Massachusetts Electric and Gas Program Administrators

Top-down Modeling Methods Study—Final Report

March 31, 2015

Prepared by:

DNV GL and NMR

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. EXECUTIVE SUMMARY

The work presented in this document is part of a multi-year initiative designed to assess the utility of top-down modeling as a viable technique for evaluating energy efficiency programs in Massachusetts. This document presents a summary of the Year 1 investigation into possible top-down methods for net impact evaluations, as a supplement to techniques currently used. Top-down techniques use a holistic approach by estimating program impacts across all energy-efficiency programs in a given geographical region or service territory, rather than running separate studies for each program (or measure/end-use within a program). Top-down models attempt to measure changes in energy consumption over time that are attributable to programmatic interventions by the utilities. The goal of this type of modeling is to isolate the effect of program activity from other natural changes and policy variables.

1.1 STUDY COMPONENTS

This Year 1 research included the following elements:

- *Assessment of top-down modeling methods.* Section 4 of this document discusses advantages, disadvantages, and necessary properties of top-down methods, and reviews 15 top-down research studies that were used to estimate impacts associated with energy efficiency programs. Based on this literature review, we developed recommendations for specific methods to be used in Massachusetts.
- *PA-Municipal utility pilot study.* Section 5 of this document summarizes the Year 1 research concerning the first pilot study. This PA-Muni pilot study contrasted changes in consumption in the residential and C&I sectors relative to programmatic activity, and compared results for the PAs and municipal utilities. This model relied on aggregate consumption and program expenditure data at the PA or municipal utility level. Some preliminary models from the pilot study were able to generate statistically significant and substantial net savings estimates. In principle, the models provide estimates of program-attributable savings, net of naturally occurring reductions to energy consumption, by contrasting consumption in the PA territories (where there has been substantial programmatic activity) to that in municipal utility territories (where there has been minimal to no programmatic activity), while controlling for other differences between the muni and PA territories.
- *PA Data pilot study.* Section 6 of this document discusses considerations for and preliminary results from the PA Data pilot study C&I models. For this study, we used PA-provided consumption and program tracking data to construct a set of C&I top-down models.[1] In contrast to the PA-Muni model, this study relied on account-level billing and program tracking data provided by the PAs rather than data already aggregated to the PA level. This allowed the model evaluation team to consider different levels of aggregation of the data, flexibility in the measures of programmatic activity, and use of weather normalization techniques to attempt to remove weather-related consumption changes from the models. Another fundamental difference

[1] The PA Data pilot study for the residential sector will be provided in 2015, as the three years of consumption data required for modeling is not yet available for analysis.

between the PA Data pilot study and PA-Muni pilot study, is that rather than capitalizing on the lack of programs in municipal territories, the PA Data model capitalized on variations in the level of program activity across different portions of PA service territories. Model estimation in this stage of evaluation did not attempt to estimate net savings using these models because the evaluation team did not believe the study would produce a stable model with the limited data series. Rather, the pilot study investigated whether the data provided evidence of sufficient signal between programmatic activity and consumption to warrant further study.

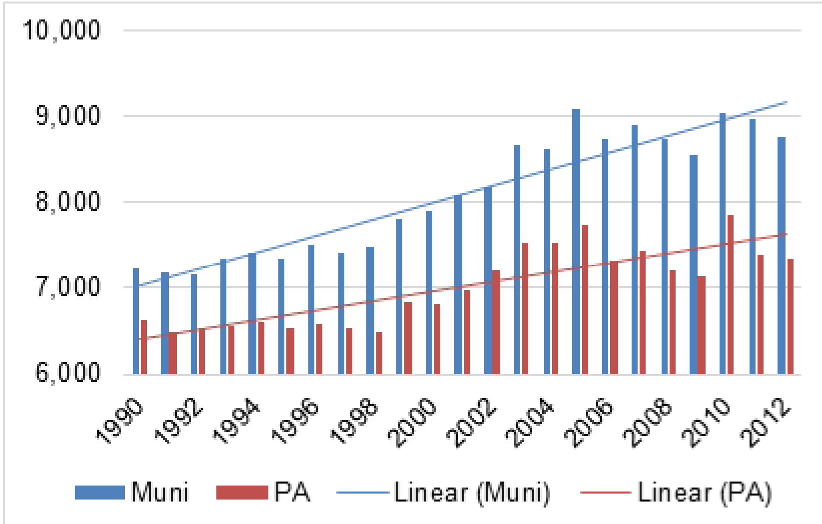
1.2 KEY FINDINGS

The top-down research yielded a number of key findings stemming from the pilot studies and review of methods. In the sections that follow, we first review the findings from the PA-Muni model pilot study, as these provided the most compelling evidence that top-down research can play a role in the evaluation of energy efficiency programs going forward. Then we review the key findings from the literature review portion of the study. These findings provide a framework for understanding the role that top-down modeling can serve in evaluating programmatic activity and the range of constraints it faces. Finally, we review the findings from the PA Data C&I model. This pilot study showed some promise, but also faced key data challenges.

1.2.1 Key findings from the PA-Muni model

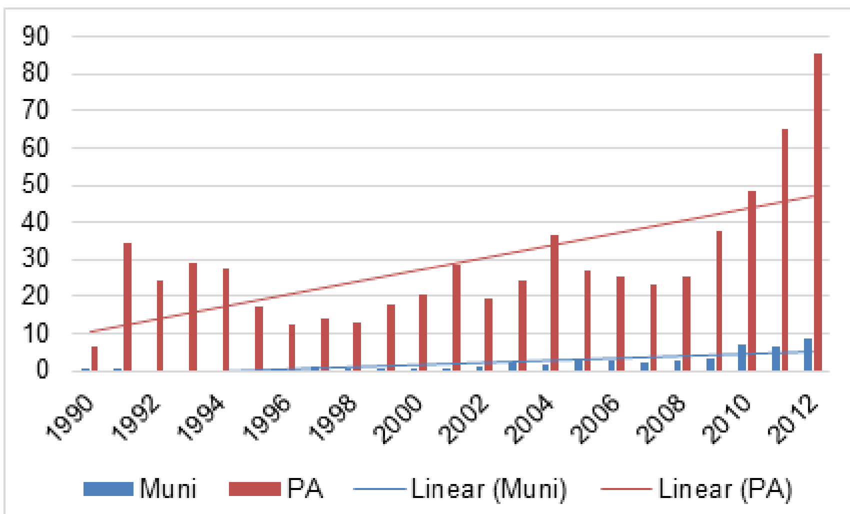
A primary motivation for the PA-Muni top-down approach was to further analyze apparent trends in residential energy consumption and energy program expenditures in the territories served by the PAs and the municipal utilities in Massachusetts. As shown in Figure 1-1, the average annual residential consumption per customer for both PAs and municipal utilities increased from 1990 to 2012, but the rate of increase was greater for the municipal utilities than for the PAs.

Figure 1-1. Residential Electricity Consumption per Customer (in kWh)



While most municipal utilities had residential energy efficiency programs during the same period, the funding levels for the municipal utilities were significantly below those of the PAs, as shown in Figure 1-2.

Figure 1-2. Residential Electric Program Expenditures per Customer (in $)



The PA-Muni approach explored the extent to which the lower rate of increase in consumption in PA territories is due to greater programmatic activity. The analysis attempted to control for the economic, time-series, and weather-related factors to isolate the effect of program activity from naturally occurring changes and the effects of other factors.

For both the residential and C&I models, the team investigated three families of models. The first family included current and past program expenditures for individual years and other control variables. The second family cumulated the program expenditures into a single variable and was otherwise identical to the first family. If one of the first-family models is a good representation of reality, the expected effect of substituting cumulative expenditures for individual year expenditures in what would otherwise be the same model would be to blur and smooth out the individual year effects. This cumulating approach would likely reduce the variation in the program activity variable and lower the coefficients that measure program impact. The third family dropped the other control variables from the models simply to see how sensitive the results are to the inclusion of these control variables. The first family of models is the team's preferred approach.

Table 1-1 provides a comparison of annual savings estimates from the first and second family of residential top-down models, with lags included.[2] The top-down models accounted for the leakage of upstream PA lighting program rebate dollars to municipal utility service territories. The table shows the annual net savings estimates and the corresponding lower and upper bounds of the 90% confidence intervals. The table also expresses top-down estimated net savings as a percent of the annual bottom-up net saving estimates. The four- and six-lag models, respectively, account for the impact of up to four and six previous years' programmatic activity on the current year's consumption. The four-lag model, which provided the best statistical fit to the data,[3] shows a top-down to bottom-up savings ratio of 187%, but the 90% confidence interval ranges from 92% to 282%. When the individual year expenditures are cumulated into a single variable, the ratio from the four-lag residential model reduces to 85%, with a confidence interval on it ranging from 2% to 168% of annual bottom-up savings. The fact that four-year and six-year lag models produce comparable results suggests that the results are stable across models with different lag lengths. However, further research is needed to understand the differences in the estimates from the individual-year and cumulated program expenditure models.

Table 1-1. PA-Muni Residential Top-down and Bottom-up Net Savings Comparisons, 2003–2012

		Top-down Annual Net Saving Estimates (GWh)			Top-down Annual Net Saving Estimates (% of Net Bottom-up Estimates)		
Model Family	#Lags	Lower Bound	Point Estimate	Upper Bound	Lower Bound	Point Estimate	Upper Bound
Individual Year	Four	1,851	3,762	5,674	92%	187%	282%
Cumulated	Four	41	1,714	3,387	2%	85%	168%
Individual Year	Six	2,829	3,821	4,814	141%	190%	240%
Cumulated	Six	1,075	2,233	3,391	53%	111%	169%

Similarly, Table 1-2 provides a comparison of annual saving estimates from the first family of C&I top-down models. The three-lag model, which provided the best statistical fit to the data,[4] shows a top-down to bottom-up savings ratio of 101%, while the 90% confidence interval ranges from 28% to 174%. When the individual year expenditures are aggregated into a

[2] A group of models with no lags was also tested, but were considered less meaningful since program participation affects consumption over a period of several years.

[3] While the estimate for the fourth lag was statistically significant in both the four- and six-lag models, the estimates for the fifth and the sixth lags in the six-lag model were not statistically significant.

[4] While the estimate for the third lag was statistically significant in other lagged models, the estimate for the fourth lag was not. The fifth and the sixth lags in the six-lag model were statistically significant but they had the opposite (positive) sign.

single variable, this ratio for the three-lag C&I model reduces to 95%, with a confidence interval on the top-down saving estimates ranging from 22% to 168% of annual bottom-up savings. The fact that C&I models with different lag lengths produce different results indicates that the C&I model results were less robust compared to residential. While this is expected given that consumption in the C&I sector is more volatile than that in the residential sector, and the customer base is more heterogeneous, further research is needed to understand the high degree of volatility of the results with respect to lag length.

Table 1-2. PA-Muni C&I Top-down and Bottom-up Net Savings Comparisons, 2003–2012

Model Family	#Lags	Top-down Annual Net Saving Estimates (GWh)			Top-down Annual Net Saving Estimates (% of Net Bottom-up Estimates)		
		Lower Bound	Point Estimate	Upper Bound	Lower Bound	Point Estimate	Upper Bound
Individual Year	Three	925	3,342	5,758	28%	101%	174%
Cumulated	Three	742	3,158	5,574	22%	95%	168%
Individual Year	Four	-207	2,142	4,491	-6%	65%	136%
Cumulated	Four	307	2,656	5,005	9%	80%	151%
Individual Year	Six	-2,850	-573	1,703	-86%	-17%	51%
Cumulated	Six	-3,204	-277	2,651	-97%	-8%	80%

The team considers the top-down to bottom-up estimate ratio of 1.9 for the residential sector and 1.0 for the C&I sector to be preliminary indicators. While these indicators suggest that the program effects identified by bottom-up approaches are real—and may even be understating the program-induced savings—the team does not recommend using these preliminary indicators as program metrics. Further research is being conducted to explore the stability and sensitivity of the PA-Muni model results. This further research includes, but is not limited to, the following:

- Additional model diagnostic tests
- Alternative model specifications
- Identification of outliers or influential observations
- Investigation of time-varying differences between the territories served by the munis and the PAs
- Exploration of other ways that the explanatory variables could be constructed.

If this further research establishes that the results are stable against alternative specifications, then the team recommends exploring ways to reduce the width of the confidence intervals around the estimates.

As a higher level research finding, the results of the PA-Muni residential model suggest that further improvements to this top-down research should be viewed alongside other efforts regarding spillover and market effects. By contrasting top-down model results to net impacts from bottom-up approaches that include spillover and market effects, the top-down research can be used to provide evidence to support or refute the existence of market effects and spillover..

1.2.2 Other key findings

In this section we review the key findings from the literature review and the second pilot study, the PA Data C&I model.

A. Findings from the Literature Review

- While there have been a variety of applications of top-down modeling of energy impacts, only a limited number of studies were directly relevant to the objectives of the present study—two pilot studies done in California. The more successful studies used a variety of techniques to account for challenges associated with measuring changes over time, including:
  - Models use a variety of measures of programmatic activity as explanatory variables, such as program expenditures, ex-ante savings, and measures of market transformation.
  - Models must account for the cumulative impact of programmatic activity over time, which can be accomplished by using terms that measure the amount of programmatic activity in previous periods (lagged effects).
- A major challenge to the present analysis is that much of the programmatic activity is consistent across the PA territories, as are the socioeconomic characteristics. This limits the amount of variation within and across observational units. National-level models tend to be affected by consistency in the reporting of data, while state-level models tend to be impacted by the availability of data.
- Top-down models must include a sufficiently long time series to capture variations in programmatic activity within and across observational units. However, it is equally important to consider major changes to the economy and/or level of programmatic activity over the period used in the analysis.
- The time series must be long enough to capture changes in programmatic activity over time. However, there must be a sufficiently long time series after that change to measure its true impact.

B. Findings from the PA Data pilot study

The PA Data pilot study showed that, aside from a limited time series, the models may be impacted by the following additional modeling considerations:

- Weather normalization process[5]

[5] To determine if the weather normalization method was responsible for the model results, we also constructed a set of models that used actual consumption as the independent variable instead of normalized annual consumption (NAC). These models included heating and cooling degree-days as explanatory variables. However, none of the models tested showed a statistically significant

- Industry level differentiation[6]
- Changes in the level of energy-consuming technology over time.

Based on this pilot study, we developed recommendations for modeling techniques that may contribute to longer ongoing evaluation efforts, and established the preferred model specifications and data requirements for the next phase of top-down modeling efforts.

1.3 RECOMMENDATIONS

This Year 1 top-down research provided a number of key recommendations for conducting the next phase of pilot studies in Massachusetts. We summarize these as follows:

- Continue refinement of the PA-Muni model to investigate the stability of models and possible changes to model specification that may reduce confidence intervals as outlined above.
- For the PA Data model, continue to collect data through the C&I database to extend the available data series to include five years of consumption and program tracking data, then continue collecting the necessary data going forward for future analysis. Continue to refine the existing models to further explore approaches to weather normalization, industry segmentation, and inclusion of other key explanatory variables such as technology trends; and incorporate multiple lag periods of the program and consumption variables.

---

relationship between consumption and degree-days, nor did the significance of the other model parameters improve.

[6] While we did run separate models for small commercial, large commercial, and industrial sectors, the publicly available economic data did not allow for industry level data for all relevant variables in the model. Further research would need to consider the appropriate way to sub-segment key economic variables.

2. SUMMARY OF YEAR 1 TOP-DOWN RESEARCH

2.1 BACKGROUND

Traditionally, energy efficiency program savings have been estimated using a bottom-up approach that incorporates a range of techniques to estimate gross and net energy savings for individual measures/end-uses, programs, or groups of programs. In Massachusetts, there is some concern that current attribution methods do not fully capture net impacts because programs are large and the effects of multiple programs may contribute jointly to aggregate savings, making it difficult to isolate the effects of any one program using just a bottom-up approach. Stakeholders are concerned that the bottom-up approach may be understating (or overstating) net program impacts.

Top-down modeling is an econometric approach to measure program impacts using aggregate cross-sectional and time series data. The top-down models measure changes to aggregate energy consumption relative to changes in energy efficiency programmatic activity, prices, and other economic factors. The goal of this type of modeling is to isolate the effect of program activity from other natural changes and policy variables. Top-down studies seek to capture net program effects regardless of specific program activities. A top-down approach may provide a complementary approach to bottom-up methods for estimating net impacts. Evaluators and other stakeholders are intrigued by top-down techniques because of their potential to provide low-cost supplemental or alternative estimates of net program savings. In principle, top-down methods are capable of capturing the full program effect, including free-ridership, spillover, market effects, and snapback. However, savings estimates derived from top-down models include many of the shortcomings observed in bottom-up model results; top-down methods also introduce other, significant technical challenges.

The Massachusetts program administrators (PAs) and the Energy Efficiency Advisory Council (EEAC) engaged Tetra Tech and its subcontractors—NMR and DNV GL—to explore the potential for top-down methods to help address shortcomings in the current approach to measuring net energy impacts. The goal of this multi-year study is to develop and apply multiple top-down methods for Massachusetts, and to understand the strengths and limitations of those methods relative to the traditional bottom-up approach to measuring net energy impacts.

2.2 STUDY COMPONENTS

The work presented in this document is part of a multi-year initiative designed to assess the utility of top-down modeling as a viable technique for evaluating energy efficiency programs in Massachusetts. This document presents a summary of the Year 1 investigation into possible methods for employing top-down modeling as a supplementary technique for evaluating net impacts associated with energy efficiency programs. This final report discusses the Year 1 research activities, that include the following:

- *Assessment of top-down modeling methods.* Section 4 of this document discusses advantages, disadvantages, and necessary properties of top-down methods, and reviews 15 top-down research studies that were used to estimate impacts associated with energy efficiency programs. Based on this literature review, we developed recommendations for specific methods to be used in Massachusetts.

- *PA-Municipal utility pilot study.* Section 5 of this document summarizes the Year 1 research concerning the first pilot study. This pilot study contrasted changes in consumption in the residential and C&I sectors relative to programmatic activity, and compared results for the PAs and municipal utilities. This model relied on aggregate consumption and program expenditure data at the PA or municipal utility level. The evaluation team was able to include a 15-year time series, allowing the model to overcome some modelling limitations resulting from shorter time series. Consequently, some preliminary models from the pilot study were able to generate statistically significant and substantial net savings estimates. In principle, the models provide estimates of program-attributable savings, net of naturally occurring reductions to energy consumption, by contrasting consumption in the PA territories (where there has been substantial programmatic activity) to that in municipal utility territories (where there has been minimal to no programmatic activity), while controlling for other differences between the muni and PA territories. There are a number of models that show particularly promising results, namely models that account for the lagged and cumulative impact of programmatic activity. However, further research is needed to refine model specifications, establish the stability of model results, and improve model precision.

- *PA Data pilot study.* Section 6 of this document discusses considerations for and preliminary results from the PA Data pilot study C&I models. For this study, we used PA-provided consumption and program tracking data to construct a set of C&I top-down models.[7] Model estimation in this stage of evaluation did not attempt to estimate net savings using these models. Rather, the pilot study investigated whether the data provided evidence of sufficient signal between programmatic activity and consumption to warrant further study. Based on this pilot study, we developed recommendations for modeling techniques that may contribute to longer ongoing evaluation efforts, and established the preferred model specifications and data requirements for the next phase of top-down modeling efforts.

2.3 OVERVIEW OF TOP-DOWN MODELING

Top-down modeling is an econometric approach to measure program impacts using aggregate cross-sectional and time series data. The top-down models measure changes to aggregate energy consumption relative to changes in energy efficiency programmatic activity, prices, and other economic factors. The goal of this type of modeling is to isolate the effect of program activity from other natural changes and policy variables.

Top-down techniques use a holistic approach by estimating program impacts across all energy-efficiency programs in a given geographical region or service territory, rather than running separate studies for each program (or measure/end-use within a program). Top-down models attempt to measure structural changes in energy consumption over time that are directly attributable to programmatic interventions by the utilities. Energy efficiency is a form of technological change. Utility energy efficiency (EE) programs most often serve as an accelerant to the pace of technological change.

[7] The PA Data pilot study for the residential sector will be provided in 2015, as the three years of consumption data required for modeling are not yet available for analysis.

The premise of the top-down approach is that energy consumption (E) for a given area is a function of program activity (P) in that area, and other identifiable factors (X).

Equation 2-1. General Form of Top-down Models

$$E_{at} = \beta_0 + \beta_p P_{at} + \Sigma_j \beta_j X_{jat} + \varepsilon_{at}$$

This condensed equation is presented for purposes of discussion. In practice, the program activity metric (P) may be multi-dimensional and include lag terms or cumulative activity. The program variable(s) may be expressed as a set of index variables to represent the cumulative penetration of utility EE into the markets, over time. If the model is well specified, the exogenous factors (Xj) control for all the non-program differences among areas. The coefficient βp, expected to be negative, estimates the change in consumption per unit of program activity, controlling for all other factors. The program effect for a particular area and time would then be estimated as βpPat.

2.4 TOP-DOWN METHOD ASSESSMENT: APPROACH AND FINDINGS

In order to assess existing top-down modeling techniques and recommend specific methods to use in Massachusetts, the evaluation team reviewed 15 top-down research studies that have been used to estimate impacts associated with energy efficiency programs. These studies employed different units of analysis for varying levels of aggregation, and used a range of techniques to provide a variety of programmatic impacts, including:

- Realization rate on ex-ante savings
- Cost-effectiveness of program expenditures
- Gross and net savings estimates
- Measures of market transformation
- Changes to market share of energy efficient products

Of the 15 studies reviewed, only two were directly relevant to fulfilling the objectives of the present pilot studies to examine the impacts associated with energy efficiency programmatic activity within a state. For the remainder of the studies:

- Six studies estimated national-level impacts based on data aggregated at the state level.
- Two studies provided reviews of national-level impact studies.
- Four studies provided top-down analyses associated with specific technologies only.
- One study measured in-state changes to consumption resulting from changes to building codes only, and did not consider energy efficiency programmatic activity.

Table 2-1, Table 2-2, and Table 2-3 provide an overview of the studies reviewed by our evaluation team.

Table 2-1. Overview of Top-down Studies Reviewed—National Level

Study	Summary	Pros	Cons
Demand-Side Management and Energy Efficiency in the United States (Loughran and Kulick (2004))	National time-series cross sectional model of state level energy consumption and program expenditures data. The model sought to estimate the cost effectiveness of energy efficiency programs.	The model estimated the cost effectiveness of energy efficiency programs accounting for the lagged impact of expenditures on savings and other fixed effects. Model properly addresses fixed effects and econometric considerations.	Program impacts limited to return on expenditures. Model could not measure the effectiveness of program designs and relies on highly aggregated data with reporting inconsistencies.
Demand-Side Management and Energy Efficiency Revisited (Affhammer et al. (2007))	Provided a review of Loughran and Kulick study. Re-estimated results weighting observations based on the relative size of utilities. Provided confidence intervals around parameter estimates.		
How Many Kilowatts are in a Negawatt? Verifying Ex Post Estimates of Utility Conservation Impacts at the Regional Level (Rivers and Jaccard (2011))	National time-series cross sectional model of utility and province level energy consumption and program expenditures data. The model sought to estimate the cost effectiveness of energy efficiency programs.	The model attempted to estimate the cost effectiveness of energy efficiency programs accounting for the lagged impact of expenditures	Voilette demonstrate that applying Loughran and Kulick's model to Rivers and Jaccard's data results in savings that are sufficiently high to justify expenditures. Illustrate the importance of accounting for the lag in program activity and fixed effects.
Review of a Top-Down Evaluation Study: Rivers and Jaccard 2011) (Violette (2012))	Provided a review of Rivers and Jaccard study. Applied Rivers and Jaccard data to Loughran and Kulick's model.		
Electricity Intensity in the Commercial Sector: Market and Public Program Effects. (Horowitz (2004))	Estimated a national time-series fixed effects model using state level energy consumption data. The attempted to estimate the effects of energy programs that directly target customers from up-stream (market transformation) programs.	Model demonstrates the importance of considering different types of programmatic activity on savings. Model estimated using data from 42 states of 12 years of varying programmatic activity.	Measure of market transformation derived using data from a variety of loosely connected sources, leads to questionable interpretation of results.
Changes in Electricity Demand in the United States from the 1970s to 2003 (Horowitz (2007))	Study uses a difference of differences approach to construct a national model that contrasts pre- and post-program consumption for states with strong-to-moderate programmatic activity to states with weak programmatic activity.	Provides an approach for developing the counterfactual conditions and estimating net savings.	Difference of differences approach requires many assumptions regarding the selection of treatment and control states as well as pre- and post-periods.
Measuring the savings from energy efficiency policies: a step beyond program evaluation (Horowitz (2010))	Demonstrates that top-down models can be developed at different levels of analysis to provide estimates of programmatic impacts based on data aggregated at the account, utility, and state levels.	Illustrates ability of top-down methods to be applied to different levels of analysis using data aggregated at different levels.	Reduction in energy intensity assumed to result from corresponding increases in energy efficiency activity without direct causality being established.

Table 2-2. Overview of Top-down Studies Reviewed—Regional and State Level

Study	Summary	Pros	Cons
How Many Kilowatts are in a Negawatt? Verifying Ex Post Estimates of Utility Conservation Impacts at the Regional Level (Parfomak and Lave (1996))	Uses utility level consumption and ex-ante savings to estimate the realization rate on savings across utilities in New England and California.	Provides a realization rate on ex ante savings.	Model does not account for many factors that may also result in reductions to energy consumption over time.
CPUC Macro Consumption Metric Pilot Study (Final Report) (Cadmus (2012))	Used energy efficiency expenditures and a series of explanatory variables to predict changes to energy use for commercial and residential for a utility service territories in California.	The model used an extended time series, 1990 – 2010. While this may not provide for a true "No Program" baseline, the level of activity in the early 1990's should be sufficiently different to provide a meaningful point of comparison.	The model does not distinguish between types of programmatic activity. The model was not able to produce statistically significant results.
Macro Consumption Metrics Pilot Study Technical Memorandum – Preliminary Findings (Demand Research (2012))	This study uses a two-way fixed effects model that aggregates consumption and economic variables to either the census tract level for residential customers or industry by county for nonresidential customers. Annual consumption per location is set equal to a set of time series variables that reflect the ratio of ex-ante savings to consumption, the ratio of measure costs to fuel expenditures, and incentive costs to fuel expenditures.	This study is one of two existing studies that focus specifically on measuring programmatic net impacts from utility sponsored programs within a single state. This study includes multiple measures of programmatic activity including ex-ante savings, incentive and measure costs. The model uses weather normalized consumption as the dependent variable which is the same as the PA data model being developed through the current study	The model limits impacts to in-state that occur over a 5-year time series.
Are Building Codes Effective at Saving Energy? Evidence From Residential Billing Data in Florida (Jocobsen and Kotchen (2009))	Uses account level utility data to estimates a pooled time-series cross-sectional model that is used to construct a difference of differences comparison of the effect of building codes on energy consumption.	Demonstrate the importance of building codes on reductions in energy consumption.	Model does not consider the effect of energy efficiency programs on consumption. Scope of model is limited to the utility service territory.

Table 2-3. Overview of Top-down Studies Reviewed—Technology Specific Studies

Study	Summary	Pros	Cons
The Impact of Regional Incentive and Promotion Programs on the Market Share of ENERGY STAR® Appliances (Rosenberg (2003))	Estimated multi-state linear regression models to predict the impact of incentive programs and regional demographic variables on market shares for separate ENERGY STAR® appliances	Models demonstrate the ability to employ a variety of data sources and statistical techniques to estimate programmatic impacts.	Models provide measure specific results only
Modelling the Effects of U.S. ENERGY START Appliance Programs. (Feldman et al. (2005))	Used ANOVA and linear regression analysis to first estimate the market penetration of ENERGY STAR appliances by state as a function of the presence of program activity and then used the change in market shares over time to predict cumulative effects of ENERGY STAR programs		
Results of the Multistate CFL Modeling Effort (NMR Group, Inc. (2011))	Used CFL saturations from survey data along with energy efficiency program budget information, number CFLs receiving incentives and program types to predict CFL purchases over multiple years		
Economic Indicators of Market Transformation: Energy Efficient Lighting and EPA's Green Lights (Horowitz (2001))	Used data from the Census' "Manufacturing and Construction database" from 1959 – 2000 to construct a model that estimates the market share for energy efficient lighting based on product price, the price of electricity and a vector of macroeconomic variables		

A review of this literature provided the following insights.

Relevance of existing studies to current Massachusetts pilot studies:

- While there have been a variety of applications of top-down modeling of energy impacts, only a limited number of studies were directly relevant to the objectives of the present study—the two California pilot studies.
- Many of the models employed time series, cross-sectional models that measured change to aggregate consumption resulting from changes in program expenditures and/or ex-ante savings, which the two-pilot studies also attempt to employ.

Approaches to addressing the time series effects of estimates from top-down models:

- The more successful studies used a variety of techniques to account for challenges associated with measuring changes over time, including:
  - Measuring year-over-year change in consumption relative to the year-over-year change in programmatic activity.
  - Including terms that accounted for fixed characteristics of a population, utility service territory, or year.
  - Using a log transformation of the consumption variable to reduce the amount that the variance increases with the level of consumption.
  - Using a “difference in differences” approach to contrast consumption between groups of observations with high and low programmatic activity. This can help minimize the need for consistency across observational units.

Approaches for measuring the variety of programmatic influences:

- It is important that the model use a variety of measures of programmatic activity as explanatory variables, such as program expenditures, ex-ante savings, and measures of market transformation.
- Models must account for the cumulative impact of programmatic activity over time, which can be accomplished by using terms that measure the amount of programmatic activity in previous periods (lagged effects).
- A number of studies showed that it is important to include terms that allow for discernment of impacts associated with changes in building codes from program impacts.
- A major challenge to the present analysis is that much of the programmatic activity is consistent across the PA territories, as are the socioeconomic characteristics. This limits the amount of variation within and across observational units. National-level models tend to be affected by consistency in the reporting of data, while state-level models tend to be impacted by the availability of data.

Factors related to the length of time series:

- Top-down models must include a sufficiently long time series to capture variations in programmatic activity within and across observational units. However, it is equally important to consider major changes to the economy and/or level of programmatic activity over the period used in the analysis.

- The time series must be long enough to capture changes in programmatic activity over time. However, there must be a sufficiently long time series after that change to measure its true impact.
- Using an extended time series creates additional concerns with regard to changes to the overall economy and demographics. For example, the present analysis must necessarily contain the recent recession in any time series before 2010.
- Other structural changes include shocks to energy prices, major structural changes to the economy, and large influxes of government expenditures.

2.5 PA-MUNICIPAL UTILITY PILOT STUDY APPROACH AND FINDINGS

2.5.1 PA-Muni model review of available data

The evaluation team collected time-series data on residential electricity consumption and factors that could affect consumption for all Massachusetts PAs/utilities and towns from 1990 to 2012. The team also developed a panel database for the study, which included the following data elements:

- *Electricity consumption and price data* – The team collected data on the total residential electricity sales, revenue, and customers in Massachusetts from the EIA's 861 files for 1990-2012 for each PA and municipal utility.
- *Energy efficiency programmatic activity* – Due to limitations identified in program expenditure data that were publicly available through EIA, the evaluation team collected the energy efficiency program expenditures data by sector and year from the PAs, the municipal utilities, and their association directly. Collecting consistent electric program data across all PAs and municipal utilities was a substantial challenge to this study. While the evaluation team attempted to collect detailed time-series data on program activity, the only consistent piece of data that the team was able to gather across all PAs and municipal utilities was the annual total electric program expenditures. The collection of data from the municipal utilities was especially challenging because municipal utility participation was completely voluntary and many municipal utilities were unable to compile historical program data due to time and staffing constraints.
- *Weather data* – The evaluation team gathered daily temperature data for all weather stations in Massachusetts from the National Oceanic and Atmospheric Administration (NOAA) from 1990 through 2012.
- *Economic and demographic data* – The evaluation team gathered town-level economic and demographic data from the US Census American Community Survey (ACS), US Decennial Census, US Census Building Permits Survey, and Bureau of Labor Statistics.

2.5.2 PA-Muni model summary of methodology

The team specified a fixed-effects panel regression model. This type of regression model allows each individual to act as its own control. The unique effects of the stable, but unmeasured characteristics of each utility are their “fixed effects” from which this method takes its name. These fixed effects are held constant in the model. The fixed-effects nature of the model means the model does not need to include unchanging characteristics. In a model of households, for example, these characteristics might include square footage, number of floors, direction the home faces, etc. In this study’s model, this includes characteristics of these areas that do not change over time. These might include that Boston is the home of the state capitol with the state’s tallest buildings, that the Cape gets sea breezes and usage that varies with vacation travel, and that western MA has the Berkshire Mountains, more rural areas, and the greatest differences in topography, etc. Including fixed effects in the model controls the amount of variance (noise) that the model must address to explain electricity consumption. This approach also provides for a much closer fit to the data than other types of regression models.[8]

The team initially considered running the models at the town level because the economic and demographic data were available at that gradation. This would have allowed for a better comparison of PAs and municipal utilities given that most municipal utilities serve only a single town, while the PAs serve a large number of towns. However, because the PAs’ energy consumption and energy efficiency program data were available only at the PA level, the team aggregated the town-level economic and demographic data to the PA level.

For both the residential and C&I sector models, we estimated separate models that varied by (1) program expenditures entered into the model as individual years or as cumulative expenditures and (2) whether the models included adjustments for leakage of upstream program expenditures to municipal utility territories.

[8] The inclusion of fixed effects in the model ensures that the estimated regression coefficients are not biased due to non-time-varying (i.e., PA/utility-specific) characteristics. A random-effects specification is more efficient, but using random effects does not fully control for all utility-specific characteristics. Hausman tests were used to determine which model specification to use. The findings from those tests showed that fixed effects were more appropriate for this analysis.

Equation 2-2 below shows the residential PA-Muni top-down model specification. Since there is a significant variation in the size of PAs and municipal utilities, the models were weighted by the amount of residential electricity sales to properly represent the different magnitudes of spending and potential savings across the PAs and municipal utilities in Massachusetts.

Equation 2-2. PA-Muni Residential Top-down Model

$$\mathbf{log}(\boldsymbol{EC_{it}}) = \boldsymbol{\beta_1}\, \boldsymbol{log}(\boldsymbol{P_{it}}) + \boldsymbol{\beta_2}\, \mathbf{log}(\boldsymbol{HDD_{it}}) + \boldsymbol{\beta_3}\, \mathbf{log}(\boldsymbol{CDD_{it}}) + \boldsymbol{\beta_4}\, \mathbf{log}(\boldsymbol{I_{it}}) + \boldsymbol{\beta_5}\, \boldsymbol{EH_{it}} + \boldsymbol{\beta_6 VAL_{it}}$$

$$+ \beta_7\, NC_{it} + \beta_8\, SF_{it} + \beta_9\, RENT_{it} + \beta_{10}\, EMP_{it} + \sum_{j=0}^{n} \alpha_j EE_{it-j} + \beta_{11}\tau_t + \delta_i + \varepsilon_{it}$$

Where:

$\log(EC_{it})$	=	Natural logarithm of annual consumption per residential customer in PA/utility service area *i* and year *t*.
$\log(P_{it})$	=	Natural logarithm of electricity price in 2012 dollars in PA/utility service area *i* and year *t*.[9] The coefficient β_1 measures the price elasticity of electricity consumption.
$\log(HDD_{it})$	=	Natural logarithm of annual heating degree days (base 65) in PA/utility service area *i* and year *t*. The coefficient β_2 measures the elasticity of electricity consumption with respect to heating degree days.
$\log(CDD_{it})$	=	Natural logarithm of annual cooling degree days (base 70) in PA/utility service area *i* and year *t*. The coefficient β_3 measures the elasticity of electricity consumption with respect to cooling degree days.
$\log(I_{it})$	=	Natural logarithm of median household income in 2012 dollars in PA/utility service area *i* and year *t*.
EH_{it}	=	The share of households using electricity as the primary heating fuel in PA/utility service area *i* and year *t*.
VAL_{it}	=	The median house values in 2012 dollars in PA/utility service area *i* and year *t*.
NC_{it}	=	The share of new construction in residential housing, computed as the total number of residential new construction permits divided by the total number of housing units in PA/utility service area *i* and year *t.*
SF_{it}	=	The share of single-family homes in residential housing, computed as the total number of single-family housing units divided by the total number of housing units in PA/utility service area *i* and year *t*.
$RENT_{it}$	=	The share of renters in PA/utility service area *i* and year *t*.

[9] Nominal prices were adjusted to reflect 2012 dollars using the GDP implicit price deflator from the Federal Reserve Economic Data.

EMP_{it}	=	The employment rate, computed as the number of employees divided by the number of people in the labor force in PA/utility service area *i* and year *t*.
EE_{it-j}	=	Total residential electric energy efficiency program expenditures per residential customer in PA/utility service area *i* and year *t-j.* The coefficient α_j measures the percentage change in electricity consumption in year *t* from a one-dollar change in energy efficiency program expenditures in year *t-j*. The sum of α_0 through α_n measures the percentage change in electricity consumption in year *t* from a one-dollar change in energy efficiency program expenditures in year *t* and the previous *n* years.[10]
τ_t	=	Time-trend variable that is equal to 1 in 1990 and increasing by one unit annually. The coefficient β_{11} captures the naturally occurring change in electricity consumption not captured by the variables included in the model.[11]
δ_i	=	PA/utility fixed effects that capture time-invariant PA/utility-specific fixed effects in electricity consumption. There may be a certain PA/utility-level variation in the data that is not necessarily related to energy efficiency programmatic activity, such as changes to the local economy resulting from local businesses closing.
ε_{it}	=	Regression error term in PA/utility service area *i* and year *t*.

Equation 2-3 below shows the C&I PA-Muni top-down model specification. Again, we initially considered running the C&I models at the town level because the economic and demographic data were available at that level. However, because the PAs' energy consumption and energy efficiency program data were available only at the PA level, the team computed a weighted average of economic and variable factors at the PA/utility level using the number of employees in each town as the weight.

Equation 2-3. PA-Muni C&I Top-down Model

$$\log(EC_{it}) = \beta_1\, log(P_{it}) + \beta_2 \log(HDD_{it}) + \beta_3 \log(CDD_{it}) + \beta_4 \log(EINC_{it}) + \beta_5 NC_{it} + \beta_6 EMP_{it} + \sum_{k=1}^{20} \gamma_k\, NAICS_{k,it} + \sum_{j=0}^{n} \alpha_j EE_{it-j} + \beta_7 \tau_t + \delta_i + \varepsilon_{it}$$

[10] The team also tested specifications with distributed lag models with a special parameterization of lagged energy efficiency expenditures variables in order to account for the possible non-linear and delayed effects of energy efficiency program activity on consumption. The results were similar.

[11] As a robustness check, the team also tested specifications with non-linear (a natural cubic spline, or some second- or third-degree polynomials) time trends. This had little effect on the results. Similarly, including the indicator variables for individual years instead of a time trend did not result in a significant change in the model results.

Where:

$\log(EC_{it})$	=	Natural logarithm of annual consumption per customer, per establishment, or per employee in PA/utility service area *i* and year *t*.
$\log(P_{it})$	=	Natural logarithm of electricity price in 2012 dollars in PA/utility service area *i* and year *t*.[12]
$\log(HDD_{it})$	=	Natural logarithm of annual heating degree days in PA/utility service area *i* and year *t*.
$\log(CDD_{it})$	=	Natural logarithm of annual cooling degree days in PA/utility service area *i* and year *t*.
$\log(EINC_{it})$	=	Natural logarithm of mean annual employment income per employee, in 2012 dollars, computed as total annual payroll divided by total number of employees in PA/utility service area *i* and year *t*.
NC_{it}	=	Square footage of C&I new construction per customer, per establishment, or per employee in PA/utility service area *i* and year *t*.
$\mathrm{NAICS}_{k,it}$	=	The percent of establishments in a two-digit NAICS industry code *k* in PA/utility service area *i* and year *t*. The establishments in Massachusetts belonged to 21 different two-digit NAICS codes. The $\gamma_k$is a vector of coefficients that capture the differences in building energy use by business type.
EMP_{it}	=	The employment rate, computed as the number of employees divided by the number of people in the labor force, in PA/utility service area *i* and year *t.*
EE_{it-j}	=	Total commercial and industrial energy efficiency program expenditures per C&I customer, per establishment, or per employee in PA/utility service area *i* and year *t-j.* The coefficient α_j measures the percentage change in electricity consumption in year *t* from a one-dollar change in energy efficiency program expenditures in year *t-j*. The sum of α_0 through $\alpha_n$measures the percentage change in electricity consumption in year *t* from a one-dollar change in energy efficiency program expenditures in year *t* and the previous *n* years.
τ_t	=	Time-trend variable that is equal to 1 in 1990 and increasing by one unit annually. This time-trend variable captures the naturally occurring change in electricity consumption not accounted for by the variables included in

[12] Nominal prices were adjusted to reflect 2012 dollars using the GDP implicit price deflator from the Federal Reserve Economic Data.

the model.[13]

δ_i	=	PA/utility fixed effects that capture time-invariant, PA/utility-specific fixed effects in electricity consumption. There may be a certain PA/utility-level variation in the data that is not necessarily related to energy efficiency programmatic activity, such as changes to the local economy resulting from local businesses closing.
ε_{it}	=	Regression error term in PA/utility service area *i* and year *t*.

2.5.3 PA-Muni model summary of results

A. PA-Muni residential model results

The residential top-down models provide estimates of net energy savings as a result of ratepayer-funded energy efficiency programs in Massachusetts. Such estimates can be used in concert with other estimates to arrive at a picture of the true impact of such programs. As such, top-down saving estimates at their best are not intended to replace the traditional bottom-up estimates, but ideally can help validate them.

For both the residential and C&I sectors, the team investigated three families of models. The first family included current and past program expenditures for individual years and other control variables. The second family cumulated the program expenditures into a single variable and was otherwise identical to the first family. If one of the first-family models is a good representation of reality, the expected effect of substituting cumulative expenditures for individual year expenditures in what would otherwise be the same model would be to smooth out the individual year effects. This cumulative approach would likely reduce the variation in the program activity variable and lower the coefficients that measure program impact. The third family dropped the other control variables from the models simply to see how sensitive the results are to the inclusion of these control variables. Hence the first family of models is the team's preferred approach.

Table 2-4 provides a comparison of annual savings estimates from the first and second family of residential top-down models, with lags included.[14] The top-down models accounted for the leakage of upstream PA lighting program rebate dollars to municipal utility service territories. The table shows the annual net savings estimates and the corresponding lower and upper

[13] The team also tested specifications with non-linear (a natural cubic spline, or some second or third degree polynomials) time trends. This had little impact on the results.

[14] A model with no lags was tested. Most of the residential energy savings would occur at the end of a calendar year. This means that the savings in any calendar year would not line up with the usage of that year. The residential upstream lighting program generates a large proportion of residential savings. The first several months of each year are slow as new Memorandum of Understanding Agreements are put in place. The result is that more than one-half and, generally, two-thirds of savings or more are in the latter half of the year. Therefore, the consumption impacts of program expenditures on average would be part of the referenced calendar year and the first part of the following. Lag models can accommodate this mismatch while no-lag models cannot.

bounds of the 90% confidence intervals. The table also expresses top-down estimated net savings as a percent of the annual bottom-up net saving estimates to provide a top-down to bottom-up estimate ratio. The four- and six-lag models account for the impact of up to four and six previous years' programmatic activity on the current year's consumption, respectively. The four-lag model, which provided the best statistical fit to the data,[15] shows a top-down to bottom-up ratio of 187%, but the 90% confidence interval ranges from 92% to 282%. When the individual year expenditures are cumulated into a single variable, this ratio from the four-lag residential model reduces to 85%, with a confidence interval ranging from 2% to 168% of annual bottom-up savings. The fact that four-year and six-year lag models produce comparable results suggests that the results are stable across models with different lag lengths. However, further research is needed to understand the differences in estimates from the individual-year and cumulated program expenditure models.

Table 2-4. PA-Muni Residential Top-down and Bottom-up Net Savings Comparisons, 2003–2012

Model Family	#Lags	Top-down Annual Net Saving Estimates (GWh)			Top-down Annual Net Saving Estimates (% of Net Bottom-up Estimates)[16]		
		Lower Bound	Point Estimate	Upper Bound	Lower Bound	Point Estimate	Upper Bound
Individual Year	Four	1,851	3,762	5,674	92%	187%	282%
Cumulated	Four	41	1,714	3,387	2%	85%	168%
Individual Year	Six	2,829	3,821	4,814	141%	190%	240%
Cumulated	Six	1,075	2,233	3,391	53%	111%	169%

[15] While the estimate for the fourth lag was statistically significant in both the four- and six-lag models, the estimates for the fifth and the sixth lags in the six-lag model were not statistically significant.

[16] The source of residential electric program reported net savings and expenditures is Massachusetts Division of Energy Resources' (DOER's) PARIS database. Annual net savings claims from 2003 to 2012 are cumulated and then divided by 10 (the number of years) to compute an average annual bottom-up estimate. The cumulative model estimate from the top-down individual-year models was divided by the number of lags included in the model plus 0.5 (to account for the partial-year effect of the current-year expenditures) to arrive at an average annual top-down estimate.

Similarly, Table 2-5 provides a comparison of annual saving estimates from the first family of C&I top-down models. The three-lag model, which provided the best statistical fit to the data,[17] shows a top-down to bottom-up estimate ratio of 101%, and the 90% confidence interval ranges from 28% to 174%. When the individual-year expenditures are cumulated into a single variable, the ratio from the three-lag C&I model reduces to 95%, with a confidence interval on it ranging from 22% to 168% of annual bottom-up savings. The fact that C&I models with different lag lengths produce different results indicates that the C&I model results were less robust compared to residential. While this is expected, given that consumption in the C&I sector is more volatile than that in the residential sector and the customer base is more heterogeneous, further research is needed to understand the large volatility of the results with respect to lag length.

Table 2-5. PA-Muni C&I Top-down and Bottom-up Net Savings Comparisons, 2003–2012

Model Family	#Lags	Top-down Annual Net Saving Estimates (GWh)			Top-down Annual Net Saving Estimates (% of Net Bottom-up Estimates)[18]		
		Lower Bound	Point Estimate	Upper Bound	Lower Bound	Point Estimate	Upper Bound
Individual Year	Three	925	3,342	5,758	28%	101%	174%
Cumulated	Three	742	3,158	5,574	22%	95%	168%
Individual Year	Four	-207	2,142	4,491	-6%	65%	136%
Cumulated	Four	307	2,656	5,005	9%	80%	151%
Individual Year	Six	-2,850	-573	1,703	-86%	-17%	51%
Cumulated	Six	-3,204	-277	2,651	-97%	-8%	80%

The team considers the top-down to bottom-up ratio of 1.9 for the residential sector and 1.0 for the C&I sector to be preliminary indicators. While these indicators suggest that the program effects identified by bottom-up approaches are real—and may even be understating the program-induced savings—the team does not recommend using these preliminary indicators as program metrics. Further research is being conducted to explore the stability and sensitivity of the PA-Muni model results. This further research includes, but is not limited to, the following:

- Additional model diagnostic tests
- Alternative model specifications

[17] While the estimate for the third lag was statistically significant in other lagged models, the estimate for the fourth lag was not. The fifth and the sixth lags in the six-lag model were statistically significant but they had the opposite (positive) sign.

[18] The source of C&I electric program reported net savings and expenditures is Massachusetts DOER's PARIS database. Annual net savings claims from 2003 to 2012 are cumulated and then divided by 10 (the number of years) to compute an average annual bottom-up estimate. The cumulative model estimate from the top-down individual-year models was divided by the number of lags included in the model plus 0.5 (to account for the partial-year effect of the current-year expenditures) to arrive at an average annual top-down estimate.

- Identification of outliers or influential observations
- Investigation of time-varying differences between the territories served by the munis and the PAs
- Exploration of other ways that the explanatory variables could be constructed.

If this further research establishes that the results are stable against alternative specifications, then the team recommends exploring ways to reduce the width of the confidence intervals around the estimates.

2.6 TOP-DOWN PA DATA PILOT STUDY APPROACH AND FINDINGS

2.6.1 Review of available data

The evaluation team reviewed data available for constructing the C&I and residential top-down models for the PA Data pilot study. We reviewed the available C&I consumption and program activity data in terms its availability and consistency at different geographic and temporal levels of aggregation. In addition, we reviewed building codes, macroeconomic variables, and energy price data that are relevant to both C&I and residential models. The focus of this review was to determine whether the necessary data were available and consistently reported at different levels of geographic aggregation and time intervals, and to determine the length of time series that was available for each series.

The data review provided the following insights:

- Data are generally available for constructing a historical series of consumption and programmatic activity at the account, town, and county level. However, additional years of these data are not easily accessible and require considerable processing to construct the necessary time series.
- Additional years of data are required to make reliable estimates of the impact of programmatic activity on consumption. The PAs already collect the necessary billing and tracking data to extend the time series going forward. The data elements that are required include:
  - Account level monthly billing data that contains:
    - Monthly meter reads
    - Start and end read dates for each read
    - Customer name and address
  - Program tracking data:
    - Upstream and downstream measure descriptions and expenditures (incentive costs, total costs, other marketing costs)
    - Downstream account and premise numbers
    - Upstream customer name and address or account number, if possible
    - Firm-o-graphic information: NAICS code, SIC code

- The level of geographic resolution for different macroeconomic series varies, which prevents the use of some series (e.g., GDP) at more disaggregate levels, while allowing them to be used at higher levels of aggregation (e.g., county or state level).
- Due to the escalation in programmatic activity over the past four to five years, even with sufficiently long historical time series, there is insufficient post-escalation-period consumption to measure the impact of this change in the level of activity.

2.6.2 Summary of methodology

This section describes the modeling specifications and process we used to construct the C&I electric top-down model for the PA Data pilot study. For the PA Data pilot study, we constructed C&I models at the following levels of resolution.

- **Geographic data resolution.** We constructed models at two levels of geographic aggregation:
  - *County-level models* **–** Models at the county level offer widely available time series data for the commercial and industrial models, as well as the forthcoming residential models.
  - *Town-level models* **–** We selected town-level models over census-tract level because the data series available to explain the variation in energy consumption are more limited at the census-tract level than at the town level. The Census publishes a series on wages and employment by industry code at the town level.
- **Sector-level resolution.** We estimated separate models for the following C&I sectors to allow for greater differentiation across observational units:
  - *Small commercial* **–** Customers identified as non-industrial customers, based on their NAICS codes, that had annual demand less than 300 kW, based on the PA billing and tracking data.
  - *Medium to large commercial* **–** Customers identified as non-industrial customers, based on their NAICS codes, that had annual demand greater than 300 kW, based on the PA billing and tracking data. Program activity for these customers was restricted to tracking records for the small business and upstream programs.
  - *All commercial* **–** All non-industrial customers, based on their NAICS codes, regardless of size.
  - *Industria*l **–** Customers identified as industrial based on their NAICS codes.
- **Temporal resolution.** There are two possible levels of temporal resolution: quarterly and annual. There was too much uncertainty in the data at the quarterly level to recommend attempting a quarterly model. The choice of using annual data is consistent with the findings of the literature discussed in Section 4 of this document.

Equation 2-4 presents the general form of the PA Data model. We use δ (delta) to highlight the use of the first difference (i.e., year-over-year change) in variables used in the model specification. This is a standard billing analysis approach, extended to the cross-sectional economic aggregate, that works well when there is a limited number of time series observations, as exists during the initial model estimation phase in this study. Equation 2-4 presents the general form of the PA Data C&I electric top-down model.

Equation 2-4. The PA Data Commercial/Industrial Model

$\delta(NAC)_{tsgf} = \beta0_{sgf} + \beta1*[\delta Employment]_{tsg} + \beta2*[\delta EE\ \$\ Program\ Activity]_{tsgf} + \beta4*£_{sgf} + \beta4*¥_{tsf}$
ε_{tsgf}

Where each variable in Equation 2-4 is defined as follows:

$\beta0_{sgf}$	=	A fixed effects variable for sector (s), within geographic region (g), and by fuel type (f).
$(NAC)_{tsgf}$	=	Normalized (C&I) Annual Energy Consumption in year (t), sector (s), within geographic region (g), and by fuel type (f). For the county-level models, all variables are divided by gross domestic product to provide a measure of energy intensity per unit of output. For the town-level models, population is used in place of GDP due to data limitations.
$Employment_{tsg}$	=	Economic activity measured as the total employment per GDP or population, for county and town-level models, respectively, within year (t), sector (s), and geographic region (g).
Program activity	=	We considered two separate measures of programmatic activity separately: • EE $ Program Expenditure Vbl(s) $_{tsgf}$ is one or more EE program variables measured in $s, reflecting program expenditures as reported in the PA program tracking data, in year (t), sector (s), within geographic region (g), and by fuel type (f); and • EE Program Energy Savings Vbl $_{tsgf}$ is a measure of estimated EE savings, as reported in the PA program tracking data, in year (t), sector (s), within geographic region (g), and by fuel type (f = electricity or natural gas).
$*£_{sgf}$	=	Parameter for geographic fixed effects for county or town g in sector s, and fuel type f.
$¥_{tsf}$	=	Parameter for annual fixed effects for year t in sector s, and fuel type f.

We used the following steps to develop the county- and town-level models:

- *Model of total NAC versus economic activity* – Before introducing program activity and other variables, we first investigated whether changes to NAC could be explained by changes in employment, as well as geographic and annual fixed effects. We constructed these simplified consumption models at two separate levels of analysis:
  - *County-level model* – This model used NAC per unit of GDP as a dependent variable.
  - *Town-level model* – Due to lack of available GDP and payroll data at more granular levels of analysis, the town-level model used NAC per capita as a dependent variable. Since only three years of data were available, we explored

the non-differenced version of each model to provide for estimation across all three years of available data.

- *Introduce measures of program activity* – After determining that we could successfully model NAC as a function of the economic variable, we introduced the following two measures of programmatic activity separately. We considered the measures separately to limit collinearity:
  - Aggregate energy efficiency expenditures per unit – As is discussed in detail in Section 6.1.3, we obtained account- and measure-level downstream program expenditures and measure- and location-specific upstream data from the PA tracking data.
  - Aggregate ex-ante savings per unit – As is discussed in detail in Section 6.1.3, we obtained account- and measure-level downstream program savings and measure- and location-specific upstream data from the PA tracking data.

  Due to data limitations, we were not able to include measure of the lag in program activity. We did attempt to construct lagged variables based on data provided within the PAs' annual reports. However, we concluded that the data contained in the PA annual reports and in the program tracking data were too dissimilar. We could not, with confidence, use the allocation of program tracking data by geography to allocate the data contained in the annual reports without making arbitrary assumptions regarding the differences between these two series.

- *Separate program activity by program type* – The evaluation team examined the impact of separating program expenditures and ex-ante savings into upstream and downstream activity, and examined the impact of lighting and non-lighting program activity on NAC.

- *Estimate first-difference form of each model* – The evaluation team estimated each of the models specified in their first differenced form (i.e., the year-over-year change in the dependent variable and corresponding independent variables).

The evaluation team employed these steps to estimate both county- and town-level non-differenced and differenced forms of each of the models identified in Table 2-6.

Table 2-6. Alternative Model Descriptions for PA Data C&I models

Model	Model Name	Model Description
Model 1	Employment Only	NAC is a function of employment plus time and geography fixed effects only
Model 2	Employment Plus Ex Ante Savings	NAC is a function of employment plus ex ante savings and time and geography fixed effects
Model 3	Employment Plus Total Expenditures	NAC is a function of employment plus time total program expenditures and geography fixed effects
Model 4	Upstream Plus Total Downstream Expenditures	NAC is a function of employment plus upstream and downstream program expenditures and time and geography fixed effects
Model 5	Upstream plus Lighting and Non-lighting Downstream Expendutures	NAC is a function of employment plus upstream and downstream expenditures and time and geography fixed effects. Downstream expenditures not are separated into lighting and non-lighting.
Model 6	Upstream Plus Total Downstream Savings	NAC is a function of employment plus upstream and downstream program savings and time and geography fixed effects
Model 7	Upstream plus Lighting and Non-lighting Downstream Savings	NAC is a function of employment plus upstream and downstream ex ante savings and time and geography fixed effects. Downstream savings are separated into lighting and non-lighting.

2.6.3 PA Data model summary of results

A. County-level model results

- There was no consistent pattern in the sign (+/-) or statistical significance of program variables or economic variables used to predict the level of estimated NAC.
- While employment only was a good predictor of NAC for small businesses, as the parameter estimate was both statistically significant and positive, this was not the case for both the large commercial and industrial models.
- For small commercial models, the "employment plus total program expenditure" model showed that when you include program expenditures, the model no longer provided a significant parameter on employment, but did provide a significant and negative parameter on the program variable. Further, as program expenditures were further split into upstream and downstream expenditures, the parameter estimates on expenditures were also negative and significant; however, the estimated parameter for employment was still not significant, which may suggest collinearity with program expenditures since businesses are more likely to be able to make capital expenditures as production expands.
- Separating expenditures into downstream lighting and non-lighting and upstream expenditures provided for significant parameter results; however, the sign on downstream non-lighting expenditures was positive. The savings models showed a similar pattern.

- The county-level results suggest that there is evidence that with a sufficiently long time series, it may be possible to isolate changes due to programmatic activity. In addition to expanding the time series, it may be desirable to expand the scope of the analysis to a regional analysis as opposed to just Massachusetts. This is because the introduction of variables that provide greater differentiation of programmatic activity across units (i.e., upstream and downstream lighting and non-lighting activity) shows greater significance of program variables. However, it should be noted that the timeframe being studied would be particularly challenging given the period of economic decline and recent growth, and given the recent escalation in programmatic activity.

B. Town-level model results

- None of the models for the large commercial PA Data model provided significant results for both economic activity and program variables.
- Three of the industrial models showed statistically significant parameters on the program variables, and one of the models showed a significant parameter on the economic variables, but the coefficient was negative.

2.7 CONCLUSIONS

This section provides conclusions from the review of methods and the two pilot studies.

2.7.1 Conclusions from the review of methods

The literature review illustrates that there have been a range of approaches employed to measure programmatic impacts at different levels of analysis using a range of data inputs. Each study offers different strengths and weaknesses relative to its ability to address the desirable properties of top-down models. While none of the studies provides a single approach for isolating net programmatic impacts from other influences on consumption, these approaches provided the following guidance for the Year 1 pilot studies.

- *Length of time series* – It is important to have a long enough time series to isolate changes to programmatic activity. The studies reviewed suggest that ten or more years of data are required. In terms of the Massachusetts programs, the level of program activity began sharply accelerating about four years ago. Therefore, even if the amount of available history was extended to ten years or more, there is only a limited time series following the acceleration period to measure changes resulting from the increase in programmatic activity. While this phenomenon may limit the ability to measure programmatic impacts in the near term, top-down analysis may become more viable the longer we keep running at the higher level of program activity.
- *Lagged program impacts* – It is important to consider the effect that programmatic activity in previous periods has on consumption of later periods. This is the lagged effect of programmatic activity on consumption. The effects of energy efficiency programmatic activity are not limited to the year that the activity occurred, but are actually cumulative over time, such that program expenditures made in some prior year may impact consumption in years following those expenditures. For example, if

program expenditures led to installation of energy efficient lighting two or three years ago, impacts from those installations would continue to occur in the current period. Spillover or market transformation is also seen as a potential supplemental impact to program expenditures expected to occur over time after the program intervention. These considerations require using multiple lagged terms for programmatic impacts. Work by Cadmus (2012) shows that consumption in the current period may be impacted by programmatic activity of up to five years previous. This recommendation held for the PA-Muni pilot study discussed in Section 5 that found statistically significant lagged effects for up to four years.

- *Lagged consumption impacts* –There is often variance in the timing of in which savings are realized from newly installed measures. This may result from differences between the date that measures are actually installed and the date they are recorded in the PA tracking data, learning curves associated with properly using the new technology, and other factors that cause a delay in the realization of savings. The literature suggests including a term for consumption in previous periods as one approach for addressing this lag in the realization of savings.

- *Measures of differing program types* – Horowitz (2004) shows that accounting for energy efficiency program expenditures or ex-ante savings alone may provide artificially low estimates of programmatic impacts by not accounting for market transformation impacts or impacts associated with upstream programs. Market transformation programs bleed across observational units, so models that do not specifically account for market transformation may understate effects observable from cross-unit analysis. In Massachusetts, the PAs have an extensive history of measuring and reporting program expenditures and savings across a wide range of measures and programs. These data will facilitate measuring changes resulting from expansion of programs covering non-lighting measures, custom measures, and upstream program offerings. These data will improve the ability of models to capture programmatic impacts within the PAs' territories in Massachusetts and regionally. However, extending the analysis to municipal territories or other states may be more difficult.

- *Use of scaled dependent variable* – The consensus among studies is that top-down models should seek to measure changes to energy consumption per unit (e.g., GSP, employee, household) in order to standardize estimates across locations and times.

- *Weather normalization* – Most of the existing studies include heating degree days and cooling degree days as explanatory variables, and use non-normalized consumption as the dependent variable. Further, these studies do not attempt to distinguish among heating and cooling impacts of programmatic activity. The recent Demand Research (2012) California pilot study marks a departure from this shortcoming, and uses normalized consumption based on utility records as the dependent variable. There are some advantages and disadvantages of this method but it is not possible to accurately assess their merits with the data limitations faced by the PA Data pilot study at this time. Use of weather-normalized consumption as the dependent variable mitigates some modeling concerns, but may introduce others.

- *Dependent variable (consumption) data series* – Our review of data shows that it is possible to employ account-level billing records to construct the dependent variable for the PA Data model at multiple levels of analysis. Because we normalized data at the account level—constructing normalized annual heating, cooling, and base loads—we were able to aggregate these series to any level of geographic resolution to construct top-down models. Further, using customer information, we aggregated the data to different sectors to construct sector-level models. Provided the PAs are able to provide sufficiently long data series (e.g., 10+ years of billing and tracking records), these data should provide an excellent source for constructing top-down models. Several of the prior top-down modeling efforts used aggregate data as reported to the Energy Information Agency (EIA), but that data was provided the PA level, and cannot be used to construct models at different levels of aggregation. Finally, provided the PAs are able to make data available from other states, it is possible to combine the Massachusetts PA data with that of other states to construct cross-regional models.

- *Programmatic impact variables* – The PA billing and tracking data reviewed demonstrates that it is possible to construct program impact variables at any level of aggregation. Further, program data can be segmented to provide differing impacts for upstream and downstream programs, as well as categorized by measure type to isolate heating, cooling, and base load impacts.
- *Lag in program and dependent variables* – The available time series at present limit our ability to construct lag program and dependent variables for earlier time periods. The PA-Muni model was able to assess the impact of programmatic activity from previous periods on current consumption, however, absent the necessary data series from previous periods, the PA Data model was not.
- *Exogenous variables* – There are a variety of sources that provide the necessary data to account for exogenous influences on consumption. The choice of data source and available series depends largely upon the desired level of aggregation. Finer resolution in the level of observation restricts access to some variables, as some data are not reported as proprietary information, such as sales volume and payroll.
- *Level of analysis* – The desired level of analysis for detecting programmatic effects should extend beyond Massachusetts. While the current year's analysis is only limited to measuring changes that occur within Massachusetts, to effectively capture sufficient differentiation in programmatic activity the longer-term study could be expanded to include neighboring states included in the PAs' service territories, including Connecticut, Rhode Island, New Hampshire, Maine, and New York. However, expanding to a regional model will likely result in some states having less data available than in Massachusetts. Because data availability is a primary concern for this approach, the prospect for expanding to a regional model should be weighed relative to the necessary trade-offs regarding data availability.
- *Source of consumption and program tracking data* – PA billing and tracking databases provided the most robust set of data for conducting top-down models. Because records could be captured at the account level and cleaned and weather normalized, many issues associated with consistency in measuring and reporting data discussed in this memo can be avoided. However, relying on billing and tracking records restricted the analysis to data that could be obtained from utilities. Our analysis showed that the data provided by the PAs to the C&I evaluation team to date was appropriate for conducting the PA Data pilot study, but that a longer time

series would be required to determine whether this approach will be fruitful in producing statistically significant results.

2.7.2 Pilot study conclusions

The two pilot studies presented in this report demonstrate that top-down modeling may provide a valuable tool in the set of tools used to evaluate net-energy impacts associated with energy efficiency programs.

The two approaches used in the pilot studies have differing strengths and weaknesses in terms of addressing the desirable properties of top-down models and modeling concerns identified in the existing literature. The PA-Muni pilot study employed a relatively long time-series, 15 years, which allowed the model to examine possible cumulative effects of programmatic activity on consumption over time through use of various lagged program expenditure terms. This was a key finding of the literature review, and the model results indicated that these lagged terms were, in fact, instrumental in developing a model that produced statistically significant results. The PA Data model had a much more limited time series, 3 years, and consequently was not able to account for the cumulative impact of programmatic activity. Similarly, the PA-Muni study was able to address a number of other influential factors related to the time-series, which the PA Data study was not able to address, such as the impact of building codes, technology trends, and time-specific fixed effects. Due to the overarching restrictions on the PA data resulting from the limited time series, the evaluation team did not address a number of other modeling concerns that may also limit the success of this technique, such as industry-level segmentation, the impact of building codes and technology trends, and the most appropriate treatment of weather normalization.

Both modeling approaches rely on differences in program activity across geographies and time to isolate the effect of program activity on consumption. The PA-Muni model contrasted consumption in the PA territories, which have relatively high levels of programmatic activity, to consumption in municipal utility territories, which have relatively low levels of programmatic activity. This contrast provides a stronger basis for measuring net impacts. In effect, the low-program muni territories represented a comparison area that was used to remove naturally occurring energy savings from gross impacts. Because the PA Data model relies exclusively on data within the PAs' territories, the PA Data models have a weaker program signal in their contrasts across time and units; the PA Data models have the advantage of more detailed data that can help in controlling for non-program factors and support the isolation of program attributable impacts from naturally occurring savings.

The PA Data model offers an approach to address many questions that are important for planning, policy, and implementation of energy efficiency programs, which the PA-Muni approach cannot address. Because the PA Data models were developed from account-level billing and tracking data, separate models can be developed to examine the impact of differing program offerings, or the relative contribution of various customer segments to savings. While the models were not statistically significant, the PA Data pilot study showed that the ability to model different customer segments (i.e., large commercial, small commercial, and industrial customers) provide differing measures of programmatic impacts. Further, the ability to break out various measure and program types may also influence savings estimates. This information is important for policy, planning, and implementation, as it allows for the development and implementation of targeted program offerings. The PA Data approach provides this level of flexibility in modeling, while the PA-Muni approach does not. Both studies face differing, but substantial data limitations.

In summary, the evaluation team explored two forms of analysis though this first year of the top-down research study. While differing in their technique, these studies followed best practices identified in the literature. The PA-Muni model benefitted from a wider range of variation of activity and a longer time series. Its results indicate that this is a promising approach, but further work is needed to ensure that robust results can be obtained. The PA Data model had the advantage of being able to consider a finer level of granularity, but was disadvantaged by generally less variability of program activity across geographic units within PA territories, and a shorter time series available at this time. The PA Data approach has not yet been applied to the residential sector, as the data were not available earlier, but it will be shortly. For both modeling approaches, data assembly was a non-trivial effort. The conclusion of this research is that top-down modeling as a set of modeling approaches is promising, but needs more study.

A. Conclusions from the PA-Muni data pilot study

While the findings from this pilot study are preliminary, initial model results look promising as a supplemental approach to the bottom-up methods used to estimate net-energy impacts. A number of model results—both for residential and C&I—indicate that energy efficiency program expenditures had a statistically significant effect on reducing electricity consumption, but the effect was less consistent for the C&I sector. While additional research is necessary to refine the PA-Muni top-down models, understand the stability of the model results, and reduce the size of the confidence intervals and model specification uncertainties, this technique appears to offer a potential means of validating the bottom-up estimates of gross and net savings. The team has identified a number of next steps for model refinement that should help improve the precision of the savings estimates.

The results showed that savings estimates are sensitive to model specification, particularly the inclusion of various lagged measures of programmatic activity. Among the various residential model specifications tested, the fixed-effects model with four lags for energy efficiency expenditures appeared to perform the best, as the coefficients of the energy efficiency expenditures variables for the current year and the past four years were jointly statistically significant. The model accounted for the lagged impact of energy efficiency program expenditures on electricity consumption and the leakage of PA lighting program rebate dollars to municipal utility service territories. The findings indicate that the impact of cumulated energy efficiency program expenditures on current consumption increases with the number of previous years included in the cumulated sum. This reflects the importance of including lagged program year effects in the model.

Among the various C&I model specifications tested, the fixed-effects model with three lags for energy efficiency program expenditures appeared to perform the best. Adding more lags to the model yielded results that were not in the expected direction. The estimated impact of one dollar spent for the C&I energy efficiency programs was somewhat smaller than that for the residential sector, but the difference was not statistically significant. Finally, top-down savings estimates from this model were very close to the corresponding reported annual savings for the PA C&I programs, but savings estimates were highly dependent on model specification, which warrants further study.

This study also draws attention to an inherent limitation of macro-consumption methods. While the team was able to detect energy savings, further model refinement is necessary to reduce the confidence intervals around the top-down estimates. The team has identified a number of next steps for model refinement that should help improve the precision on savings estimates.

In addition, traditional bottom-up approaches measure energy impacts at the individual customer level, capturing customer-level detail that is beneficial in program design, marketing, and policymaking, such as the importance of certain program designs or measure offerings to overall portfolio savings and measurement of free ridership and spillover effects. While explanatory variables measure the relative influence of demographic and economic factors on savings, top-down estimates based on data aggregated at too high of a geographic level (i.e., PA/utility level) may lose the ability to provide meaningful estimates of variables important to all interested parties. This conclusion suggests that this top-down method offers an additional tool for triangulating the overall impact of energy efficiency programs, but this technique is not a one-size-fits-all approach for addressing all program evaluation needs.

An important challenge for this study was to collect consistent electric program data across all PAs and municipal utilities. While the evaluation team attempted to collect detailed time-series data on program activity, the only consistent piece of data that the team was able to gather across all PAs and municipal utilities was the annual total electric program expenditures.

If the PAs and the EEAC decide to move forward with the second phase of the study, which would entail the collection of PA energy consumption and energy efficiency program data at the town or city level, then these data could potentially help increase the precision of the savings estimates and thus improve both the residential and C&I models. Moreover, these PA data would allow commercial sector data to be separated from industrial sector data, which should improve the results because the electricity consumption for the commercial sector should be more stable than that of the combined commercial and industrial sectors due to high volatility of consumption in the industrial sector. Finally, if the gas PAs can provide gas consumption and energy-efficiency program data at the town or city level, then comparable gas macro-consumption models could be run to estimate the impact of gas program activity on gas consumption in the residential and commercial sectors in Massachusetts.[19]

B. Conclusions from the PA Data pilot study

This study sought to determine whether top-down methods should play a role in the overall portfolio of attribution methods both in terms of the recommended role on an ongoing basis as well as the methodological approaches that are recommended. The methods review portion of this study concluded that top-down modeling may provide an additional tool in the set of tools used to evaluate the portfolio of programs. However, the top-down approach cannot replace bottom-up approaches, as bottom-up techniques provide much information that top-

[19] These data would also allow for developing a model to test the impact of the PA residential upstream lighting programs on residential gas consumption. If the interactive effects are large, this test might provide corroborating evidence to suggest that further research is needed to ensure that the most accurate assumptions for program planning and savings claims are made.

down techniques cannot provide. Information pertaining to the relative contribution of different activities to overall savings can assist in the allocation of resources across the portfolio of programs, or help with program design. Such information cannot be obtained from top-down approaches. Moreover, the review of data suggested that top-down techniques face a variety of challenges pertaining to the reporting and availability of data that limited the effectiveness of these techniques. The PA Data model confirms that data availability was a primary obstacle to successful estimation of the models presented in this report. While our review of methods indicated that load forecasters within each PA interviewed used relatively simplistic models to estimate demand, forecasters reported that they were not able to tease out program effects from their load forecasts.

The model results estimated in this study were consistent with this finding; however, these results were limited to just three years of data. One factor that may lead to meaningful estimates is the availability of a longer time series. With only three years of data, the PA Data pilot study portion of this report shows inconclusive evidence that the approach we employed is able to detect programmatic impacts. Our analysis demonstrates the ability to construct the necessary variables at the desired levels of aggregation, and the ability to systematically test a variety of models. Some of the models showed statistically significant parameter estimates for measures of either programmatic and/or economic activity; however, these results were not consistent across model specifications or levels of geography.

One could conclude that the statistical significance of parameter estimates for some models is an indication that the models would perform well given a sufficiently long time series. However, one could also argue that the significance of terms measuring programmatic activity is the result of noise in the model, and the true models are ones in which there are no program effects. While our analysis does not provide sufficient information to make a determination that program effects can be detected with certainty, some model results do show statistically significant parameter estimates on the program variables. Further, our review of the available literature suggests that effective top-down modeling of energy impacts requires a sufficiently long time series to account for:

- *Variation in the level of program data over time* – Our time series included only three years of data, which all occur during a period of economic recovery and rapid increase in programmatic activity.
- *Multiple lags in programmatic activity* – Previous research, as well as the PA-Muni pilot study, illustrate the importance of using multiple lags in both the program variables and dependent variable.
- *Use of first-difference in the dependent and independent variables* – By including only three years of data in the model, the first-difference models included in this study contain only two years of data for unit of observation.

Absent these measures, it is not surprising that the model results did not provide statistically significant parameter estimates, that the results were not consistent across levels of aggregation, and that the results were not stable in terms of the significance of variables or their sign. Despite the lack of significant results at this point, the evaluation team believes that the PA Data pilot study model approach will likely improve given a long enough time series. We draw this conclusion based on the following evidence:

1. Loughran and Koulick (2007) and Violette (2014) demonstrated that successful top-down models with at least ten years of data can successfully account for programmatic impacts.

2. The PA-Muni model, which includes more than ten years of data over the same population, was able to provide statistically significant savings estimates. Given the PA Data model has the ability to examine impacts associated with more specified programmatic activity, it is likely that given a sufficiently long time series, the PA Data model would also produce significant results.

3. The PA Data model is able to capture variation across program and customer types that provides valuable information for program planning and implementation, and allows program evaluators to determine the effectiveness of differing program offerings and/or marketing strategies.

However, compiling a sufficiently long historical time series retrospectively would be costly, and may not be possible due to limitations in electronic record keeping. Therefore, a more practical approach may be to construct the historical series back five years and continue collecting the necessary data going forward for future analysis.

Apart from these time-series related limitations, the following factors may also be responsible for the lack of significance in the model:

- The current C&I models do not account for differences in consumption and programmatic activity by industry, which the literature has shown is an important factor for isolating program impacts.

- The model results could be impacted, in part, by the normalization process. However, the evaluation team did test a set of models that used non-normalized consumption as the dependent variable and included HDD and CDD as independent variables. These models did not perform better than the models using normalized consumption, and did not have statistically significant parameters for the weather terms.

- The models did not include terms to measure the impacts of changes to building codes or technology.

2.7.3 Limitations

This section reviews important limitations to the analysis.

A. Limitations of the PA-Muni pilot study

- *Aggregate data* – The primary limitation of this study is that it relies on data aggregated at the PA or muni level to estimate energy impacts. Consequently, separate models cannot be developed for different customer segments or program offerings. Lack of segmentation may be partially responsible for the wide confidence intervals for savings estimates.

- *Municipalities are not a true comparison area* – While there is relatively limited programmatic activity in the muni territories, there is still some activity. Moreover, there is likely leakage or cross-unit spillover from the PA to muni territories. These factors limit the ability of the model to adjust for naturally occurring adoption of energy efficiency technology.

- *Data availability* – The evaluation team expended considerable effort to capture programmatic data from the PAs and municipal utilities. While the publicly available energy consumption data was relatively clean, the publicly available programmatic activity data was not. Obtaining the program expenditure data required considerable effort from the evaluation team, the PAs, and municipal utilities. A number of municipal utilities were not able to obtain the necessary data, as it would require coding data reported on paper forms.

B. Limitations of the PA Data pilot study

- *Fuel prices not reported at the same level of granularity as unit of analysis* – The evaluation team did not identify any data for actual average electricity prices ($/kWh) at the county or town level. The DNV GL billing data set contained rate codes, and billing amounts could therefore be imputed, but there were many missing values and other data quality concerns.
- *Absence of lagged program activity and consumption variables* – The literature review identified the importance of incorporating lagged program and consumption variables into the models. Because the existing time series was limited to just three years of data, we were not able to construct lagged variables using the consumption and program tracking data. The evaluation team attempted to construct lagged series based on data available through the PAs' annual reports; however, we were unable to construct a series that did not introduce bias into the model.
- *Absence of building codes* – The evaluation team attempted to construct variables to account for the impact of building codes on consumption; however, due to the limited time series, there was insufficient variation in the building code data in include in the model.
- *Limited time series during periods of rapid expansion of both economic and programmatic activity* – Our review of the available consumption, program tracking, and economic activity variables revealed a fundamental limitation of the present analysis. During the three years of observation for this study, the three critical time series for the analysis underwent a period of rapid expansion. Figure 2-1 presents the change in employment and NAC for 2012 and 2013 relative to 2011. Figure 2-2 presents the change in program expenditures and ex ante savings for 2012 and 2013 relative to 2011. Given the limited time series, it is likely that the model results will be impacted by the corresponding increase in these three time series. Without a longer time series, or substantial variation between observational units, it is likely that the model will not be able to differentiate between increases in programmatic activity and reductions in consumption. It is important to have a long enough time series to isolate changes to programmatic activity. The studies reviewed suggest that ten or more years of data are required. In terms of the Massachusetts programs, the level of program activity began sharply accelerating about four years ago. Therefore, even if the amount of available history was extended to ten years or more, there is only a limited time series following the acceleration period to measure changes resulting from the increase in programmatic activity. While this phenomenon may limit the ability to measure programmatic impacts in the near term, top-down analysis may

become more viable the longer the PAs keep running at the higher level of program activity.

Figure 2-1. Percent Change in Annual Employment and Actual and Normalized Annual Consumption (2011–2013)

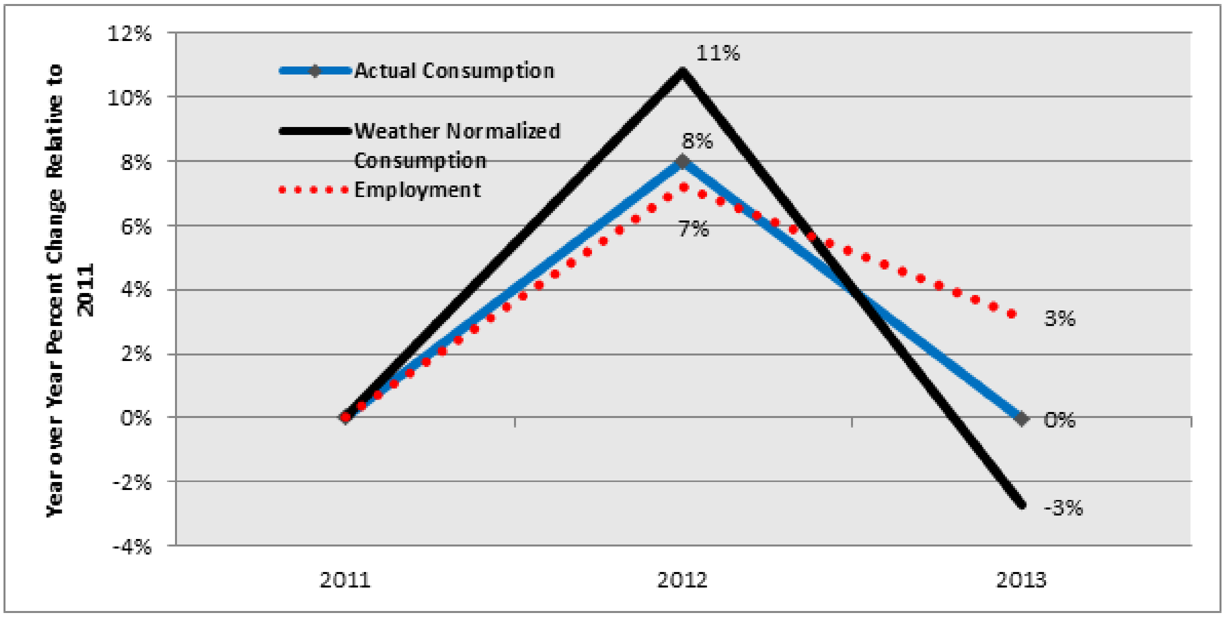
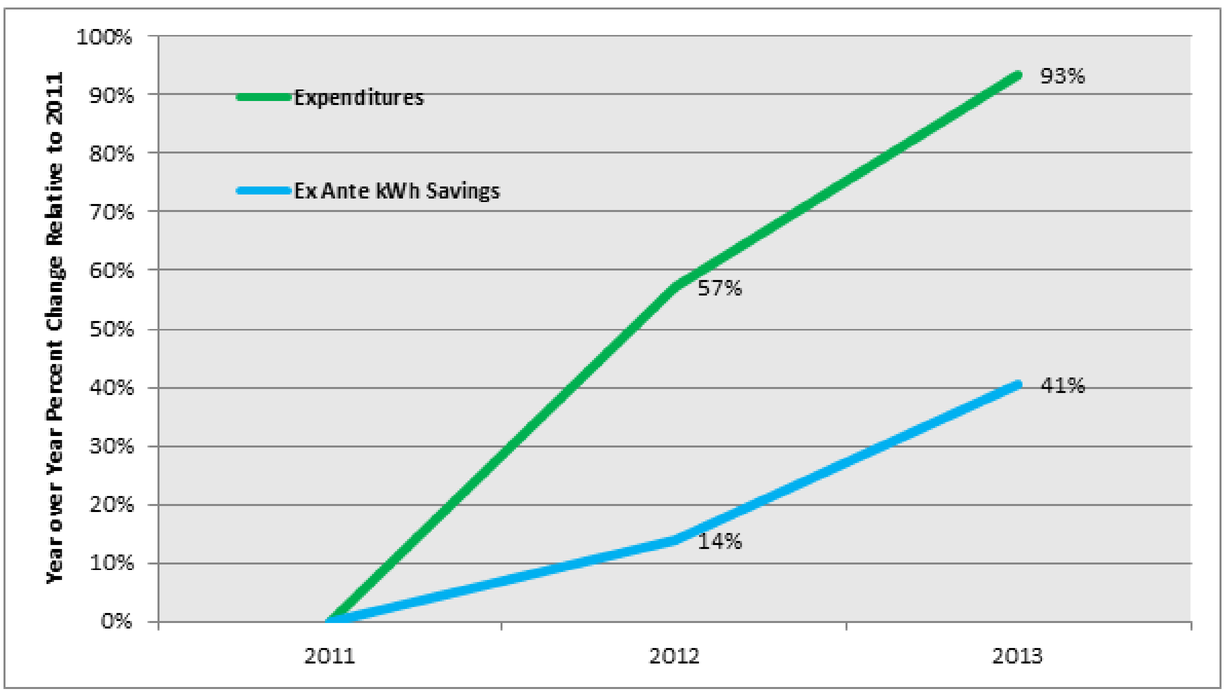


Figure 2-2. Percent Change in Program Expenditures and Ex-ante Savings (2011–2013)



- *Limitations in use of per capita income* – Income is not very indicative of C&I economic conditions, as it can be skewed by individuals with relatively high salaries, such as CEOs. Consequently, this variable is not used by the load forecasters at the PAs.

- *Isolating industry- or sector-level differences* – There may be considerable variation in the savings and consumption by industry sector. However, economic series by sector are only available at the county level or for the major metropolitan areas.[20] Population, a variable that is available at all levels of aggregation, could theoretically serve as a measure of market size, but it is more closely associated with residential consumption than commercial or industrial, which is also true of per capita income. The evaluation team believes that per capita income and population is likely to be correlated to employment if they are both included jointly in a statistical model. Population is available at the town level or census-track level from the American Community Survey data set.

[20] http://www.bls.gov/cew//.

3. INTRODUCTION TO THE YEAR 1 TOP-DOWN WORK[21]

The work presented in this document is part of a multi-year initiative designed to assess the utility of top-down modeling as a viable technique for evaluating energy efficiency programs in Massachusetts. Specific efforts undertaken by the evaluation team in 2013/2014, which are discussed in this report, include the following:

- *Assessment of top-down modeling methods.* Section 4 of this document discusses advantages, disadvantages, and necessary properties of top-down methods, and evaluates 15 top-down research studies that were used to estimate impacts associated with energy efficiency programs. Based on this literature review, we developed recommendations for specific methods to be used in Massachusetts.
- *PA-Municipal Utility pilot study.* Section 5 of this document summarizes the Year 1 research concerning the first pilot study. This pilot study contrasted changes in consumption in the residential and C&I sectors relative to programmatic activity, and compared results for the PAs and municipal utilities. This model relied on aggregate consumption and program expenditure data at the PA or municipal utility level. The evaluation team was able to include a 15-year time series, allowing the model to overcome some modelling limitations resulting from shorter time series. Consequently, some preliminary models from the pilot study were able to generate statistically significant and substantial net savings estimates. In principle, the models provide estimates of program-attributable savings, net of naturally occurring reductions to energy consumption, by contrasting consumption in the PA territories (where there has been substantial programmatic activity) to that in municipal utility territories (where there has been minimal to no programmatic activity), while controlling for other differences between the muni and PA territories. There are a number of models that show particularly promising results, namely models that account for the lagged and cumulative impact of programmatic activity. However, further research is needed to refine model specifications, establish the stability of model results, and improve model precision.
- *PA Data pilot study.* Section 6 of this document discusses considerations for and preliminary results from the PA Data pilot study C&I models. For this study, we used PA-provided consumption and program tracking data to construct a set of C&I top-down models.[22] Model estimation in this stage of evaluation did not attempt to estimate net savings using these models. Rather, the pilot study investigated whether the data provided evidence of sufficient signal between programmatic activity and consumption to warrant further study. Based on this pilot study, we developed recommendations for modeling techniques that may contribute to longer

[21] This report is split into two major segments: a summary (found in Section 2) and a detailed discussion that starts with this chapter and continues through the end of Section 7. Since Section 2 is a summary of Sections 3 through 7, some of the content will be redundant for readers who are reading this document from start to finish.

[22] The PA Data pilot study for the residential sector will be provided in 2015, as the three years of consumption data required for modeling are not yet available for analysis.

ongoing evaluation efforts, and established the preferred model specifications and data requirements for the next phase of top-down modeling efforts.

3.1 MOTIVATION FOR INVESTIGATING TOP-DOWN MODELING

Traditionally, energy efficiency savings have been computed using a bottom-up approach that incorporates a range of techniques to estimate gross and net energy savings for individual measures/end-uses, programs, or groups of programs. In Massachusetts, there is some concern that these attribution methods are not fully capturing net impacts because programs are large and interactive, making it difficult to isolate the effects of any one program using just a bottom-up approach. Consequently, the bottom-up approach may be understating (or overstating) net program impacts. A top-down approach may provide a complementary approach to bottom-up methods for estimating net impacts. There is considerable interest among Massachusetts stakeholders in exploring top-down methods in order to help assess this issue.

While the specific bottom-up techniques may vary from evaluation to evaluation, these evaluations typically involve one or more of the following:

- *Develop engineering estimates with assumed parameters based on program tracking data* – Estimate savings by applying engineering formulas and per-unit savings from technical resource manuals to program tracking data and assumed parameters (e.g., hours of use, baseline efficiency level).
- *Verification/measurement* – Conduct participating customer surveys, on-site interviews, and/or metering studies to verify measures installed and to capture actual usage patterns.
- *Compute gross realized savings* – Employ engineering and/or statistical techniques to develop verified gross savings, incorporating information from verification, on-site observations, and/or consumption data analysis.
- *Attribution studies* – Conduct customer and/or supply-side surveys to estimate program attribution (1 - free-ridership + spillover) either by program or across a range of programs.

Despite the common use of the bottom-up approach in program evaluation, evaluators have long recognized the following challenges:

- *Evaluation scope* – Each evaluation often requires separate primary research efforts for verification and on-site data collection.
- *Program impacts may not be additive* – The validity of the bottom-up technique relies on the assumption that program impacts are additive, such that the sum of individual measure-level savings from each program can be aggregated to provide an estimate of total savings. This would imply that program impacts are independent, meaning that impacts associated with one program do not influence the impacts of other programs.
- *Difficulties in estimating net adjustments* – Evaluators expend considerable resources attempting to estimate net adjustments such as spillover, free-ridership, snapback, and impacts associated with codes and standards. There are also

concerns among stakeholders about relying on self-reported information, alone, to estimate what would have occurred absent the program.

A 2013 study by Ralph Prahl et al. notes that "Market effects research studies seek to find evidence of and measure spillover savings that reflect significant program-induced changes in the structure or functioning of energy efficiency markets."[23]

Market effects studies frequently use primary data obtained from a range of market actors to measure the impacts of program activities on awareness and acceptability of energy efficient products, as well as changes in promotional activity, stocking practices, product availability, and costs. Where possible, these studies are used to provide estimates of program-attributable savings not tracked by programs or by traditional spillover studies. Compared to bottom-up studies, they tend to include a larger range of net impacts (e.g., spillover, market effects, snap-back, and free-ridership), but their results are not as specific to particular measures or, in some cases, even programs.

In contrast, top-down studies seek to capture net program effects regardless of specific program activities. The top-down modeling approach captures the combined effect of all programs, reflecting any non-additive effects (i.e., impacts from different programs or measures) that should not be added together because the savings are correlated). For example, measures installed to improve refrigeration units may result in lower consumption from central air-conditioning as the refrigeration units generate less heat. All net effects, in principle, are captured in the measured program effect.

In addition to program evaluation, top-down modeling may provide a useful tool for a number of other regulatory considerations. Because top-down models can provide estimates of the total energy savings across the range of programmatic activities, they may be valuable tools for environmental regulators' assessment of the overall carbon impact of energy programs. Further, the ability to model the total energy consumption for a service area can be used to supplement energy forecasts used for long-term resource planning.

3.2 OVERALL STUDY GOAL

The PAs and EEAC expressed interest in determining whether top-down modeling should play a role in net energy impact evaluation. As leaders in energy-efficiency programs and evaluation, the PAs and EEAC are exploring the potential for top-down techniques to be an additional tool to complement bottom-up approaches to measuring impacts from energy efficiency programs, as top-down techniques have the potential to capture interactive effects between programs, and market effects. However, these approaches present a number of challenges that limit their ability to serve as a viable alternative to the traditional bottom-up approach for measuring net-to-gross impacts from existing energy efficiency programs.

The goal of this study is to develop and apply multiple top-down methods for Massachusetts, and to understand the strengths and limitations of those methods relative to the traditional bottom-up approaches to measuring net energy impacts. This is the first year in a multi-year top-down research and development project during which the evaluation team investigated the use of various modeling approaches and initiated a longer term set of pilot studies. We initiated these pilot studies to identify and begin to compile the necessary data inputs,

[23] Prahl, Ridge, Hall & Saxon. "The Estimation of Spillover: EM&V's Orphan Gets a Home." IEPEC. 2013.

evaluate different model specifications, and estimate model parameters to suggest the course of the subsequent years' research activities.

3.3 RESEARCH OBJECTIVES

The specific research questions that will be addressed through the multi-year research project include the following:

- What role should top-down methods should play in the overall portfolio of attribution methods both in terms of the recommended role on an ongoing basis and the methodological approaches that are recommended?
- How can top-down methods be linked to long-term demand forecasts and/or provide useful adjustments to such forecasts, if at all?
- What data are required to support recommended methods on an ongoing basis?
- Given data limitations in the near term (i.e., the current evaluation year), does one or more top-down modeling techniques provide sufficient evidence to support pursuing a multi-year research plan?
- Which approaches, or aspects of those approaches, should the Massachusetts PAs employ for a multi-year evaluation effort?
- What are the appropriate model specifications, data requirements, and expected outcomes for the Massachusetts PAs to implement a top-down multi-year modeling research plan?
- What role can a national-level top-down model serve in 2015 and beyond, and how would the evaluation team construct national models, from which data, and over what period?

To help address these questions, the evaluation team focused on the following activities during the 2013–2014 research period; the results of these activities are discussed in this document.

- Reviewed existing top-down modeling techniques, and recommend specific methods to be used in Massachusetts.
- Obtained the necessary data for employing one or more agreed-upon approaches.
- Implemented multiple agreed-upon approaches in parallel to evaluate and recommend top-down modeling techniques that may contribute to longer ongoing evaluation efforts.
- Established the preferred model specifications and data requirements for the Year 2 and beyond top-down modeling efforts for regional and national models using different approaches. Due to time and data limitations, we restricted the Year 1 analysis to electric consumption models only.

For the first year of the study, two separate top-down modeling approaches were selected to provide a range of net savings estimates in the near term. The first approach uses 15 years of aggregate electricity consumption data for the PAs and municipal utilities in Massachusetts. The purpose of adding data from the municipal utilities to the model was to provide a measure of the baseline level of program activity, because municipal utilities had no or significantly lower levels of energy program activity than the PAs until recently. The disadvantage of this approach is that top-down estimates based on data aggregated at too high of a geographic level (i.e., PA/utility level) may lose the ability to provide meaningful estimates of variables important to all interested parties.

The second pilot study used individual-level data provided by the PAs. The advantage of this approach is that the use of PA data provides detailed information regarding program activity level and the ability to aggregate by whatever dimensions are useful (such as PA territory or town). However, the disadvantage to this approach is that it lacks a “no-program” situation, and has a short time series of PA consumption and program activity data available for the first year of the study. These pilot studies are presented in Sections 5 and 6 of this report.

4. ASSESSMENT OF TOP-DOWN MODELING METHODS

Top-down modeling is an econometric approach to measure program impacts using aggregate cross-sectional and time series data. The top-down models measure changes to aggregate energy consumption relative to changes in energy efficiency programmatic activity, prices, and other economic factors. The goal of this type of modeling is to isolate the effect of program activity from other natural changes and policy variables.

Top-down techniques use a holistic approach by estimating program impacts across all energy efficiency programs in a given geographical region or service territory, rather than running separate studies for each program (or measure/end-use within a program).

In the following sections, we discuss the analytical framework for assessing top-down modeling methods, and review the existing literature focused on top-down research.

4.1 ANALYTICAL FRAMEWORK FOR REVIEWING TOP-DOWN MODELING APPROACHES

This section summarizes the expected outcomes and requirements for the use of a top-down method in evaluating the expected savings from a portfolio of programs. First, we discuss the possible benefits of including top-down modeling in the set of tools used to evaluate energy efficiency programs. Next, we discuss some important limitations to the use of top-down methods in program evaluation. Finally, we discuss the necessary requirements for successful top-down models, which will form the basis for our comparison of techniques in the following sections.

4.1.1 What could top-down modeling provide to the evaluation of energy efficiency programs?

Evaluators and commissioners are intrigued by the potential of top-down techniques because of their potential to provide low-cost supplemental or alternative estimates of net program savings. The appeal of this structure is that, in principle, it captures the full program effect, including free-ridership, spillover, market effects, and snapback. The cited benefits of top-down techniques are summarized as follows.

- Assuming researchers can properly control for the exogenous factors, regression-based techniques capture the aggregate effect of program activities.
- Because models typically employ aggregate consumption and economic data, they typically do not rely on survey responses to hypothetical questions.
- Regression models can be developed to capture the combined effect of multiple measures and programs simultaneously.
- By using time series data, they can provide estimates of the cumulative effect of programs to date, rather than isolating the effects of a particular period of program activity.

A properly structured top-down model can potentially make the following contributions to the set of tools used to evaluate energy efficiency programs:

- Provide relatively inexpensive estimates of program-induced savings for all units in the study. Because top-down models can examine the change in total energy consumption over time, relative to changes in the level of program activity, it is theoretically possible to measure the overall impact of all programmatic activity across the portfolio of energy efficiency programs net of free-riders, spillover, and snapback. However, as is discussed in Section 4.1.2, results are not necessarily net of free-riders. Rather they include incremental program induced savings as well as some level of naturally occurring savings due to the inability to fully remove free-riders from self-selected program participants.

- Provide expected program-induced savings for a unit with particular characteristics. Top-down models are essentially macroeconomic models that relate energy consumption to a series of program and non-program (exogenous) variables. The non-program variables are used to describe the relative impact of a population's demographic or firm-o-graphic characteristics. Consequently, by using assumed values for these characteristics, evaluators, program designers, and policy makers can estimate the expected level of savings from populations with the prescribed characteristics, given an assumed level of program activity.

- Provide combined effects of all cumulative program activity for a particular unit, including spillover and snapback. By measuring the change in aggregate consumption relative to the level of program activity across the combined set of programmatic activities over time, top-down models naturally account for the interaction among program impacts. Consequently, they provide measures of the cumulative effect of the portfolio over time, including spillover (i.e., untracked program attributable savings).

- Provide confidence intervals and precision levels for net energy savings from the portfolio of programs. The statistical estimation of top-down models yields estimates of the standard error that can be used to construct confidence intervals around the savings estimates.

4.1.2 What top-down modeling cannot provide to the evaluation of energy efficiency programs

Despite the potential advantages discussed above, savings estimates derived from top-down models present a number of technical challenges—some shared with bottom-up estimates, and some unique—that may limit their ability to provide a single solution for program evaluation. There are several challenges to developing robust top-down models:

- Models provide average, not specific, program effects. The extent to which some programs are more efficient than others will not be reflected in the model results.

- Data availability limits possible model specifications. To estimate such a model, we need economic, price, and programmatic activity data for each area and time period. Compromise is typically necessary between the ideal specification and the types of data available at various levels of aggregation. Before producing savings estimates using a set of top-down models, researchers must articulate the potential limitations and biases of available data and alternative choices.

- Spillover between study units reduces apparent program effects. To the extent that program activity in one area and time period reduces consumption in another area or time period, the program effect estimated by the regression will be dampened. The difference in consumption per unit across areas and time periods with different levels of program activity will be reduced, so that the measured program effect will be less. In the current pilot studies, the evaluation team defined units of measurement (geographic areas and time intervals) in a way that limited the potential for spillover among units, or explicitly incorporate a cross-observation metric to limit the potential for bias due to spillover. Inclusion of a cumulative activity metric at least partially accounts for spillover from one time period to the next. Similarly, future research may include a statewide or regional activity metric to account for cross-area spillover. If correctly specified, the non-spillover program-induced savings will be smaller, as the spillover will account for some of the program impacts in the model. However, the true program impact (non-spillover program induced savings + spillover) will likely result in larger overall impacts.

- Omitted or incorrectly specified variables can bias the results in unknown ways. A key premise of the top-down model is that all non-program factors affecting consumption are accounted for in the model. However, model specifications tend to be limited by available data. To the extent that there are other drivers of consumption not accounted for in the model structure, or not incorporated with the best metric or form of relationship, all the model coefficients including the program coefficient can be biased. To reduce omitted variable bias, the evaluation team considered and, to the extent possible, tested alternative drivers in the models. Our ongoing research is focused, in part, on identifying the types of drivers that may be missing to determine the potential effect of their omission.

- Self-selection effects can bias the results. If program activity tends to be higher in areas and times where customers would have a natural tendency to adopt more (or less) energy efficiency on their own, and there is no good metric of this natural tendency that can be incorporated to control for it, the estimated program effect will be overstated (or understated). This is a special case of omitted variable bias. This challenge can be mitigated if the regression includes areas and time periods with and without program availability, or at least with substantially different programs available, which was done in the PA-Muni model. To control for self-selection bias, assuming no spillover, further research can also develop instrumental variables for the activity level, similar to the use of self-selection correction terms for individual customer regression models. These variables can be used to predict program activity level as a function of known aggregate customer characteristics and economic factors. We would then enter the predicted activity level from these variables into the model as another independent variable in the top-down consumption model. This approach does not completely resolve the self-selection issue, but it can provide a useful perspective.

These technical constraints result in the following limitations in top-down models:

- Inability to obtain savings estimates net of free-riders. While some authors have argued that top-down models provide true measures of net savings, net of free-riders, this is not necessarily the case for a number of reasons. First, top-down models that look at varying levels of program activity within a geographical region (i.e., a state or PA territory) are unlikely to contain a true measure of the “no program” scenario. In the PA-Muni modeling approach, using municipal utilities with limited programmatic activity may provide a close approximation to a true control group. As has been discussed extensively regarding bottom-up approaches, nonparticipants are not good measures of participant behavior absent the program, as participants are inherently different from nonparticipants. While it is possible to use comparison groups (i.e., other states), comparison groups are likely to have a lower percentage of energy efficient equipment than a true control group within the population of Massachusetts customers. This is due to demographic and socioeconomic differences between Massachusetts and comparison-area states, and cultural similarities among Massachusetts residents concerning energy conservation and other environmental issues. Further, if it were possible to extend the time series far enough back in history to capture the “no-program” scenario, such a time series would be so long that it would be difficult, if not impossible, to control for all the relevant exogenous changes.

- Inability to provide separate free-ridership, spillover, and market effects estimates. Estimates of free-ridership, spillover, and market effects are important measures for policy makers and program designers. Top-down models may provide measures of “near net” savings, but they do not produce separate estimates for free-ridership, spillover, and market effects. Rather, these metrics are imbedded in the final “near net” savings estimates. In contrast, bottom-up techniques implement separate studies to measure free-ridership, spillover, and market effects.
- Inability to provide an isolated effect of a particular program and year. Information concerning the relative contribution of separate programs to the overall savings associated with a portfolio of programs is important to policy makers and program designers. This information provides for more efficient allocation of resources to meet savings goals.
- Inability to identify which groups of measures are performing better, or worse, given their characteristics. Estimates of the total savings from a portfolio of programs derived from bottom-up methods incorrectly assume that savings are additive, and thus are likely to over- or underestimate true net savings. However, bottom-up techniques do provide for estimates of the relative contribution of different measures on overall savings. Such information is valuable to policy makers, program designers, and implementers.

4.1.3 Desirable properties of top-down modeling studies: what is needed for top-down methods to work well

Top-down models must meet the criteria listed below to provide a useful tool for evaluating energy efficiency programs. While it is unlikely that any one tool will contain all of these attributes, these criteria provide a basis for assessing the tradeoffs among methods. The choice of method will result in different strengths and weaknesses. No one model will provide a “silver bullet” to address all the relevant concerns. Rather, running a variety of models over time is most likely to provide a more comprehensive view of net program savings that can be used in conjunction with bottom-up techniques to triangulate the net impact of programmatic activity on energy consumption.

The criteria include the following:

- *Ability to establish the counterfactual (no-program) scenario* – Understanding the true extent of a program's impacts requires information regarding the level of consumption absent from any programmatic activity. Without using a random experimental design with a true "control group," the counterfactual scenario is typically simulated using a number of techniques, each of which has certain limitations. One option is to extend the timeframe long enough to include a period with no programmatic activity. However, this option is seldom possible because energy efficiency programs have been offered for the past several decades. Therefore, it is not possible to identify a time period when there has been no programmatic activity. Extending the time period back to a pre-program period would require a time series of 30 years or more. Such an extended time series makes it impossible to control for all relevant exogenous factors. Even if comparison areas are used, there are very few areas with no programmatic activity for which exogenous differences in the population would make them desirable comparison groups. Consequently, most modeling exercises measure changes in the degree of programmatic activity relative to some base level.

- *Diversity of program activity levels across units of observation (time-geography combinations)* – The unit of observation refers to the level at which data are aggregated for use in the model as observational units (i.e., the data points used to estimate the model). Energy consumption data can be aggregated to the PA, census tract, ZIP code, town, county, or state level, so that each aggregate unit serves as an observation in the model.[24] In order to detect the impact of programmatic changes on consumption, the level of programmatic activity for each unit of observation must be reported consistently. Moreover, the program variables must provide sufficient variation from one location-time combination to relate changes in consumption to the program variables.

- *Consistent relationship between program activity and savings* – The influence of program variables and consumption must be consistent across units of observation. Data that are reported differently for separate units of observation (time-geography combination) will impact the predictive capability of the model. This may result in over- or underestimating program savings when the observed changes in consumption or program variables are due to differences in reporting or some other structural difference in the data itself.

- *Minimal effect of one area on another (cross-area spillover)* – Non-program tracked spillover from one area to another will assign exaggerated program effects in the region experiencing the spillover, and reduce the apparent impacts of the region that is the source of the spillover.

[24] The concept of "unit of observation" differs from the level of analysis. The unit of observation refers to the data points that serve as observations in the model. The level of analysis refers to the society being studied. The society being studied could be the state of Massachusetts, in which the units of observation are consumption and programmatic activity aggregated to the census tract, town, or county level.

- *Appropriate and consistent use of exogenous explanatory variables* – The model must account for exogenous[25] differences between units of observation. This requires inclusion of relevant explanatory variables that are reported at a comparable level of analysis as the dependent variable, scaling variable, program variables, and other independent variables used in the model.

- *Ability to measure program activity at the most granular geographic level* – Traditional bottom-up approaches measure energy impacts at the individual account level, which captures account-level detail that is beneficial in program design, marketing, and policy making, such as the importance of certain program designs or measure offerings to overall portfolio savings, measurement of free-ridership, and spillover effects. While explanatory variables measure the relative influence of demographic factors on savings, top-down estimates based on data aggregated at too high of a geographic area (i.e., state) may lose the ability to provide meaningful estimates of variables important to all interested parties. To retain this desirable level of information, top-down models should seek to measure program activity at the most granular geographic level possible. However, this poses a particular challenge in terms of the ability to measure impacts associated with upstream programs, which may only be recorded at the utility or state level.

- *Large enough sample size* – It is important that the sample size is large enough to detect the size and variability of program effects. Smaller effects that are not consistently realized across observational units need bigger sample sizes, while larger, more homogeneous effects require only small sample sizes. For example, assume we are estimating a model that examines town-level impacts across a state. If only limited savings are realized for relatively few towns, then a larger sample size will be required to detect these impacts. However, if program changes result in larger impacts that are reported throughout many or all towns in the state, then the model can be constructed with a smaller sample.

- *Long enough time series to detect and isolate program impacts* – Related to sample size, it is important that the time series is sufficiently long to contain adequate variability in the program variable(s) that is distinguishable from exogenous factors. In the event of major change to the level and/or structure of programmatic activity, it is important that there is a sufficiently long time series following the change to be able to measure impacts. In terms of the Massachusetts programs, even if the amount of available history is extended to ten years or more, the level of program activity began sharply accelerating about four years ago. Therefore, there is only limited post-acceleration to measure changes resulting from the increase in programmatic activity. While this phenomenon may limit the ability to measure

[25] Exogenous variables account for the influence of factors that influence energy consumption but are not associated with the relationship being studied. In the case of a model measuring the impacts of energy efficiency programs on consumption, exogenous variables are factors such as the level of employment that may influence consumption, but are not necessarily tied to the level of energy efficiency programmatic activity. In contrast, the amount of expenditures on energy efficiency programs are considered “endogenous” to the relationship being studied, as they provide a measure of the amount of energy efficiency programmatic activity.

programmatic impacts in the near term, top-down analysis may become more viable the longer we keep running at the higher level of program activity.

- *Account for the lag structure of program impacts* – The time series must be sufficiently long to allow for the lag structure in program impacts, relative to the level of program activity. There is a lag between program marketing efforts and the realization and reporting of program savings. There must be sufficient data in the time series to capture programmatic activity from at least one or two periods prior to the first observation in the consumption history.

4.1.4 Econometric characteristics for successful top-down modeling

Top-down models are actually aggregate versions of time series, cross-sectional (TSCS) models that are used to conduct billing analysis of individual customer data in the evaluation of separate programs. Consequently, all the known challenges of individual customer billing analysis apply to top-down models, as well. Table 4-1 outlines the challenges of TSCS models as identified in the Uniform Methods project, and describes how they manifest in aggregate top-down models.[26] TSCS billing analysis is a regression-based approach that produces a single overall model that accommodates individual household base electricity loads, and controls for local weather and systematic changes across all households. Any change in consumption due to program-related measure installations is captured as an outcome of the program. This approach is particularly advantageous because it models changes to the whole program population, and therefore, any uncertainty in the results is not due to biases in the small sample size.

[26] Agnew, Ken and Mimi Goldberg. *Whole Building Retrofit with Consumption Data Analysis Evaluation Protocol*. The Uniform Methods Project: Methods for Determining Energy Savings for Specific Measures. Prepared by DNV GL. Prepared for National Renewable Energy Laboratory. April 2013.

Table 4-1. Limitations of TSCS Billing Analysis Applicable to Aggregate Top-down Models

Limitation of TSCS models	Description of Billing Analysis Issue	Implication for Top Down
Ability to measure no-program option	• The estimated program effect is not necessarily net. It may be somewhere between net and gross because non-participants or non-program areas are not necessarily a good proxy for what the participants or program areas would have done absent the program. • Inclusion of additional explanatory variables may help minimize this but it is never perfect. • Self-selection Bias further complicates the ability to isolate naturally occurring consumption (i.e. consumption absent any program intervention).	Within state Top Down models can measure the impact of differing levels of program activity on energy consumption, but it is unlikely that the time series will be long enough to measure no-program conditions. Using other states with limited programmatic activity will provide comparison area, but not true control groups.
Spillover across geographic regions or time periods	• Spillover from the program area (or time period) to non-program area (or time period) reduces the apparent program effect.	Spillover between geographic regions and states is still present with top down methods.
Heterogeneity within the study population	• Differences in the characteristics of populations increase the possibility of specification bias as it is challenging to account for all exogenous factors influencing savings estimates. This limits the ability to fit model to data, except for relatively homogeneous populations (i.e. residential). • Size normalization can improve models by standardizing the data series, but much variability among non-residential customers remains.	More aggregated data may decrease the ability to discern heterogeneity within a population and increases the risk of specification bias.
Weather normalization	• Weather is primary determinant of energy consumption. In order to isolate program impacts from natural changes in consumption due to weather, weather normalizations is essential. • However, normalization must to be done in a manner that makes physical sense or the results can be substantially distorted.	Weather normalization should be specific to each location. If possible, aggregate data should reflect the sum of normalized consumption across the population.

Despite the similarities between top-down and TSCS modeling, there are a number of important differences. First, if units of observation have unrelated program offerings (e.g., unit = state, municipality, or service territory), then self-selection effects are minimized because the model includes customers with the different units of observation (i.e., states or towns, or other areas that do not have access to the same set of program offerings as those in another unit). For example, in a state-level analysis, if the comparison-area states have no programmatic activity, or the time series is expanded to include observations prior to program offerings, then self-selection is minimized because the model includes true measures of the "no program" scenario for the same or comparable population(s). Second, multi-year analysis used in top-down modeling is not commonly used for individual TSCS models, which increases the amount of data included in the analysis, adding variability, and increases the potential for measuring spillover.

It is important to consider the following econometric characteristics when estimating top-down models. These include both statistical and economic considerations.

- *Use of first-difference in the dependent variable* – Because top-down models analyze changes to energy consumption for a population over time, it is important to control for variation in the data resulting from time-series-specific patterns, or autocorrelation. When autocorrelation is present in a model, the model estimation procedure will produce biased estimates of the parameters and predicted values. Using the first-difference in the dependent and independent variables provides an important step to help reduce autocorrelation. The first-difference refers to the difference between the variable in two adjacent time periods; for example, the difference in energy DE can be defined as the difference between energy consumption (E) in time period "t" and "t-1," such that: DE = Et – E(t-1).

- *Use of differences in program variable and other explanatory variables* – A differenced model should include taking the first difference in the program and other explanatory variables.

- *Account for heteroscedasticity* – Heteroscedasticity occurs when the error from a model is correlated to the level of the dependent and independent variables. When heteroscedasticity is present in a model, the model estimation procedure will produce unbiased estimates of the parameters and predicted values, but the standard errors of those estimates will be inflated. This will limit the ability to make inferences with the estimated results, because the statistical significance of the estimates will be limited. There are a number of techniques for correcting this issue. A common practice is to use a log transformation of the dependent variable. Similarly, one may use the log transformation of the independent variables. If the log transformation does not correct the problem, then a third technique is to estimate the model using the weighted least squares estimation approach.

- *Account for annual fixed effects* – There may be specific abnormalities in the data for a given year caused by different exogenous factors such as general changes to the economy in a year. The impact of these factors on consumption can be isolated by including indicator variables for each time period in the analysis.

- *Account for geographic fixed effects* – Similar to annual fixed effects, there may be a certain county-, town-, or census-tract-level variation in the data that is not necessarily related to energy efficiency programmatic activity, such as changes to the local economy resulting from local businesses closing. The impact of these factors on consumption can be isolated by including indicator variables for each level of geography in the analysis.

- *Difference in differences approach* – The difference in differences approach compares energy consumption for two groups (i.e., a treatment and control group) over two time periods (e.g., a program and non-program time period). This approach can be beneficial to the top-down modeling because it may minimize errors in the reporting of data over time and across geographies by contrasting areas of high and low programmatic activity during pre- and post-treatment time periods.

- *Allow for differences in types of programmatic activity* – Changes in the portfolio of energy efficiency activities may have a substantial influence on energy savings. For example, there may be differing levels of influence for upstream and downstream programs on savings. Similarly, the amount of savings in a community resulting from non-lighting measures may be reflective of a community that is more receptive to energy efficiency programs. Expenditures in these communities may result in greater savings.

- *Account for lag between program activity and savings* – The effects of programmatic activity in one period may not be realized immediately, but in a later period of time. The effects of energy efficiency programmatic activity are not limited to the year that the activity occurred, but are actually cumulative over time, such that program expenditures made in some prior year may impact consumption in years following those expenditures. For example, if program expenditures led to the installation of energy efficient lighting two or three years ago, impacts from those installations will be realized in the current period. This requires using multiple lagged terms for programmatic impacts. Cadmus (2012) shows that consumption in the current period may be impacted by programmatic activity of up to five years previous. Similar to the lag in program impacts, there is often variance surrounding the realized savings from newly installed measures. This may result from differences between the date that measures are actually installed and the date they are recorded in the PA tracking data, learning curves associated with properly using the new technology, and other factors that cause a delay in the realization of savings. To ensure the program impact variable in the model is isolating changes that are due to programmatic activity and not natural changes in consumption over time, it is important to include a term for the amount of consumption in previous periods. Including this term will isolate changes resulting from natural changes to year-over-year consumption from those attributable to programmatic activity. Therefore, it is important to investigate whether the current period's consumption is dependent upon the previous period's programmatic activity. This can be accomplished by including variables for the amount of programmatic activity in previous periods, such as the programmatic activity from one, two, or three periods prior.

- *Account for lag structure of the dependent variable* – It is common in time series analysis for the value of the dependent variable to be correlated to its value in the previous period. In this case, using the value of the dependent variable from the previous period as an explanatory variable can eliminate the possibility of autocorrelation in the error terms, which leads to biased estimates of the parameter coefficients.

- *Multiple measures of program influence* – Specification bias exists when key variables are omitted from a model causing other variables, the constant, or error terms to absorb the variation associated with the relevant omitted terms. It is unlikely that a single term, such as total expenditures, can capture all impacts associated with different programmatic activities. This is particularly due to uncertainty in the lag structure of each program variable and consumption. Therefore, it is important to explore the use of various measures of program influence in order to determine the combination of explanatory variables that produces the best modeling results. The explanatory variables may include factors such as ex-ante savings, incentive costs, and total program costs.

- *Account for changes in consumption resulting from building codes* – Programmatic activity may follow or overlap changes in codes and standards for buildings. The relative influence of codes must be removed from true programmatic impacts to provide an accurate accounting of program-attributable savings.

- *Energy prices* – Economic theory postulates that the demand for any good is dependent upon the price of that good and the price of the next best substitute. While energy is considered to have a relatively inelastic demand, consumption and savings attributable to energy efficiency activities should depend, in part, on the price of electricity and substitute fuels.

In addition to the desirable properties of top-down models, we will evaluate each respective study reviewed in Section 4.2 in terms of these economic characteristics.

A. How the choice of level of analysis impacts criteria for successful modeling of net energy impacts

For the discussion that follows, we have defined five “levels of analysis,” which refers to the geographic region or population being studied. In addition, we have identified the unit of observation or level at which consumption, programmatic activity, and other data are aggregated (i.e., the census tract, town, or county level). The levels of analysis and the associated units of observation are listed below:

- *National analysis (state unit of observation)* – This option constructs a national model, for which the unit of analysis is aggregated to the state level.
- *Regional analysis (state unit of observation)* – This refers to an option for constructing a model across a set of states (e.g., New England), for which the unit of analysis is aggregated to the state level.
- *Regional analysis (town or census-tract unit of observation)* – This refers to an option for constructing a model across a set of states (e.g., New England), for which the unit of analysis is data aggregated to the town or census tract level.
- *In-state analysis (town or census-tract unit of observation)* – This refers to an option for constructing a model across Massachusetts only, for which the unit of analysis is data aggregated to the town or census tract level.
- *In-state analysis (individual account-level unit of observation*) – This refers to an option for constructing a model across Massachusetts only, for which the unit of analysis is data aggregated at the individual account level.

Figure 4-1 shows the tradeoffs the evaluation team must make when choosing between the levels of analysis used to construct a set of top-down models. The figure also shows the desirable properties for models discussed in Section 4.1. The ranking reported in the table indicates the following:

- *Low* – The level of aggregation results in a low level of the desirable property being achieved.
- *Moderate* – The level of aggregation results in a moderate level of the desirable property being achieved.

- *High* – The level of aggregation results in a high level of the desirable property being achieved.

The diagram illustrates that the choice of the level of analysis will result in different strengths and weaknesses, and no one model will provide a “silver bullet” to address all the relevant concerns. Rather, running a variety of models over time is most likely to provide a more comprehensive view of net program savings that can be used in conjunction with bottom-up techniques to triangulate the net impact of programmatic activity on energy consumption.

Figure 4-1. Level of Desirable Properties for Top-down Models for Different Levels of Analysis

Desirable Property	Level of Analysis				
	National Analysis (State level aggregation)	Regional Analysis (State level aggregation)	Regional Analysis (town or census tract aggregation)	In-state Analysis (Town or census tract aggregation)	Individual Account Level Aggregation
Ability to establish the counter-factual (no-program) scenario	**High:** Use states with no program activity	**Moderate:** With long enough time series may have no programs in some states in region (i.e., New England)		**Low:** No program condition prior to start of available data series. Model can only capture different levels of programmatic activity across observational units relative to some non-zero level of activity.	
Diversity of program activity level across units of observation (time-geography combination)	**Moderate:** Differences in program activity high, but reporting inconsistent	**Moderate:** Some variation in levels of program activity. Regional program data more likely to have consistent reporting than national.		**Low:** Within state, variation in programs is low. Requires longer time series	
Large enough sample size to account for the lag structure of program impacts	**Moderate:** Publicly available data with long time series, inconsistent reporting over time		**Unknown**	**Low:** In short term (2014), only 3 years of data **High:** Long term (2015 and beyond) 10+ years of data	
Long enough time series to detect and isolate program impacts	**Moderate:** Publicly available data with long time series, inconsistent reporting over time		**Unknown**	**Low**: In short term (2014), only 3 years of data; **High:** Long term (2015 and beyond) 10+ years of data	

Desirable Property	Level of analysis				
	National Analysis (State level aggregation)	**Regional Analysis (State level aggregation)**	**Regional Analysis (town or census tract aggregation)**	**In State Analysis (Town or census tract aggregation)**	**Individual account level aggregation**
Ability to measure program activity at the most granular geographic level	**Low:** Highly aggregate data limits the ability to detect how incremental changes in program activity and other explanatory variables impact consumption. Upstream program data are available and not a limiting factor.		**Moderate:** More disaggregate data improves the ability to measure how incremental changes in program activity and other explanatory variables impact consumption over state level aggregation. Upstream program data are available and not a limiting factor.		**High:** Changes in program activity can be measured at the account level. Availability of upstream data may be limiting factor.
Consistent relationship between program activity and savings	**Low:** Changes in program activity and consumption reported at a high level of aggregation may correlate with other factors.	**Moderate:** Regional differences in reporting may exist; Aggregate data more difficult to control for exogenous factors		**Moderate:** Aggregate data more difficult to control for exogenous factors	**High:** Changes in consumption tie directly to individual consumption
Minimal effect of one area on another (cross-area spillover)	**Low:** Savings estimates account for cross-state spillover	**Moderate:** Able to control for activity in some neighboring states.		**High:** No information on program activity in other states.	
Appropriate and consistent use of *exogenous explanatory variables*	**Low:** Inconsistent reporting of variables aggregated at a high level		**High:** Variables consistently reported		

4.2 LITERATURE REVIEW

This section provides a detailed review of the existing literature on top-down modeling. The literature review was conducted to provide important background information for the PA-Muni and PA Data pilot studies. The literature review:

- Reviewed alternative approaches in terms of their ability to estimate programmatic impacts for a single state
- Identified important economic and statistical concerns for developing top-down models
- Identified key variables for isolating impacts and potential sources of data
- Explored techniques for addressing important statistical, time series, and other technical concerns for developing effective top-down models.

This section provides a review of 15 top-down research studies that were used to estimate impacts associated with energy efficiency programs employing different units of analysis for varying levels of aggregation. The studies reviewed also use a range of techniques to provide a variety of programmatic impacts that include:

- Realization rate on ex-ante savings
- Cost effectiveness of program expenditures
- Gross and net savings estimates
- Measures of market transformation
- Changes to market share of energy efficient products

Of the 15 studies reviewed, only two studies were directly relevant for fulfilling the objective of the PA-Muni and PA Data pilot studies, which was to assess the impacts associated with energy efficiency programmatic activity within a state. Among the remaining studies:

- Six studies estimate national level impacts based on data aggregated at the state level.
- Two studies provide reviews of two national-level impact studies.
- Four studies provide top-down analyses associated with specific technologies only.
- One study measures in-state changes to consumption resulting from changes to building codes only, and does not consider energy efficiency programmatic activity.

Table 4-2, Table 4-3, and Table 4-4 provide a brief overview of the top-down studies reviewed in the remainder of this section. Following these summary tables, we provide a more thorough review of these studies, separating them by the level of analysis for which the studies attempted to measure impacts.

Table 4-2. Overview of Top-down Studies Reviewed—National Level

<table>
<tr><th>Study</th><th>Summary</th><th>Pros</th><th>Cons</th></tr>
<tr><td>Demand-Side Management and Energy Efficiency in the United States (Loughran and Kulick (2004))</td><td>National time-series cross sectional model of state level energy consumption and program expenditures data. The model sought to estimate the cost effectiveness of energy efficiency programs.</td><td rowspan="2">The model estimated the cost effectiveness of energy efficiency programs accounting for the lagged impact of expenditures on savings and other fixed effects. Model properly addresses fixed effects and econometric considerations.</td><td rowspan="2">Program impacts limited to return on expenditures. Model could not measure the effectiveness of program designs and relies on highly aggregated data with reporting inconsistencies.</td></tr>
<tr><td>Demand-Side Management and Energy Efficiency Revisited (Affhammer et al. (2007))</td><td>Provided a review of Loughran and Kulick study. Re-estimated results weighting observations based on the relative size of utilities. Provided confidence intervals around parameter estimates.</td></tr>
<tr><td>How Many Kilowatts are in a Negawatt? Verifying Ex Post Estimates of Utility Conservation Impacts at the Regional Level (Rivers and Jaccard (2011))</td><td>National time-series cross sectional model of utility and province level energy consumption and program expenditures data. The model sought to estimate the cost effectiveness of energy efficiency programs.</td><td rowspan="2">The model attempted to estimate the cost effectiveness of energy efficiency programs accounting for the lagged impact of expenditures</td><td rowspan="2">Voilette demonstrate that applying Loughran and Kulick's model to Rivers and Jaccard's data results in savings that are sufficiently high to justify expenditures. Illustrate the importance of accounting for the lag in program activity and fixed effects.</td></tr>
<tr><td>Review of a Top-Down Evaluation Study: Rivers and Jaccard 2011) (Violette (2012))</td><td>Provided a review of Rivers and Jaccard study. Applied Rivers and Jaccard data to Loughran and Kulick's model.</td></tr>
<tr><td>Electricity Intensity in the Commercial Sector: Market and Public Program Effects. (Horowitz (2004))</td><td>Estimated a national time-series fixed effects model using state level energy consumption data. The attempted to estimate the effects of energy programs that directly target customers from up-stream (market transformation) programs.</td><td>Model demonstrates the importance of considering different types of programmatic activity on savings. Model estimated using data from 42 states of 12 years of varying programmatic activity.</td><td>Measure of market transformation derived using data from a variety of loosely connected sources, leads to questionable interpretation of results.</td></tr>
<tr><td>Changes in Electricity Demand in the United States from the 1970s to 2003 (Horowitz (2007))</td><td>Study uses a difference of differences approach to construct a national model that contrasts pre- and post-program consumption for states with strong-to-moderate programmatic activity to states with weak programmatic activity.</td><td>Provides an approach for developing the counterfactual conditions and estimating net savings.</td><td>Difference of differences approach requires many assumptions regarding the selection of treatment and control states as well as pre- and post-periods.</td></tr>
<tr><td>Measuring the savings from energy efficiency policies: a step beyond program evaluation (Horowitz (2010))</td><td>Demonstrates that top-down models can be developed at different levels of analysis to provide estimates of programmatic impacts based on data aggregated at the account, utility, and state levels.</td><td>Illustrates ability of top-down methods to be applied to different levels of analysis using data aggregated at different levels.</td><td>Reduction in energy intensity assumed to result from corresponding increases in energy efficiency activity without direct causality being established.</td></tr>
</table>

Table 4-3. Overview of Top-down Studies Reviewed—Regional and State Level

Study	Summary	Pros	Cons
How Many Kilowatts are in a Negawatt? Verifying Ex Post Estimates of Utility Conservation Impacts at the Regional Level (Parfomak and Lave (1996))	Uses utility level consumption and ex-ante savings to estimate the realization rate on savings across utilities in New England and California.	Provides a realization rate on ex ante savings.	Model does not account for many factors that may also result in reductions to energy consumption over time.
CPUC Macro Consumption Metric Pilot Study (Final Report) (Cadmus (2012))	Used energy efficiency expenditures and a series of explanatory variables to predict changes to energy use for commercial and residential for a utility service territories in California.	The model used an extended time series, 1990 – 2010. While this may not provide for a true "No Program" baseline, the level of activity in the early 1990's should be sufficiently different to provide a meaningful point of comparison.	The model does not distinguish between types of programmatic activity. The model was not able to produce statistically significant results.
Macro Consumption Metrics Pilot Study Technical Memorandum – Preliminary Findings (Demand Research (2012))	This study uses a two-way fixed effects model that aggregates consumption and economic variables to either the census tract level for residential customers or industry by county for nonresidential customers. Annual consumption per location is set equal to a set of time series variables that reflect the ratio of ex-ante savings to consumption, the ratio of measure costs to fuel expenditures, and incentive costs to fuel expenditures.	This study is one of two existing studies that focus specifically on measuring programmatic net impacts from utility sponsored programs within a single state. This study includes multiple measures of programmatic activity including ex-ante savings, incentive and measure costs. The model uses weather normalized consumption as the dependent variable which is the same as the PA data model being developed through the current study	The model limits impacts to in-state that occur over a 5-year time series.
Are Building Codes Effective at Saving Energy? Evidence From Residential Billing Data in Florida (Jocobsen and Kotchen (2009))	Uses account level utility data to estimates a pooled time-series cross-sectional model that is used to construct a difference of differences comparison of the effect of building codes on energy consumption.	Demonstrate the importance of building codes on reductions in energy consumption.	Model does not consider the effect of energy efficiency programs on consumption. Scope of model is limited to the utility service territory.

Table 4-4. Overview of Top-down Studies Reviewed—Technology Specific Studies

Study	Summary	Pros	Cons
The Impact of Regional Incentive and Promotion Programs on the Market Share of ENERGY STAR® Appliances (Rosenberg (2003))	Estimated multi-state linear regression models to predict the impact of incentive programs and regional demographic variables on market shares for separate ENERGY STAR® appliances	Models demonstrate the ability to employ a variety of data sources and statistical techniques to estimate programmatic impacts.	Models provide measure specific results only
Modelling the Effects of U.S. ENERGY START Appliance Programs. (Feldman et al. (2005))	Used ANOVA and linear regression analysis to first estimate the market penetration of ENERGY STAR appliances by state as a function of the presence of program activity and then used the change in market shares over time to predict cumulative effects of ENERGY STAR programs		
Results of the Multistate CFL Modeling Effort (NMR Group, Inc. (2011))	Used CFL saturations from survey data along with energy efficiency program budget information, number CFLs receiving incentives and program types to predict CFL purchases over multiple years		
Economic Indicators of Market Transformation: Energy Efficient Lighting and EPA's Green Lights (Horowitz (2001))	Used data from the Census' "Manufacturing and Construction database" from 1959 – 2000 to construct a model that estimates the market share for energy efficient lighting based on product price, the price of electricity and a vector of macroeconomic variables		

We divide the studies reviewed in this section as follows:

- Section 4.2.1 reviews studies that measure national-level impacts.
- Section 4.2.2 reviews studies that measure regional-level impacts.
- Section 4.2.3 reviews studies that measure state-level impacts.

4.2.1 Studies that measure national-level impacts

The majority of existing top-down studies attempt to measure impacts resulting from energy efficiency programs at the national level. One benefit of the studies that are conducted at the national level is they clarify the amount by which free-ridership impacts savings estimates by increasing the amount of variation in programmatic activity. These studies use states with low levels of programmatic activity to provide a measure of the amount of naturally occurring adoption of energy efficiency measures. Further, because the models capture changes to consumption and programmatic activity across the entire nation, if properly specified, they will account for spillover and rebound effects. However, aggregating data to the state or regional level limits the ability to obtain estimates of these net-impact factors, as well as estimates at the specific utility, program, or measure level.

In this section, we review seven studies that were conducted at the national level. We review each of the following national-level studies in terms of study design, outcomes, and the ability to address the criteria for successful modeling of net energy impacts.

A. Demand-Side Management and Energy Efficiency in the United States (Loughran, 2004) and Demand-Side Management and Energy Efficiency Revisited (Auffhammer, University of California, Blumstein, & Fowelie, 2007)

Loughran and Kulick is one of the most widely cited top-down studies at the national level. The study includes many of the criteria for successful top-down modeling of programmatic impacts, as well as the econometric characteristics discussed earlier. The study uses annual consumption data across 324 utilities in the United States from 1989 to 1999. Loughran and Kulick's model estimates the impact of year-over-year changes in utility energy efficiency expenditures on year-over-year changes to energy consumption per unit of gross state product (GSP). The unit of analysis for this study is the utility territory within a state. The study uses a number of variables and variable transformations to account for autocorrelation time series and fixed effect considerations including:

- Uses first difference of the natural log of energy consumption as the dependent variable
- Includes utility, state, and annual level fixed effects
- Program impacts enter the model as three terms for each utility, state, and time period:
  - The first difference in the natural log of the current time period's total expenditures
  - The first difference in the 12-month lag of total expenditures
  - The first difference in the 24-month lag of total expenditures.

In addition, the model includes variables for year-over-year change in GSP, fuel prices, heating degree days and cooling degree days, and the number of customers. Finally, the model identifies the percent of total annual consumption associated with the residential, commercial, and industrial sectors.

The model estimates the return on energy efficiency expenditures over a period in which there is a large increase and then decrease in energy efficiency expenditures. The time series also captures a period in which 60% of the utilities contained in the study reported zero energy efficiency expenditures to serve as counterfactual conditions. The authors found that energy efficiency expenditures do lower consumption, but much less than reported by utilities. However, Auffhammer et al. show that energy savings provided by Loughran and Kulick (2004) are artificially low and estimates of costs are artificially high because they do not weight expenditures by the relative size of utility. Auffhammer et al. also use a non-parametric bootstrap[27] method to construct confidence intervals around savings and cost estimates, and found that the utility-reported consumption reduction cannot be rejected based on the findings of the revised model results with 95% certainty.

Loughran and Kulick make the following contributions to the Year 1 pilot studies:

- *Time series adjustments* – The model demonstrates the importance of using a variety of techniques to isolate true programmatic impacts. Those used in this model include the first difference of the dependent variable; annual, utility, and state level fixed effects; and the 12- and 24-month lag of program expenditures. The lag in program expenditures is used to account for the cumulative effect of programmatic activity over time as discussed in Section 4.1.4.

- *Accounting for lag effects of program activity* – While also a time series issue, it is important to discuss this issue separately. The result of not including lagged programmatic activity is that the program impacts will be restricted to those attributed to the current period's programmatic activity, thereby potentially underrepresenting the true extent of program impacts. The lag in program expenditures is used to account for the cumulative effect of programmatic activity over time as discussed in Section 4.1.4.

- *Appropriate weighting of energy data and modeling error* – In their review of the Loughran and Kulick study, Auffhammer et al. demonstrated the importance of considering the relative size of observational units to the overall contribution of parameter estimates. In relation to the present study, this finding demonstrates that differences between small and large PAs are likely to result in overall results that differ from results at the individual PA level. Care should be taken when interpreting results to isolate these differences.

- *Use of a range of explanatory variables* – The study shows the importance of isolating effects resulting from changes to a range of exogenous variables and customer segments. The authors found significant changes to energy consumption

[27] This technique accounts for predicted savings not being independent within utilities. To implement this technique, Auffhammer et al. used a sample of observations to re-estimate multiple iterations of the model using Monte Carlo simulation.

resulting from variations in aggregate productivity, fuel prices, and weather. Savings estimates also differed according to the number of customers and customer segments.

B. How Many Kilowatts are in a Negawatt? Verifying Ex Post Estimates of Utility Conservation Impacts at the Regional Level (Rivers and Jaccard) and Review of a Top-down Evaluation Study: Rivers and Jaccard 2011 (Violette, 2012)

Rivers and Jaccard examine the cost-effectiveness of expenditures for energy efficiency programs using a national model of energy consumption in Canada. The authors use annual utility level consumption in Canada from 1990 to 2005 to estimate a first difference linear regression model. The model estimates the impact of year-over-year changes in province-level expenditures for energy efficiency programs on year-over-year changes to province-level energy consumption per customer. In addition, the model includes variables for gross domestic product (GDP), fuel prices, and heating degree days and cooling degree days. Finally, the model includes a variable for the percent of consumption in the residential sector, but does not distinguish between commercial and industrial sector consumption.

The model makes some adjustments to account for time series related concerns, such as using the first difference in energy consumption and program expenditures. However, the model does not consider the effect that the previous period's program expenditures have on current period consumption. Instead, the model includes a term for the lag in the dependent variable, the previous period's difference in consumption. Finally, the model does not consider terms that capture annual fixed effects.

Violette reviews Rivers and Jaccard's analysis and demonstrates that the results were impacted by a number of methodological shortcomings, including those found in the Auffhammer et al. review of the Loughran and Kulick study. Violette demonstrates that Rivers and Jaccard's approach suffers from a number of methodological limitations that prevent it from producing savings estimates sufficient to justify program costs:

- The model does not account for annual fixed effects.
- While the model includes the 12-month lag in the dependent variable, it does not include the lag in program expenditures, which prevents detection of inter-temporal effects of program expenditures, thereby assuming all impacts occur within the same period in which they are reported.

After reviewing the Rivers and Jaccard study, Violette uses data developed by Rivers and Jaccard to estimate the model developed by Loughran and Kulick, which does account for fixed effects and the lag effect of program expenditures. The results demonstrate the importance of time dependent fixed effects and use of the 12-month lag in program expenditures. The true impact of program spending in Canada (included in Rivers and Jaccard's analysis) was that savings estimates were sufficiently high to justify energy efficiency expenditures based on costs.

The contributions of this study to the Year 1 pilot studies include:

- *Time series adjustments* – By re-estimating Loughran and Kulick's model using Rivers and Jaccard's data, Violette demonstrates the importance of accounting for time series adjustments, particularly annual fixed effects and the lagged effect of programmatic activity on consumption.

- *Lagged programmatic activity instead of lagged dependent variable* – The Rivers and Jaccard model attempts to account for autocorrelation of the dependent variable by only including a term for consumption lagged one period. This is a useful technique if there is reason to believe that the difference in consumption in one period influences the consumption in the next period that is not explained by the other variables in the model. Violette shows that it is important to account for changes in consumption influenced more by energy efficiency programmatic activity of previous periods and time-specific fixed effects.

C. Horowitz (2004, 2007, 2010)

Horowitz provides three different approaches to national level top-down models through three separate studies (Horowitz 2004, 2007, and 2010). These studies use a variety of techniques to demonstrate impacts resulting from energy efficiency programs at the national level.

i. Electricity Intensity in the Commercial Sector: Market and Public Program Effects (Horowitz, 2004)

In the first study, Horowitz (2004) uses a weighted least squares linear regression model to estimate the change in energy intensity (natural log of annual kWh per unit of GSP) for energy efficiency programmatic activity. The unit of analysis in this model is the state-level commercial energy consumption, for 42 states within the US. Horowitz distinguished between programmatic activities resulting from the following two types of programs:

- *Energy efficiency programs* – Energy efficiency activity as average energy efficiency savings resulting from audits, technical assistance, and financial incentives. To measure the impacts associated with these programs, Horowitz (2004) uses aggregate energy efficiency savings from utilities aggregated to the state level, as provided by EIA form 861.

- *Market transformation programs* – Horowitz (2004) defines market transformation programs as those that target both the supply and demand side through public information, marketing, and education. To measure impacts associated with these programs, Horowitz (2004) constructs a market transformation index based on census data regarding product shipments, survey data identifying rebated sales, and a model developed by Horowitz (2001) (Horowitz, Economic Indicators of Market Transformation: Energy Efficient Lighting and EPA's Green Lights, 2001), to allocate a fraction of lighting sales to market transformation programs.

In addition to program impacts, the model includes variables that account for one period lag in the energy consumption index, and electric and natural gas prices. The percent of a state's generation is derived from natural gas and non-conventional sources, heating and cooling degree days, and a time trend for technology-based business based on Federal Reserve data. These data were intended to isolate impacts associated with increased uses of information technology, which includes an index of annual construction plus state and annual fixed effects indicator variables. While the study did consider the impact of the previous period's consumption on the current period, it did not account for the lagged impact of energy efficiency expenditures on consumption. Finally, the model was not estimated using the first difference in consumption and explanatory variables.

The study results showed average savings attributable to energy efficiency programs of 1.9%, which suggest a realization rate of 54% on energy efficiency programs. However, the study found an additional 5.8% savings attributable to market transformation programs.

Strengths and weaknesses of this study relative to the Year 1 pilot studies include:

- *Accounts for differing types of program influence on consumption* – This study illustrates that it is important to consider different measures of programmatic influence and the variety of different types of efficiency programs. For the present pilot studies, this finding highlights the importance of considering both downstream and upstream program impacts. Where possible, the current pilot studies incorporated measures that allow for these impacts to be measured separately.

- *Accounts for changes in energy intensities* – The model measures the total changes in energy intensity as a function of changes in programmatic activity and other variables, rather than considering the first difference in consumption or programmatic activity and other variables. Further, the model does not account for the lagged impact of energy efficiency expenditures on consumption.

- *Perfect correlation of states with no programmatic activity* – The model does not address issues concerning perfect correlation between fixed effects and states with no programmatic activity.

ii. Changes in Electricity Demand in the United States from the 1970s to 2003 (Horowitz, 2007)

This study provides an alternative approach to examining programmatic impacts and market transformation associated with energy efficiency programs in the US. This is based on a difference in differences of states. The approach compares modeled consumption for states with strong programmatic activity to states with weak programmatic activity in two periods: a base period with little to no energy efficiency programs, and a treatment period with more advanced programs. This paper first uses EIA-861 to obtain sector-level energy consumption by states from 1979 to 2003, and splits the states into two groups:

- *Strong to moderate program state* – based on EIA's quartile rankings of 36 states with strong to moderate programmatic activity (quartiles 1–3) as determined by reported savings estimates.

- *Weak program states* – based on EIA's quartile rankings of 12 states with weak programmatic activity (quartile 4) as determined by reported savings estimate.

By using EIA rankings to identify states with strong to moderate (SM) and weak (W) levels of programmatic activity, this approach avoids relying on EIA's reported energy savings or expenditure data directly, thereby avoiding issues surrounding inconsistencies in data reporting across units of observation. Once states with strong and weak programmatic activity are identified, state-level data are separated into pre-program and post-program periods. However, because there is no definitive "year" in which all energy efficiency programs began, the study considers all data prior to 1992 to be pre-program, as 1992 marks a point in time when many states began accelerating energy efficiency programs.

Using this framework, Horowitz develops separate residential, commercial, and industrial regression models for pre- and post-periods. The models estimate changes in sector-level energy intensity (annual consumption/GSP for commercial and industrial sectors and annual consumption/per capita for the residential sector as a function of whether a state is designated as having strong to moderate (SM) or weak programmatic activity (W), which enters the models as an indicator variable (where 1=SM; 0=W)). Separate models are constructed for the pre- and post-program periods that include explanatory variables for energy prices (electricity, natural gas), per capita income/GSP, technology trend from Federal Reserve data on energy using equipment, and weather impacts (heating degree days and cooling degree days). Comparing the regression coefficients from pre- and post-period models demonstrates the effect of a state having high or low programmatic activity. This allows models to fit different underlying structures to two contiguous periods. Comparing the parameter estimates would give a good indication if the whole construct makes sense or not. Further, plugging the average values for the SM states into the W state's model provides an estimate of the counterfactual condition, or the amount of naturally occurring consumption. The difference in differences approach measures the difference between treatment and base period energy intensity for SM states assuming the factual conditions, less the difference between treatment pre- and post-period energy intensity for SM states assuming the counterfactual conditions, where the counterfactuals are calculated by imputing the mean values for the SM states into the W state's equations.

Strengths and weaknesses of this study relative to the Year 1 pilot studies include:

- *Viable approach for measuring relative magnitudes of impacts* - This method provides a viable approach for understanding the relative magnitude of program impacts, including spillover.
- *Selection bias* – The assignment of states to weak and strong-to-moderate groups creates considerable potential for selection bias. The choice of pre- and post-periods is also a potential source for specification bias.
- *No-levels of programmatic activity* – The model treats the pre-program year as an absolute point in time when programs began. This assumption is incorrect as some programmatic activity was present in the 1980s and earlier. Further, the model cannot allocate the acceleration in the adoption of energy efficiency programs to specific years. Adoption of energy efficiency programs increased gradually after 1992, where some states may have had no programs, and others had more advanced programs. While the study does possess these limitations, the ability to apply this research design to investigate programmatic impacts is a major contribution of this study.

iii. Measuring the Savings from Energy Efficiency Policies: A Step beyond Program Evaluation (Horowitz, 2010)

Horowitz (2010) presents multiple approaches for employing top-down modeling to estimate impacts associated with energy efficiency programs based on different levels of analysis. For one of these methods, Horowitz (2010) uses two groups of states to project national level program impacts, similar to Horowitz (2007). Using this technique, Horowitz (2010) uses 1970–1991 pre-program data to develop a model that explains residential energy intensity (annual consumption per customer by state) as a function of heating and cooling degree days, energy prices, equipment stock, and per capita income. Horowitz (2010) then estimates a model for states with high programmatic activity in the pre-program period, and forecasts the level of consumption in the program period. The difference between pre- and post-consumption estimates reflects gross savings. The author then applies the program states' models to non-program states' data to provide counterfactual conditions. Removing the forecasted counterfactual savings (naturally occurring) from the gross savings provides an estimate of net savings. Horowitz also demonstrates how top-down models can be developed using a single state, multiple utilities, and multiple states.

The strengths and weaknesses of this study relative to the Year 1 pilot studies include:

- *Use of top-down studies to measure impacts for different levels of analysis* – A key contribution of this study is that it demonstrates how it is possible to conduct top-down analyses to measure impacts at different levels of analysis. The study shows impacts can be measured at the national, state, and regional levels using a variety of techniques.
- *Segmenting study area to develop counterfactual conditions* – This study demonstrates that splitting study populations into areas of high and low programmatic activity can help simulate counterfactual conditions. This approach is comparable to the PA-Muni model pilot study presented in this report.
- *Identifies relative strength of programmatic activity* – The method relies on the choice of strong and weak programmatic states as measures of program impacts between specific points in time. This poses substantial concerns regarding the choice of states and time period. Further, the model does not control for major differences in economic conditions of the different locations or between the two time periods, thereby assigning all changes to consumption to differences in programmatic activity.

4.2.2 Studies that measure regional-level impacts

A. How Many Kilowatts are in a Negawatt? Verifying Ex Post Estimates of Utility Conservation Impacts at the Regional Level (Parfomak & Lave, 1996)

Parfomak and Lave construct a time series cross sectional model to compute the realization rate on ex-ante energy savings across 39 utilities in the 10 states in the Northeast and California. The authors develop a first difference model that estimates the change in utility level consumption from 1970 to 1993 as a function of the difference in ex-ante energy savings, fuel prices, manufacturing and non-manufacturing employment, installed technology, and utility-level fixed effects. The authors found that the weighted least squares model estimates provided for a 99% realization rate on ex-ante savings estimates.

This model faces a number of methodological concerns that were noted in the development of the PA-Muni and PA Data pilot studies. First, the model measures changes to electric consumption rather than electric intensity (or consumption per measure of output). When considering changes to consumption over such an extended period of time and across geographies, it is important to measure changes to consumption relative to measures of production. The model also lacks key explanatory variables to account for technological changes over the extended time period. Specifically, the model does not attempt to account for the expansion of information technology between 1970 and 1993, which is likely to account for changes to the electric intensity in the economy. The model includes utility-level fixed effects, but not state or annual fixed effects. Finally, the model does not include terms to address year-over-year differences in the measurement and reporting of energy savings.

Strengths and weaknesses of this study relative to the Year 1 pilot studies include:

- *Use of ex-ante energy savings* – In contrast to most other top-down studies reviewed, this study attempts to estimate a realization rate on savings directly by using utility reported ex-ante savings as the key explanatory variable.
- *The model excludes a number of key explanatory variables that are likely responsible for changes in consumption over time.* It also does not include year-over-year differences in the program variables, annual fixed effects, or other explanatory variables.
- *Perfect correlation of states with no programmatic activity* – The model does not address issues concerning perfect correlation between fixed effects and states with no programmatic activity.

4.2.3 Studies that measure state-level impacts

A. Measuring the Savings from Energy Efficiency Policies: A Step beyond Program Evaluation (Horowitz, 2010)

Horowitz (2010) not only looked a national analysis, but also considered a number of other levels of analysis, including measuring state-level impacts. In the state-level analysis, Horowitz contrasts use of annual and monthly consumption estimates of savings associated with energy efficiency programs for four states separately. Similar to the national model discussed earlier, data from the “pre-program” period is used to construct the model of energy consumption assuming no programmatic activity. The pre-program period is defined as prior to 1992. Data for this exercise are derived from EIA’s State Energy Data System (SEDS) database, and cover 1970 to 2009. Explanatory variables include fuel prices, real personal income, heating and cooling degree days, and equipment stock.

Horowitz estimates similar models using annual and monthly consumption history. The general finding of this approach demonstrates that there are differences in consumption detected at the monthly level that are not isolated at the annual level. One could argue these savings are attributable to monthly fluctuations in temperature that cannot be detected from an annualized model. One possible means of addressing this concern is to normalize consumption at a disaggregate level and then construct models using the normalized series. However, the differences may simply reflect variances in measurement at the monthly level that result in alternate estimates of savings.

Strengths and weaknesses of this study relative to the Year 1 pilot studies include:

- *Importance of temperature changes* – This study demonstrates the importance of adjusting for monthly fluctuations in temperature.
- *Use of weather normalization* – While this can be accomplished by constructing models at the monthly level, many data employed by top-down models are not available at the monthly level. Weather normalization of the dependent variable

reflects an alternative approach to accounting for changes in monthly weather conditions.

B. Macro Consumption Metrics Pilot study Technical Memorandum – Preliminary Findings (Demand Research, 2012)

This study uses a two-way fixed effects model that aggregates consumption and economic variables to either the census tract level for residential customers, or the industry by county level for nonresidential customers. Annual consumption per location is set equal to a set of time series variables that reflect the ratio of ex-ante savings to consumption, the ratio of measure costs to fuel expenditures, and incentive costs to fuel expenditures. In addition, the residential model incorporates explanatory variables that adjust for factors such as weather, energy prices, average income, household size, median age in households, level of education, and housing type. The commercial/industrial sector model adjusts for factors such as industry, utility, revenue, energy prices, and employment.

The authors argue that the coefficients on ex-ante savings and incentive costs can be used as policy impact indicators. This approach estimates the cumulative impact of program effects over time. The model does not distinguish among individual programs; consequently, program impacts are estimated in aggregate. This approach reduces errors associated with aggregating impacts across multiple programs. Because the model is estimated using a time series, indicators for changes to codes and standards can be used to isolate their impacts. This method also allows for the use of trend variables to capture the cumulative effect of programs over time.

Strengths and weaknesses of this study relative to the Year 1 pilot studies include:

- *Single state analysis* – This study is one of two existing studies that focus specifically on measuring programmatic net impacts from utility sponsored programs within a single state.
- *Choice of dependent variable* – The model uses weather normalized consumption as the dependent variable, which is the same as the PA Data model being presented in this report.
- *Variety of measures of programmatic activity* – This study includes multiple measures of programmatic activity, including ex-ante savings, and incentive and measure costs.
- *Limited time span* – The model limits impacts to in-state impacts that occur over a five-year time series.

C. CPUC Macro Consumption Metric Pilot study (Final Report) (Cadmus Group, 2012)

This study uses energy efficiency expenditures and a series of explanatory variables to predict changes to energy use per capita or housing unit (residential) and energy use per square foot (commercial) for a utility service territory. The authors also use a series of variables that identified the amount of new construction in a service territory occurring after various buildings codes were instituted to isolate the effects of building codes.

Based on this approach, the authors argue that the coefficient on energy efficiency expenditures can be used to estimate the energy efficiency program impact on savings.

Further, the coefficient on the building code variables can be used to estimate the amount of savings attributable to codes and standards. The model resulted in statistically significant impacts for building codes and energy efficiency in commercial and industrial sectors, but not for residential programs. This method is simpler than the approach used by Demand Research, but has only produced statistically significant results for nonresidential programs.

Strengths and weaknesses of this study relative to the Year 1 pilot studies include:

- *Single state analysis* – Similar to the Demand Research LLC study, the model only considered in-state (California) data.
- *Length of time series* – The model uses an extended time series, 1990–2010. While this may not provide for a true “no program” baseline, the level of activity in the early 1990s should be sufficiently different to provide a meaningful point of comparison.
- The model does not distinguish between types of programmatic activity.
- The model is not able to produce statistically significant results for residential programs.

D. Are Building Codes Effective at Saving Energy? Evidence from Residential Billing Data in Florida (Jacobsen & Kotchen, 2009)

This study is an example of the use of pooled time series cross-sectional (TSCS) billing analysis to construct a model from account-level billing data that examines the impact of residential construction code changes on aggregate consumption for a region. The approach uses monthly account-level billing data to estimate changes to household consumption as a function of heating and cooling degree days, household square footage, number of bathrooms and bedrooms, ZIP code, central air conditioning (CAC) indicator, and an indicator for whether the corresponding observation month occurs before or after code changes in that area. The model also interacts with the code change variable with heating and cooling degree days to account for the impact of code changes on temperature sensitive loads. The authors use the model results to conduct a difference in differences analysis, similar to Horowitz (2007), to show that code changes correspond to a 4% decrease in electric consumption and a 6% decrease in gas consumption.

While this method applies to only a limited geographic area—as it does not employ aggregate consumption or macroeconomic data, as do the other top-down models reviewed—it demonstrates the similarities between top-down modeling and traditional time series cross-sectional billing analysis, as was discussed in the Abbreviated Methods Review (DNV GL, Abbreviated Review of Methods for the Draft Top-down Modeling Methods Study, 2014). The model also demonstrates the relative importance of incorporating indicators of codes changes to explain variations in consumption over time. However, there is no attempt to address other programmatic influences on energy consumption over time, such as utility sponsored efficiency programs or changes to federal standards.

Strengths and weaknesses of this study relative to the Year 1 pilot studies include:

- Demonstrates the importance of considering building codes.
- Demonstrates the similarities between top-down and TSCS models.

- Since this is a limited geographical area, code change variation across subpopulations is presumably small, meaning this variable would tend to pick up any general downward trend not picked up elsewhere. The authors assume the time fixed effect would capture all economic trends.

4.2.4 Technology-specific national studies

In addition, the following studies use top-down models to estimate market shares and programmatic impacts for specific energy efficient technologies. These studies were some of the earlier top-down modeling efforts. They demonstrate the ability to use data aggregated at the state and regional levels to estimate programmatic impacts associated with energy efficient programs, specifically lighting. We mention these studies briefly because of their importance in the history of top-down modeling. The studies introduce the use of top-down models to measure specific programmatic impacts. However, these studies only isolate impacts associated with lighting measures, and therefore cannot be used to assess programmatic impacts across the portfolio of measures. Therefore, we only mention these studies briefly and do not provide more detailed descriptions of the approaches used by each.

- *The Impact of Regional Incentive and Promotion Programs on the Market Share of ENERGY STAR® Appliances (Rosenberg, 2003)* – Estimates multi-state linear regression models to predict the impact of incentive programs and regional demographic variables on market shares for separate ENERGY STAR appliances.
- *Modeling the Effects of U.S. ENERGY STAR Appliance Programs (Feldman S. L.-W., 2005)* – Uses ANOVA and linear regression analysis to first estimate the market penetration of ENERGY STAR appliances by state as a function of the presence of program activity, and then uses the change in market shares over time to predict cumulative effects of ENERGY STAR programs.
- *Results of the Multistate CFL Modeling Effort (NMR Group, 2011*) – Uses CFL saturations from survey data along with energy efficiency program budget information, number of CFLs receiving incentives, and program types to predict CFL purchases over multiple years.
- *Economic Indicators of Market Transformation: Energy Efficient Lighting and EPA's Green Lights (Demand Research, 2012)* – Uses data from the Census' "Manufacturing and Construction database" from 1959 to 2000 to construct a model that estimates the market share for energy efficient lighting based on product price, the price of electricity, and a vector of macroeconomic variables.

4.2.5 Comparison of studies

Table 4-5 below illustrates how each of the studies reviewed above ranked relative to the desirable properties of top-down models. The table lists each of the modeling attributes and ranks each study on a scale of 1 to 10, with 1 being the least favorable ranking and 10 the most favorable. The table also indicates whether a national- or individual-level analysis would be more appropriate for each attribute.

The ranking of studies according to the properties for successful top-down models was based on both objective and subjective judgment. For the objective portion, we scored each study based on whether the study included the important characteristics listed below. If the study included the factor, it received a score of one for that factor, otherwise a zero. This provided a rough estimate of the approximate ranking of the study. Then, we used subjective analysis to place the study on the scale of 1 to 10 based on how it compared to studies that had similar rankings. In conducting this assessment, we asked the following questions:

Ability to establish the counterfactual

- Is a comparison area used (i.e., an area with no programmatic activity)?
- Is the time series long enough to have variability in the level of programmatic activity?
- Are there areas with no programmatic activity during some of the time periods?
- Does the study use difference in differences?

Diversity of programmatic activity across units of observation

- Are multiple utility territories considered?
- Are multiple regions considered?
- Is the time series sufficiently long to capture varying levels of activity?
- Are there units with no programmatic activity for some time periods?

Consistent relationship between program activity and savings

- Do the data used allow researchers to confirm energy consumption and program activity measured for the same level of aggregation?
- Do the data used allow researchers to confirm energy consumption and program activity measured for the same time series?
- Do the data used allow researchers to confirm program activities are consistent across geographic units?
- Do the data used allow researchers to confirm data reported for the same time interval across geographic units?

Ability to measure program activity at the most granular geographic level

- Can program activity and consumption be measured at the account level?
- Can program activity and consumption be measured at the census tract or town level?
- Can program activity and consumption be measured at the utility level?
- Can program activity and consumption be measured at the state level?

Minimal effect of one area on another (cross-area spillover)

- Does the level of aggregation account for spillover across towns within a state?
- Does the level of aggregation account for spillover across counties within a state?
- Does the level of aggregation account for spillover across states within a region?

- Does the level of aggregation account for spillover across states nationally?

Appropriate and consistent use of exogenous explanatory variables

- Does the model include energy prices?
- Does the model account for Gross Domestic Product?
- Does the model include changes to energy-using technology over time?
- Does the model consider segmentation of impacts by sector?

Large enough sample size to account for the lag structure of program impacts

- Did the model include the one-period lag of program variables?
- Did the model include the two-period lag of program variables?
- Did the model include multiple-periods lags of the program variable?
- Did the model include the lag of the dependent variable as an explanatory variable?

Long enough time series to detect and isolate program impacts

- Does the model include at least 20 years of data?
- Does the model include at least ten years of data?
- Does the model include at least five years of data?
- Is program activity minimal or approaching zero for some observations during some years of the study?

Table 4-5. Comparison of Studies Relative to Top-down Modeling Attributes

Ranking (1 worst; 10 best)									
1	2	3	4	5	6	7	8	9	10
Ability to establish the counter-factual (no-program) scenario									
Cadmus; Demand Research							Rivers and Jaccard; Loughran and Kulick; Horowitz (2004); Parfomak and Lave); Jocobsen and Kotchen		Horowitz 2007, 2010;
Diversity of program activity level across units of observation (time-geography combination)									
		Cadmus; Demand Research		Jocobsen and Kotchen			Horowitz 2004, 2007, 2010; Rivers and Jaccard; Loughran and Kulick; Parfomak and Lave		
Consistent relationship between program activity and savings									
				Horowitz 2007					Jocobsen and Kotchen;
Ability to measure program activity at the most granular geographic level									
Horowitz 2004, 2010; Rivers and Jaccard; Loughran and Kulick; Parfomak and Lave				Horowitz 2007					Jocobsen and Kotchen; Cadmus; Demand Research
Minimal effect of one area on another (cross-area spillover)									
		Cadmus; Demand Research		Parfomak and Lave; Loughran and Kulick; Rivers and Jaccard; Horowitz (2004);					Horowitz 2007, 2010
Appropriate and consistent use of exogenous explanatory variables									
		Demand Research; Parfomak and Lave; Rivers and Jaccard; Jocobsen and Kotchen		Horowitz 2010; Loughran and Kulick			Horowitz 2004, 2007; Cadmus		
Large enough sample size to account for the lag structure of program impacts									
Jocobsen and Kotchen; Horowitz 2007, 2010; Parfomak and Lave		Horowitz 2004; Rivers and Jaccard		Horowitz 2004	Demand Research	Demand Research	Loughran and Kulick; Cadmus		
Long enough time series to detect and isolate program impacts									
		Demand Research; Cadmus		Jocobsen and Kotchen			Rivers and Jaccard; Loughran and Kulick		Parfomak and Lave; Horowitz 2004, 2007, 2010
Overall Rating (1-10)									
			Parfomak and Lave; Demand Research	Horowitz (2004); Jocobsen and Kotchen; Cadmus; Rivers and Jaccard; Loughran and Kulick;	Horowitz 2010	Horowitz 2007			

The following summarizes the information presented in Table 4-5.

A. Ability to establish the counterfactual (no-program) scenario

The models described in the Horowitz 2007 and 2010 studies are the most effective at establishing the counterfactual conditions based on the criteria discussed earlier. Horowitz

(2007) estimates differences in consumption between weak and strong program states for the pre- and post-periods of advanced programmatic activity, which allows for inferences concerning what would have happened if there were no programs in the program states. Furthermore, this study includes information from 36 states over a 33-year period. Horowitz (2010) used a similar approach to his earlier study with a shorter time series. The same approach is used by Jacobsen and Kotchen; however, that study does not incorporate other programmatic activities that may influence consumption, such as changes to the local economy over time. Further, it only measures changes to consumption relative to a uniform code change across the state, so differences in the level of code compliance are not measured. Therefore, the model does not account for instances in which code compliance may be less than 100%.

The next group of studies all use a linear regression based approach to isolate savings in the absence of a counterfactual condition. Loughran and Kulick rely on state-level differences in expenditures over a 10-year period where the majority of states in the sample had zero energy efficiency expenditures for at least one year, providing counterfactual conditions. Similarly, Horowitz's (2004) use of 12 years of data across 42 states provides for greater variation in the level of programmatic activity, thereby allowing the model to project consumption based on minimal programmatic activity (i.e., the counterfactual). An important strength of these analyses is that the models look at differences in two types of efficiency programs over time. Similarly, Parfomak and Lave consider changes in utility ex-ante savings over a 24-year period. However, changes to consumption over an extended period are not the same as counterfactual conditions during the same period. A true counterfactual would contrast what consumption is in the presence of programmatic activity to what it would have been absent programmatic activity, but under the same conditions. Since there is a naturally occurring change in the technology mix over time, not related to programmatic activity, earlier periods do not provide an "apples to apples" comparison of the factual condition during the current period.

The Demand Research study examines five years of in-state (California) time series data. The model results report the change in consumption relative to changes in programmatic activity starting in 2008. This does not provide a true measure of the counterfactual (no-program) scenario because the study lacks a non-program area of comparison and does not extend to a pre-program period. Similarly, the Cadmus (2012) study also only considers in-state (California) data, but uses a longer time series (1990 to 2010). While this may not provide for a true "no program" baseline, the level of activity in the early 1990s should be sufficiently different to provide a meaningful point of comparison.

B. Diversity of program activity level across units of observation (time-geography combinations)

Horowitz's (2004) use of 12 years of data across 42 states provides for greater variation in the level of programmatic activity. Given the range of programmatic activity across the states included in the model, the model is able to use states with little to no programmatic activity as the counterfactual to states with programmatic activity during the same period. Further, by controlling for changes to fuel prices and other exogenous factors, such as GDP, it can use changes in the level of programmatic activity and consumption over time to estimate gross program impacts. Rivers and Jaccard, Parfomak and Lave, and Loughran and Kulick also provide for a greater diversity in programmatic activity as they construct models with over 10 years of data across regions that are diverse with respect to programmatic activity.

In Horowitz (2007), program activity is only defined in terms of states having “weak” or “high to moderate” programmatic activity. Horowitz’s model also does not provide evidence of the impact of different levels of programmatic activity on savings. Horowitz assigns states to the strong/moderate or weak program activity categories based on EIA rankings, only. Because Horowitz contrasts strong/moderate and weak program states during the same period and controls for differences in GDP, energy prices, and other exogenous factors, the approach does allow for isolation of changes in consumption due to variations in the level of programmatic activity over time. An extension of this approach would allow interaction of the program variable with heating degree days and cooling degree days to differentiate between heating and cooling programs. This would allow the model to detect temperature sensitive consumption changes relative to changes in consumption due to energy efficiency activities. Alternatively, the model could include terms for the presence of upstream, behavioral, or demand response programs to determine whether these programs had a greater influence on energy savings.

The Parfomak and Lave study is focused on utilities within the Northeast and California, where energy consumption and savings is more consistent over 24 years because they are reported by relatively few utilities with similar reporting requirements and procedures. However, the model does not include states with relatively low levels of programmatic activity. The model employs ex-ante savings as a measure of programmatic activity. While the model includes utility level fixed effects, it does not include any time-oriented fixed effects to explain annual differences in consumption not attributable to conservation efforts.

Because they are conducted for a single state only, the Demand Research (2012) and Cadmus (2012) studies have only limited differences in program activity level across units of observation, as in-state programmatic activity levels only vary slightly.

C. Consistent relationship between program activity and savings

Loughran and Kulick (2004) and Rivers and Jaccard both rely on publicly reported data across states, which can suffer from inconsistent measurement and reporting issues. Due to variability in the reporting of program expenditures and consumption, and variability in program offerings across states, the model does not distinguish between types of energy efficiency programs. Consequently, the approach does not provide information regarding the type of programmatic activity and savings. For programs that received relatively little funding, but resulted in substantial savings compared to other program types, the impacts may not be detected because the model aggregates program expenditures into a single value. Alternatively, Parfomak and Lave’s study is focused on utilities within the Northeast and California; therefore, reporting of energy consumption and savings is likely to be more consistent.

Demand Research (2012) only considers in-state (California) data. Due to consistency in the reporting of energy efficiency expenditures and consumption levels and consistency in the program offerings within the state, the approach provides a strong opportunity to detect the relationship between program activity and savings. Similarly, the Cadmus study only considered in-state (California) data. The results of this study report the change in utility-level consumption relative to changes in programmatic expenditures and buildings codes.

However, the Cadmus study does not include other program variables such as differences in marketing or incentive costs, or ex-ante savings. Because the reporting of program offerings and consumption across utilities in the state is standardized, the approach provides a strong opportunity to detect the relationship between program activity and savings.

Horowitz (2004) uses multiple measures of programmatic activity including energy efficiency savings per unit of GSP, and a market transformation index. These metrics are likely to suffer from considerable reporting issues across a wide geography. Meanwhile, Horowitz (2007) breaks states into high to moderate programmatic activity and weak programmatic activity, which avoids consistency issues.

D. Minimal effect of one area on another (cross-area spillover)

Horowitz (2007) attempts to estimate spillover using regression models from states with low levels of programmatic activity to represent counterfactual conditions. Horowitz (2010) uses a similar approach to measure spillover. However, the linear regression approach used by Horowitz (2004)—which aggregated data at the state level—obscures the ability to detect changes that result from differences in local conditions and changes in the types and/or magnitude of programmatic activity.

Rivers and Jaccard (2011) obtained consumption and program expenditure data at the utility service territory level; however, they chose to conduct the analysis at the provincial level in Canada. This reduces cross-unit spillover effects (spillover from one unit of analysis to the next) because the units where spillover may occur are aggregated However, using more aggregate data prevents detection of programmatic impacts at more granular levels of analysis, such as programmatic or service territory differences across utilities. Loughran and Kulick (2004) aggregate data to the state level, which minimizes cross unit spillover, but at the expense of detecting impacts at a less granular level.

The Cadmus (2012) analysis only provides for limited differences in program activity levels across units of observation within the state, and does not provide for detection of spillover from other states. Similarly, Demand Research (2012) does not provide for detection of spillover from other states. However, both of these studies do allow for within-state spillover.

E. Appropriate and consistent use of exogenous explanatory variables

Demand Research incorporates a number of key exogenous explanatory variables to control for exogenous factors and uses weather-normalized consumption as the dependent variable. The Cadmus study also incorporates a number of key exogenous explanatory variables to control for exogenous factors, but only includes heating and cooling degree days as explanatory variables rather than using weather-normalized consumption as a dependent variable.

The models with a greater national focus require exogenous variables aggregated to the national level, which suffers from inconsistent measurement and reporting issues. Loughran and Kulick and Rivers and Jaccard employ variables for energy prices, number of customers, and GSP. The models do not include variables to model the structural shift of the economy towards a more technology-based production, nor does the model account for changes to building codes. Horowitz (2004) relies on publicly reported data across states—including fuel costs, the share of generation by fuel type, price of electricity and natural gas, and trends of electronic business—which can suffer from inconsistent measurement and reporting issues as states may not track the same data and/or employ differing techniques for measuring and reporting on the data they do track. Similarly, Horowitz (2007) employs fuel prices, per capita income/GSP, and a technology trend.

F. Ability to measure program activity at the most granular geographic level with a large enough sample size

Demand Research uses normalized consumption based on utility billing data to construct the dependent variables at the customer account level. Jacobsen and Kotchen also examined account-level data, while Parfomak and Lave (1996) examined data at the utility level within states. Loughran and Kulick, Horowitz (2004), and Horowitz (2007) all look at state-level analysis, which obscures the ability to detect changes that result from differences in local conditions and changes in the types and/or magnitude of programmatic activity.

G. Long enough time series to detect and isolate program impacts

Horowitz (2007) and Parfomak and Lave employ the longest time series at 33 years and 24 years, respectively. Cadmus uses a 20-year time series, Rivers and Jaccard use a 15-year time series, and Horowitz (2004) uses 12 years of data. Demand Research includes only 5 years of data, which may be sufficient to account for the lag in savings and isolate program impacts, but a longer time series would be more desirable.

H. Account for the lag structure of program impacts

Loughran and Kulick include terms for a 12- and 24-month lag in program expenditures. Both Rivers and Jaccard and Horowitz (2004) include the lag of the dependent variable, but not the lag of the program variable, which Violette identifies as a flaw in the model design. The Cadmus (2012) California pilot study accounts for one to five periods of lags of the program variable. The Demand Research California pilot study only used five years of data, but this study uses cumulative savings and expenditures as indicators of program activity, which accounts for lag in program impacts on consumption. Horowitz (2007), Horowitz (2010), and Jacobsen and Kotchen do not consider the lag structures, as these studies use a pre- and post-difference in differences (pre- and post-program) research approach that employs a cutoff year for when programs ramped up.

4.2.6 Summary of econometric considerations for successful top-down model

Table 4-6 below summarizes the information presented above in terms of whether each of the studies reviewed addresses the econometric considerations for successful top-down modeling.

Table 4-6. Review of Studies Relative to the Econometric Criteria for Successful Top-down Modeling

Econometric considerations for successful top-down models	Loughran and Kulick	Rivers and Jaccard	Horowitz 2004	Horowitz 2007	Parfomak and Lave	Horowitz 2010	Demand Research	Cadmus	Jocobsen and Kotchen
Use of Differences Dependent	√	√			√		√		
Use of Differences in Explanatory Variables									
Program Variable	√	√			√		√	√	
Other Explanatory Variables	√				√		√	√	
Account for heteroscedasticity									
Natural Log of Dependent Variable	√				√	√	√		
Natural Log of Program Variable	√				√	√	√		
Natural Log of Other Explanitory Variables					√	√	√	√	
Estimated Using Weighted Least Squares			√		√	√	√	√	
Lag of Depdent Variable		√	√						
Account for Fixed Effects									
Annual Fixed Effects	√		√		√		√	√	√
Geographic Fixed Effects	√		√		√		√	√	
Difference in Differences Approach				√		√			√
Allow for Differences in Types of Programmatic Activity			√				√		
Multiple Measures of Program Influence							√	√	
Account for Changes in Consumption Resulting from Building Code Changes							√		√
Account for Energy Prices									
Electricity Prices	√	√	√	√	√	√	√	√	
Substiture Fuel Prices	√	√	√	√	√	√	√	√	

4.2.7 Summary of variables used in previous top-down studies

Before moving on to the model specification presented in Section 5, this section summarizes the data used in each of the previous top-down studies. The information used to construct the consumption (or savings) variable, as well as the program impact variables and other macro-economic (exogenous) variables, are presented in Table 4-7, Table 4-8, and Table 4-9 below. Each table lists the variables used in each study, identifies the source for each variable, and indicates the geographic and time series covered by the respective data series. Finally, we identify the various pros and cons of using each variable in the Year 1 pilot studies.

Table 4-7. Inventory of Dependent Variables Employed Existing Top-down Studies

Variable Type	Variable	Source	Geography	Time series available	Loughran and Kulick (2004)	Rivers and Jaccard	Horowitz 2004	Horowitz 2007	Parfomak and Lave 1996	Horowitz 2010	Demand Research 2012	Cadmus 2012	Jacobsen and Kotchen	Discussion
Dependent Variable	Monthly Energy Consumption	IOU billing database	Account level	Monthly						√	√		√	Pros: Data can be weather normalized at the account program level and industry level Cons: Processing time/cost—data may only be available for limited number of years.
	Annual Energy Consumption	IOU billing database	Account level	Monthly					√					Pros: Data can be weather normalized at the account, program, and industry level Cons: Processing time/cost—limits time series
		EIA 861	Utility/State/Sector	1970 to 2012	√		√	√		√				Pros: Long time series—wide geographic coverage Cons: Inconsistent reporting; only segmented by sector; not viable for lower levels of aggregation
		Stats Canada	Utility/Province/Sector	1970 to 2012		√								Pros: Long time series—wide geographic coverage Cons: Inconsistent reporting; only segmented by sector; not viable for lower levels of aggregation
		CEC database	Utility/County by sector	1990 to 2012								√		Pros: Detailed data reporting Cons: CA only; cannot weather normalize at account level

Table 4-8. Inventory of Program Impact Variables Employed Existing Top-down Studies

Variable Type	Variable	Source	Geography	Time series available	Loughran and Kulick (2004)	Rivers and Jaccard	Horowitz 2004	Horowitz 2007	Parfomak and Lave 1996	Horowitz 2010	Demand Research 2012	Cadmus 2012	Jacobsen and Kotchen	Discussion
Program Variables	Energy Efficiency Expenditures	EIA 861	Utility/State	1970 to 2012	√		√	√		√				Pros: National coverage Cons: Cannot differentiate by program or customer type; inconsistent reporting
		Stats Canada	Utility/Province/Sector	1970 to 2012		√								Pros: National coverage Cons: Cannot differentiate by program or customer type; inconsistent reporting
		Utility records	Account level	Varies							√			Pros: Allows for segmentation by program and customer type Cons: Processing time/cost
		EE Groupware Application	Utility by sector									√		Pros: Detailed data reporting Cons: CA only; cannot weather normalize at account level
	Ex-ante Savings	EIA 861	Utility/State	1970 to 2012			√							Pros: Long time series—wide geographic coverage Cons: Inconsistent reporting; only segmented by sector; not viable for lower levels of aggregation
		Utility records	Account level	Varies							√			Pros: Allows for account to state level analysis and segmentation by customer, program, or measure type or geography Cons: Processing time/cost—data may only be available for limited number of years.
	Market Transform Index	US Census Horowitz 2001 Survey data	Custom	Custom			√							Pros: Accounts for spillover Cons: Many assumptions to construct
	Code Change: Pre-Post-Code Change	Record of historical code changes	Account level	Historical									√	Pros: Simple approach Cons: Requires assumption that code impacts immediate.
	Code Change: Index of New Construction by Year	Dodge Players Database	ZIP code	1996–2013								√		Pros: Capture change in building stock over time. Cons: Limited to New Construction; data processing and cost.
	Pre- Post Program Indicator	Based on historical escalation in programs	State level	Used 1992						√				Pros: Simple approach Cons: Requires assumption that program impacts immediate.
	Grouping units into High and Low Activity	EIA rankings	State level	Varies				√		√				Pros: Eliminates errors associated with reporting savings and expenditures. Cons: Does not allow for detection to program differences, or relative impact of expenditures; not aviable for lower levels of aggregation

Table 4-9. Inventory of Macro-Economic Variables Employed Existing Top-down Studies

Variable Type	Variable	Source	Geography	Time series available	Loughran and Kulick (2004)	Rivers and Jaccard	Horowitz 2004	Horowitz 2007	Parfomak and Lave 1996	Horowitz 2010	Demand Research 2012	Cadmus 2012	Jacobsen and Kotchen	Discussion
Macro Economic Variables	Gross State product	US Census	State/County/ Town				√	√		√				Pros: Available for County by NAICs; or down to Census Tract Cons: Not available for Town or Census Tract by NAICS
	Personal Income	US Census	County					√		√		√		
	Energy Prices	CEC database	Utility/County by sector									√		Pros: Local prices Cons: Limited to CA
		FERC and EIA			√		√	√	√	√	√			Pros: Extensive price history Cons: More generalized
	Equipment Stock - Technology trend of energy using equipment	Federal Reserve	State Level	Annual				√						Pros: Shows changes in energy using equipment Cons: Aggregate geography
	Residential Appliance Saturations	Census: American Community Survey	Census tract	Annual								√		Pros: Data reported annually at census tract level Cons: Annual series modeled off rolling 5-year survey.
	Share of generation by fuel type	FERC and EIA databases	Utility/State	Annual			√							Pros: Allows for detection of changed in generation Cons: Not available for lower levels of geography
	Household Characteristics	American Community Survey	Census tract	1990–present							√			Pros: Detailed household information. Cons: Aggregate data only; data modeled off annual survey

4.2.8 Summary of literature review findings and limitations

The literature review illustrates that there have been a range of approaches employed to measure programmatic impacts at different levels of analysis, and a range of data inputs. Each study offers different strengths and weaknesses relative to its ability to address the desirable properties of top-down models. While none of the studies provided a single approach for isolating net programmatic impacts from other influences on consumption, these approaches provided much guidance for the Year pilot studies. In that regard, the following conclusions can be made from the existing literature:

- *Account for fixed effects* – The studies employ multiple techniques to account for fixed effects that include using the first difference in the dependent and explanatory variables, and including annual and geographic unit fixed effects terms. For longer time series, the studies also show that it is important to consider periods of major structural changes in the energy economy.
- *Lagged program impacts* – It is important to consider the lagged effect of programmatic activity on consumption. This requires using multiple lagged terms for programmatic impacts. Cadmus shows that consumption in the current period may be impacted by programmatic activity of up to five years prior.
- *Measures of differing program types* – Horowitz (2004) shows that accounting for energy efficiency program expenditures or ex-ante savings alone may provide artificially low estimates of programmatic impacts by not accounting for market transformation impacts or impacts associated with upstream programs.
- *Use of scaled dependent variable* – The consensus among studies is that top-down models should seek to measure changes to energy consumption per unit (e.g., GSP, employee, household) in order to standardize estimates across locations and times.
- *Weather normalization* – Most of the existing studies include heating degree days and cooling degree days as explanatory variables and use non-normalized consumption as the dependent variable. Further, these studies do not attempt to distinguish among heating and cooling impacts of programmatic activity. The recent Demand Research California pilot study marks a departure from this shortcoming, and uses normalized consumption based on utility records as the dependent variable. This is a desirable property provided the data are available for a sufficiently long time series.

This methods review provided a number of key insights for conducting the Year 1 pilot studies in Massachusetts. We summarize these as follows:

- *Level of analysis* – The desired level of analysis for detecting programmatic effects should extend beyond Massachusetts. While the Year 1 analysis is only limited to measuring changes that occur within Massachusetts, to effectively capture sufficient differentiation in programmatic activity, the longer term study should be expanded to include neighbouring states in which the PAs' service territories extend, including Connecticut, Rhode Island, New Hampshire, Maine, and New York. In addition, the evaluation team should investigate opportunities for partnering with utilities from other regions of the country to obtain data for areas with limited programmatic activity.

- *Need to obtain additional billing and tracking history* – The analysis of the CPUC pilot studies indicates that a limited time series is unlikely to result in adequate variation in programmatic activity to construct statistically significant estimates of net savings. The Loughran and Kulick study and the re-analysis of the Rivers and Jaccard study by Violette illustrate that models should have at least ten years of billing and tracking history in order to produce reliable top-down estimates of program impacts. As such, the PA Data model pilot study, which includes only three years of consumption and program activity data, focuses on development of the approach. We explored whether the modeling approach using the available data provided sufficient signal to recommend further research with a longer time series.
- *Need for time series adjustments to models* – The literature review indicates that it is important for top-down studies to use the first-difference in the dependent and independent variables rather than estimating the model using the annual values themselves.
- *Account for heteroscedasticity* – Statistical analysis of energy consumption often requires adjustments for heteroscedasticity. This can be accomplished using a number of statistical techniques, but the most common are using the log transformation of the dependent and independent variables, and using weighted least squares regression. The evaluation team included the log transformation in the current pilot studies. Once the evaluation team refines the current techniques to select a final model specification, further research should test for heteroscedasticity to make corrections if necessary.
- *Account for fixed effects* – The studies reviewed point to the importance of including annual and location-specific fixed effects.
- *Multiple measures of programmatic activity* – A number of studies illustrate the importance of a variety of measures of programmatic activity on consumption. The pilot studies considered changes in program expenditures and ex-ante savings as well as the effects of codes and standards on consumption. Where possible, the models examined differences in the type of program offerings, such as upstream and downstream programs, or measures of market maturity, such as the percentage of savings resulting from non-lighting measures.

5. PA-MUNICIPAL TOP-DOWN MODEL

In this pilot study, the evaluation team developed a version of a macro-consumption model using aggregate electricity consumption data for PAs (at the PA level) and municipal utilities in Massachusetts.[28] The team modeled these data as a function of exogenous variables including program activity, price, weather, and other demographic and economic factors affecting consumption. The evaluation team ran separate electric models for the residential and commercial and industrial (C&I) sectors. By controlling for other factors that could cause the diverging trends in electricity consumption between the PAs and municipal utilities, this top-down model sought to isolate the effect of energy efficiency programs on consumption. The substantial differences in energy efficiency program expenditures across the PAs and the municipal utilities in a given year—and within PA and municipal utilities over time—provided the identifying variation for the model.

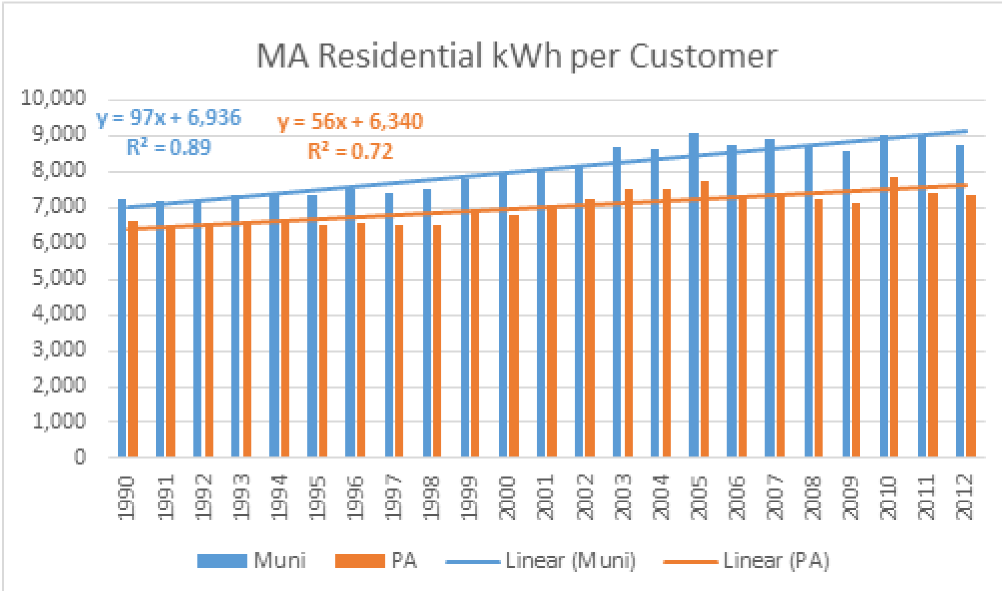
One of the primary motivations for the PA-Muni top-down approach was an important initial analysis conducted by Lawrence Masland of the Massachusetts Department of Energy Resources. Mr. Masland examined trends in per-customer residential energy consumption for PAs and municipal utility customers for 1990 through 2011 using the data from the Energy Information Administration's (EIA) Annual Electric Power Industry Reports (EIA-861 data files). His analysis showed that average annual residential electricity consumption per customer for both PAs and municipal utilities has increased from 1990 to 2011, but the rate of increase was significantly higher for the municipal utilities. While he hypothesized that the lower levels of increase in electricity consumption in the PA territories could be due to greater programmatic activity, his Massachusetts work used a simple linear regression-based approach that did not control for other exogenous factors. Greater accuracy in isolating programmatic impacts would use additional variables to control for structural and exogenous trend factors, which would better isolate the effect of program activity from natural changes and policy variables.

Another primary motivation for this model was to establish the counterfactual (no-program) scenario. Understanding the true extent of a program's impacts requires information regarding the level of consumption absent any programmatic activity. This PA-Muni approach extends the timeframe long enough to include a period with no programmatic activity, at least for municipal utilities.

As shown in Figure 5-1, average annual residential electricity consumption per customer for both PAs and municipal utilities increased from 1990 to 2012, but the rate of increase is significantly higher for the municipal utilities than for the PAs.

[28] The evaluation team also considered a version of a macro-consumption model that used aggregate natural gas consumption data for investor-owned and municipal utilities in Massachusetts. However, there was not enough variation in the program activity variables for that model to provide a reliable estimate because there are only four municipal utilities providing gas service in Massachusetts.

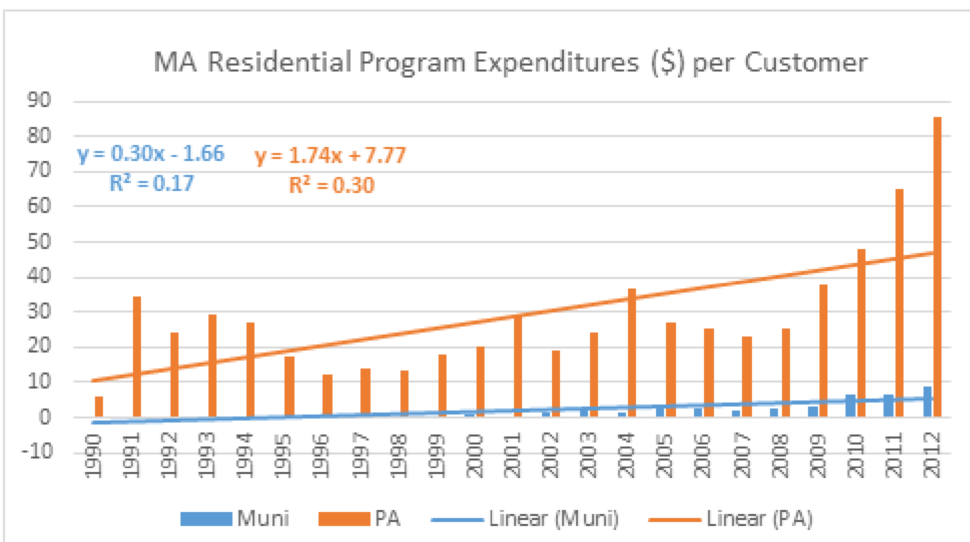
Figure 5-1. Trends in Annual Residential Electricity Consumption per Customer (in kWh), 1990–2012, Massachusetts



Source: Annual Electric Power Industry Report (EIA-861 data file)

While most, if not all, municipal utilities had residential energy efficiency programs during the period Masland examined (1990 to 2011), the municipal utilities were slow to embrace the funding of energy efficiency programs, and funding levels were significantly below those of the PAs, as shown in Figure 5-2.

Figure 5-2. Trends in Annual Residential Energy Efficiency Program Expenditures per Customer (in $), 1990–2012, Massachusetts



Source: Massachusetts Program Administrators, municipal utilities, and the Massachusetts Municipal Wholesale Electric Company (MMWEC)

While the lower rate of increase in consumption in PA territories could be due to greater programmatic activity, accurately estimating programmatic impacts would require controlling for structural and exogenous trend factors, which would allow for isolation of the effect of program activity from natural changes and policy variables.

5.1 PA-MUNI TOP-DOWN MODEL SPECIFICATION

The team specified a fixed-effects panel regression model. This type of regression model allows each individual to act as its own control. The unique effects of the stable, but unmeasured characteristics of each utility are their "fixed effects" from which this method takes its name. These fixed effects are held constant in the model. The fixed-effects nature of the model means the model does not need to include unchanging characteristics. In a model of households, for example, these characteristics might include square footage, number of floors, and direction the home faces, etc. In this study's model, this includes characteristics of these areas that do not change over time. These might include that Boston is the home of the state capitol with the state's tallest buildings, that the Cape gets sea breezes and usage that varies with vacation travel, and that western Massachusetts has the Berkshire Mountains, more rural areas, and the greatest differences in topography, etc. Including fixed effects in the model controls the amount of variance (noise) that the model must address to explain electricity consumption. This approach also provides for a much closer fit to the data than other types of regression models.[29]

The fixed-effects panel regression model allowed the team to estimate electricity consumption differences across PAs and municipal utilities over time. Program activity, the variable of interest for estimating program impact, was incorporated through program expenditure data. The model was specified with lagged program activity variables to account for the time between program implementation and program-induced electricity consumption reduction, and because of the fact that energy efficiency investments continue to yield savings for the life of measures installed.

5.1.1 PA-Muni residential top-down model specification

The residential models sought to estimate the impact of energy-efficiency program expenditures on electricity consumption by separating that effect from other causes of changes in usage. These other causes were controlled for in the analysis by incorporating the following factors:

- Price of electricity
- Heating degree days
- Cooling degree days

[29] The inclusion of fixed effects in the model ensures that the estimated regression coefficients are not biased due to non-time-varying (i.e., PA/utility-specific) characteristics. A random-effects specification is more efficient, but using random effects does not fully control for all utility-specific characteristics. Hausman tests were used to determine which model specification to use. The findings from those tests showed that fixed effects were more appropriate for this analysis.

- Household income
- Home values
- Proportion of households using electricity as the primary heating fuel
- Amount of residential new construction in the housing stock
- Proportion of single-family homes in the housing stock
- Proportion of renter-occupied housing in the housing stock
- Employment/unemployment rate
- Time trends.

Other explanatory variables considered by the team included home square footage, age of housing stock, household education level, and Green Community status. The data on home square footage were not available. The team used home values as a proxy for home square footage. Age of housing stock would be highly correlated with the amount of new residential construction. Similarly, household income would be highly correlated with household education level. In order to avoid the issue of multicollinearity in explanatory variables, only one of the two correlated variables was included in the model. Finally, the town-level Green Community status indicator was excluded from the model because the models were run at the PA/utility level.

A. PA-Muni residential top-down model data

The evaluation team collected time-series data on residential electricity consumption and factors that could affect consumption for all Massachusetts PAs/munis and towns from 1990 to 2012. The team also developed a panel database for the study, which included the following data elements:

- *Electricity Consumption and Price Data* – The team collected data on the total residential electricity sales, revenue, and customers in Massachusetts from the EIA's 861 files for 1990-2012 for each PA and municipal utility. The team derived the annual energy consumption per customer and average price per kWh using these data.
- *Energy Efficiency Programmatic Activity* – The evaluation team then assessed the quality of demand-side management program data reported to EIA on the EIA-861 form. The assessment revealed that data were missing and/or inconsistent for some PAs and municipal utilities for some years. Moreover, these data did not separate energy efficiency program expenditures by sector until 2008. Because it was crucial to gather accurate information for the main explanatory variables of interest for the model in order to produce reliable estimates, the evaluation team made a substantial effort to collect the energy efficiency program expenditures data by sector and year from the PAs, the municipal utilities, and their association.[30]

[30] Despite these efforts, the team could not get program data for 12 municipal utilities in the state. These utilities were excluded from the analysis. We would especially like to thank Kim Boas of the

- *Weather Data* – The evaluation team gathered daily temperature data for all weather stations in Massachusetts from the National Oceanic and Atmospheric Administration (NOAA) from 1990 through 2012. The team first computed the annual heating degree days (HDDs) and cooling degree days (CDDs) for each station. Next, the team matched each town to the nearest weather station. Finally, the team computed a weighted average of annual HDDs and CDDs for each PA/utility service area using the number of housing units in each town as the weight.

- *Economic and Demographic Data* – The evaluation team gathered town-level economic and demographic data from the following sources:
  - US Census American Community Survey (ACS) – Contains annual residential socioeconomic data at the census block, the smallest geographic unit used by the US Census Bureau, level of granularity since 2005.
  - US Decennial Census – Contains residential socioeconomic data at the census block level of granularity. Conducted in 1990, 2000, and 2010.[31]
  - US Census Building Permits Survey – Contains annual construction statistics by permit-issuing place (usually the township) on new privately owned residential housing units authorized by building permits.
  - Bureau of Labor Statistics – Contains annual labor force, employment, and unemployment counts at the town level of granularity.

The team initially considered running the models at the town level because the economic and demographic data were available at that gradation. This would have allowed for a better comparison of PAs and municipal utilities given that most municipal utilities serve only a single town, while the PAs serve a large number of towns. However, because the PAs' energy consumption and energy efficiency program data were available at the PA level only, the team aggregated the town-level economic and demographic data to the PA level.

A significant challenge for this study was to collect consistent electric program data across all PAs and municipal utilities. While the evaluation team attempted to collect detailed time-series data on program activity, the only consistent piece of data that the team was able to gather across all PAs and municipal utilities was the annual total electric program expenditures. The collection of data from the municipal utilities was especially challenging because municipal utility participation was completely voluntary. While the team requested program data from 1990 through 2012, many utilities noted that the older program records

---

Massachusetts Municipal Wholesale Electric Company (MMWEC) for providing data for its members and the Massachusetts PAs for providing the data for their utilities.

[31] Decennial Census data (1990 and 2000) were used for the period before the annual ACS data were available. In order to make the decennial data fit into a data set with yearly time points, the team estimated the difference between the two points (1990 and 2000, for example) and evenly distributed the difference annually between the two data collection points, thereby forcing the decennial data to vary from year to year.

were most likely in a paper file somewhere and that electronic filing systems had undergone changes making data retrieval a laborious process. Many municipal utilities told the team that they did not have the time or resources necessary to compile historical program data. When municipals were willing to provide program data, they were limited in the sort of data they had access to and were usually only able to provide total program expenditure at the utility level.

B. PA-Muni residential top-down model detailed specification

Equation 5-1 below shows the residential PA-Muni top-down model specification. Since there is a significant variation in the size of PAs and municipal utilities, the models were weighted by the amount of residential electricity sales to properly represent the different magnitudes of spending and potential savings across the PAs and municipal utilities in Massachusetts. The regression model uses the natural logarithm of each of the variables as is recommended in a number of the studies reviewed in Section 4.2 above, to control for extreme values and so the relationship between the dependent variable to independent variables can be interpreted as elasticities. Since there is a significant variation in the size of PAs and municipal utilities, the models were weighted by the amount of residential electricity sales to properly represent the different magnitudes of spending and potential savings across the PAs and municipal utilities in Massachusetts.

Equation 5-1. PA-Muni Residential Top-down Model

$$\boldsymbol{\log(EC_{it}) = \beta_1 \, log(P_{it}) + \beta_2 \log(HDD_{it}) + \beta_3 \log(CDD_{it}) + \beta_4 \log(I_{it}) + \beta_5 \, EH_{it} + \beta_6 VAL_{it}}$$

$$+ \beta_7 \, NC_{it} + \beta_8 \, SF_{it} + \beta_9 \, RENT_{it} + \beta_{10} \, EMP_{it} + \sum_{j=0}^{n} \alpha_j EE_{it-j} + \beta_{11}\tau_t + \delta_i + \varepsilon_{it}$$

Where:

$\log(EC_{it})$	=	Natural logarithm of annual consumption per residential customer in PA/utility service area *i* and year *t*.
$\log(P_{it})$	=	Natural logarithm of electricity price in 2012 dollars in PA/utility service area *i* and year *t*.[32] The coefficient β_1 measures the price elasticity of electricity consumption.
$\log(HDD_{it})$	=	Natural logarithm of annual heating degree days (base 65) in PA/utility service area *i* and year *t*. The coefficient β_2 measures the elasticity of electricity consumption with respect to heating degree days.
$\log(CDD_{it})$	=	Natural logarithm of annual cooling degree days (base 70) in PA/utility service area *i* and year *t*. The coefficient β_3 measures the elasticity of electricity consumption with respect to cooling degree days.
$\log(I_{it})$	=	Natural logarithm of median household income in 2012 dollars in PA/utility service area *i* and year *t*. The coefficient β_4 measures the elasticity of

[32] Nominal prices were adjusted to reflect 2012 dollars using the GDP implicit price deflator from the Federal Reserve Economic Data.

electricity consumption with respect to household income.

EH_{it}	=	The share of households using electricity as the primary heating fuel in PA/utility service area *i* and year *t*. The coefficient $\beta_5$captures the effect of electricity used for heating on electricity consumption.
VAL_{it}	=	The median house values in 2012 dollars in PA/utility service area *i* and year *t*. The coefficient $\beta_6$captures the effect of the home values on electricity consumption.
NC_{it}	=	The share of new construction in residential housing, computed as the total number of residential new construction permits divided by the total number of housing units in PA/utility service area *i* and year *t.* The coefficient β_7 captures the effect of new construction on electricity consumption.
SF_{it}	=	The share of single-family homes in residential housing, computed as the total number of single-family housing units divided by the total number of housing units in PA/utility service area *i* and year *t.* The coefficient β_8 captures the effect of housing type on electricity consumption.
$RENT_{it}$	=	The share of renters in PA/utility service area *i* and year *t.* The coefficient β_9 captures the effect of home ownership on electricity consumption.
EMP_{it}	=	The employment rate, computed as the number of employees divided by the number of people in the labor force in PA/utility service area *i* and year *t.* The coefficient β_{10} captures the effect of employment/unemployment rate on electricity consumption.
EE_{it-j}	=	Total residential electric energy efficiency program expenditures per residential customer in PA/utility service area *i* and year *t-j.* The coefficient α_j measures the percentage change in electricity consumption in year *t* from a one-dollar change in energy efficiency program expenditures in year *t-j*. The sum of α_0 through $\alpha_n$measures the percentage change in electricity consumption in year *t* from a one-dollar change in energy efficiency program expenditures in year *t* and the previous *n* years.[33]
τ_t	=	Time-trend variable that is equal to 1 in 1990 and increasing by one unit annually. The coefficient β_{11} captures the naturally occurring change in electricity consumption not captured by the variables included in the model.[34]

[33] The team also tested specifications with distributed lag models with a special parameterization of lagged energy efficiency expenditures variables in order to account for the possible non-linear and delayed effects of energy efficiency program activity on consumption. The results were similar.

[34] As a robustness check, the team also tested specifications with non-linear (a natural cubic spline, or some second- or third-degree polynomials) time trends. This had little effect on the results. Similarly,

δ_i	=	PA/utility fixed effects that capture time-invariant PA/utility-specific fixed effects in electricity consumption. There may be a certain PA/utility-level variation in the data that is not necessarily related to energy efficiency programmatic activity, such as changes to the local economy resulting from local businesses closing.
ε_{it}	=	Regression error term in PA/utility service area *i* and year *t*.

Massachusetts PAs have had residential upstream lighting programs since 1998, and these programs have accounted for a significant share of program-claimed savings for the PAs. The incentive structure of these programs does not allow for assurances that each purchaser of a program bulb is a residential customer in the sponsoring PA's service territory. Therefore, some program bulbs may have been purchased by customers served by municipal utilities. This leakage means that some of the program expenditures in the neighboring area are affecting consumption in the municipal utility. The team re-aligned program expenditure and the efficiency it is purchasing to appropriately estimate impacts on consumption from program expenditures. The team used the work from a component of the Massachusetts Residential Customer Profile study, which allocated total upstream program rebate dollars to census block groups in Massachusetts from 2010 through 2013, and reallocated a portion of PA electricity program expenditures to municipal utilities from 1998 through 2012 using the following steps:

1. Compute the amount and percentage of total rebate dollars that ended up in each PA and municipal utility service territory in 2010 through 2012 after matching census block groups to PA and municipal utility service territories.
2. Compute the share of lighting program rebate dollars in total energy efficiency program expenditures by PAs from 2010 through 2012.
3. Compute the amount of leaked rebate dollars from PAs to municipal utilities in each year from 1998 through 2012, assuming the same average ratio of rebate dollars to total expenditures and the same average leakage rates.
4. Subtract these leaked rebate dollars from PAs' energy program expenditures and add them to those of the municipal utilities based on their share of total leaked dollars from Step 1.

The evaluation team's modeling approach is similar to the one that Cadmus took in California, with two important differences.[35] First, in the comparable PA-Publicly Owned Utility (POU) models, Cadmus relied on the DSM expenditures data reported to EIA on the EIA 861 form

including the indicator variables for individual years instead of a time trend did not result in a significant change in the model results.

[35] Cadmus Group, Inc. "CPUC Macro Consumption Metric Pilot study (Final Report)." Prepared for the California Public Utilities Commission. October 19, 2012.

and acknowledged reporting, consistency, and other issues with these data. The evaluation team observed similar issues with the EIA-861 DSM expenditures data for the Massachusetts PAs and municipal utilities, and decided not to use them. Instead, the evaluation team made a substantial effort to collect more accurate and reliable energy efficiency expenditures data by sector from the PAs, municipal utilities, and the municipal utilities' association in Massachusetts. Second, the evaluation team's approach is the first attempt to account for the upstream lighting program CFLs that may have been purchased by customers served by non-sponsoring utilities.

5.1.2 PA-Muni C&I top-down model specification

The C&I models sought to estimate the impact of energy efficiency program expenditures on electricity consumption by separating that effect from other causes of changes in usage. These other causes were controlled for in the analysis by incorporating the following factors:

- Price of electricity
- Heating degree days
- Cooling degree days
- Average employment income
- Square footage of C&I new construction
- C&I building types (per NAICS classification)
- Employment/unemployment rate
- Time trends.

We first considered using C&I electricity consumption intensity—i.e., electricity use per square foot of floor space—as the dependent variable. However, reliable information on square footage could not be gathered from the public data sources. As a result, electric C&I consumption was expressed as per customer, per establishment, or per employee in the models.

A. PA-Muni C&I top-down model data

In parallel to the residential model data collection, the evaluation team collected time-series data on C&I electricity consumption and factors that affect consumption for all Massachusetts utilities and towns for 1990 to 2012. The team developed a panel analysis database that included the following data elements:

- *Electricity Consumption and Price Data* – The data on the total C&I electricity sales, revenue, and customers in Massachusetts by PA/utility were collected from the EIA's 861 files for 1990 to 2012. An inspection of the data showed that some utilities changed the classification of non-residential customers over time. Many year-to-year changes occurred in commercial sales, and an almost equal and opposite change occurred in industrial sales, suggesting that utilities reported sales as industrial in the previous year and as commercial in the current year. Given this inconsistency, the team aggregated the commercial and industrial sectors into a single category. The team then derived the annual energy consumption per C&I customer and average price per kWh from these data.

- *Energy Efficiency Programmatic Activity* – First, the evaluation team assessed the quality of demand-side management program data reported to EIA on the EIA-861 form. These data contained missing and/or inconsistent information for some PAs and municipal utilities for some years. Moreover, these data did not separate energy efficiency program expenditures by sector until 2008. Since it was crucial to gather accurate information for the main explanatory variables of interest for the model in order to generate reliable estimates, the evaluation team made a substantial effort to collect the energy efficiency program expenditures data by sector and year from the PAs, municipal utilities, and the municipal utility association. The data indicated that municipal utilities did not have C&I energy efficiency programs until 2009, and only four municipal utilities had C&I programs after 2009.

- *Weather Data* – The evaluation team gathered daily temperature data for all weather stations in Massachusetts from the National Oceanic and Atmospheric Administration (NOAA) from 1990 through 2012. The team first computed the annual heating degree days (HDDs) and cooling degree days (CDDs) for each station. Next, the team matched each town to the nearest weather station. Finally, the team computed a weighted average of annual HDDs and CDDs for each PA/utility service area using the number of employees[36] in each town as the weight.

- *Economic and Firmographic Data* – The evaluation team gathered town-level economic and demographic data from the following sources:
  - US Census ZIP Business Patterns – Contains business (establishment) and employee counts by size and by North American Industry Classification System (NAICS) industry type, summaries by ZIP code (without industry breakdown) for employment, payroll, and counts by employment size.
  - Bureau of Labor Statistics – Contains annual labor force, employment, and unemployment counts at the town level of granularity.
  - McGraw Hill Dodge C&I New Construction Database – This database, purchased as part of the Massachusetts C&I Program Evaluation, contains information on project square footage, value, type, and location for all nonresidential new construction projects in Massachusetts from 1996 through 2011.

B. PA-Muni C&I top-down model detailed specification

Equation 5-2 below shows the C&I PA-Muni top-down model specification. We initially considered running the C&I models at the town level because the economic and demographic data were available at that level. This would have allowed for both a better assessment of the relationship between energy consumption and energy efficiency expenditures, after controlling for other factors, and a better comparison of PAs and municipal utilities given that most municipal utilities serve only a single town, while the PAs serve a large number of towns. However, because the PAs' energy consumption and energy efficiency program data

[36] The information on the number of employees is from the US Census ZIP Business Patterns data.

were available only at the PA level, the team computed a weighted average of economic and variable factors at the PA/utility level using the number of employees in each town as the weight.

Equation 5-2. PA-Muni C&I Top-down Model

$$\log(EC_{it}) = \beta_1 \log(P_{it}) + \beta_2 \log(HDD_{it}) + \beta_3 \log(CDD_{it}) + \beta_4 \log(EINC_{it}) + \beta_5 NC_{it} + \beta_6 EMP_{it} + \sum_{k=1}^{20} \gamma_k NAICS_{k,it} + \sum_{j=0}^{n} \alpha_j EE_{it-j} + \beta_7 \tau_t + \delta_i + \varepsilon_{it}$$

Where:

$\log(EC_{it})$	=	Natural logarithm of annual consumption per customer, per establishment, or per employee in PA/utility service area *i* and year *t*.
$\log(P_{it})$	=	Natural logarithm of electricity price in 2012 dollars in PA/utility service area *i* and year *t*.[37] The coefficient β_1 measures the price elasticity of electricity consumption.
$\log(HDD_{it})$	=	Natural logarithm of annual heating degree days in PA/utility service area *i* and year *t*. The coefficient β_2 measures the elasticity of electricity consumption with respect to heating degree days.
$\log(CDD_{it})$	=	Natural logarithm of annual cooling degree days in PA/utility service area *i* and year *t*. The coefficient β_3 measures the elasticity of electricity consumption with respect to cooling degree days.
$\log(EINC_{it})$	=	Natural logarithm of mean annual employment income per employee, in 2012 dollars, computed as total annual payroll divided by total number of employees in PA/utility service area *i* and year *t*.
NC_{it}	=	Square footage of C&I new construction per customer, per establishment, or per employee in PA/utility service area *i* and year *t*.
$\mathrm{NAICS}_{k,it}$	=	The percent of establishments in a two-digit NAICS industry code *k* in PA/utility service area *i* and year *t*. The establishments in Massachusetts belonged to 21 different two-digit NAICS codes. The γ_k is a vector of coefficients that capture the differences in building energy use by business type.
EMP_{it}	=	The employment rate, computed as the number of employees divided by the number of people in the labor force, in PA/utility service area *i* and year *t.*
EE_{it-j}	=	Total commercial and industrial energy efficiency program expenditures per C&I customer, per establishment, or per employee in PA/utility service

[37] Nominal prices were adjusted to reflect 2012 dollars using the GDP implicit price deflator from the Federal Reserve Economic Data.

		area *i* and year *t-j.* The coefficient α_j measures the percentage change in electricity consumption in year *t* from a one-dollar change in energy efficiency program expenditures in year *t-j*. The sum of α_0 through α_n measures the percentage change in electricity consumption in year *t* from a one-dollar change in energy efficiency program expenditures in year *t* and the previous *n* years.
τ_t	=	Time-trend variable that is equal to 1 in 1990 and increasing by one unit annually. This time-trend variable captures the naturally occurring change in electricity consumption not accounted for by the variables included in the model.[38]
δ_i	=	PA/utility fixed effects that capture time-invariant, PA/utility-specific fixed effects in electricity consumption. There may be a certain PA/utility-level variation in the data that is not necessarily related to energy efficiency programmatic activity, such as changes to the local economy resulting from local businesses closing.
ε_{it}	=	Regression error term in PA/utility service area *i* and year *t*.

5.2 PA-MUNI TOP-DOWN MODEL RESULTS: RESIDENTIAL MODEL

The team estimated the models with and without the lagged energy efficiency program expenditure variables using the software package *Stata*. The estimation sample included data from all Massachusetts electric PAs (Cape Light Compact, National Grid, NSTAR, Unitil, and WMECo)[39] and 28 municipal utilities that had consistent energy efficiency program data available from 1994 through 2012. In all models, observations were weighted by PA/utility annual total residential sales (in GWh).

Table 5-1 shows summary statistics for key model variables for PAs and municipal utilities included in the estimation sample for the years 2000 through 2012.[40] The statistics are weighted by PA/utility total residential sales. For PAs and munis combined, the average annual electricity consumption per residential customer was 7,533 kWh and the average

[38] The team also tested specifications with non-linear (a natural cubic spline, or some second or third degree polynomials) time trends. This had little impact on the results.

[39] EIA-861 energy consumption and price data were available separately for former National Grid (Massachusetts Electric Co. and Nantucket Electric Co.) and NSTAR (Boston Edison Co., Cambridge Electric Light Co., and Commonwealth Electric Co.) companies. In order to take advantage of these data and improve the precision of the estimates, the team included them in the analysis as five separate companies. The annual energy efficiency program expenditures per customer by National Grid and NSTAR were distributed equally among their former companies. Finally, the energy program expenditures by Cape Light Compact were counted under the Commonwealth Electric Co.

[40] Consistent data on model variables are available starting in 1994. However, the inclusion of lagged values of energy efficiency program expenditures (up to 6 lags) requires that the analysis start in year 2000 because that is the first year in which the data for the energy efficiency expenditures for the previous six years are available.

annual energy efficiency program expenditure per customer was $32.40. The municipal utilities, on average, had higher per-customer electricity consumption (8,766 kWh) and lower per-customer energy efficiency expenditures ($3.00) than did PAs (7,392 kWh and $32.40, respectively).

Table 5-1. Residential PA-Muni Model Summary Statistics, Weighted, 2000–2012

Variable	All	PA	Municipal
Residential annual electricity consumption (kWh) per customer	7,533 (845)	7,392 (659)	8,766 (1,219)
Price of electricity (cents per kWh in 2012 $)	15.3 (2.5)	15.5 (2.5)	13.3 (1.8)
Annual HDDs (Base 65)	6,044 (469)	6,023 (461)	6,225 (518)
Annual CDDs (Base 70)	275 (89)	277 (88)	264 (93)
Median household income (in 2012 $)	69,211 (11,028)	67,729 (7,346)	82,220 (23,048)
Median home values (in 2012 $)	335,811 (101,865)	332,344 (95,203)	366,256 (146,897)
Percent of homes using electricity as the main heating fuel	12.9 (3.2)	13.1 (2.6)	11.3 (6.5)
Percent of residential new construction	0.5 (0.4)	0.5 (0.3)	0.8 (1.0)
Percent of single-family homes	57.7 (12.7)	56.3 (11.9)	69.8 (13.5)
Percent of renters	36.0 (8.6)	37.2 (7.2)	25.7 (11.0)
Percent employed	93.7 (2.0)	93.5 (2.0)	94.8 (1.9)
Annual residential energy efficiency program expenditures per customer ($)	32.4 (21.9)	35.7 (20.7)	3.0 (5.6)

Notes: All values are averages across the 34 PAs and municipal utilities (7 PAs and 28 municipal utilities in the years between 2000 and 2012). Standard deviations are given in parentheses. The statistics are weighted by PA/utility annual total residential sales.

Our model-building criteria prioritized theoretical relevance over observed explanatory power. Therefore, we kept the same explanatory variables in all models even if some explanatory variables did not come out to be statistically significant in some models. Table 5-2 shows the results from six different residential models. In each model, the dependent variable is the annual average electricity consumption per customer. A summary of the results from each model is described below.

Model 1 shows the results in which current electricity consumption is modeled as a function of current-year energy efficiency expenditures and other factors affecting electricity consumption.[41] As expected, weather is a significant factor in residential electricity consumption. For example, a 10% increase in heating degree days would increase electricity consumption by about 1.7%. Median home value is also a significant factor in electricity consumption. The coefficient -0.00014 of annual residential energy efficiency program expenditures per customer in year *t* is not statistically significant at a 90% confidence level. This model does not capture the lagged impact of the energy efficiency programs on energy consumption. In addition, the impact of current program expenditures on current consumption could be twice as large if expenditures were distributed uniformly in a given year because, in that case, each dollar of current-year expenditures would affect only one-half of current-year consumption.

Model 2 shows the results in which current electricity consumption is modeled as a function of current-year energy efficiency expenditures and those of the previous four years, as well as other factors affecting electricity consumption. The lagged energy efficiency expenditures included in the model capture the impact of the measures installed in the previous four years on current consumption, as well as the market effects. While all of the lagged energy efficiency program expenditure coefficients have the expected negative sign, the current year energy efficiency program expenditures have a positive sign. The first and the fourth year lag coefficient is statistically significant at a 99% confidence level. The coefficients of energy efficiency program expenditures are also jointly significant at a 99% confidence level ($F(5,34)=8.1$, $p=0.000$). The sum of the current and four lagged energy efficiency expenditure coefficients is -0.00252 with a standard error of 0.0007, which is also statistically significant at a 99% confidence level. The average annual residential electricity consumption in Massachusetts for years 2000 through 2012 was 7,533 kWh per customer. The model suggests that one dollar spent in energy efficiency expenditures per customer this year would decrease per-customer residential electricity consumption by a total of 18.98 kWh over the next four and one-half years with a 95% confidence interval of [8.2 kWh, 29.8 kWh] or 4.2±2.4 kWh per year.

Model 3 shows the results in which current electricity consumption is modeled as a function of current-year energy efficiency expenditures and those of the previous six years, as well as other factors affecting electricity consumption. The lagged energy efficiency expenditures included in the model capture the impact of the measures installed in the previous six years on current consumption, as well as the market effects. While all of the lagged energy efficiency program expenditure coefficients have the expected negative sign, the current year energy efficiency program expenditures have a positive sign. The third and the fourth year lag coefficients are statistically significant at a 99% confidence level. The coefficients of energy efficiency program expenditures are also jointly significant at a 99% confidence level ($F(7,34)=23.3$, $p=0.000$). The sum of the current and six lagged energy efficiency expenditure coefficients is -0.00363 with a standard error of 0.0006 (significant at a 99% confidence level). The model suggests that one dollar spent in energy efficiency expenditures per customer this year would decrease per-customer residential electricity consumption by a total of 27.34 kWh

[41] Model 1 was run with data for 2000 through 2012 so that the results can directly be compared with Model 2. When Model 1 is run with all available data from 1994 through 2012, the coefficient estimates remain similar.

over the next six and one-half years with a 95% confidence interval of [18.7 kWh, 35.9 kWh] or 4.2±1.3 kWh per year.

Models 4 through 6 repeat Models 1 through 3, except that Models 4 through 6 contain the adjustments to energy efficiency program expenditures by PAs and municipal utilities to account for the PA-supported program bulbs that were purchased by municipal utility customers, using the procedure described in2.5.2. The results indicate that these adjustments improve the estimates of the impact of energy efficiency program expenditures on consumption, but only slightly.

The relatively long time-series data allowed the evaluation team to test several finite distributed lag models to empirically determine the appropriate lag length. The team selected the four-lag model as the most appropriate model through a statistical significance test.[42] Among the six models whose results are shown in Table 5-2, Model 5 with a 4-year lag is the team's preferred model, because:

1. It accounts for the lagged impact of energy efficiency program expenditures on energy consumption;
2. It accounts for the leakage of PA lighting program rebate dollars to municipal utility service territories;
3. The coefficients of the first, the third, and the fourth lag are statistically significant; and
4. The coefficients of current and lagged energy efficiency expenditure variables are jointly statistically significant.

This being said, the fact that the 6-year lag model produces very similar results indicates that the fixed-effects model produces stable results across models with different lags.

In Model 5, the sum of the current and four lagged energy efficiency expenditure coefficients is -0.00259. This suggests that one dollar spent in energy efficiency expenditures per customer this year would decrease per-customer residential electricity consumption by a total of 19.5±12.2 kWh over the next four and one-half years, or 4.3±2.7 kWh per year.

Table 5-2. Residential PA-Muni Model Results with Individual Year Program Expenditures

Variable	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Price of electricity	-0.10670 (0.0675)	-0.13813** (0.0471)	-0.15799** (0.0261)	-0.10591 (0.0675)	-0.13671** (0.0487)	-0.15808** (0.0278)
Annual HDDs	0.17483** (0.0574)	0.20179** (0.0665)	0.13418+ (0.0757)	0.15674** (0.0576)	0.20555** (0.0675)	0.13842+ (0.802)

[42] In this method, the way to choose the length of a lag is to start with a long lag, test the statistical significance of the coefficient at the longest lag—the "trailing lag"—and shorten the lag by one period if one cannot reject the null hypothesis that the effect at the longest lag is zero. One continues shortening the lag until the trailing lag coefficient is statistically significant.

Variable	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Annual CDDs	0.05798** (0.0069)	0.06530** (0.0073)	0.05359** (0.0081)	0.05806** (0.0069)	0.06570** (0.0077)	0.05404** (0.0084)
Median household income	-0.14926 (0.1288)	-0.06627 (0.1154)	0.01145 (0.1162)	-0.15027 (0.1294)	-0.07936 (0.1163)	-0.00740 (0.1176)
Median home values	0.31034** (0.1085)	0.43072** (0.0943)	0.47439** (0.0873)	0.31171** (0.1081)	0.42720** (0.0929)	0.46868** (0.0859)
Percent of homes using electricity as the main heating fuel	0.59148 (0.6564)	0.55555 (0.5391)	0.63794 (0.5731)	0.59436 (0.6555)	0.55943 (0.5349)	0.64519 (0.5652)
Percent of residential new construction	1.67497* (0.7537)	1.43743+ (0.7368)	1.30648 (0.8444)	1.67464* (0.7514)	1.48740+ (0.7474)	1.35890 (0.8498)
Percent of single-family homes	0.61429 (0.6459)	0.37955 (0.5210)	0.05182 (0.4358)	0.61542 (0.6464)	0.37419 (0.5160)	0.03805 (0.4458)
Percent of renters	1.07745+ (0.5385)	0.90910+ (0.5072)	0.50765 (0.5180)	1.07797+ (0.5376)	0.89290+ (0.4982)	0.48268 (0.5056)
Percent employed	0.23166 (0.2036)	0.94816** (0.2292)	0.90222** (0.3022)	0.23357 (0.20327)	0.95572** (0.2438)	0.92497** (0.3129)
Annual residential energy efficiency program expenditures per customer in year t	-0.00014 (0.0002)	0.00038+ (0.0002)	0.00031 (0.0003)	-0.00012 (0.0002)	0.00040 (0.0002)	0.00032 (0.0003)
Annual residential energy efficiency program expenditures per customer in year t-1		-0.00046** (0.0001)	-0.00033 (0.0003)		-0.00049** (0.0001)	-0.00037 (0.0003)
Annual residential energy efficiency program expenditures per customer in year t-2		-0.00028 (0.0004)	-0.00030 (0.0003)		-0.00029 (0.0004)	-0.00032 (0.0003)
Annual residential energy efficiency program expenditures per customer in year t-3		-0.00066* (0.0003)	-0.00073** (0.0003)		-0.00068* (0.0003)	-0.00078** (0.0003)
Annual residential energy efficiency program expenditures per customer in year t-4		-0.00150** (0.0004)	-0.00128** (0.0002)		-0.00153** (0.0004)	-0.00132** (0.0003)
Annual residential energy efficiency program expenditures per customer in year t-5			-0.00110 (0.0011)			-0.00111 (0.0011)

Variable	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Annual residential energy efficiency program expenditures per customer in year t-6			-0.00019 (0.0009)			-0.00023 (0.0009)
Time Trend	-0.00077 (0.0026)	0.00029 (0.0020)	-00081 (0.0027)	-0.00086 (0.0026)	0.00051 (0.0021)	-00031 (0.0027)
Constant	4.06623* (1.5456)	0.98565 (1.5494)	0.68102 (1.6504)	4.04847* (1.5509)	1.13566 (1.5344)	0.91356 (1.6436)
Estimation method	FE	FE	FE	FE	FE	FE
Cumulative residential energy efficiency program expenditures per customer in years t-4 through t	N/A	-.00252** (0.0007)	-.00234** (0.0005)	N/A	-.00259** (0.0008)	-.00247** (0.0005)
Cumulative residential energy efficiency program expenditures per customer in years t-6 through t	N/A	N/A	-.00363** (0.0006)	N/A	N/A	-.00380** (0.0006)
Observations	438	422	414	438	422	414
Within R^2	0.64	0.69	0.71	0.64	0.69	0.71
Years included	2000-2012	2000-2012	2000-2012	2000-2012	2000-2012	2000-2012
Account for leakage of PA-supported CFLs to municipal utility customers	NO	NO	NO	YES	YES	YES
Number of utilities	35	35	35	35	35	35

Notes: In all models, the dependent variable is the natural logarithm of annual electricity consumption per customer. All independent variables are in natural log forms except the variables expressed as percentages and energy efficiency expenditures variables. Observations are weighted by PA/utility annual total residential sales.
+ p<0.10, * p<0.05, ** p<0.01

Figure 5-3 and Figure 5-4 show the smoothed relationship between energy efficiency program expenditures and residential consumption based on Model 2 of Table 5-2, along with corresponding 95% confidence bands, shown in gray. Figure 5-3 summarizes the relationship at each specific lag, while Figure 5-4 summarizes the cumulative impact of energy efficiency program expenditures over time. Similarly, Figure 5-5 and Figure 5-6 correspond to Model 3 in Table 5-2, but otherwise have the same interpretation as Figure 5-3 and Figure 5-4. The plots for Models 5 and 6 are nearly identical to those for Models 2 and 3, and are not presented here.

Figure 5-3. Association with a One-dollar Increase in EE Expenditures—Model 2

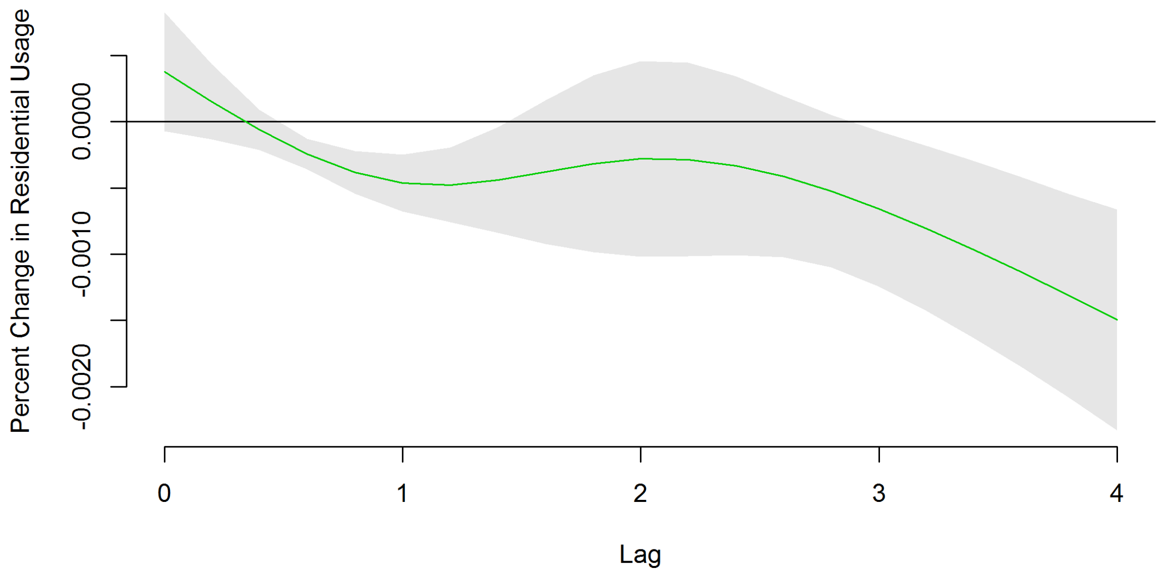


Figure 5-4. Cumulative Association with a One-dollar Increase in EE Expenditures—Model 2

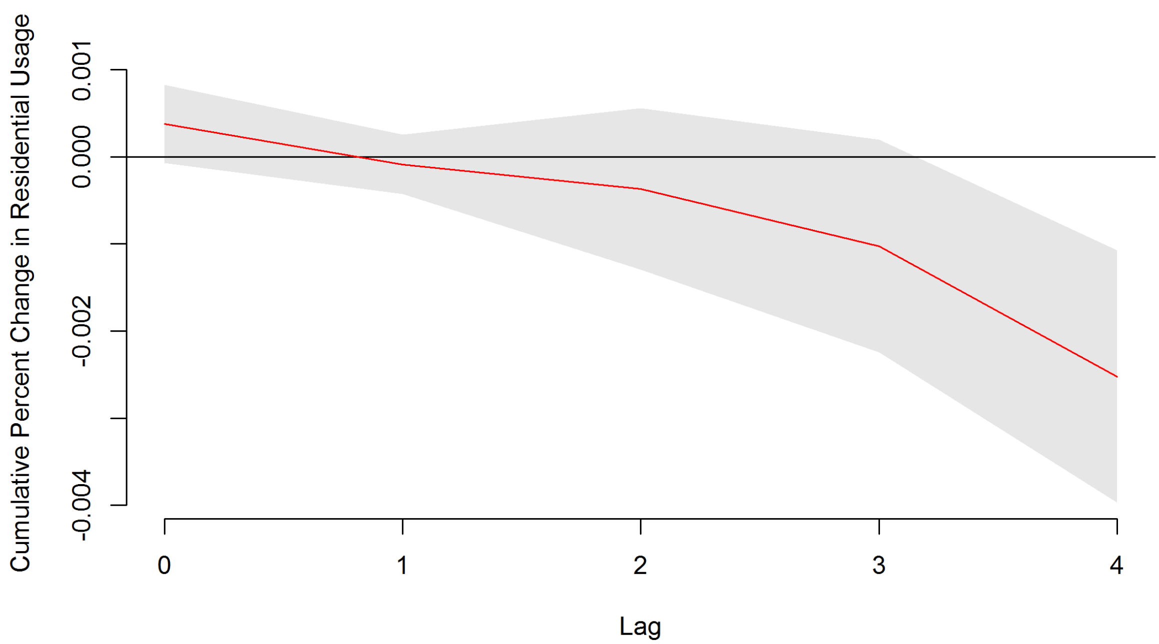


Figure 5-5. Association with a One-dollar Increase in EE Expenditures—Model 3

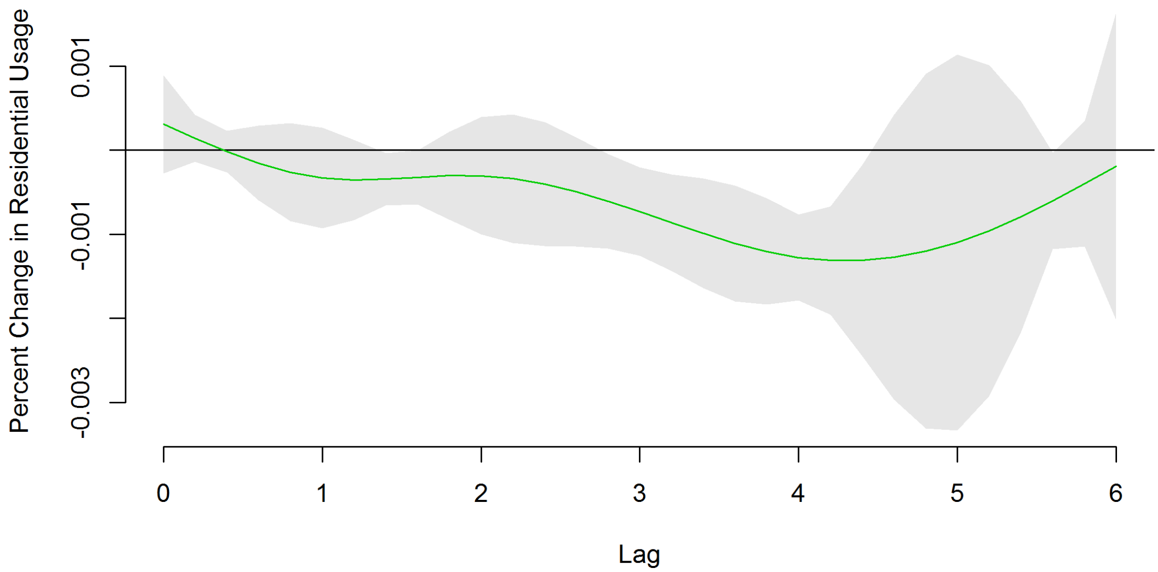
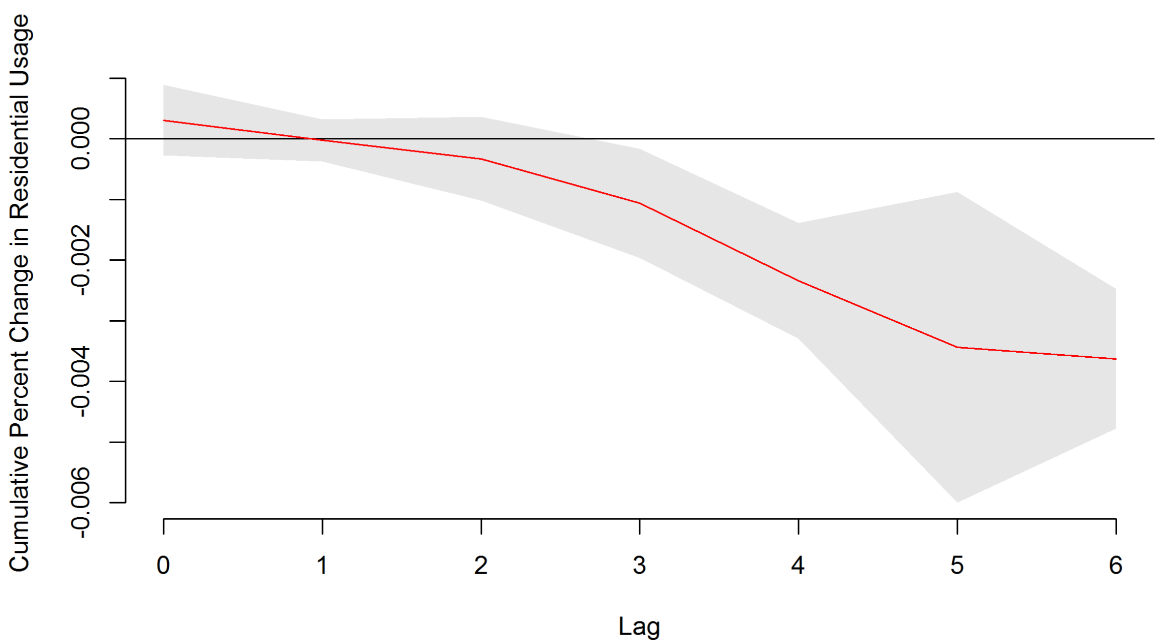


Figure 5-6. Cumulative Association with a One-dollar Increase in EE Expenditures—Model 3



The residential model results shown in Table 5-2 draw attention to an inherent limitation of macro-consumption methods. While the team was able to detect energy savings, these savings were not estimated precisely for each lagged year when individual-year lagged energy efficiency expenditure variables were included in the model. As an alternative model specification, the team cumulated the energy efficiency expenditures for the current year and up to six previous years into a single variable and reran the models. Table 5-3 shows the results of the fixed effects models with cumulated energy efficiency program expenditures. Models 10 through 12 repeat Models 7 through 9, except that Models 10 through 12 contain the adjustments to energy efficiency program expenditures by PAs and municipal utilities to account for the PA-supported program bulbs that were purchased by municipal utility customers.

The cumulated residential energy expenditures come out to be statistically significant at a 90% level in all models. The findings suggest that one dollar spent in energy efficiency expenditures per customer this year would decrease per-customer residential electricity consumption by a total of 8.5 kWh over the next three years; 13.1 kWh over the next five years; and 16.4 kWh over the next six years. The six-year estimate is slightly smaller than the corresponding estimate produced by the models with individual-year expenditure variables.

As expected, the impact of cumulated energy efficiency expenditures on current consumption increases with the number of previous years included in the cumulated sum. This reflects the importance of including sufficiently long lagged program activity in the model, as measures installed by the energy efficiency programs continue to save energy well beyond the current year.

Table 5-3. Residential PA-Muni Model Results with Cumulated Program Expenditures

Variable	Model 7	Model 8	Model 9	Model 10	Model 11	Model 12
Price of electricity	-0.11740+ (0.0690)	-0.12718+ (0.0681)	-0.14864* (0.0597)	-0.11658 (0.0694)	-0.12643+ (0.0687)	-0.14904* (0.0608)
Annual HDDs	0.16539** (0.0604)	0.16241** (0.0589)	0.13408* (0.0535)	0.16612** (0.0607)	0.16326** (0.0591)	0.13424* (0.0545)
Annual CDDs	0.05933** (0.0070)	0.05957** (0.0067)	0.05676** (0.0069)	0.05930** (0.0070)	0.05955** (0.0068)	0.05666** (0.0069)
Median household income	-0.11887 (0.1235)	-0.09395 (0.1180)	-0.05191 (0.1144)	-0.12122 (0.1236)	-0.09723 (0.1182)	-0.05615 (0.1156)
Median home values	0.30965** (0.1098)	0.31258** (0.1113)	0.33180** (0.1057)	0.30914** (0.1090)	0.31356** (0.1102)	0.32845** (0.1043)
Percent of homes using electricity as the main heating fuel	0.53969 (0.6328)	0.52695 (0.6158)	0.51898 (0.6110)	0.54400 (0.6325)	0.53083 (0.6140)	0.52412 (0.6059)
Percent of residential new construction	1.74507* (0.8079)	1.74538* (0.8138)	1.63163* (0.7724)	1.75613* (0.8122)	1.76880* (0.8223)	1.66921* (0.7825)
Percent of single-family homes	0.64128 (0.6218)	0.61853 (0.5957)	0.47997 (0.5479)	0.64757 (0.6234)	0.62836 (0.5336)	0.48614 (0.5422)
Percent of renters	1.16691* (0.5279)	1.17613* (0.5333)	1.04400+ (0.5307)	1.17327* (0.5287)	1.18840* (0.5336)	1.05727+ (0.5297)

Variable	Model 7	Model 8	Model 9	Model 10	Model 11	Model 12
Percent employed	0.33024 (0.2539)	0.41445 (0.2692)	0.49667* (0.2164)	0.32709 (0.2536)	0.41275 (0.2710)	0.50391* (0.2196)
Cumulated residential energy efficiency program expenditures per customer in years t-3 to year t	-0.00066 (0.0005)			-0.00066 (0.0005)		
Cumulated residential energy efficiency program expenditures per customer in years t-4 through year t		-0.00115+ (0.0006)			-0.00118 (0.0007)	
Cumulated residential energy efficiency program expenditures per customer in years t-6 through t			-0.00210** (0.0006)			-0.00222** (0.0007)
Time Trend	0.00062 (0.0021)	0.00158 (0.0020)	0.00239 (0.0020)	0.00063 (0.0021)	0.00169 (0.0020)	0.00276 (0.0020)
Constant	3.69559* (1.5964)	3.32586+ (1.6445)	3.02751* (1.4771)	3.71613* (1.5978)	3.36422* (1.6431)	3.09792* (1.4826)
Estimation method	FE	FE	FE	FE	FE	FE
Observations	426	422	414	426	422	414
Within R^2	0.65	0.66	0.67	0.65	0.66	0.67
Years included	2000-2012	2000-2012	2000-2012	2000-2012	2000-2012	2000-2012
Account for leakage of PA-supported CFLs to municipal utility customers	NO	NO	NO	YES	YES	YES
Number of utilities	35	35	35	35	35	35

Notes: In all models, the dependent variable is the natural logarithm of annual electricity consumption per customer. All independent variables are in natural log forms except the variables expressed as percentages and energy efficiency expenditures variables. Observations are weighted by PA/utility annual total residential sales.
+ p<0.10, * p<0.05, ** p<0.01

Finally, the team ran the residential models without any additional control variables and compared the results with those that had these controls. Table 5-4 shows the results for four fixed-effects models with no additional controls except a time trend. The results from these models were different than those for the corresponding models that controlled for additional time-varying characteristics. This suggests that it was important to control for such variables in the model in order to isolate the impact of the energy efficiency program activity on energy consumption from other natural and policy changes.

Table 5-4. Residential PA-Muni Model Results without Additional Controls

Variable	Model 13	Model 14	Model 15	Model 16
Annual residential energy efficiency program expenditures per customer in year t	-0.00041** (0.0001)			
Cumulated residential energy efficiency program expenditures per customer in years t-3 to year t		-0.00041 (0.0003)		
Cumulated residential energy efficiency program expenditures per customer in years t-5 through year t			-0.0042 (0.0003)	
Cumulated residential energy efficiency program expenditures per customer in years t-6 through t				-0.00078+ (0.0004)
Time Trend	0.00604** (0.0006)	0.00552** (0.0007)	0.00545** (0.0008)	0.00594** (0.0009)
Constant	8.83133** (0.0091)	8.83780** (0.0089)	8.83863** (0.0089)	8.83805** (0.0088)
Estimation method	FE	FE	FE	FE
Observations	439	427	419	425
Within R^2	0.20	0.19	0.19	0.19
Years included	2000-2012	2000-2012	2001-2012	2000-2012
Account for leakage of PA-supported CFLs to municipal utility customers	NO	NO	NO	NO
Number of utilities	34	34	34	34

Notes: In all models, the dependent variable is the natural logarithm of annual electricity consumption per customer. Observations are weighted by PA/utility annual total residential sales. + $p<0.10$, * $p<0.05$, ** $p<0.01$.

Among all different models tested, the team's preferred model specification is Model 5 in Table 5-2, which is a fixed-effects panel regression model in which current electricity consumption is modeled as a function of current-year energy efficiency expenditures and those of the previous four years, as well as other factors affecting electricity consumption. In this model, the coefficients of the current and lagged energy efficiency expenditure variables were jointly statistically significant in addition to the most of the individual-year energy efficiency expenditure coefficients. The model accounts for the lagged impact of energy efficiency program expenditures on energy consumption and presents the impact of individual year lags. Finally, this model accounts for the leakage of PA lighting program rebate dollars to municipal utility service territories.

5.2.1 A comparison of residential top-down and bottom-up saving estimates

Table 5-5 provides a comparison of annual savings estimates from the first and second families of residential top-down models with lags.[43] The top-down models accounted for the leakage of upstream PA lighting program rebate dollars to municipal utility service territories. The table shows the annual net savings estimates and the corresponding lower and upper bounds of the 90% confidence intervals. The table also expresses top-down estimated net savings as a percent of the annual bottom-up net saving estimates to provide a top-down to bottom-up estimate ratio. The four-and six-lag models account for the impact of up to four and six previous years' programmatic activity on the current year's consumption, respectively. The four-lag model, which provided the best statistical fit to the data,[44] shows a top-down to bottom-up ratio of 187%, but the 90% confidence interval ranges from 92% to 282%. When the individual year expenditures are cumulated into a single variable, this ratio from the four-lag residential model reduces to 85%, with a confidence interval ranging from 2% to 168% of annual bottom up savings. The fact that four-year and six-year lag models produce comparable results suggests that the results are stable across models with different lag lengths. However, further research is needed to understand the differences in estimates from the individual-year and cumulated program expenditure models.

Table 5-5. PA-Muni Residential Top-down and Bottom-up Net Savings Comparisons, 2003–2012

Model Family	#Lags	Top-down Annual Net Saving Estimates (GWh)			Top-down Annual Net Saving Estimates (% of Net Bottom-up Estimates)[45]		
		Lower Bound	Point Estimate	Upper Bound	Lower Bound	Point Estimate	Upper Bound
Individual Year	Four	1,851	3,762	5,674	92%	187%	282%
Cumulated	Four	41	1,714	3,387	2%	85%	168%
Individual Year	Six	2,829	3,821	4,814	141%	190%	240%
Cumulated	Six	1,075	2,233	3,391	53%	111%	169%

[43] A model with no lags was tested. Most of the residential energy savings would occur at the end of a calendar year. This means that the savings in any calendar year would not line up with the usage of that year. The residential upstream lighting program generates a large proportion of residential savings. The first several months of each year are slow as new Memorandum of Understanding Agreements are put in place. The result is that more than one-half and, generally, two-thirds of savings or more are in the latter half of the year. So the consumption impacts of program expenditures on average would be part of the referenced calendar year and the first part of the following. Lag models can accommodate this mismatch while no-lag models cannot.

[44] While the estimate for the fourth lag was statistically significant in both the four- and six-lag models, the estimates for the fifth and the sixth lags in the six-lag model were not statistically significant.

[45] The source of residential electric program reported net savings and expenditures is Massachusetts Division of Energy Resources' (DOER's) PARIS database. Annual net savings claims from 2003 to 2012 are cumulated and then divided by 10 (the number of years) to compute an average annual bottom-up estimate. The cumulative model estimate from the top-down individual-year models was divided by the number of lags included in the model plus 0.5 (to account for the partial-year effect of the current-year expenditures) to arrive at an average annual top-down estimate.

5.3 PA-MUNI TOP-DOWN MODEL RESULTS: C&I MODEL

The team estimated the models with and without the lagged energy expenditure variables using the software package *Stata*. The estimation sample was restricted to Massachusetts investor-owned electric utilities (National Grid, NSTAR, Unitil, and WMECo)[46] and 29 municipal utilities that had C&I energy efficiency program data available from 1996 through 2012.[47] The team indexed energy consumption and energy efficiency program expenditures to the number of customers, the number of establishments, and the number of employees in a PA/utility service territory for a given year. In all models, observations were weighted by PA/utility annual total C&I sales (in GWh). Table 5-6 shows summary statistics for key model variables for PAs and municipal utilities included in the estimation sample for the years 2002 through 2012.[48] The statistics are weighted by PA/utility total C&I sales. For PAs and munis combined, the average annual electricity consumption per employee was 12,560 kWh and the average annual energy efficiency program expenditure per employee was $43.47. The municipal utilities, on average, had slightly higher per-employee electricity consumption (12,959 kWh) and significantly lower per-customer energy efficiency expenditures ($1.58) than those of PAs (12,186 and $47.53, respectively).

Table 5-6. C&I PA-Muni Model Summary Statistics, Weighted, 2002–2012

Variable	All	PA	Municipal
C&I annual electricity consumption (kWh) per customer	99,549 (32,050)	97,110 (27,495)	124,972 (56,818)
C&I annual electricity consumption (kWh) per establishment	217,599 (59,809)	215,471 (54,528)	239,780 (98,276)
C&I annual electricity consumption (kWh) per employee	12,560 (4,332)	12,186 (1,707)	12,959 (1,218)
Price of electricity (cents per kWh in 2012 $)	11.8 (2.9)	11.7 (3.0)	12.2 (1.8)
Annual HDDs (Base 65)	5,999 (482)	5,980 (478)	6,193 (514)

[46] EIA-861 energy consumption and price data were available separately for former National Grid (Massachusetts Electric Co. and Nantucket Electric Co.) and NSTAR (Boston Edison Co., Cambridge Electric Light Co., and Commonwealth Electric Co.) companies. The team included these as five separate companies in the modeling. The annual energy efficiency program expenditures per customer by National Grid and NSTAR were distributed equally among their former companies. Finally, the energy program expenditures by Cape Light Compact were counted under the Commonwealth Electric Co.

[47] Only 4 of these 29 municipal utilities had any C&I programs.

[48] Consistent data on model variables are available starting in 1996. However, the inclusion of lagged values of energy efficiency program expenditures (up to 6 lags) requires that the analysis start in 2002 because that is the first year in which the data for the energy efficiency expenditures for the previous six years are available.

Variable	All	PA	Municipal
Annual CDDs (Base 70)	303 (89)	305 (89)	283 (94)
Mean annual wage (in 2012 $)	53,646 (12,898)	54,107 (12,605)	48,838 (15,376)
Percent employed	93.5 (1.7)	93.4 (1.7)	94.3 (1.7)
C&I annual new construction per customer (in sq. ft.)	86.5 (87.4)	88.1 (80.9)	70.5 (139.7)
C&I annual new construction per establishment (in sq. ft.)	183.0 (164.5)	188.8 (158.5)	122.5 (214.0)
C&I annual new construction per employee (in sq. ft.)	11.5 (11.3)	11.8 (10.8)	8.1 (15.4)
Annual C&I energy efficiency program expenditures per customer ($)	313.30 (195.6)	342.47 (179.6)	11.94 (51.3)
Annual C&I energy efficiency program expenditures per establishment ($)	710.80 (463.4)	777.2 (431.5)	24.53 (112.4)
Annual C&I energy efficiency program expenditures per employee ($)	43.47 (34.9)	47.53 (34.0)	1.58 (7.4)

Notes: All values are averages across the 36 utilities (7 PAs and 29 municipal utilities in the years between 2002 and 2012). Standard deviations are given in parentheses. The statistics are weighted by PA/utility annual total C&I sales.

Our model-building criteria prioritized theoretical relevance over observed explanatory power. Therefore, we kept the same explanatory variables in all models even if some explanatory variables did not come out to be statistically significant in some models. Table 5-7 shows the results from four different fixed-effects C&I models with different lag lengths. In each model, the dependent variable is the annual average C&I electricity consumption per employee. The results (not shown) were similar when consumption was defined on a per-customer or per-establishment basis. A summary of the results from each model is described below.

Model 1 shows the results in which current electricity consumption is modeled as a function of current-year energy efficiency expenditures and other factors affecting electricity consumption.[49] As expected, cooling degree days is a significant factor in C&I electricity consumption. A 10% increase in cooling degree days would increase electricity consumption per employee by about 0.6%. The coefficient -0.00029 of annual C&I energy efficiency program expenditures per customer in year *t,* which is statistically significant at a 90% confidence level, suggest that a one-dollar increase in C&I program expenditures per

[49] Model 1 was run with data for 2002 through 2012 so that the results can directly be compared with other models containing lagged program expenditures. When Model 1 is run with all available data from 1996 through 2012, the coefficient estimates remain similar.

employee in a given year would decrease the per-employee electricity consumption by about 0.029% in that year. The average annual C&I electricity consumption per employee in Massachusetts for years 2002 through 2012 was 12,560 kWh. This suggests that one dollar spent in energy efficiency expenditures per employee in a given year would decrease per-employee C&I electricity consumption by a total of 3.6 kWh in that year with a 95% confidence interval of [0.5 kWh, 6.8 kWh]. This model does not capture the lagged impact of the energy efficiency programs on energy consumption. In addition, the impact of current program expenditures on current consumption could be twice as large if expenditures were distributed uniformly in a given year because, in that case, each dollar of current-year expenditures would affect only half of current-year consumption.

Model 2 adds the previous three years' energy efficiency expenditures to the specification. The lagged energy efficiency expenditures included in the model capture the impact of the measures installed in the previous three years on current consumption, as well as the market effects. While the current and lagged energy efficiency program expenditure coefficients all have the expected negative sign, only the first and the third year lag coefficients are statistically significant at a 95% confidence level. The coefficients of energy efficiency program expenditures are also jointly significant at a 99% confidence level ($F(4,35)=6.63$, $p=0.0004$). The sum of the current and three lagged energy efficiency expenditure coefficients is -0.00091 with a standard error of 0.0004, which is also statistically significant at a 95% confidence level. The model suggests that one dollar spent in energy efficiency expenditures per employee this year would decrease per-employee C&I electricity consumption by a total of 11.4 kWh over the next three and one-half years with a 95% confidence interval of [2.55 kWh, 20.3 kWh] or 3.2±2.54 kWh per year.

Model 3 adds the previous four years' energy efficiency expenditures to the specification. While the current and previous three years' energy efficiency program expenditure coefficients all have the expected negative sign, the fourth year's coefficient is positive. Similar to Model 2, the first and the third lag coefficients are negative and statistically significant. In addition, the sum of the current and four lagged energy efficiency expenditure coefficients is not statistically significant.

Model 4 adds the previous six years' energy efficiency expenditures to the specification. While the current and previous three years' energy efficiency program expenditure coefficients all have the expected negative sign, the fourth, fifth, and sixth years' coefficients are positive. Moreover, the fifth and sixth year coefficients are statistically significant. The sum of the current and four lagged energy efficiency program expenditure coefficients is also positive but not statistically significant.

In general, the C&I model results with individual year program expenditure variables were less consistent than the corresponding residential models. Moreover, the R^2 values for the C&I models were lower than those for the residential models, suggesting that the explanatory variables included in the models explain less variation in electricity consumption in the C&I sector than the residential sector.

Table 5-7. C&I PA-Muni Model Results with Individual Year Program Expenditures

Variable	Model 1	Model 2	Model 3	Model 4
Price of electricity	-0.00886 (0.0294)	-0.003644 (0.0363)	-0.03208 (0.0404)	-0.01523 (0.0401)
Annual HDDs	-0.00904 (0.660)	0.00553 (0.899)	0.00758 (0.0921)	0.01743 (0.780)
Annual CDDs	0.05498** (0.0089)	0.05274** (0.0116)	0.05203** (0.0113)	0.04864** (0.0078)
Mean annual wage (in 2012 $)	-0.02201 (0.0924)	-0.06962 (0.1393)	-0.06342 (0.1441)	-0.04237 (0.1332)
Percent employed	-0.51223+ (0.3170)	-0.28939 (0.3589)	-0.25874 (0.3600)	-0.40118 (0.3630)
C&I annual new construction per employee (in sq. ft.)	0.00046 (0.0004)	0.00028 (0.0003)	0.00026 (0.0004)	0.00022 (0.0003)
Annual C&I energy efficiency program expenditures per employee in year t	-0.00029+ (0.0002)	-0.00018 (0.0001)	-0.00018 (0.0001)	-0.00017 (0.0001)
Annual C&I energy efficiency program expenditures per employee in year t-1		-0.00025* (0.0001)	-0.00024* (0.0001)	-0.00018+ (0.0002)
Annual C&I energy efficiency program expenditures per employee in year t-2		-0.00011 (0.0002)	-0.00009 (0.0002)	-0.00008 (0.0002)
Annual C&I energy efficiency program expenditures per employee in year t-3		-0.00036** (0.0001)	-0.00033* (0.0001)	-0.00026+ (0.0002)
Annual C&I energy efficiency program expenditures per employee in year t-4			0.00010 (0.0001)	0.00011 (0.0002)
Annual C&I energy efficiency program expenditures per employee in year t-5				0.00044** (0.0001)
Annual C&I energy efficiency program expenditures per employee in year t-6				0.00043* (0.0002)
Time Trend	-0.00851 (0.0053)	-0.00645 (0.0054)	-0.00671 (0.0056)	-0.01127+ (0.0064)
Constant	-11.68837 (15.0934)	-11.96775 (15.9607)	-11.96775 (15.9837)	-2.29631 (12.8869)
Estimation Method	FE	FE	FE	FE
Cumulative C&I energy efficiency program expenditures per customer in years t-3 through t		-0.00091* (0.0004)		

Variable	Model 1	Model 2	Model 3	Model 4
Cumulative C&I energy efficiency program expenditures per customer in years t-4 through t			-0.00075 (0.0005)	
Cumulative C&I energy efficiency program expenditures per customer in years t-6 through t				0.00029 (0.0007)
Observations	379	379	379	379
Within R^2	0.39	0.40	0.40	0.43
Years Included	2002-2012	2002-2012	2002-2012	2002-2012
Number of Utilities	36	36	36	36

Notes: In all models, the dependent variable is the natural logarithm of annual electricity consumption per employee. All independent variables are in natural log forms except the variables expressed as percentages, C&I new construction, and energy efficiency expenditures variables. Observations are weighted by PA/utility annual total C&I sales. + p<0.10, * p<0.05, ** p<0.01

Figure 5-7 and Figure 5-8 show the smoothed relationship between energy efficiency program expenditures and C&I consumption based on Model 2 of Table 5-7, along with corresponding 95% confidence bands, shown in gray. Figure 5-7 summarizes the relationship at each specific lag, while Figure 5-8 summarizes the cumulative impact of energy efficiency program expenditures over time. Figure 5-9 and Figure 5-10 correspond to Model 4 in Table 5-7, but otherwise have the same interpretation as Figure 5-7 and Figure 5-8.

Figure 5-7. Association with a One-dollar Increase in Program Expenditures—Model 2

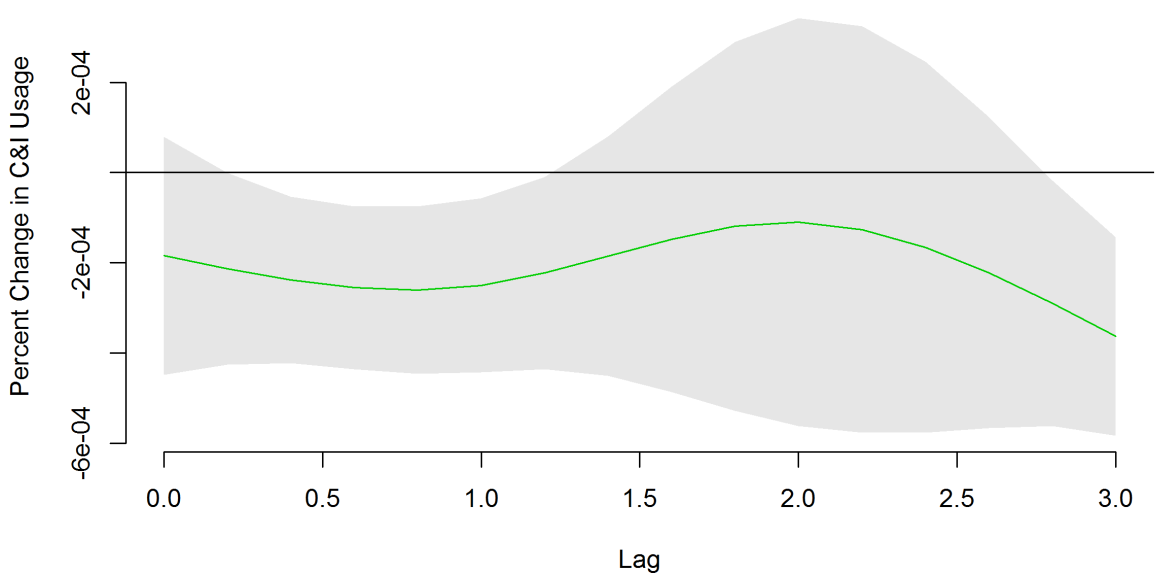


Figure 5-8. Cumulative Association with a One-dollar Increase in Program Expenditures—Model 2

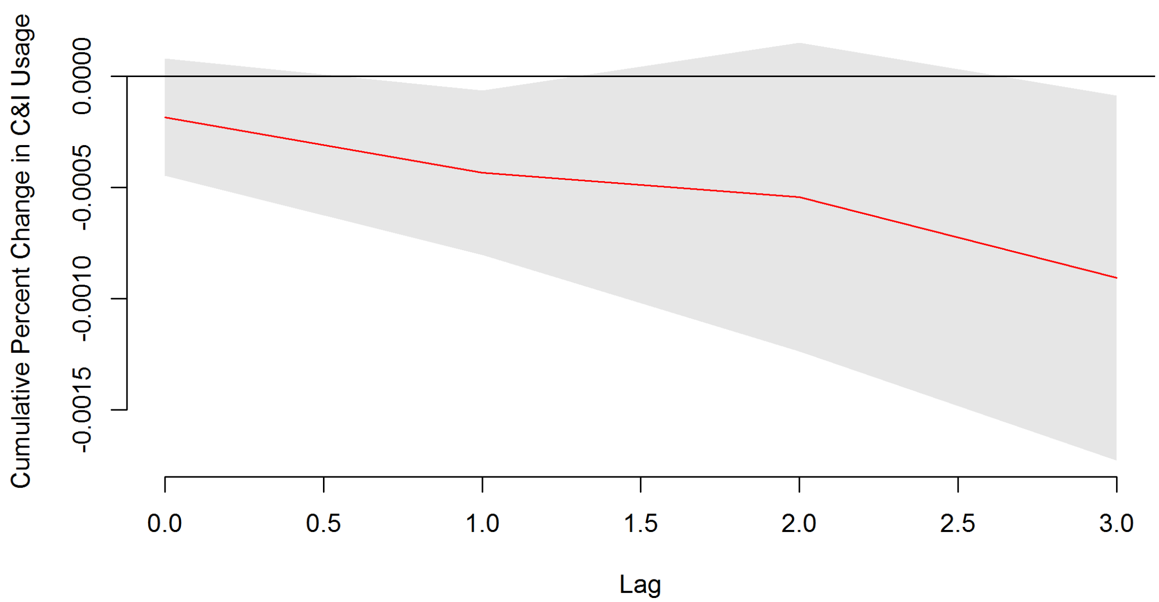


Figure 5-9. Association with a One-dollar Increase in Program Expenditures—Model 3

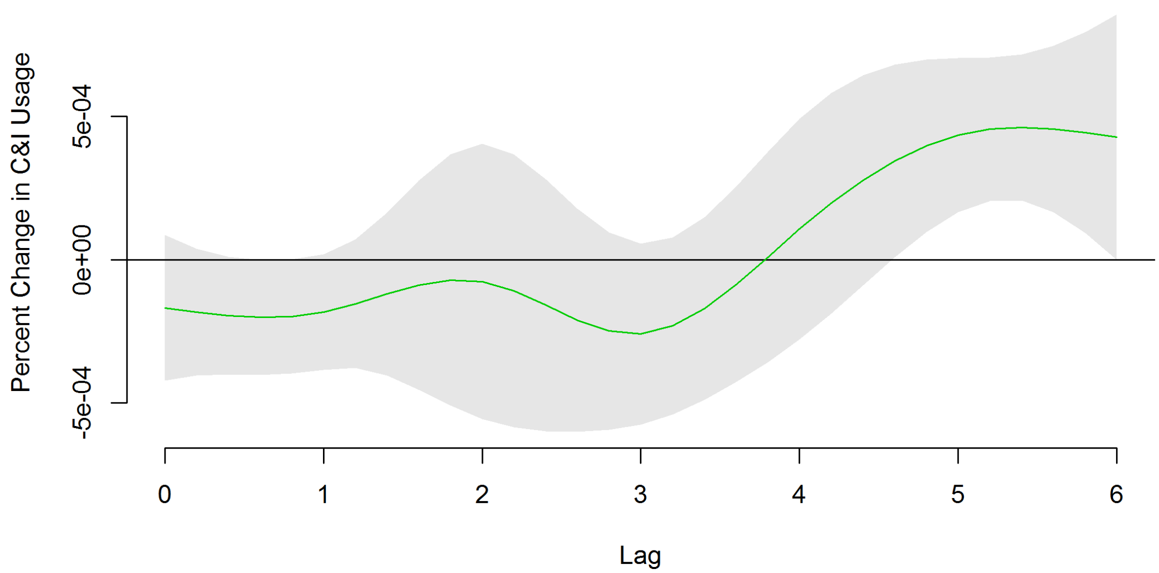
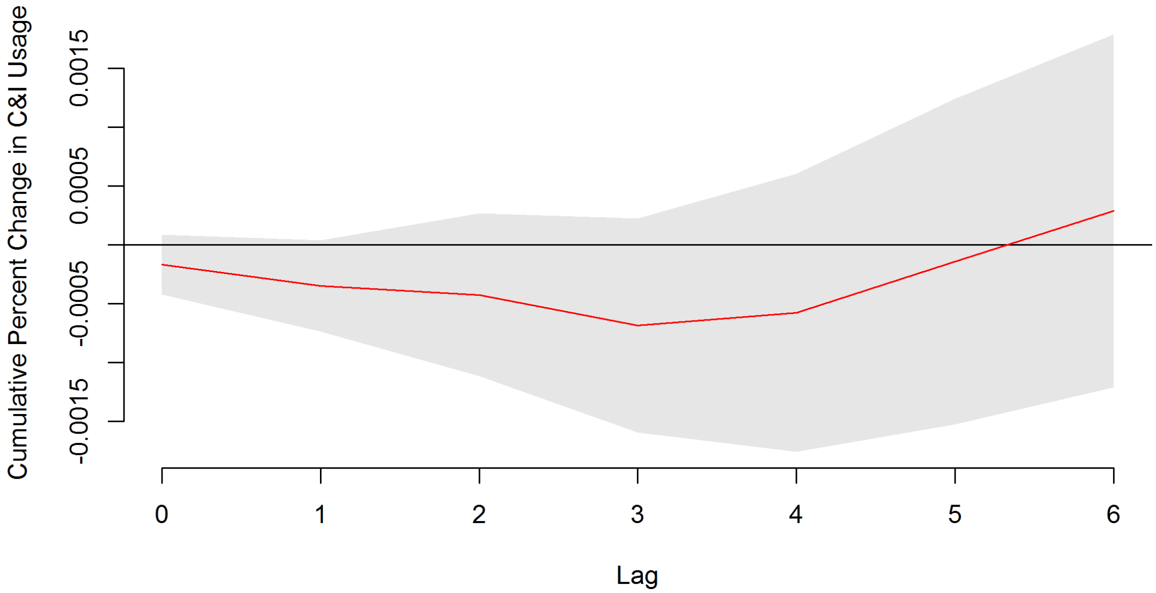


Figure 5-10. Association with a One-dollar Increase in Program Expenditures—Model 3



As an alternative model specification, the team summed the C&I energy efficiency program expenditures for the current year and up to six previous years into a single variable and reran the models. Table 5-8 shows the results for select models with cumulated energy efficiency program expenditures.

The cumulated C&I energy expenditures come out to be statistically significant at a 90% level in the models that include cumulated energy efficiency expenditures for up to four previous years but not beyond that. The results for Model 5 through 7 are comparable to those for Model 2 through 4 in Table 5-7. Model 5, which included the cumulated C&I program expenditures for the current and the previous three years, appears to provide the best results. The coefficient for the cumulated C&I energy efficiency program expenditures per employee in years t-3 to t is -0.00086 with a standard error of 0.0003. This suggests that one dollar spent in energy efficiency expenditures per employee this year would decrease per-employee C&I electricity consumption by a total of 10.8 kWh over the next three and one-half years with a 95% confidence interval of [3.6 kWh, 17.9 kWh] or 3.1±2.0 kWh per year. These are very similar to the findings from Model 2 with individual year program expenditures.

Table 5-8. C&I PA-Muni Model Results with Cumulated Program Expenditures

Variable	Model 5	Model 6	Model 7
Price of electricity	-0.03428 (0.0354)	-0.03453 (0.0377)	-0.00273 (0.0417)
Annual HDDs	-0.00499 (0.0760)	-0.00192 (0.0762)	0.04173 (0.0829)
Annual CDDs	0.05310** (0.0095)	0.05558** (0.0089)	0.05575** (0.0072)
Mean annual wage (in 2012 $)	-0.06154 (0.1427)	-0.05209 (0.1464)	-0.03995 (0.1470)
Percent employed	-0.24794 (0.2942)	-0.36732 (0.3236)	-0.53052 (0.3156)
C&I annual new construction per employee (in sq. ft.)	0.00029 (0.0003)	0.00038 (0.0003)	0.00042 (0.0004)
Cumulated C&I energy efficiency program expenditures per employee in years t-3 to t	-0.00086** (0.0004)		
Cumulated C&I energy efficiency program expenditures per employee in years t-4 through t		-0.00093+ (0.0005)	
Cumulated C&I energy efficiency program expenditures per employee in years t-6 through t			0.00014 (0.0009)
Time Trend	-0.00629+ (0.0054)	-0.00673 (0.0057)	-0.01109 (0.0069)
Constant	-11.60873 (15.6961)	-11.38312 (15.4941)	-10.95285 (13.1589)
Estimation Method	FE	FE	FE
Observations	379	379	379
Within R^2	0.40	0.39	0.38
Years Included	2002-2012	2002-2012	2002-2012
Number of Utilities	36	36	36

Notes: In all models, the dependent variable is the natural logarithm of annual electricity consumption per employee. All independent variables are in natural log forms except the variables expressed as percentages, C&I new construction, and energy efficiency expenditures variables. Observations are weighted by PA/utility annual total C&I sales. + p<0.10, * p<0.05, ** p<0.01

The team also ran the C&I models without any additional control variables and compared the results with those that had these controls. Table 5-9 shows the results for four fixed-effects models with no additional controls except a time trend. The results from these models were somewhat different than those for the corresponding models that controlled for additional time-varying characteristics. This suggests that it was important to control for such variables in the model in order to isolate the impact of the energy efficiency program activity on energy consumption from other natural and policy changes.

Table 5-9. C&I PA-Muni Model Results without Additional Controls

Variable	Model 8	Model 9	Model 10
Cumulated C&I energy efficiency program expenditures per employee in years t-3 to t	-0.00061 (0.0004)		
Cumulated C&I energy efficiency program expenditures per employee in years t-4 through t		-0.00050 (0.0005)	
Cumulated C&I energy efficiency program expenditures per employee in years t-6 through t			0.00018 (0.0006)
Time Trend	-0.00148 (0.0018)	-0.00194 (0.0017)	-0.00303** (0.0018)
Constant	9.46467** (0.0318)	9.46825** (0.0342)	9.46332** (0.0359)
Estimation Method	FE	FE	FE
Observations	379	379	379
Within R^2	0.06	0.05	0.04
Years Included	2002-2012	2002-2012	2002-2012
Number of Utilities	36	36	36

Notes: In all models, the dependent variable is the natural logarithm of annual electricity consumption per employee. All independent variables are in natural log forms except the variables expressed as percentages, C&I new construction, and energy efficiency expenditures variables. Observations are weighted by PA/utility annual total C&I sales. + p<0.10, * p<0.05, ** p<0.01

Among all C&I models tested, Model 2 and Model 5, which included current year and the previous three years' energy program expenditures, provided the best results. When the previous fifth year and sixth year's expenditures were added to the specification, the model results were no longer statistically viable.

In general, C&I model results were less stable across different model specifications than the residential ones. Moreover, the explanatory variables added to the models explained less variation in energy consumption in the C&I sector than the residential sector. Finally, the estimated impact of one dollar spent for the C&I energy efficiency programs was somewhat smaller than that for the residential sector, but the difference was not statistically significant.

5.3.1 A comparison of C&I top-down and bottom-up saving estimates

The top-down estimation methods employed by the researchers in this report provide an estimate of net energy savings as a result of ratepayer-funded energy efficiency programs in Massachusetts, which can be used in concert with other estimates to arrive at a picture of the true impact of such programs. As such, top-down saving estimates are not intended to replace the traditional bottom-up estimates, but can help validate them.

Table 5-10 provides a comparison of annual savings estimates from the first family of C&I top-down models. The three-lag model, which provided the best statistical fit to the data,[50] shows a top-down to bottom-up estimate ratio of 101%, and the 90% confidence interval on the top-down savings estimates ranges from 28% to 174%. When the individual-year expenditures are aggregated into a single variable, this ratio for the three-lag C&I model reduces to 95%, with a confidence interval on ranging from 22% to 168% of annual bottom-up savings. The fact that C&I models with different lag lengths produce different results indicates that the C&I model results were less robust compared to residential. While this is expected given that consumption in the C&I sector is more volatile than that in the residential sector, and given that the customer base is more heterogeneous, further research is needed to understand the high degree of volatility of the results with respect to lag length.

Table 5-10. PA-Muni C&I Top-down and Bottom-up Net Savings Comparisons, 2003–2012

		Top-down Annual Net Saving Estimates (GWh)			Top-down Annual Net Saving Estimates (% of Net Bottom-up Estimates)[51]		
Model Family	#Lags	Lower Bound	Point Estimate	Upper Bound	Lower Bound	Point Estimate	Upper Bound
Individual Year	Three	925	3,342	5,758	28%	101%	174%
Cumulated	Three	742	3,158	5,574	22%	95%	168%
Individual Year	Four	-207	2,142	4,491	-6%	65%	136%
Cumulated	Four	307	2,656	5,005	9%	80%	151%
Individual Year	Six	-2,850	-573	1,703	-86%	-17%	51%
Cumulated	Six	-3,204	-277	2,651	-97%	-8%	80%

[50] While the estimate for the third lag was statistically significant in other lagged models, the estimates for the fourth lag was not. The fifth and the sixth lags in the six-lag model were statistically significant but they had the opposite (positive) sign.

[51] The source of C&I electric program reported net savings and expenditures is Massachusetts Division of Energy Resources' (DOER's) PARIS database. Annual net savings claims from 2003 to 2012 are cumulated and then divided by 10 (the number of years) to compute an average annual bottom-up estimate. The cumulative model estimate from the top-down individual-year models was divided by the number of lags included in the model plus 0.5 (to account for the partial-year effect of the current-year expenditures) to arrive at an average annual top-down estimate.

6. PA DATA TOP-DOWN MODEL PILOT STUDY

In this section, we discuss the development of the PA Data pilot study top-down models. First we describe the general form of these macro-consumption models, then discuss details of the C&I PA Data models.

The consumption history and the program tracking data for this pilot study came from the C&I customer database. However, as of this pilot study, this database currently included just three years of data. Model estimation in this current evaluation period did not attempt to estimate net savings using these models because of this data limitation. Rather, the pilot study in this evaluation cycle investigated whether the data provided evidence of a sufficient signal between programmatic activity and consumption to warrant further study.

For the PA Data pilot study, the evaluation team developed and estimated a set of statewide macro-economic time series cross sectional (TSCS) consumption models for C&I sectors.[52] We used these models to explore whether we could detect a sufficient signal between changes to aggregate consumption per unit (e.g., gross domestic product or population) over time (i.e., delta consumption), for a geographic region (e.g., county or towns), as a function of delta changes in programmatic activity and economic conditions.[53] Because this modeling approach used PA billing data for the dependent variable and was restricted to PA territories, the model results compared consumption for regions with higher and lower program activity levels within PA territories. The advantage of this approach is that the use of PA data provided detailed information regarding program activity level and the ability to aggregate by whatever dimensions are of interest, such as PA territory, cities, or towns. The disadvantage to this approach is that it lacked a “no-program” situation; thus, spillover and/or program self-selection effects may influence the results.

6.1 PA DATA MODEL SPECIFICATION

6.1.1 General form of the PA Data top-down model study

As defined in Equation 6-1 below, the dependent variable for the PA Data models was weather-normalized annual energy consumption per unit (i.e., per square foot, per capita, per employee, or per household). We performed weather normalization at the premise level to capture individual heating and cooling responses to weather. We also considered aggregating the data at different geographical levels depending upon the availability of other data required by the model(s), and the PAs’ treatment of geographies in program design, implementation, marketing, and reporting. Moving to larger geographies allowed the model to better capture spillover effects—one of the key motivations behind top-down modeling—at the expense of reducing the amount of available data. We coordinated with the MA C&I research team to obtain customer-level consumption and tracking data for use in separate commercial, industrial, and residential models, by fuel type (electricity & natural gas).

[52] The residential PA Data models will be developed in a later stage of this multi-year project.
[53] Cross sectional time series models include variables that describe attributes of each observation that are fixed (constant), as well as those that vary over time, as well as by cross section.

Equation 6-1. PA Data Macro-Consumption Model: General Form

$$\delta NAC_{tsgf} = \beta_1(\delta A)_{tsgf} + \beta_2(\delta P)_{tfg} + \beta_3(\delta E)_{tsg} + \varepsilon_{sgft}$$

Where:

NAC_{tsgf}	=	Normalized annual consumption per household (residential), per square foot by building type (commercial), or per employee by industry in year (t), sector (s) in geographic region (g), and fuel type (f)[54]
A_{tsgf}	=	Vector of programmatic activity variables in year (t), sector (s) in geographic region (g), and fuel type (f)
E_{tsg}	=	Vector of economic variables for sector (s) in geographic region (g) during year (t)
P_{tfa}	=	Vector of average (real) electric, gas, and heating oil prices during time period (t) in territory (g)
β_1	=	Effect of programmatic activity on NAC
β_2	=	Effect of energy prices on NAC
β_3	=	Effect of economic conditions on NAC
ε_{im}	=	Regression residual

As shown in Equation 6-1, the delta symbol (δ) precedes NAC, the dependent variable, and each model variable on the right-hand-side to denote that equations were estimated as first-differences model specification(s). In this first year of the PA Data pilot study, we explored both differenced and non-differenced forms of top-down models. We also included a set of constant terms representing fixed-effects for the cross sections, in a time series regression model in which the dependent variable of the model was normalized annual consumption aggregated across geographic region(s) and sectors (C&I) within the Massachusetts PA territories.[55]

We tested a range of programmatic variables to measure programmatic influences, such as program tracked energy savings estimates, total program ($) expenditures, and models that split program expenditures and savings into upstream and downstream as well as lighting and non-lighting programs. Because impacts associated with expenditures and ex-ante savings are likely to be collinear, the models included either ex-ante savings or expenditures, and did not include both terms at the same time. We also tested alternative lead-lag relationships by

[54] The evaluation team will examine other segmentation criteria, such as segmenting customers by consumption patterns and estimating separate models for each consumption group.

[55] This set of fixed effects coefficients (i.e., constant terms) are not displayed in Equation 6-1 and Equation 6-2.

building lags (variables that reflect the value of a metric in previous periods) into the program variables that capture time differentials between the dates when program data and energy efficiency (EE) measures are usually booked and/or fully installed and when the impacts on consumption gain visibility in the metered consumption (billing) data.

We scaled the program variables to a common metric including, for example, population, number of households, employment, and/or gross state product. In addition, where possible, we explored variables that account for qualitative differences in the types of marketing efforts employed by the PAs.

A. Geographic data resolution

The level of geographic resolution of the model has implications for the data availability and other cross-cutting factors. Smaller geographic units offer the advantage of more data points to explain the variation in energy consumption. However, smaller geographies diminish the ability to capture spillover—a primary motivation for use of top-down models—and are faced with data availability constraints. Public sources, such as the US Census, obfuscate data to preserve confidentiality. Larger geographies necessarily mean fewer data points, which generally results in less certainty around model estimates.

This subsection presents the general trade-offs for the possible geographic aggregations. Details about specific data sources are in the subsequent subsections. The following are possible geographic aggregations, from largest to smallest:

- State-level models look at the energy demand in the state as a response to total activity in the state. Examples include Horowitz (2007), which considered energy consumption throughout the United States. The advantage of this geography level is the wealth of economic and energy time series data that are available. The disadvantages are that state-level models are not able to reflect differences within the state, and they cannot make use of a data series that is only available in one state.

- Utility- or program-level models compare energy consumption of one utility or program administration area to all others in the geographic scope of the model. The advantage of this geographic resolution is that measures of programmatic activity are readily available, making it particularly suited for comparing a large number of efficiency programs. The disadvantage is that this geographic resolution offers too few data points to develop a model for Massachusetts.

- County-level models compare energy consumption of one county to other counties. Models at this level of aggregation can operate within a state or potentially across multiple states. The county level is often the smallest level of reporting with complete detail from the census. The disadvantage to using counties as the geographic level is that program areas are not coincident with county boundaries. This is particularly an issue when the county contains one or more municipal utilities that are not part of a study area. Top-down models need to account for the amount of residential or commercial/industrial activity that does not occur in the study area.

- Town or census-designated place models offer a couple of advantages in comparison to other geographic levels. Town-level geographic boundaries are typically coincident with program administration areas. This enables a cleaner definition of the programmatic activity variables. Another advantage is that models at this level offer more data points than county-level models, which should better explain the variation in energy consumption. A primary disadvantage is that economic series from the census at this level do not always offer the same degree of resolution as the county level due to data confidentiality concerns. Further, town-level models are not able to capture spillover as effectively as models with larger geographies.

- Census-tract models have similar advantages and disadvantages to town-level models. There are more data points to explain the variation in energy consumption, but less data resolution due to confidentiality.

- Account-level models are the smallest possible geographic level. Models at this level look at changes in energy consumption as a response to programmatic activity and other factors that drive consumption at the account level. The challenge of modeling at this geographic level is the lack of time series data at the premise level. Proxies for production (e.g., manufacturing employment and wages) are available at town and county levels of geography, but are not typically known at the premise level.

For the C&I PA Data pilot study, we constructed models at the following two levels of geographic aggregation:

- County-level models – Models at the county level offer widely available time series data for the commercial and industrial models, as well as the forthcoming residential models.

- Town-level models – We selected town-level models over census-tract level because the data series available at the census-tract level to explain the variation in energy consumption are more limited than the town level. The Census publishes a series on wages and employment by industry code at the town level.

The geographic resolution is also tied to the geographic scope of the model. An extension of the models presented in this report would be to develop a model that goes beyond the boundaries of the Commonwealth. The benefit of such a model is that it would be able to compare program activity within Massachusetts to other geographies. However, such a model would necessarily be constrained by data availability for the larger geography.

The geographic scope of the model has implications on the model specification. One of the novel approaches in our pilot is the use of NAC (normalized annual consumption) as a dependent variable. NAC is a translation of each household’s monthly (or bimonthly) consumption history into the estimated annual usage for each year, under consistent temperature conditions over the years. There are several advantage to normalizing each household and year individually prior to fitting the time series cross-sectional models.

1. Individual households vary widely in their response to temperature. Coefficients of degree-day variables and the degree-day base temperature are all different across customers. Fitting individual models gives better granularity compared to fitting a single set of degree-day coefficients across all households and time periods.

2. Most of the economic and program data available for the time series-cross-sectional model is available on an annual basis. The degree-day response is estimated much

more accurately using monthly data than using annual data only. Using monthly data for the full time series cross-sectional model would substantially increase the complexity of that model.

3. One of the ways programs affect energy consumption is through measures that reduce heating and cooling usage. Capturing these effects appropriately in the time series analysis without normalizing to consistent weather would require more complex interactions between weather and program terms.

4. The basis for NAC is a physical model of household energy consumption. Putting degree-days into a pooled model along with various economic variables sacrifices the physical underpinnings of the degree-day relationship, leading to a variety of potential model mis-specificaton biases.

5. Annualizing the consumption data puts the staggered meter reading series onto a common calendar.

Use of NAC also has the following limitations.

6. To the extent the monthly model used to construct NAC is incomplete, the normalization may be inaccurate and introduce some other inaccuracies to the full model. A particular potential concern is that the relationship between consumption and degree-days may have some non-linearity. While this is true, often the observed non-linearity is related to the use of a fixed degree-day base across all homes. Further, every modelling approach is an approximation of a complex relationship with a more simple structure.

7. Home physical parameters that determine NAC do not change annually at the start of the year. Household structure, equipment, and occupancy patterns change throughout the year, with and without program engagement. Thus, an overall modelling structure that allows for gradual changes in individual household consumption patterns over time could have advantages. As noted, capturing those changes effectively would require that the full model use monthly rather than annual data.

8. Normalizing for weather response at the account level and then aggregating is possible only where we have access to account-level billing-period information. We will need to secure cooperation and funding from regulators and utilities/program administrators outside of the Commonwealth in order to develop a model with a broader geographical extent. It would also be possible to construct NACs for utility-wide aggregate data, provided monthly utility-wide aggregate data are available, however, the current analysis did not explore this option.

With these considerations in mind, our primary modelling approach used NAC normalized at the account level then aggregated to town or county as the dependent variable in an annual time series-cross-sectional model.

B. Sector-level resolution

We estimated separate models for the following C&I sectors to allow for greater differentiation across observational units:

- *Small commercial* – Customers identified as non-industrial customers, based on their NAICS codes, that had annual demand less than 300 kW, based on the PA billing and tracking data. Program activity for these customers was restricted to tracking records for the small business and upstream programs.
- *Medium to large commercial* – Customers identified as non-industrial customers, based on their NAICS codes, that had annual demand greater than 300 kW, based on the PA billing and tracking data. Program activity for these customers was restricted to tracking records for the small business and upstream programs.
- *Total commercial* – All non-industrial customers, based on their NAICS codes, regardless of size.
- *Industrial* – Customers identified as industrial based on their NAICS codes.

C. Temporal data resolution

There are two possible levels of temporal resolution: quarterly and annual. Like models with smaller levels of geographic aggregation, models based on quarterly data have the advantage of more data points to explain energy consumption variation. The disadvantage of quarterly models is one of data availability. The particular concern with Massachusetts data was program tracking data. There was too much uncertainty in the data at the quarterly level to recommend attempting a quarterly model. The choice of using annual data was consistent with all of the literature reviewed in the earlier section.

6.1.2 Detailed residential PA Data model specification

The residential PA data model is forthcoming, pending analysis of data to be provided by the MA Residential Contractor.

6.1.3 Detailed C&I PA Data model specification

This section describes the modeling specifications and process we employed to construct the PA Data pilot study C&I model. Again, note that we use δ (delta) to highlight the use of a differences-of-differences model specification, as discussed above. This is a standard billing analysis approach, extended to the cross-sectional economic aggregate, that works well when there is a limited number of time series observations, as exists during the initial model estimation phase in this study. Equation 6-2 presents the general form of the C&I PA Data model. We discuss the elements of this model in the sections that follow.

Equation 6-2. The PA Data Commercial/Industrial Model

$$\delta(NAC)_{tsgf} = \beta 0_{sgf} + \beta 1^*[\delta Employment]_{tsg} + \beta 2^*[\delta EE\ \$\ Program\ Activity]_{tsgf} + \beta 4^*£_{sgf} + \beta 4^*¥_{tsf}\ \varepsilon_{tsgf}$$

Where each variable in Equation 6-2 is defined as follows:

$\beta 0_{sgf}$	=	A fixed effects variable for sector (s), within geographic region (g), and by fuel type (f).
$(NAC)_{tsgf}$	=	Normalized (C&I) Annual Energy Consumption in year (t), sector (s), within geographic region (g), and by fuel type (f). For the county-level

		models, all variables are divided by gross domestic product to provide a measure of energy intensity per unit of output. For the town-level models, population is used in place of GDP due to data limitations.
$Employment_{tsg}$	=	Economic activity measured as the total employment per GDP or population, for county and town-level models, respectively, within year (t), sector (s), and geographic region (g).
Program activity	=	We considered two separate measures of programmatic activity separately: • EE $ Program Expenditure $Vbl(s)_{tsgf}$ is one or more EE program variables measured in $s, reflecting program expenditures as reported in the PA program tracking data, in year (t), sector (s), within geographic region (g), and by fuel type (f); and • EE Program Energy Savings Vbl_{tsgf} is a measure of estimated EE savings, as reported in the PA program tracking data, in year (t), sector (s), within geographic region (g), and by fuel type (f = electricity or natural gas).
*$£_{sgf}$	=	Parameter for geographic fixed effects for county or town g in sector s, and fuel type f.
$¥_{tsf}$	=	Parameter for annual fixed effects for year t in sector s, and fuel type f.

We used the following steps to develop the county- and town-level models:

- *Model of total NAC verses economic activity* – Before introducing program activity and other variables, we first investigated whether changes to NAC could be explained by changes in employment, as well as geographic and annual fixed effects. We constructed these simplified consumption models at two separate levels of analysis:
  - *County-level model* – This model used NAC per unit of GDP as a dependent variable.
  - *Town-level model* – Due to lack of available GDP and payroll data at more granular levels of analysis, the town-level model used NAC per capita as a dependent variable. Since only three years of data were available, we explored the non-differenced version of each model to provide for estimation across all three years of available data.
- *Introduce measures of program activity* – After determining that we could successfully model NAC as a function of the economic variable, we introduced the following two measures of programmatic activity separately. We considered the measures separately to limit collinearity:
  - Aggregate energy efficiency expenditures per unit – We obtained account- and measure-level downstream program expenditures and measure- and location-specific upstream data from the PA tracking data.

- Aggregate ex-ante savings per unit – We obtained account- and measure-level downstream program savings and measure- and location-specific upstream data from the PA tracking data.

  Due to data limitations, we were not able to include measure of the lag in program activity. We did attempt to construct lagged variables based on data provided within the PAs' annual reports. However, we concluded that the data contained in the PA annual reports and in the program tracking data were too dissimilar. We could not, with confidence, use the allocation of program tracking data by geography to allocate the data contained in the annual reports without making arbitrary assumptions regarding the differences between these two series.

- *Separate program activity by program type* – The evaluation team examined the impact of separating program expenditures and ex-ante savings into upstream and downstream activity, and examined the impact of lighting and non-lighting program activity on NAC.

- *Estimate first-difference form of each model* – The evaluation team estimated each of the models specified in their first differenced form.

The evaluation team employed these steps to estimate both county- and town-level non-differenced and differenced forms of each of the models identified in Table 6-1. The evaluation team's logic was to first determine whether we could actually model NAC as a function of the economy, as was suggested by the PAs' load forecasting groups. Next, we tried to add a single measure of program activity, either total ex ante savings or total expenditures, to determine if the models could pick up the program effects in aggregate with the limited time series. Then we split out the program variables into upstream and downstream, lighting and non-lighting to see whether differentiating between programs would improve the model performance. If there were multiple models that had significant results, we would have used a "goodness of fit (f-test)" to identify the appropriate model; however, for this exercise, we did not find statistically significant models, which preempted the need for a formal test.

Table 6-1. Alternative Model Descriptions for PA Data C&I Models

Model	Model Name	Model Description
Model 1	Employment Only	NAC is a function of employment plus time and geography fixed effects only
Model 2	Employment Plus Ex Ante Savings	NAC is a function of employment plus ex ante savings and time and geography fixed effects
Model 3	Employment Plus Total Expenditures	NAC is a function of employment plus time total program expenditures and geography fixed effects
Model 4	Upstream Plus Total Downstream Expenditures	NAC is a function of employment plus upstream and downstream program expenditures and time and geography fixed effects
Model 5	Upstream plus Lighting and Non-lighting Downstream Expendutures	NAC is a function of employment plus upstream and downstream expenditures and time and geography fixed effects. Downstream expenditures not are separated into lighting and non-lighting.
Model 6	Upstream Plus Total Downstream Savings	NAC is a function of employment plus upstream and downstream program savings and time and geography fixed effects
Model 7	Upstream plus Lighting and Non-lighting Downstream Savings	NAC is a function of employment plus upstream and downstream ex ante savings and time and geography fixed effects. Downstream savings are separated into lighting and non-lighting.

In the sections that follow, we discuss the specific data used in each of these models.

A. PA Data C&I model dependent variable (NAC): energy consumption data

The dependent variable for the PA Data pilot study C&I model was normalized annual electric consumption per (scaling) unit aggregated to geography. The source for the dependent variable was the PA consumption history contained in the C&I customer database, which records account-level consumption history for all C&I customers between 2011 and 2013. Below, we discuss the dependent variable in terms of weather normalization, our choice of timeframe (annual), scaling variable (to be determined), and unit of analysis (i.e., level of geography to which the data is aggregated).

Weather normalization. Weather is a primary determinant of energy consumption. Therefore, it was essential to account for weather-dependent impacts on energy consumption in a given time period to eliminate the possibility of those impacts obscuring programmatic impacts. There were essentially two options for weather normalizing savings estimates. One was to include terms for weather conditions as explanatory variables—specifically, heating degree days (HDD) and cooling degree days (CDD). While this approach would account for impacts attributable to changes in weather conditions, many explanatory variables are closely correlated to weather, which could have greatly complicated the ability to isolate programmatic impacts. The alternative approach was to remove weather-induced change from the dependent variable before constructing the model. This is frequently done in traditional billing analysis at the customer or account level using a degree-day normalization technique (i.e., the Princeton Scorekeeping Method, or PRISM™) that models each individual account's consumption as a function of HDD and CDD. Once the relationship between the customer's consumption and HDD and CDD is identified, consumption for each account under normal weather conditions is determined by imputing "normal HDD and normal CDD" conditions into the formula estimated from PRISM for that account or customer. While not used in the present analysis due to data and time limitations, another benefit of this technique is that account-level models could also be used to estimate heating and cooling loads for each customer, which could be used to construct indicatory variables for the relative level of heating and cooling sensitivity in an observational unit (i.e., town, census tract, or county). Normalization of the dependent variable for the effects of weather variation represents a noteworthy improvement over prior studies, which have typically included CDD/HDD as explanatory variables in the annual model or failed to account for weather effects altogether. This process requires account-level monthly billing data.

Time unit of the dependent variable. While the PA consumption history did contain data at the monthly level, the normalized consumption was constructed at the annual level. A primary reason for this decision was that some explanatory variables could not be easily obtained at a monthly level. A number of the exogenous data series available from public sources only reported data annually, and while others provided monthly estimates, such data were often interpolated from seasonal trends in other variables. Further, while variables reporting energy efficiency programmatic activity were reported at the month in which costs were accounted or the date the installation was reported, there is often a lag between these reported dates and the dates when savings will result. Further, different programs, measures, and accounts may have different lags. Attempting to isolate the programmatic activity to a single month would likely provide misleading results. Using annual data circumvented these issues, to an extent.[56]

Scaling variable(s) for NAC. The evaluation team scaled (i.e., divided) the county-level variables by gross domestic product. Due to data limitations, we scaled the town-level models by population. That choice was used as a scaling variable was dependent upon data availability and quality, at the desired level of granularity.

The evaluation team used account-level C&I billing data for the five electric PAs from 2011 through 2013 to construct the dependent variable: county- or town-level normalized annual consumption (NAC). Our methodology for developing the NAC used the site-level modeling approach originally developed for the PRISM software (Fels M. F., 1986). The theory regarding the underlying structure was discussed in materials for and articles about the software (Fels, Kissock, Marean, & Reynolds, 1995). PRISM was based on the concept that building operators or households cool (or heat) their building or home when temperatures were above (or below) a reference temperature. The PRISM model calculates the optimal reference temperature and responds to heating and cooling needs.

Using NAC as the measure of energy consumption had the following advantages:

[56] Future research should explore the possible effects of monthly differences between the reporting and realization of energy-efficiency expenditures, as there may be up to a four- to six-month difference between when expenditures are reported and when savings are realized.

- It separated weather-related energy consumption variation from other effects using higher resolution data. Monthly data gave the model more visibility into heating- and cooling-related energy consumption compared to an annual data series.
- Premises responded differently to heating and cooling needs. Some businesses, for instance, did not have air conditioning and, therefore, did not respond to cooling needs. PRISM accounts for premise-level differences in heating and cooling reference temperatures and responds to heating and cooling. The alternative of relying on an annual model to account for heating- and cooling-related consumption would have forced the entire population into the same linear response to temperature variation.

The fundamental PRISM regression model for a single premise is provided in Equation 6-3.

Equation 6-3. Basic PRISM Regression Model

$$E_t = \beta_0 + \beta_h H(\tau_h) + \beta_c C(\tau_c) + \varepsilon_t$$

Where:

E_t = Energy, measured in kWh, therms, or BTU, consumed at time period t

$H_t(\tau_h)$ = Calculated heating degree days using actual observed temperature at time period t and its deviation from reference temperature τ_h

$C_t(\tau_c)$ = Calculated cooling degree days using actual observed temperature at time period t and its deviation from reference temperature τ_c

$\beta_0, \beta_h, \beta_c$ = Regression coefficients measuring the marginal effect of base load, heating load, and cooling load, on a single site's energy consumption, respectively

ε_t = Regression residual in time period t

A PRISM analysis used cooling and heating degree days to measure the variation in a site's energy consumption that can be attributed to variation in weather conditions. These cooling and heating variable constructs were calculated using the following equations:

Equation 6-4. Cooling Degree Day Computation

$$C(\tau_c) = \begin{cases} 0, & x_t - \tau_c < 0 \\ x_t - \tau_c, & x_t - \tau_c \geq 0 \end{cases}$$

Equation 6-5. Heating Degree Day Computation

$$H(\tau_h) = \begin{cases} \tau_h - x_t, & x_t - \tau_h < 0 \\ 0, & x_t - \tau_h \geq 0 \end{cases}$$

In other words, if the observed temperature was above the cooling threshold τ_c, then that difference in degrees Fahrenheit was treated as cooling degree days, and vice versa for heating. The heating and cooling degree days for a particular billing period was determined by calculating the heating or cooling degree days for each day within the billing period and aggregating across all days. The aggregation of degree days was then associated with time period t. Once the reference temperature parameters (τ_c & τ_h) were chosen, and the model parameters $\beta_0, \beta_h,$ & β_c were estimated, Equation 6-6 provided predicted values of consumption.

Equation 6-6. Predicted Consumption Based on Account-level PRISM Model

$$\widehat{E_t} = \widehat{\beta_0} + \widehat{\beta_h} H_t(\tau_h) + \widehat{\beta_c} C_t(\tau_c)$$

These estimated coefficients were then used to make inferences on the degree to which a premise's energy consumption varies with respect to variation in temperature. In addition, TMY (typical meteorological year) weather constructs were developed to estimate what consumption would have been during a ""typical" year.

The success and appropriateness of a PRISM analysis involved several modeling choices and considerations regarding their possible implications. One such consideration was the choice of heating and cooling reference temperatures used in calculating degree days for each premise. Responses in energy consumption patterns to changes in weather were not homogenous across businesses, and therefore a range of reference temperatures should be considered for each business in the study. A popular range used for τ_h is between 40 and 55 degrees Fahrenheit, and for τ_c between 65 and 80 degrees Fahrenheit. The ranges chosen were inherently defined by the climatic conditions and diversity of Massachusetts.

While selecting a pair of heating and cooling reference temperatures for each site was a critical component of undertaking a PRISM analysis, so was the decision to even use heating and cooling terms to explain variation. There are several clear instances when a premise's energy consumption is not best explained by variations in weather. An example of this may have been a manufacturing or industrial facility that consumed a relatively steady pattern of energy throughout the course of the year. There were also instances when a premise was sensitive to heating, but not to cooling, and vice versa. For this reason, it was necessary to estimate models that represented each of these possible site-specific scenarios. The "full" model, "reduced" model, and base load models are described below.

- *Full Model* – This set of models estimated a fitted regression for each premise for each combination of reference temperatures for heating and cooling degree-day days. For example, if there were 5 heating degree day bases, and 5 cooling degree day bases, there were 5 * 5 = 25 models estimated and evaluated for each premise.
- *Reduced Cooling Model* – Estimates a regression model for each premise for a combination of cooling degree-day reference temperatures only.
- *Reduced Heating Model* – Estimates a regression model for each premise for a combination of heating degree-day reference temperatures only.
- *Base Load Model* – Estimates a regression model for each premise using only an intercept treated as base load.

The final choice of degree-day bases for the model was optimized across all candidate degree-day bases for each premise in the full model using chosen statistical measures such as Coefficient of Determination (R^2, Adjusted R^2), Akaike's Information Criterion (AIC), or Bayesian Information Criterion (BIC).

Using one or more of these statistical metrics, an optimal pair(τ_h, τ_c) was chosen for each premise. The model corresponding to these heating and cool reference temperatures was selected as the site's full model.

A single model was eventually selected for each premise. A joint F-test using each model's R-squared or adjusted R-squared was the preferred means to make this selection. It should be noted that the regression did not inherently account for the degrees of freedom associated with estimating (τ_h, τ_c). The degrees of freedom were adjusted to reflect these additional parameters, which are estimated.

Once a model was selected for each premise, TMY degree days could then be used in conjunction with the estimated model coefficients to provide normalized annual consumption for each premise. Again, this NAC was a measure of consumption the premise would have consumed during a "typical" year.

B. PRISM modeling results

The evaluation team subjected each account contained in the C&I consumption database to the PRISM analysis. We matched each unique site or premise to a corresponding NOAA weather station based on the premise's ZIP code. Figure 6-1 below shows the assignment of each five-digit ZIP code in Massachusetts to a NOAA weather station. Premises that could not be matched to a NOAA station because their ZIP code was missing or incorrect were assigned to the weather stations seen in Table 6-2 below, based on the corresponding electric PA.

Figure 6-1. ZIP Codes in Massachusetts Mapped to a NOAA Weather Station

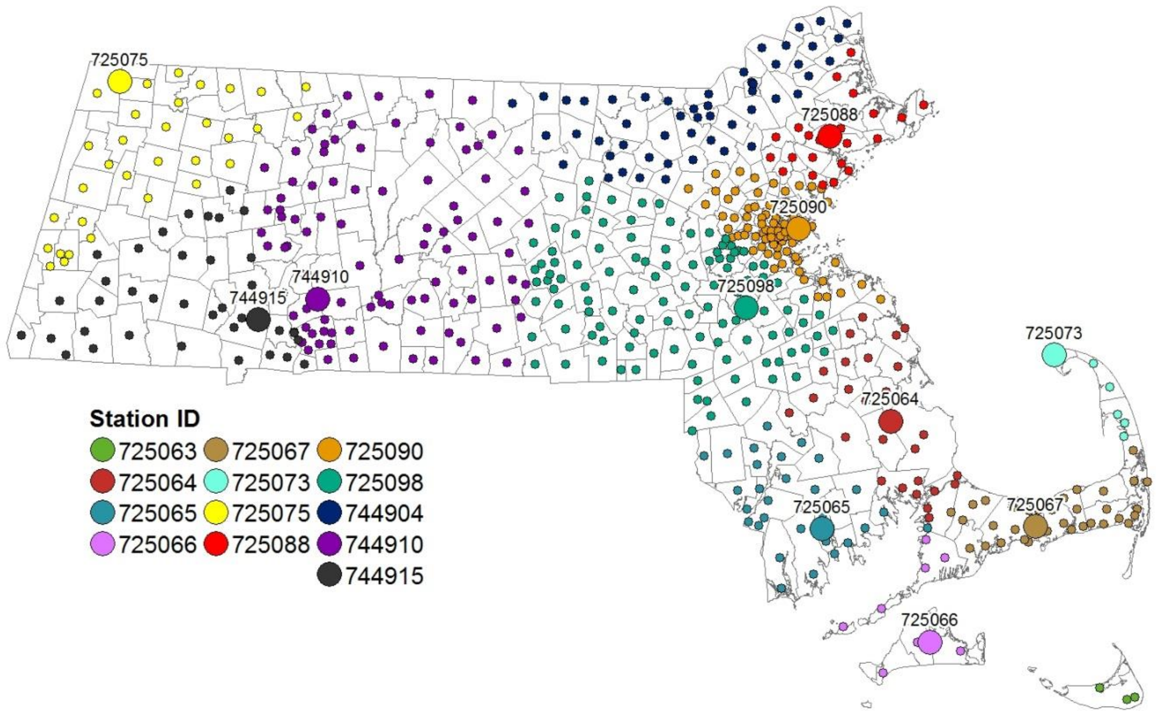


Table 6-2. Default NOAA Weather Stations for each Massachusetts PA

Program Administrator	NOAA Station ID
Berkshire Gas	725075
Cape Light Compact	725067
Columbia	744910
Liberty Gas	725065
NStar	725098
National Grid	725090
Unitil	744904
WMECO	744910

Once premises were assigned to a NOAA station, consumption records were matched with the corresponding actual and typical temperature observations for that station. For each billing period and for each premise, heating and cooling degree day variables were constructed. For HDD variables, a range of 45 to 55 degrees Fahrenheit was used as reference temperatures, while 65 to 75 degrees Fahrenheit was used for CDD.

The energy consumption CDD and HDD variables were then used to estimate a fitted regression model for each premise in each year (2011, 2012, and 2013). The Coefficient of Determination was used for model selection. For each year and in total, the portfolio of selected models are provided in the following table.

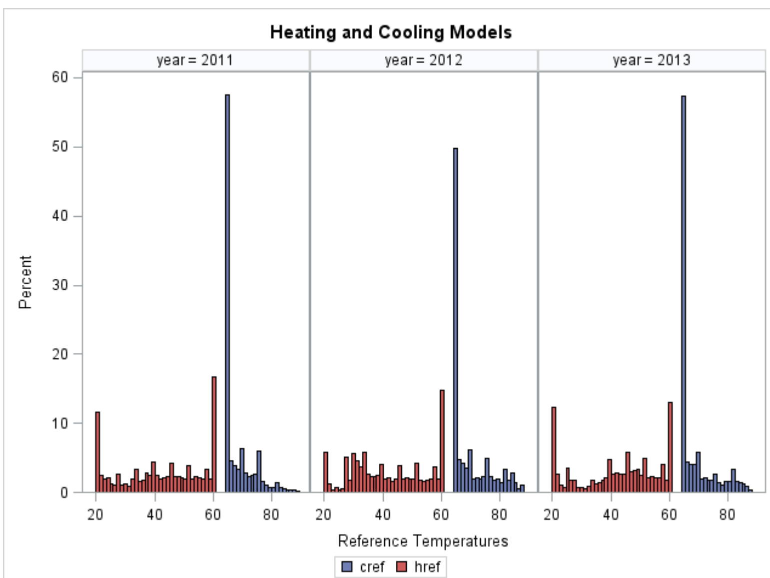
Table 6-3. Prism Model Selection Results

Final Selected Model	Parameters	2011	2012	2013	Total
Base Load	β_0	18%	17%	16%	**51%**
Full	$\beta_0, \beta_h, \beta_c$	2%	2%	2%	**6%**
Reduced	β_0, β_h or β_c	15%	14%	14%	**43%**

The optimization of PRISM model selection resulted in the majority of premises being best represented by a base load only model. This is not necessarily surprising, as both commercial and industrial sectors are significantly less sensitive to variations in temperature than the residential sector. This limited resulting weather normalization points to the advantage of using NAC as the dependent variable compared to the approach of including explanatory variables to measure the effect of weather. This analysis demonstrates that the base loads for roughly half of the C&I population were substantial, limiting the models' ability to detect weather dependent variation in the consumption series.

The evaluation team further explored the model selection process to determine the underlying cause for roughly 50 percent of models selected with base load. Figure 6-2 below shows the distribution of reference temperatures of fitted PRISM models used to calculate NACs for each account. The figure illustrates that the evaluation team attempted a wide range of reference temperatures to compute heating and cooling degree days in order to determine whether a significant heating or cooling effect could be identified.

Figure 6-2. Heating and Cooling Reference Temperatures for Fitted PRISM models



Next, we examined the process employed as noted above. Once the best-fitting set of base temperatures for heating and cooling were identified for each account, we employed a goodness-of-fit (f-test) to determine whether the heating or cooling, or both heating and cooling, added any explanatory power over the base load only models. Our analysis of the f-test results showed that, while many accounts did have fitted models for either heating, cooling, or both, the goodness-of-fit test determined that these models did not explain sufficient variation in consumption to provide improvements over the base load only model.

The fitted models were used to produce NACs for each premise and each year. This result was used as a dependent variable in the macroeconomic models described in the next section. The evaluation team applied the PRISM approach to the 2011, 2012, and 2013 monthly billing history contained in the C&I customer profile database. Figure 6-3 and Figure 6-4 below show the normalized annual heating and cooling loads for C&I customers as a percentage of total load by different levels of geographic resolution in the state. These data allowed us to construct top-down models that measure programmatic impacts on heating measures (i.e., heating systems), cooling measures (i.e., central air conditioning), and base load measures (i.e., lighting).

Figure 6-3. Heating Load Index: Number of Standard Deviations from the State Mean Percent of Total Load that is Due to Heating by ZIP Code (C&I Customers)

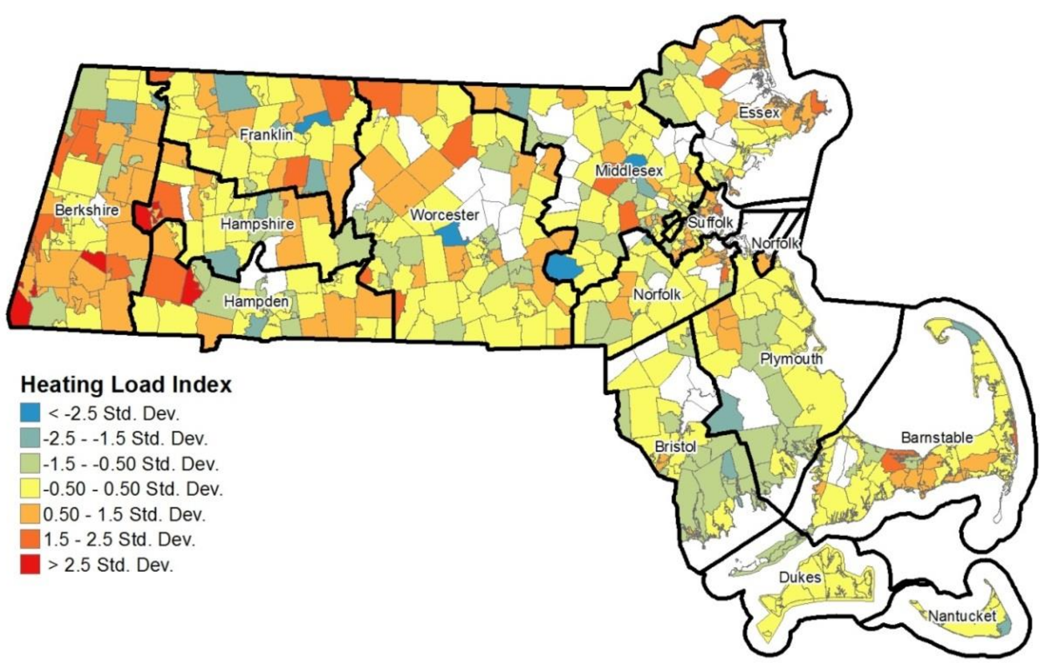
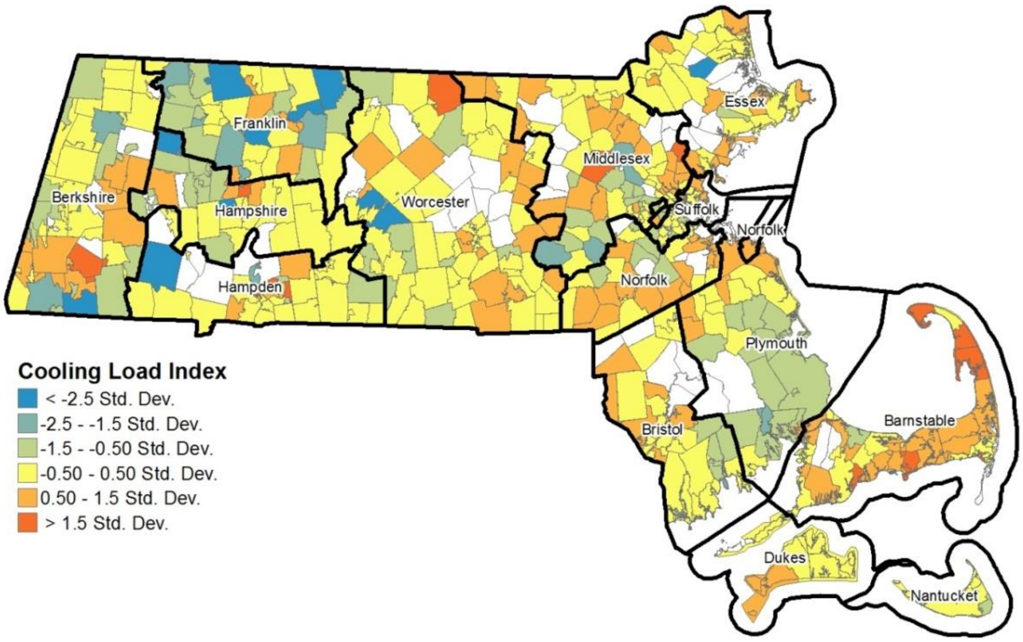


Figure 6-4. Cooling Load Index by ZIP Code: Number of Standard Deviations from the State Mean Percent of Total Load that is Due to Cooling by ZIP Code (C&I Customers)



C. PA Data model measures of programmatic activity

This section describes the range of variables to measure programmatic activity that the evaluation team considered in the PA Data model. The source for the program variables was tracking data contained in the C&I customer database. For downstream programs, the PA tracking data contained incentive costs, ex-ante savings, and number of rebates by measure type. For upstream programs, the database contained program rebates and fixture types that were mapped to individual customer names. The C&I team matched these customer names to the customer's ZIP code, census track, town, and county.

We explored the following variables for this study:

- Estimates of kWh energy savings delivered through the EE programs.
- Rebate $ expenditures paid to EE program participants for installation of EE measures.
- Total $ expenditures on EE programs delivered through the PAs.
- EE $ expenditures by other broad program categories (e.g., marketing/advertising, administrative, etc.).
- Program variables were scaled by the same unit of measure and aggregated to the same geography as the dependent variable. We tested different lagged relationships between the program variables and the dependent variables if program data was obtained a few months to one year prior to the periodicity of the dependent variable(s).

The PAs track programmatic activity for their energy efficiency portfolios. The C&I and residential program tracking data available for this project began in 2011. The variables in the program tracking data included the year of the program, a description of the measures, and a description of the participants. The program tracking data also contained the two measurements of programmatic activity that were consistent with the reviewed literature in Section 4:

- *Ex-ante savings* – Similar to Parfomak and Lave, the use of ex-ante savings determined by deemed savings computations of existing bottom-up evaluation studies provides a measure of the realization rate for program savings. Ex-ante savings estimates can be further separated into upstream and downstream program savings. The evaluation team explored estimating differences by these program types as well as measure type (i.e., percentage of savings attributable to non-lighting).
- *Program expenditures* – As discussed in Loughran and Kulick and Rivers and Jaccard, the use of program expenditures provides a measure of the cost effectiveness of programmatic activity. Program expenditures can be further separated into incentive costs and other program costs, as well as expenditures for upstream and downstream programs and possibly by measure type.

Figure 6-5 and Figure 6-6 below illustrate the total C&I energy efficiency program expenditures and the percent of expenditures that were from upstream programs by county and block group. The evaluation team used these data to construct county- and town-level measures of ex-ante savings and program expenditures; we also split those values out according to upstream, downstream, lighting, and non-lighting programs.

Figure 6-5. Total Expenditures for Energy Efficiency Program by Town (C&I Customers)

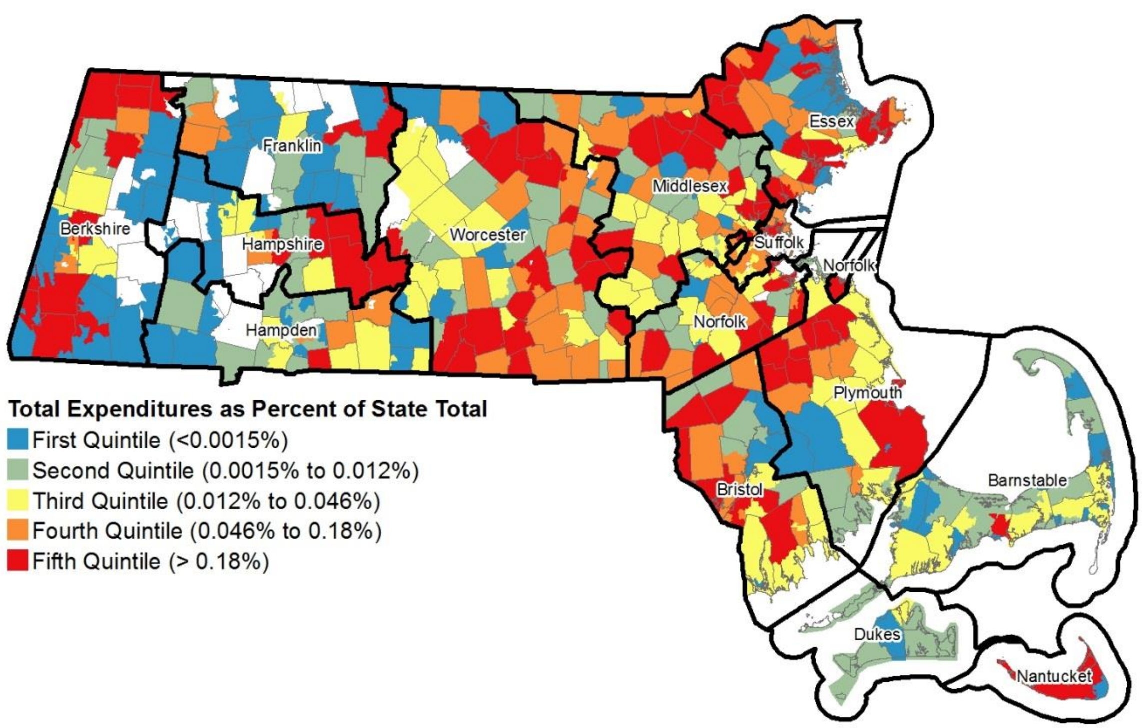
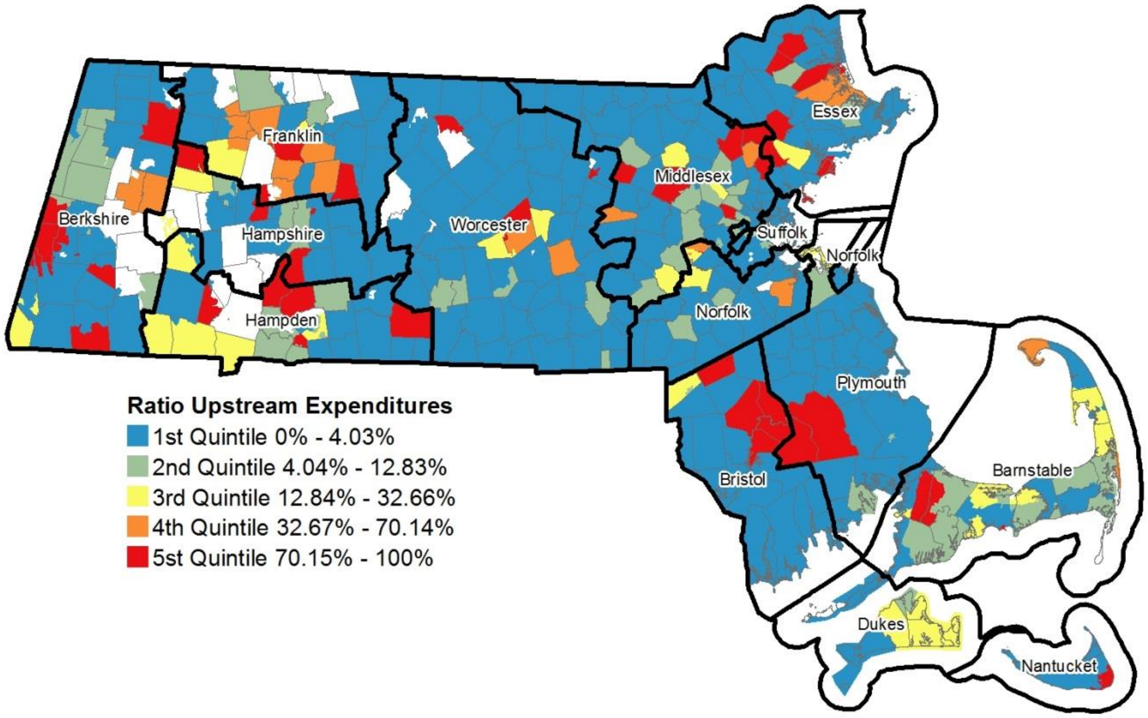


Figure 6-6. Percent of Total Expenditures for Energy Efficiency Program that are Upstream by Town (C&I Customers)



D. PA Data model consideration of building codes

While not included in the present models due to limited variation in the data, the evaluation team also considered the impact of building code changes on consumption. Building codes define standards of building performance and are a key driver of energy consumption as discussed in Cadmus and in Jacobsen and Kotchen. Incorporating the effect of building codes on energy consumption requires tracking the number of buildings by code edition. We reviewed various code changes and compliance rates associated with the commercial energy code identified (IECC 2009/ASHRAE 90.1 2007) in the ARRA legislation. For projects permitted before IECC 2009 went into effect, code compliance was based on Massachusetts Building Code Chapter 13/IECC 2006. These key dates can be used to construct metrics of the total amount of new construction prior to 2006, from 2007 to 2008, 2008 to 2009, and after 2009.

The evaluation team examined options for using the percent of square footage by building vintage as an indicator of code changes. We identified three sources of information around the building vintage:

- *Census Building Permits Survey* – This is a monthly county-level series and an annual town-level series of new residential home construction. These data can be used to estimate the number of homes by building vintage. These data are applicable to C&I and residential models.

- *Dodge Construction Market Research* – These data track construction activity at the project level and can be aggregated to the ZIP code, census tract, or county levels. These data are applicable to C&I models only.
- *Massachusetts Tax Assessor Data* – This database contains important information for describing the state's building stock, particularly construction dates and square footage of buildings at a parcel level for the 351 municipalities in Massachusetts. These data have the advantage of being used for revenue gathering, which implies a high level of scrutiny. These data are applicable to both the C&I and residential models.

The advantages of these three sources of information are as follows:

- *Data Reporting* – The data are transactional (i.e., they record changes to parcels) and can be transformed and aggregated into time series data at both the county and census-tract level.
- *Time Series Considerations* – These data are reported for all buildings in the Commonwealth in a year. The year-over-year building stock increases by the number of new units built and decreases by the number of units destroyed. Examples of the latter include tear-downs of smaller homes to facilitate larger buildings.
- *Level of Aggregation* – The data are at the parcel level, which has the advantage of cleanly aggregating to census tracts and counties.

Ultimately, we were not able to construct a meaningful measure of change to building codes over the three-year time period for this study. Given a longer time series, we recommend introducing changes to building codes into the models.

E. PA Data C&I model exogenous variables

Table 6-4 describes the data that are unique to the commercial and industrial model. The following describes the data sources considered for non-programmatic variables and the scaling variable:

- Tax assessor data contains the construction date and square footage of buildings at a parcel level, as described in Section 6.1.3. Unlike the residential model, we have not been able to identify alternative data sources. Unfortunately, these data do not provide sufficient variation over the three-year time period, as 2013 data are not available. Similarly, we considered using the Dodge Player's database to examine trends in new construction; however, the available data series ended in 2011.
- Moody's provides county-level data combined from the Bureau of Labor Statistics (BLS), Bureau of Economic Analysis, and Census Bureau. We have the data as an annual series beginning in 2010. The employment data are by NAICS, which allows for specifying separate models for commercial and industrial.
- The Quarterly Census of Employment and Wages (QCEW) is a product of BLS. The data reports wages and employment by NAICS code at the Census Designated Place or town level. These quarterly series data (going back to the 1930s) are the official government estimates of economic activity by sector of the economy, and one

of Moody's sources. We were not able to find a viable alternative data source at the town level.

- Energy Information Agency (EIA) tracks electricity prices at a state level. The energy price data source is the same as that for the residential sector model. Unfortunately, these data do not provide sufficient variation over the three-year time period, as 2013 data are not available. Consequently, we were not able to include energy prices in the models. Given a longer time series, we will include energy prices.
- County business pattern data – This census product provides an estimate of the employment and GDP at the county level.

Our review of the data identifies that different sources of information must be used to provide the most comprehensive and precise measures of economic data at different levels of geography. For confidentiality purposes, many data series available at the county-NAICS code level are not available by NAICS code at a finer geographic resolution. Further, data series such as Gross State Product (GSP) are only available at the county-NAICS level and are not available at town or census block group, regardless of whether the series are subdivided by industry. Table 6-4 identifies the sources for different data series at the county and town level.

Table 6-4. Recommended C&I Model Data

Scaling Variable	Description	Level of Aggregation	Source
Energy Consumption	Normalized annual consumption	County	Monthly account level billing data
		Town	
Programmatic Activity	Ex ante savings	County	PA program tracking data
		Town	
	Downstream program expenditures	County	
		Town	
	Upstream program expenditures	County	Address level data available from Ecova
		Town	
Building Vintage	Year build	County	Construction dates from tax assessor data at a building level
		Town	
Population	Total residents by age group	County	Non-group quarters population from 1-year American Community Survey
		Town	
Energy Prices	Prevailing residential rates	County	N/A - EIA reports 2011 and 2012 by PA only
		Town	
Output	Gross Domestic Product	County	Moody's data
		Town	N/A
Employment	Total residents by age group	County	County Business Patterns
		Town	
Payroll	Total residents by age group	County	County Business Patterns
		Town	N/A

Data availability complicated our ability to use a consistent scaling variable for both the town- and county-level models. Ideally, GSP provides the truest measure of energy intensity, as it speaks directly to production efficiency, making it a desirable scaling variable. However, the US Census reports that the most granular level GSP is tracked at the county level, which is seen in Figure 6-7 below. Possible alternative scaling variables include employment and square footage. The evaluation team obtained square footage from the tax assessment data. We also obtained employment by block group from County Business Patterns, as seen in Figure 6-8. However, it should be noted that employment cannot enter the model as both a scaling variable and as a measure of non-programmatic activity. These data are provided by County Business Patterns on an annual basis through 1986.

Figure 6-7. Gross State Product by County

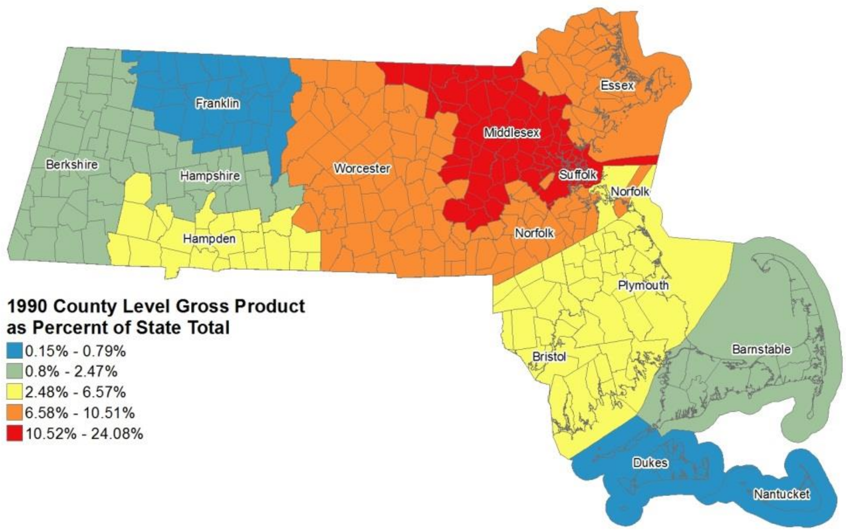
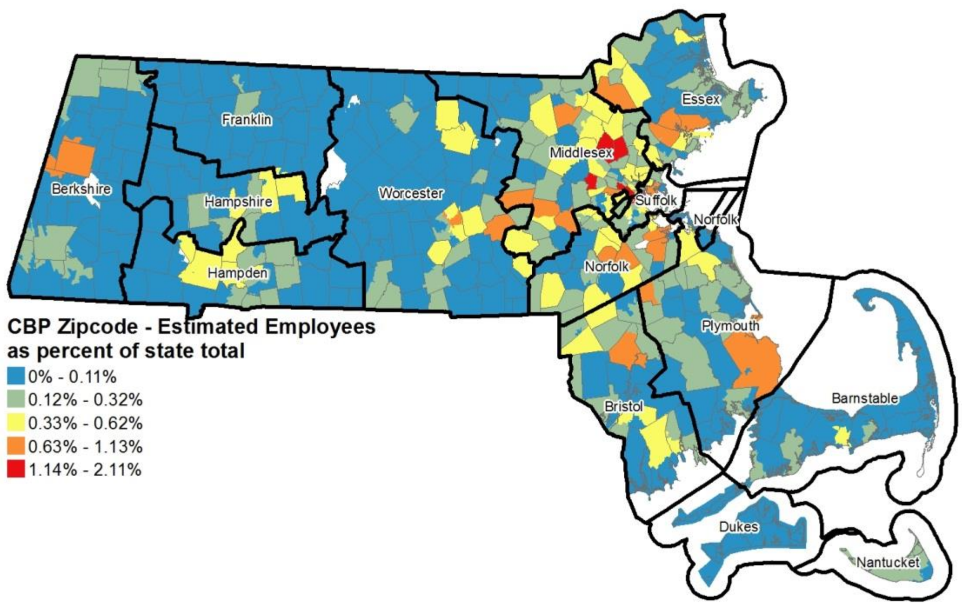


Figure 6-8. Percent of Total State Employees by County and Block Group (2012)



County Business Patterns data contained a number of additional macro-economic series at the block group level that we explored in the models to control for differences in the local economy that may account for changed in consumption. Figure 6-9 and Figure 6-10 below illustrate the distribution of businesses and payroll by census block group and county.

Figure 6-9. Number of Businesses by County and Block Group (2012)

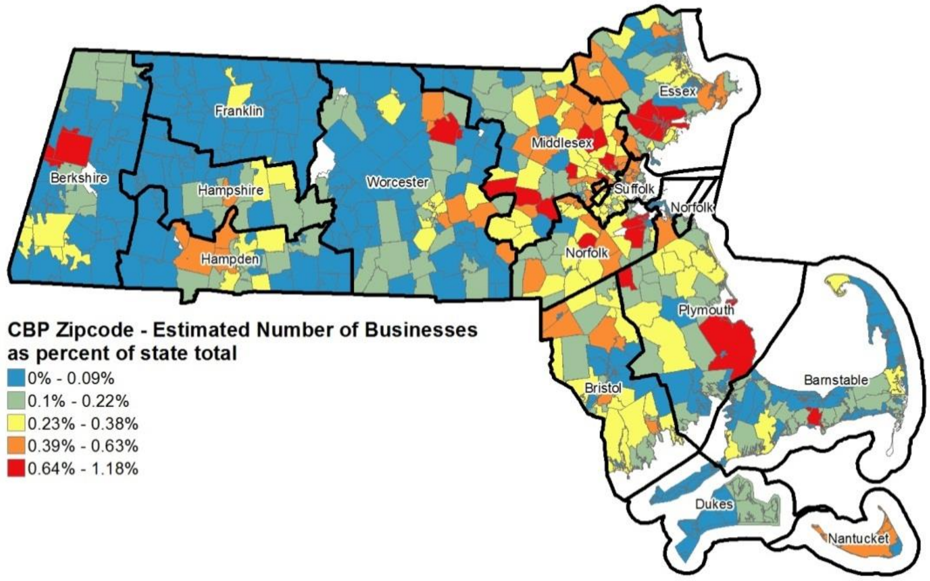
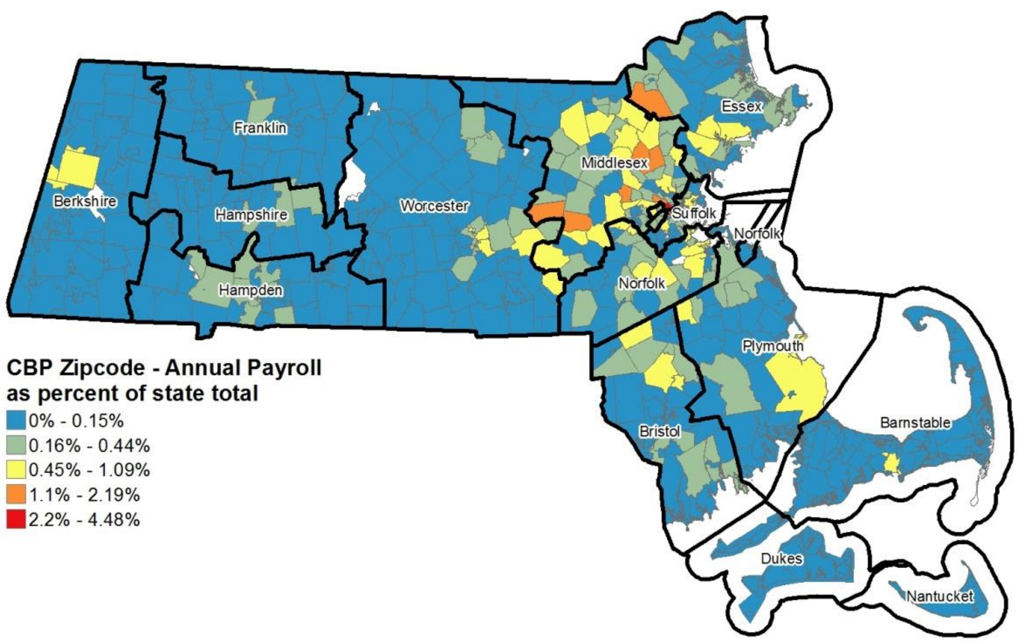


Figure 6-10. Percent of State Payroll by County and Block Group (2012)



6.2 PA DATA MODEL RESULTS

This section reviews the estimation results for the county-level models first, examining both the non-differenced and the differenced model results for small and large commercial and industrial sectors. Then, we examine results for these models estimated using the first difference in dependent and independent variables. Once we review the county-level results, we present the same set of results for the town-level models.

6.2.1 PA Data residential model results

The residential model results are forthcoming.

6.2.2 PA Data C&I model results

Table 6-5 shows the results for the seven non-differenced county-level models. The first model estimated (Model 1) used only employment to predict the estimated level of NAC. The results show that for small businesses, employment-only was a good predictor of NAC, as the parameter estimate was both statistically significant and positive. This model also was significant for large commercial and total commercial, but not for the industrial sector.

Model 2 adds in the total ex ante savings. The Model 2 results show that it was not possible to have statistically significant terms for both savings and NAC. The same is true for Model 3 (employment plus total program expenditure). As program expenditures are further split into upstream and downstream expenditures (Model 4), the parameter estimates on expenditures are negative and significant; however, the estimated parameter for employment is still not significant. When downstream expenditures are separated into lighting and non-lighting expenditures (Model 5), the parameters are significant, but the sign on downstream non-lighting expenditures is positive.

None of the savings models show statistically significant results for the program variables for any sector. One possible reason for this result is that atypical changes in both dependent and predictor variables associated with the economic recovery made it difficult to obtain stable coefficients over the period studied. A longer time series may potentially mitigate this problem and produce more reliable results.

To determine if the weather normalization method was responsible for the model results, we also attempted a set of models that used actual consumption as the independent variable instead of NAC. These models included heating and cooling degree-days as explanatory variables. However, none of the models tested showed a statistically significant relationship between consumption and degree-days, nor did the significance of the other model parameters improve.

Table 6-5. Non-Differenced Commercial and Industrial Model—County Level

	Model Specification 1		Model Specification 2		Model Specification 3		Model Specification 4		Model Specification 5		Model Specification 6		Model Specification 7	
	Employment Only		Employment Plus		Employment Plus		Upstream Plus		Upstream plus		Upstream Plus		Upstream plus	
Economic Sector	**Estimate**	**P-Value**	**Estimate**	**P-Value**	**Estimate**	**P-Value**	**Estimate**	**P-Value**	**Estimate**	**P-Value**	**Estimate**	**P-Value**	**Estimate**	**P-Value**
Small Commercial														
Total kWh Savings			0.074	0.231										
Downstream Total Savings													0.014	0.860
Downstream Lighing Savings											-0.026	0.722		
Upstream Total/Lighting Savings											-0.004	0.916	-0.016	0.570
Downstream Non-lighting Savings											0.021	0.481		
Total kWh Expenditures					**0.067**	**0.025**								
Downstream Total Expenditures									-0.005	0.890				
Downstream Lighing Expenditures							**-0.077**	**0.084**						
Upstream Total/Lighting Expenditures							-0.011	0.563	-0.029	0.209				
Downstream Non-lighting Expenditures							**0.035**	**0.071**						
Total Employment (Economic Condition)	**0.589**	**0.065**	-0.248	0.772	-0.259	0.749	0.484	0.523	-0.450	0.570	-0.043	0.966	-0.355	0.684
Large Commercial														
Total kWh Savings			0.298	0.102										
Downstream Total Savings													0.191	0.399
Downstream Lighing Savings											-0.141	0.176		
Upstream Total/Lighting Savings											0.030	0.741	0.040	0.721
Downstream Non-lighting Savings											0.079	0.512		
Total kWh Expenditures					0.305	0.138								
Downstream Total Expenditures									0.005	0.974				
Downstream Lighing Expenditures							**-0.222**	**0.002**						
Upstream Total/Lighting Expenditures							0.029	0.456	-0.027	0.751				
Downstream Non-lighting Expenditures							**0.179**	**0.012**						
Total Employment (Economic Condition)	**2.164**	**0.020**	**3.845**	**0.006**	**3.355**	**0.013**	1.527	0.337	0.705	0.846	1.470	0.585	1.299	0.700
All Commercial														
Total kWh Savings			0.015	0.880										
Downstream Total Savings													-0.150	0.117
Downstream Lighing Savings											-0.170	0.110		
Upstream Total/Lighting Savings											0.022	0.856	0.003	0.978
Downstream Non-lighting Savings											-0.044	0.327		
Total kWh Expenditures					0.021	0.793								
Downstream Total Expenditures									**0.145**	**0.000**				
Downstream Lighing Expenditures							**-0.176**	**0.062**						
Upstream Total/Lighting Expenditures							0.041	0.698	**0.055**	**0.069**				
Downstream Non-lighting Expenditures							0.042	0.402						
Total Employment (Economic Condition)	**3.409**	**0.000**	**4.667**	**0.000**	**4.684**	**0.000**	**6.391**	**0.000**	**0.625**	**0.000**	**3.538**	**0.000**	**3.508**	**0.000**
Industrial														
Total kWh Savings			0.043	0.369										
Downstream Total Savings													0.060	0.450
Downstream Lighing Savings											0.014	0.862		
Upstream Total/Lighting Savings											**0.218**	**0.062**	**0.217**	**0.029**
Downstream Non-lighting Savings											0.055	0.431		
Total kWh Expenditures					-0.013	0.792								
Downstream Total Expenditures									0.205	0.307				
Downstream Lighing Expenditures							0.145	0.299						
Upstream Total/Lighting Expenditures							-0.092	0.707	0.093	0.430				
Downstream Non-lighting Expenditures							-0.066	0.792						
Total Employment (Economic Condition)	0.184	0.632	0.018	0.989	-0.341	0.780	-0.107	0.986	4.332	0.358	3.672	0.260	3.548	0.236

Table 6-6 shows the results for the same set of models using the first-difference series. This table shows that differencing does improve the statistical significance of some of the commercial models. This is likely because differencing reduces the impact associated with year-over-year changes in the economy to allow the model to isolate impacts that result from changes in the program variables relative to NAC.

The county level PA Data model was not able to detect a statistically significant relationship between programmatic activity and consumption. However, based on the results of the PA-Muni model as well as previous top-down research, we speculate that it is necessary to have a sufficiently long time-series to account for the cumulative effects of program expenditures over time. A longer time series would allow the evaluation team to test for what the PA-Muni model found to be a lagged effect of program spending and energy savings. In addition to expanding the time series, it may be desirable to explore sector level models such as retail, manufacturing, and public sector. Another possibility would be to expand the scope to a regional analysis as opposed to just Massachusetts. This is because the introduction of variables that provides greater differentiation of programmatic activity across units (i.e., upstream and downstream lighting and non-lighting activity) shows greater significance of program variables. However, we note that the timeframe being studied would be particularly challenging given the period of economic decline and recent growth, as well as the recent escalation in programmatic activity.

Table 6-6. Differenced Commercial and Industrial Model—County Level

	Model Specification 1		Model Specification 2		Model Specification 3		Model Specification 4		Model Specification 5		Model Specification 6		Model Specification 7	
	Employment Only		Employment Plus Ex Ante Savings		Employment Plus Total Expenditures		Upstream Plus Total Downstream Expenditures		Upstream plus Lighting and Non-lighting Downstream Expendutures		Upstream Plus Total Downstream Savings		Upstream plus Lighting and Non-lighting Downstream Savings	
Economic Sector	Estimate	P-Value	Estimate	P-Value	Estimate	P-Value	Estimate	P-Value	Estimate	P-Value	Estimate	P-Value	Estimate	P-Value
Small Commercial														
Total kWh Savings			0.043	0.493										
Downstream Total Savings													0.014	0.860
Downstream Lighing Savings											-0.026	0.722		
Upstream Total/Lighting Savings											-0.004	0.916	-0.016	0.570
Downstream Non-lighting Savings											0.021	0.481		
Total kWh Expenditures					**0.057**	**0.063**								
Downstream Total Expenditures									-0.005	0.890				
Downstream Lighing Expenditures							**-0.077**	**0.084**						
Upstream Total/Lighting Expenditures							-0.011	0.563	-0.029	0.209				
Downstream Non-lighting Expenditures							**0.035**	**0.071**						
Total Employment (Economic Condition)	**0.658**	**0.007**	0.111	0.927	-0.159	0.887	0.484	0.523	-0.450	0.570	-0.043	0.966	-0.355	0.684
Large Commercial														
Total kWh Savings			**0.368**	**0.049**										
Downstream Total Savings													0.191	0.399
Downstream Lighing Savings											-0.141	0.176		
Upstream Total/Lighting Savings											0.030	0.741	0.040	0.721
Downstream Non-lighting Savings											0.079	0.512		
Total kWh Expenditures					0.218	0.183								
Downstream Total Expenditures									0.005	0.974				
Downstream Lighing Expenditures							**-0.222**	**0.002**						
Upstream Total/Lighting Expenditures							0.029	0.456	-0.027	0.751				
Downstream Non-lighting Expenditures							**0.179**	**0.012**						
Total Employment (Economic Condition)	**1.774**	**0.028**	**5.256**	**0.017**	**6.078**	**0.008**	1.527	0.337	0.705	0.846	1.470	0.585	1.299	0.700
All Commercial														
Total kWh Savings			**0.288**	**0.012**										
Downstream Total Savings													0.131	0.237
Downstream Lighing Savings											**-0.155**	**0.004**		
Upstream Total/Lighting Savings											0.018	0.647	-0.001	0.982
Downstream Non-lighting Savings											**0.033**	**0.401**		
Total kWh Expenditures					0.114	0.167								
Downstream Total Expenditures									-0.024	0.678				
Downstream Lighing Expenditures							**-0.192**	**0.000**						
Upstream Total/Lighting Expenditures							0.018	0.462	-0.030	0.467				
Downstream Non-lighting Expenditures							**0.091**	**0.000**						
Total Employment (Economic Condition)	1.608	0.002	**5.947**	**0.000**	**6.546**	**0.000**	**1.862**	**0.052**	0.223	0.886	0.946	0.452	0.407	0.790
Industrial														
Total kWh Savings			0.051	0.377										
Downstream Total Savings													0.060	0.450
Downstream Lighing Savings											0.014	0.862		
Upstream Total/Lighting Savings											**0.218**	**0.062**	**0.217**	**0.029**
Downstream Non-lighting Savings											0.055	0.431		
Total kWh Expenditures					-0.012	0.836								
Downstream Total Expenditures									0.205	0.307				
Downstream Lighing Expenditures							0.145	0.299						
Upstream Total/Lighting Expenditures							-0.092	0.707	0.093	0.430				
Downstream Non-lighting Expenditures							-0.066	0.792						
Total Employment (Economic Condition)	0.221	0.570	0.976	0.641	0.668	0.739	-0.107	0.986	4.332	0.358	3.672	0.260	3.548	0.236

Table 6-7 shows the results for the seven non-differenced town-level models, and Table 6-8 shows the model results for the same set of models using the first-difference series.

The program variables for the town-level models perform less well than those for the county-level models. While a number of models do have statistically significant parameters for the program variables, none of the models have statistically significant parameters for both the program and employment variables. Differencing does not appear to improve the fit of these models.

While the fixed effects terms are not shown in the tables below to conserve space, it should be noted that the town-level fixed effects were largely statistically significant, whereas the county-level fixed effects were not. Further research should explore whether town level fixed effects serve as a proxy for industry level differences.

Table 6-7. Non-Differenced Commercial and Industrial Model—Town Level

	Model Specification 1		Model Specification 2		Model Specification 3		Model Specification 4		Model Specification 5		Model Specification 6		Model Specification 7	
	Employment Only		Employment Plus		Employment Plus		Upstream Plus		Upstream plus		Upstream Plus		Upstream plus	
Economic Sector	Estimate	P-Value	Estimate	P-Value	Estimate	P-Value	Estimate	P-Value	Estimate	P-Value	Estimate	P-Value	Estimate	P-Value
Small Commercial														
Total kWh Savings			-0.013	0.400										
Downstream Total Savings													-0.027	0.402
Downstream Lighing Savings											-0.009	0.688		
Upstream Total/Lighting Savings											-0.018	0.243	-0.017	0.238
Downstream Non-lighting Savings											-0.008	0.532		
Total kWh Expenditures					-0.020	0.306								
Downstream Total Expenditures									-0.029	0.288				
Downstream Lighing Expenditures							-0.018	0.426						
Upstream Total/Lighting Expenditures							-0.008	0.656	-0.009	0.636				
Downstream Non-lighting Expenditures							-0.005	0.650						
Total Employment (Economic Condition)	0.003	0.905	0.003	0.958	-0.019	0.806	0.070	0.416	0.060	0.491	0.080	0.350	0.058	0.516
Large Commercial														
Total kWh Savings			0.107	0.114										
Downstream Total Savings													**0.064**	**0.027**
Downstream Lighing Savings											-0.029	0.200		
Upstream Total/Lighting Savings											0.002	0.878	0.007	0.653
Downstream Non-lighting Savings											**0.036**	**0.018**		
Total kWh Expenditures					**0.110**	**0.095**								
Downstream Total Expenditures									0.022	0.366				
Downstream Lighing Expenditures							**-0.045**	**0.005**						
Upstream Total/Lighting Expenditures							0.010	0.601	0.025	0.209				
Downstream Non-lighting Expenditures							0.026	0.106						
Total Employment (Economic Condition)	-0.150	0.340	-0.073	0.914	-0.263	0.687	0.401	0.475	0.266	0.655	0.560	0.353	0.598	0.320
All Commercial														
Total kWh Savings			**-0.085**	**0.010**										
Downstream Total Savings													**-0.083**	**0.020**
Downstream Lighing Savings											-0.008	0.811		
Upstream Total/Lighting Savings											0.009	0.829	0.012	0.773
Downstream Non-lighting Savings											**-0.040**	**0.039**		
Total kWh Expenditures					**0.130**	**0.000**								
Downstream Total Expenditures									**0.103**	**0.005**				
Downstream Lighing Expenditures							**0.141**	**0.000**						
Upstream Total/Lighting Expenditures							0.030	0.538	0.028	0.583				
Downstream Non-lighting Expenditures							-0.024	0.216						
Total Employment (Economic Condition)	0.003	0.968	0.064	0.760	0.099	0.681	0.229	0.455	0.184	0.569	0.070	0.824	-0.023	0.943
Industrial														
Total kWh Savings			0.007	0.709										
Downstream Total Savings													-0.018	0.416
Downstream Lighing Savings											0.018	0.209		
Upstream Total/Lighting Savings											0.019	0.249	0.020	0.241
Downstream Non-lighting Savings											-0.018	0.201		
Total kWh Expenditures					-0.007	0.695								
Downstream Total Expenditures									-0.010	0.694				
Downstream Lighing Expenditures							0.027	0.077						
Upstream Total/Lighting Expenditures							**0.039**	**0.038**	**0.041**	**0.044**				
Downstream Non-lighting Expenditures							-0.024	0.127						
Total Employment (Economic Condition)	**0.191**	**0.004**	-0.358	0.262	-0.266	0.355	0.139	0.814	0.312	0.629	-0.053	0.933	0.214	0.732

Table 6-8. Differenced Commercial and Industrial Model—Town Level

Economic Sector	Model Specification 1: Employment Only		Model Specification 2: Employment Plus Ex Ante Savings		Model Specification 3: Employment Plus Total Expenditures		Model Specification 4: Upstream Plus Total Downstream Expenditures		Model Specification 5: Upstream plus Lighting and Non-lighting Downstream Expendutures		Model Specification 6: Upstream Plus Total Downstream Savings		Model Specification 7: Upstream plus Lighting and Non-lighting Downstream Savings	
	Estimate	P-Value	Estimate	P-Value	Estimate	P-Value	Estimate	P-Value	Estimate	P-Value	Estimate	P-Value	Estimate	P-Value
Small Commercial														
Total kWh Savings			-0.003	0.865										
Downstream Total Savings													-0.027	0.402
Downstream Lighing Savings											-0.009	0.688		
Upstream Total/Lighting Savings											-0.018	0.243	-0.017	0.238
Downstream Non-lighting Savings											-0.008	0.532		
Total kWh Expenditures					0.000	0.977								
Downstream Total Expenditures									-0.029	0.288				
Downstream Lighing Expenditures							-0.018	0.426						
Upstream Total/Lighting Expenditures							-0.008	0.656	-0.009	0.636				
Downstream Non-lighting Expenditures							-0.005	0.650						
Total Employment (Economic Condition)	0.003	0.922	-0.004	0.961	0.002	0.983	0.070	0.416	0.060	0.491	0.080	0.350	0.058	0.516
Large Commercial														
Total kWh Savings			**0.219**	**0.007**										
Downstream Total Savings													0.064	0.027
Downstream Lighing Savings											-0.029	0.200		
Upstream Total/Lighting Savings											0.002	0.878	0.007	0.653
Downstream Non-lighting Savings											0.036	0.018		
Total kWh Expenditures					0.088	0.236								
Downstream Total Expenditures									0.022	0.366				
Downstream Lighing Expenditures							**-0.045**	**0.005**						
Upstream Total/Lighting Expenditures							0.010	0.601	0.025	0.209				
Downstream Non-lighting Expenditures							0.026	0.106						
Total Employment (Economic Condition)	-0.203	0.242	0.925	0.493	1.181	0.384	0.401	0.475	0.266	0.655	0.560	0.353	0.598	0.320
All Commercial														
Total kWh Savings			0.132	0.005										
Downstream Total Savings													0.024	0.248
Downstream Lighing Savings											-0.022	0.149		
Upstream Total/Lighting Savings											-0.005	0.672	-0.005	0.643
Downstream Non-lighting Savings											0.012	0.220		
Total kWh Expenditures					0.048	0.281								
Downstream Total Expenditures									0.002	0.924				
Downstream Lighing Expenditures							**-0.037**	**0.004**						
Upstream Total/Lighting Expenditures							0.004	0.736	0.007	0.607				
Downstream Non-lighting Expenditures							0.008	0.410						
Total Employment (Economic Condition)	-0.051	0.395	0.180	0.561	0.149	0.652	0.042	0.606	0.088	0.300	0.066	0.429	0.127	0.139
Industrial														
Total kWh Savings			-0.008	0.707										
Downstream Total Savings													-0.018	0.416
Downstream Lighing Savings											0.018	0.209		
Upstream Total/Lighting Savings											0.019	0.249	0.020	0.241
Downstream Non-lighting Savings											-0.018	0.201		
Total kWh Expenditures					-0.013	0.578								
Downstream Total Expenditures									-0.010	0.694				
Downstream Lighing Expenditures							**0.027**	**0.077**						
Upstream Total/Lighting Expenditures							**0.039**	**0.038**	0.041	0.044				
Downstream Non-lighting Expenditures							-0.024	0.127						
Total Employment (Economic Condition)	**0.136**	**0.091**	0.190	0.819	0.192	0.817	0.139	0.814	0.312	0.629	-0.053	0.933	0.214	0.732

7. SUMMARY REMARKS

This report provided a summary of the Year 1 pilot study research concerning the PA-Muni and PA Data C&I top-down modeling pilot studies. Through the initial year of this study, the evaluation team completed the following research activities:

- Developed an analytical framework for developing a multi-year top-down research agenda.
- Conducted an extensive review of existing literature and available data for the Massachusetts PA top-down modeling pilot studies.
- Prepared preliminary model specifications for two pilot studies.
- Obtained and reviewed the available data for constructing pilot studies.
- Estimated net preliminary C&I and residential models for the PA-Muni pilot study.
- Produced preliminary net savings estimates and confidence intervals based on PA-Muni pilot study models.
- Estimated preliminary C&I models at multiple levels of aggregation for the PA Data pilot study to advise on further development of this approach given a sufficiently long time series.

7.1 CONCLUSIONS

This section provides conclusions from the review of methods and the two pilot studies.

7.1.1 Conclusions from the review of methods

The literature review illustrates that there have been a range of approaches employed to measure programmatic impacts at different levels of analysis using a range of data inputs. Each study offers different strengths and weaknesses relative to its ability to address the desirable properties of top-down models. While none of the studies provides a single approach for isolating net programmatic impacts from other influences on consumption, these approaches provided the following guidance for the Year 1 pilot studies.

- *Length of time series* – It is important to have a long enough time series to isolate changes to programmatic activity. The studies reviewed suggest that ten or more years of data are required. In terms of the Massachusetts programs, the level of program activity began sharply accelerating about four years ago. Therefore, even if the amount of available history was extended to ten years or more, there is only a limited time series following the acceleration period to measure changes resulting from the increase in programmatic activity. While this phenomenon may limit the ability to measure programmatic impacts in the near term, top-down analysis may become more viable the longer we keep running at the higher level of program activity.
- *Account for fixed effects* – The studies employ multiple techniques to account for fixed effects, which include using the first difference in the dependent and explanatory variables, and including annual and geographic unit fixed effects terms.

For longer time series, the studies also show that it is important to consider periods of major structural changes in the economy.

- *Lagged program impacts* – It is important to consider the effect that programmatic activity in previous periods has on consumption of later periods. This is the lagged effect of programmatic activity on consumption. The effects of energy efficiency programmatic activity are not limited to the year that the activity occurred, but are actually cumulative over time, such that program expenditures made in some prior year may impact consumption in years following those expenditures. For example, if program expenditures led to installation of energy efficient lighting two or three years ago, impacts from those installations would continue to occur in the current period. Spillover or market transformation is also seen as a potential supplemental impact to program expenditures expected to occur over time after the program intervention. These considerations require using multiple lagged terms for programmatic impacts. Work by Cadmus (2012) shows that consumption in the current period may be impacted by programmatic activity of up to five years previous. This recommendation held for the PA-Muni pilot study discussed in Section 5 that found statistically significant lagged effects for up to four years.

- *Lagged consumption impacts* – There is often variance in the timing of in which savings are realized from newly installed measures. This may result from differences between the date that measures are actually installed and the date they are recorded in the PA tracking data, learning curves associated with properly using the new technology, and other factors that cause a delay in the realization of savings. The literature suggests including a team for consumption in previous periods as one approach for addressing this lag in the realization of savings.

- *Measures of differing program types* – Horowitz (2004) shows that accounting for energy efficiency program expenditures or ex-ante savings alone may provide artificially low estimates of programmatic impacts by not accounting for market transformation impacts or impacts associated with upstream programs. Market transformation programs bleed across observational units, so models that do not specifically account for market transformation may understate effects observable from cross-unit analysis In Massachusetts, the PAs have an extensive history of measuring and reporting program expenditures and savings across a wide range of measures and programs. These data will facilitate measuring changes resulting from expansion of programs covering non-lighting measures, custom measures, and upstream program offerings. These data will improve the ability of models to capture programmatic impacts within the PAs' territories in Massachusetts and regionally. However, extending the analysis to municipal territories or other states may be more difficult.

- *Use of scaled dependent variable* – The consensus among studies is that top-down models should seek to measure changes to energy consumption per unit (e.g., GSP, employee, household) in order to standardize estimates across locations and times.

- *Weather normalization* – Most of the existing studies include heating degree days and cooling degree days as explanatory variables, and use non-normalized consumption as the dependent variable. Further, these studies do not attempt to distinguish among heating and cooling impacts of programmatic activity. The recent Demand Research (2012) California pilot study marks a departure from this shortcoming, and uses normalized consumption based on utility records as the dependent variable. There are some advantages of this method but only can be accomplished if monthly data are available for individual accounts or for each geographic unit of analysis.

- *Dependent variable (consumption) data series* – Our review of data shows that it is possible to employ account-level billing records to construct the dependent variable for the PA Data model at multiple levels of analysis.[57] Because we normalize data at the account level—constructing normalized annual heating, cooling, and base loads—we are able to aggregate these series to any level of geographic resolution to construct top-down models. Further, using customer information, we can aggregate the data to different sectors to construct sector-level models. Provided the PAs are able to provide sufficiently long data series (e.g., 10+ years of billing and tracking records), these data should provide an excellent source for constructing top-down models. Finally, provided the PAs are able to make data available from other states, it is possible to combine the Massachusetts PA Data with that of other states to construct regional models.

- *Programmatic impact variables* – The PA billing and tracking data reviewed demonstrates that it is possible to construct program impact variables at any level of aggregation. Further, program data can be segmented to provide differing impacts for upstream and downstream programs, as well as categorized by measure type to isolate heating, cooling, and base load impacts.

- *Lag in program and dependent variables* – The available time series at present limit our ability to construct lag program and dependent variables for earlier time periods. The PA-Muni model was able to assess the impact of programmatic from previous periods on current consumption, however, absent the necessary data series from previous periods, we the PA Data model was not.

- *Exogenous variables* – There are a variety of sources that provide the necessary data to account for exogenous influences on consumption. The choice of data source and available series depends largely upon the desired level of aggregation. Finer resolution in the level of observation restricts access to some variables, as some data are not reported as proprietary information, such as sales volume and payroll.

- *Level of analysis* – The desired level of analysis for detecting programmatic effects should extend beyond Massachusetts. While the current year's analysis is only limited to measuring changes that occur within Massachusetts, to effectively capture sufficient differentiation in programmatic activity the longer-term study could be expanded to include neighboring states included in the PAs' service territories, including Connecticut, Rhode Island, New Hampshire, Maine, and New York. However, expanding to a regional model will likely result in some states having less data available than in Massachusetts. Because data availability is a primary concern for this approach, the prospect for expanding to a regional model should be weighed relative to the necessary trade-offs regarding data availability.

---

[57] Several of the prior top-down modeling efforts used aggregate data as reported to the Energy Information Agency (EIA).

- *Source of consumption and program tracking data* – PA billing and tracking databases provide the most robust set of data for conducting top-down models. Because records can be captured at the account level and cleaned and weather normalized, many issues associated with consistency in measuring and reporting data discussed in this memo can be avoided. However, relying on billing and tracking records does restrict the analysis to data that can be obtained from utilities. Our analysis showed that the data provided by the PAs to the C&I evaluation team to date is appropriate for conducting the PA Data pilot study, however, the length of time series was not sufficiently long to provide a true assessment of the model specification.

7.1.2 Pilot study conclusions

The two pilot studies presented in this report demonstrate that top-down modeling may provide a valuable tool in the set of tools used to evaluate net-energy impacts associated with energy efficiency programs.

The two approaches used in the pilot studies have differing strengths and weaknesses in terms of addressing the desirable properties of top-down models and modeling concerns identified in the existing literature. The PA-Muni pilot study employed a relatively long time-series, 15 years, which allowed the model to examine possible cumulative effects of programmatic activity on consumption over time through use of various lagged program expenditure terms. This was a key finding of the literature review, and the model results indicated that these lagged terms were, in fact, instrumental in developing a model that produced statistically significant results. The PA Data model had a much more limited time series (three years) and consequently was not able to account for the cumulative impact of programmatic activity. Similarly, the PA-Muni study was able to address a number of other influential factors related to the time-series, which the PA Data study was not able to address, such as the impact of building codes, technology trends, and time-specific fixed effects. Due to the overarching restrictions on the PA data resulting from the limited time series, the evaluation team did not address a number of other modeling concerns that may also limit the success of this technique, such as industry-level segmentation, the impact of building codes and technology trends, and the most appropriate treatment of weather normalization.

Both modeling approaches rely on differences in program activity across geographies and time to isolate the effect of program activity on consumption. The PA-Muni model contrasted consumption in the PA territories, which have relatively high levels of programmatic activity, to consumption in municipal utility territories, which have relatively low levels of programmatic activity. This contrast provides a stronger basis for measuring net impacts. In effect, the low-program muni territories represented a comparison area that was used to remove naturally occurring energy savings from gross impacts. Because the PA Data model relies exclusively on data within the PAs' territories, the PA Data models have a weaker program signal in their contrasts across time and units; the PA Data models have the advantage of more detailed data that can help in controlling for non-program factors and support the isolation of program attributable impacts from naturally occurring savings.

The PA Data model offers an approach to address many questions that are important for planning, policy, and implementation of energy efficiency programs, which the PA-Muni approach cannot address. Because the PA Data models were developed from account-level billing and tracking data, separate models can be developed to examine the impact of differing program offerings, or the relative contribution of various customer segments to savings. While the models were not statistically significant, the PA Data pilot study showed that the ability to model different customer segments (i.e., large commercial, small commercial, and industrial customers) provide differing measures of programmatic impacts. Further, the ability to break out various measure and program types may also influence savings estimates. This information is important for policy, planning, and implementation, as it allows for the development and implementation of targeted program offerings. The PA Data approach provides this level of flexibility in modeling, while the PA-Muni approach does not. Both studies face differing, but substantial data limitations.

In summary, the evaluation team two forms of analysis though this first year of the top-down research study. While differing in their technique, these studies followed best practices identified in the literature. The PA-Muni model benefitted from a wider range of variation of activity and a longer time series. Its results indicate that this is a promising approach, but further work is needed to ensure that robust results can be obtained. The PA Data model had the advantage of being able to consider a finer level of granularity, but was disadvantaged by generally less variability of program activity across geographic units within PA territories, and a shorter time series available at this time. The PA Data approach has not yet been applied to the residential sector, as the data were not available earlier, but it will be shortly. For both modeling approaches, data assembly was a non-trivial effort. The conclusion of this research is that top-down modeling as a set of modeling approaches is promising, but needs more study.

A. Conclusions from the PA-Muni data pilot study

While the findings from this pilot study are preliminary, initial model results look promising as a supplemental approach to the bottom-up methods used to estimate net-energy impacts. A number of model results—both for residential and C&I—indicate that energy efficiency program expenditures had a statistically significant effect on reducing electricity consumption, but the effect was less consistent for the C&I sector. While additional research is necessary to refine the PA-Muni top-down models, understand the stability of the model results, and reduce the size of the confidence intervals and model specification uncertainties, this technique appears to offer a potential means of validating the bottom-up estimates of gross and net savings. The team has identified a number of next steps for model refinement that should help improve the precision of the savings estimates.

The results showed that savings estimates are sensitive to model specification, particularly the inclusion of various lagged measures of programmatic activity. Among the various residential model specifications tested, the fixed-effects model with four lags for energy efficiency expenditures appeared to perform the best, as the coefficients of the energy efficiency expenditures variables for the current year and the past four years were jointly statistically significant. The model accounted for the lagged impact of energy efficiency program expenditures on electricity consumption and the leakage of PA lighting program rebate dollars to municipal utility service territories. The findings indicate that the impact of cumulated energy efficiency program expenditures on current consumption increases with the number of previous years included in the cumulated sum. This reflects the importance of including lagged program year effects in the model.

Among the various C&I model specifications tested, the fixed-effects model with three lags for energy efficiency program expenditures appeared to perform the best. Adding more lags to the model yielded results that were not in the expected direction. The estimated impact of one dollar spent for the C&I energy efficiency programs was somewhat smaller than that for the residential sector, but the difference was not statistically significant. Finally, top-down savings estimates from this model were very close to the corresponding reported annual savings for the PA C&I programs, but savings estimates were highly dependent on model specification, which warrants further study.

This study also draws attention to an inherent limitation of macro-consumption methods. While the team was able to detect energy savings, further model refinement is necessary to reduce the confidence intervals around the top-down estimates. The team has identified a number of next steps for model refinement that should help improve the precision on savings estimates.

In addition, traditional bottom-up approaches measure energy impacts at the individual customer level, capturing customer-level detail that is beneficial in program design, marketing, and policymaking, such as the importance of certain program designs or measure offerings to overall portfolio savings and measurement of free ridership and spillover effects. While explanatory variables measure the relative influence of demographic and economic factors on savings, top-down estimates based on data aggregated at too high of a geographic level (i.e., PA/utility level) may lose the ability to provide meaningful estimates of variables important to all interested parties. This conclusion suggests that this top-down method offers an additional tool for triangulating the overall impact of energy efficiency programs, but this technique is not a one-size-fits-all approach for addressing all program evaluation needs.

An important challenge for this study was to collect consistent electric program data across all PAs and municipal utilities. While the evaluation team attempted to collect detailed time-series data on program activity, the only consistent piece of data that the team was able to gather across all PAs and municipal utilities was the annual total electric program expenditures.

If the PAs and the EEAC decide to move forward with the second phase of the study, which would entail the collection of PA energy consumption and energy efficiency program data at the town or city level, then these data could potentially help increase the precision of the savings estimates and thus improve both the residential and C&I models. Moreover, these PA data would allow commercial sector data to be separated from industrial sector data, which should improve the results because the electricity consumption for the commercial sector should be more stable than that of the combined commercial and industrial sectors due to high volatility of consumption in the industrial sector. Finally, if the gas PAs can provide gas consumption and energy-efficiency program data at the town or city level, then comparable gas macro-consumption models could be run to estimate the impact of gas program activity on gas consumption in the residential and commercial sectors in Massachusetts.[58]

[58] These data would also allow for developing a model to test the impact of the PA residential upstream lighting programs on residential gas consumption. If the interactive effects are large, this test might provide corroborating evidence to suggest that further research is needed to ensure that the most accurate assumptions for program planning and savings claims are made.

B. Conclusions from the PA Data pilot study

This study sought to determine whether top-down methods should play a role in the overall portfolio of attribution methods both in terms of the recommended role on an ongoing basis as well as the methodological approaches that are recommended. The methods review portion of this study concluded that top-down modeling may provide an additional tool in the set of tools used to evaluate the portfolio of programs. However, the top-down approach cannot replace bottom-up approaches, as bottom-up techniques provide much information that top-down techniques cannot provide. Information pertaining to the relative contribution of different activities to overall savings can assist in the allocation of resources across the portfolio of programs, or help with program design. Such information cannot be obtained from top-down approaches. Moreover, the review of data suggested that top-down techniques face a variety of challenges pertaining to the reporting and availability of data that limited the effectiveness of these techniques. The PA Data model confirms that data availability was a primary obstacle to successful estimation of the models presented in this report. While our review of methods indicated that load forecasters within each PA interviewed used relatively simplistic models to estimate demand, forecasters reported that they were not able to tease out program effects from their load forecasts.

The model results estimated in this study were consistent with this finding; however, these results were limited to just three years of data. One factor that may lead to meaningful estimates is the availability of a longer time series. With only three years of data, the PA Data pilot study portion of this report shows inconclusive evidence that the approach we employed is able to detect programmatic impacts. Our analysis demonstrates the ability to construct the necessary variables at the desired levels of aggregation, and the ability to systematically test a variety of models. Some of the models showed statistically significant parameter estimates for measures of either programmatic and/or economic activity; however, these results were not consistent across model specifications or levels of geography.

One could conclude that the statistical significance of parameter estimates for some models is an indication that the models would perform well given a sufficiently long time series. However, one could also argue that the significance of terms measuring programmatic activity is the result of noise in the model, and the true models are ones in which there are no program effects. While our analysis does not provide sufficient information to make a determination that program effects can be detected with certainty, some model results do show statistically significant parameter estimates on the program variables. Further, our review of the available literature suggests that effective top-down modeling of energy impacts requires a sufficiently long time series to account for:

- *Variation in the level of program data over time* – Our time series included only three years of data, which all occur during a period of economic recovery and rapid increase in programmatic activity.
- *Multiple lags in programmatic activity* – Previous research, as well as the PA-Muni pilot study, illustrate the importance of using multiple lags in both the program variables and dependent variable.
- *Use of first-difference in the dependent and independent variables* – By including only three years of data in the model, the first-difference models included in this study contain only two years of data for unit of observation.

Absent these measures, it is not surprising that the model results did not provide statistically significant parameter estimates, that the results were not consistent across levels of aggregation, and that the results were not stable in terms of the significance of variables or their sign. Despite the lack of significant results at this point, the evaluation team believes that the PA Data pilot study model approach will likely improve given a long enough time series. We draw this conclusion based on the following evidence:

1. Loughran and Koulick (2007) and Violette (2014) demonstrated that successful top-down models with at least ten years of data can successfully account for programmatic impacts.
2. The PA-Muni model, which includes more than ten years of data over the same population, was able to provide statistically significant savings estimates. Given the PA Data model has the ability to examine impacts associated with more specified programmatic activity, it is likely that given a sufficiently long time series, the PA Data model would also produce significant results.
3. The PA Data model is able to capture variation across program and customer types that provides valuable information for program planning and implementation, and allows program evaluators to determine the effectiveness of differing program offerings and/or marketing strategies.

However, compiling a sufficiently long historical time series retrospectively would be costly, and may not be possible due to limitations in electronic record keeping. Therefore, a more practical approach may be to construct the historical series back five years and continue collecting the necessary data going forward for future analysis.

Apart from these time-series related limitations, the following factors may also be responsible for the lack of significance in the model:

- The current C&I models do not account for differences in consumption and programmatic activity by industry, which the literature has shown is an important factor for isolating program impacts.
- The model results could be impacted, in part, by the normalization process. However, the evaluation team did test a set of models that used non-normalized consumption as the dependent variable and included HDD and CDD as independent variables. These models did not perform better than the models using normalized consumption, and did not have statistically significant parameters for the weather terms.
- The models did not include terms to measure the impacts of changes to building codes or technology.

7.1.3 Limitations

This section reviews important limitations to the analysis.

A. Limitations of the PA-Muni pilot study

- *Aggregate data* – The primary limitation of this study is that it relies on data aggregated at the PA or muni level to estimate energy impacts. Consequently, separate models cannot be developed for different customer segments or program offerings. Lack of segmentation may be partially responsible for the wide confidence intervals for savings estimates.

- *Municipalities are not a true comparison area* – While there is relatively limited programmatic activity in the muni territories, there is still some activity. Moreover, there is likely leakage or cross-unit spillover from the PA to muni territories. These factors limit the ability of the model to adjust for naturally occurring adoption of energy efficiency technology.
- *Data availability* – The evaluation team expended considerable effort to capture programmatic data from the PAs and municipal utilities. While the publicly available energy consumption data was relatively clean, the publicly available programmatic activity data was not. Obtaining the program expenditure data required considerable effort from the evaluation team, the PAs, and municipal utilities. A number of municipal utilities were not able to obtain the necessary data, as it would require coding data reported on paper forms.

B. Limitations of the PA Data pilot study

- *Fuel prices not reported at the same level of granularity as unit of analysis* – The evaluation team did not identify any data for actual average electricity prices ($/kWh) at the county or town level. The DNV GL billing data set contained rate codes, and billing amounts could therefore be imputed, but there were many missing values and other data quality concerns.
- *Absence of lagged program activity and consumption variables* – The literature review identified the importance of incorporating lagged program and consumption variables into the models. Because the existing time series was limited to just three years of data, we were not able to construct lagged variables using the consumption and program tracking data. The evaluation team attempted to construct lagged series based on data available through the PAs' annual reports; however, we were unable to construct a series that did not introduce bias into the model.
- *Absence of building codes* – The evaluation team attempted to construct variables to account for the impact of building codes on consumption; however, due to the limited time series, there was insufficient variation in the building code data in include in the model.
- *Limited time series during periods of rapid expansion of both economic and programmatic activity* – Our review of the available consumption, program tracking, and economic activity variables revealed a fundamental limitation of the present analysis. During the three years of observation for this study, the three critical series for the analysis underwent a period of rapid expansion. Figure 7-1 presents the change in employment and NAC for 2012 and 2013 relative to 2011. Figure 7-2 presents the change in program expenditures and ex ante savings for 2012 and 2013 relative to 2011. Given the limited time series, it is likely that the model results will be impacted by the corresponding increase in these three series. Without a longer time series, or substantial variation between observational units, it is likely that the model will not be able to differentiate between increases in programmatic activity and reductions in consumption. It is important to have a long enough time series to

isolate changes to programmatic activity. The studies reviewed suggest that ten or more years of data are required. In terms of the Massachusetts programs, the level of program activity began sharply accelerating about four years ago. Therefore, even if the amount of available history was extended to ten years or more, there is only a limited time series following the acceleration period to measure changes resulting from the increase in programmatic activity. While this phenomenon may limit the ability to measure programmatic impacts in the near term, top-down analysis may become more viable the longer we keep running at the higher level of program activity.

Figure 7-1. Percent Change in Annual Employment and Normalized Annual Consumption (2011–2013)

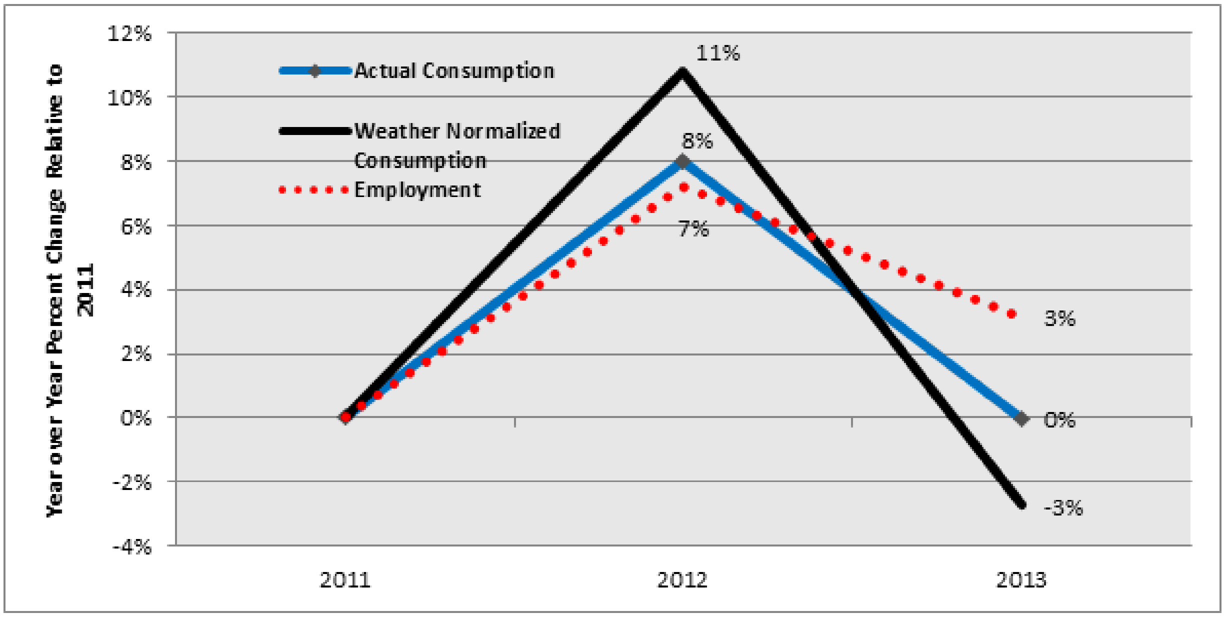
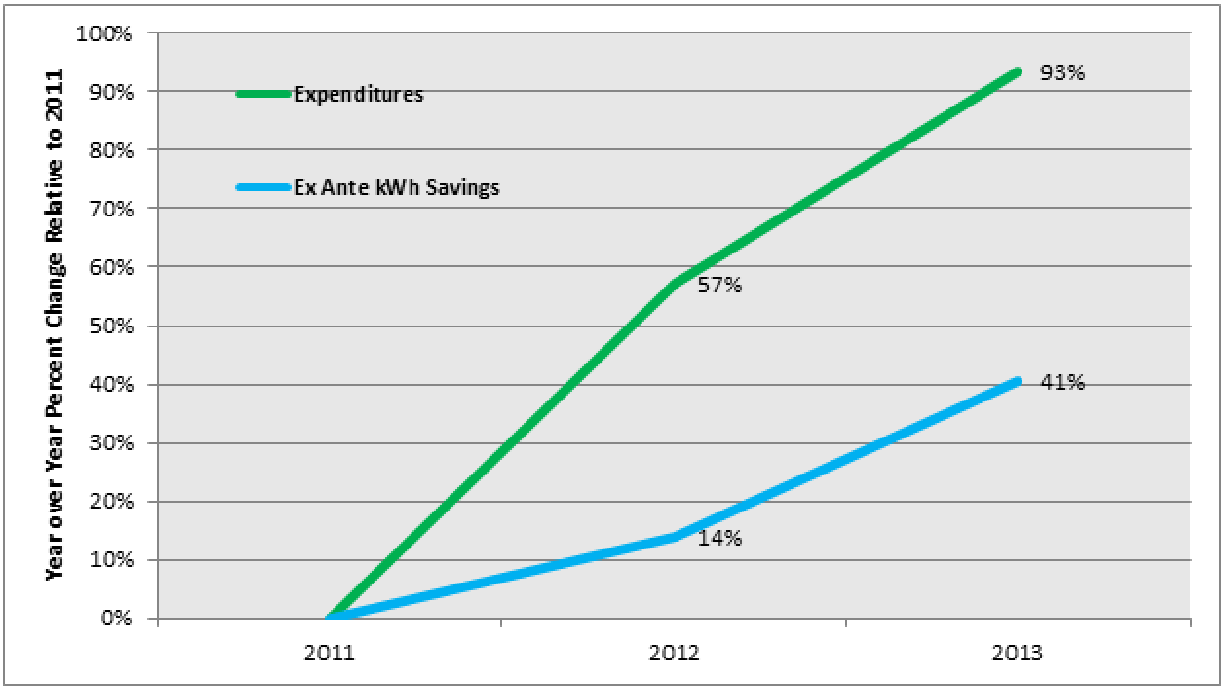


Figure 7-2. Percent Change in Program Expenditures and Ex-ante Savings (2011–2013)



- *Limitations in use of per capita income* – Income is not very indicative of C&I economic conditions, as it can be skewed by individuals with relatively high salaries, such as CEOs. Consequently, this variable is not used by the load forecasters at the PAs.
- *Isolating industry- or sector-level differences* – There may be considerable variation in the savings and consumption by industry sector. However, economic series by sector are only available at the county level or for the major metropolitan areas.[59] Population, a variable that is available at all levels of aggregation, could theoretically serve as a measure of market size, but it is more closely associated with residential consumption than commercial or industrial, which is also true of per capita income. The evaluation team believes that per capita income and population is likely to be correlated to employment if they are both included jointly in a statistical model. Population is available at the town level or census-track level from the American Community Survey data set.

7.2 RECOMMENDATIONS

This Year 1 top-down research provided a number of key recommendations for conducting the next phase of pilot studies in Massachusetts. We summarize these as follows:

- Continue refinement of the PA-Muni model to investigate the stability of models and possible changes to model specification that may reduce confidence intervals as outlined above.

[59] http://www.bls.gov/cew//.

- For the PA Data model, continue to collect data through the C&I database to extend the available data series to include five years of consumption and program tracking data, then continue collecting the necessary data going forward for future analysis. Continue to refine the existing models to further explore approaches to weather normalization, industry segmentation, and inclusion of other key explanatory variables such as technology trends; and incorporate multiple lag periods of the program and consumption variables.

APPENDIX A: WORKS CITED

Auffhammer, M., University of California, B., Blumstein, C., & Fowelie, M. (2007). Demand-Side Management and Energy Efficiency Revisited.

CADMUS Group, I. (2012). *CPUC Macro Consumption Metric Pilot study (Final Report).* Prepared for the California Public Utilities Commission.

Demand Research, L. (2012). *Macro Consumption Metrics Pilot study Technical Memorandum – Preliminary Findings.* Prepared for the California Public Utilities Commission.

DNV GL. (2014, May 22). Abbreviated Review of Methods for the Draft Top Down Modeling Methods Study. Prepared for the Massachusetts Program Administrators and EEAC Consultants.

DNV GL. (2014, June 14). Initial PA Data Top Down Model Specification. Prepared for the Massachusetts Program Administrators and EEAC Consultants.

Feldman, S. L.-W. (2005).

Feldman, S. L.-W. (2005). Modelling the Effects of U.S. ENERGY START Appliance Programs.

Fels, M. F. (1986). Fels et al. 1995. *Energy Buildings: Special Issue Devoted to Measuring Energy Savings: The Scorekeeping Approach*.

Fels, M., Kissock, K., Marean, M., & and Reynolds, C. (1995). *PRISM (Advance Version 1.0) Users' Guide, Center for Energy and Environment Studies*.

Horowitz, M. J. (2001). Economic Indicators of Market Transformation: Energy Efficient Lighting and EPA's Green Lights. *The Energy Journal, Vol. 22, No. 4*.

Horowitz, M. J. (2004). Electriciy Intensity in the Commerical Sector: Market and Public Program Effects. *The Energy Journal, Vol. 25, No. 2*.

Horowitz, M. J. (2007). Changes in Electricy Demand in the United States from the 1970s to 2003. *The Energy Journal, Vol. 28, No. 3*.

Horowitz, M. J. (2010). Measuring the savings from energy efficiency policies: a step beyond program evaluation. *Springer Science & Business Media B.V. 2010*.

Jacobsen, G. D., & Kotchen, M. J. (2009). Are Building Codes Effective at Saving Energy? Evidence From Residential Billing Data in Florida. *Conference on Green Building, the Economy, and Public Policy.*

Loughran, D. S. (2004). Demand-Side Management and Energy Efficiency in the United States. *The Energy Journal, Vol. 25 No. 1*.

NMR Group, I. a. (2011). Results of the Multistate CFL Modeling Effort . *Submitted to: EEAC Consultants and the Massachusetts Program Administrators*.

Parfomak, P. W., & Lave, L. B. (1996). How Many Kilowatts are in a Negawatt? Verifying Ex Post Estimates of Utility Convservation Impacts at the Regional LEvel. *The Energy Journal, Vol. 17, No. 4*.

Rivers, N., & Jaccard, M. (2011). Electric Utility Demand Side Management in Canada. *The Energy Journal, Vol. 32, No. 4*.

Rosenberg, M. (2003). The Impact of Regional Incentive and Promotion Programs on the Market Share of ENERGY STAR® Appliances.

Violette, D. (2012). *Review of a Top-down Evaluation Study.* Navigant.