AIM

Association for Information and Image Management

1100 Wayne Avenue, Suite 1100
Silver Spring, Maryland 20910
301/587-8202

Centimeter

Inches

MANUFACTURED TO AIM STANDARDS
BY APPLIED IMAGE, INC.

1 of 1

**Extension to Phase 2 Energy Savings Evaluation of the Commercial
Direct Investment Program**

David Trumble
Michael MacDonald

August 1994

Prepared for COM/Electric

Prepared by the
OAK RIDGE NATIONAL LABORATORY
Oak Ridge, Tennessee 37831
managed by
MARTIN MARIETTA ENERGY SYSTEMS, Inc.
for the
U.S. DEPARTMENT OF ENERGY
under contract DE-AC05-84OR21400

ACKNOWLEDGEMENTS

The authors would like to thank the staff of COM/Electric who helped provide information on the Commercial Direct Investment program. These include John Farley, Tina Torres, Robert Hardy, Jeff Luce, and their subcontractors. We would also like to thank Gil Peach, Jeff Schlegel, Scott Pigg, and Mike Blasnik for advice on the approach for the evaluation.

Table of Contents

List of Tables	iv
List of Figures	v
List of Equations	v
SUMMARY	vi
ABSTRACT	ix
INTRODUCTION	1
PHASE 2 ENERGY SAVINGS IMPACTS	2
ESTIMATED VS MEASURED IMPACTS	5
EXTENSION OF PHASE 2 RESULTS	5
EVALUATION ISSUES	5
EVALUATION MODEL SPECIFICATION	7
NEED FOR NEW METHODS	8
EXTENDED MODEL SPECIFICATION	9
ESTIMATION OF SAVINGS	13
DATA DESCRIPTION	14
ENERGY SAVINGS ESTIMATES — ANNUAL AGGREGATE METHOD	18
SENSITIVITY TO TIME AND OUTLIERS	18
INVESTMENT COST EFFECTS	21
CONTRACTOR EFFECTS	21
CONTROL GROUP	26
NET ENERGY SAVINGS	29
ENERGY SAVINGS ESTIMATES — BUILDING SPECIFIC MODEL RESULTS .	33
ABNORMALITY OF NOVEMBER/DECEMBER 1991	34
FIXED COEFFICIENT APPROACH	35
FIXED COEFFICIENT APPROACH WITH POOLING CONSTRAINTS ..	38
TEST OF POOLING CONSTRAINTS	40
RANDOM COEFFICIENT MODEL	43
CONCLUSIONS AND RECOMMENDATIONS	44
REFERENCES	47
APPENDIX A, Error Estimates for Model [5]	48
APPENDIX B, Abnormality of Data for Nov. and Dec. 1991	49

List of Tables

Table 1. Number of Buildings in the Phase 2 Treatment and Control Groups	2
Table 2. Energy Savings per Square Foot for the Small Commercial Program	3
Table 3. Aggregate Energy Use for Pre-retrofit and Post-retrofit	4
Table 4. Average Revenue, ¢/kWh	15
Table 5. Sub Sample Definition, Number of Buildings	17
Table 6. ←Sensitivity of Cost Effectiveness to Building Sample and Evaluation Period	19
Table 7. Sensitivity of Savings to Possible Outliers for S7	20
Table 8. Outlier Large Changes in kWh Per SQFT	20
Table 9. Cost Effectiveness by Investment Level, Sample S7, Treatment Group	21
Table 10. Effect of Screening Assumption on Cost Effectiveness	22
Table 11. Investment Cost by Contractor and Level of Cost, \$/kWh	22
Table 12. Number of Contractor Projects by County	23
Table 13. Contractor Performance as a Function of Cost and SQFT	24
Table 14. Pooled Estimates of Contractor Performance	25
Table 15. Control Group Annual Total Consumption	26
Table 16. Control Group Annual Total Consumption by County, Sample S7	27
Table 17. Control Group Energy Survey Responses	27
Table 18. ←Control Group Consumption Difference Model using Energy Survey Information	28
Table 19. Aggregate Treatment and Control Group Coefficients and Effects with Net Savings, Sample S7	32
Table 20. Building Types by Contractor (row %)	33
Table 21. Unconstrained Building Specific Model Types Matrix	34
Table 22. Estimates of Building Specific Savings for Model 3	37
Table 23. Initial Pooling Constraint Coefficients — Model 5 — Treatment Group Only	41
Table 24. FINAL Pooling Constraint Model Coefficients — Model 5b Treatment Group Only	42
Table 25. Test Results of Pooling Constraints on Treatment Effects	42

List of Figures

Figure 1	Schematic of attrition from treatment and control sample pools	2
Figure 2	County unemployment rates corresponding to available energy use data	15
Figure 3	Nominal electricity price	15
Figure 4	Treatment group energy use (N=1251) for 12-month moving average	18
Figure 5	Control group energy use (N=259) for 12-month moving average	18
Figure 6	Model of contractor performance and cost limit for contractors 1 & 2	26
Figure 7	Model of contractor performance and cost limit for contractors 3 & 4	26

List of Equations

Equation 1		7
Equation 2		8
Equation 3		10
Equation 4		10
Equation 5		10
Equation 6		11
Equation 7		11
Equation 8		11
Equation 9		12
Equation 10		23
Equation 11		30
Equation 12		39

SUMMARY

The commercial direct investment program of the Commonwealth Energy System (COM/Electric) is targeted toward small commercial and industrial (C&I) buildings, and under the program primarily comprehensive lighting retrofits were installed in customer facilities. Several million dollars were invested in the program in the year 1990 to install retrofits in customer facilities. This report documents the results of an analysis that extends results obtained in a previous impact evaluation of energy savings. The analysis includes over 1,000 buildings that participated in the program.

The previous evaluation showed net program savings, after a control group adjustment, of about 8.7 GWh/yr (14% of pre-retrofit energy use) for an investment of \$4.4 million, which leads to an investment cost ratio (investment divided by first year savings) of about \$0.50/kWh. This savings was estimated using simple, annual aggregate statistical measures of change in electricity use, which is typical of many C/I evaluations that are performed.

The analysis reported here extends the phase 2 evaluation results to look more closely at factors affecting program performance, including investment level, size of building, location, and weather and economic factors. Multivariate regression methods are used to identify important impacts on program performance related to these data. The evaluation did not examine electric demand data, because previous work by COM/Electric in coordination with state utility regulators had led to an energy-based index to determine program cost effectiveness.

The main drawback with the aggregate method is the information loss from adding up monthly bills. When using a year of pre and post data, the relationship between temperature and consumption is obscured when only two annual data points are available. Similarly, irregular variations in consumption which are clearly not related to retrofit measures such as a short term vacancy, will be confused with an energy savings effect. Methods which model individual bills instead of modeling annual aggregate data can be adjusted to allow for possible outlier observations by either deleting or weighting them down.

The PRinceton Scorekeeping Method (PRISM) focuses on modeling the relationship between energy and temperature, while leaving the influence of other factors to be estimated in a second stage. The strength of PRISM lies in determining the thermal parameters, but its weakness lies in using the two-stage regression. In particular, the key parameters of interest, those directly related to retrofit energy savings, are not estimated jointly with temperature effects. Because of this shortcoming, estimates of these key parameters are at best, inefficient, and at worst, biased. In addition, the implementation of this method with a year of data pre and post is somewhat difficult. The model for buildings with both heating and cooling must be estimated separately for both the pre and post period, which requires estimating five parameters with twelve or less observations.

An econometric approach (pooled factors) used by some evaluators yields joint estimates of all model parameters but imposes a large number of improbable parameter constraints. Sometimes this type of approach requires the marginal effect of each variable to be the same for all buildings. This assumption is virtually never satisfied by most samples of commercial or even residential buildings. Thus, the pooled factors approach has the strength of estimating thermal and other parameters jointly but the weakness of not allowing specific

types of buildings to be identified and treated for differences in the regression. The significant thermal differences between electricity use of buildings that heat with electricity as compared to those that do not heat with electricity is dramatic, and pooling thermal parameters is questionable.

A number of the problems associated with these approaches can be addressed by extending the PRISM model. In the analysis presented in this report, economic as well as thermal factors are included in a regression where parameters are jointly determined. Building thermal parameters are determined individually and then held constant while pooled parameters are estimated. This two-step process is iterated until the pooled parameters reach convergence.

The treatment group sample for this impact evaluation is large, with the initial pool of about 1,700 treatment facilities representing about 85% of the approximately 2,000 facilities treated under the program during the period August - December 1990. Attrition of the treatment pool due to data gaps and change of customer meter reduced the final sample to about 50% of all buildings treated during the period of interest. The control group was selected from facilities that are on the waiting list for this program, to improve comparability and overall results. The control group is much smaller, with about 260 facilities, but little attrition was experienced.

Energy savings for this program are estimated using both simple, aggregate data and the two-stage building specific parameter modeling. The aggregate data provide important insight about program performance, while the two-stage regression modeling helps reduce variability in results. Overall, the impact evaluation results indicate that the net program investment cost ratio, the ratio of total costs to net electricity savings, is close to \$0.60/kWh when weather and cost factors are included. Economic factors were included in savings estimate calculations, but the results indicated either data endogeneity or high variance problems, so economic effects cannot be inferred as being estimable for the data set on this small C/I program with the methods used here. Modeling of building thermal type gives some indication of heating penalty and cooling benefit for lighting retrofits. Cooling benefits could be modeled directly, while heating penalties seem to be apparent in the bottom-line first-year investment cost.

Survey results for a control group, together with actual energy use data, were used to determine whether customer responses could be used as indicators of actual changes in energy use. The control group was selected from waiting lists for the small C/I program, so the customers should be comparable to those in the treatment group. Analysis results show that customer responses to the survey questions **are not reliable indicators of actual energy reduction impacts**. The analysis results are empirical evidence that free rider effects are not present, because even though potential future program participants stated that they had installed energy saving measures, no corresponding energy use reduction could be found.

Program performance is at the margin of acceptable performance, and because of this marginality, improvements in program execution are desirable. Program performance at the margin could be improved significantly by controlling for potentially poor retrofit choices. These poor choices are indicated by estimates of savings that exceed 50% of pre-retrofit electricity use. Whatever the cause of these high estimates, a simple check of actual consumption before approving larger jobs should be instituted. Overall program performance could also apparently be improved by instituting increased checks on job approval in cases

where the invested \$/sq.ft. or total dollars exceeds certain cutoff criteria. We can suggest that whenever investment costs exceed \$1⁰⁰/sq.ft., that job approval should be more meticulous. However, the exact level of this cutoff probably requires more careful field and engineering examination of the jobs that exceeded this cutoff. About half of the jobs in our sample during the period of August - December 1990 exceeded this limit, so better definition of a reasonable cutoff value might help limit program cost increases for administration.

The methods presented here provide some useful methodological insight. Further work could also be conducted to more completely test these methods. The importance of using the correct error estimates on building energy use data sets was highlighted by this study. Specifically, failure to account for heteroscedasticity when evaluating building energy use data sets over time will likely lead to incorrect selection of significant parameters, which can lead to biased (and incorrect) results. Further work is needed to better understand the methods used in this study, and if incentives for program performance depend on reasonably close estimates of energy savings, careful modeling with several methods is probably necessary.

Extension to Phase 2 Energy Savings Evaluation of the Commercial Direct Investment Program

D. A. Trumble

J. M. MacDonald

ABSTRACT

The results of an analysis of the commercial direct investment program of the Commonwealth Energy System (COM/Electric) is presented. The analysis extends results obtained in a previous impact evaluation of energy savings. The analysis includes over 1,000 buildings that participated in the program. The previous evaluation showed net program savings, using simple, aggregate annual measures of energy use and a control group adjustment, of about 8.7 GWh/yr (14% of pre-retrofit energy use) for an investment of \$4.4 million, which leads to an investment cost ratio (investment divided by first year savings) of about \$0.50/kWh. Inherent difficulties with simple, aggregate calculations, with the PRinceton Scorekeeping Method (PRISM), and with econometric approaches for these types of evaluations are discussed and an extension to PRISM proposed as an alternative. Energy savings for this program are estimated using both simple, aggregate data and the extended PRISM approach (two-stage regression modeling). The aggregate data provide important insight about program performance, while the two-stage regression modeling helps reduce variability in results. Overall, the impact evaluation results indicate that the net program investment cost ratio, the ratio of total costs to net electricity savings, is close to \$0.60/kWh when weather and cost factors are included. Economic factors were included in savings estimate calculations, but the results indicated either data endogeneity or high variance problems, so economic effects cannot be inferred as being estimable for the data set on this small C/I program with the methods used here. Program performance is at the margin of acceptable performance, and because of this marginality, improvements in program execution are desirable. Program performance at the margin could be improved significantly by controlling for potentially poor retrofit choices.

INTRODUCTION

The commercial direct investment program of the Commonwealth Energy System, through its two electric utility companies, Commonwealth Electric Co. and Cambridge Electric Light Co., is an innovative, energy conservation program for commercial and industrial buildings in its electric service territories in Massachusetts. The program is targeted toward small C&I buildings, and under the program primarily **comprehensive** lighting retrofits were installed in customer facilities. Several million dollars were invested in the program in the year 1990 to install retrofits in customer facilities. For convenience the two utility companies will be referred to as COM/Electric in this report. This report documents the results of an examination of energy savings evaluation methods using the electric billing data available for over 1,000 buildings that participated in the program—an extension to the phase 2 evaluation. The energy savings achieved under this commercial direct investment program that are estimated using these evaluation methods are also reported.

A phase 1 evaluation of energy savings for the startup portion of the COM/Electric small commercial direct investment program was conducted in 1991 to provide the first estimates of energy savings for the program. An acceptable summary of these results can be found by Morris (1991). The phase 1 evaluation showed a measurable reduction of 11% of total electric energy use for a sample of 529 buildings that participated in the program. A control group showed a reduction in energy use also between the pre-retrofit and post-retrofit periods. The net difference between these two groups (the net savings) was about 7% of the total energy use of the participants. These savings did not meet cost effectiveness criteria, and based on this initial evaluation, the program was stopped, until better controls on implementation could be instituted.

The Morris article on the phase 1 evaluation has two factual errors. First, the ORNL report submitted to COM/Electric was based on the best available data at that time, and the results from this study were used to determine that the data contained in the program data base for one contractor were in error. The table in the article shows how unreasonable the predicted savings appear for the contractor in question, and later investigation showed that both the installation costs and the predicted savings for this contractor in the program data base were consistently incorrect (high). After correcting the errors, calculations showed the investment cost for conserved energy to be much lower: \$0.62 instead of \$0.74 per kWh for gross savings and \$0.84 instead of \$1.09 for net savings (all because of the incorrect data). Our concern over potential data problems led us to keep these initial reports informal, without publishing anything in the open literature. Despite the improvement in investment cost, cost effectiveness criteria were still not met.

The second point of correction relative to the Morris article is that the phase 1 evaluation did **not** represent a "reasonably clear picture of the energy savings impacts of the program as well as capture any adjustments and learning that were necessary." The report stated [we could] "surmise what the results would be if a full year of data were available for the post-retrofit period." The analysis was based on only six months of data, and we were examining the startup portion of the program. The evaluation was performed with less data than we would have liked because the Department of Public Utilities wanted some quick results that

they could examine. Initial results showed that corrections in the program were probably needed, but these results were not definitive because of the uncertainties.

Because of the data corrections needed, a phase 2 evaluation was conducted. The control group for phase 2 is selected from waiting lists for the small commercial (direct investment) program (buildings that have not yet received program services) in contrast with the control group selected from the general population of small commercial customers in phase 1. The phase 1 control group exhibited a significant decrease in energy use over the evaluation period (possibly due to economic conditions), and the buildings in the phase 1 control were larger than treatment buildings by a factor of about two. The phase 2 treatment group was selected from businesses that received program services later than the startup months of the program. The phase 1 evaluation had to use buildings from the startup phase to obtain any results, but that restriction no longer applied for the phase 2 evaluation.

PHASE 2 ENERGY SAVINGS IMPACTS

The phase 2 evaluation used aggregate statistical results to estimate gross and net savings for program retrofits installed in customer buildings during the period August - December 1990 (report available from COM/Electric). A pre-/post-retrofit analysis was used. The treatment group includes 1251 buildings. The control group consists of 259 buildings on the program waiting list (Table 1). These samples were selected from all available data for the program, so sampling was essentially at 100% of available data, depending on criteria for data completeness.

Table 1. Number of Buildings in the Phase 2 Treatment and Control Groups		
	Original pool	Final set
Treatment group	1710	1251
Control group	264	259

The entire population of buildings that received retrofits during the period from August 1990 through December 1990 was selected as the sample. From the total of about 2,000 retrofit installations that participated in the program during this period, supporting data were available for about 1,700 participants. Based on this population definition, sampling size was not an issue that was considered. Utility billing records for 1710 building "meters" were obtained. Data time coverage extended from October 1988 through March 1992, but not all meters had data over the entire time period. Of the 1710 meters, there are 307 meters that were "turned on" (changed customers) after the data starting point of October 1988, and these 307 were deleted from the sample (Fig. 1). The energy analysis made use of data beginning in October 1988, so these observations were removed to avoid potential problems with invalid data due to whatever changes occurred for these meters. An additional 152 meters had either missing pre-retrofit or post-retrofit data and were dropped from the sample. Total deletions are 459 (Fig. 1, $1710 - 459 = 1251$).

The control group pool had 264 meters originally, and all 264 of these meters could have been used in the analysis. However, there were some buildings that were much larger than typical of the buildings in the treatment group, and 5 buildings were removed simply to make the mean building size of both treatment and control groups similar. The control group was not stratified. The energy use results are not sensitive to this removal (Fig. 1).

Figure 1 — Schematic of attrition from treatment and control sample pools

The impact evaluation aggregate statistical results on a per-square-foot basis are summarized below (Table 2). These results are determined without

weather correction, and weather correction is addressed later in this report. The savings values for each observation in the data are weighted based on floor area, i.e., the values for each building were multiplied by the ratio of the floor area of the building to the mean floor area for the sample. The floor area weighting changes the values from the mean values obtained without the weighting, but the floor area weighted values can be multiplied by total floor areas to arrive at total annual energy (kWh) savings values ($\text{kwh/yr} = \text{kWh/ft}^2 \times \text{ft}^2$).

Table 2. Energy Savings per Square Foot for the Small Commercial Program			
Annual electricity use and savings (KWh/ft ²), and building size, ft ²	Mean	95% confidence interval	Standard deviation
Treatment group (N = 1251)			
Building size (ft ²)	4014	±332	5980
Pre-retrofit use	12.17	±0.88	15.8
Post-retrofit use	10.30	±0.84	15.1
Change in use (gross savings)	-1.87	±0.28	5.1
Control group (N = 259)			
Building size (ft ²)	4386	±699	5706
Pre-retrofit use	7.72	±1.61	13.1
Post-retrofit use	7.58	±1.54	12.6
Change in use	-0.14	±0.28	2.3
Net program impact	-1.73	±0.40	

Values are weighted based on floor area of each building compared to the mean.

Based on energy use rates for the pre-retrofit year of October 1988 to September 1989 and the post-retrofit year of October 1990 to September 1991.

Median lies outside 95% C.I. for change in use.

The data show that the annual energy use in the treatment group decreased 15% from the pre-retrofit use (compared with 11% in the phase 1 evaluation). The control group showed a reduction of about 2%, which is lower than the reduction of 4% estimated in the phase 1 evaluation. The net program impact is estimated to be 14% of total pre-retrofit electric energy use, which is much higher than the 7% net savings estimated in the Phase 1 evaluation. These results also showed the program was estimated to meet cost effectiveness criteria. However, the program design is still being evaluated, and the program has not restarted.

The calculations are based on a post-retrofit year that includes some of the period during which retrofits were occurring: October 1990 to September 1991. The pre-retrofit year is October 1988 to September 1989. Savings are clearly evident during this period. Anomalous billing data records from October through December 1991 (discussed later) obscured the savings, so this period was excluded from the analysis.

The energy use per square foot in the control group in 1988/89 is only about 60% of the energy use of the treatment group. This difference means that energy savings could be adjusted by percentages, instead of using absolute change in energy use. However, since the change in energy use for the control group is small, the percentage adjustment would lead to only a slightly lower net savings (1.65 kWh/ft²). Use of the percentage adjustment is clouded by the fact that the change in energy use of the control group cannot be distinguished from zero with reasonable confidence. Thus, the absolute change is used in the calculations here.

Aggregate data for the program participants for the 1251 meters analyzed are presented in Table 3. The investment cost for conserved energy determined in this analysis is shown below, together with total investment and floor area involved. The predicted savings values obtained from contractor site estimates are also shown.

Table 3. Aggregate Energy Use for Pre-retrofit and Post-retrofit			
Group	Pre-retrofit energy use (GWh/yr)	Post-retrofit energy use (GWh/yr)	Change in energy use (GWh/yr)
Treatment	61.1	51.7	9.4
Control	8.8	8.6	0.2

1251 buildings:	Total floor area:	5.0 million ft ²
	Predicted savings:	20.2 GWh/yr
	Total cost:	\$4.4 million
	Measured gross savings:	9.4 GWh/yr
	Measured net savings:	8.7 GWh/yr

Cost of predicted savings:	\$0.22 /kWh
Cost of measured gross savings:	\$0.47 /kWh
Cost of net savings:	\$0.51 /kWh

EXTENSION OF PHASE 2 RESULTS

The analysis reported here extends the phase 2 evaluation results to look more closely at factors affecting program performance. Data were available to identify specific contractors that installed retrofits and also identify the four counties where retrofits were installed. Multivariate regression methods are used to identify important impacts on program performance related to these data. In addition, extension of methods for estimating energy savings from demand-side management programs with large sample sizes are examined.

EVALUATION ISSUES

This evaluation of, primarily, lighting retrofits is based on the use of whole-building utility billing data for estimating benefits. Other methods are being developed and used, but the whole-building evaluation serves as an important real check on actual total energy savings benefits. Fels (1991a) discusses the importance of whole-building evaluation vs some newer methods and indicates the growing importance of commercial buildings in demand-side management (DSM) programs, "whose evaluation will require new techniques." Because new techniques are required, some uncertainty exists over the appropriate approaches to use and the appropriate data needed. This report examines the whole-building data obtained for the phase 2 evaluation of the COM/Electric commercial direct investment program to begin to answer the questions being raised by the Massachusetts Department of Public Utilities, intervenors in the process that establishes guidelines for utility programs, and evaluation staff at COM/Electric. These results should be useful to others involved in evaluations of energy savings for utility DSM programs.

Fels (1991a) also discusses the potential pitfalls and weaknesses of relying too heavily on more complicated end use metered data, where sample sizes may be small and understanding of the energy behavior of a utility customer population of commercial buildings (such as small commercial buildings) may be obscured. Fels also warns against the undesirability of "free-for-all and fearless" evaluation methodology development. This report describes our effort to put past evaluation methods in perspective, while demonstrating how extensions to current methods can be applied.

The evaluation did not examine electric demand data, because previous work by COM/Electric in coordination with state utility regulators had led to an energy-based index to determine program cost effectiveness. The energy index resulted because reduced generation has important benefits and averaging of electric demand benefits was considered adequate for assessing overall benefits. Thus, the evaluation was geared toward measurement of energy benefits.

The phase 2 evaluation did not specifically measure free driver bias impacts, although free rider impacts are examined. Free rider and free driver estimation is difficult. An estimate of free rider impact on energy savings for a similar small C/I lighting program at

Massachusetts Electric Company indicated that, based on application of several survey methods and estimation of free rider fractions for several lighting technologies, the free rider fraction impact on energy use was about 5% (Miller 1992).

Evaluation of a survey of the control group used in this study indicated little energy savings impact from any measures, except reducing the heating temperature (see Table 18 and companion text). This result indicates that no significant free rider effect appears to be present, since the control group did not demonstrate energy savings that correspond to stated installation of energy savings measures.

Self-selection or selectivity bias can also be an issue for conservation programs. The high participation rates for this program argue against considering selectivity bias. If such a program were operated over five years, the Companies could conceivably have program penetration exceeding 80% of all eligible customers. With such high penetration, selectivity bias would no longer be an issue, because most customers would have been selected. Selectivity bias was not considered at this point because so many customers are interested in the benefits they can realize from this program. Program interest and participation in the year 1990, which is the year evaluated in the phase 2 evaluation, were extremely high and led to long waiting lists. Whether future participants will realize the same savings as those in this 1990 sample is an open question.

Overall, any potential adjustments to program savings for free riders and self-selection, or for free drivers, are known or expected to be small. Empirical evidence from the survey of the control group (which was selected from waiting lists for the program) suggests that the free rider effect is small and not present. Since free drivers are not estimated and would potentially cancel penalties from free riders, any adjustment for free riders is probably so small as to not merit further consideration. Control group adjusted estimates are also presented.

The increasing complexity of evaluation methods will be evidenced by the discussion in the following sections that provide the description of past evaluation methods and of our extension to the COM/Electric phase 2 evaluation. Because of this increased complexity, standardization of the appropriate methods to use for evaluating program energy savings for commercial building retrofits will require careful examination and consideration of the strengths and weaknesses of each approach (for example, see Prah 1992). We hope the effort reported here provides a service to all those working to achieve effective DSM program implementation.

EVALUATION MODEL SPECIFICATION

Evaluation studies based primarily on whole-building billing data can generally be identified with one of the following three model type categories:

1. Annual or Seasonal Aggregate Model
2. Building Specific Thermal Model (e.g. PRISM)
3. Pooled Factors Model

The aggregate approach focuses on the unadjusted annual or seasonal difference, or percent difference, between pre and post periods for both a treatment and control sample. These differences are usually summarized by measures such as a simple average or a more complicated regression model to incorporate additional factors. A number of variations of this approach, such as including engineering energy savings estimates are discussed by Violette (1991). The phase 2 evaluation described in the Introduction to this report presents annual aggregate results for treatment and control samples.

The best known building specific thermal model approach is the Princeton Scorekeeping Method (PRISM) (Fels 1986). This method has recently been extended to include the following model with both heating and cooling degree day variables with estimated reference temperatures (Fels 1991b):

$$y_t = \alpha + \beta_h H(\tau_h) + \beta_c C(\tau_c) + \epsilon_t$$

where y_t - kWh/day

α - base-level consumption, kWh/day

β_h - heating slope, kWh/ F-day

β_c - cooling slope, kWh/ F-day

τ_h - heating reference temperature, F

τ_c - cooling reference temperature, F

[1]

The estimated model is typically used to calculate normalized consumption by using a long term average estimate of heating and cooling degree days. The normalized estimates are then aggregated to an annual or seasonal total. Analysis of these temperature adjusted totals then proceeds as in the previous aggregate model approach.

The pooled factors approach attempts to model the individual billing observations for a sample of buildings with a single regression model. These models usually include factors such as heating and cooling degree days (at a fixed reference temperature, see Eqn 1), square foot, price, and a retrofit variable to estimate energy savings. One model of this type estimated by Train (1987) for lighting retrofits is:

$$y_t = \beta_0 + \beta_1 C(70) + \beta_2 H(60) + \beta_3 SQFT + \beta_4 HOURS + \beta_5 PRICE \\ + \beta_6 TIME + \beta_7 RF + \beta_8 RF * SUN + \gamma SBIAS + \epsilon_t$$

$$\epsilon_t = \rho \epsilon_{t-1} + v_t$$

[2]

where y_t - kWh/day

RF = 1 post retrofit, 0 otherwise

SUN - sunlight index

SBIAS - selectivity bias correction

ρ - first order autocorrelation coefficient

The estimated monthly energy savings is then $\beta_7 + \beta_8 (SUN)$. The PRICE variable is the building's average price defined as the monthly bill divided by the corresponding kWh consumption. Train recognized that since average price is endogenous, the estimated price response is not reliable.¹ The selectivity bias correction term (SBIAS) is based on the method described by Dubin and McFadden (1984) using the probability of being in the sample, which is estimated as a function of building characteristics.²

NEED FOR NEW METHODS

The main drawback with the aggregate method is the information loss from adding up monthly bills. When using a year of pre and post data, the relationship between temperature and consumption is obscured when only two annual data points are available. Similarly, irregular variations in consumption which are clearly not related to retrofit measures such as a short term vacancy, will be confused with an energy savings effect. The phase 2 evaluation completed previously for the COM/Electric small commercial direct investment program uses an aggregate model. Methods which model individual bills instead of modeling annual aggregate data can be adjusted to allow for possible outlier observations by either deleting or weighting them down.

The PRISM approach focuses on modeling the relationship between energy and temperature, while leaving the influence of other factors to be estimated in a second stage. Fels (1991a) demonstrates a method for using PRISM with commercial DSM program data that identifies the energy use behavior of commercial buildings with respect to temperature as: heating only, cooling only, heating and cooling, flat (without heating and cooling), and unreliable, using specific reliability and flatness criteria. The strength of PRISM lies in determining the thermal parameters, but its weakness lies in using a two-stage approach. In particular, the key parameters of interest, those directly related to retrofit energy savings, are not estimated

¹ Train (1987) concludes that the estimated response of firms' energy consumption to weather was not affected by the specification of the price variable, and analogous results are expected for estimated energy savings.

² Rate structure dummy variables are the most significant factor determining the probability of being in the sample. Unfortunately, the authors do not report the effect on estimated savings of excluding this correction term from equation [2]. The benefit of this procedure is therefore difficult to judge.

jointly with temperature effects. Because of this shortcoming, estimates of these key parameters are at best, inefficient, and at worst, biased. In addition, the implementation of this method with a year of data pre and post is somewhat difficult. Equation [1] must be estimated separately for both the pre and post period, which requires estimating five parameters with twelve or less observations.

The pooled factor approach described above yields joint estimates of all model parameters but imposes a large number of improbable parameter constraints. The model as presented in Eqn [2] requires the marginal effect of each variable to be the same for all buildings. This assumption is virtually never satisfied by most samples of commercial or even residential buildings. Thus, the pooled factors approach has the strength of estimating thermal and other parameters jointly, but this approach is limited by not allowing specific types of buildings to be identified and treated for differences in the regression. For example, just as was demonstrated for PRISM, the sample of buildings in this study can be divided into the following four categories, with a sizable number in each:

- Heating and Cooling Load
- Cooling Load Only
- Heating Load Only
- No Systematic Thermal Load

Even within each category, the marginal effect of temperature is clearly not constant across buildings. This observation suggests that the first order autocorrelated term in [2] is most likely due to model misspecification. Fels (1991a) uses similar categories of buildings for PRISM analysis based on thermal load, but the last category is different. Our category here includes all buildings with no systematic thermal load, which means buildings with a flat consumption profile, as well as buildings with erratic energy use with respect to temperature are included. The PRISM analysis used by Fels addresses buildings with a flat profile as the fourth category, based on a flatness criterion. Buildings which do not meet the flatness criterion or other criteria for other types of thermal loads are classified as "unreliable."

EXTENDED MODEL SPECIFICATION

A number of the problems associated with these approaches can be addressed by extending the PRISM model. In the case of lighting retrofits, the PRISM model could be expanded as shown in Eqn [3], where all coefficients are constrained to be non-negative. In this specification, the retrofit energy saving coefficients that determine the post retrofit base and cooling load are estimated jointly with the temperature effects.

However, a number of other factors, such as price or economic activity, may also cause a building energy consumption to change over time. To incorporate such additional factors the model can be further extended to that shown in Eqn [4], where UNEMP is the county unemployment rate. At this point the number of coefficients makes it difficult to estimate the model on a building by building basis. Similarly, if the model specifies all coefficients as building specific, then estimated retrofit energy savings will be determined independent of the control group. One solution is to pool some of the coefficients. For example, one might reasonably want to pool the coefficients α_2 , α_3 , β_{A2} , β_{A3} , β_{C2} , β_{C3} , τ_A , τ_C , leaving

$$y_t = \alpha - \alpha_r RF + \beta_h H(\tau_h) + \beta_c C(\tau_c) - \beta_r RF * C(\tau_c) + \epsilon_t$$

where y_t - kWh/day

α - base-level consumption

α_r - base-level retrofit savings

β_h - heating slope, kWh/ F-day

β_c - cooling slope, kWh/ F-day

β_r - cooling slope retrofit savings

τ_h - heating reference temperature, F

τ_c - cooling reference temperature, F

[3]

$$y_t = \alpha - \alpha_r RF + \beta_h H(\tau_h) + \beta_c C(\tau_c) - \beta_r RF * C(\tau_c) + \epsilon_t$$

$$\alpha = \alpha_1 + \alpha_2 PRICE + \alpha_3 UNEMP$$

$$\beta_h = \beta_{h1} + \beta_{h2} PRICE + \beta_{h3} UNEMP$$

$$\beta_c = \beta_{c1} + \beta_{c2} PRICE + \beta_{c3} UNEMP$$

[4]

$\alpha_1, \beta_{h1}, \beta_{c1}, \alpha_r, \beta_r$ as building specific parameters. Alternatively, specification [5]

$$y_t = \alpha(1 - \gamma RF) + \beta_h H(\tau_h) + \beta_c(1 - \eta RF)C(\tau_c) + \epsilon_t$$

$$\alpha = \alpha_1 + \alpha_2 PRICE + \alpha_3 UNEMP$$

$$\beta_h = \beta_{h1} + \beta_{h2} PRICE + \beta_{h3} UNEMP$$

$$\beta_c = \beta_{c1} + \beta_{c2} PRICE + \beta_{c3} UNEMP$$

$$\gamma = \gamma_1 + \gamma_2 COST/SQFT + \gamma_3 COST$$

$$\eta = \eta_1 + \eta_2 COST/SQFT + \eta_3 COST$$

[5]

could reasonably be estimated with building specific parameters $\alpha_1, \beta_{h1}, \beta_{c1}$. This model assumes that the proportion of base load saved, γ , and the proportion of cooling load saved, η , are functions of both total retrofit cost and cost per square foot of floor area.

The research objectives are an important consideration in deciding on these and other specification issues. If our only objective is to evaluate the cost effectiveness for this particular sample of building retrofits then pooling constraints such as γ and η are of secondary importance. Models with all building specific coefficients such as [3] or [4] will yield consistent estimates of treatment effects. Additional pooling constraints, if correct, will yield more efficient estimates, and if incorrect, will yield inconsistent results. However, if our objective is to make inference on the population characteristic, such as the cost effectiveness of continuing the program or modifying it, then a model that specifically incorporates cost factors in a pooled model is more valuable and building specific coefficient estimates are of less value.

To extend the model specification further, if the building specific model [3] is expressed as

$$y_i = Z_i \Gamma_i + e_i$$

[6]

where $y_i = (y_{i1}, \dots, y_{iT})'$
 $\Gamma_i = (\alpha^i, \alpha_r^i, \beta_h^i, \beta_c^i, \beta_r^i)'$
 $Z_i = (1, -RF, H(\tau_h), C(\tau_c), -RF * C(\tau_c))$

then the coefficients, Γ_i , can be treated as be either fixed or random. One well known random specification examined by Swamy (1970) and others is

$$\Gamma_i = \Gamma + \mu_i$$

[7]

where $E(\mu_i) = 0$
 $E(\mu_i \mu_i') = \Delta$
 $E(\mu_i \mu_j') = 0, \quad i \neq j$
 $E(Z_i, \mu_j') = 0$

To evaluate a particular sample of building retrofits the Γ_i coefficients may be estimated as fixed parameters. To make inference beyond this sample, the population parameters Γ and Δ should be estimated.

Since we are interested in making both conditional and unconditional inference regarding the cost effectiveness of retrofits, both approaches are examined here. The fixed parameter approach is considerably easier to estimate. If the residuals in [6] are heteroscedastic (the model variance for individual buildings is not constant and may be correlated with the variance of explanatory variables) by building and uncorrelated across buildings:

$$E(e_i e_i') = \sigma_i^2 I_T$$

[8]

$$E(e_i e_j') = 0, \quad i \neq j$$

Under these conditions, the general least squares estimator of Γ_i is the same as applying least squares separately to the time-series of observations on each building. One limitation with such a procedure is that the non-negativity constraints on the building specific coefficients are not imposed. These inequality restrictions are a way of incorporating prior information. If correct, the inequality restricted estimator under squared error loss is uniformly superior to the maximum likelihood estimator (Judge et al 1985 pg 91). In addition, the interpretation of the model coefficients is made clear in that an increase in electricity consumption is not associated with retrofit savings, nor a decrease in thermal load requirements.

Estimation methods and results are presented in the section on Building Specific Model Results for these fixed parameter building specific models, both unrestricted and restricted. The estimates are consistent under a range of alternative assumptions which lead to more efficient estimators. However, if these alternative assumptions are not met then the corresponding estimator which would have been more efficient are inconsistent. In particular, if the variable coefficients are correlated with the explanatory variables,

$$E(Z_i \mu_i) \neq 0 \quad [9]$$

then the random coefficient model [7] would yield inconsistent estimates and the annual model would have a problem with aggregation bias (Zellner 1969).

With additional pooling assumptions, the fixed parameter approach can also be used for unconditional inference. In model [5] for example, the thermal load variables are building specific, but retrofit savings are specified as a proportion which varies as a function of cost and size. Unfortunately, heteroscedasticity by building in combination with this non-linear pooling constraint makes it difficult to test against competing models. For example, a test of this pooled model against the building specific fixed coefficient model cannot be constructed as the usual F-test of constrained and unconstrained sum of squared residuals due to the heteroscedasticity. Swamy (1974) has proposed an estimation procedure for random coefficient models as well as a test for pooling constraints. A variation of this test for our purposes is

$$\sum_i (\Gamma_i - \Gamma^*_i)' V_i^{-1} (\Gamma_i - \Gamma^*_i)$$

where $V_i = \sigma_i^2 (Z_i' Z_i)^{-1}$

$$\Gamma^*_i = (\alpha^i, \gamma \alpha^i, \beta_h^i, \beta_c^i, \eta \beta_c^i)$$

which is asymptotically chi-square-distributed with $2(N-1)$ degrees of freedom. Under the null hypothesis:

$$\alpha_r^i = \gamma \alpha^i \quad \text{and} \quad \beta_r^i = \eta \beta_h^i$$

Unconditional inference of treatment effects can also be based on the random coefficient model. If condition [8] is satisfied, then the unconditional expected mean and variance of retrofit savings can be computed from the random coefficient model estimates of Γ and Δ .

If [9] is not satisfied, Mundlak (1978) has suggested that the bias could be corrected by specifying the coefficient vector, μ_i , as a function of other building characteristics such as retrofit costs. This class of models, where random coefficients are functions of other explanatory variables, is discussed by Hsiao (1986). Within this class of models, one possible specification is Model [5], where each of the supplementary equations include an additive error term. Initial estimates for this extended model can be obtained with auxiliary regression equations specifying the building specific retrofit coefficient estimates as a function of cost and size using the coefficient variance estimate as a weighting factor. However, the random coefficient model results appear to yield inconsistent estimates, possible due to the explanatory variables being correlated with the variable coefficients. This result is discussed later in the section on Building Specific Model Results.

ESTIMATION OF SAVINGS

Savings are estimated for the COM/Electric small commercial program using several methods in the following sections. The savings were previously estimated in the phase 2 evaluation using an aggregate annual approach, with two data points for each building: pre- and post-treatment annual consumption. The phase 2 evaluation did not consider the influence of other variables, such as cost per square foot of specific installations and total cost of installations. The phase 2 evaluation did examine contractor performance, based on the pre- and post-retrofit annual consumptions, but further examination of factors affecting program performance were not explored due to time constraints.

The estimation of savings for this report includes an extended look at other factors affecting program performance, using both annual aggregate models and the time series models specified in Eqns. [3] and [5], and extensions of [5] to cover random coefficient model results. The annual aggregate models provide important insight into factors affecting program performance, and the building specific time series models provide insight into possible effects of factors outside the program such as weather, economic activity levels (with unemployment as a proxy), and price of electricity.

The following sections provide a description of the data, the analysis using annual aggregate models, and the analysis using time series models.

DATA DESCRIPTION

The phase 2 evaluation is based on data for both a treatment and control sample of buildings. The extension to this evaluation, which is documented here, uses the sample pool of buildings for treatment and control groups (see Introduction), but different final samples are selected. The treatment sample of buildings had retrofits installed during the period August 1990 to December 1990. The control sample is a set of buildings scheduled to receive retrofits whenever the program resumes. The building retrofits in the treatment sample were preformed in four different counties by four different contractors. The term building is used loosely here, in that in some cases we are actually referring to a parking lot.

Both groups of data have information for the following variables:

- SQFT - Building gross floor area, square feet, sometimes missing
- EST - Annual electricity savings as estimated by the retrofit contractor, sometimes missing
- COST - Total cost of installed retrofits
- CNTY - County ID - Bristol, Barnstable, Plymouth, or Cambridge
- CNTR - Contractor ID - 1 to 4

The explanation and interpretation of missing values is not obvious. A missing value for SQFT can be an indication that the building is a parking lot. Both groups also had one additional piece of information. For the treatment group, the actual retrofit cost and completion date were available. For the control group, survey results regarding energy conservation behavior were available.

The electricity consumption history for each building was obtained from billing account data. Retrofit projects with multiple meters were aggregated to form a total for the billing period. Daily temperature information was available for each county. Both a heating degree day and a cooling degree day measure can be calculated for each building using a building-specific billing period and the daily average temperature for one location in each county. Common bases of 65 and 70 degrees were used for the heating and cooling degree day calculations, which is a simplification of the PRISM approach wherein building-specific base temperatures are calculated.

Additional information regarding the county unemployment rate and the price of electricity were also considered. Three counties, Bristol, Barnstable and Plymouth, all experienced a high seasonal unemployment rate in 1991, reaching a high of 12%-16% in January and a low of 6%-8% in June. Unemployment in Cambridge is considerably lower, usually in a range of 3%-6%. Figure 2 illustrates the extreme levels of unemployment experienced in the post retrofit evaluation period.

The pricing of electricity for Cambridge is also different than the other counties, which all have common pricing. The nominal average revenue presented in Figure 2 clearly illustrates the difference. The high volatility in average revenue is attributed to the fuel cost adjustment portion of price. Both series show a large increase for October 1990, which coincides with the retrofit installation period of September to December 1990. Table 4 summarizes the average revenue before and after this increase for the two pricing structure.

Figure 2 — County unemployment rates corresponding to available energy use data

Figure 3 — Nominal electricity price

Table 4. Average Revenue, ¢/kWh		
Time Period	Cambridge	All Other
1/89 - 9/90	8.65	9.78
10/90 - 12/91	9.61	12.56

The total sample of both treatment and control buildings, denoted sample S1 in Table 5, contains 1667 buildings. Complete data for roughly 20% of these buildings, mostly in the treatment group, are not available. Table 5 enumerates the number of observed buildings depending on data requirements. Sample S2 in Table 5 includes only buildings that had some billing data in all three years for which data were available. Almost 98% of the buildings were matched with at least have some billing data within the three year study period.

Sample S3 shows buildings that had at least 12 billing data points over the three years. S4 shows that most of the control group and about 80% of the buildings had billing histories covering at least 330 days for both 1989 and 1991, the pre- and post-treatment years. The sample used for the phase 2 evaluation is closest to S4, but the criterion for number of days was relaxed slightly. Thus, 113 more meters than the 1138 treatment buildings (total of 1251) in S4 were included in the phase 2 calculations.

S5 shows that about 95% of the original control group and about 75% of the treatment group could be matched with billing histories that covered the whole three years and having at least 330 days of data coverage for each year. Requiring square foot information to also be available lowers this availability figure, as shown by S7. S6 is the subsample of S4 that has no missing square feet data, and S7 is the subsample of S5 that has no missing square feet data.

Table 5. Sub Sample Definition, Number of Buildings			
S1 -	1667 - Total Sample, without treatment group meters that were turned on too late (see Fig. 1)		
	1403	- Treatment, 1710 minus 307 = 1403 (Fig. 1)	
		1357 - no missing square feet	
		1337 - no missing contractor savings estimates	
		1293 - no missing square feet & contractor savings estimates	
	264 - Control		
S2 -	1633 - Sample with billing data in all 3 years		
	1369 - Treatment		
		1326 - no missing square feet	
		1303 - no missing contractor savings estimates	
		1262 - no missing square feet & contractor savings estimates	
	264 - Control		
S3 -	1598 - Sample with at least 12 bills over 3 years		
	1334 - Treatment		
		1293 - no missing square feet	
		1270 - no missing contractor savings estimates	
		1231 - no missing square feet & contractor savings estimates	
	264 - Control		
S4 -	1400 - Sample with complete (at least 330 days) years 89 & 91 Consumption		
	1138 - Treatment		
		1101 - no missing square feet	
		1080 - no missing contractor savings estimates	
		1045 - no missing square feet & contractor savings estimates	
	262 - Control		
S5 -	1340 - Sample with complete (at least 330 days) years 1989, 1990, 1991 Consumption		
	1086 - Treatment		
		1050 - no missing square feet	
		1031 - no missing contractor savings estimates	
		997 - no missing square feet & contractor savings estimates	
	254 - Control		
S6 -	1363 - S4 & No missing Square feet		
		1101 - Treatment	
		1045 - no missing contractor savings estimates	
	262 - Control		
S7 -	1304 - S5 & No missing Square feet		
		1050 - Treatment	
		997 - no missing contractor savings estimates	
	254 - Control		

ENERGY SAVINGS — ANNUAL AGGREGATE METHOD

The energy savings calculated for the phase 2 evaluation showed the investment cost for conserved energy to be about \$0.50/kWh. The evaluation also showed that anomalous billing consumption data patterns, that are clearly out of line with historical data, were evident after September 1991 for the treatment group. These data, which are drastically anomalous based on inspection, were not included in the evaluation calculations. The data were excluded by shifting the pre- and post-treatment years back three months, running October through September: 1988/89 - pre, and 1990/91 - post. This shift means that retrofits were still occurring during the post-treatment year. (See Appendix B for further discussion.)

SENSITIVITY TO TIME AND OUTLIERS

Further examination of the billing data for the treatment group after September 1991 confirmed its anomalous nature. The energy savings are very sensitive to the evaluation period. Inspection of the consumption data indicates that bills for the month of November and December of 1991 are usually high, even for buildings which have no identifiable heating load. Figures 4 and 5 illustrate this phenomenon, where the treatment group exhibits a distinct, large increase in annual energy use beginning at this time, while the control group remains consistent. Table 6 shows the effect on overall investment cost of including these observations in late 1991. The ratio of costs to savings (investment cost) actually increases from 1.1 to 2.4 \$/kWh with the more restricted samples. The ratio is unstable and shifts upward markedly as the sample is restricted. On the other hand, excluding these observations yields a more stable ratio in the range of 0.53 to 0.55 \$/kWh, which is in the same range as reported for the phase 2 evaluation. (Later examination showed that the data for 11/91 and 12/91 were incorrect, see Appendix B.)

Figure 4 — Treatment group energy use (N=1251) for 12-month moving average

Figure 5 — Control group energy use (N=259) for 12-month moving average

Savings estimates for both of these measures are simply the difference between pre and post annualized consumption, which is an annual aggregate measure of program performance, the same type of result reported in the phase 2 evaluation. In this extension to the evaluation,

we examine annual aggregate results more thoroughly and also report on the results obtained with the extension to PRISM (Eqns [3] and [5]).

Table 6. Sensitivity of Cost Effectiveness to Building Sample and Evaluation Period									
Annual Total Consumption & Retrofit Cost (Treatment Group)									
Sample	N	1989	1990	1991	Cost	Cost/ Savings	1989	1991	Cost/ Savings
		(a)	(a)	(a)	(b)	(c)	(d)	(d)	(e)
S1	1403				4.9				
S2	1369				4.8				
S3	1334	62.9	59.4	58.6	4.6	1.07			
S4	1138	55.8	51.8	53.7	4.1	1.96	55.3	47.5	0.53
S5	1086	54.3	50.6	52.6	3.9	2.22	53.8	46.5	0.54
S6	1101	54.1	50.4	52.3	3.9	2.13	53.6	46.3	0.54
S7	1050	52.7	49.2	51.2	3.8	2.43	52.1	45.3	0.55

(a) GWh/yr - Calendar Years

(b) Millions of 1990\$

(c) 1990\$ / kWh, Savings = 1989(a) - 1991(a), DEL1

(d) GWh/yr - 1989 = 10/1988 through 9/1989; 1991 = 10/1990 through 9/1991

(e) 1990\$ / kWh, Savings = 1989(d) - 1991(d), DEL

Table 6 also shows the breakdown of total retrofit cost by sample size. In terms of accounting for the total retrofit cost of \$4.9 million, restricting the evaluation sample from S1 to S4 excludes 16% of costs from the estimate of savings, which is unavoidable because of missing consumption data. The additional requirement of three complete years and no missing square foot information, sample S7 over S4, results in an additional 6% of costs excluded.

To further examine the robustness of these ratio estimates the effect of possible outliers was considered. Table 7 presents the ratio of retrofit costs to estimated annual savings for subsets of sample S7. The Bowman/Shenton Normality test statistic is also presented to show the effect of omitting these observations. This test statistic indicates that for sample S7, the assumption that S7 has a "normal" distribution is rejected at the 95% confidence level for measures of annual change. Rejecting observations that change by more than 10 W/sq.ft. allow the sample to pass the normality distribution assumption. Surprisingly, rejecting observations which change by more than 10 W/SQFT has little effect on the cost measure based on savings. Alternative criteria for outliers percent change in consumption in excess of 35% or 50% yield a significantly higher cost ratio, but the normality test fails for these outlier assumptions. A review of buildings with consumption changes in excess of 10 W/SQFT as identified in Table 8 shows that the majority of these observations actually represent a relatively small percent change.

Table 7. Sensitivity of Savings to Possible Outliers for S7

SAMPLE	N	AVERAGE		INVESTMENT COST			
		DEL	DEL1	DEL	DEL1	DEL	DEL1
		(a)	(b)	(c)	(c)	(d)	(d)
S7	1050	3.85	1.23	7.9	7.2	0.55	2.4
REJECT IF							
ABS(DEL) > 90	1041	3.15	1.23	0.28	0.36	0.56	2.7
ABS(DEL1) > 90							
REJECT IF							
ABS(%DEL) > 50%	823	3.24	0.18	9.3	7.6	0.63	12.6
ABS(%DEL1) > 50%							
REJECT IF							
ABS(%DEL) > 35%	706	2.28	1.05	10.8	9.5	0.71	4.3
ABS(%DEL1) > 35%							
(a) kWh per sqft, DEL = Years - 1988/10:1989/9 - 1990/10:1991/9							
(b) kWh per sqft, DEL1 = Calendar Years - 1989 - 1991							
(c) Bowman/Shenton Normality test, Critical Value for 95% confidence level = 6.0							
(d) 1990\$ / kWh							

Table 8. Outlier Large Changes in kWh Per SQFT
S7 — ABS(DEL) > 90 OR ABS(DEL1) > 90

Indicator (see below)	Floor area sq. ft.	Cost \$1,990	DEL (a)	DEL1 (b)	1989 MWh/yr	1991 MWh/yr
**	1400	10025	75	93	400	270
*	500	214	164	140	169	99
**	160	101	202	-222	145	181
*	150	101	336	345	172	120
**	100	6577	-50	-126	47	59
	100	100	-103	-117	4	16
**	100	201	2	-197	158	177
	100	582	141	106	19	8

* - % change less than 35% for both DELs
 ** - % change less than 50% for both DELs
 (a) DEL = Change in kWh/sq.ft., Years - 1988/10:1989/9 - 1990/10:1991/9
 (b) DEL1 = Change in kWh/sq.ft., Calendar years 1989 - 1991

INVESTMENT COST EFFECTS

Investment cost of conserved energy by level of retrofit cost is presented in Table 9 for sample S7. Almost half of the buildings in this sample, 45%, had retrofit costs less than \$2000 and are brilliantly cost effective (except \$1,000-2,000 for DEL1) using either pre- and post-year. However, this subset of low cost retrofits only account for 14% of the sample's total retrofit costs of \$3.8 million. The remaining 55% of the sample, buildings which have retrofit costs in excess of two thousand, account for 86% of total costs and are substantially less cost effective, with the choice of evaluation period again showing strong instability and a negative influence.

Table 9. Cost Effectiveness by Investment Level, Sample S7, Treatment Group				
Building Retrofit Cost (\$)	N	Total Cost	Investment Cost	Investment Cost
		\$1990 (000s)	DEL1 (a)	DEL (b)
100 - 500	61	20	0.08	0.05
500 - 1000	131	97	0.30	0.16
1000 - 2000	279	405	0.61	0.28
2000 - 5000	367	1192	-5.00	0.60
5000 - 10000	150	1040	5.20	0.73
> 10000	62	1037	2.90	0.96
SUM	1050	3791		
(a) 1990\$ / kWh, DEL1 = Calendar years 1989 - 1991				
(b) 1990\$ / kWh, DEL = Years - 1988/10:1989/9 - 1990/10:1991/9				

Compared to annual energy use of many buildings with high retrofit cost, a significant number have very high contractor estimated savings. Simply imposing a requirement that contractor estimated savings cannot exceed 80% of a building's pre-retrofit annual consumption would make investment costs for DEL (last column in Table 9) \$0.60 or less for all categories except the one that is greater than \$10,000. From a project standpoint, more expensive projects are more risky for COM/Electric and should be scrutinized more carefully. As demonstrated, simple rules of thumb on acceptable investments would significantly increase cost effectiveness. This result is presented in Table 10.

CONTRACTOR EFFECTS

An examination of contractor performance using this measure of cost effectiveness reveals considerable variation. Contractor 1 has an investment cost of \$0.61/kWh for retrofits in excess of \$10,000, whereas, Contractor 2 has a cost of \$0.73/kWh and higher for retrofits in excess of \$2,000. Contractor 3 has a cost of \$0.85/kWh for retrofits in excess of \$5,000. Contractor 4 exhibits some of the best performance, but this contractor may have benefitted most by simply not doing many large jobs that were not cost effective. The major point is that the program is very cost effective for small jobs, but larger jobs are apparently not selected as well. The larger jobs are causing the program to perform less well than it could. Thus, perhaps 10% of the overall number of jobs are reducing program performance.

Table 10. Effect of Screening Assumption on Cost Effectiveness				
Building Retrofit Cost (\$)	N	Total Cost	Investment Cost	Investment Cost
		\$1990 (000s)	DEL1 (a)	DEL (b)
100 - 500	51	17	0.07	0.05
500 - 1000	124	93	0.32	0.17
1000 - 2000	233	340	0.57	0.26
2000 - 5000	280	907	-4.10	0.54
5000 - 10000	104	1040	4.70	0.60
> 10000	40	631	2.50	0.69
SUM	832	3791		
65 - Rejected by Screening Criteria				
53 - Missing Contractor Estimates				
SUM	1050 - Sample S7, Treatment Group			
* Annual savings estimated by contractor must be less than 80% of 1989 annual consumption				
(a) 1990\$ / kWh, DEL1 = Calendar years 1989 - 1991				
(b) 1990\$ / kWh, DEL = Years - 1988/10:1989/9 - 1990/10:1991/9				

Table 11. Investment Cost by Contractor and Level of Cost, \$/kWh NOT ADJUSTED FOR CONTROL GROUP					
Building Retrofit Cost (\$)	Contractor				SUM
	1	2	3	4	
100 - 500	0.16	0.04	0.06	0.06	0.05
N	2	18	32	9	61
500 - 1000	0.15	0.20	0.18	0.09	0.16
N	21	22	71	17	131
1000 - 2000	0.26	0.33	0.30	0.23	0.28
N	46	54	137	42	279
2000 - 5000	0.52	0.73	0.55	0.45	0.60
N	95	102	122	15	367
5000 - 10000	0.59	0.92	0.85	0.59	0.73
N	50	46	39	15	150
> 10000	0.61	0.97	1.58	1.17	0.96
N	16	21	19	6	62
SUM, Sample S7					1050

However, an evaluation of contractor performance should take into account the different counties in which the retrofit measures were installed. Table 12 shows the number of retrofits in each county for the four contractors. Contractor 1 worked primarily in Bristol and contractors 3 & 4 worked primarily in Barnstable. Contractor 2 worked in both Plymouth and Cambridge. Therefore, distinguishing between county and contractor effects becomes more difficult. At this point, the sample size shrinks from 1050 to 1042 for the treatment group, to delete the outliers identified in Table 7 and improve statistical measures of results.

Table 12. Number of Contractor Projects by County					
Contractor	Bristol	Plymouth	Barnstable	Cambridge	Total
1	217	2	11	0	230
2	11	147	17	81	256
3	1	3	414	0	418
4	1	0	136	0	137
Total	230	152	578	81	1041

Unadjusted contractor performance was examined by estimating the following regression model (Eqn. [10]) separately for each contractor.

$$DEL = \alpha_0 + \alpha_1 COST/SQFT + \alpha_2 SQFT + \alpha_3 COST + \epsilon \quad [10]$$

To allow for possible county effects of Cambridge being different than Plymouth, buildings retrofitted by contractor 2 were divided into a Non-Cambridge (2a) and Cambridge (2b) sub-sample. (Cambridge electricity pricing differs from the other counties, and analysis of the control group shows differences between Cambridge and non-Cambridge control group buildings.) The results of these contractor specific models are presented in Table 13 (coefficients are in W/SQFT). The t-statistics reported in this table are based on the White heteroskedasticity consistent covariance matrix, since the Breusch/Pagan test for heteroskedasticity clearly rejects the constant variance assumption. Tests of whether the estimated coefficients are the same for all contractors show that statistically we strongly reject the hypothesis. In fact, coefficient estimates for contractor 2 in and out of Cambridge are statistically different. Contractor 2 exhibits different performance in the two different locations, probably showing differences in county consumption patterns. However, the hypothesis that the estimated coefficients are the same for contractor 1 and contractor 2 outside of Cambridge, (2a) cannot be rejected (Chow test results are shown as F(num,denom) in far right column). Similarly, a difference between contractor 3 and 4 is not shown statistically. Results are NOT ADJUSTED for the control group, so Cambridge performance is not as poor as a quick glance would suggest.

The R^2 for the contractor model is low, which shows the tremendous variation in energy use changes for buildings in this sample. This variation is typical of the types of changes found in buildings, and especially in commercial buildings over a three year period. The negative coefficients for COST again illustrate the negative impact of more costly jobs on overall savings. The negative coefficients for SQFT also show the trend toward difficulties with

Table 13. Contractor Performance as a Function of Cost and SQFT							
	CON=1	CON=2	CON=2a	CON=2b	CON=3	CON=4	Test
REG S.E.	0.585	0.889	0.822	0.971	0.674	0.565	CON 2a = CON 2b
SSE	77.225	198.938	115.521	72.529	187.895	42.418	F(4,248) = 3.589
R ²	0.064	0.066	0.050	0.206	0.122	0.133	CON 1 = CON 2a
ADJ R ²	0.051	0.055	0.034	0.175	0.115	0.113	F(4,397) = .096
Heterosced. Test	46.044	134.208	9.118	306.713	51.211	55.440	CON 3 = CON 4
Constant, W/SQFT	0.290	0.280	0.311	-0.110	0.261	0.353	F(4,447) = .263
t-statistic	4.037	3.449	4.164	-0.250	3.614	2.974	
COST/SQFT, W/\$	0.084	0.099	0.070	0.380	0.223	0.224	
t-statistic	2.327	1.974	2.973	1.218	4.102	1.644	
SQFT, W/1000-SQFT	-0.006	-0.008	-0.005	0.021	-0.004	-0.011	
t-statistic	-2.224	-1.154	-1.054	0.435	-0.530	-0.946	
COST, W/SQFT-\$1000	-0.011	-0.013	-0.018	-0.040	-0.039	-0.044	TOTAL
t-statistic	-1.186	-1.195	-2.299	-0.915	-4.261	-1.716	
N	230	256	175	81	418	137	
Total Savings, kWh	1949000	1518308	895295	623013	2219847	1031029	
Total Cost, \$	\$977,766	\$1,162,814	\$726,291	\$436,523	\$1,222,057	\$420,261	
Investment Cost, \$/kWh	0.502	0.766	0.811	0.701	0.551	0.408	0.563
NOTE: Positive coefficients show savings, e.g., for CON=1 the constant level of savings is 0.29 W/SQFT, and other coefficients adjust from that value. (W/SQFT x 8760/1000 = kWh/SQFT)							

larger buildings. Based on the statistics regarding differences between contractors, the pooled coefficient model estimates showing the contractors pooled together are presented in Table 14.

Table 14. Pooled Estimates of Contractor Performance				
	CON=1&2a	CON=2b (Cambridge)	CON=3&4 (SQFT not significant)	CON=3&4 (Final)
REG S.E.	0.694	0.966	0.647	0.647
SSE	192.932	72.816	230.757	230.987
R ²	0.055	0.203	0.122	0.121
ADJ R ²	0.048	0.182	0.117	0.118
Heterosced . Test	47.821	277.523	85.123	86.266
Constant	0.310	-0.017	0.282	0.262
t-statistic	6.226	-0.071	4.488	5.700
COST/SQFT	0.073	0.340	0.221	0.233
t-statistic	3.530	1.510	4.393	5.497
SQFT	-0.005		-0.006	
t-statistic	-2.080		-0.952	
COST	-0.015	-0.027	-0.039	-0.044
t-statistic	-2.416	-1.614	-4.470	-5.502
N	405	81	555	555
Total Savings, kWh	2844295	501386	3250876	3593842
Total Cost, (\$000s)	\$1,704,057	\$436,523	\$1,642,318	\$1,642,318
Inv. Cost, \$/kWh	0.599	0.871	0.505	0.457

Figures 6 and 7 show the impact of the data in Table 14. The model coefficients are used to plot straight lines of "linear" contractor performance. These lines are "trend" lines. The cost limit shown in these figures is the kWh per square foot energy savings required so that total retrofit cost divided by annual kWh savings is less than or equal to 0.6. For example, the model predicts that contractors 1 & 2 would exceed the ratio of 0.6 when retrofitting a thousand square foot building if retrofit cost per square foot exceeds \$2.75. Similarly, contractors 1 & 2 are predicted to exceed the ratio of 0.6 with a three thousand square foot building when cost per square foot exceeds \$1.75. Comparing the two figures indicates that contractors 3 & 4 were better than 1 & 2 with smaller buildings and worse than 1 & 2 with larger buildings.

Figure 7 — Model of contractor performance and cost limit for contractors 3 & 4

Overall, total annual consumption for the control group has decreased by 2% to 3% from 1989 to 1991. However, an examination by county shows that consumption in Cambridge has increased by 8%, with the other two counties decreasing by roughly 5%. The previously discussed differences in price and economic activity are two possible reasons. Table 15 shows aggregate data on these changes for the entire sample. A breakdown by county is shown in Table 16.

Sample	N	SQFT	1989	1990	1991	Percent	1989	1991	Percent
		(a)	(b)	(b)	(b)		(c)	(d)	
S1	264	1.83	15.1	15.1	14.7	-2.65%	15.0	14.7	-2.00%
S4	262	1.82	14.9	14.9	14.5	-2.68%	14.9	14.5	-2.68%
S7	254	1.79	14.7	14.7	14.3	-2.72%	14.7	14.4	-2.04%

(a) Millions of square feet
 (b) GWh/yr - Calendar Years
 (c) GWh/yr - 1988/10:1989/9
 (d) GWh/yr - 1990/10:1991/9

Table 16. Control Group Annual Total Consumption by County, Sample S7									
County	N	SQFT	1989	1990	1991	Percent	1989	1991	Percent
		(a)	(b)	(b)	(b)		(c)	(d)	
Plymouth	92	0.88	4.3	4.2	4.1	-4.65%	4.3	4.1	-4.65%
Barnstable	123	0.75	8.1	7.9	7.6	-6.17%	7.8	7.6	-2.56%
Cambridge	39	0.17	2.5	2.7	2.7	8.00%	2.5	2.7	8.00%
(a) Millions of square feet									
(b) GWh/yr - Calendar Years									
(c) GWh/yr - 1988/10:1989/9									
(d) GWh/yr - 1990/10:1991/9									

A survey of buildings in the control group was performed to determine if any activities undertaken in this group could influence evaluation results. The survey asked questions about changes in the buildings and about energy saving behaviors. The survey contained the following six questions:

1. Have energy saving measures/equipment been installed in either 1990 or 1991?
2. Have operating hours been reduced since 1990?
3. Has the heating temperature been reduced since 1990?
4. Has the cooling temperature been increased since 1990?
5. Has the heating fuel changed since 1990?
6. Have any other changes to reduce energy costs been made since 1990?

The survey covered 212 buildings in the complete control group of 264, of which 205 are in the sub sample S7. A tabulation of these survey results in Table 17 for sample S7 shows the number of respondents who answered yes to the question.

Table 17. Control Group Energy Survey Responses							
		Survey Question Number					
	N	1	2	3	4	5	6
Plymouth	85	25	4	25	5	2	8
Barnstable	90	36	12	49	13	6	11
Cambridge	30	8	0	10	1	2	4
	205	69	16	84	19	10	29

The effect of these measures on the change in annual consumption was examined with a regression model. The model results are summarized in Table 18 with binary variables Q_j representing a "yes" response to question j. The left most column presents the most general model considered for the control group. Columns to the right present more restricted models

with insignificant variables omitted. Surprisingly, the coefficient of Q1 is positive, indicating that energy measures/equipment have a mean effect of increasing consumption in Barnstable and Cambridge, while having only a negligible negative effect in Plymouth of -.01 to -.04. Similarly, the coefficient of Q2 is positive, also implying that reducing operating hours led to small (but statistically insignificant) increases in energy use. Reducing heating temperature (Q3) was the only measure which had a negative effect on consumption, significant at the .05 level. When taking this measure into account, the overall change in consumption is not significantly different from zero. However, increased consumption in Cambridge is evident in all of these models.

Table 18. Control Group Consumption Difference Model using Energy Survey Information					
Dependent Variable (all coefficients) = kWh/SQFT					
N	205	205	205	205	254
SSR	97.86	98.07	99.57	100.17	143.8
Adj R ²	0.0425	0.05	0.055	0.059	0.044
Hetero. Test	214.2	195.1	124.2	121.9	80.6
Constant	0.44	0.06	0.17	0.21	-0.42
t-statistic	0.56	0.12	0.372	0.483	-1.161
Plymouth	-0.51				
t-statistic	-0.58				
Cambridge	3.94	4.17	4.37	4.21	4.11
t-statistic	1.769	2.015	2.051	2	2.25
SQFT	-0.000022				
t-statistic	-0.696				
Q1	1.27	1.38			
t-statistic	0.673	0.861			
Q1 *	-1.36	-1.72			
Plymouth	-0.607	-0.938			
Q2	1.11	1.17	1.39		
t-statistic	1.331	1.402	1.817		
Q3	-2.11	-2.07	-1.69	-1.28	
t-statistic	-2.275	-2.147	-1.85	-1.669	
Q4	1.08	1.07748	0.92856		
t-statistic	1.178	1.185	1.068		
Q5	1.75	1.74			
t-statistic	0.919	0.9			
Q6	-0.49	-0.43			
t-statistic	-0.685	-0.615			

The final control group model with the survey questions accounted for is the second-to-last column in Table 18, where the energy use coefficient is positive and significant for Cambridge, positive but insignificant for the base level (which includes all buildings and means the base change was essentially zero), and the response to Q3, which shows that energy use decreased when respondents stated they reduced heating temperature. The coefficient for Q2 is not included in the model because, although it is significant, it indicates what appears to be an answer contrary to expectations, that energy use increased when operating hours were reduced. The net effect for the 74 non-Cambridge buildings with a "yes" response to question 3 is a 12.2% decrease in annual energy use (kWh per square foot). For the control group as a whole this represents a decrease of roughly .4 kWh per square foot since these 74 buildings account for less than half of total control group square feet. The results concerning the survey questions argue strongly against any presence of free riders in the program (basically what people say is different from what they do), and no free rider correction is reported for net savings following. (See also Evaluation Issues in the Introduction.)

Unfortunately, we do not have similar information for the treatment group. The control group results inform us that energy use did not decrease in the control group, except for buildings that reduced their heating temperature, but we do not know whether any similar action occurred for the treatment group. However, with energy saving retrofits installed that are expected to reduce future bills, the treatment group is less likely to perform such an action. The section on Building Specific Model Results indicates that buildings with electric heating behavior had much worse savings results than those without electric heating, which in addition to highlighting the potential savings problems for buildings with electric heat, also supports the assumption that the treatment group probably did much less of temperature reduction.

An alternative model for the complete control group (in sample S7) is presented as the right-most column in Table 18. This model approximates the energy use reduction in the control group by including a constant reduction in energy use for all buildings, with an increase from the constant base for Cambridge buildings. This model averages the energy use reduction for the control group across all buildings and yields a similar estimate of non-Cambridge change in kWh per square foot as the model that includes the Q3 coefficient. Comparing this model to the treatment group probably represents a worst case for the aggregate modeling approach for net energy savings.

NET ENERGY SAVINGS

A total net savings estimate can be computed by combining the treatment and control models formulated previously. As described previously, selection bias, free riders and free drivers are not expected to have much, if any, impact on net savings estimates for this comprehensive retrofit program. As a result, the estimates of net energy savings do not include additional adjustments to account either for free riders and self selection or for free drivers.

The pooled treatment/control model estimates are presented in Table 19. The equation for the pooled model is shown in Eqn [11]. The resulting ratio of costs to estimated annual benefits (investment cost) shown in the left-hand column at the bottom reflects an assumption of no change in the control group, which corresponds to the assumption that treatment group buildings outside of Cambridge did not reduce heating temperature during the evaluation period as the control group did. The second estimate appearing on the bottom in the right-

$$\begin{aligned}
 y_i = & \beta_0 + \beta_1 \text{CNTY4} \\
 & + [\alpha_{01} + \alpha_{11} \text{COST/SQFT} + \alpha_{21} \text{SQFT} + \alpha_{31} \text{COST}] \times (\text{TR}) \times (\text{CON1,2a}) \\
 & + [\alpha_{02} + \alpha_{12} \text{COST/SQFT} + \alpha_{32} \text{COST}] \times (\text{TR}) \times (\text{CON2b}) \\
 & + [\alpha_{03} + \alpha_{13} \text{COST/SQFT} + \alpha_{33} \text{COST}] \times (\text{TR}) \times (\text{CON3,4}) + e_i
 \end{aligned}
 \tag{11}$$

where

y_i = kWh/SQFT, average for each billing period

CNTY4 = 1 if Cambridge, 0 otherwise

TR = 1 if treatment group, 0 otherwise

CON1,2a, etc., designate contractors, (using 1s; 0s)

NOTE: α_{22}, α_{23} have t-statistics < 1

hand column is based on inclusion of the estimated reductions in energy use for the control group. These control group changes consist of a decrease of 0.42 kWh per square foot in non-Cambridge buildings and an increase of 4.11 kWh per square foot in Cambridge buildings.

The coefficients in Table 19 show change in kWh/sq.ft. for the sample based on a change in parameter such as COST/SQFT. The "Constant" coefficients are simple kWh/sq.ft., those for COST/SQFT are kWh/sq.ft. saved for each \$1.00 per sq.ft. invested, those for SQFT are change in kWh/sq.ft. for each 1,000 SQFT of building, and those for COST are kWh/sq.ft. saved per \$1,000 of total investment. Savings are positive and increases in energy use are negative. The unweighted R^2 for the regression is presented in the table, which shows the unweighted degree of fit attributable to the model coefficients.

Total kWh savings are calculated by first estimating the energy savings, in terms of kWh/sq.ft., for the pooled treatment group model and for a pooled model with both treatment and control group buildings included. The pooled treatment group only model is Eqn [6] and uses the pooled coefficients shown in Table 14. The combined model for both treatment and control group buildings uses the model parameters previously defined for treatment group (Table 14 and Eqn [7]) and control group (right-most column, Table 18), and refined model coefficients are estimated jointly from the pooled treatment and control groups.

The results for the control group demonstrate that the decrease in energy use for all control group buildings, including Cambridge, is not statistically significant (the constant term has a t-statistic of 1.168, which means this value is statistically distinguishable from zero with a confidence level of 0.75). The coefficient for Cambridge is highly significant and indicates a noticeable positive increase. Thus, different results are obtained for Contractor 2 when working in Cambridge compared to Contractor 2 working outside of Cambridge. These results indicate that the investment cost based on net savings is about \$0.55/kWh saved. The 95% confidence limits on the treatment group alone (N=1041) are \$0.51-0.58/kWh saved. The 95% confidence limits on the combined treatment and control groups (N=1295) are \$0.53-0.61/kWh saved.

However, these results are somewhat sensitive to two control group buildings that exhibited large increases in energy use during the evaluation period. Each of these two buildings had increases in energy use exceeding 10 W/sq.ft., and using the outlier criterion developed for the treatment group, these buildings could be considered outliers and deleted from the group. Deletion of these two buildings leads to an estimate of investment cost of \$0.64/kWh, which implies that the actual investment cost may be around \$0.60/kWh when consideration of control group behavior is added. Overall, the aggregate annual evaluation results indicate that the total program net investment cost is about \$0.60/kWh, and that the program was much more cost effective for smaller buildings and smaller total dollar investment per retrofit project than for larger buildings and larger total investments. The program can achieve significant savings, but special care must be taken to determine how to significantly improve performance for larger buildings and to prevent excessive investments (\$/sq.ft.) in specific buildings.

Simple screening criteria, such as checking estimated savings against total energy use, would appear to offer important benefits to consider. More thorough examination of some of the larger buildings that received retrofits also appears needed to allow better explanations to be developed of the causes for high savings estimates and degradation of program performance.

Table 19. Aggregate Treatment and Control Group Coefficients and Effects with Net Savings, Sample S7				
		Treatment	Control	Both
	N	1041	254	1295
	REG S.E.	0.694	0.756	0.707
	SSE	496.736	143.95	589.173
	R ²	0.112	0.048	0.135
	ADJ R ²	0.104	0.044	0.128
	Heterosk. Test	1134	81	1165
CONTROL	Constant		0.42	0.42
	t-statistic		1.161	1.161
	Cambridge		-4.11	-4.11
	t-statistic		-2.25	-2.25
CON = 1 or 2a	Constant	2.72		2.30
	t-statistic	6.226		4.029
	COST/SQFT	0.64		0.64
	t-statistic	3.53		3.53
	SQFT	-0.04		-0.04
	t-statistic	-2.08		-2.08
	COST	-0.13		-0.13
	t-statistic	-2.416		-2.416
CON = 2b Cambridge	Constant	-0.15		3.53
	t-statistic	-0.071		1.273
	COST/SQFT	2.98		2.98
	t-statistic	1.51		1.51
	COST	-0.24		-0.24
	t-statistic	-1.614		-1.614
CON = 3 or 4	Constant	2.30		1.87
	t-statistic	5.7		3.439
	COST/SQFT	2.04		2.04
	t-statistic	5.497		5.497
	COST	-0.39		-0.39
	t-statistic	-5.502		-5.502
N		1041		1295
TOTAL SAVINGS, kWh		6,939,524		6,697,871
TOTAL COST, \$		\$3,782,898		\$3,782,898
INVESTMENT COST, \$/kWh		0.545		0.565

ENERGY SAVINGS ESTIMATES — BUILDING SPECIFIC MODEL RESULTS

As discussed previously, the aggregate annual savings reduce the available data to two data points for each building, one pre- and the other post-treatment. To account for the effects of weather, PRISM develops building specific parameters of building behavior with respect to temperature and also aggregates this temperature domain data to two annual data points for each building. The basic strategy used in this study for developing and estimating building specific models is to include both temperature and retrofit savings parameters, as well as other important time dependent factors such as price and economic activity. These building specific models provide a check against the annual aggregate results presented previously to determine whether some refinement of the annual aggregate values is needed.

The first step in estimating this building specific class of models was to identify each building as one of the following types: Heating & Cooling, Cooling Only, Heating Only, and Constant, where Constant includes buildings with no systematic thermal load. The basic model used for this identification process is similar to the PRISM heating and cooling model (Fels 1991b), but with fixed reference temperatures. In order for a building to be identified as Both Heating and Cooling, coefficients must be positive and significant at the 90% level. Similarly, both the Cooling Only and Heating Only categories have a positive and significant load coefficient requirement. The remaining buildings are classified in the Constant category. The three different approaches to developing building specific models: fixed coefficient, fixed coefficient with pooling, and random coefficient were then applied to each of these four categories. For example, with the fixed coefficient approach, five building specific coefficients, as expressed in equation [3], are estimated for each building in the heating and cooling category. However, only two coefficients (α , α_c) are estimated for each building in the constant category.

Table 20 presents a breakdown of building types by contractor. The proportions of each building type are reasonably constant across contractors. Thus, differences in energy savings by building type probably does not account for noticeable differences in contractor performance. Surprisingly, some of the largest difference are between contractors 3 and 4, who both operated primarily in Barnstable County.

Table 20. Building Types by Contractor (row %)					
BUILDING TYPES	Heating & Cooling	Cooling Only	Heating Only	Constant	Total
CON = 1	0.23	0.46	0.21	0.10	229
CON = 2a	0.27	0.41	0.19	0.14	182
CON = 2b	0.14	0.54	0.25	0.08	80
CON = 3	0.22	0.37	0.27	0.13	416
CON = 4	0.16	0.46	0.26	0.13	135
SQFT	4147	3890	4661	4044	
Total	224	439	251	128	1042

The building specific modeling process begins once again with sample S7, defined in Table 5. The treatment group has 1050 buildings, and the control group has 254 buildings. Buildings are lost from S7 in the aggregate results presented earlier due to being classified as an outlier, leading to a final total of 1042. Buildings are lost from S7 for different reasons in the building specific modeling process, although the reduced sample is the same size as with the aggregate data. For the building specific process, 6 treatment buildings are deleted because the energy consumption is very low, often zero, and causes difficulty with the modeling process. Also, because the modeling process is sensitive to the actual date when the retrofits are installed, 2 additional treatment group buildings were deleted because they had no energy use data covering the date of the retrofit. Thus, the treatment group sample is 8 less than 1050, for a total of 1042 (Table 20). The control group has 2 buildings deleted because of very low energy consumption data, often zero, that causes difficulty with the modeling, so the control group total for this modeling is 252.

ABNORMALITY OF NOVEMBER/DECEMBER 1991

The building specific modeling process allows additional examination of the abnormality of the billing data for November and December 1991. Because this modeling process proceeds month by month, with cognizance of seasonality for specific building types, statistical measures of the months of November and December 1991 as compared to the previous three years of data can be developed. Table 21 shows how extensive the problem is with bills for November and December of 1991. Recall that previous aggregate results showed instability and abnormal distributions when these last two months of 1991 were included. For the time series data, with heating and cooling effects included, abnormal statistics are also shown for these two months. For example, 87% of the Cooling Only buildings show a significant increase for these two months. The table shows the number of buildings (and what percentage they represent) for each building thermal load type that would have the months November and December of 1991 rejected at the 95% confidence level for being significantly different from what model expectations would be for these months. Overall, these two months represent a three year maximum consumption for any time period for an alarming number of these buildings, all during a mild winter. Tests for abnormality indicate that these last two months should not be included in the building specific models based on the building specific monthly model, and thus these two months are excluded in the results presented here. (As before, later examination showed that these data were incorrect, see Appendix B.)

Table 21. Unconstrained Building Specific Model Types Matrix (Treatment Group Only - Sample S7)					
	Heating & Cooling	Cooling Only	Heating Only	Constant	Total
# BUILDINGS	224	439	251	128	1042
# REJECT	197	380	191	84	852
% REJECT	88%	87%	76%	66%	82%
N1	102	263	0	0	365
N2	80	80	221	114	495
N3	22	50	0	0	72
N4	20	46	30	14	110

where

REJECT = Number of buildings for which observations for 11/91 & 12/91 are significantly different from monthly model expectations at the 95% confidence level, based on previous three years of data

N1 = Number of building with estimated base and cooling savings.

N2 = Number of buildings with estimated base savings only.

N3 = Number of buildings with estimated cooling savings only.

N4 = Number of buildings with no estimated savings.

FIXED COEFFICIENT APPROACH

The second step in the building specific modeling procedure is to estimate the net savings and net investment cost ratio, using the models specified previously as model [3] and model [5]. First we examine results for the unconstrained model [3] (presented earlier as equation [3]). This model has five building specific parameters and no pooled parameters. The building specific parameters include adjustments for weather effects. Because this model assumes the form of changes in energy use, all coefficients are constrained to be positive. This means that the retrofit savings coefficients are constrained also (i.e., negative savings are not allowed). However, only 10% of the sample is affected by this constraint, and the estimated savings are still compared to total investment for all buildings. In addition, we compare those buildings which exhibit a change in energy use at the time of retrofit that is significant at the 90% confidence level to buildings in the control group.

The results of the modeling are shown in Table 22 by thermal type of building: Heating & Cooling, Cooling Only, Heating Only, and Constant. There are 252 control group buildings and 1041 treatment group buildings. Buildings that exhibit a change that is 90% significant between treatment and control group periods include 127 control group buildings and 808 treatment group buildings.

Please note that savings are presented differently in this table than in previous tables. Table 22 shows savings as decreases in energy use or as negative values. Increases are positive. This convention makes explanation of the control group easier. As a result, investment ratios (\$/kWh) are negative when a savings occurs.

Because all parameters are building specific, with no pooled coefficients, each treatment and each control group building is modeled separately. Results of the modeling are shown using several different comparisons between treatment and control buildings in Table 22. Two sections of control group results are shown, for ALL buildings and for buildings showing a change between 1989 and 1991 that is significant at the 90% level. Two savings values are also shown for ALL treatment group buildings and for 90% significant change buildings: aggregate annual results and building specific results. The 90% significant change results are presented because the model coefficients are constrained to be positive, so the 90% significant results provide a check against the overall results.

Readers are cautioned to examine the net savings for ALL buildings, shown above the heavy black lines, before studying the results for the 90% significant change values. The results for the 90% significant change buildings may be unfair in the handling of the control group, which limits the control group only to buildings with a significant change.

The abbreviations used in Table 22 are:

SQFT	- Building gross floor area, square feet
TSAV1	- Annual aggregate change in electricity use
TSAV2	- Modeled building specific change in electricity use
COST	- Total cost of installed retrofits
NSAV1	- Net annual aggregate change in electricity use, treatment minus control
NSAV2	- Net modeled building specific change in electricity use, treatment minus control

The aggregate annual results, which DO NOT INCLUDE THE EFFECTS OF WEATHER and are equivalent to the aggregate values reported in the last column of Table 6 for sample S7, show that Heating & Cooling buildings and Cooling Only buildings have a net investment cost ratio (COST/NSAV1, after control group adjustment) higher than \$0.70/kWh. Heating Only and Constant buildings each show an aggregate annual investment cost ratio of about \$0.45/kWh. These results for Heating Only and for Cooling Only buildings are counterintuitive, in that we would expect the opposite. These counterintuitive results are driven by the control group for these building types. The results for all buildings show an aggregate total investment cost of \$0.615/kWh.

However, when the effects of weather are included through building specific modeling (COST/NSAV2), the Heating & Cooling buildings are the only buildings shown to be higher than \$0.60/kWh. The building specific models included weather effects that were driving the energy use of the control group. The overall total investment cost, after weather correction, is \$0.51/kWh.

When we look at buildings which show a change after retrofit that is significant at the 90% confidence level, we also see similar results, but with higher net investment cost ratio. Overall, these results show that the program has marginal cost effectiveness performance, but this performance could be corrected. Previously, the negative effect of higher investment levels and higher total investment cost was demonstrated. The building specific model results also confirm the importance of heating and cooling on overall results. Many utilities have known over the years the importance of lighting impacts on electric heating requirements, and the results shown here indicate that savings are diminished when buildings have electric heating present. Examination of the TSAV2/SQFT values for ALL buildings shows the building specific weather effect on changes in kWh/SQFT, where the treatment group reduction is 1.8 kWh/SQFT for Cooling Only and Constant load buildings and is 1.4 kWh/SQFT for Heating & Cooling and Heating Only buildings. Thus, savings for buildings with electric heating appear to be about 20% less than for buildings without electric heating.

Table 22. Estimates of Building Specific Savings for Model 3					
	HEATING & COOLING	COOLING ONLY	HEATING ONLY	CONSTANT	TOTAL
CONTROL					
ALL BUILDINGS, N	50	114	69	19	252
SQFT	248039	436676	998880	103011	1786606
TSAV1, kWh	-133119	21836	-119202	-41989	-272474
TSAV2, kWh	-93784	139386	41502	-38517	48587
TSAV1, kWh/SQFT	-0.537	0.05	-0.119	-0.408	-0.153
TSAV2, kWh/SQFT	-0.378	0.319	0.042	-0.374	0.027
90% SIGNIFICANT, N					
ALL BUILDINGS, N	22	46	46	13	127
TSAV1, kWh	-135533	-113705	10675	-45630	-284193
TSAV2, kWh	-103499	11943	64311	-37360	-64605
TSAV1, kWh/SQFT	-0.546	-0.26	0.011	-0.443	-0.159
TSAV2, kWh/SQFT	-0.417	0.027	0.064	-0.363	-0.036
TREATMENT					
ALL BUILDINGS, N	224	440	250	127	1041
SQFT	928883	1708106	1168859	516677	4322525
COSTS, \$	787718	1567385	957312	423036	3735451
TSAV1, kWh	-1619698	-1791609	-2287565	-1141024	-683989
TSAV2, kWh	-1352570	-3236296	-1687680	-942241.4	-721878
COST/TSAV1, \$/kWh	-0.486	-0.875	-0.418	-0.371	-0.546
COST/TSAV2, \$/kWh	-0.582	-0.484	-0.567	-0.449	-0.517
TSAV1, kWh/SQFT	-1.744	-1.049	-1.957	-2.208	-1.582
TSAV2, kWh/SQFT	-1.456	-1.895	-1.444	-1.824	-1.67
NSAV1, kWh	-1121180	-1877022	-2148079	-930417.8	-607669
NSAV2, kWh	-1001358	-3781518	-1736245	-749052.2	-726817
COST/NSAV1, \$/kWh	-0.703	-0.835	-0.446	-0.455	-0.615

COST/NSAV2, \$/kWh	-0.787	-0.414	-0.551	-0.565	-0.514
90% SIGNIFICANT, N	175	307	213	113	808
TSAV1, kWh	-1576833	-2242184	-2282294	-1097500	-719881
TSAV2, kWh	-1310429	-2821593	-1668873	-924295.5	-672519
COST/TSAV1, \$/kWh	-0.5	-0.699	-0.419	-0.385	-0.519
COST/TSAV2, \$/kWh	-0.601	-0.555	-0.574	-0.458	-0.555
TSAV1, kWh/SQFT	-1.698	-1.313	-1.953	-2.124	-1.665
TSAV2, kWh/SQFT	-1.411	-1.652	-1.428	-1.789	-1.556
NSAV1, kWh	-1069274	-1797413	-2294786	-868629.3	-603010
NSAV2, kWh	-922836.6	-2868310	-1744128	-736907.8	-627218
COST/NSAV1, \$/kWh	-0.737	-0.872	-0.417	-0.487	-0.619
COST/NSAV2, \$/kWh	-0.854	-0.546	-0.549	-0.574	-0.596

NOTE: The most important results appear above the thick black lines.

The net savings values for Table 22 indicate that the control group adjustment for Heating Only buildings makes those buildings also look favorable. The control group adjustment for Cooling Only buildings leads to some important changes between the total sample and the buildings with a change that is 90% significant. The weather corrected results for ALL buildings show Cooling Only buildings to be exemplary performers, with a net investment cost of \$0.41/kWh. The control group for ALL buildings in this category has a significant increase in energy use for the evaluation period. For the buildings with a 90% significant change, the control group shows a much lower increase in energy use, so the Cooling Only buildings net investment ratio increases to \$0.55/kWh.

The lower portion of Table 21 shows the frequency with which the estimated coefficients were constrained to be non-negative (i.e., N4 in Table 21). Overall, 10% of the sample exhibited no month to month consistent energy savings. This fraction is fairly constant across the different building types. Of the buildings with an identified cooling load, approximately half exhibited both base and cooling related savings.

This modeling process shows the net investment cost ratio for the program is \$0.51/kWh, similar to the phase 2 evaluation results and the annual aggregate results presented earlier in Table 19. When only buildings with a change in energy use after retrofit that is 90% significant are examined, the net investment cost ratio is slightly over \$0.60/kWh. However, the results for the 90% significant change buildings are probably unfair in the handling of the control group, which limits the control group only to buildings with a significant change. This handling is offset to some degree by the constraint of model coefficients to all be positive. Overall, the two sets of results (ALL buildings and 90% significant change) present a bracket of potential results that shows savings are acquired for an investment in the range of \$0.55/kWh.

FIXED COEFFICIENT APPROACH WITH POOLING CONSTRAINTS

In this section the pooling constraints included in model [5] are imposed on the fixed coefficient estimates of the previous section. The primary objective is to express the building specific estimates of savings as a function of retrofit costs. If appropriate, this estimated relationship can be useful in developing more cost effective retrofit programs.

The basic model for the treatment group has three building specific parameters for base and heating and cooling load, as well as twelve pooled parameters. All retrofit energy savings effects are modeled with pooled parameters as a function of cost and size. In addition, the effect of other time dependent factors, unemployment and price, on base and heating and cooling loads are specified as a function of pooled parameters. The control group was modeled in a similar manner, which included a pseudo treatment effect to test for any systematic change in this group. These control group estimates provide a baseline by which to measure net energy savings for the treatment group.

When modeling N buildings over T periods as

$$y_{it} = f_{it}(\gamma_j, \beta_{ik}) + e_{it}, \quad e \sim N(0, \sigma_i^2) \quad [12]$$

where $\gamma_j, j=1, K_1$, are the pooled parameters

and $\beta_{ik}, k=1, K_2$, are the building specific parameters for building i

the total number of model parameters is $K_1 + NK_2$, the number of pooled parameters plus N times the number of building specific parameters. Because of the large number of parameters inherent in such an approach, an iterative two-step estimation procedure is employed based the general procedure proposed by Oberhofer and Kmenta (1974). Starting with initial estimates of the pooled coefficients, a regression model is used to estimate the building specific parameters for each building. Using the building specific coefficient estimates, the pooled coefficients are then estimated for all buildings as a weighted non-linear regression model with the building specific coefficients held constant. For this pooling procedure, the variances from the building specific regressions, σ_i^2 are used as weights. The process is iterated until the pooled parameter estimates change by less than a specified tolerance level, usually 0.01%.

Ideally the degree of pooling is strictly an empirical issue. However, data availability and computational resources are also issues when modeling three years of monthly bills for over a thousand buildings. For these reasons, equation [5] is the most general pooled specification considered. For a sample of N heating and cooling type buildings this pooling constraint reduces the number of estimated coefficients

$$\begin{aligned} &\text{from } 5N \quad (N(\alpha, \alpha_r, \beta_h, \beta_c, \beta_r)) \\ &\text{to } 12 + 3N \quad (\alpha_2, \alpha_3, \beta_{h2}, \beta_{h3}, \beta_{c2}, \beta_{c3}, \gamma_1, \gamma_2, \gamma_3, \eta_1, \eta_2, \eta_3 + N(\alpha_1, \beta_{h1}, \beta_{c1})). \end{aligned}$$

The initial results for this model, which includes both treatment and control group buildings together, showed model coefficients with positive price coefficients, which is a problem. Positive price coefficients mean that electricity use increases when the electricity price increases, which is not logical for a commodity such as electricity. This problem is likely due to the endogeneity problem of higher prices in the form of fuel cost adjustments resulting from higher consumption. This particular specification with a price variable calculated as average revenue was therefore not examined further. Unfortunately, it was not possible obtain each individual building's pricing schedule history. Model 5 was modified to exclude PRICE as a variable. We then found that unemployment (UNEMPLOY) was no longer a significant variable in either the base load or thermal load components when PRICE was removed.

Elimination of the pooled parameters, PRICE and UNEMPLOYMENT, means that effects of the control group cannot be estimated using model [5] directly in that there are no common parameters. In which case, the parameters (model coefficients and variances) for the two groups of treatment and control are more easily estimated separately.

Of special note, after the PRICE parameter was dropped because of incorrect sign on the coefficient, the UNEMPLOYMENT parameter was dropped because it was no longer significant. However, the standard error estimates used to estimate significance in model [5]

were initially incorrect and therefore only a rough approximation (see Appendix A for discussion of error estimates for Model [5]). However, using the error estimating procedure described in Appendix A, PRICE and UNEMPLOYMENT were still dropped from the model.

The control group results, using the Appendix A error estimating procedure, indicated that NO CHANGE IN ENERGY USE could be detected that is significantly different from zero under the pooling constraints. This finding means that the change in the treatment group is the only result needed to examine results from Model [5].

Initial results for the pooling of the treatment group using Model [5] are presented in Table 23. Again in this table, decreases in energy use, or savings, will have negative values, and negative investment cost means that energy savings were acquired. The correct standard errors were NOT calculated for these initial results, so the t-statistics are high by a factor of 5-10. Using this initial model, the overall ratio of cost to net energy savings is considerably higher than in earlier estimates. As with the previous unconstrained estimates, the two non-heating categories are the most cost effective. However, PRICE still shows a positive coefficient and must be removed.

Model [5] was constrained to remove PRICE and UNEMPLOYMENT. The results are shown in Table 24. The coefficients for COST/SQFT under Base Savings for Heating Only and Constant buildings are not significant, and the coefficient for COST/SQFT under Base Savings for Cooling Only buildings is not significant. However, removal of these parameters from the model does not change the bottom line investment cost much, and retaining them here shows their relative importance.

As with Table 19, the coefficients in Table 24 show change in kWh/sq.ft. for the sample based on a change in parameter such as COST/SQFT. The "CONSTANT" coefficients are simple kWh/sq.ft., those for COST/SQFT are kWh/sq.ft. saved for each \$1.00 per sq.ft. invested, and those for COST are kWh/sq.ft. saved per \$1,000 of total investment. The unweighted R^2 for the second stage nonlinear regression is presented in the table, which shows the unweighted degree of fit attributable to the pooled coefficients.

These results exhibit strong retrofit savings behavior, with the buildings that have some form of heating leading to significant reductions in overall performance. The pooled results are important because they are based on building specific models and include the COST/SQFT and (total) COST parameters in the model. Thus, we can infer results from this pooled model to the general population of possible program participants. The pooling employed in this model may be overconstrained, but the basic result is the same as before: the program offered in 1990 was operating very near the margin of \$0.60 invested per kWh saved.

TEST OF POOLING CONSTRAINTS

Two different tests of this pooling specification, a Likelihood ratio test and a Wald test proposed by Swamy, are computed for the overall sample. A Lagrange multiplier test is also computed for each building, to identify for which buildings the hypothesis is clearly rejected. In addition, specification [5] is further examined with a series of nested hypotheses tests to identify a more restricted specification which is not rejected by the data.

Table 23. Initial Pooling Constraint Coefficients — Model 5 — Treatment Group Only						
NOTE: t-statistics are INCORRECT in this table (see text)						
		Heating & Cooling	Cooling Only	Heating Only	Constant	Total
Buildings, N		224	440	250	127	1041
Monthly values, N		7745	15064	8674	4356	
Regression SE		0.0611	0.109	0.035	0.032	
Weighted R ²		0.995	0.999	0.996	1	
BASE	PRICE	2.93	5.45	1.53	0.78	
	t-statistic	5.73	19.11	9.31	5.95	
	UNEMPLOY	-0.02	-0.07	-0.02	-0.01	
	t-statistic	-2.53	-19.49	-7.14	-3.84	
HEATING	PRICE	0.04		0.01		
	t-statistic	1.49		1.03		
	UNEMPLOY	0.00		0.00		
	t-statistic	-2.15		0.05		
COOLING	PRICE	0.00	0.33			
	t-statistic	-0.03	3.49			
	UNEMPLOY	0.00	-0.01			
	t-statistic	-1.37	-5.35			
BASE SAVINGS	CONSTANT	-1.25	-0.57	-1.46	-2.32	
	t-statistic	-23.95	-18.34	-35.02	-33.67	
	COST/SQFT	-0.01	-0.15	-0.11	-0.02	
	t-statistic	-0.32	-9.45	-5.40	-0.55	
	COST	0.05	0.04	0.04	0.06	
	t-statistic	7.07	7.27	7.14	6.06	
	CONSTANT	-0.88	-1.18			
	t-statistic	-4.75	-10.82			
COOLING SAVINGS	COST/SQFT	-0.23	-0.26			
	t-statistic	-1.91	-5.05			
	COST	0.02	0.03			
	t-statistic	0.95	1.44			
COST, \$		\$787,718	\$1,567,385	\$957,312	\$423,036	\$3,735,451
Savings, kWh		-787,179	-2,476,190	-884,305	-706,303	-4,853,977
Investment Cost, \$/kWh		-1.00	-0.63	-1.08	-0.60	-0.77

Table 24. FINAL Pooling Constraint Model Coefficients — Model 5b Treatment Group Only						
		Building Thermal Model Type				Total
		Heating & Cooling	Cooling	Heating	Constant	
Buildings, N		224	440	250	127	1041
Monthly Values, N		7745	15064	8674	4356	
SE STAGE 1		0.073	0.121	0.042	0.038	
Unweighted R ²		0.285	0.246	0.286	0.344	
BASE SAVINGS	CONSTANT	-.169	-0.079	-.185	-.265	
	t-statistic	-30.41	-24.02	-35.37	-30.32	
	COST/SQFT	0.016	0.003	-0.002	-0.001	
	t-statistic	3.87	6.62	-1.23	-0.22	
COOLING SAVINGS	CONSTANT	-0.030	-.134			
	t-statistic	-1.20	-10.28			
	COST/SQFT	-0.034	-0.031			
	t-statistic	-2.10	-6.59			
	COST	-0.007	0.003			
	t-statistic	-2.11	1.36			
Cost, \$millions		\$0.79	\$1.57	\$0.96	\$0.42	\$3.74
Savings, GWh		-1.02	-3.18	-1	-0.79	5.83
Investment Cost, \$/kWh		-\$0.77	-\$0.49	-\$0.96	-\$0.54	\$0.64

The pooling constraints imposed for this model may be misspecified, because the two different tests gave different results regarding these constraints. The two tests used were Likelihood ratio test and Wald test. The results of these tests are shown in Table 25. Comparing the likelihood ratio test to the CHI(.99) value shows that we would clearly reject the pooling hypothesis at the 99% confidence level for all building types. This test statistic, which can be expressed as

$$\sum_{i=1}^N T_i \left(\ln \left(\sum_{j=1}^{T_i} \bar{e}_{ij}^2 \right) - \ln \left(\sum_{j=1}^{T_i} e_{ij}^2 \right) \right),$$

where $\bar{e}$ are the residuals from the constrained model, is essentially a comparison of model fit.

Table 25. Test Results of Pooling Constraints on Treatment Effects				
	Heating & Cooling	Cooling Only	Heating Only	Constant
No. of Buildings	224	440	250	127
Unrestricted Likelihood Value	4082	-735	4997	1596
Likelihood Ratio Test	1908	3716	2146	1023
WALD Test	190	154	189	188
Degrees of Freedom	442	874	247	124
CHI(.90)	480	928	276	144
CHI(.99)	513	973	301	163

However, the Wald test only rejects the pooling hypothesis for the building category denoted Constant. This test is based on whether the building specific treatment effects are significantly different from those predicted by the pooled model determined by cost and square feet. The two tests are therefore not strictly comparable in that this Wald test is not based on differences in non-treatment effect coefficients. In addition, the Wald test is likely to be more robust to some forms of model misspecification. If for example, the residuals were correlated across buildings, then the likelihood functions would be misspecified, whereas, the Wald test would still be valid.

The mixed results on the pooling constraint tests show that the model may be overconstrained, but since the pooling constraints passed one test reasonably well and the results are reasonably consistent with our other results, we find the results to provide corroborating information. These fixed coefficient with pooling constraint results again show that the program was operating in the neighborhood of \$0.60/kWh saved and that the program would likely achieve that level of savings if continued as it was.

Further, based on these pooling test results, we surmise that pooling constraints as used by Train (1987) would not be likely to pass either of the tests used here. Of major importance, we specifically tested whether all buildings could be constrained to be the same thermal type (i.e., all buildings pooled together to determine overall sample response to heating and cooling degree days). Train (1987) used this type of pooling, but our tests overwhelmingly rejected the validity of this type of pooling constraint. Thus, results presented previously by evaluators that involve significant sample pooling may have erroneous results if the pooling constraints bias results.

RANDOM COEFFICIENT MODEL

A random coefficient model was also developed to estimate savings, but the results were found to be variable and inconsistent, indicating that the requirements of Equation [9] are probably violated. Thus, the random coefficient model was not applied successfully to these data, and no results are presented.

CONCLUSIONS AND RECOMMENDATIONS

The energy savings for the small commercial/industrial (C/I) program of COM/Electric have been evaluated using several methods. Overall, the impact evaluation results indicate that the net program investment cost ratio, the ratio of total costs to net electricity savings, is close to \$0.60/kWh when weather and cost factors are included. Economic factors were included in savings estimate calculations, but the results indicated either data endogeneity or high variance problems, so economic effects cannot be inferred as being estimable with the methods used here. Savings estimates based on simple, aggregate annual results (which are typically used for impact evaluations of C/I programs), indicate that this investment ratio is about \$0.55/kWh. This estimate is similar to the previous aggregate results reported in the phase 2 evaluation. The building specific model that corrects for weather (model 3) also indicates that the net investment cost ratio is about \$0.55/kWh (\$0.51-\$0.60/kWh). However, coefficients were constrained to be positive in this model, and this constraint probably lowered the cost ratio (perhaps 10%). The building specific model with pooled coefficients indicated that the investment cost ratio is \$0.64/kWh, but the pooling constraints for this model may have been too restrictive. Thus, several modeling approaches have indicated that the net investment cost ratio for this program is about \$0.60/kWh. Cooling benefits were modeled directly, while heating penalties seem to be apparent in the bottom-line first-year investment cost.

The sample of treatment group buildings for this impact evaluation is large, with the initial pool of about 1,700 treatment facilities representing about 85% of the approximately 2,000 facilities treated under the program during the period August - December 1990. Attrition of the treatment pool due to data gaps and change of customer meter reduced the final sample to about 50% of all buildings treated during the period of interest. The control group was selected from facilities that are on the waiting list for this program, to improve comparability and overall results. The control group is much smaller, with about 260 facilities, but little attrition was experienced.

The evaluation used models that account for weather effects, impacts of different contractors, effects of location of facilities, impacts of facility size and retrofit investment level per facility, and examined potential effects of electricity price and the overall economy (using unemployment as a surrogate for the overall economic climate). The evaluation indicates areas for program improvement, especially the need for screening larger jobs to determine that predicted savings are not above some threshold level of current electricity use. The program could also benefit from field examination of facilities where the total investment and the investment level per square foot were high, to better understand causes of poor performance of these buildings.

Survey results for a control group, together with actual energy use data, were used to determine whether customer responses could be used as indicators of actual changes in energy use. The control group was selected from waiting lists for the small C/I program, so the customers should be comparable to those in the treatment group. Analysis results show that customer responses to the survey questions are **not reliable indicators of actual energy reduction impacts**. The analysis results are empirical evidence that free rider effects may not be present, because even though potential future program participants stated that they had installed energy saving measures, no corresponding energy use reduction could be found. Also, the energy use of the control group serves as a correction for free riders.

The only reliable indicator of reductions in energy use found for the control group survey was the response that heating temperatures had been reduced, which is not related to the types of lighting retrofits installed. The empirical evidence indicates that, if participants stated that they would install some measures without participating in the program, the actual effect for the control group was that no savings occurred. This issue is complicated, and further testing of these results are needed to extend these findings to other assessments of free riders. However, based on our empirical results for this impact evaluation, the free rider effect for this program is estimated to be zero.

We conclude that the net investment cost (total \$ invested, divided by net kWh saved in the first year) is about \$0.60/kWh. Net kWh (sometimes called "adjusted") is for treatment group impacts minus changes observed in the control group. The control group is selected from waiting lists for the program, so that possible bias related to comparability between groups is minimized as best possible.

Program performance is at the margin of acceptable performance, and because of this marginality, improvements in program execution are desirable. Program performance at the margin could be improved significantly by controlling for potentially poor retrofit choices. These poor choices are indicated by estimates of savings that exceed 50% of pre-retrofit electricity use. Whatever the cause of these high estimates, a simple check of actual consumption before approving larger jobs should be instituted. Overall program performance could also apparently be improved by instituting increased checks on job approval in cases where the invested \$/sq.ft. or total dollars exceeds certain cutoff criteria. Although Figs. 6 and 7 of this report suggest that these cutoff criteria may be known, the regression models provide only trend information. We can suggest that whenever investment costs exceed \$1⁰⁰/sq.ft., that job approval should be more meticulous. However, the exact level of this cutoff probably requires more careful field and engineering examination of the jobs that exceeded this cutoff. About half of the jobs in our sample during the period of August - December 1990 exceeded this limit, so better definition of a reasonable cutoff value might help limit program cost increases for administration.

Significant difficulty with installing cost effective retrofits in larger buildings is found. The term "larger" means that some larger buildings had poor performance. Buildings that were 5,000-10,000 sq.ft. in size appear to be the major cause of overall poor program performance. Buildings that were larger than 10,000 sq.ft. had reasonably good performance. Field examination of some of the larger facilities where retrofits were installed is recommended. A field study of a sample of these facilities should be focused on probable causes of poor savings performance. Until the causes of poor performance in larger buildings are understood better, program efforts in any buildings over 5,000 square feet should be approved only after reasonable study.

The modeling results with different models indicate that annual aggregate impact evaluation results, which are the easiest to generate, of lighting retrofits in small C/I buildings probably provide adequate results compared to multivariate modeling of impacts, provided the cost effectiveness tests do not show marginal results. In other words, as would be expected by many evaluators, weather and economic factor correction is probably not needed to confirm overall cost effectiveness in cases where cost effectiveness performance is about 1.2 times as good as required minimums. For example, if a levelized cost of conserved energy of \$0.12/kWh were required to meet a cost effectiveness test, annual aggregate impact results estimating program impacts at \$0.10/kWh would likely be adequate in many cases for stating

that weather and other corrections are not needed to ensure program cost effectiveness. The 1.2 value is **approximately determined** based on results generated for this study, and only offered as an approximate answer for bounding potential further weather-corrected analysis. Future work by others could consider examining this bounding value.

Regardless of whether simple aggregate calculations provide adequate bottom line results, multivariate models are still valuable for developing an understanding of the factors affecting program performance, as they provide a direct estimate of the impacts of specific cost and building factors that may affect program performance. Further work with these types of models on other evaluation data sets would be useful for improving our knowledge of the best types of models to use and what important evaluation design factors should be considered when performing an evaluation.

The methods presented here provide some useful methodological insight. Further work could also be conducted to more completely test these methods. The importance of using the correct error estimates on building energy use data sets was highlighted by this study. Specifically, failure to account for heteroscedasticity when evaluating building energy use data sets over time will likely lead to incorrect selection of significant parameters, which can lead to biased (and incorrect) results. Further work is needed to better understand the methods used in this study, and if incentives for program performance depend on reasonably close estimates of energy savings, careful modeling with several methods is probably necessary.

Our test of pooling constraints indicates that evaluators must be careful when using pooled models. Specifically, tests of the validity of the pooling constraints should be performed and reported when presenting evaluation results involving pooled models. Building thermal type, similar to those used here or as used by Fels (1991a), should be an important sample stratification parameter when estimating building specific models.

REFERENCES

- Amemiya, T. (1983). "Nonlinear regression models," in Handbook of Econometrics Vol I, edited by Z. Griliches and M. Intriligator, pp 333-89 Amsterdam, North-Holland
- Dubin, J. and D. McFadden (1984). "An econometric analysis of residential electrical appliance holdings and consumption," Econometrica 52, 345.
- Fels, M. (1986). "PRISM: An introduction," Energy and Buildings, 9(1 & 2).
- Fels, M. and C. Reynolds (1991a). "Toward standardizing the measurement of whole-building energy savings in DSM programs." Energy Program Evaluation: Uses, Methods, and Results, Proceedings of the 1991 International Energy Program Evaluation Conference, pp 75-85. Conference coordinator, Gail Ettinger, 309 Davis St, Evanston, IL 60201.
- Fels, M. et al (1991b). Intermediate HC PRISM (Draft) Center for Energy and Environmental Studies, Princeton Univ.
- Hisao, C. (1986). Analysis of Panel Data, Cambridge: Cambridge University Press.
- Judge, G., W. Griffiths, R. Hill, H. Lutkepohl, and T. Lee (1985). The Theory and Practice of Econometrics. New York: Wiley.
- Miller, M. et al. (1992). "Development of measured savings for a small commercial/industrial lighting program," Proceedings of the ACEEE 1992 Summer Study on Energy Efficiency in Buildings, 7: p 7.151.
- Morris, K. (1991). "Comm/Elec suspends commercial/industrial program." Demand-Side Monthly (December). 747 Leigh Mill Rd, Great Falls, VA 22066.
- Mundlak, Y. 1978 "Models with variable coefficients: integration and extension," Annales de l'INSEE 30-1:483-509
- Oberhofer, W. and J. Kmenta (1974) "A general procedure for obtaining maximum likelihood estimates in generalized regression models," Econometrica, 42: 579-590
- Prahl, R. (1992). "In speaking with ..., 'Center Focus, DSM on Stage, 1(3), p 3. Newsletter of Wisconsin Center for Demand-Side Research, 595 Science Drive, Ste A, Madison, WI 53711-1060.
- Swamy, P. A. V. B. (1970). "Efficient inference in a random coefficient regression model," Econometrica 38: 311-23.
- Swamy, P. A. V. B. (1974). "Linear models with random coefficients," in Frontiers in Econometrics, edited by P. Zarembka, pp 143-68. New York: Academic Press.
- Train, K., and P. Ignelzi (1987). "The economic value of energy-saving investments by commercial and industrial firms." Energy 12: 543-53.
- Violette, D. et al. (1991). Impact Evaluation of Demand-Side Management Programs, Vol. 1: A Guide to Current Practice, EPRI CU-7179. Palo Alto: Electric Power Research Institute.
- Zellner, A. (1962). "An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias," Journal of the American Statistical Association 57: 348-68.

APPENDIX A **Error Estimates for Model [5]**

An important issue with the iterative estimation approach is the problem of estimating the standard errors of the model's parameters. The usual covariance estimator for this non-linear heteroscedastic model is

$$[G' \Omega^{-1} G]^{-1}$$

where Ω is the NT x NT error covariance matrix
and G is the NT x $(K_1 + K_2 N)$ matrix of partial derivatives

which is computationally difficult to evaluate when N is reasonably large. Since the standard error estimates of the pooled parameters are of primary interest, an alternative, computationally simpler, method is employed. By treating the building specific parameters as an implicit function of the pooled parameters the dimension of the equation above reduces to $K_1 \times K_1$ with the elements of G expressed as

$$\frac{\partial f_{i,t}(\gamma, \beta(\gamma))}{\partial \gamma_j} = \frac{\partial f_{i,t}(\gamma, \beta)}{\partial \gamma_j} + \frac{\partial f_{i,t}}{\partial \beta_{i,1}} \frac{\partial \beta_{i,1}}{\partial \gamma_j} + \dots + \frac{\partial f_{i,t}}{\partial \beta_{i,K_2}} \frac{\partial \beta_{i,K_2}}{\partial \gamma_j}$$

where $\gamma_j, j=1, K_1$, are the pooled parameters
and $\beta_{i,k}, k=1, K_2$, are the building specific parameters for building i

An approximation of this expression is obtained numerically by varying γ_j and then re-estimating the building specific parameters to evaluate the total change in $f_{i,t}$. The pooled parameter covariance matrix is then estimated as

$$[H]^{-1}$$

$$\text{where } H_{j,k} = \sum_{i=1}^N \sum_{t=1}^T \frac{\partial f_{i,t}}{\partial \gamma_j} \frac{\partial f_{i,t}}{\partial \gamma_k} \frac{1}{\sigma_i^2}$$

APPENDIX B
Abnormality of Data for Nov. and Dec. 1991

We requested new billing data for the treatment buildings in the phase 2 evaluation that extended from the middle of 1991 through the middle of 1992. The exclusion of data for November and December 1991 from the evaluation needs further explanation. We were able to obtain data from 1064 of the accounts. The table below shows the difference between the data we received for the phase 2 evaluation and the later data obtained from COM/Electric for the same customers. The differences for November and December 1991 are dramatic. The top part of the table shows the raw values for the cases where we received billing data for a particular month (N indicates how many meters had valid data). The bottom part of the table shows the differences by month (positive values indicate that the original data were high by that amount). As can be seen, the data for January 1992 are good, with less than one-half percent difference, but the data check for November and December 1991 indicated that the original data were almost double what current computer records indicate.

Comparison of Initial and Billing Data from Early 1993					
Date	N	Initial kWh	Early '93 kWh	Delta kWh	Delta Percent
Oct-91	1,028	3,401,214	3,271,489	129,725	4%
Nov-91	995	5,598,681	2,934,054	2,664,627	91%
Dec-91	1,019	6,311,287	3,181,105	3,130,182	98%
Jan-92	1,040	3,362,198	3,366,761	-4,563	0%

A quick examination of the impact on the evaluation was performed using aggregate statistics for the pre- and post-retrofit years of 1989 and 1991 as compared to Oct 1988–Sept 1989 and Oct 1990–Sept 1991. Almost no difference was shown in overall aggregate statistical results. The lack of difference is likely due to the fact that heating season penalty occurs during these months, and the possible extra benefits from having retrofits installed is cancelled by the loss in savings that occurs for buildings with electric heating during this time of year. We did NOT redo all our analyses using the new data because the significant time and expense involved did not appear justified based on the quick analytical check.

APPENDIX C **Building Specific Models With Estimated Reference**

Temperatures and Constrained Retrofit Effects

Both heating and cooling reference temperatures are estimated for the general model

$$WS(i,t) = A(i) - BLR(i) + CL(i)*CD(t,CREF) - CLR(i)*CD(t,CREF) + HL(i)*HD(t,HREF)$$

$$A > 0, BLR \geq 0, CL \geq 0, CLR \geq 0, HL \geq 0$$

where

WS = WATTS / SQFT

A = Base Load

BLR = Base Load Retrofit Effect

CL = Cooling Load

CLR = Cooling Load Retrofit Effect

HL = Heating Load

CREF = Cooling Reference Temperature

HREF = Heating Reference Temperature

The reference temperatures are varied in one degree increments so as to maximize the models likelihood value. The results savings estimates are presented in the following table.

Estimated Reference Temperature Results			
	Heating &	Cooling	Heating
	Cooling	Only	Only
Total Buildings	224	440	221
Total Cost	787718	1567385	957312
Estimated Savings			
Fixed Ref Temp	1352570	3236297	1687680
Estimated Savings			
Estimated Ref Temp	1368080	3281979	1686473

Test of Fixed Reference Temperature

The following two different test statistics were computed

T1 - Number of Building for which the hypothesis: $CREF = 65$, $HREF = 60$, is rejected at the 90 % Level with a Likelihood ratio test

T2 - Likelihood ratio test that: $CREF = 65$, $HREF = 60$ for all N buildings

The first test statistic, T1, is based on an individual test for each building of the fixed reference temperature hypothesis. The total number of rejections is denoted T1. The second test statistic is a test of the joint hypothesis that all individually estimated reference temperature are not significantly different from the fixed reference temperatures. The results presented in the following table indicate that the fixed reference temperature hypothesis is most strongly rejected for Cooling Only buildings. However, the resulting savings show above appear to be fairly insensitive to the choice of reference temperatures.

Test of Fixed Reference Temperature Hypothesis			
	N	T1	T2
Heating & Cooling Load Buildings, $CL > 0$, $HL > 0$			
$BLR(i) > 0$, $CLR(i) > 0$	102	14	247.6
$BLR(i) > 0$, $CLR(i) = 0$	85	9	177.3
$BLR(i) = 0$, $CLR(i) > 0$	29	2	65.0
Cooling Only Load Buildings, $CL > 0$, $HL = 0$			
$BLR(i) > 0$, $CLR(i) > 0$	264	123	925.1
$BLR(i) > 0$, $CLR(i) = 0$	87	27	205.6
$BLR(i) = 0$, $CLR(i) > 0$	71	33	222.3
Heating Only Load Buildings, $CL = 0$, $HL > 0$			
$BLR(i) > 0$, $CLR(i) = 0$	221	15	204.2

INTERNAL DISTRIBUTION

1	L.G. Berry, 4500N, MS-6206	14	M. Schweitzer, 4500N, MS-6206
2	M.A. Brown, 4500N, MS-6206	15	R.B. Shelton, 4500N, MS-6187
3	R.S. Carlsmith, 4500N, MS-6188	16	M.P. Ternes, 3147, MS-6070
4	G.E. Courville, 3147, MS-6070	17	B.E. Tonn, 4500N, MS-6207
5	T.R. Curlee, 4500N, MS-6205	18-26	D.A. Trumble, 4500N, MS-6205
6	W. Fulkerson, 4500N, MS-6247	27	D.P. Vogt, 4500N, MS-6205
7	L.J. Hill, 4500N, MS-6205	28	D.L. White, 4500N, MS-6206
8	E. Hirst, 4500N, MS-6206	29-107	Section Distribution, 3147, MS-6070
9	M.A. Kuliasha, 4500N, MS-6189	108	ORNL Patent Office 4500N, MS-6258
10	J.M. MacDonald, 3147, MS-6070	109	Central Research Library, 4500N, MS-6191
11	W.R. Mixon, 3147, MS-6070	110-112	Laboratory Records (3), 4500N, MS-6285
12	D.E. Reichle, 4500N, MS-6253	113	Laboratory Records - RC, 4500N, MS-6285
13	A.C. Schaffhauser, 4500N, MS-6186		

EXTERNAL DISTRIBUTION

- 114 D.R. Bohi, Director, Energy and Natural Resources Division, Resources for the Future,
1616 P Street, N.W., Washington, DC 20036
- 115 T.E. Drabek, Professor, Department of Sociology, University of Denver,
Denver, CO 80208-0209
- 116-127 R.M. Hardy, COM/Electric, 2421 Cranberry Highway, Wareham,
MA 02571-1091
- 128 C.D. MacCracken, President, Calmac Manufacturing Corporation,
101 West Sheffield Ave., P.O. Box 710, Englewood, NJ 07631
- 129 Ralph Nader, Post Office Box 19367, Washington, DC 20036
- 130 Office of Assistant Manager for Energy Research and Development,
DOE Oak Ridge Field Office, P.O. Box 2008, Oak Ridge, TN 37831-6269
- 131-132 OSTI, U.S. Department of Energy, P.O. Box 62, Oak Ridge, TN 37831
- 133 J.B. Shrago, Director, Office of Technology Transfer,
Vanderbilt University, 405 Kirkland Hall, Nashville, TN 37240
- 134 G.F. Sowers, P.E., Senior Vice President, Law Companies Group, Inc.,
114 Townpark Drive, Suite 250, Kennesaw, GA 30144-5599
- 135 C.M. Walton, Ernest H. Cockrell Centennial Chair in Engineering and Chairman,
Department of Civil Engineering, College of Engineering, The University of Texas,
Cockrell Hall, Suite 4.2, Austin, TX 78712

DATE

FILMED

10 / 31 / 94

END