate installation of equipment 7. Manufacturer performance estimates that do not reflect in-field operating conditions.	1. Business turnover 2. Remodeling 3. Varying maintenance 4. Operating hours and conditions 5. Inappropriate installation of equipment 6. Manufacturer performance estimates that do not reflect in-field operating conditions

Sensitivity analyses using the benefit/cost models can highlight those measures for which adjustments in persistence will have the largest impact. This information can then be used to prioritize persistence evaluation efforts. Thus, before deciding whether additional analyses are needed, test the sensitivity of NPV benefits to potential changes in the persistence of savings. This can help determine whether the impact may be large enough to merit a substantial study effort, or sufficiently small, requiring only a modest retention study.

1.2 State of the Practice in Assessing Persistence

Professional judgment plays a significant role in selecting a method for assessing persistence. The California EE Evaluation Protocols (California Public Utilities Commission, 2006) has several types of retention, degradation, and measure life/EUL studies from which to select, based on the priority given to the issue by regulatory staff or other stakeholders.

Evaluators seem to rely on the following two processes for developing estimates of persistence:

- ***Database or Benchmarking Approach.*** This entails developing and regularly updating[8] a database of information on measure life and performance degradation.

[8] As it is important that these benchmarking studies be updated on a regular basis, the cost of these updates should be included in the cost estimate for using this approach. While these studies may not appear costly on a one-time basis, the effort required to update the database regularly can be significant. This is important, as these

- ***Periodic In-Field Studies.*** This entails performing selected in-field studies of program participants from earlier years.

These two approaches are not necessarily mutually exclusive. The database/benchmarking approach is often used when: (1) there is a large number of EE measures; (2) there are concerns about the sample sizes required for in-field studies; and (3) the cost of conducting in-field persistence studies is an issue. Periodic studies may be used for updating a database of measure life and performance degradation. Such studies are also useful when focusing only on those measures that account for a large fraction of the savings. Additionally, in-field studies of program participants that are conducted a number of years after participation provide direct information on persistence of savings for that program.

1.3 Database/Benchmarking Approaches

The three examples of database/benchmarking approaches presented below are based on:

- Engineering judgment
- Experience with the EE measures, and
- Information on local and regional conditions to develop tables of measure lives for use in EE program planning.

These values are often used as deemed values for persistence and applied to produce estimates of the energy savings over time (as inputs to benefit/cost calculations). An assessment of this approach follows the examples. (References to each study are provided for those wanting more information on the methods used beyond the short descriptions provided below.)

1.3.1 Example Study 1: GDS Associates (2007)

Objective: The measure life values presented in this report were developed to meet the following conditions:

- Accurately reflect conditions for measures installed by EE programs in the New England states that have supported this research effort;
- Satisfy any ISO-NE requirements (e.g., for definition and documentation sources); and,
- Work as common values, accepted by all New England states for the FCM (i.e., the ISO-NE forward capacity market).

Methodology: "Reviewed all secondary data collected and developed a preliminary list of potentially applicable residential and C&I measures. This list was then distributed to program administrator staff within the SPWG for review and to obtain additional program-specific measure life values and associated documentation sources. GDS compiled all responses and developed initial measure life recommendations for SPWG member consideration."

---

databases are sometimes the source of deemed values for measure life and persistence of savings that are used in evaluation efforts.

1.3.2 Example Study 2: KEMA (2009)

Objective and Methodology: "The principal objective of this study was to update the current measure life estimates used by the Focus Evaluation Team and the Focus Program. **The evaluation team's approach to this study consisted entirely of secondary research**; the team did not conduct primary research, fieldwork, or produce a savings persistence study." (Emphasis added.)

1.3.3 Example Study 3: Energy & Resource Solutions (2005)

Objective: "The primary goals of the Common Measure Life Study were as follows:

- Define measure life and related terms, such as persistence
- Review the provided table of current measure lives
- Survey other utility EE programs
- Develop a table of technological measure lives
- Recommend common measure lives and persistence assumptions to be used by the sponsors."

Methodology: "ERS reviewed the tables of agreed-upon and disputed measure lives provided by the sponsoring utilities. As tasked in our proposal, we researched several sources to use in support of selecting individual measure lives. We first thoroughly researched the CALMAC database. The CALMAC database provides a public depository for all persistence, technical degradation factor (TDF) and other related studies performed in the State of California. Next, we surveyed many electric utilities and state utility commissions throughout the nation, obtaining other utilities' tables of measure lives. We obtained measure life tables used in 8 states by at least 14 different utilities. Finally, we performed a literature search, referenced technical sources and consulted equipment manufacturers to establish a table of technical lives for each measure. In conjunction with these efforts, we specifically researched the effect of New Construction versus Retrofit status on measure lives, as well as the effect of Small versus Large businesses."

1.4 The Challenges of New Technologies and Measures

The methods in the three examples above have produced useful estimates for a wide number of measures where practical information exists from measure installations and field work. However, new technologies and measures installed less frequently pose greater challenges for this judgment-based benchmarking approach. For many widely implemented EE measures, both the evaluation work and additional on-site engineering work (such as installation and maintenance) provide a basis for the use of informed engineering judgment. A series of retention/survival rate studies in California—conducted from 1994 to 2006—found that most *claimed* estimates could not be rejected by the in-field studies. However, the in-field studies often had small sample sizes for certain measures and short time frames that did not allow for many failures to occur in the dataset.

Some important measures in these engineering and expert-developed measure life tables may not have fared well. Both residential lighting and commercial lighting have provided a large fraction of savings, and the persistence of these savings has been controversial. Nexus (2008) found that the life for certain lighting measures depends not only on the equipment, but also on the program design.

Skumatz (2009, 2012) critiques the database/benchmarking approach, which is based on engineering judgment combined with literature reviews. Skumatz (2012) identifies strengths and weaknesses in this approach compared to on-site data collection, and she offers suggestions for improving current estimates. Skumatz notes that measure life values existing in tables often vary by more than 25%, and that this has “precisely the same impact on a measure’s or program cost-benefit ratio” as savings values that are off by 25%.

While this comment has merit, the measure life and persistence factors will start at 1.0 in the initial years of the program and then gradually change. This change in savings is offset to some degree by the discounting of benefits from five, 10, and 15 years out. Also, this single measure with varying measure life values across engineering-based tables may not represent the composite effective life of a group of measures that comprise a program.

1.5 In-field Persistence Studies (survey and on-site data approaches)

Methods that make use of in-field data collected on program participants at some point after they participated in an EE program generally rely on the following:

- Surveys or on-site visits to determine whether the measure is still in place and operable, or, if the measure was removed, when and why.[9]
- Statistical analyses using regression-based methods to estimate retention/survival models that produce estimates of the survival or failure rates of EE measures.

The California EE Evaluation Protocols[10] specified these three categories of methods used for in-field studies of persistence:

- ***Retention Studies*** provide the percentage of the measures that are in place and operable at a point in time. Retention studies identify technology design, define operable conditions, and describe how operable conditions could be measured.
- ***Measure Life/EUL*** estimates the median numbers of years that the measures installed under the program are still in place and operable. This value is calculated by estimating the amount of time until half of the units will no longer be in place and operable.

[9] One reviewer suggested that the surveys referred to in this section should specifically include online approaches. The topics of using online surveys to obtain customer-specific information and combining online surveys with other methods are discussed in the chapter on Survey Research.

[10] The methodology language from the California EE Evaluation Protocols (California Public Utilities Commission 2006) has been adapted to fit the measure life definition and persistence structure used in this chapter. One difference is the use of *persistence* as the overarching term for all types of changes in energy savings over time, which the California Protocols document addresses in the *Effective Useful Life Protocol* section (p. 105). The California Protocols still contain the most comprehensive discussion of methods for assessing persistence.

- ***Performance Degradation*** uses both technical and behavioral components to measure time-related and use-related changes in energy savings relative to a standard efficiency measure or practice. In general, both standard equipment and EE equipment become less efficient over time, regardless of the equipment measure life. This factor is a ratio reflecting the decrease in savings due to performance degradation from the initial year savings.

1.5.1 Retention and Measure Life Studies

A retention study determines the number of installed and operable measures at a given point in time. A measure life study is an extension of a retention study, where there is adequate data to allow for the development of a statistical model (commonly called a "survival analysis") to estimate failures that might occur after the data are measured.

Information from the retention model provides an estimate of the measures that were installed and operating at a point in time, which allows the evaluator to calibrate the *claimed* savings and produce adjusted *evaluated* estimates of savings over time. The current estimates of persistence are adjusted to account for the new information *and* the stream of savings over the year. These estimates could, for example, be adjusted in year 4 to be consistent with the retention study. This ratio for year 4 would then be used to adjust the savings in all subsequent years.

The measure life estimation methods, which are based on survival analysis, provide more information. However, estimating measure life requires a much larger sample—one that contains an adequate number of both installed and missing (that is, uninstalled or replaced) equipment.

The following are two types of retention and measure life methods, which have been used to estimate the survival models that produce estimates of measure life. (Studies using these methods are described later in this section.)

1.5.1.1 In-place and operable status assessment (using on-site inspections)

The in-place assessment studies are verified through on-site inspections of facilities. Typically, the measure, make, and model number data are collected and compared to participant program records, as applicable. As-built construction documents may also be used to verify selected measures when access is difficult or impossible (such as wall insulation). Spot measurements may be used to supplement visual inspections—such as solar transmission measurements and low e-coating detection instruments—to verify the optical properties of windows and glazing systems.

Correct measure operation is observed and compared to the project's design intent. Often, this observation is a simple test of whether the equipment is running or can be turned on. However, the observation and comparison can extend to changes in application or sector, such that the operational nature of the equipment no longer meets the design intent. For example, working gas-cooking equipment that had been installed in a restaurant but is now installed in the restaurant owner's home is most likely no longer generating the expected energy savings, so it would not be counted as a program-induced operable condition.[11]

1.5.1.2 Non-Site Methods

Typical non-site methods include telephone surveys/interviews, analysis of consumption data, or the use of other data (such as from energy management systems). The goal is to obtain essentially the same data as would be gotten through an on-site verification; however, there is the potential for collecting inaccurate data, due to a number of factors (and discussed in the Sample Design chapter).

1.5.1.3 Examples of Retention and Measure Life Studies

Two examples of these types of studies were performed by KEMA and by Nexus Market Research.

- KEMA (2004) used a telephone survey to gather information on refrigerators at years 4 and 9, as part of a review of an appliance recycling program.
- Nexus Market Research (2008) used on-site verification data to conduct a measure life study of residential lighting measures.

Both studies provide good examples of collecting information for a basic retention study, and they serve as illustrations of the statistics necessary to estimate a survival model.[12] Each is discussed below.

Example Study 1: KEMA (2004). Conducted with program participants from the years 1994 through 1997, this study looked at retained savings over this period.

For each year, the measure life/EUL estimate reflects the following factors:

- The time at which half of the recycled appliances are from participating premises that have added an appliance, and
- The time at which half of the recycled appliances would have been out of service without the program influence.

The KEMA study illustrates one way in which the *claimed* and *evaluated* measure life values can be used. As stated in the study:

---

[11] In addition to this language, the California EE Evaluation Protocols outline certain sampling criteria that must be met in California. However, these criteria may vary in accordance with the requirements of different jurisdictions.

[12] To assist evaluators, the California EE Evaluation Protocols state: "Multiple statistical modeling packages (SAS®, Stata®, SPSS®, R®, S+®, and others) provide survival analysis programs. There are several commercial and graduate textbooks in biostatistics that are excellent references for classic survival analysis. One of these used as reference for some of the prior EUL studies in California is the SAS® statistical package and the reference *Survival Analysis Using the SAS® System: A Practical Guide* by Dr. Paul D. Allison, SAS® Institute, 1995. Several model functional forms are available and should be considered for testing. These forms include logistic, logistic with duration squared (to fit expected pattern of inflection point slowing of retention losses), log normal, exponential, Weibull, and gamma."

For each of the program years from 1994 through 1997, both refrigerators and freezers have a claimed (or *ex ante*) estimate of measure life/EUL of six years, which has been used in the earnings claims to date. A measure's evaluated measure life/EUL is the value estimated by a persistence study. If a measure's claimed measure life/EUL is outside the 80% confidence interval, the measure's evaluated measure life/EUL may be used for future earnings claims. Otherwise, the measures claimed value will continue to be used in earnings claims.

Figure 2 is a replication of Table E-1 from the KEMA study, which shows the comparison between the *claimed* and *evaluated* measure life/EUL estimates. In this case, the measure life results showed that the program was under-estimating the measure life/EUL values *and* that the realization rate exceeds 1.0.

Figure 2. KEMA (2004) Table E-1

Table E-1
1994-1997 Appliance Recycling Program
Summary of Effective Useful Life Estimates

			EUL (years)					
						80% Confidence Interval		
Program Year	Measure	End Use	Ex Ante	Ex Post (estimated from study)	Adopted ex post (to be used in claim) Lower Bound	Lower Bound	Upper Bound	EUL Realization Rate (adopted ex post/ex ante)
1994	Freezer	Refrigeration	6.0	8.0	8.0	8.0	11.0	1.33
	Refrigerator		6.0	8.0	8.0	8.0	11.0	1.33
1995	Freezer	Refrigeration	6.0	8.0	8.0	8.0	11.0	1.33
	Refrigerator		6.0	8.0	8.0	8.0	11.0	1.33
1996	Freezer	Refrigeration	6.0	8.0	8.0	8.0	8.0	1.33
	Refrigerator		6.0	8.0	8.0	8.0	8.0	1.33
1997	Freezer	Refrigeration	6.0	8.0	8.0	8.0	8.0	1.33
	Refrigerator		6.0	8.0	8.0	8.0	8.0	1.33

Example Study 2: Nexus Market Research (2008). This study examined the measure life of lighting products distributed through EE programs in New England.

The definition of measure life is the same as presented above in the Addressing Persistence section and used in Energy & Research Solutions (2005) example application presented above. Specifically, Nexus states that:

> [T]he measure life estimates do not distinguish between equipment life and measure persistence; our estimates—one for each measure category—include both those products that were installed and operated until failure (i.e., equipment life) as well as those that were retired early and permanently removed from service for any reason, be it early failure, breakage, or the respondent not liking the product (i.e., measure persistence).

Nexus drew a random sample of participants based on the type and number of products they had obtained through the programs. The report states, "We collectively refer to these sample products as the 'measure life products.'"

Auditors visited 285 homes to inventory lighting products, and Nexus designed a respondent survey to learn more about the measure life products and other lighting products found in the home. These survival analyses were based on the following methods and, ultimately, Nexus used estimates resulting from Method 3.

- Method 1: Measure Life Tables
- Method 2: Logit Regression
- Method 3: Parametric Regression Models of Survival Analysis

The results showed that the measure life for CFLs varies by program design (that is, whether the program was coupon-based, direct install, or a markdown at a retail facility). The results of the Nexus (2008) study are shown in Table 2.

Table 2. Nexus (2008) "Recommended Estimates of Measure Life – Decimals"

Product	Measure Life	80% Confidence Interval	
		Low	High
Coupon CFLs	5.48	5.06	5.91
Direct Install CFLs	6.67	5.97	7.36
Markdown CFLs (all states)	6.82	6.15	7.44
Coupon and Direct Install Exterior Fixtures	5.47	5.00	5.93
Markdown Exterior Fixtures	5.88	5.24	6.52
All Interior Fixtures	Continue using current estimates of measure life		

Nexus deemed a representation of the results—at an 80% confidence interval—as being accurate enough for the purposes of this study. Nexus recommended measure life estimates for three measures: one for compact fluorescent lamps (CFLs; coupon, direct install, and markdown)[13] and two for exterior fixtures (markdown and all other programs).

[13] Due to the diversity of program types throughout the region, Nexus used the term "markdown" to refer to both markdown programs (offered in all of the states) and buy-down programs (offered in some of the states).

Nexus did not recommend an estimate of measure life for interior fixtures, as the timing was too early in the measure lifecycle to provide a reliable estimate. This occurs with a number of measure life studies that are conducted too early (before there have been enough failures or un-installs to allow for statistical modeling of measure life).

1.5.2 Examples of Degradation Studies

While there are few reports that directly focus on the degradation of savings, two types of studies are available, and they are described below:

- Focusing on technical degradation. (One of the clearest examples is by Proctor Engineering in1999.)
- Performing a billing analyses at some point after participation to capture all of the factors that impacted persistence of savings. (In 2011, Navigant performed a billing analysis of a customer information program, which was used to examine persistence of impacts across two years for a behavioral program.)

Example Study 1: Proctor Engineering (1999). The purpose of this project was "to examine the relative technical degradation of demand side management (DSM) measures compared to standard efficiency equipment. This project covers two major DSM measures: commercial direct expansion air conditioners (Comm. DX AC) and EMS [energy management systems]."

Proctor Engineering's methodology involved establishing a time-series estimate—derived from available research—for condenser and evaporator coil fouling rates. Proctor used laboratory testing to modify the estimated fouling rates and establish a profile for coil fouling. It tested both high-efficiency and standard-efficiency coils in a controlled laboratory environment, and both were subjected to continuous fouling. Proctor then monitored the efficiency of the air conditioner at various intervals to document the effects.

The results of this study found that: (1) the impact on standard equipment was greater, and (2) the high-efficiency units actually had a higher level of savings persistence. The end result was that "testing shows that the TDF [technical degradation factor] for this measure is greater than one." This is an example of degradation needing to be conducted with reference to standard efficiency equipment. EE measures may have performance degradation, but so does standard equipment. If the EE measures have a lower rate of degradation, then savings increase (as measured against the standard equipment baseline).

To assess EMS, Proctor used an on-site methodology rather than laboratory testing. The research data showed that although there is some EMS savings degradation at some locations, other locations show increasing savings. Some of the causes for this persistence are:

- No instances of disconnected or non-operational EMSs were found.
- The vast majority of EMSs appeared to be operated in a competent and professional manner.
- EMS operators had found that the EMS was a useful tool in performance of their jobs.

Proctor Engineering contrasted its work with other EMS studies showing greater degradation due to operational issues. Proctor explained the comparatively high level of persistence it found as being due to the high interest of the program participants in saving energy. The more random group of facilities in the comparison may not have been involved in EMS-related EE programs.

Proctor also conducted a billing analysis to confirm these findings. For this billing analysis, it combined the consumption data from all of the sites and then estimated the persistence of savings over time. The regression process provided statistically significant estimations at the 95% level.[14]

The primary purpose of this research was to establish the technical degradation factors (TDF), estimated for each measure. The results from Proctor's study, seen in Figure 3, shows that the degradation factors are greater than 1.0 for the high-efficiency DX AC equipment. This indicates the degradation was less for the high-efficiency DX AC equipment than for the standard-efficiency equipment.

Figure 3. Proctor Engineering (1999) Table ES-1

Table ES-1 TDF

Year	EMS	Comm DX AC
1	1.00	1.00
2	1.00	1.00
3	1.00	1.00
4	1.00	1.01
5	1.00	1.01
6	1.00	1.01
7	1.00	1.01
8	1.00	1.01
9	1.00	1.01
10	1.00	1.02
11	1.00	1.02
12	1.00	1.02
13	1.00	1.02
14	1.00	1.02
15	1.00	1.02
16	1.00	1.02
17	1.00	1.02
18	1.00	1.02
19	1.00	1.06
20	1.00	1.08

Still, the difference is small through year 18, and this size of effect might not show up in benefit/cost analyses due to the discounting required to obtain an NPV of savings benefits.

[14] References to statistically significant results in regression analyses must be carefully interpreted. The analysis may have been a test to determine if the effect was significantly different from zero (±100% precision). Alternatively, the test may have actually established a precision level of ±10% or another level of precision, (for example, 30%). A statement of statistically significant results should be accompanied by an explanation for interpreting that statement in terms of the level of precision being used in the test of significance.

Example Study 2: Navigant (2011). This study examined the short-term persistence of a behavioral information program using billing data across multiple years, as short-term persistence may be an important factor for these programs.

The program was designed to assist and encourage customers to use less energy. These types of programs are increasing in the industry; for example, OPOWER, Inc., offers residential customers regular Home Electricity Reports about their electricity consumption to help those customers manage their electricity. In combination with other information, these reports compare a household's electricity use to that of its neighbors and then suggest actions to reduce electricity use. It is hypothesized that presenting energy use in this comparative fashion creates a social nudge that induces households to reduce their consumption.

Navigant evaluated the first 29 months of the program, with an emphasis on the second program year. The following main research questions were addressed in the evaluation and presented in this report:

- Does the program continue to generate savings?
- What is the trend in program savings? Is there a ramp-up period to savings? If so, for how long? Are savings now relatively stable, increasing, or falling?
- Do program savings increase with usage?

The evaluation of this program entailed developing a random control group and conducting a fixed-effects regression analysis, which is a common evaluation method. This regression method is discussed in the Whole House Retrofit Chapter of this UMP report.

Navigant's results showed that the effects of slightly more than 2% of the energy savings persisted across the 29 months examined in the study, after an initial ramp-up period of approximately 10 to 12 months. The small effect size required a large sample of customers for the regression analysis to produce reliable results. For this behavioral program evaluation, there were over 20,000 treatment customers and a control group of over 30,000 customers. Thus, large samples are needed to identify small effect sizes from EE programs.

This regression framework can be applied to a third and fourth year of data to assess longer-term participation.

1.6 Persistence Recommendations and Conclusions

Evaluators address the issue of persistence of savings from EE programs because of the impact that the stream of savings estimates has on the benefit/cost tests of measures and programs. While some measure life values are estimated at more than 20 years, most benefit/cost assessments are estimated out at least 10 years or, more commonly, 15 to 20 years.

The approaches discussed in this chapter include methods to address measure life and savings performance, which may be impacted by operating conditions, behavioral changes, turnover in building occupancy, changes in measure use, and other factors. To date, the tools and methods used which comprise the recommended tool kit for evaluators include:

- Benchmarking and database development for measure life values and savings persistence.
- On-site analyses of equipment.
- Survey methods for select measures amenable to survey techniques.
- Single-year estimations of equipment retention and operation.
- Multiyear statistical analyses based on survival models.
- Technical degradation studies based on engineering review.
- Technical degradation based on laboratory testing.
- Billing analyses that capture overall persistence (that is, that assess savings directly and capture all changes in savings for the time period being analyzed).

The review of methods illustrates the different ways persistence can be addressed. Research is continuing in this area, and methods have been adopted in different jurisdictions. As with any area of evaluation, there will always be improvements. Appendix A to this chapter presents tables outlining program and measure persistence study challenges and issues.

This balance of this section presents practical recommendations for assessing the persistence of savings. The goal of evaluation is to help stakeholders make good decisions about investments in EE programs, and this requires both an understanding of the techniques and applied judgment.

1.6.1 Recommendations

1. ***Before determining whether to undertake a large-scale persistence study of a program or measure (or even to undertake such a study at all), consider whether the results of the study are likely to have a material impact on the economics of the program.*** Persistence of savings refers to the stream of savings expected from a measure or program over a period of years. If the study's revised persistence of savings is expected to be small and to occur 10 or more years or more in the future, then the impact of that change may not have a large effect on the cost-benefit economics.

   Keep these considerations in mind when deciding:

   - Benefit-cost tests are based on NPVs that discount the streams of benefits and costs. A change in measure life by a year or two *and* changes for long-lived measures may not have much impact after they are discounted.
   - The performance degradation of EE measures should be assessed relative to that of the standard-efficiency equipment, as both will have performance degradation. The difference between these two values determines the impact on savings.

2. ***Select the methodology that best fits the individual circumstances of the measure/program being evaluated.***

   - Pick the method most appropriate to the magnitude of the effect expected. Before conducting the study, take a forward-looking view of what might be learned. While this may seem difficult, researchers across the evaluation community and the industry make these decisions on a regular basis. The key is to ensure that the information produced is worth the effort expended to produce it. The goal is to obtain information that decision makers need for making good decisions regarding EE investments.

- Measures that may have persistence impacts within the first three to seven years are the most important to study because of their near-term effects and their potential to influence the benefit/cost tests and program designs.
- As benchmarking uses the expertise of engineers who have been working in the field for years, it may be a good approach for many measures, particularly given the large number of measures across all EE programs. However, past work can be improved upon through the use of more systemized approaches, such as a Delphi-type of analysis.[15]
- Although the billing analyses method addresses the issue of persistence most comprehensively, there are cautions to consider. The effect may be small, which will require large sample sizes. Also, it may be difficult to control for other factors outside the program that cause changes in energy use across a five- or 10-year period. Where quality data exist, a billing analysis is a good method for assessing persistence, but it requires an appropriate data platform for it to be reliable.[16]

3. ***It is important to be open to the new methods and approaches being developed.*** Specifically, a panel of participants established at the time of program participation could be used in cross-sectional, time-series models. This involves incorporating the evaluation of persistence in program design and implementation planning. This type of forward thinking will make persistence easier to address, particularly in near-term years when it is most important.[17]
4. ***Certain types of persistence studies, particularly database/benchmarking approaches, might best be addressed on a regional basis that includes numerous specific programs.*** Assessing persistence across a number of regional programs can provide information on the influence of program design on persistence, which might not be found using a series of program-specific studies. In identifying these regional opportunities, it is important to consider the influence of program design on persistence. (For example, in the study Nexus performed across New England in 2008, program-specific elements had a large influence on the persistence of lighting measures.)

---

[15] Skumatz (2012) presents a number of ways these studies can be improved, including the use of Delphi approaches. An expert-panel approach was used in an evaluation of the Northwest Energy Efficiency Alliance's market transformation programs by Violette and Cooney (2003).

[16] Billing data analyses that try to estimate small effects reliably (e.g., 2% savings) without the required sample sizes and accurate data for the independent variables (i.e., little measurement error) have often not been successful. Quantum (1998) discusses this issue in the context of using a billing analysis to assess persistence for new home construction.

[17] Panel data methods are suggested as a potential approach in both Skumatz (2012) and Nexus (2008).

2 Other Evaluation Issues

This section briefly addresses these evaluation issues: (1) synergy; (2) errors in variables, measurement error, and program tracking; (3) dual baselines; and (4) rebound.

2.1 Addressing Synergies Across Programs

Evaluators are often asked about potential synergies across programs. For example, certain information programs may result in direct savings impacts, but the programs may also be designed to lead participants into other programs. In addition, there may be effects across programs. For example, a whole-house retrofit program may influence the uptake of measures offered in other residential programs. These synergies are useful for designing programs and portfolios. Synergies that increase the overall savings from a portfolio of programs are valuable even if one specific program has lower savings due to these synergies.

The industry practice is to use approximate information to assess the relative importance of synergies. Even this level of analysis has generally been limited in evaluations. However, useful information on synergies can be developed by having evaluators do the following:

1. Identify what they believe may be positive and negative synergies (i.e., direction); and
2. Determine the rough magnitude of these synergies by benchmarking them as a fraction of the programs' savings.

With this material, portfolio models designed to assess the importance of synergies can produce information useful for assessing investments in EE and future program/portfolio designs.[18]

2.1.1 Conclusion

At the present time, the state-of-the practice involves identifying and assessing the potential importance of specific synergies across programs, although this is not always requested of evaluators. If assessing synergies becomes part of an evaluator's reporting requirements, the evaluator could modify surveys to provide useful information on potentially important EE program design considerations.[19]

---

[18] This approach does not have to be information intensive in terms of developing useful data for analyzing synergies and benchmarking their magnitude. Two pieces of information are needed: (1) an estimated range of effects, e.g., from 5% of program savings to 20% of program savings; and (2) an estimate of where the most likely value falls within this range. Based on these three points—the lower bound, the upper bound and an estimate of where within this range the most likely value falls— Monte Carlo methods can be used to test the importance and sensitivity of program impacts to identified synergies using Excel-based tools. An example of this range-based method can be found in Violette and Cooney (2003), and a version of this method is discussed in EPRI (2010, p. 5-4). This information can be used by the program administrator to inform the design of future EE portfolios.

[19] One reviewer of this chapter pointed out the potential complexities of determining program-specific synergies and their direction "...to the extent that synergies are increasingly observed or acknowledged, policies regarding the use of individual program cost-benefit analysis results for justifying the retention of programs may need to be changed in favor of portfolio level benefit cost analyses." This section was not intended to delve into benefit-cost methods. However, increased attention on synergies across programs is likely to prove useful. Monte-Carlo

2.2 Errors in Variables, Measurement Errors, and Tracking Systems

This section outlines the issues of errors in the input variables to an energy savings calculation. Such errors could be caused by an incorrect engineering calculation or by inaccurate values of the independent variables used in the regression analyses.

It is important that evaluators consider the accuracy of the input data and use the best quality data possible. In this context, data accuracy issues include data that are unbiased on average, but are subject to measurement error. Biased data clearly poses issues for any analysis; however, measurement error in itself poses challenges for evaluation. This is true even when the measurement error may be uncorrelated with the magnitude of the value of the variable, and the error may be equally distributed above and below the true value.

Program implementers need to be aware that the designs of the data tracking system and the data collection processes have a substantial influence on the accuracy and reliability of data. In turn, the accuracy and completeness of the data influence the estimated realization rates and the ability to achieve the target levels of confidence in these estimates.

While errors in variables can bias the evaluation results either up or down, there are several practical factors in EE evaluations that tend to result in lower realization rates and lower savings estimates. A typical realization rate study uses information from the tracking system to verify that the equipment is in place, working as expected, and achieving the energy savings predicted in the tracking system. Tracking system errors can include not properly recording the site location, contact information, equipment information, location where the equipment is installed, and the operating conditions of the equipment. This will make any associated field verification more difficult and the variance around the realization rate greater.

Different data issues will have different impacts on the estimates; however, improved data quality will usually decrease the variance of the realization rate estimate and increase confidence and precision. When stakeholders have set high target confidence-and-precision levels, it is important to track accurately the essential data (such as the installed measures' location, size, model number, date, contact person) required to produce the initial tracking system estimate of savings at that site.

The issue of errors in variables and measurement error can be important.

- Kennedy (2003) states: "Many economists feel that the greatest drawback to econometrics is the fact that the data with which econometricians work with are so poor."
- Similarly, Chen et al. (2007) states: "The problem of measurement errors is one of the most fundamental problems in empirical economics. The presence of measurement errors causes biased and inconsistent parameter estimates and leads to erroneous conclusions to various degrees in economic analysis."

---

models that use different scenarios regarding the magnitude and direction of synergies can help assess the robustness of program and portfolio cost-effectiveness.

Errors in measuring the dependent variable of a regression equation are incorporated in the equation's error term and are not a problem. The issue is with errors in measuring the independent variables used in a regression model. This violates the fixed independent variables assumption of classical linear regression models: the independent variable is now a stochastic variable.[20] A good source for approaches to address the errors-in-variables issue is Chapter 9 in Kennedy (2003).

The program tracking system data used in regression analyses can be a source of potential issues. For example, the inability to track customer participation in multiple programs can cause a number of problems. In these instances, data can be very accurate at the program level, but there is no mechanism to ascertain the effects of participating in multiple programs. For example, if a billing analysis is being conducted of a high-efficiency residential HVAC replacement program but the tracking system is not linked to the residential audit and weatherization program that feeds participants into the HVAC program, this will cause bias. When customers first participate in a feeder program but that information is not conveyed in the tracking system used by the HVAC evaluator, then the HVAC program's savings analysis will be biased, most likely on the low side.

Another well-known errors-in-variables issue relates to models that use aggregate data on DSM expenditures and energy consumption in analyzing the relationship between expenditures on EE activities and changes in energy use.[21] Developing the appropriate datasets poses challenges. For example, Rivers and Jaccard (2011) note that:

> [O]ur data on demand side management expenditures include all demand side management—in particular it includes both load management expenditures as well as energy efficiency expenditures. Since load management expenditures are not aimed at curtailing electricity demand explicitly… (p. 113).

The report then states that they do not believe this is a problem since

> ...utilities that were able to provide us with data (as well as in US utilities), load management expenditures amounted to less than 25% of the total, so error in our estimates should not be too severe, and in particular should not change the nature of our conclusions.

The authors may be correct, but their assessment was based on judgment with little real analysis of the degree of the issue.

The work by Rivers and Jaccard (2011) and by Arimura et al. (2011) illustrates the degree of effort often required to develop a useful set of aggregate state/province-level data or utility-level DSM. Using the Energy Information Administration forms, Arimura states: "The original data

---

20 The assumption is that observations of the independent variable can be considered fixed in repeated samples (i.e., that it is possible to repeat the sample with the same independent variable values; Kennedy, 2003, p.49).

21 Two recent publications with examples of this are Rivers and Jaccard (2011) and Arimura et al. (2011).

set has many observations with missing values for DSM spending, even after our meticulous efforts to find them from various sources."[22]

Another issue concerns the fact that numerous states have both utility and third-party program providers, which complicates the development of data that can be used to examine the relationship between utility EE program expenditures and aggregate energy consumption.

Attenuation bias is a potential issue when there is measurement error in the independent variables used in regression analyses. Simply stated, the implications are these: (1) more noise in the data due to measurement errors will make it more difficult to find significant impacts, and (2) those impacts will tend to be biased downwards.[23]

Attenuation bias can be an important problem in regression models using independent variables that might have large amounts of measurement errors due to:

- Differences in reporting of values in databases compiled across utilities, or
- Assignment/allocation of values at a utility service territory level down to a county level to create more observations.

Chen et al. (2007, 2011) and Satorra (2008) present a graphical example of this bias using a measurement error model developed for a simple one-variable regression.

- Using the model $Y = \beta X + e$ and
- having X measured with error,
- the measurement error model $X = x + u$, with x uncorrelated with u, $\mathrm{var}(X) = \mathrm{var}(x) + \mathrm{var}(u)$ can be used to assess the reliability of the estimated coefficient.

The reliability of X is defined as $rel = 1 - \mathrm{var}(u)/\mathrm{var}(X)$ (which results in a number between 0 and 1).

Satorra performed a set of simulations for a sample size equal to 10 and used different values for the reliability of the regressor X: 1 (accurate), 0.86, 0.61, and 0.50 (considerable measurement error).

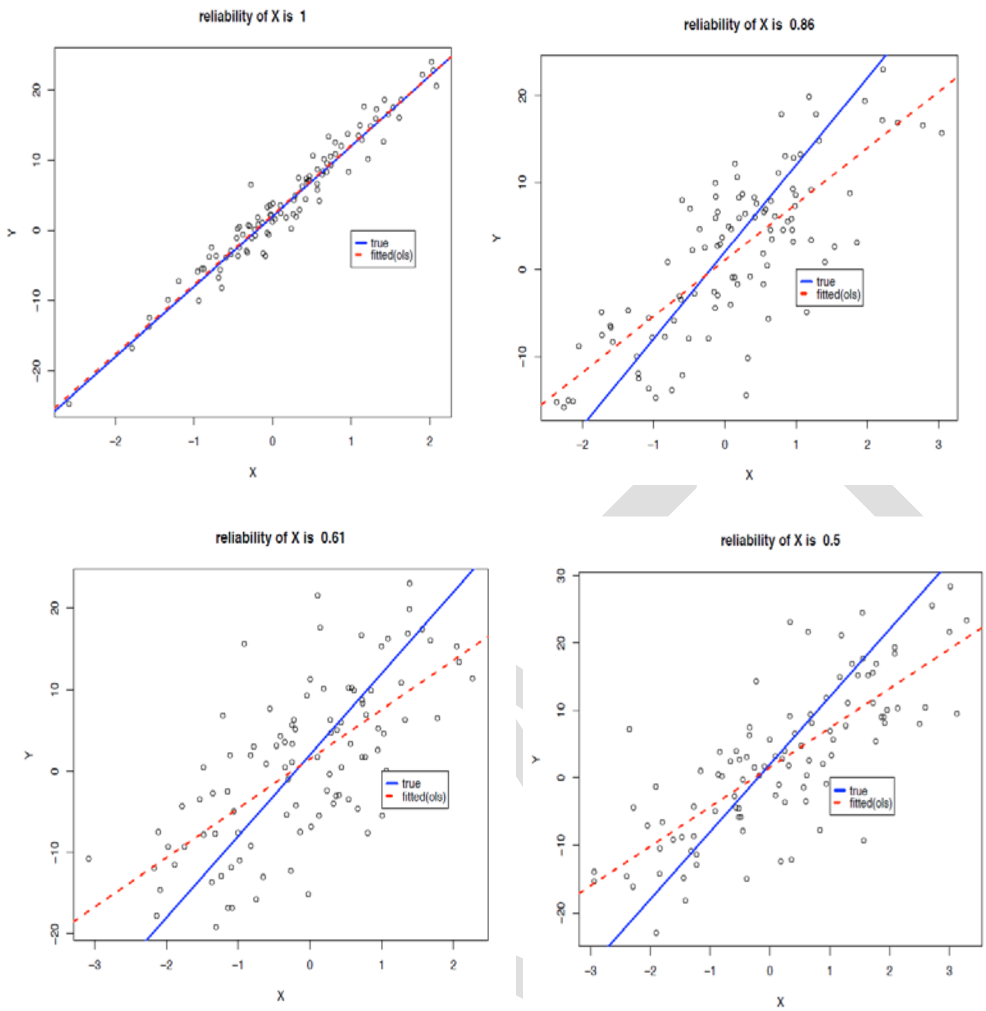
Each simulation is shown in Figure 5.

---

22 See footnotes 15, 16 and 17 in Arimura et al. (2011) for a discussion of the challenges they addressed in developing values of the key variables (i.e., the utility's EE expenditures that could explain changes in energy use and be used to assess cost-effectiveness in terms of cost per kWh saved).

23 This is not a new problem. Chen (2007 and 2011, p. 901) discusses how one of the most famous studies in economics had to address attenuation bias. In his famous book *A Theory of the Consumption Function*, Milton Friedman (1957) shows that, because of the attenuation bias, the estimated influence of income on consumption would be underestimated.

Figure 1. Satorra (2008) Simulation Results



As shown in Figure 5, the bias in the coefficient increases as the reliability of X decreases (that is, measurement error increases), even if this measurement error is uncorrelated with the variance of X. The slope of the coefficient declines as the reliability of X declines. This represents the attenuation bias associated with measurement error.

2.2.1 Conclusion

Issues associated with measurement error are often unavoidable in applied regression analysis. On occasion, data collected for one purpose with one level of accuracy may be used as a variable in a model testing for different types of effects. The solution is to reduce measurement error in the independent variables (the regressors) as much as possible.

Errors in variables, measurement errors, and general issues with data in tracking systems will make it more difficult for the evaluator to identify energy savings at a desired level of confidence. Kennedy (2003) states, “In the spirit of fragility analysis, econometricians should report a range of estimates corresponding to a range of values of measurement variance.” Kennedy presents examples of how this can be accomplished, but this extra effort is best reserved for large-scale efforts, and it goes beyond current industry standard practice in EE evaluation.

Nevertheless, having a good data platform from which EE savings are evaluated is important and needs more emphasis in practical evaluation work.

2.3 Dual Baselines

There are several evaluation issues caused by changes—during the lifetime of that measure—in the baseline against which savings are estimated. One issue, "remaining useful life" (RUL), occurs when a program is focused on replacing existing (lower-efficiency) equipment with EE equipment before the old equipment ceases to function *or* before it would otherwise have been replaced.

- The savings could be calculated simply as the difference between energy use for the replaced measure and the new EE measure; or
- The savings could be based on the difference between the new standard measures available in the market as compared to the new EE measure.

These savings would be constant for the assumed life of the measure—that is, no adjusted baseline for that measure is considered for the period after the RUL.

In theory, the use of two baselines can be argued to be the appropriate approach in certain applications. The baseline for the replaced low-efficiency measures that still had useful life would be the difference in efficiency between the replaced measure and the high-efficiency measure for the RUL of the replaced measure. For the period *after* the replaced measure's RUL, the baseline should shift to the difference between the installed high-efficiency equipment and the currently available standard equipment. (This would be the baseline for the balance of the assumed life of the new high-efficiency measure.) In practice, this is not often done. (See the conclusions for this section).

A similar situation occurs when a replacement is made of equipment that has a measure life spanning a point when a new code requires higher-efficiency equipment. In this case, evaluators must decide whether the baseline should be the efficiency of the equipment replaced and, in that event, change to a new baseline after the new code or standard is adopted. In general, the working assumption is that the baseline should reflect the energy use of the replaced equipment. If, however, that equipment would have been replaced within a few years by new equipment that meets the new code, then there is a question about whether the baseline should shift.

2.3.1 Conclusion

These dual baseline questions are beginning to receive more attention. Two opinions are expressed in the literature.

- The first and most common is that the complexity and uncertainty entailed in estimate the RULs of the equipment being replaced are excessive compared to their effects on energy savings calculations.
- The second opinion is that dual baseline the issues are important to address for some certain select measures, such as lighting, where the impacts may be large.

These dual-baseline issues have been addressed in some program evaluations, but have not generally been viewed as important for overall EE program evaluation because of their complexity and uncertainty regarding customer actions. However, the topic of dual baselines deserves more research to assess those specific situations in which accounting for the two baselines might have a substantive effect on energy savings.

2.4 Rebound Effects

Rebound occurs when the costs of using energy are reduced due to EE programs. When families spend less money to cool their home in the summer because of more efficient equipment, they might change their temperature setpoint to increase their comfort and their energy use.

Rebound is discussed in the literature according to the following two types:

- ***Type 1: Rebound is used essentially synonymous with take-back*** and happens at the participant level. It involves the question of whether participants who experience lower costs for energy because of an EE program measure—such as the installation of a high-efficiency air conditioner) then "take back" some of those savings by using more energy.[24]
- ***Type 2: Rebound takes place in the larger economy.*** EE programs have reduced the cost of energy across a number of uses, stimulating the development and use of energy-using equipment.

With the exception of low-income programs, Type 1 rebound has not been found to be significant in most EE program evaluations.[25] When consumers match marginal benefits with marginal costs, the concepts of bounded rationality and compartmentalized decision-making are being recognized as one theory of consumer behavior and decision-making.[26] (This is contrary to pure economic theory.) Consumers optimize, but only to the point when the complexity of the decision and the cost of the information become too high. For example, although the efficiency of an air conditioning (AC) unit varies daily with temperature and load; however, a consumer setting the thermostat on the AC unit is probably not going to examine the cost of running that unit each day and then adjust the thermostat accordingly.

Most customers set their thermostats at a comfortable level, regardless of whether they participate in an AC equipment program (whether for maintenance or new equipment) that increases the efficiency of the unit. In other words, consumers generally do not change their thermostat setting as a result of participating in an EE program.

---

24 A reviewer pointed out that, for many customers, the lower costs of energy are not reflected in the price of a kWh or a Therm of natural gas. Instead, customers use less energy, resulting in a lowering of their monthly bills. This results in customers spending less on energy over the course of a season or year.

25 This chapter is focused on EE programs. Take-back is more common in demand response and load management programs where AC units or other equipment are cycled to reduce peak demand for several hours on a few select days. This can result in a warming of the house or building, and the equipment automatically runs a bit more after the cycling event to return the temperature to the original set point. More efficient operational and cycling designs for AC load management programs can greatly reduce take-back, and take-back is a more common effect for event-based load management programs than for EE programs that influence all hours of a season.

26 The primary reference for this concept is Simon (1957), but it is also discussed in Kahneman (2003).

Low-income customers can be the exception, as they may change their thermostat setpoints for both AC and heating after participating in an EE program designed to increase the efficiency of the equipment. The change in energy price is more important to low-income customers, who may have been sacrificing comfort to meet their household budget before they participated in the EE program. Lowering the costs of AC and heating may allow them to set their thermostats at a level that provides more comfort, which may result in greater energy use for this participant segment. While this may cause an increase in the overall energy use for these low-income customers, it can provide a large welfare gain and even improved health and safety for low income customers.

Going beyond the program participants' actions, Type 2 rebound assesses the economy as a whole, as lowering the cost of energy through aggressive EE programs may make energy more economical for many new uses. There has been a recent resurgence of interest in this type of rebound, but a full analysis is beyond the scope of this chapter which focuses on EE program evaluation. (Gavankar and Geyer [2010] present a review of this larger rebound issue.) There is substantial literature on this economy-wide concept of rebound, and addressing most of the key theses in the discussion requires economy-wide models with energy as one of the inputs for the a wide variety of products and services.[27]

Searching on the terms "energy efficiency" and "rebound" results in many policy papers that present theses on how rebound may be an influence in the larger economy. The issue seems not to be economic welfare, but other policy goals. Using resources as efficiently and cost-effectively as possible always seems like a good policy, unless there is some other constraint. Reducing the cost of energy and allowing people to use energy in additional applications may increase overall welfare. Still, if the goal is to not increase energy use at all, then the downside of reducing energy costs may be concerns about carbon emissions. (It is not the purpose of this chapter, however, to detail this literature, other than noting it exists and offering some practical places to begin a review.

Using resources as efficiently as possible should be a good start towards any policy designed to reduce energy consumption that may contribute to carbon emissions. This policy could complement pricing and other policies designed to reduce energy use. Starting from a platform of efficient energy use should not hinder the applicability of other policies.

[27] The Rocky Mountain Institute has posted information at http://blog.rmi.org/blog_Jevons_Paradox discussing recent hypotheses about this type-two rebound effect. Other references are Tierney J. (2011), which presents the issue of rebound as being important, and a counterpoint paper by Afsah (2011).

3 References and Resources

Ahmad, M. , Deng, A., Spoor, S., Usabiaga, M., and I. Zhao (2011). *Persistence of Energy Savings in Industrial Retrocommissioning Projects*. ACEEE Summer Study on Energy Efficiency in Industry.

Allison, P.A. (1995).*Survival Analysis Using the SAS® System: A Practical Guide.* SAS® Institute.

Arimura, T.H., S. Li, R.G. Newell, and K. Palmer (2011). *Cost-Effectiveness of Electricity Energy Efficiency Programs.* National Bureau of Economic Research, NBER Working Paper No. 17556.

California Public Utilities Commission (2006). *California Energy Efficiency Evaluation Protocols: Technical, Methodological, and Reporting Requirements for Evaluation Professionals*. Oregon: TecMarket Works. Available online: http://www.calmac.org/publications/EvaluatorsProtocols_Final_AdoptedviaRuling_06-19-2006.pdf.

Chen, X., H. Hong, and D. Nekipelov (2007). Measurement Error Models. Prepared for the *Journal of Economic Literature*. Available online: http://www.stanford.edu/~doubleh/eco273B/survey-jan27chenhandenis-07.pdf.

Chen, X., H. Hong, and D. Nekipelov (2011). Nonlinear Models of Measurement Errors. *Journal of Economic Literature*, 49(4): 901–37.

Decision Sciences Research Associates (1999). *1994 Commercial CFL Manufacturers' Rebate Persistence Study.* Report ID 529D, Pasadena: Decision Sciences Research Associates.

Energy and Resource Solutions (2005). *Measure Life Study.* Prepared for The Massachusetts Joint Utilities, November.

EPRI (2010)*, Methodological approach for Estimating the Benefits and Costs of Smart Grid Demonstration Projects*. EPRI, Palo Alto, CA: 2010. 1020342.

Friedman, M. (1957). *A Theory of the Consumption Function*. Princeton University Press.

GDS Associates (2007). *Measure Life Report: Residential and Commercial/Industrial Lighting and HVAC Measures*. Prepared for The New England State Program Working Group for use as an Energy Efficiency Measures/Programs Reference Document for the ISO Forward Capacity Market.

Gavankar, S. and R. Geyer (2010). The Rebound Effect: State of the Debate and Implications for Energy Efficiency Research. Institute of Energy Efficiency (UCSB), June. Available online: http://iee.ucsb.edu/files/pdf/Rebound%20Report%20for%20IEE-UCSB_0.pdf.

Kahneman, D. (2003). Maps of Bounded Rationality: Psychology for Behavioral Economics. *The American Economic Review,* 93 (5): 1449–75.

KEMA (2004). *Residential Refrigerator Recycling Ninth Year Retention Study*. Study ID Nos. 546B, 563, Madison. 2004.

KEMA (2006). *2005 Smart Thermostat Program Impact Evaluation*. Final Report prepared for San Diego Gas and Electric Company.

KEMA (2009). *Focus on Energy Evaluation: Business Programs: Measure Life Study Final Report. P*repared for the Public Service Commission of Wisconsin, August 25.

Kennedy, P. (2003). *A Guide to Econometrics.* Chapter 9, MIT Press.

Navigant Consulting, Inc. (2010). *Kaizen Blitz Pilot. Report 1*. Prepared for Energy Trust of Oregon.

Navigant Consulting, Inc. (2011). *Evaluation Report: OPower SMUD Pilot Year 2*. Presented to OPower, February.

Nexus Market Research, Inc. (2008). *Residential Lighting Measure Life Study*. Cambridge: New England Residential Lighting Program.

Ontario Power Authority (2011). *EM&V Protocols and Requirement*. Ontario: 2011-2014.

Pacific Gas & Electric (1999). *Commercial Lighting Study*. CALMAC, pp. 4-14.

Peterson, G., R. deKieffer, J. Proctor, and T. Downey (1999). *Persistence #3A: An Assessment of Technical Degradation Factors: Commercial Air Conditioners and Energy Management Systems, Final Report*. CADMAC Report # 2028P.

Rocky Mountain Institute (2011). *Jevons Paradox: The Debate That Just Won't Die*. By Potts, M. and C. Burns.

Proctor Engineering (1998). *Statewide Measure Performance study #2: An Assessment of Relative Technical Degradation Rates. Final Report.* For California Measurement Advisory Committee.

Proctor Engineering (1999). *Summary Report of Persistence Studies: Assessments of Technical Degradation Factors, Final Report*. CADMAC Report #2030P.

Quantum Consulting (1998). *PG&E Statewide Multi-Year Billing Analysis Study: Commercial Lighting Technologies Final Report*. Berkeley.

Rivers, N. and M. Jaccard (2011). Electric Utility Demand Side Management in Canada*. The Energy Journal*, 32 (4), pp. 93-116.

RLW Analytics (1998). *SCE Non-Residential New Construction Persistence Study. Final Report.* For Southern California Edison. Study # SCE0064.01; 554; 530

San Diego Gas & Electric (1999). *Commercial Energy Efficiency Incentives Ninth Year Retention Evaluation.* 1996 and 1997.

Satorra, A. (2008). *Theory and Practice of Structural Equation Modeling, Department d'Economiai Empresa*. Barcelona. Available online: http://statmath.wu.ac.at/courses/TPStrucEqMod/errorinvariables.pdf.

Simon, H. (1957). *A Behavioral Model of Rational Choice in Models of Man, Social and Rational: Mathematical Essays on Rational Human Behavior in a Social Setting.* New York: Wiley.

Skumatz et al. (2009). *Lessons Learned and Next Steps in Energy Efficiency Measurement and Attribution: Energy Savings, Net to Gross, Non-Energy Benefits, and Persistence of Energy Efficiency Behavior.*

Skumatz, L. A. (2012). *What Makes a Good EUL? Analysis of Existing Estimates and Implications for New Protocols for Estimated Useful Lifetimes (EULs).* International Energy Program Evaluation Conference, Rome, Italy.

U.S. Environmental Protection Agency (2007). *Model Energy Efficiency Program Impact Evaluation Guide.*

Violette, D. and K. Cooney (2003), "Findings and Report: Retrospective Assessment Of The Northwest Energy Efficiency Alliance," Prepared for: Northwest Energy Efficiency Alliance Ad Hoc Retrospective Committee, Summit Blue/Navigant Consulting. December. Web: http://www.theboc.info/pdf/Eval-BOC_SummittBlue_NEEA_2003.pdf

3 Appendix A: (sent as a separate document – with edits couldn't get it in right)