Status and Directions for Top-Down Work (TXC 43)

FINAL

March 19, 2018

SUBMITTED TO:
Massachusetts Program Administrator Evaluation Group

SUBMITTED BY:
DNV GL, Inc.
NMR Group, Inc.
Table of Contents

EXECUTIVE SUMMARY................................................................................................................ 1
PURPOSE OF THIS DOCUMENT ................................................................................................. I
THE VALUE AND USE OF TOP-DOWN ESTIMATION ............................................................ I
ACTIVITIES .................................................................................................................................. II
PATH FORWARD .......................................................................................................................... III

SECTION 1  INTRODUCTION .................................................................................................. 1
1.1 PURPOSE OF THIS DOCUMENT ......................................................................................... 1

PART 1: STATUS AND DIRECTIONS FOR TOP-DOWN WORK ............................................. 1
1.2 WHAT IS TOP-DOWN ANALYSIS? ....................................................................................... 1
1.3 GENERAL ELEMENTS AND FORM OF TOP-DOWN ANALYSIS ...................................... 2
1.4 THE APPEAL OF TOP-DOWN ESTIMATION FOR NET SAVINGS .................................... 3
1.5 CHALLENGES AND LIMITATIONS .................................................................................... 5
  1.5.1 Accuracy ...................................................................................................................... 5
  1.5.2 Ability to Capture All Net Savings Components .......................................................... 6
  1.5.3 Level of Detail ............................................................................................................. 6
  1.5.4 Persistence .................................................................................................................. 7
1.6 USE OF TOP-DOWN ESTIMATION IN OTHER REGIONS ............................................... 7
1.7 THE VALUE AND USE OF TOP-DOWN ESTIMATION ..................................................... 7

SECTION 2  KEY WORK TO DATE ......................................................................................... 9
2.1 WORK PRODUCTS ................................................................................................................. 9
2.2 SUMMARY OF APPROACHES EXPLORED ..................................................................... 10
2.3 MAJOR FINDINGS AND LESSONS LEARNED FROM WORK TO DATE .......................... 12
  2.3.1 Ability to Provide Useful Estimates .............................................................................. 12
  2.3.2 Limitations and Lessons from the Residential Modeling ............................................ 12
  2.3.3 Commercial-Industrial Analysis Results ...................................................................... 13
  2.3.4 Data Challenges ......................................................................................................... 13
  2.3.5 Technical Issues Explored ......................................................................................... 14

SECTION 3  NEW IDEAS ........................................................................................................ 16
3.1 NEW HIGH-LEVEL APPROACHES ..................................................................................... 16
  3.1.1 Individual Customer Observations .............................................................................. 16
  3.1.2 Machine Learning Methods ....................................................................................... 17
3.2 NEW IDEAS FOR IMPROVEMENTS WITHIN THE SAME GENERAL TOP-DOWN MODELING APPROACH .............................................................................................................. 18

SECTION 4 POTENTIAL PATHS FORWARD .................................................................................................................. 20

4.1 DATA NEEDED .................................................................................................................................................... 20
  4.1.1 PA Data ...................................................................................................................................................... 20
  4.1.2 Municipal Data ........................................................................................................................................ 21
  4.1.3 Data from Other Sources ........................................................................................................................ 21
  4.1.4 Compiling and Maintaining Data ............................................................................................................... 22

4.2 CONTINUING TOP-DOWN ANALYSIS OVER TIME ............................................................................................ 24

4.3 IMPROVEMENTS INCORPORATED RELATIVELY EASILY WITHIN THE CURRENT FRAMEWORK ................................................................. 24

4.4 IMPROVEMENTS REQUIRING GREATER EFFORT WITHIN THE CURRENT FRAMEWORK ............................................................ 25

4.5 METHODS REQUIRING ADDITIONAL DEVELOPMENT .................................................................................. 26

SECTION 5 KEY ANALYTIC FINDINGS FROM WORK TO DATE .................................................................................. 27

PART 2: ADDITIONAL TECHNICAL DETAIL ............................................................................................................. 27

5.1 MODELS DEVELOPED .................................................................................................................................... 27
  5.1.1 Residential Model ..................................................................................................................................... 27
  5.1.2 Commercial-Industrial Model .................................................................................................................. 29
  5.1.3 Basic Sensitivity Testing .......................................................................................................................... 30
  5.1.4 Exploring the Effect of the Recession ...................................................................................................... 32

5.2 UPSTREAM LIGHTING PROGRAM EFFECTS ...................................................................................................... 33

SECTION 6 EARLIER GUIDELINES FOR IMPROVING MODELING ........................................................................... 35

6.1 WEATHER MODELING ...................................................................................................................................... 35

6.2 LAG EFFECTS .................................................................................................................................................... 35

6.3 RECESSION ....................................................................................................................................................... 35

6.4 FIXED EFFECTS ................................................................................................................................................ 36

6.5 ENDOGENEITY OF ELECTRICITY AND GAS PRICES ..................................................................................... 36

6.6 ENDOGENEITY OF PROGRAM SPENDING .................................................................................................... 36

6.7 ERROR BOUNDS .............................................................................................................................................. 36

6.8 CONSTRUCTING PORTFOLIO SAVINGS ......................................................................................................... 37

6.9 SELF-SELECTION .......................................................................................................................................... 37

6.10 SPIral OVER .................................................................................................................................................. 37

SECTION 7 IDEAS FROM NEW CONVERSATIONS ................................................................................................. 38
7.1 PA FORECASTING APPROACHES .......................................................... 38
7.1.1 Forecasting Approaches ............................................................... 38
7.1.2 Implications .................................................................................. 39
7.2 SUGGESTIONS FROM THE ACADEMIC EXPERT AND RELATED INTERNAL TEAM DISCUSSIONS ......................................................... 39

SECTION 8 REFERENCES .......................................................................... 42

APPENDIX A PA-MUNI TOP-DOWN MODEL RESULTS FROM THE FIRST TOP-DOWN EFFORT ................................................................. A1

Figures

FIGURE 1: CONSUMPTION PER CUSTOMER (MWh) FOR MASSACHUSETTS AND THE REST OF NEW ENGLAND, 2001 THROUGH 2016 ............................................................ 2
FIGURE 2: RESIDENTIAL ELECTRICITY CONSUMPTION PER CUSTOMER (kWh) .............................................................. 28
FIGURE 3: RESIDENTIAL ELECTRIC PROGRAM EXPENDITURES PER CUSTOMER ($) .............................. 28

Tables

TABLE 1: SAVINGS ESTIMATES FROM RESIDENTIAL PA-MUNI MODELS WITH VARYING TREATMENT OF CUMULATIVE SAVINGS EFFECTS .............................................. 29
TABLE 2: SAVINGS ESTIMATES FROM COMMERCIAL-INDUSTRIAL PA-MUNI MODELS WITH VARYING TREATMENT OF CUMULATIVE SAVINGS EFFECTS .............................................................. 30
TABLE 3: RESIDENTIAL PA-MUNI MODEL RESULTS WITH INDIVIDUAL YEAR PROGRAM EXPENDITURES .................................................................................. A1
TABLE 4: RESIDENTIAL PA-MUNI MODEL RESULTS WITH CUMULATED PROGRAM EXPENDITURES .................................................................................. A3
TABLE 5: C&I PA-MUNI MODEL RESULTS WITH INDIVIDUAL YEAR PROGRAM EXPENDITURES .................................................................................. A5
TABLE 6: C&I PA-MUNI MODEL RESULTS WITH CUMULATED PROGRAM EXPENDITURES A7
Executive Summary

Purpose of this Document

Top-down methods for estimating net savings offer a supplemental perspective to the traditional “bottom-up” methods for the Massachusetts Program Administrators’ energy-efficiency portfolios. Top-down estimation of net savings refers to an approach to estimating net impacts of an energy-efficiency portfolio by looking at aggregate consumption over time, or across geographic areas, or both (e.g., across service territories and across years).

The top-down work conducted by the Special and Cross-Cutting evaluation team to date has shown some promise, and has also faced some challenges. This document provides a high-level summary of the top-down work undertaken to date, including lessons learned, and offers a potential path forward for future work. It also offers supporting technical details in the second part of the document.

The Value and Use of Top-Down Estimation

The value of top-down estimation is that it can provide high-level validation of program effects based on concrete observations of the key quantity targeted by the programs, namely energy consumption. The result does not depend on survey responses to hypothetical questions, nor on engineering estimates based on limited information about equipment and its use. All factors that need to be taken into account – from free-ridership to market effects to takeback to interactive effects among measures and systems, to gross savings corrections and installation rates – are included in the overall observed consumption and changes in consumption. If we can analytically determine the effect of the program portfolio on overall consumption – that is, pull out the difference between the with-program and without-program scenarios – this total net effect is captured. From a portfolio-level impact perspective, there is no need to attempt to estimate separate, hard-to-measure effects.

The top-down savings estimate gives the cumulative effect of a program portfolio over multiple years. This is a useful perspective for planning and assessing a long-term effort. The analysis also provides a confidence interval for the estimated multi-year portfolio-level savings.

Nonetheless, top-down methods are still dependent on debatable assumptions and are subject to uncertainties and challenges, such as:

- **Accuracy.** Sources of inaccuracy and uncertainty in top-down results include the following:
  - Model specification uncertainty – different reasonable specifications can give very different results
  - Estimation precision – error bands for any given model are typically wide
o Data limitations – models require consistent data on key explanatory variables across the geographies and time frame of interest

o Data anomalies – even if data are accurate, there is the potential for pockets of anomalous observations to distort the results.

- Completeness. Spill-over, self-selection, and untracked upstream savings may not be fully accounted for, depending on the data available and structure of the analysis.

- Lack of detail.
  o Top-down analysis does not provide separate estimates of individual savings components or adjustments, including free-ridership, participant and non-participant spillover, take-back, interactive effects, and errors in gross savings estimates or EULs.
  o Top-down analysis does not provide indicators of the effectiveness of programs for individual geographies or individual years, or provides insights into what works and does not work for individual programs.

The results of the top-down estimation need to be considered in the context of these known limitations. The goal of the top-down analysis should not be viewed as providing a single definitive estimate of savings achieved by a portfolio of programs. Rather, the goal of the analysis is to provide another perspective to substantiate the value of the portfolio. To the extent the top-down estimate differs significantly from the bottom-up estimate, it is worth exploring what might have been omitted from either estimate that needs to be considered.

In particular, if top-down analysis produces a savings estimate substantially larger than the corresponding sector-level bottom-up evaluated estimate, that finding alone is not good evidence of large-scale unaccounted-for market effects. A best effort should first be made to estimate the market effects and upstream impacts based on a solid program theory, then to explain the remaining discrepancies.

**ACTIVITIES**

Developing this report included the following steps:

- Preparation of an initial summary of findings and lessons learned from top-down work in Massachusetts to date.

- Discussion of the top-down approaches with PA forecasting staff to share insights and ideas.

- Review of the initial work by an independent academic expert economist, with follow-up discussions.
PATH FORWARD

The report provides a path forward for top-down estimation of net savings in Massachusetts. This path includes the following:

1. Suggestions for ongoing data compilation, leveraging the existing customer profile database work, to facilitate periodic top-down analysis

2. Suggestions for a next round of top-down analysis in 2018 or 2019 building on the previous models and databases, and incorporating certain technical improvements

3. Suggestions for additional improvements that could be explored in future work, under the same general analysis approach

4. Some alternative high-level approaches for future consideration
Section 1  Introduction

1.1  Purpose of This Document
Since 2013, the Massachusetts Special and Cross-Cutting Evaluation team has been exploring “top-down” methods for estimating net savings from the Program Administrators’ (PAs) energy-efficiency portfolios. These methods offer a supplemental perspective to the traditional “bottom up” savings estimates. To date, the top-down work has shown some promise and has faced some challenges.

This document provides a high-level summary of the top-down work undertaken to date, including lessons learned, and offers a potential path forward for future work.

Part 1 is designed primarily for readers who want a big-picture understanding of what top-down savings estimation is, the value it can bring, and its limitations. Part 1 may be of interest to program managers, regulators, or others who need a general appreciation of the approach without delving into technical details. Part 2 (starting with Section 5) builds on Part 1 and summarizes some of the more technical findings and guidance from the work to date. Part 2 also discusses new ideas. For ease of exposition, some details of interest to technically inclined readers are included in Part 1, set off in text boxes.

Part 1: Status and Directions for Top-Down Work

1.2  What is Top-Down Analysis?
Top-down estimation of net savings refers to an approach to estimating net impacts of an energy-efficiency portfolio by looking at aggregate consumption over time, or across geographic areas, or both (for example, across service territories and across years). The idea is to account analytically for all the other factors that affect consumption, so that the effect of the energy-efficiency program activity can be isolated.

The term “top-down” is used in this context to contrast the approach with the more common “bottom-up” approach. The bottom-up approach determines savings for individual measures installed at individual customer locations, and aggregates these to the program and portfolio levels. Included in the bottom-up approach are estimates of gross savings for individual units, and estimates of the influence of the programs on measure adoption by program participants as well as outside of programs.

Figure 1 shows electric consumption per capita for Massachusetts and for the rest of New England for the past several years. Beginning roughly with the major increase in Program Administrator (PA) program funding initiated in 2009, Massachusetts has exhibited greater declines in per capita consumption than its neighboring states – averaging a drop of 1.4% per year in Massachusetts, compared to 0.7% per year for the rest of New England. The key question is, how much of this differential reduction is attributable to the Massachusetts
energy-efficiency programs? How much is related to other factors, such as changes in industrial base, growth of customer-sited solar generation, or improved stock efficiency due to natural turnover of equipment and buildings? The top-down analysis attempts to separate the effect of the programs from the combination of these other factors.

**Figure 1: Consumption per Customer (MWh) for Massachusetts and the Rest of New England, 2001 through 2016**

Source: developed from EIA data.

### 1.3 **General Elements and Form of Top-Down Analysis**

Top-down analysis can take many forms. The simplest type of analysis is a comparison between two regions. For instance, in Figure 1, we could use the consumption trend for the rest of New England as the baseline against which to measure the Massachusetts changes, ascribing the entire difference to the difference in efficiency program spending. However, we know there are many other factors that could affect each region’s consumption pattern, including energy-efficiency programs in other states. It is preferable to use an approach that accounts for these other factors to the extent possible.

Conceptually, the top-down analysis we consider involves the following general steps:

- Compile aggregate consumption data, together with a measure of customer population size (e.g., number of customers, employment, or population), for a set of geographic units (e.g., service territories) over a set of time periods (e.g., years).
- For the same geographic units and time periods, compile economic data (e.g., fuel prices and GDP), weather data, and a measure of program activity (e.g., spending, or ex ante savings).
- Fit a model across geographies and time to describe aggregate consumption as a function of the economic factors and the program activity.
• Use the program activity coefficient(s) as the basis for estimating the net savings due to the program portfolio.

Key economic variables typically include a measure of economic activity, such as GDP or personal income, and energy price. Data on uptake of disruptive technologies, such as distributed generation or electric vehicles, could also be included.

Alternative approaches exist that look only across regions for a given time period or change between two periods, or that look only across time for a single geography. Looking across geographies and time periods jointly can improve the ability of the analysis to isolate the program activity effect of interest.

1.4 THE APPEAL OF TOP-DOWN ESTIMATION FOR NET SAVINGS

There are several reasons top-down estimation for net savings is conceptually appealing.

**Savings based on direct observation of the key quantity targeted by energy efficiency portfolios.** The goal of energy-efficiency programs is to reduce energy consumption in the territory where they operate. We would therefore like to see that consumption is in fact reduced, in comparison to what it would have been without the programs. If we look over enough time, all the direct and indirect effects of the efficiency portfolio are bundled into the observed aggregate consumption. Increased adoption of efficiency measures due to program efforts, via non-free-riding participants, as well as via spillover, shows up in the aggregate consumption. While the program ex ante estimates of the gross savings for a particular

---

**Technical Detail**

An illustrative model form would be as follows:

\[
E_{it} = \beta_o + \beta_p EE_{it} + \sum_j \beta_j X_{jit} + \epsilon_{it}
\]

where

- \(E_{it}\) = Energy consumption per unit, for area i, time t
- \(EE_{it}\) = EE program activity metric, for area i, time t
- \(X_{jit}\) = economic variable j, for area i, time t
- \(\epsilon_{it}\) = residual
- \(\beta_o, \beta_p, \beta_j\) = regression coefficients estimated from the data.

The coefficient of the program activity variable tells us the incremental change in aggregate consumption per incremental change in the energy-efficiency activity. Multiplying this coefficient by the program activity for each year gives the estimated savings.

The specific model form applied is typically somewhat more complicated. The terms accounting for different factors may be nonlinear. The model fit may be on a logarithmic basis. Multiple variables may be included for different kinds of program activity. Lag terms may be included as predictors.
measure may be over- or understated, the actual savings from the measure is included in the observed aggregate consumption. Effects of behavioral changes in use patterns resulting from the measure, and physical interactions with other systems, are also included in the observed consumption.

Thus, all effects of interest – from free-ridership to market effects to takeback to interactive effects among measures and systems, to gross savings corrections and installation rates – are included in the overall observed consumption and changes in consumption. If we can analytically separate the effect of the program portfolio from the effects of other factors on overall consumption, this total net portfolio effect is captured. If we can pull out the difference between the with-program and without-program scenarios, there is no need to attempt to estimate separate, hard-to-measure effects.

**Cumulative effects of a long-term effort.** The top-down savings estimate gives the cumulative effect of a program portfolio over multiple years. This is a useful perspective for planning and assessing a long-term effort. The analysis also provides a confidence interval for the estimated multi-year portfolio-level savings.

**Relation to Randomized Control Trials.** The gold standard of energy-efficiency evaluation (or any other causal inference) is a randomized control trial (RCT). However, use of a true RCT is possible only if the program can be randomly made available to some customers and not to others, as with common Home Energy Reports programs. For more general program delivery approaches, and for an overall portfolio of programs, RCT is not possible.

The success of the Home Energy Reports model has led to increased interest in the use of random assignment methods for program evaluation. The recent burgeoning literature on causal inference has relied on the use of Randomized Encouragement Designs Recruit and Deny/Delay Designs, as well as the use of natural and quasi experiments where RCT is not possible. Each of these departures from the RCT gold standard has its limitations and required assumptions. Moreover, random assignment methods, while more rigorous and less subject to assumptions, may take years to implement and can be costly to conduct.

Top-down approaches use analysis similar to what might be used with many random assignment designs. The top-down models do not have the underlying rigor of the random assignment design, but allow for a quicker, lower cost evaluation of energy-efficiency programs than methods using random assignment.

**Avoiding the need for other potentially contentious methods.** Top-down methods do not rely on survey responses to hypothetical questions to assess program influence, nor on engineering analysis and assumptions.

**Cost Savings.** Top-down approaches can not only be less expensive to implement than a randomized assignment study, but can be much less expensive for portfolio-level results than a typical bottom-up portfolio evaluation involving data collection and analysis for each program. However, as discussed below, top-down methods are best considered as a supplement to bottom-up analysis, not as an alternative. Cost savings would result only if the bottom-up studies were eliminated, which is not recommended.
1.5 Challenges and Limitations

While the desirability of such an all-in estimate discernible from territory-level consumption data is clear, the approach has several challenges and limitations.

1.5.1 Accuracy

The key challenge for the top-down method is how reliably it can isolate the effect of the program portfolio from all other factors affecting consumption over time. A key concern in this regard is model specification uncertainty. Different modeling details will often produce very different estimates of the effect of the program. The analyst needs to explore reasonable model alternatives and acknowledge the uncertainties they indicate. At the same time, a top-down study needs to build a case as to why certain model choices make more sense than others, giving more credence to the associated savings estimates. There are multiple examples where improperly specified top-down models have been shown to spuriously demonstrate that the efficiency portfolio is ineffective, or conversely to produce inflated savings estimates.

Even assuming the right model structure is known, the error bands around the savings coefficient tend to be fairly broad. While the error bands for individual Bottom-Up program estimates may be wide, the aggregation of many separate, unrelated estimates has a much smaller relative error than any one estimate alone. With the single all-in top-down estimate, we do not have the same benefit of multiple sources of error cancelling each other.

Data access is another factor that limits the method’s accuracy. It is necessary to rely substantially on existing data sources, such as the Census, at the levels of aggregation for which they are available. Factors that might theoretically be more direct drivers of consumption cannot be included if they are not consistently available across the time periods and geographic areas observed. Moreover, access to the basic consumption and program data itself may be a challenge since historic data may not be retained consistently or with the desired level of granularity.

Pockets of anomalous data can distort the estimates, but may be hard to detect unless we know what to look for. For example, it is generally recognized that economic recessions are times when the relationships between consumption and economic factors might not follow their usual patterns.
1.5.2 Ability to Capture All Net Savings Components

Another set of limitations are related to how fully top-down methods can capture the components of net savings. These issues are partly related to some of the factors discussed above that affect accuracy.

If the geographic units are entire service territories or states, there may be limited expectation of spillover from one geography to another, and therefore limited concern about biases due to self-selection, spillover, or untracked upstream savings. All of these effects will be appropriately captured in the top-down analysis if cross-geography spillover is negligible. However, larger units mean fewer observations, and a broader range of effects and conditions that need to be accounted for.

On the other hand, if the geographies are smaller areas, such as towns, counties, or districts within service territories, we have essentially all the sources of bias that can affect individual-customer billing analysis:

**Self-selection:** If higher program activity is associated with lower consumption, that could be because higher activity leads to lower consumption, or because a greater tendency to adopt efficiency leads to higher program engagement. For example, if customers who are inclined to reduce consumption on their own are more inclined to join energy efficiency programs, then towns where there are more consumption-reducing customers will also have more program activity. The observed association across towns of lower consumption with higher program activity combines this underlying relationship with the actual effect of the program on consumption. The same would be true using service territories as the geographic units, if Massachusetts had a single statewide program such that differences across territories in program activity rates were due to customer self-selection. However, for the most part the total level of program activity is set by the service territory budgets, not by differential participation rates across territories.

**Non-participant Spillover:** If measure adoption outside the program spills over to neighboring geographic units, the cross-geography comparison built into the analysis will tend to lower the estimate of net savings compared to the true program effects.

**Untracked savings from upstream programs:** The effect of untracked savings can only be measured if we have a good estimate of how these savings are distributed across geographic units, and there is variability in that distribution.

**Participant Spillover** is captured in either large or small geographic units, provided we get the timing effects approximately correct. Since participant spillover can occur over multiple years after the initial program participation, the first years of analysis include spillover from any earlier programs, and the last years do not yet reflect the full spillover effects that will occur.

1.5.3 Level of Detail

While the top-down method ideally reflects the effects of free-ridership, participant and non-participant spillover, take-back, interactive effects, and errors in gross savings estimate, the method does not provide separate estimates of any of these factors. It also does not in
general give separate indicators of the effectiveness of programs for individual geographies or individual years, or provide any insights into what works and does not work for individual programs.

1.5.4 Persistence

In principle, the top-down analysis reflects actual measure retention and persistence, since the effects of the installed measures on consumption will continue for as long as the measure remains in place and operable. In practice, the analysis is not sensitive enough to determine appropriate Effective Useful Life (EUL) values or correct for incorrect EUL estimates.

1.6 Use of Top-Down Estimation in Other Regions

Most prior uses of aggregate consumption analysis to estimate net effects of energy efficiency programs have been national studies looking at overall spending and consumption across multiple utilities. Many of these studies have used the savings results to assess the overall cost-effectiveness of energy efficiency program activity. Detailed methods have varied. Some authors have concluded that overall utility energy efficiency programs are less cost-effective than reported by utilities, while others have concluded the opposite, and still others have said the uncertainties are too large to make conclusive statements.

Application of top-down methods to assess impacts for individual service territories or states are less common. Stakeholders in various regions have indicated interest in top-down analysis for their jurisdictions. To date limited work of this type has been published.

The California Public Service Commission (CPUC) sponsored pilot studies completed in 2012, summarized in the Massachusetts cross-cutting evaluation team’s first report on the topic. The CPUC subsequently has had the data compiled to support another round of analysis, but the analysis has not yet been funded. NYSERDA has included top-down methodology studies in its evaluation research agenda; this work also has not yet been funded. Informally, utility staff in multiple states have indicated regulatory interest in methods of this type, generally related to the advantages described in Section 1.4. The method has been used to evaluate savings of a non-utility market transformation program, though published reports are not available.

As other jurisdictions begin to produce top-down savings analysis, the guidance in this document may help Massachusetts PAs and regulators assess and interpret those studies.

1.7 The Value and Use of Top-Down Estimation

The value of top-down estimation comes from its primary appealing features, described in Section 1.4. In particular, the method can provide high-level validation of program effects

---


based on concrete observations of the key quantity targeted by the programs, namely energy consumption. Moreover,

- The effects of large-scale programs over multiple years should be observable in the aggregate consumption.
- The results in principle capture all net-to-gross components.
- The approach is a type of quasi-experimental design, an alternative to random assignment studies for a situation where the random assignment is not possible.
- The result does not depend on engineering models nor on customer or market surveys.
- The cost for producing portfolio-level results can be much less than via a typical bottom-up evaluation.

However, as described in Section 1.5, there are limits to how fully these appealing features are borne out. Time and cost savings would apply only if top-down methods were used instead of bottom-up methods, which is not recommended. Further, top-down methods do not require primary data collection, but do require a non-trivial investment in compiling data across geographies and over time. For top-down models using multiple years of analysis, it may take several years before a long enough series of reliable data is compiled.

In terms of the results provided, top-down methods are dependent on debatable assumptions, just as other study methods are. They are also subject to relatively high levels of uncertainty, even for a given set of assumptions. Top-down methods do not provide detail about the performance and improvement opportunities for individual programs and areas. Additionally, the methods do not necessarily account fully for non-participant spillover and upstream programs, and may be subject to self-selection bias.

Thus, the results of the top-down estimation need to be considered in the context of these known limitations. The goal of the top-down analysis should not be viewed as providing a single definitive estimate of portfolio-level net savings. Rather, the goal of the analysis is to provide another perspective to substantiate the credibility of portfolio savings. To the extent the top-down estimate differs significantly from the bottom-up estimate, it is worth exploring what might have been omitted from either estimate that needs to be considered.
Section 2  Key Work to Date

2.1 WORK PRODUCTS

The Cross-cutting evaluation team has produced the following deliverables related to top-down modeling since the beginning of 2015:

- March 2015 Top-Down Modeling Methods Study Report
  - Consolidated analytic framework, literature review, and pilot analyses; consolidated Residential and Commercial with PA-Muni and PA-only data
- June 2015 Results Memo – PA-Muni Top-Down Model Additional Investigation
  - Provided additional testing of alternative specifications and data screening for the PA-Muni model from the March 2015 report, with related recommendations
- Feb 2017 Top-Down Modeling Extended Methods Review
  - Expanded literature review scope, interviewed experts in different fields, and recommended approaches to several technical issues
- June 2017 Memo on Top-Down Model with New Lighting Distribution Results
  - Re-ran the March 2015 Residential PA-Muni analysis using results from a new lighting distribution model developed in a separate project under the Residential contract

In the work listed, we completed the following:

- A literature review of similar methods applied for estimating net program savings
- Pilot analyses with Massachusetts data
  - Modeled PA data only, for the commercial-industrial sector, using three years of data, with counties or towns as the geographic unit
  - Modeled PA data together with municipal utility data, for both the residential and commercial-industrial sectors, using 13 years of data, with service territories as the geographic unit
  - Incorporated a model of the distribution of upstream lighting program savings into the residential PA-Muni analysis
  - Re-ran the PA-Muni residential analysis with an updated lighting distribution model
- Extended “tire-kicking” of the PA-Muni analysis to assess the robustness of results
- An extended literature review together with expert interviews, considering approaches for modeling with aggregate data outside the energy industry, to recommend approaches for dealing with several technical issues in our context
2.2 SUMMARY OF APPROACHES EXPLORED

The analyses conducted in Massachusetts to date have all used the same general structure of fitting a regression model across time and geographic areas. Models were tested for electricity only.

One model, for the commercial-industrial sector, used PA data only, with only three years of data. A longer series of detailed PA data was not available, nor were residential data available for the analysis. This analysis did not develop reliable estimates, likely due to the limited number of observations.

The other model combined PA data with municipal data. This analysis used data for 13 years (2000 through 2012). By including data from a period of relatively low PA activity, as well as the more recent period of high PA activity, and including service territories with low or no activity across the entire study period, this approach offered a better opportunity to isolate the effects of the energy-efficiency program activity. However, the wider range of time and geographic areas also introduce more variability for which the model needed to account.
Technical Detail

For the PA-Muni models, residential consumption per customer was modeled as a function of the following:

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

Additional terms considered for the PA-Muni residential model included the following:

- Home square footage (not available, used home value as a proxy)
- Age of housing stock (correlated with new construction which was included)
- Household education level (correlated with income, which was included)
- Green Community status (not applicable at the service territory level)

For the Commercial-Industrial PA-Muni model, the dependent variable was consumption per establishment, per customer, or per employee. The predictors were the following:

- 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

For the PA-only Commercial-Industrial model, which had only three years of data, the analysis did not attempt to estimate net savings. It only established whether there was sufficient “signal” from the program activity variable to indicate that, with a longer data series, net savings estimation could be possible. Separate models were fit for small commercial, medium-large commercial, total commercial, and industrial sectors. Data were not available to conduct corresponding analysis for the Residential sector. Terms included in the model were the following:

- Employment
- Program activity
- Fixed effects

Weather was accounted for in the PA-only model by weather-normalizing each year of data for each customer before fitting the model across geographic units and years.
2.3 MAJOR FINDINGS AND LESSONS LEARNED FROM WORK TO DATE

2.3.1 Ability to Provide Useful Estimates

The pilot analyses indicated that top-down estimation can provide useful indicators of portfolio-level net impacts, and provide high-level confirmation that the savings are “real.” On the whole, the PA-Muni residential electric models substantiated that savings are associated with the programs, of a magnitude roughly consistent with the bottom-up estimates. The estimates from the primary models developed ranged from about the same as the bottom-up estimate to nearly twice that, with 90% confidence intervals on the order of ±50% to ±100% of the estimates themselves.

The initial analysis and the “PA-Muni Top-Down Model Additional Investigation” also found that the PA-Muni residential net savings estimates were fairly stable under several alternative specifications and some outlier exclusions. Some informative exceptions follow, and are discussed further in Section 5.

2.3.2 Limitations and Lessons from the Residential Modeling

The analysis also demonstrates that considerable care is needed to ensure that the results developed are both robust (i.e., stable against moderate perturbations in the data and the models) and well determined (i.e., produce savings estimates with narrow confidence bands). Obtaining reliable estimates requires several years of data, identification and treatment of anomalous time periods or geographic areas, sensitivity testing, appropriate estimation of standard errors, and appropriate handling of several technical issues.

The testing of alternative residential PA-mini model specifications and data exclusion revealed some important issues and model sensitivities.

- Appropriately capturing the effect of measures over their lifetimes is an important specification issue. Across different treatments of lag effects, the PA-Muni residential net savings estimates ranged from similar magnitude to almost twice the bottom-up estimates.

- Excluding the Munis from the analysis, or allowing a separate trend line for them, reduced the net savings estimate substantially.

- Careful, explicit treatment is needed for anomalous economic periods, such as the 2008-2010 recession. Excluding the 2008-2010 recession years substantially lowered the net savings estimate. The need for careful treatment of this period was further identified in the subsequent Expanded Methods Review. Section 6.3 identifies some approaches to address this issue.
Capturing upstream program effects correctly in the top-down model requires a reliable estimate of cross-geography program leakage. As described further in Section 5.2, development of the lighting distribution model (LDM) used to estimate leakage was challenging. Moreover, with the LDM ultimately developed, the effect of the upstream program was essentially undetectable by the top-down model. Thus, we consider that the upstream lighting effects are substantially excluded from the PA-Muni net savings estimate.

Technical Detail

The final lighting distribution model produced very similar estimated lighting spending per capita across all service territories (Muni and PA) and years. As a result, the top-down model cannot identify a difference in consumption associated with higher or lower upstream lighting activity. The upstream lighting effects are essentially absorbed into year and service-territory fixed effects. The uniform estimated upstream lighting spending per capita was due to a combination of the lighting distribution itself not varying over the years, as well as limitations in the data, where lighting spending could not be separated from the overall program budget.

2.3.3 Commercial-Industrial Analysis Results

The commercial-industrial results were less definitive than for the residential sector. This is not surprising given the greater variability of the former sector. Overall, our findings indicate that obtaining good results for this sector will likely require both the ability to segment into broad business/industry types and several years of data.

- The PA-Muni commercial-industrial net savings estimates ranged from essentially zero to the same magnitude as the bottom-up estimates. The estimated 90% confidence intervals were on the order of $\pm 100\%$ of the magnitude of the bottom-up estimates.

- Thus, the PA-Muni commercial-industrial model gives some validation to the bottom-up savings, but with high variability. The PA-Muni analysis was not able to separate commercial from industrial accounts, let alone further separate subsectors. The PA-only commercial analysis was overall unsuccessful in providing meaningful stable results. This analysis used data at the town level, but had only three years of data.

2.3.4 Data Challenges

Fitting the top-down models requires that the necessary data be available on a consistent basis across the geographic regions and time periods included. The necessary data includes program activity, aggregate consumption, population size, weather, and economic variables, all by sector.

The top-down analyses faced several challenges related to data availability, as follows:

- The analysis relies on existing government series such as Census data for most of the economic variables. Variables that might be meaningful but are not consistently available cannot be included. For example, square footage could be useful, but is not directly available for the Muni data. Square footage, building vintage, ownership, sale
dates, and property value are all available from the tax data that have been used to
develop the PA customer profile data bases. We have these data for the whole state,
not just the PA territories, and can map them to the municipal territories. Also, the info
USA data contains site level data for most of the state, including municipal territories.
These variables can all be added into the Muni model, with some additional efforts.

- As noted, addressing upstream lighting effects in the residential (PA-Muni) model
required separate estimates of lighting and non-lighting program activity. This
information was not available for years 2010 and earlier.

- For the PA-only analysis, program activity and customer counts were needed for
towns or counties within service territories. This level of granularity was not available
for years 2010 and earlier. Within the time when the study was being conducted, the
granular data were also not available for residential customers for any years.

- The PA-Muni model required energy-efficiency activity data from the municipal
utilities. We were able to obtain these data through 2012 for the initial analysis.
However, a planned update to the PA-Muni analysis was canceled because we were
unable to obtain consistent updated data.

- The PA-Muni analysis relied on EIA aggregate consumption data by sector. Separate
analysis of commercial and industrial sectors would potentially have given more
accurate results. However, the EIA data exhibited inconsistencies over time that
suggested accounts may have been shifted between the commercial and industrial
reporting categories in different years. In addition, reporting of energy-efficiency
activity is not mandatory for municipal utilities, and the EIA energy-efficiency data
were inconsistent with the data provided by the municipal utilities.

- Greater detail of all the data elements can enhance the sensitivity of the model.
However, we often had to rely on high-level totals and generic allocation factors, as
was the case with the lighting distribution model.

2.3.5 Technical Issues Explored

We explored several model specification issues, both via the pilot model development and in
the literature reviews and interviews with experts. In Section 4, we discuss potential paths
forward given the lessons learned from exploring these issues. In Part 2, Section 6, we offer
additional technical detail on these suggestions.

Key issues explored were the following:

- How should aggregate consumption be normalized, if at all – per customer, per capita,
per household, per unit of employment, etc.?

- Which variables should be included in log form versus not logged?

- Should fixed effects terms be included for both time and location, or should a trend
term be used?
• How should the effect of measures surviving over time be represented – by explicit lag terms, or by cumulative program activity?

• How should weather effects be handled – by annual weather terms in the single regression, or by first weather-normalizing individual years using monthly data?

• How should the recession period starting in 2008 be handled?

• How should upstream program activity be accounted for?

Investigating these questions led to several suggestions for the next phase of analysis. The specific technical suggestions from the earlier work are summarized below in Section 6 of Part 2. Technical detail on some new ideas from more recent thinking and discussions is given in Section 7 of Part 2. Section 5 of Part 2 provides additional detail on the analytic results to date, and Appendix A contains tables from those results.

In Section 3, which follows here, we provide a high-level description of the new ideas, and in Section 4, we summarize approaches for future work, combining the prior lessons and the new ideas.
Section 3  New Ideas

To develop a path forward, we engaged two kinds of outside experts to identify high-level alternatives or improvements to our approaches.

1. We interviewed forecasting staff at two of the PAs. Forecasters use models that can be similar to the kind of model that is useful for isolating efficiency program effects, and face many similar challenges.

2. We had a leading academic econometrician experienced in energy analysis review our summary of work to date.

A key goal of broadening the conversation in these ways was to identify whether there was a big-picture approach that has been overlooked in our work so far. New ideas we identified are described in Section 3.1 New High-Level Approaches. We also wanted to identify additional improvements possible within the general approach we have used. The additional improvements we suggest are summarized in Section 3.2.

As described in Section 1, the key feature of top-down analysis is that it considers total consumption for a geographic region and uses analysis across regions, across time, or both to separate effects of energy-efficiency programs from other factors affecting consumption. In the work thus far, the analysis has been by sector, and used data across years and across service territories. We have also explored smaller aggregations within service territories.

Our recent discussions led to two suggestions for approaches that would depart from the general structure of observing aggregate consumption across time and across geographies.

3.1 NEW HIGH-LEVEL APPROACHES

3.1.1 Individual Customer Observations

Related to our top-down approaches to date, but qualitatively different, would be to use individual customers as the “geographic” unit of analysis, while still conducting analysis across several years and across all customers within a sector or segment of interest. The advantage of analysis at this level is that the effects of tracked program activity should appear more clearly since they are tied to specific customers at specific points in time.

This approach would be possible only for a PA-only model, where the individual customer data are available. This analysis would be similar to the kinds of billing analysis regression often conducted on individual programs. However, the analysis would be for multiple programs and the time frame would extend over multiple years.

Several details would need to be worked out to frame such an analysis. In particular, when we move to analysis at the level of individual customers, it is necessary to account explicitly for potential non-participant spillover and self-selection. Likewise, the model structure would need to account for upstream program activity and market effects, or else identify such activity as excluded from the measured impacts. An additional challenge for this approach is the need for good quality, consistent consumption and tracking data at the individual customer level for multiple years.
3.1.2 Machine Learning Methods

Recent developments in computer science have quickly made their way into energy-efficiency program evaluation. Techniques, commonly referred to as “machine learning,” leverage the availability of “big data” (e.g., large volumes of billing data, often including interval data) and program participation tracking to evaluate the effectiveness of programs for the average participating household or by household type (e.g., affluent versus low-income).

Machine learning means allowing patterns and relationships in the data to emerge using a very general set of search algorithms, without a pre-specified model structure. This automated pattern recognition can be incorporated into the analysis at various stages. Models of this type have the potential to provide more accurate predictions under certain conditions as compared to structural models. To the extent a machine-learning algorithm can provide a better prediction of future consumption absent any program activity, this algorithm gives a more accurate estimate of savings associated with the program. Thus, depending on how the learning is done, machine learning may provide substantial reductions in uncertainty around the program effectiveness estimates compared to the conventional regression approach taken to date.

An example of the machine learning approach in the context of energy-efficiency program evaluation is a recent evaluation of a California schools program (Burlig et al, 2013). This analysis was in many ways similar to a conventional bottom-up billing analysis, but used machine learning rather than an explicit degree-day model to predict usage in the post-treatment period absent the program participation. The study also uses several years of data. In this respect, it is similar to the individual customer approach described in Section 3.1.1, restricted to the schools segment. However, the machine learning does not utilize econometric variables. Effectively, the analysis relies on the non-participating schools to represent the effects of these non-program factors.

The schools analysis demonstrates some advantages of machine learning compared to a more conventional matched comparison approach. However, a key self-selection risk of conventional billing analysis with comparison group remains unaddressed: if schools tend to join energy-efficiency programs at times when they are undertaking other kinds of capital investments, the pre-program model, no matter how well trained, is not a good predictor for a participating school’s post-period consumption absent program effects.

“Machine learning” can mean many different things, and moving in this direction would be a research effort. An important drawback is that modeling flaws may be more difficult to detect with an approach that has less explicit model structure. Nevertheless, this idea may be worth further consideration in the future.
3.2 New Ideas for Improvements within the Same General Top-Down Modeling Approach

Other ideas that emerged from the recent discussions provided additional suggestions related to many of the previously explored topics, listed in Section 2.3.5. Our outside expert offered the following further suggestions:

- Consider using household counts rather than number of accounts as the basis for normalizing consumption.
- Use a formal test to determine whether to model on a log, non-log, or alternative scale.
- When fitting a model on a log or other non-linear scale and mapping results back to the natural scale of interest, there is a bias correction formula that should be applied.
- Inclusion of too many fixed effects terms can mask the effects of other factors of interest, resulting in downward bias of the key coefficients.
- Using cumulative bottom-up savings as a single program activity metric rather than estimating individual coefficients for multiple lag years has the advantage that calculation of savings from the estimated coefficients is much less complex.
- To ensure appropriate standard error calculations, test for auto-regressive structure, a good reference is provided for standard error calculations in the contexts of models like ours.

In addition, the evaluation team’s discussions with the outside expert produced suggestions on the following topics:

- **How finely should customer sectors be split into separate segments for analysis?**
  
  If we use municipal utility data, we cannot split customers more finely than residential versus commercial-industrial sectors. As noted, it appears that the commercial and industrial sectors are not consistently assigned in the available data. In a PA-only analysis, finer segmentation is possible, and may have advantages.

- **What geographic unit of observation should be used?**
  
  PA data allow disaggregation to smaller geographic units and business subgroups, but the data are limited in number of years available. Muni data strengthen the analysis by including geographies with minimal program activity, but are available only at the service territory and broad sector level. Both PA and Muni service-territory level data are available for many historic years, but Muni data later than 2012 may not be made available.

  Consider fitting models with mixed geographic detail to take advantage of the geographic detail available for the PAs together with the benefit of the low-program-activity observations provided by the Muni data. For the PAs, use within-PA geographies supported by the residential and commercial-industrial customer databases that are being developed. For the municipal utilities, continue to use the
service-territory data available. If a mixed geography approach is taken, where different observations have different levels of aggregation, more attention is needed to estimate the standard errors. The previous PA-Muni analysis did attempt a town-level analysis, treating each municipal utility as a single town and breaking PAs into separate towns. In that earlier effort, reliable customer count data were not available for towns within PAs. With a better-developed customer profile data base this problem should be reduced.

As an extension of the mixed geographic-level approach, if it is not possible to obtain Muni data for more recent years, consider fitting a model with both PA and Muni data through the period where these are already compiled, and with PA data only for later years. If this “unbalanced” data set is used, in which Muni observations are not available for later years, care is needed in use of fixed effects terms and trend lines.

- What, if anything, needs to be done to address the “circularity” of two of our predictor variables:

  1. Average energy price is used as a predictor of consumption (sales) but is itself calculated as the ratio of sector revenue to sector sales.

  2. Program activity is used as a predictor of consumption (sales), but program spending levels are set as a percentage of revenue, which is driven by sales.

Some attention is suggested to each of these questions, but neither is considered likely to be a major source of error.

- What are recommended approaches to selecting among a variety of competing models with different sets of predictors and somewhat different form?

Key suggestions include not to rely on R², but to consider overall F statistics and use out-of-sample testing. Other formal model selection criteria can also be considered.
Section 4  Potential Paths Forward

Below are suggestions for how to plan and prepare for any future top-down analysis.

4.1  DATA NEEDED

The key to this analysis is data availability. Even in years when no new analysis is immediately planned, it is important to continue to compile the data that will support the analysis. Maintaining the data consistently over service territories and over time is important to the success and validity of the analysis.

4.1.1  PA Data

4.1.1.1  PA-Level Aggregates

The minimum data sets needed from the PAs would include the following aggregate data for each PA as a whole, on an annual basis, for each sector (Residential, Commercial-Industrial):

- Program activity
  - Expenditures, broken out by broad measure type (lighting vs. non-lighting), major program class (small business/direct install vs. large C&I programs or low-income vs. moderate income), or cost type (incentive vs. other program costs)
  - Ex ante and ex post savings
- Consumption
- Number of customers or accounts (number of customers is more useful than number of accounts, but utility data systems do not consistently consolidate accounts to customers)
- Average fuel price (ratio of annual revenue to annual delivered energy)
- Estimated untracked program activity, with allocation formulas or principles

4.1.1.2  Within-PA Aggregates

A richer data set can be created based on the Residential and Commercial-Industrial Customer Profile data sets. Through these profiles, the PAs have already invested in data standardization and organization that can support future analysis. Information available from these data sets is described in Section 4.1.4. These data sets have customer-specific information, as well as the ability to map nearby geographic averages to individual customers. Use of these data would allow breaking analysis into finer geographic or other segments, as well as incorporating additional or more meaningful predictor variables in the models.
4.1.2 Municipal Data

For the PA-Muni model, we need the following at a minimum from each municipal utility, for each sector (Residential and Commercial-Industrial):

- Annual consumption
- Number of customers or accounts. As noted above, we are more likely to be able to have number of accounts accurately than number of customers.
- Program expenditures

The municipal data for the original analyses were obtained from the Massachusetts Municipal Wholesale Electric Company (MMWEC). However, the updated data delivered for a planned 2017 analysis were inconsistent with the data provided previously through 2012, and the discrepancies were not resolved within a reasonable timeframe for the 2017 analysis.

An alternative to obtaining the municipal data from MMWEC would be to use EIA data, which include these same data elements. However, in our prior work we found that the EIA data on energy-efficiency were substantially different from the corresponding data provided by MMWEC for the 1996-2013 period. There were some differences between the data provided to us by the PAs and that in the EIA data series, but the discrepancies were much smaller.

We did use the EIA data for sales by sector, but the changing definition of what consisted of commercial and industrial led to inaccuracies in those sector sales. Therefore, for many years, it was not possible to separate those sectors in the study data set.

We suggest reaching out to MMWEC again to try to establish an agreement and process for regular updates to their data. It is worth exploring whether there is a data collection and transfer process we can facilitate, and how; whether updating would be easier for MMWEC if done routinely on an annual basis, or is easier undertaken every few years; and whether there are ways to provide value to MMWEC members in exchange for maintaining our access to this data series.

If it is not possible to get more recent data from MMWEC, alternatives are as follows:

1. Rely entirely on EIA (or MA DPU) data for the Munis, perhaps excluding Munis where the MMWEC data we do have indicates the EIA or MA DPU data are likely to be less reliable
2. Fit PA-Muni models that include Muni data only through 2012, but PA data through later years, as described in Section 3.2
3. Fit PA-only models

4.1.3 Data from Other Sources

Data from other sources will need to be compiled and linked to the PA data by mapping service territory boundaries, finer segments, or individual customers to available data sources. Most of these data will be available after the fact and do not need to be maintained annually. Such data include the following:

- Degree-days or daily temperature by assigned weather stations
Several of the PAs maintain their own weather stations, used for forecasting and other analysis. If a PA wished to have their own weather stations used for the analysis, that can be accommodated provided the available data span the desired period of analysis. Using NOAA weather stations for all customers is likely to be simpler and will provide consistency across the analysis. For example, the zip code identifier in the Customer Profile data can be used to map NOAA weather stations and data.

- Economic variables based on Census data. Data available at the Census block group level can be appended to individual customer records.
  - Personal Income
  - Gross Domestic Product
  - Median home value
  - Proportion of single-family homes in the housing stock
  - Proportion of renter-occupied housing in the housing stock
  - Employment/unemployment rate
  - Household education level
  - Average employment income
  - Square footage of C&I new construction

Some economic data, such as GDP and payroll, are only available at the county level by industry or town without industry breakdown. Revenue data from InfoUSA could be used as a proxy to apportion GDP by industry. Tax data also provides valuable information regarding building age, size, home values, and ownership at the tax parcel level. Employment data from InfoUSA can be used to approximate the distribution of aggregate data by model segments, such as industry sector or program type.

4.1.4 Compiling and Maintaining Data

While the top-down analysis may be undertaken only every few years, we recommend compiling and maintaining the necessary data on a regular basis. Aggregate data for each PA as a whole are available in various reports. To ensure consistency over time, it may be worth maintaining a file of these data in one place with consistent definitions. The individual customer-level data and associated mapped geographic data are currently being compiled in the Customer Profile data bases. For the Residential sector, PA-validated data are currently available for 2013 through 2015. Data for 2016 are expected to be ready in February or March of 2018, and data for 2017 around August of 2018. The PAs have been working to improve the quality of the Residential data for 2013 forward, and the data may be available more frequently or with less lag in the future. Data from 2010-2012 were previously developed under a different contract, but have some discrepancies – in particular with savings and incentives – that would need to be resolved to align with 2013 onwards.
Commercial-Industrial profile data are available for the years 2011 through 2016, with annual updates planned, similar to the Residential data.

The compiled profile data include the following for each account:

- Tracking data, including gross savings, EUL, measure group, incentive payment, and date
- Coding to allow mapping to any geographic level of interest, such as zip code, town, and county
- Tax assessors’ data, including year of construction, property value, square footage, number of stories, dwelling unit type (residential), last sale data, use code, and sometimes heating fuel
- Monthly consumption data with meter read dates and flags, and rate class
- Green Community Status and year status obtained

**For Residential accounts**
- American Community Housing Survey data (census block averages) for demographic characteristics, including household composition, income, and ethnicity

**For Commercial-Industrial accounts**
- InfoUSA data including NAICS code, Employment category, and organization type
- County Business Patterns data (zip code level), including total employment and payroll by NAICs group (currently through 2015)

Even with this compilation, there will still need to be data set preparation work at the start of any new top-down analysis. The data quality and completeness for each item depends on the source data, such as local tax records. For a particular analysis, some individual fields will still need to be scrubbed and filled, and the analysts will need to review known issues with particular sources and geographic areas. For example, some PAs have indicated that the “installation date” may simply be the data entry date. InfoUSA coverage is high but misses some small areas.

For any future top-down analysis, to address these issues, the data management team can work with the analysis team to help prepare the data at the start of the analysis, and to ensure that the modeling approach is consistent with the data available. For the PA-only top-down analysis, the analysis team worked with the data management team to assemble the analysis data set.

Other program data needed that are not customer-level information should be available from regular PA reporting. This information includes the following:

- Program operational (non-incentive) expenditures
- Upstream program expenditures and ex ante and ex post savings estimates

Data from government sources not included in the customer profiles can typically be obtained when needed at the desired level of granularity.
Municipal data should be obtained at whatever frequency the provider is willing to make them available. Moreover, the same data sources that are accessed to produce the Customer Profile data can also provide data for the Munis. While we do not expect to have customer-specific data from the Munis, the data management team can use the tax data and InfoUSA data, for example, to develop average values for the same variables compiled individually for the PA customers. These data could be used to enhance the Muni model, providing insights into the building size, vintage, number of residents/employees, and income level/revenue.

4.2 Continuing Top-Down Analysis Over Time

We suggest that the PAs consider continuing the general modeling approaches explored to date, with specific elements and technical improvements described below. Specifically, we suggest that the PAs consider performing the next round of top-down analysis in 2018 or 2019. In 2018, we will have PA data available for 2011 through 2017. For a PA-only analysis, based on successful top-down modeling performed in other states, seven years of data should be sufficient to produce estimates for each sector using town or county-level geographies. For a PA-Muni analysis, we will have seven post-recession years. These can be combined with a similar number of pre-recession years to give estimates unclouded by the recession period itself. If it is not possible to obtain more recent energy-efficiency data from the municipal utilities, options are to use EIA data for the entire time span, or else to conduct analysis that includes the municipal utility data only through 2012.

In the remainder of this Section, we provide three sets of suggestions for improving the analysis:

1. Improvements that can be incorporated into the recently used models relatively easily.
2. Improvements within the same general framework, but requiring more effort to implement.
3. Alternative modeling approaches, requiring further methodological development.

We suggest that the next rounds of top-down analysis incorporate the suggestions from Section 4.3 at minimum, and further consider some of the suggestions from Section 4.4. These methods are described more fully in Part 2, Section 6 and Section 7.

4.3 Improvements Incorporated Relatively Easily Within the Current Framework

- Variable to represent the level of program activity:
  - Use cumulative bottom-up ex ante or ex post savings as the program activity variable, rather than estimating lag effects of energy-efficiency. With this approach, the analysis can be set up so that the coefficient of the bottom-up estimate represents an adjustment or realization rate for that estimate. If the time frame of the analysis extends over many years, this approach may require “retiring” some savings from the cumulative total, based on estimated measure life.
- **Recession:**
  - Include explicit terms to account for recession effects. At a minimum, allow for separate program activity coefficients for the recession and non-recession periods. A more illuminating, but more challenging, approach would be to include terms in the model that would account more fully for different responses to other factors during the recession period.

- **Time period effects:**
  - Explore and test for linear time trends rather than estimating individual time period effects.

- **Access to municipal utility data:**
  - If new data are not available from the municipal utilities, expand the PA-Muni data set with additional years of PA data. That is, the analysis data set would have both Muni and PA data through 2012, and only PA data for later years.

- **Scale of analysis:**
  - Determine the appropriate scale for the analysis (e.g., log or non-log) using the Box-Cox parameter transformation test.
  - If the scale is logarithmic or otherwise transformed, apply Goldberger’s method to eliminate bias when the estimates are transformed back to the original scale.

- **Standard error calculations:**
  - Test for auto-regressive error structure
  - Apply standard error estimation methods described in Abadie et al. (2017)

- **Interpretation of differences between top-down and bottom-up savings estimates:**
  - If top-down analysis produces a savings estimate substantially larger than the corresponding sector-level bottom-up evaluated estimate, that can’t be assumed to be evidence of large-scale unaccounted-for market effects, unless there’s a credible theory and additional evidence that market effects could be that large.

### 4.4 Improvements Requiring Greater Effort Within the Current Framework

- **Finer geographic levels of aggregation:**
  - Use data at the level of town or Census tract within PA territories.

- **Finer segmentation:**
  - In a PA-only analysis, explore conducting the analysis separately for different customer segments with the residential and commercial-industrial sectors. This approach could produce more accurate and more informative results for many segments, while potentially isolating less tractable segments.
• Units for scaling the analysis:
  o Use household count rather than number of accounts to scale consumption. This requires mapping of geographic units to Census data sources to obtain the counts.

• Circularity of energy price and program expenditures as predictors:
  o Review in more detail how spending levels are set. However, this is likely to be a minor issue, particularly if we move away from program expenditures to cumulative bottom-up savings estimate as the program activity metric. The circularity between consumption and the price variable is not a major concern since the focus of our analysis is not on developing an accurate estimate of the price response.

• Model selection:
  o Do not use $R^2$ as a criterion for selecting among competing plausible models. Use overall F statistics and out-of-sample tests. Keep models simple to the extent possible, following approaches of Houthakker and Taylor (1970). Other formal model selection criteria, such as Akaike’s Information Criterion, may also be considered.

• Geographic span of the analysis:
  o Broadening the range of geographic units included in the analysis could improve the analysis or might make savings harder to pinpoint. We suggest consideration of incorporating data from service territories outside Massachusetts in the analysis. One option would be to use data from neighboring out-of-state operating companies of the multi-state utilities operating in Massachusetts. The other option is to conduct state- or utility-level analysis using data from a broader section of the United States, and relying on EIA data.

### 4.5 Methods Requiring Additional Development

• Individual customer observations:
  o Explore approaches that model individual customers across time periods, or across an entire segment or sector. The model would include participants and non-participants, and would incorporate expected reductions for each customer beginning with each instance of program participation and continuing for the life of the measure. Explicit structure would need to be developed to address self-selection and spillover in this framework.

• Machine learning:
  o Machine learning methods may provide a basis for a richer and more powerful analysis incorporating individual customer observations. However, a means of addressing spillover and self-selection in this context must still be developed.
5. Key Analytic Findings from Work to Date

5.1 Models Developed

As described in Section 2.3.5, only three years of data were available for the PA-only analysis, and only for commercial-industrial customers. However, the PA-only data could be compiled at any level of aggregation of interest. The PA-Muni analysis had data for several years, for both residential and commercial-industrial customers, but only aggregated to the service-territory level. Only electricity models were tested in either analysis.

The PA-Muni analysis, particularly the residential analysis, was the most successful. The data included time periods with high and with low PA activity levels, and service territories with high and with low program activity across the later years. This range of variability in the data made it easier to identify effects of program activity.

With a limited time-span and only high-activity years included, as well as the greater variability of the commercial-industrial sector, the PA-only models were inconclusive.

Results of the PA-Muni analysis are summarized here. A large number of variations were tested.

5.1.1 Residential Model

Figure 2 shows residential energy consumption per customer for the PAs and the Munis from 1990 through 2012. As the figure indicates, the overall trend line has been flatter for the PAs than for the Munis. Figure 3 shows the energy-efficiency program spending for the PAs and Munis over the same period. PA expenditures have been much higher. The top-down analysis seeks to isolate the contribution of the program spending to the slower growth rate among the PAs.
Figure 2: Residential Electricity Consumption per Customer (kWh)

Figure 3: Residential Electric Program Expenditures per Customer ($)
Models initially tested for the residential sector included the following:

  Models 1-3: Models were tested with only current-year program activity, with current and four lag years of program activity, and with current and six lag years of activity.

  Models 4-6: Each of the lag 1-3 models was tested with an adjustment to account for upstream lighting.

  Models 7-12: As an alternative to modeling each lag year separately, as in Models 1-6 ("individual year"), models were run with the cumulated program activity for three, four, or six years as the single program activity variable. These models were tested with (10-12) and without (7-10) the upstream lighting adjustment included.

The estimated coefficients for each model are provided in Appendix A.

The table below summarizes the savings estimates for the models with varying levels of lag (four years or six), and with cumulated and individual-year lag terms.

<table>
<thead>
<tr>
<th>Model Family</th>
<th>#Lags</th>
<th>Top-down Annual Net Saving Estimates (GWh)</th>
<th>Top-down Annual Net Saving Estimates (% of Net Bottom-up Estimates)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Lower Bound</td>
<td>Point Estimate</td>
</tr>
<tr>
<td>Individual Year</td>
<td>Four</td>
<td>1,851</td>
<td>3,762</td>
</tr>
<tr>
<td>Cumulated</td>
<td>Four</td>
<td>41</td>
<td>1,714</td>
</tr>
<tr>
<td>Individual Year</td>
<td>Six</td>
<td>2,829</td>
<td>3,821</td>
</tr>
<tr>
<td>Cumulated</td>
<td>Six</td>
<td>1,075</td>
<td>2,233</td>
</tr>
</tbody>
</table>

The results indicate a realization rate compared to net bottom-up estimates on the order of one for the cumulated models and on the order of two for the individual year models. As described below, when the anomalous recession years are dropped from the model, the factor of two for the individual-year model reduces to roughly a factor of one.

On the whole, the residential model results substantiate that savings are associated with the programs, of a magnitude roughly consistent with the bottom-up estimates.

5.1.2 Commercial-Industrial Model

A similar set of commercial-industrial models were tested to those for the residential sector. No upstream adjustments were attempted for this sector.

  Models 1-4: Models were tested with only current-year program activity, with current and three lag years of program activity, with current and four lag years of program activity, and with current and six lag years of activity.

  Models 5-7: As an alternative to modeling each lag year separately, as in Models 1-4, models were run with the cumulative program activity for three, four, or six years as the single program activity variable.
The estimated coefficients for each model are provided in Appendix A.

The table below summarizes the savings estimates for the models with varying levels of lag, cumulated and individual year.

Table 2: Savings estimates from Commercial-Industrial PA-Muni Models with varying treatment of cumulative savings effects

<table>
<thead>
<tr>
<th>Model Family</th>
<th>#Lags</th>
<th>Top-down Annual Net Saving Estimates (GWh)</th>
<th>Top-down Annual Net Saving Estimates (% of Net Bottom-up Estimates)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Lower Bound</td>
<td>Point Estimate</td>
</tr>
<tr>
<td>Individual Year</td>
<td>Three</td>
<td>925</td>
<td>3,342</td>
</tr>
<tr>
<td>Cumulated</td>
<td>Three</td>
<td>742</td>
<td>3,158</td>
</tr>
<tr>
<td>Individual Year</td>
<td>Four</td>
<td>-207</td>
<td>2,142</td>
</tr>
<tr>
<td>Cumulated</td>
<td>Four</td>
<td>307</td>
<td>2,656</td>
</tr>
<tr>
<td>Individual Year</td>
<td>Six</td>
<td>-2,850</td>
<td>-573</td>
</tr>
<tr>
<td>Cumulated</td>
<td>Six</td>
<td>-3,204</td>
<td>-277</td>
</tr>
</tbody>
</table>

There is little consistency across the estimates from the different commercial-industrial models. For the best-fitting model, dropping the recession period increased the point estimate, but rendered the estimate not statistically significant. Thus, the commercial-industrial model gives some validation to the bottom-up savings, but there appears to be too much variability in this sector to derive reliable estimates with this structure. With further model detail that can account for that variability, better estimates may be possible.

5.1.3 Basic Sensitivity Testing

As noted in Section 2, we performed additional investigation of the PA-Muni top-Down Model to assess model robustness in several ways. Specifically, we conducted the following evaluations:

- We conducted a sensitivity analysis by dropping individual Munis and PAs out of the model one at a time to determine whether the exclusion of any one Muni or PA caused disruption in the model.
- We conducted a sensitivity analysis by dropping individual years out of the model one at a time to determine whether the exclusion of any one year caused disruption in the model.
- We evaluated the within R-squared and the adjusted R-squared of these tire-kicking models compared to the models in the PA/Muni report.
- We conducted and examined F-tests.
- We examined the confidence intervals around the estimated model savings.
- We created residual plots based on the Residential PA-Muni model to evaluate the fit of the recommended model.
• We ran the model with some changes to the predictor variables, including using alternative variables, including interactive terms, deleting some years from the model, and fitting separate coefficients for some subgroups.

The PA-Muni Residential model proved to be robust in response to the exclusion of individual Munis and PAs. The PA Muni model and outputs were still significant despite dropping the Munis and PAs out of the model one at a time. We did not see any sign of shifts or large coefficient value changes, which is evidence that the estimated savings in the original PA-Muni model were not driven by one or a small number of Munis.

The sensitivity analysis of the PA-Muni C&I model showed a similar level of stability compared to that found in the PA-Muni Residential model analysis, with the exception of a single Muni outlier, as we identified one highly influential Muni. Excluding this outlier from the model reduces the savings estimates by about 30%, indicating that the C&I model is somewhat sensitive to outliers.

The PA-Muni Residential model was not sensitive to having certain individual years dropped from the analysis. We systematically dropped individual years from the residential model, and the model continued to perform as it did when all years were included in the model, proving that this model is robust to excluding years from the analysis.

The C&I PA Muni model did not prove to be robust when excluding individual years. When we dropped (one at a time) 2004, 2005, 2007, 2010, and 2011, the model failed to produce a significant estimate, and the estimated coefficient varied wildly. This is evidence that the C&I model is sensitive to which years are excluded from the analysis, that certain years have a greater impact on the model than do other years, and that the original C&I results are not reliable. This exercise reveals the inherent difficulty in detecting energy savings due to ratepayer programs in the C&I sector, and due to the high variation in C&I consumption and in the many drivers of consumption in this sector.

Most of the changes in model specifications did not result in major changes to the net savings estimates, or they did change the estimates but with a worse overall fit. Therefore, that model change would not be recommended.

• For the residential model, using median number of rooms instead of house value as a proxy for house size substantially reduced the estimated net savings. The model with house value had a better fit. House value is related not only to dwelling size but also to economic conditions. While we cannot assume that the house value term is capturing a size effect, it appears to be an appropriate variable to help account for consumption.

• For both the residential and commercial-industrial models we allowed separate trend terms for municipal utilities than for PA territories. Compared to the “preferred” original models (Model 5 of Table A-1 and Model 2 of Table A-3 below), the residential net savings was 30% lower, and the commercial-industrial estimate was similar but no longer statistically significant.
• Excluding all the municipal utilities from the model reduced the net savings estimate by about half, from nearly double the bottom-up estimate to about the same magnitude.

• Excluding the recession years 2008-2010 changed both the residential and commercial-industrial net savings estimates substantially, as described below.

In principle, the inclusion in the model of territories with little or no program activity across the study period should give us a more reliable estimate of the effect of the programs. The lower estimate when the Munis are excluded may indicate that PA-only analysis may underestimate savings. On the other hand, the drop in savings when the Munis are allowed a separate trend coefficient illustrates the importance of appropriately accounting for the ways Munis may be different from the PAs, apart from the ways reflected in the economic variables that are currently included.

The reduction in savings when the Munis have a separate trend line, as well as when the recession years are excluded, may be related. Given that the recession years have a large effect on the model fits, it is likely that these years are also affecting the trend terms and that the recession response itself is different for Munis than for the PAs. As noted in Section 2, the recession period is a key consideration that needs more attention based on the sensitivity analysis.

5.1.4 Exploring the Effect of the Recession

All the experts we interviewed for the extended methods review, as well as the forecasters we talked to, indicated that the recession was an anomalous period that could distort the results if not addressed separately. To explore the impact of the recession years on the model, we removed the years 2008-2010 from the model in the additional investigation.

We were concerned that the recession years were anomalous in ways that may not be well controlled for by the available economic variables, and that the disruptions of this period could have affected Muni and PA areas differently due to the different economic mixes. A possible result would be that declines in consumption due to the recession might be spuriously associated with the increased PA spending starting in 2009. Since there were three post-recession years (2011-2013), we believed there should be sufficient numbers of high and low program activity years to provide good estimates of the effect of greater PA activity, even if we excluded the recession years.

When we drop the recession years, the model fit does not include any of the consumption observations for 2008-2010. The model does use the program spending lag variables created from the recession year program spending, to estimate impacts on usage for later years.

For the Residential sector, the result of this exercise supports the conjecture that some of the reduction in consumption that the original model attributes to PA program activity appears to have been associated with the recession itself. When recession months are removed, the estimated program effect is still present, and we still have a well determined model overall, but the estimated program effect is only half as large – and in fact close to the bottom-up estimate.
For the C&I sector, dropping the recession years increases the estimated program effect, but that estimate is no longer significant. Since the C&I model was not well determined to begin with, removing a large proportion of the observations results in estimates too uncertain to be meaningful.

5.2 **Upstream Lighting Program Effects**

The residential upstream lighting program accounts for a sizeable fraction of the overall residential portfolios. However, because the program is upstream, discounted lamps may end up within the service territory of the PA that provided the discounts, or may “leak” into neighboring territories, including municipal territories.

For upstream effects to be captured in a top-down savings estimate requires that the leakage can be reliably estimated (or credibly assumed to be zero). With or without leakage, it is also necessary to have variation in the levels of upstream activity across geographic units and time.

In particular, for the residential upstream lighting programs, it was necessary to estimate the amount of leakage across territories, including from PAs to Munis. Developing the lighting distribution models was challenging.

Two efforts have been undertaken in association with the top-down modeling to model the geographic distribution of the benefits of the upstream lighting program. The first effort (2014) estimated the geographic distribution of bulbs using concentric drive-time rings around participating retail locations in the state. This work assigned bulbs sold by a retail location to the census block groups surrounding that location, with proportionally more bulbs assigned to nearby block groups and fewer bulbs assigned to block groups farther away. The proportion of bulbs assigned to each concentric drive-time ring around each location was informed by a residential customer profile study in which respondents reported location, travel mode, and travel time associated with their lighting purchasing behavior. Program leakage from PA territory to municipal territory was estimated based on the amount of total rebate dollars that ended up in municipal territories in 2010-2013. A ratio of lighting expenditure to total EE program expenditure was computed and applied to the leakage expenditure. This was subtracted from PA territories and applied to municipal territories using a constant ratio applied to each year represented in the model, and finally included in the model as a portion of the municipal EE budget.

An initial top-down model using the first lighting distribution estimates found a slight effect on the savings results when the lighting distribution model was incorporated. However, a subsequent effort to update the first LDM results was not able to reproduce those lighting allocations. Rather than attempt to resolve the discrepancies, effort was focused on an alternative lighting distribution model approach.

The second set of lighting distribution models produced estimates of the geographic distribution of upstream program activity at the Census tract level. These Census tract distributions were aggregated up to the Muni and PA service territory level.
The second lighting distribution model estimated the effects of several predictors of efficient bulb uptake using a multi-year panel of in-home lighting surveys of household bulb inventories. The work included two pathways for estimating the geographic distribution of program bulbs across the state: one based on observed in-home socket saturation of efficient lighting products for a sample of MA households, and the other based on observed year-over-year inferred purchases of efficient lighting products by the occupants of a sample of MA households. Included in the set of demographic predictors for both pathways, were measures of household proximity to retail locations for program bulb types.

The results were consistent with a conclusion that, within the state, there was little effect (null finding) of geographic proximity. Some other demographic predictors were found to be associated with differences in efficient bulb use at the household level, but the overall effect sizes at reasonable levels of geographic aggregation within the state were relatively minor.

The result of this modeling effort was very similar per capita program spending effects across all service territories. Additionally, the lighting distribution results were not estimated separately for each year of program activity.

As a result, the upstream lighting effects would be nearly completely absorbed into the year and service-territory fixed effects. Hence, the top-down analysis detects almost no difference in savings with the upstream lighting activity included or excluded. Essentially, the upstream lighting program effects are largely excluded from the estimated portfolio impacts.
Section 6  Earlier Guidelines for Improving Modeling

The following are technical recommendations to improve future modeling, from the extended methods review.

6.1  WEATHER MODELING

Two approaches to weather modeling are as follows:

1. Weather normalize each year’s consumption separately using a monthly regression of consumption on degree-days. Use weather normalized consumption as the dependent variable in the top-down model of annual savings.

2. Include annual weather terms as explanatory variables in the top-down model of annual consumption.

Both methods are acceptable, with some preference for annual weather normalization, as long as the time period of observations is annual.

6.2  LAG EFFECTS

1. Use the cumulative surviving savings as the program activity variable if possible, considering the EULs of different types of measures.

2. Test for the need for lagged predictors other than program activity, and consider the structural reasons these may be important.

3. Whatever model is fit, fit the same model in first differences and see if the coefficients are stable. This suggestion is useful to address several issues. In particular, taking first differences can reduce serial correlation of the residuals, reducing the risk that standard errors are underestimated and the reliability of the results thereby overstated.

6.3  RECESSION

1. At a minimum, include a separate program activity coefficient for the 2008-2010 recession period and test whether it is significant.

2. With several years of data post-recession, explore models that include years before and after the formal definition of the recession period, but exclude the recession period itself.

3. Explore including variables that account explicitly for what was different during the recession. In particular, incorporate consumer spending and real income. Also consider using rates of change in key economic indicators, such as GDP or capital investment.
6.4 Fixed Effects
1. Include fixed effect terms, particularly for geographic units.
2. Explore specifications that omit time fixed effects, but use a linear trend term in place of individual time effects.
3. Conduct formal testing of the need for the fixed effects.
4. Conduct the analysis with first differences, as well as in the original form.

6.5 Endogeneity of Electricity and Gas Prices
1. Do not put a lot of effort into fixing this problem.
2. If performing analysis with geographic units is finer than service territory, use service territory price, lagged service territory price, or a marginal price if available.
3. Consider constructing expected price over some reasonable measure lifetime horizon.
4. Consider constructing an approximate marginal price by modeling the revenue/consumption ratio versus consumption. This model would need to be for a different year than the observation year, or it will create a different endogeneity.

6.6 Endogeneity of Program Spending
1. Explore how spending levels for each year are determined and to what extent they are directly related to current year consumption.
2. Use predicted savings, not spending, as the program activity variable. Predicted savings is likely to give more meaningful results.
3. If we need to rely on spending as the activity metric and current year consumption is a strong determinant of current year spending, consider decomposing the spending into major types of spending (such as lighting or non-lighting). Each of these would be much less correlated with the total consumption than total spending would be.
4. Consider using cumulative surviving savings, or a spending proxy that is not the current year’s.
5. If the dependent variable is weather-normalized consumption, this will help break the relationship between the dependent variable and the spending variable.

6.7 Error Bounds
1. Use robust estimation of standard errors. Ensure that the estimated savings or realization rates have error bound calculations that are appropriate to the estimation process, and provide a transparent description of how these bounds are determined.
6.8 **Constructing Portfolio Savings**

1. Ensure that savings from the top-down model are estimated consistently with how each savings coefficient appears in the estimated model.
2. Ensure that top-down savings are compared with bottom-up estimates that cover the same period and the same cumulative effects.
3. Be very transparent on how the top-down and bottom-up savings are calculated, what period and cumulative effects they represent, and why these are comparable.
4. Using cumulative savings rather than estimating individual year lag effects makes the translation from estimated coefficients to estimated savings more transparent.

6.9 **Self-Selection**

1. For analysis using geographic units that are subsets of service territories, explore adapting the recent advances in self-selection correction for individual customer regressions to top-down aggregate analysis methods.

6.10 **Spillover**

1. Explore inclusion of spillover terms in the model. For example, include the effect of program activity in surrounding areas or in the territory overall as a predictor along with the geography-specific activity.
Section 7  Ideas from New Conversations

7.1  PA Forecasting Approaches

The forecasters we interviewed observed that in recent years actual consumption has been lower than forecast, even after accounting for program estimates of energy-efficiency. They speculated that this may be due to a combination of sources, including additional efficiency, increasing customer-sited solar generation, changing codes and standards, and modeling errors related to the recession period.

7.1.1 Forecasting Approaches

The forecasters model consumption for each sector and operating company over time as a function of economic variables. Key differences from our approach are as follows:

1. They model total service territory sector consumption over time only, not across geographic units.

2. They use months rather than years as the time increment, and include monthly degree-days as predictors. With monthly modeling, degree-days become the primary driver of variation.

3. They tend to include fewer economic factors, focusing primarily on energy price (ratio of revenue to MWh or therm sales) and GDP or income.

4. They need to forecast total consumption, not consumption per customer. One approach is to model total consumption per customer, similar to our modeling, and combine that with projected number of customers to produce the total forecast. Another approach is to model consumption in total directly.

5. They do not attempt to isolate the effect of energy efficiency. They identified two different ways of accounting for energy efficiency in the projections.
   
   a. Model energy consumption with no explicit energy-efficiency variable. Assume the program activity is built into the “business as usual” represented by the model. After using this business-as-usual model to predict future consumption, add or subtract the incremental planned energy-efficiency program savings below or beyond the business-as-usual average.
   
   b. Add the cumulative program estimates of historical energy savings to the historical observed consumption, and model this adjusted consumption amount. Then subtract the projected future cumulative savings from the resulting projections. For some sectors and territories this approach has been found to produce more stable and meaningful models.

6. They do not attempt to fit models over large numbers of years. One comment forecasters made was that the series needed either to begin after the recession, or else to go back far enough (2006) to establish non-recession patterns.
7. They sometimes include explicit auto-regressive terms in the model (i.e., the residual – the portion of consumption that is not accounted for by the model – from one period affects consumption in the next). This may be more important with monthly modeling than with annual.

7.1.2 Implications
Most of these approaches do not have direct application to addressing our goal of estimating net savings.

Increased attention to the error structure is something we should explore. If the errors are not appropriately represented, the calculated precision can be off.

We could consider using monthly models. However, since the effect of weather on consumption changes over time and in response to efficiency measures implemented, we would likely need to include some interactive terms that could result in a complicated model. We have tended to use annual data because many economic variables, as well as program spending, are available only annually.

7.2 SUGGESTIONS FROM THE ACADEMIC EXPERT AND RELATED INTERNAL TEAM DISCUSSIONS
Our academic review provided several suggestions related to questions previously explored, as well as some additional issues. One of these suggestions led to the consideration of machine learning methods described in Section 3.1.2. Other suggestions from the reviewer were ways to improve on the models within the same general approach of an explicit regression model fit across time and geographies. These points are noted below, and are included in the recommendations in Section 4.

1. Scale:
   o The choice of whether to model on a log or unlogged scale, or some alternative, can be addressed by use of the Box-Cox (1964) parameter transformation test, described in many standard econometric texts.

2. Scale change bias correction:
   o If analysis is done on a log scale, then results are translated back to the normal scale, the bias correction term due to Goldberger (1968) should be included, and is often ignored.

3. Interpretation of top-down versus bottom-up estimates:
   o If top-down analysis produces a savings estimate substantially larger than the corresponding sector-level bottom-up evaluated estimate, that cannot be assumed to be evidence of large-scale unaccounted-for market effects without a credible theory and additional evidence that market effects could be that large. A best effort should first be made to estimate the market effects and upstream impacts, then to explain the remaining discrepancies.
4. Use of lagged versus cumulative program activity:
   o Estimating separate coefficients for lagged program activity at each number of years of lag can lead to tricky calculation of the overall effects of interest. Use of a single cumulative variable simplifies this calculation. (In our prior work, the complexity of estimating multiple lag coefficients was recognized, but these models performed slightly better in some respects than those with a single cumulative term.)

5. Fixed effect terms:
   o When fixed effects terms are included, a lot of the variation in the consumption drivers is absorbed into the fixed effects terms. The result can be that all the coefficients, including the program activity coefficient used to estimate savings, are biased downward. Fisher et al. (2012) describe how to mitigate this problem.

6. Model selection:
   o Factors to consider in selecting among alternative model specifications include:
     a. Keep it simple as much as possible, following Houthakker and Taylor (1970)-type models.
     b. Look at the overall F statistics, and ignore $R^2$.
     c. Use out-of-sample testing, which puts more weight on predictive ability than on sample fit. If possible, leave perhaps one arbitrary year out of the analysis and check whether the mean square error of predicting the omitted year is consistent with the diagnostics from the fitted model.

7. Endogeneity (circularity between consumption and its drivers):
   o Our previous work considered potential bias due to use of predictors that are themselves determined from the dependent variable. Our reviewer suggests that neither of these is likely to be a major concern, but offers some suggestions.
     a. Since program spending is a portion of sales, there is a circular relationship between consumption and expenditures. The suggestion is to review in more detail how spending levels are set, as previously identified, but that the relationship between current year sales and current year program expenditures is probably noisy enough not to be too concerned about this effect.
     b. Since our energy price variable is calculated as revenue divided by energy sales, there’s a circular relationship between consumption and price. The suggestion is that this is not a major concern since the focus of our analysis is not on developing an accurate estimate of the price response. A solution if we had a monthly model would be simply to use the price from two months previously rather than current price. The lagged price may be a better predictor in any case since price is visible to customers primarily in bills, which lag consumption.
8. Standard errors:
   - Abadie et al. (2017) describe criteria for when special attention is needed to address standard errors accurately where clustering effects are possible by time period or by cross-sectional unit.

9. Population size metric:
   - For residential models, normalize consumption by number of households rather than number of customers or accounts. This would require household counts for the geographic units used in the analysis, which is straightforward if the units correspond to Census designations, but more complicated when mapping for service territories.

10. Analytic segmentation:
   - Conduct a more finely segmented analysis, rather than simply separating residential from commercial-industrial. In a PA-only model, residential customers can be segmented by dwelling unit type and it may further be possible to separate low-income customers or neighborhoods. The commercial-industrial sector could be segmented by size, by load factor, or by business type — at minimum separating commercial from industrial. It can also be helpful to exclude the very large customers, whose idiosyncratic changes can contribute substantial volatility to the aggregate analysis.

11. Geographic span of the analysis:
   - Broadening the range of geographic units included in the analysis could improve the analysis by sharpening the difference between geographies and times with higher versus lower program activity. A broader range could also introduce more variability in the program and non-program factors to be accounted for, making savings harder to pinpoint. We recommend consideration of two possible means to incorporate data from service territories outside Massachusetts in the analysis. One option would be to use data from neighboring out-of-state operating companies of the multi-state utilities operating in Massachusetts. Such an approach would require either (1) using service territories as the geographic units, relying on EIA and Census data, unless the utilities were able to supply more granular data, or (2) conducting state- or utility-level analysis using data from a broader section of the United States, and relying entirely on EIA data for the consumption and energy-efficiency data.
Section 8 References


Burlig, Fiona, Christopher Knittel, David Rapson, Mar Reguant, and Catherine Wolfram, 2017. "Machine Learning from Schools about Energy Efficiency" (September 2017)


Table 3: Residential PA-Muni Model Results with Individual Year Program Expenditures

<table>
<thead>
<tr>
<th>Variable</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
<th>Model 5</th>
<th>Model 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Price of electricity</td>
<td>-0.10670 (0.0675)</td>
<td>-1.38113** (0.0471)</td>
<td>-0.15799** (0.0261)</td>
<td>-0.10591 (0.0675)</td>
<td>-0.13671** (0.0487)</td>
<td>-0.15808** (0.0278)</td>
</tr>
<tr>
<td>Annual HDDs</td>
<td>0.17483** (0.0574)</td>
<td>0.20179** (0.0665)</td>
<td>0.13418+ (0.0757)</td>
<td>0.15674** (0.0576)</td>
<td>0.20555** (0.0675)</td>
<td>0.13842+ (0.0802)</td>
</tr>
<tr>
<td>Annual CDDs</td>
<td>0.05798** (0.0069)</td>
<td>0.06530** (0.0073)</td>
<td>0.05359** (0.0081)</td>
<td>0.05806** (0.0069)</td>
<td>0.06570** (0.0077)</td>
<td>0.05404** (0.0084)</td>
</tr>
<tr>
<td>Median household income</td>
<td>-0.14926 (0.1288)</td>
<td>-0.06627 (0.1154)</td>
<td>0.01145 (0.1162)</td>
<td>-0.15027 (0.1294)</td>
<td>-0.07936 (0.1163)</td>
<td>-0.00740 (0.1176)</td>
</tr>
<tr>
<td>Median home values</td>
<td>0.31034** (0.1085)</td>
<td>0.43072** (0.0943)</td>
<td>0.47439** (0.0873)</td>
<td>0.31171** (0.1081)</td>
<td>0.42720** (0.0929)</td>
<td>0.46868** (0.0859)</td>
</tr>
<tr>
<td>Percent of homes using electricity as the main heating fuel</td>
<td>0.59148 (0.6564)</td>
<td>0.55555 (0.5391)</td>
<td>0.63794 (0.5731)</td>
<td>0.59436 (0.6555)</td>
<td>0.55943 (0.5349)</td>
<td>0.64519 (0.5652)</td>
</tr>
<tr>
<td>Percent of residential new construction</td>
<td>1.67497* (0.7537)</td>
<td>1.43743+ (0.7368)</td>
<td>1.30648 (0.8444)</td>
<td>1.67464* (0.7514)</td>
<td>1.48740+ (0.7474)</td>
<td>1.35890 (0.8498)</td>
</tr>
<tr>
<td>Percent of single-family homes</td>
<td>0.61429 (0.6459)</td>
<td>0.37955 (0.5210)</td>
<td>0.05182 (0.4358)</td>
<td>0.61542 (0.6464)</td>
<td>0.37419 (0.5160)</td>
<td>0.03805 (0.4458)</td>
</tr>
<tr>
<td>Percent of renters</td>
<td>1.07745+ (0.5385)</td>
<td>0.90910+ (0.5072)</td>
<td>0.50765 (0.5180)</td>
<td>1.07797+ (0.5376)</td>
<td>0.89290+ (0.4982)</td>
<td>0.48268 (0.5056)</td>
</tr>
<tr>
<td>Percent employed</td>
<td>0.23166 (0.2036)</td>
<td>0.94816** (0.2292)</td>
<td>0.90222** (0.3022)</td>
<td>0.23357 (0.20327)</td>
<td>0.95572** (0.2438)</td>
<td>0.92497** (0.3129)</td>
</tr>
<tr>
<td>Annual residential energy-efficiency program expenditures per customer in year t</td>
<td>-0.00014 (0.0002)</td>
<td>0.00038 (0.0002)</td>
<td>0.00031 (0.0003)</td>
<td>-0.00012 (0.0002)</td>
<td>0.00040 (0.0002)</td>
<td>0.00032 (0.0003)</td>
</tr>
<tr>
<td>Variable</td>
<td>Model 1</td>
<td>Model 2</td>
<td>Model 3</td>
<td>Model 4</td>
<td>Model 5</td>
<td>Model 6</td>
</tr>
<tr>
<td>-------------------------------------------------------------------------</td>
<td>------------</td>
<td>------------</td>
<td>------------</td>
<td>------------</td>
<td>------------</td>
<td>------------</td>
</tr>
<tr>
<td>Annual residential energy-efficiency program expenditures per customer in year t-1</td>
<td>-0.000046** (0.0001)</td>
<td>-0.000033 (0.0003)</td>
<td>-0.000049** (0.0001)</td>
<td>-0.000037 (0.0003)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Annual residential energy-efficiency program expenditures per customer in year t-2</td>
<td>-0.00028 (0.0004)</td>
<td>-0.00030 (0.0003)</td>
<td>-0.00029 (0.0004)</td>
<td>-0.00032 (0.0003)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Annual residential energy-efficiency program expenditures per customer in year t-3</td>
<td>-0.00066* (0.0003)</td>
<td>-0.00073** (0.0003)</td>
<td>-0.00068* (0.0003)</td>
<td>-0.00078** (0.0003)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Annual residential energy-efficiency program expenditures per customer in year t-4</td>
<td>-0.00150** (0.0004)</td>
<td>-0.00128** (0.0002)</td>
<td>-0.00153** (0.0004)</td>
<td>-0.00132** (0.0003)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Annual residential energy-efficiency program expenditures per customer in year t-5</td>
<td>0.00110 (0.0011)</td>
<td>-0.000019 (0.0009)</td>
<td>-0.00023 (0.0009)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Annual residential energy-efficiency program expenditures per customer in year t-6</td>
<td>-0.00077 (0.0026)</td>
<td>0.00029 (0.0020)</td>
<td>-0.00081 (0.0027)</td>
<td>-0.00086 (0.0026)</td>
<td>0.00051 (0.0021)</td>
<td>-0.00031 (0.0027)</td>
</tr>
<tr>
<td>Time trend</td>
<td>4.06623* (1.5456)</td>
<td>0.98565 (1.5494)</td>
<td>0.68102 (1.6504)</td>
<td>4.04847* (1.5509)</td>
<td>1.13566 (1.5344)</td>
<td>0.91356 (1.6436)</td>
</tr>
<tr>
<td>Constant</td>
<td>FE</td>
<td>FE</td>
<td>FE</td>
<td>FE</td>
<td>FE</td>
<td>FE</td>
</tr>
<tr>
<td>Estimation method</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Cumulative residential energy-efficiency program expenditures per customer in years t-4 through t</td>
<td>N/A</td>
<td>-0.00252** (0.0007)</td>
<td>-0.00234** (0.0005)</td>
<td>N/A</td>
<td>-0.00259** (0.0008)</td>
<td>-0.00247** (0.0005)</td>
</tr>
<tr>
<td>Cumulative residential energy-efficiency program expenditures per customer in years t-6 through t</td>
<td>N/A</td>
<td>N/A</td>
<td>-0.00363** (0.0006)</td>
<td>N/A</td>
<td>N/A</td>
<td>-0.00380** (0.0006)</td>
</tr>
<tr>
<td>Observations</td>
<td>438</td>
<td>422</td>
<td>414</td>
<td>438</td>
<td>422</td>
<td>414</td>
</tr>
<tr>
<td>Within R²</td>
<td>0.64</td>
<td>0.69</td>
<td>0.71</td>
<td>0.64</td>
<td>0.69</td>
<td>0.71</td>
</tr>
</tbody>
</table>
### Status and Direction for Top-Down Work

<table>
<thead>
<tr>
<th>Variable</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
<th>Model 5</th>
<th>Model 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Account for leakage of PA-supported CFLs to municipal utility customers</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
</tr>
<tr>
<td>Number of utilities</td>
<td>35</td>
<td>35</td>
<td>35</td>
<td>35</td>
<td>35</td>
<td>35</td>
</tr>
</tbody>
</table>

**Table 4: Residential PA-Muni Model Results with Cumulated Program Expenditures**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Model 7</th>
<th>Model 8</th>
<th>Model 9</th>
<th>Model 10</th>
<th>Model 11</th>
<th>Model 12</th>
</tr>
</thead>
<tbody>
<tr>
<td>Price of electricity</td>
<td>-0.11740+</td>
<td>-0.12718+</td>
<td>-0.14864*</td>
<td>-0.11658</td>
<td>-0.12643+</td>
<td>-0.14904*</td>
</tr>
<tr>
<td></td>
<td>(0.0690)</td>
<td>(0.0681)</td>
<td>(0.0597)</td>
<td>(0.0694)</td>
<td>(0.0687)</td>
<td>(0.0608)</td>
</tr>
<tr>
<td>Annual HDDs</td>
<td>0.16539**</td>
<td>0.16241**</td>
<td>0.13408*</td>
<td>0.16612**</td>
<td>0.16326**</td>
<td>0.13424*</td>
</tr>
<tr>
<td></td>
<td>(0.0604)</td>
<td>(0.0589)</td>
<td>(0.0535)</td>
<td>(0.0607)</td>
<td>(0.0591)</td>
<td>(0.0545)</td>
</tr>
<tr>
<td>Annual CDDs</td>
<td>0.05933**</td>
<td>0.05957**</td>
<td>0.05676**</td>
<td>0.05930**</td>
<td>0.05955**</td>
<td>0.05666**</td>
</tr>
<tr>
<td></td>
<td>(0.0070)</td>
<td>(0.0067)</td>
<td>(0.0069)</td>
<td>(0.0070)</td>
<td>(0.0068)</td>
<td>(0.0069)</td>
</tr>
<tr>
<td>Median household income</td>
<td>-0.11887</td>
<td>-0.09395</td>
<td>-0.05191</td>
<td>-0.12122</td>
<td>-0.09723</td>
<td>-0.05615</td>
</tr>
<tr>
<td></td>
<td>(0.1235)</td>
<td>(0.1180)</td>
<td>(0.1144)</td>
<td>(0.1236)</td>
<td>(0.1182)</td>
<td>(0.1156)</td>
</tr>
<tr>
<td>Median home values</td>
<td>0.30965**</td>
<td>0.31258**</td>
<td>0.33180**</td>
<td>0.30914**</td>
<td>0.31356**</td>
<td>0.32845**</td>
</tr>
<tr>
<td></td>
<td>(0.1098)</td>
<td>(0.1113)</td>
<td>(0.1057)</td>
<td>(0.1090)</td>
<td>(0.1102)</td>
<td>(0.1043)</td>
</tr>
<tr>
<td>Percent of homes using electricity as the main heating fuel</td>
<td>0.53969</td>
<td>0.52695</td>
<td>0.51898</td>
<td>0.54400</td>
<td>0.53083</td>
<td>0.52412</td>
</tr>
<tr>
<td></td>
<td>(0.6328)</td>
<td>(0.6158)</td>
<td>(0.6110)</td>
<td>(0.6325)</td>
<td>(0.6140)</td>
<td>(0.6059)</td>
</tr>
<tr>
<td>Percent of residential new construction</td>
<td>1.74507*</td>
<td>1.74538*</td>
<td>1.63163*</td>
<td>1.75613*</td>
<td>1.76880*</td>
<td>1.66921*</td>
</tr>
<tr>
<td></td>
<td>(0.8079)</td>
<td>(0.8138)</td>
<td>(0.7724)</td>
<td>(0.8122)</td>
<td>(0.8223)</td>
<td>(0.7825)</td>
</tr>
<tr>
<td>Percent of single-family homes</td>
<td>0.64128</td>
<td>0.61853</td>
<td>0.47979</td>
<td>0.64757</td>
<td>0.62836</td>
<td>0.48614</td>
</tr>
<tr>
<td></td>
<td>(0.6218)</td>
<td>(0.5957)</td>
<td>(0.5479)</td>
<td>(0.6234)</td>
<td>(0.5336)</td>
<td>(0.5422)</td>
</tr>
<tr>
<td>Percent of renters</td>
<td>1.16691*</td>
<td>1.17613*</td>
<td>1.04400+</td>
<td>1.17327*</td>
<td>1.18840*</td>
<td>1.05727+</td>
</tr>
<tr>
<td></td>
<td>(0.5279)</td>
<td>(0.5333)</td>
<td>(0.5307)</td>
<td>(0.5287)</td>
<td>(0.5336)</td>
<td>(0.5297)</td>
</tr>
<tr>
<td>Percent employed</td>
<td>0.33024</td>
<td>0.41445</td>
<td>0.49667*</td>
<td>0.32709</td>
<td>0.41275</td>
<td>0.50391*</td>
</tr>
<tr>
<td></td>
<td>(0.2539)</td>
<td>(0.2692)</td>
<td>(0.2164)</td>
<td>(0.2536)</td>
<td>(0.2710)</td>
<td>(0.2196)</td>
</tr>
<tr>
<td>Cumulated residential energy-efficiency program expenditures per customer</td>
<td>-0.00066</td>
<td>-0.00066</td>
<td>-0.00066</td>
<td>-0.00066</td>
<td>0.41275</td>
<td>0.50391*</td>
</tr>
<tr>
<td>in years t-3 to year t</td>
<td>(0.0005)</td>
<td>(0.0005)</td>
<td>(0.0005)</td>
<td>(0.0005)</td>
<td>(0.0005)</td>
<td>(0.2196)</td>
</tr>
<tr>
<td>Variable</td>
<td>Model 7</td>
<td>Model 8</td>
<td>Model 9</td>
<td>Model 10</td>
<td>Model 11</td>
<td>Model 12</td>
</tr>
<tr>
<td>-------------------------------------------------------------------------</td>
<td>--------------------</td>
<td>--------------------</td>
<td>--------------------</td>
<td>--------------------</td>
<td>--------------------</td>
<td>--------------------</td>
</tr>
<tr>
<td>Cumulated residential energy-efficiency program expenditures per customer in years t-4 through year t</td>
<td>-0.00115+ (0.0006)</td>
<td>-</td>
<td>-</td>
<td>-0.00118 (0.0007)</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Cumulated residential energy-efficiency program expenditures per customer in years t-6 through t</td>
<td>-</td>
<td>-0.00210** (0.0006)</td>
<td>-</td>
<td>-</td>
<td>-0.00222** (0.0007)</td>
<td>-</td>
</tr>
<tr>
<td>Time trend</td>
<td>0.00062 (0.0021)</td>
<td>0.00158 (0.0020)</td>
<td>0.00239 (0.0020)</td>
<td>0.00063 (0.0021)</td>
<td>0.00169 (0.0020)</td>
<td>0.00276 (0.0020)</td>
</tr>
<tr>
<td>Constant</td>
<td>3.69559* (1.5964)</td>
<td>3.32586+ (1.6445)</td>
<td>3.02751* (1.4771)</td>
<td>3.71613* (1.5978)</td>
<td>3.36422* (1.6431)</td>
<td>3.09792* (1.4826)</td>
</tr>
<tr>
<td>Estimation method</td>
<td>FE</td>
<td>FE</td>
<td>FE</td>
<td>FE</td>
<td>FE</td>
<td>FE</td>
</tr>
<tr>
<td>Observations</td>
<td>426</td>
<td>422</td>
<td>414</td>
<td>426</td>
<td>422</td>
<td>414</td>
</tr>
<tr>
<td>Within R²</td>
<td>0.65</td>
<td>0.66</td>
<td>0.67</td>
<td>0.65</td>
<td>0.66</td>
<td>0.67</td>
</tr>
<tr>
<td>Account for leakage of PA-supported CFLs to municipal utility customers</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
</tr>
<tr>
<td>Number of utilities</td>
<td>35</td>
<td>35</td>
<td>35</td>
<td>35</td>
<td>35</td>
<td>35</td>
</tr>
</tbody>
</table>
Table 5: C&I PA-Muni Model Results with Individual Year Program Expenditures

<table>
<thead>
<tr>
<th>Variable</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Price of electricity</td>
<td>-0.00886 (0.0294)</td>
<td>-0.003644 (0.0363)</td>
<td>-0.03208 (0.0404)</td>
<td>-0.01523 (0.0401)</td>
</tr>
<tr>
<td>Annual HDDs</td>
<td>-0.00904 (0.660)</td>
<td>0.00553 (0.899)</td>
<td>0.00758 (0.921)</td>
<td>0.01743 (0.780)</td>
</tr>
<tr>
<td>Annual CDDs</td>
<td>0.05498** (0.0089)</td>
<td>0.05274** (0.0116)</td>
<td>0.05203** (0.0113)</td>
<td>0.04864** (0.0078)</td>
</tr>
<tr>
<td>Mean annual wage (in 2012 $)</td>
<td>-0.02201 (0.0924)</td>
<td>-0.06962 (0.1393)</td>
<td>-0.06342 (0.1441)</td>
<td>-0.04237 (0.1332)</td>
</tr>
<tr>
<td>Percent employed</td>
<td>-0.51223+ (0.3170)</td>
<td>-0.28939 (0.3589)</td>
<td>-0.25874 (0.3600)</td>
<td>-0.40118 (0.3630)</td>
</tr>
<tr>
<td>C&amp;I annual new construction per employee (in sq. ft.)</td>
<td>0.00046 (0.0004)</td>
<td>0.00028 (0.0003)</td>
<td>0.00026 (0.0004)</td>
<td>0.00022 (0.0003)</td>
</tr>
<tr>
<td>Annual C&amp;I energy-efficiency program expenditures per employee in year t</td>
<td>-0.00029+ (0.0002)</td>
<td>-0.00018 (0.0001)</td>
<td>-0.00018 (0.0001)</td>
<td>-0.00017 (0.0001)</td>
</tr>
<tr>
<td>Annual C&amp;I energy-efficiency program expenditures per employee in year t-1</td>
<td>-0.00025* (0.0001)</td>
<td>-0.00024* (0.0001)</td>
<td>-0.00018+ (0.0002)</td>
<td></td>
</tr>
<tr>
<td>Annual C&amp;I energy-efficiency program expenditures per employee in year t-2</td>
<td>-0.00011 (0.0002)</td>
<td>-0.00009 (0.0002)</td>
<td>-0.00008 (0.0002)</td>
<td></td>
</tr>
<tr>
<td>Annual C&amp;I energy-efficiency program expenditures per employee in year t-3</td>
<td>-0.00036** (0.0001)</td>
<td>-0.00033* (0.0001)</td>
<td>-0.00026+ (0.0002)</td>
<td></td>
</tr>
<tr>
<td>Annual C&amp;I energy-efficiency program expenditures per employee in year t-4</td>
<td>0.00010 (0.0001)</td>
<td>0.00011 (0.0002)</td>
<td>0.00044** (0.0001)</td>
<td></td>
</tr>
<tr>
<td>Annual C&amp;I energy-efficiency program expenditures per employee in year t-5</td>
<td>0.00043* (0.0002)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Annual C&amp;I energy-efficiency program expenditures per employee in year t-6</td>
<td>0.00044** (0.0001)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time trend</td>
<td>-0.00851 (0.0053)</td>
<td>-0.00645 (0.0054)</td>
<td>-0.00671 (0.0056)</td>
<td>-0.01127+ (0.0064)</td>
</tr>
<tr>
<td>Constant</td>
<td>-11.68837 (15.0934)</td>
<td>-11.96775 (15.9607)</td>
<td>-11.96775 (15.9837)</td>
<td>-2.29631 (12.8869)</td>
</tr>
<tr>
<td>Estimation method</td>
<td>FE</td>
<td>FE</td>
<td>FE</td>
<td>FE</td>
</tr>
<tr>
<td>Cumulative C&amp;I energy-efficiency program expenditures per customer in years t-3 through t</td>
<td>-0.00091* (0.0004)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Variable</td>
<td>Model 1</td>
<td>Model 2</td>
<td>Model 3</td>
<td>Model 4</td>
</tr>
<tr>
<td>-------------------------------------------------------------------------</td>
<td>----------------------------------</td>
<td>----------------------------------</td>
<td>----------------------------------</td>
<td>----------------------------------</td>
</tr>
<tr>
<td>Cumulative C&amp;I energy-efficiency program expenditures per customer in</td>
<td></td>
<td></td>
<td>-0.00075 (0.0005)</td>
<td></td>
</tr>
<tr>
<td>years t-4 through t</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cumulative C&amp;I energy-efficiency program expenditures per customer in</td>
<td></td>
<td></td>
<td>0.00029 (0.0007)</td>
<td></td>
</tr>
<tr>
<td>years t-6 through t</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>379</td>
<td>379</td>
<td>379</td>
<td>379</td>
</tr>
<tr>
<td>Within R²</td>
<td>0.39</td>
<td>0.40</td>
<td>0.40</td>
<td>0.43</td>
</tr>
<tr>
<td>Number of utilities</td>
<td>36</td>
<td>36</td>
<td>36</td>
<td>36</td>
</tr>
</tbody>
</table>
Table 6: C&I PA-Muni Model Results with Cumulated Program Expenditures

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