A Peek Behind the Mask: An Approach to Revealing Hidden ECM Behavior in Commercial Buildings

Jim Perich-Anderson and Joseph A. Taffe, Tacoma Public Utilities

In conducting an impact evaluation of a retrofit commercial conservation program, Tacoma City Light staff determined that classic comparison group techniques could not be used effectively to estimate program savings. There were frequent differences between audit-predicted and meter-observed savings estimates even though post-installation visits to all participants indicated that the measures were correctly installed and operating properly. These visits also indicated that at half of the businesses, significant changes in energy use intensity apparently masked performance of the installed energy conservation measures. The establishment of a relevant comparison group to adjust for these behaviors was precluded by the small program sample, and a lack of sufficient data on non-treated businesses to establish a control group.

An alternative approach, the outlier savings adjustment technique, was developed to identify net program impacts. This technique assigns savings estimates for participants whose observed savings differ substantially from predicted savings. Specifically, for participants with differences outside a set range of pre-program consumption, savings were adjusted to the average savings of those within that range.

Application of the technique indicates that effective measure performance was being masked by building behavior independent of the installed measures. Application of the technique to a larger group of program participants suggests that it may be generally applicable as an estimator of overall program savings.

Introduction

With the increasing integration of energy conservation into utility resource portfolios comes the need to quantify program related energy savings at the utility level. The usual approach to measuring program savings attempts to derive the "true" or net impact of the program by subtracting total consumption of the treatment group in a fixed post-treatment period from consumption in a pre-treatment period of the same duration. To account for program independent behavior, this savings estimate is adjusted based on the percent difference in energy consumption of a comparison group between the same time periods.

Various pooled comparison group approaches have been used effectively in assessing aggregate savings from a regionally operated commercial retrofit program when relatively large sample and comparison groups were available (Coates, 1989; Cambridge Systematics, Inc. 1990). However, examination of the measure performance and business behavior of a subset of 15 commercial retrofit projects in the Tacoma Public Utilities service area led to the conclusion that these comparison group assessments had not reveal the true performance of the 15 projects. This led to the development of an alternative approach, an outlier savings adjustment technique, to assessing program impact.

This paper discusses the applicability of the comparison group approach in assessing small programs, presents an alternative approach, and presents the results of its application to the initial 15 projects and to a larger set of 76 projects from the same conservation program.

By Comparison

The comparison approach to assessing program impact, as described in the introduction, has proven useful and acceptable for many applications. Billing histories provide inexpensive and accessible monthly consumption totals for buildings in treatment and comparison populations. They are good indicators of overall program performance if building types, treatment type, weather impacts and occupant activity are known, and can be controlled for based on the size and/or similarity of both the program and treatment groups.

However, when these conditions are not met, the comparison approach may provide misleading estimates of program effectiveness. If there are relatively few of any type or size of building participating in a program, the confidence with which a comparison group can be used to adjust consumption to indicate program performance is diminished. Likewise, if there is a range of measures offered as part of a conservation program, not all of which are applicable or required of all participants, it is difficult to match a comparison group to the participants.

This difficulty is becoming more evident as utilities begin assessing conservation programs with participants that vary widely in size and intensity of energy use and adopt treatments which range from minor operating changes to major redesigns of equipment. As the Cambridge study notes, "It is very difficult to identify exact matches between a control sample and a participant sample on a building-by-building basis, because there will almost always be some difference that can be identified (pg. 2-10)".

Commercial Conservation

The limitations of the comparison approach are particularly evident in assessing the performance of commercial buildings. Commercial conservation measures often focus on a small portion of a buildings energy use while billing histories reflect energy use in whole buildings or building subsections. At this level, there is no distinction between measure behavior and changes in energy use intensity attributable to business hour changes, equipment additions, changes in tenancy, economic conditions, weather and other factors. As a result, changes in energy consumption behavior unrelated to measures can mask measure-related change in consumption.

The idiosyncratic nature of commercial businesses, and lack of information about these idiosyncrasies, limits the ability to identify a comparison group of buildings. As noted in Evaluation of Commercial Incentives Pilot Program (Cambridge Systematics 1990), "Control groups should be designed to represent the participant sample in the aggregate, and all comparisons should be made at the aggregate level. Additionally, a control group should be as large as reasonably possible for the evaluation so that any anomalies in the factors being controlled for that occur in the control sample will be averaged out."

The Adjustment Process

The masked performance phenomenon was addressed through a strategy for identifying and treating those participants whose observed savings, pre-installation minus post-installation consumption, differed substantially from predicted savings in either direction. Our approach to the problem was to use an indicator of deviation which incorporated gross savings (pre minus post usage) and predicted savings. Deviation was calculated as the difference between predicted and observed savings compared to total pre-program consumption. In order to treat over-performers and under-performers equally, the sign of the deviation was ignored.

This indicator was same as that used an earlier assessment of masked performance in CIPP projects in the Tacoma City Light Service area (Lerman and Perich-Anderson 1991).

Symbolically, the computation of the deviation can be expressed as follows:

$$\frac{[\text{Gross Savings (kWh)} - \text{Predicted Savings (kWh)}]}{\text{Pre-audit consumption (kWh)}}$$

The resulting variable multiplied times 100 yields a deviation from predicted savings expressed as a percentage of pre-audit consumption. This normalized the differences by the magnitude of consumption and permitted us to compare the performance of both large and small project without the distortions inherent in comparisons of absolute variations from predicted or calculated savings. Poor performers, those showing low savings or even increases in post-installation consumption, were afforded the same treatment as over-performers, those with savings far exceeding audit-based predictions.

A threshold of 15% deviation from predicted savings was established. Participants with differences in savings greater than 15% of pre-program consumption and who gave evidence of successful installation and maintenance of conservation measures were defined as outliers and were assigned savings based on the average savings percentage of participants with differences in savings of less than 15%. Projects which, on the basis of follow-up visits and interviews, demonstrated substantial changes in measure installation or operation were dropped from the outlier adjustment process. Both the initial and the expanded sample each had one project dropped from analysis for this reason.

Assumptions

The selection of 15% as the outlier threshold was based on several informed assumptions about building behavior. The first assumption is that project savings estimates are an accurate representation of measure performance. They are produced by trained analysts prior to installation and are based on an accurate accounting of the type, cost and efficiency of the installed measures. These estimates provide the best available information on measure performance and are the product of multiple layers of professional effort from the manufacturer, testing laboratories, energy analysts, engineers and evaluators. Each of the project estimates were subject to a rigorous review prior to project approval. Furthermore, each project was inspected by multiple, trained inspectors to assure correct measure installation. In this structured, estimating environment performance which varies widely from estimates indicates more than estimating error.

The second assumption is that no significant measure related change occurred in the program buildings in the year following treatment. This is based on limited information from CIPP Building Follow-on Worksheets which addressed not only equipment modifications but also such issues as hours of occupancy and HVAC setpoints. This builds on the detailed inspection of the installation at project completion to assure the factors of the installation are as the estimator expected. Significant changes include changes in measure controls, removal of measures, and other factors which could render the measures ineffective. Information on significant measure changes is based on site visits and interviews with building operators approximately one year after installation.

The third assumption is that variation in consumption in commercial buildings is to be expected, and will occur over time as influenced by normal fluctuation in business practices. This includes increased or decreased operating hours, output, occupancy, incidental plug load, and power rates. This variation can mask overperformance (shortened payback period) as well as underperformance (lengthened payback period). This phenomenon is borne out in follow-on interviews and inspections at Tacoma City Light and other utilities (Perich-Anderson and Lerman 1991; Cambridge Systematics 1990; Synergic Resources Corporation 1991).

Results

The 14 participants whose examination led to the development of the outlier adjustment technique ranged in pre-treatment annual consumption from 46,839 kWh to 1,120,800 kWh with an average of 216,421 kWh. Their predicted savings ranged from 12.5% to 65.1% with and average of 30.2%. The differences between their predicted and observed savings ranged from 1.3% to 45.8% with an average of 12.0%. For the sample 7 of the 14 or 50% exceeded the 15% deviation threshold. Table 1 illustrates these participants and Figure 1 is a graphic representation of their deviation percentages plotted from smallest to largest deviation.

Application of the outlier adjustment technique increased the program savings estimate from 596,280 kWh to 675,899 kWh which indicates that 79,619 kWh of savings were possibly being masked. The increase in estimated program savings was from 19.7% to 22.3%.

Further Evaluation

While we found the approach useful in reviewing a limited number of buildings, we felt it necessary to conduct additional testing of the technique on a larger group of buildings. The intent was to examine whether the pattern of difference between predicted and observed consumption for a larger sample of buildings was similar, and test the applicability of the approach on a larger group of buildings to determine if results similar to those predicated on the initial 14 buildings would result from the analysis of the larger program population.

Data were acquired for additional projects operated under the same commercial retrofit program. Of these, 76 had consumption histories for a minimum of one year prior to and one year following the treatment date. Additional information on measure type and cost, and limited follow-on data were included for each project. One project documented as having ceased heating with a large electric boiler in the post treatment period was dropped from additional analysis for a final group of 75 buildings.

The 75 projects that were examined using the outlier adjustment technique ranged in pre-treatment annual consumption from 34,780 kWh to 9,518,400 kWh with an average of 746,146 kWh which is indicative of larger buildings participating in the program than those in the initial analysis. Their predicted savings ranged from 2.1% to 72.8% with an average of 14.6%. The differences between their predicted and observed savings ranged from 0.1% to 97.7% with an average of 16.3%. For this larger sample, 36 of the 75 or 48% exceeded the 15% deviation threshold. Application of the outlier adjustment technique to this group of 75 buildings indicated 3,015,668 kWh of savings were being masked. Program savings increased from 1.4% to 6.8%. Table 2 illustrate details of the outlier adjustment of these projects.

Additional Analysis

Additional analysis of the larger group of 75 buildings was performed to see if the differences between predicted and observed savings was a function of the size of

Table 1. Outlier Adjustment of Savings Estimates - 14 Tacoma Public Utilities Commercial Conservation Projects

(a) Project Number	(b) Annual Consumption (kWh/yr)	(c) Predicted Savings (kWh/yr)	(d) Observed Savings (kWh/yr)	(e) Savings Difference [c-d]/b	(f) Adjusted Savings (kWh/yr)	(g) Adjusted Savings (f/b)
TI-010	1,120,800	235,517	221,189	1.3%	221,189	19.7%
TI-006	79,320	24,537	22,484	2.6%	22,484	28.3%
TI-005	145,640	18,154	28,927	7.4%	28,927	19.9%
TI-007	110,302	51,447	39,415	10.9%	39,415	35.7%
TI-001	268,579	23,812	64,973	12.9%	64,973	37.1%
TI-016	83,372	22,771	34,190	13.5%	34,190	40.5%
TI-002	87,600	23,812	11,913	13.6%	11,913	13.6%
TI-013	14,560	41,339	17,937	16.1%	32,480	22.3%
TI-012	115,583	38,073	18,279	17.1%	25,784	22.3%
TI-004	255,416	102,490	54,036	19.0%	56,978	22.3%
TI-003	359,100	142,681	63,370	22.1%	80,107	22.3%
TI-015	46,839	18,520	4,964	28.9%	10,449	22.3%
TI-014	96,967	23,489	-7,326	31.8%	21,631	22.3%
TI-008	113,768	74,073	21,928	45.8%	25,379	22.3%
Averages	216,421	65,464	42,591	12.0%	47,289	22.3%

pre-project consumption. Was it easier to estimate the performance of a large consumer or a small consumer?

The 74 projects were divided into three groups with annual consumption of roughly up to 250,000 kWh, 250,000 kWh to 500,000 kWh, or over 500,000 kWh. This analysis indicated that for the 25 largest, the 25 midsized, and the 25 smallest projects, the distribution of these differences was essentially the same. Figure 2 illustrates this distribution.

Refinements

As noted above, the 15% outlier threshold used in both analyses was chosen based on informed assumptions about commercial buildings and conservation programs. Even as such, it was chosen arbitrarily as a reasonable range of performance to expect.

Intuitively it would be more satisfying if the outlier threshold was somehow statistically derived from the known data regarding the cohort to which it was applied. Several such thresholds were investigated. First was the use of the average deviation as defined for the outlier adjustment technique. Then the average predicted savings, the standard deviation of the outlier deviations and several other thresholds were explored. Finally, an analysis of all thresholds ranging from 1% to 30% was performed.

This analysis, which is graphically represented in Figure 3, indicated that when the outlier adjustment technique is applied to the 75 buildings the adjusted program savings estimates were essentially the same for all thresholds above ten per cent.

Summary

Average per cent savings of the larger group were considerably lower than that of the initial group. This is probably an indicator of larger buildings in the larger group and therefore projects which impacted less of the total building consumption. This is, however, an important distinction to make, as it may indicate a disparity in program performance and presents a challenge to applying the outlier approach. For example, project outliers in the initial study group were deemed to a savings of 22.3%, whereas if grouped with the buildings in the second sample, savings from the same projects outliers would be deemed to approximately 6.8%. Populations to which the outlier adjustment technique is applied must be true

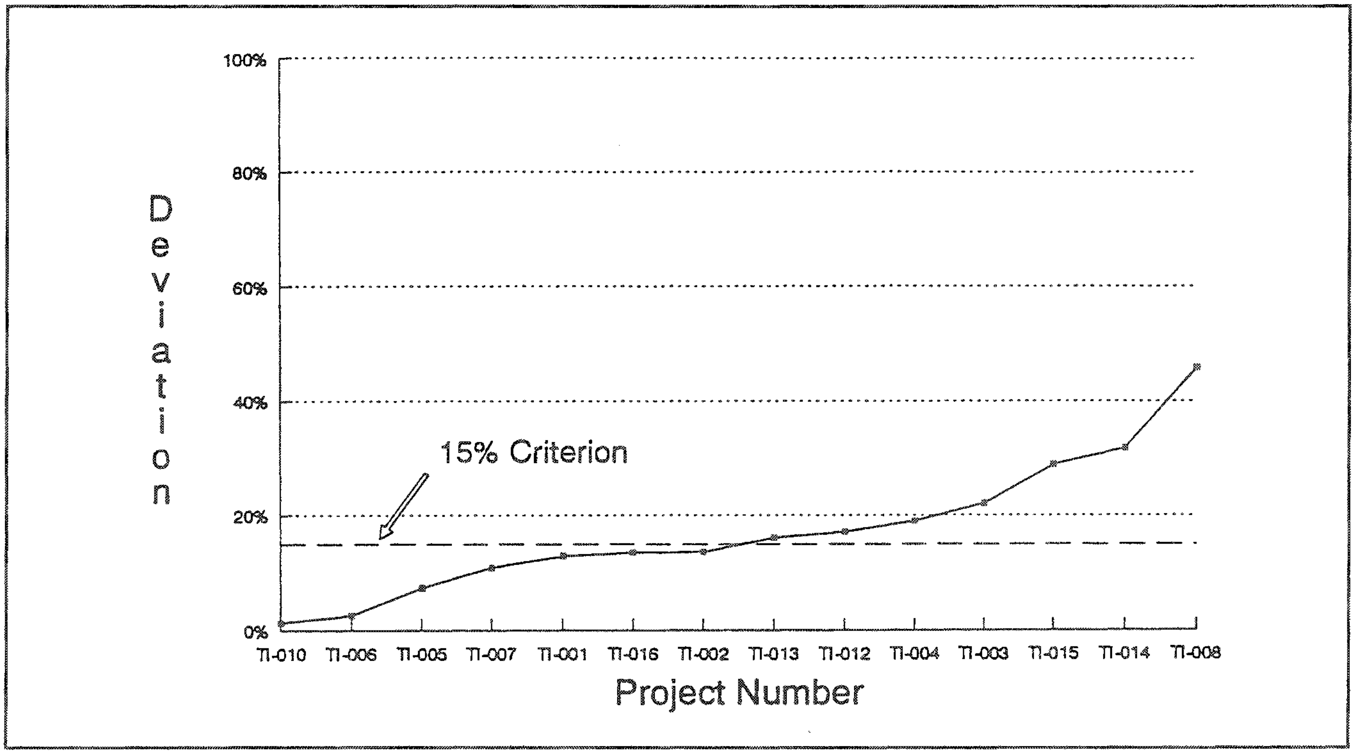

Figure 1. Outlier Analysis

cohorts, having gone through the same analysis and installation experience.

For both the smaller and larger sample, the absolute difference between observed and predicted savings was dispersed rather evenly around the average. The magnitude in difference ranged from 1.3% to 45.8% in the smaller sample and from 0.1% to 97.7% in the larger sample.

When the real difference in savings is taken into account for the larger study group, 63 of the 75 projects (84%) showed observed savings to be less than the predicted. 12 of the 14 projects (86%) in the smaller study group also had observed savings which were less than predicted. On the basis of the follow-on data, we feel that masking of ECM behavior typically appears as increased consumption in the post period.

The overall estimate of difference between total predicted and observed savings was 364,597 kWh or 12% for the small group and 9,134,198 kWh or 16.3% for the large group. If a similar array of measures was installed in each group, the difference may indicate significantly different impacts of exogenous variables in the two groups, such as business intensity, installation, engineering estimation approach, or audit procedures.

The variation between observed and predicted savings was evenly distributed in both samples. Level of variation between observed and savings appears to be independent of initial project size or estimated savings. These savings support the concept of a consistent band of acceptable variation across projects.

The analysis of deviation, in addition to being useful as a tool for calculating savings estimates, may also serve as a 'flag' for additional assessment of program effectiveness at sites where overperformance or underperformance is apparent.

The difference in apparent savings between the initial group of buildings, which were serviced by a single utility, and the expanded group, which were serviced by a several utilities, points to a liability in using the approach to evaluate performance across programs.

Conclusions

In a structured conservation program using trained energy auditors, analysts, engineers, installers and inspectors, projects which, based on billing histories, save significantly more or significantly less than predicted do not reflect measure-related consumption alone. Within a program group, these projects are classified as outliers, as

Table 2. Outlier Adjustment of Savings Estimates - 75 Commercial Conservation Projects

(a) Project Number	(b) Preaudit Consumption (kWh/yr)	(c) Predicted Savings (kWh/yr)	(d) Observed Savings (kWh/yr)	(e) Per Cent Difference [c-d]/b	(f) Adjusted Savings (kWh/yr)	(g) Adjusted Savings (f/b)
35	300,160	49,410	49,120	0.1%	49,120	16.4%
32	196,499	46,823	45,799	0.5%	45,799	23.3%
50	9,518,400	428,315	360,000	0.7%	360,000	3.8%
49	174,640	53,144	55,280	1.2%	55,280	31.7%
72	151,370	24,235	26,220	1.3%	26,220	17.3%
15	180,149	30,488	33,764	1.8%	33,764	18.7%
8	190,160	47,375	51,760	2.3%	51,760	27.2%
46	148,855	46,849	43,496	2.3%	43,496	29.2%
55	273,600	53,057	46,680	2.3%	46,680	17.1%
18	567,000	48,256	31,800	2.9%	31,800	5.6%
28	1,065,360	35,771	1,200	3.2%	1,200	0.1%
26	1,709,280	231,928	170,880	3.6%	170,880	10.0%
11	2,256,750	161,126	74,250	3.8%	74,250	3.3%
12	343,017	39,380	25,688	4.0%	25,688	7.5%
3	322,440	38,608	54,840	5.0%	54,840	17.0%
9	334,080	113,871	95,760	5.4%	95,760	28.7%
53	3,116,400	363,308	176,400	6.0%	176,400	5.7%
41	2,317,800	279,536	426,600	6.3%	426,600	18.4%
63	224,489	65,496	49,862	7.0%	49,862	22.2%
37	1,238,351	135,563	232,951	7.9%	232,951	18.8%
59	1,002,400	259,842	180,400	7.9%	180,400	18.0%
30	284,960	61,349	38,000	8.2%	38,000	13.3%
42	243,840	30,827	10,800	8.2%	10,000	4.4%
17	491,520	51,728	96,000	9.0%	96,000	19.5%
47	3,399,560	70,698	-242,040	9.2%	-242,040	-7.1%
10	259,230	43,963	18,763	9.7%	18,763	7.2%
60	241,520	61,358	36,960	10.1%	36,960	15.3%
51	63,768	22,146	28,826	10.5%	28,826	45.2%
58	365,280	51,711	90,000	10.5%	90,000	24.6%
6	632,640	56,735	-11,520	10.8%	-11,520	-1.8%
71	199,120	80,620	58,440	11.1%	58,440	29.3%

their electrical consumption lies outside the range of post treatment consumption that can reasonably be accepted as representative of the program's intervention into their energy use. Outliers are a challenge to program managers and evaluators and offer an opportunity to refine our understanding of energy conservation.

Each step of a conservation project holds the opportunity for error, but do errors in prediction of savings, measure installation and measure performance explain significant differences between predicted and observed savings? Based on data obtained from follow-up visits and interviews we conclude that when projects save significantly less or more than predicted it is the result of activities unrelated to the conservation program. The true savings for these projects are masked by the impact of business behaviors which are independent of the installed energy conservation measures.

Table 2. (contd)

(a) Project Number	(b) Preaudit Consumption (kWh/yr)	(c) Predicted Savings (kWh/yr)	(d) Observed Savings (kWh/yr)	(e) Per Cent Difference [c-d]/b	(f) Adjusted Savings (kWh/yr)	(g) (f/b)
25	1,878,000	483,486	270,600	11.3%	270,600	14.4%
22	386,160	127,347	83,280	11.4%	83,280	21.6%
62	294,560	38,360	4,840	11.4%	4,840	1.6%
19	3,367,850	276,571	-119,830	11.8%	-119,830	-3.6%
14	182,065	35,144	57,837	12.5%	57,837	31.8%
34	1,015,520	67,751	-67,840	13.4%	-67,840	-6.7%
38	245,760	103,618	70,400	13.5%	70,400	28.6%
43	192,627	42,158	13,954	14.6%	13,954	7.2%
4	282,600	33,908	77,640	15.5%	19,104	6.8%
24	319,260	73,455	23,760	15.6%	21,582	6.8%
13	343,722	96,334	37,154	17.2%	23,235	6.8%
7	920,880	136,516	-27,720	17.8%	62,251	6.8%
16	84,480	29,449	14,440	17.8%	5,711	6.8%
36	86,880	19,703	4,000	18.1%	5,873	6.8%
48	756,960	48,966	-96,640	19.2%	51,170	6.8%
45	434,560	170,489	85,600	19.5%	29,376	6.8%
67	455,280	183,037	93,960	19.6%	30.777	6.8%
56	565,988	62,569	177,233	20.3%	38,260	6.8%
27	1,432,560	39,091	-252,480	20.4%	96,840	6.8%
31	637,920	339,942	206,080	21.0%	43,123	6.8%
57	314,451	42,028	118,559	24.3%	21,257	6.8%
5	386,501	20,228	-74,492	24.5%	26,127	6.8%
61	92,920	45,368	22,240	24.9%	6,2810	6.8%
75	358,800	91,900	-480	25.7%	24,255	6.8%
2	1,011,960	296,982	33,840	26.0%	68,408	6.8%
52	266,400	101,653	30,840	26.6%	18,008	6.8%
68	34,780	25,320	15,230	29.0%	2,351	6.8%
29	172,160	53,939	3,840	29.1%	11,638	6.8%
1	150,480	62,823	17,940	29.8%	10,172	6.8%
39	74,456	36,767	13,923	30.7%	5,033	6.8%
54	352,320	157,354	47,680	31.1%	23,817	6.8%

This masking can be accounted for by establishing a range of acceptable deviation from predicted savings, and adjusting the savings estimates of those lying outside that range. The ongoing challenge is defining the acceptable range of accuracy and the level to which outliers will be adjusted to.

If the quality of a given energy conservation program is high with regard to tracking buildings, measures, and measure behavior, a narrower band of acceptable performance may be justified. Savings from each outlier could be adjusted to their predicted savings. Concerns about quality control may call for a more conservative

Table 2. (contd)

(a) Project Number	(b) Preaudit Consumption (kWh/yr)	(c) Predicted Savings (kWh/yr)	(d) Observed Savings (kWh/yr)	(e) Per Cent Difference [c-d]/b	(f) Adjusted Savings (kWh/yr)	(g) Adjusted Savings (f/b)
40	668,000	152,439	-67,200	32.9%	45,156	6.8%
21	773,288	207,925	470,408	33.9%	52,274	6.8%
20	225,840	57,569	-19,800	34.3%	15,267	6.8%
73	396,630	199,820	57,670	35.8%	26,812	6.8%
44	78,275	26,580	-3,723	38.7%	5,291	6.8%
74	125,760	24,235	-30,080	43.2%	8,501	6.8%
23	445,500	195,999	-25,380	49.7%	30,116	6.8%
33	319,200	173,548	8,160	51.8%	21,578	6.8%
76	65,100	10,403	-23,828	52.6%	4,401	6.8%
65	261,760	109,666	-38,080	56.4%	17,695	6.8%
70	485,040	308,522	33,240	56.8%	32,788	6.8%
69	224,280	159,653	-10,440	75.8%	15,161	6.8%
64	2,980,800	97,861	-2,814,000	97.7%	201,500	6.8%
Averages	746,146	108,668	10,391	16.3%	50,552	6.8%

The 16.3% average per cent difference (savings deviation) is the average of the sum of the absolute differences divided by the total pre-audit consumption.

approach of establishing a wider band of acceptable performance and adjusting to the average percent savings for cases within the band. In either case the adjusted program savings will more truly represent the impact of the conservation program than relying on billing histories alone.

Acknowledgements

The authors would like to acknowledge David Lerman, Supervisor of Planning and Evaluation for the Tacoma City Light Energy Conservation Office, for his support in the development of this evaluation process.

We also acknowledge the Bonneville Power Administration's cooperation in providing participant performance and follow-on data.

References

Coates, B. 1990. "Evaluation of Energy Savings in Commercial Retrofit Programs." Seattle City Light, Seattle.

Cambridge Systematics, Inc. and ERC Environmental and Energy Services Co. 1990. Evaluation of the Commercial Incentives Pilot Program Impact Evaluation Report. Bonneville Power Administration, Portland, Oregon 1990.

Lerman, D., and J. Perich-Anderson. 1991. "Out of Control: An Alternative Approach to Outcomes Analysis of Commercial Sector Programs." Energy Program Evaluation: Uses, Methods, and Results - Proceedings from the 1991 International Energy Conservation Evaluation Conference, pp. 168-173.

Synergic Resources Corporation 1991. Energy Conservation Measure Life Study, Bonneville Power Administration, Portland, Oregon.

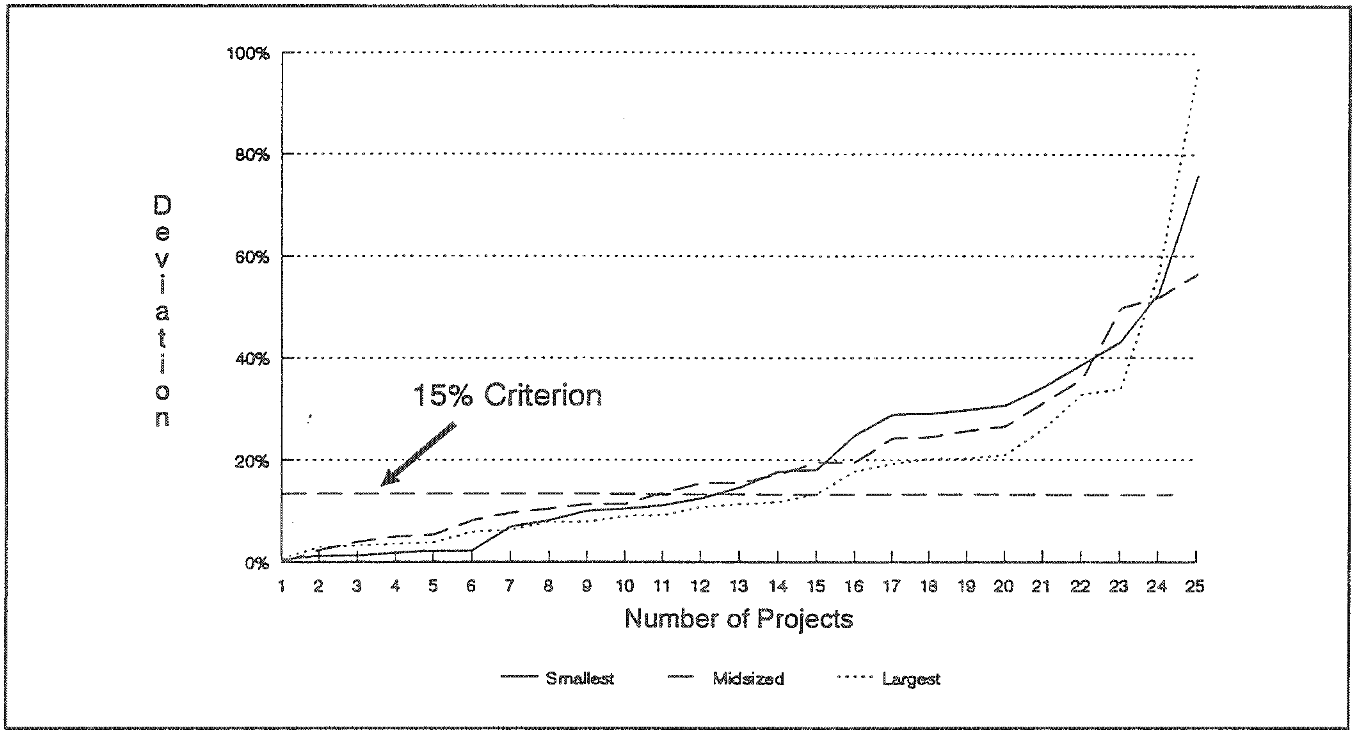

Figure 2. Consumption vs. Deviation

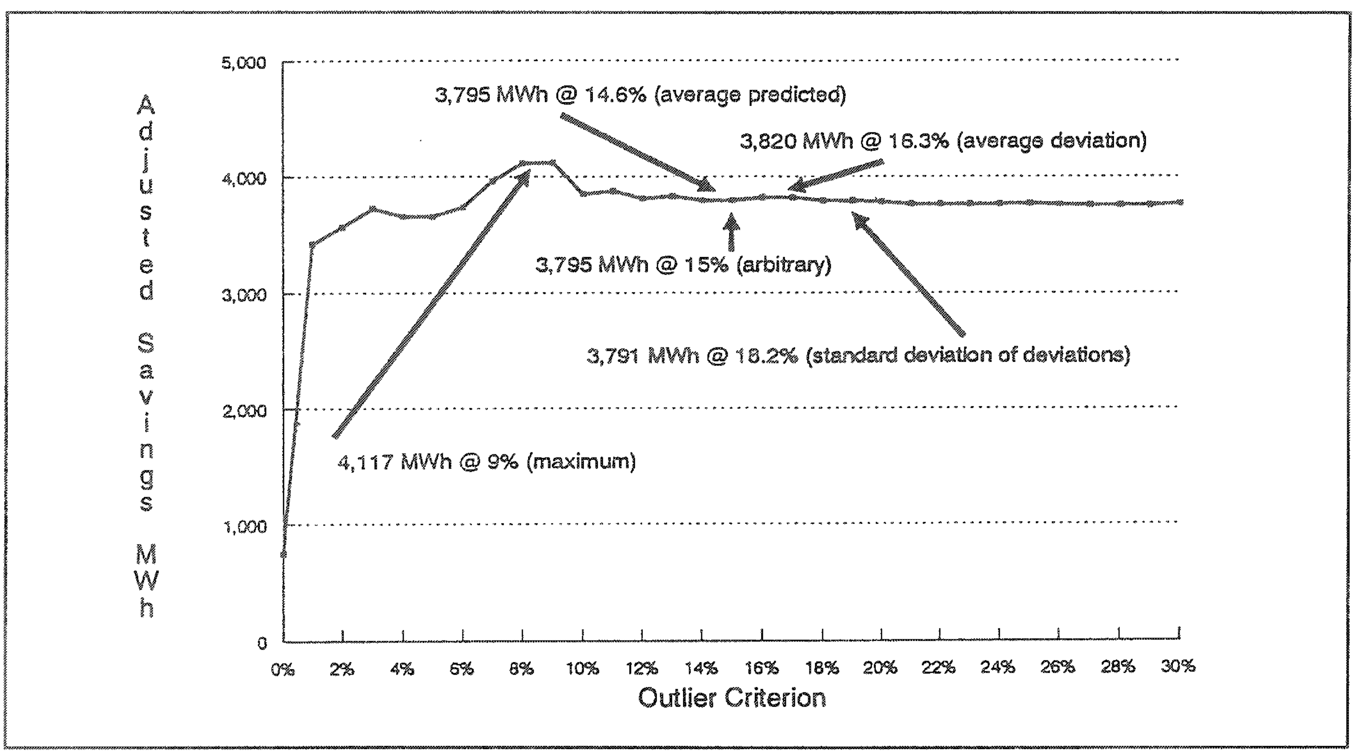

Figure 3. Threshold Sensitivity Analysis