

Program Administrators of Massachusetts

**Top-Down Modeling Extended Methods Review
Final**

February 16, 2017



TABLE OF CONTENTS

1. Introduction	1-1
2. Background and Objectives	2-1
3. Summary of Approach	3-1
3.1 Literature Review	3-1
3.2 Expert Interviews	3-1
4. Findings	4-1
4.1 Treatment of weather variation	4-1
1.1.1 Weather related issues in econometrics	4-3
4.1.1 Weather variation conclusions	4-5
4.2 Treatment of Lag Effects	4-6
4.3 Recession	4-10
4.3.1 Overview and context	4-11
4.3.2 Recession and the top-down model	4-11
4.3.3 Treating recession in the top-down model	4-12
4.4 Fixed effects	4-12
4.5 Endogeneity	4-13
4.5.1 Electricity and gas prices	4-14
4.5.2 Program spending	4-14
4.6 Error bounds	4-15
4.7 Constructing portfolio savings	4-15
4.8 Self-selection	4-15
4.9 Spillover	4-16
5. Conclusions and Guidelines	5-1
5.1.1 Weather modeling	5-1
5.1.2 Lag effects	5-1
5.1.3 Recession	5-1
5.1.4 Fixed effects	5-2
5.1.5 Endogeneity of electricity and gas prices	5-2
5.1.6 Endogeneity of program spending	5-2
5.1.7 Error bounds	5-2
5.1.8 Constructing portfolio savings	5-3
5.1.9 Self-selection	5-3
5.1.10 Spillover	5-3
APPENDIX A: List of Relevant Literature	A-1
APPENDIX B: Name, Affiliation, and Research Area of Interviewees	B-1



APPENDIX C: Top-down Modeling Interview Guide.....C-1

1. INTRODUCTION

This memo is prepared by DNV GL, NMR, and Tetra Tech (Team) for the Massachusetts Program Administrators (PAs) and the Energy Efficiency Advisory Council (EEAC) consultants as part of the Massachusetts Special and Cross Cutting Research Contract. The memo presents the findings and recommendations from: (1) an extended literature review of top-down modeling methodologies applied in studies in both the energy and non-energy sectors, and (2) the insights and findings from in-depth interviews with nationally recognized experts in the field of econometrics,¹ macroeconomics,² and top-down modeling. Based upon the results of the literature review and expert interviews, we provide an assessment of the advantages and disadvantages of different theoretical methodologies and approaches, and provide recommendations for enhancements to future top-down efforts.

This extended methods memo includes the following sections:

- Section 2 – Background and objectives
- Section 3 – Summary of approach
- Section 4 – Findings
- Section 5 – Conclusions and recommendations.

¹ Econometrics is the application of statistical and mathematical techniques in solving problems as well as in testing and demonstrating theories (Dictionary.com).

² Macroeconomics is the branch of economics dealing with the broad and general aspects of an economy, as the relationship between the income and investments of a country (Dictionary.com).

2. BACKGROUND AND OBJECTIVES

In March 2015, DNV GL and NMR completed a top-down modeling methods study to assess the usefulness of a top-down approach for evaluating energy efficiency programs in Massachusetts.³ The goal of top-down modeling is to isolate the effect of program activity from other natural changes and policy variables across groups of energy efficiency programs in a regional area or service territory, rather than estimating ex post changes by program, measure, or end-use (bottom up). Top-down models estimate changes in energy consumption over time that are attributable to programmatic interventions across multiple utilities.

As part of the 2015 study, DNV GL and NMR performed the following tasks:

1. Literature review to help assess top-down modeling methods.
2. PA-Muni pilot study (involving both residential and C&I electric models) to analyze changes in energy consumption in PA and Municipal Utility territories to isolate the effects of PA energy efficiency programs.
3. Commercial and industrial (C&I) PA-data pilot study to develop a top-down model using account-level billing and program tracking data for 2011–2013 provided by the PAs. Due to the limited C&I time series, this study did not estimate net savings but instead investigated whether the data provided evidence of a relationship between programmatic activity and consumption, and recommended modeling techniques that could be used in future evaluation efforts.

The objective of this extended methods review was to build on the 2015 study and conduct a methodological review of the macroeconomic and econometric literature to explore approaches to selected methodological issues, and how these approaches might inform the C&I PA-data and PA-Muni (both residential and C&I) models.

During the recent review, we focused on the following methodological considerations:

- What is the best approach for handling seasonal weather variation? In particular, is weather-normalized annual data preferred over weather terms in the econometric model?
- How should the top-down models account for the cumulative effects of energy efficiency programs over time? Are cumulative lagged terms, individual lagged terms, or indexed lagged terms preferred?
- How should models determine the impacts on energy usage for the recession years (2007–2009)?
- What kinds of fixed effects terms, if any, should be included in the top-down models?
- What concerns are there related to key endogeneity problems in our modeling approach, and how should they be addressed? These problems arise because an independent variable (program activity or energy price) is partially determined by the dependent variable (consumption) itself, which builds a bias into the regression if not addressed.

³ Massachusetts Electric and Gas Program Administrators Top-down Modeling Methods Study—Final Report, March 31, 2015, Prepared by: DNV GL and NMR.

- What are the considerations for determining error bounds from the models' results?

Some additional questions related to top-down model specifications were not directly addressed in the literature review or interviews. For these questions, we provide general considerations based on information indirectly gathered during our investigation of the aforementioned research questions. Detailed specification will depend on the given models. The additional questions were:

- What are the most appropriate methods for constructing portfolio-level savings estimates?
- Are there ways to mitigate self-selection bias in the modeling?
- Are there ways to mitigate bias due to cross-geography spillover in the modeling?

The findings from our research are summarized in this document.

3. SUMMARY OF APPROACH

This section summarizes the research approach employed for the literature review and expert interviews to obtain information to support our exploration into the research questions discussed in Section 2.

3.1 LITERATURE REVIEW

This review explored several unanswered questions that emerged from the 2015 study and describes the relationship of these findings to our current top-down modeling efforts. In the 2015 methods review, the team focused on literature from the energy industry and on a broader set of research topics pertaining to the use of top-down modeling of energy efficiency program impacts. The recent review expands on the 2015 review to explore the advantages and disadvantages of econometric and macroeconomic methods adopted in the energy industry and in other fields to address a more refined set of research questions. Most of the reviewed macroeconomic and econometric studies do not directly estimate or forecast energy consumption, but rather compare, contrast, and build on existing statistical methods to solve a given research question. However, the peer-reviewed articles were selected—for the most part—because parallels could be drawn to the objectives of the top-down modeling the team is aiming to conduct for the PAs. The selected literature spans fields including energy economics, finance, health, banking, and climate change modeling. It also incorporates selected top-down evaluation studies published after the 2015 study was released, as they were not available for the literature review conducted as part of that study.

While reviewing alternative econometric approaches, we focused on peer reviewed literature that informed our topic areas (construction of portfolio-level savings estimates, weather variation, the 2007–2009 recession, cumulative program effects, treatment of lags, and errors).

3.2 EXPERT INTERVIEWS

In a parallel effort, the team identified and interviewed industry experts with applied experience in econometrics and top-down modeling methods. We paid special attention to select experts who were both leaders in their respective fields and had targeted experience addressing the topics of interest, such as weather, recession, lagged variables, cumulative program effects, errors, and other topics that emerged from the literature review. We included experts who have used these methods for net savings analysis similar to our prior work for the PAs, as well as those who are experienced with other energy and non-energy applications of panel data regression analysis.

The team, together with input from the PAs and EEAC, developed an in-depth research guide for the interviews. The interview guide is provided in APPENDIX C: Top-down Modeling Interview Guide. APPENDIX B: Name, Affiliation, and Research Area of Interviewees, contains the full list of interviewees and their affiliations.

4. FINDINGS

4.1 TREATMENT OF WEATHER VARIATION

Weather is an important consideration in top-down modeling as temperature is highly correlated with consumption, particularly in the residential sector. Weather is also included in econometric and macroeconomic models that consider the causal effect weather has on the outcome of interest. The findings in this section address the research question the team posed regarding how weather data should be incorporated into the study. This question asks:

What is the best approach for handling seasonal weather variation? In particular, is weather-normalized annual data preferred over weather terms in the econometric model?

Several categories of research use weather as an independent variable. These include: macroeconomic demand and home production models, climate change forecasting and simulation, utility revenue forecasts, top-down energy efficiency program evaluations, and studies that look for the influence of weather on financial markets, disease, and mortality.

Most studies the team reviewed outside the utility and energy efficiency fields include a degree day term in the models but do not normalize within years (i.e., by month). Weather terms are always included as an independent variable, on the right hand side of the equation, which aligns with the results of the expert interviews discussed later in this section. In some cases, annual weather terms are used because researchers determined that annual data was sufficient, or they did not have access to data at higher levels of granularity. Where normalization does occur in the literature, the review confirmed that fixed-based degree day normalization is the standard proxy for weather. However, many of the articles reviewed, including studies of energy consumption, expanded on the traditional physical model. Further, most articles reviewed do not discuss the choice of either the degree-day metric or a normalization method.

Those articles that did conclude weather had a significant effect on the outcome of interest provided details concerning two topics related to the inclusion of weather in the model:

1. The linearity of weather and consumption
2. The error in this relationship and what to do about it.

The 2015 top-down study noted that under extreme temperature conditions, the relationship between ambient temperature and energy consumption does not have the same slope as during average conditions. At a very high (for cooling) or low (for heating) temperature, the consumption becomes essentially flat, as HVAC equipment operates at full capacity. An opposite effect can occur when some households only use their cooling or supplemental electric heating at extreme temperatures. The relationship will differ by household, or account. Economists noted similar distinctions between average (normal) and extreme (abnormal) weather, but expanded the standard physical model to correct for the change in slope that they attributed to occupant behavior. The way HVAC equipment is used validates these observations.

On the other hand, observed nonlinearity between consumption and degree-days also results from using a fixed degree-day base rather than identifying the appropriate breakpoint for each home or for an aggregate. Using a normalization model grounded in the physical relationship is arguably more appropriate than incorporating squared or log degree-days using a base of 65 degrees Fahrenheit. Moreover, the

extreme day conditions usually affect only a handful of days, so that the nonlinearity on those days does not necessarily introduce much error into the annual weather normalized value.

The key question regarding weather normalization addressed in our review is whether it is appropriate to weather-normalize consumption using monthly data prior to fitting the top-down model, where each observation is an annual value. We also considered different approaches practitioners took to modeling weather.

below provides a summary of the various weather-related articles reviewed and the subject areas in which those studies were conducted.

Table 4-1. Weather Related Articles Reviewed by Research Area

Author	Study	Model Specificati	Stats Technique	Energy	Macro economics	Finance	Health
Auffhammer, Maximilian	"Using weather data and climate model output in economic analyses of climate change," Cambridge, Mass: National Bureau of Economic Research, 2013.			X	X		
Filippini, Massimo, and Lester C. Hunt	"Measurement of energy efficiency based on economic foundations", Energy Economics: Supplement 1. 52 (2015): S5-S16	X	X	X	X		
Filippini, Massimo, and Lester C. Hunt.	"Measuring persistent and transient energy efficiency in the US". Energy Efficiency. 9 (2016): 663-675	X	X	X	X		
Hunt, Lester C., and Yasushi Ninomiya	"Unravelling Trends and Seasonality: A Structural Time Series Analysis of Transport Oil Demand in the UK and Japan". The Energy Journal. 24 (2003).	X	X	X	X		
Filippini, Massimo, and Lester C. Hunt	"US residential energy demand and energy efficiency: A stochastic demand frontier approach". Energy Economics. 34 (2012): 1484-1491.	X	X	X	X		
Dimitropoulos, J, Hunt, LC, and Judge, G.	Estimating underlying energy demand trends using UK annual data. Routledge Taylor & Francis LTD, 2005.	X	X	X	X		
Hunt, Lester C., and David L. Ryan.	"Economic modelling of energy services: Rectifying mis-specified energy demand functions". Energy Economics. 50 (2015): 273-285.	X	X	X	X		
Elkhafif, Mahmoud A.T	"An iterative approach for weather-correcting energy consumption data". Energy Economics. 18: (1996) 221-230.	X	X	X	X		
Won, S. J., X. H. Wang, and H. E.	"Climate normals and weather normalization for utility regulation". Energy Economics. 54 (2016):		X	X	X		

Author	Study	Model Specificati	Stats Technique	Energy	Macro economics	Finance	Health
Warren	405-416						
Alberini, Anna, and Charles Towe	"Information v. energy efficiency incentives evidence from residential electricity consumption in Maryland," Zürich: CER-ETH, 2015.	X	X	X			
Matt Kotchen, Grant Jacobsen	"Do Building Codes Save Energy? Evidence from Residential Billing Data in Florida," <i>The Review of Economics and Statistics</i> , 95 (2013): 34-49.	X	X	X			
Kotchen, Matthew J.	"Longer-run evidence on whether building energy codes reduce residential energy consumption." (2016). <i>Journal of the Association of Environmental and Resource Economists</i> .	X	X	X			
Braga AL, A Zanobetti, and J Schwartz.	"The effect of weather on respiratory and cardiovascular deaths in 12 U.S. cities". <i>Environmental Health Perspectives</i> . 110 (2002): 859-63.	X	X				X
Hirshleifer, D. and Shumway, T.	Good Day Sunshine: Stock Returns and the Weather. <i>The Journal of Finance</i> , 58 (2003): 1009–1032. doi:10.1111/1540-6261.00556	X		X		X	
Fels, Margaret F., and Miriam L. Goldberg	Using the scorekeeping approach to monitor aggregate energy conservation". <i>Energy and Buildings</i> . 9 (1986): 161-168.	X	X	X			

The following section reviews several of these articles and includes the authors' observations, and provides discussion about the pros and cons of how weather is treated in their models.

1.1.1 Weather related issues in econometrics

Use of weather data in economic analyses

In "Using weather data and climate model output in economic analyses of climate change," Aufhammer (2013) provides guidance on how to avoid common weather related mistakes he has observed in econometric models, especially in studies and simulations of the economics of climate change. Aufhammer describes the effects of missing sequential temperature data and missing weather variables. We note that Massachusetts has enough reliable weather stations, such that missing data can be collected from proximate stations. However, when weather stations close, substitution over larger geographical areas must be considered.

He explains that when two or more weather variables are used, such as temperature and precipitation, correlation can be both positive and negative depending on location and season within regional areas. Measurable bias may be introduced if one or more correlated weather variables are omitted.

Another area of caution he provides is spatial correlation, which occurs when weather variables are assumed to be correlated to the study area but are not. This effect may occur when using gridded weather products that are defined by latitude and longitude.⁴

The Massachusetts model uses town or metropolitan level weather data, so that the variables used are good reflections of the weather affecting the consumption.

Macroeconomic demand models

Energy economists have frequently used energy intensity as a proxy of energy efficiency in macroeconomic demand models. There is a large body of work from a group of economists led by Filippini and Hunt (2015, 2016) who propose a theoretical framework for energy demand models that accounts for traditional factors influencing energy demand: economic activity, energy price, and air temperature. In almost all cases, mean daily temperatures are aggregated to annual heating degree days (HDD) and cooling degree days (CDD) and placed on the right side of the equations. In each case, consumption is assumed to depend on the combined effect of weather and other factors, and these effects are estimated together within the model.

Macroeconomic methods for program evaluation

Alberini and Towe's (2015) top-down ex-post evaluation of heat pump rebates and energy audits in Maryland calculates program effects with the goal of estimating reductions in CO₂ emissions. Alberini tries to replicate a quasi-experimental design with a retrospective study and identifies a control group after treatment. To optimize the matching process when defining the control group, the study tested the combination of house characteristics and pre-period consumption. Matching reduced error, and neither consumption nor building type alone was sufficient to identify program effects.

Alberini and Towe aggregate daily HDDs and CDDs to seasonal totals at the building level, and place them on the right side of the regression equation. Seasonal totals are used because they did not know when the customers received the treatment.

The health effects of weather

Braga and Zanobetti (2002) studied "The effect of weather on respiratory and cardiovascular deaths in 12 U.S. cities." Weather is not normalized since the research investigates the direct correlation of average temperature to deaths from respiratory and cardiovascular diseases. Still, the authors controlled for the same variables described in other articles we reviewed that do normalize weather, such as cumulative effects of temperature. In this case, cumulative effects were not affecting behavioral responses, but a physical one, such as mortality. For example, hot days had twice the effect on heart attacks than did cold days. Greater temperature variance was associated with respiratory deaths.

The authors captured mean daily temperatures from the closest weather station 21 days before each death and controlled for season and temperature trend (up or down), day of the week, and barometric pressure. This mirrors many of the studies we reviewed, which extend the linear relationship between temperature and the outcome of interest. The authors also apply lag functions because temperature has an impact on deaths not only on one specific hot or cold day, but also on the days following. They also describe the opposite effect, where deaths that occur in Day 1 will depend on the temperature in Day -1.

⁴ NOAA, Physical Science Division, <https://www.esrl.noaa.gov/psd/data/gridded/tables/temperature.html>.

Explaining financial market returns

Over the last 20 years, many economists have researched the impacts of weather on the stock market. In “Good Day Sunshine: Stock Returns and the Weather,” Hirshleifer investigates the effect of morning sunshine on daily stock market returns in 26 countries from 1982–1997. Researchers found that when controlling for sunshine, rain and snow are unrelated to returns. Although the correlation with sunshine was strong, the economic effect was small and transaction costs eliminated the gains associated with weather-based investment strategies.

The model uses a dummy variable for sunny and cloudy days, which is the National Ocean and Atmospheric Administration (NOAA) variable “sky cover.” Hirshleifer states that using eight years of weather data is too short a period because of statistical noise. Instead, he found a much stronger effect when measuring multiple cities, which introduced a higher level of variability in both returns and weather effects. As of 2016, there is consensus that the relationship of weather to market returns is inconclusive.

Additional perspectives from interviews

As indicated by the discussion above, most econometric analyses do not include weather normalization as a separate step from the overall modeling. That is, weather terms are commonly included on the right-hand side of the annual model. However, most of these studies do not have ready access to monthly data that can be used for normalization. When the idea of conducting weather normalization as a first step was suggested to interviewees, most of them indicated that they saw merit in either approach. Those who recommended a focus on understanding causal relationships tended to favor the separate weather normalization step. A reason given to not weather normalize was that the response to weather is interactive with the response to other economic factors, so these effects should be estimated jointly. Several of the interviewees pointed out that the separate weather normalization step introduces additional error to the dependent variable, consumption. One interviewee cited this as a reason that the normalization step is inappropriate. Other interviewees suggested that the additional error should be addressed by using robust standard error calculations to account for the heterogeneous variance introduced by the normalization. One interviewer pointed out that the standard error of normalized annual consumption tends to be quite small (Fels and Goldberg’s 1986 report standard errors around 1 percent for aggregate regressions) and some interviewees felt this was a compelling justification where the benefits would outweigh the disadvantages. The incremental error introduced is much smaller than the level of variability that would otherwise have to be accounted for by annual weather terms.

4.1.1 Weather variation conclusions

Massachusetts top-down modeling has an advantage compared to many, if not most, macroeconomic and econometric studies in that monthly consumption data is available to the researchers, and can be normalized at a higher level of granularity (with lower error) than with data available in annual form only. Further, the ability to calculate a variable degree-day base in the normalization model accounts for the effects of underlying energy trends such as behavior, which is not possible with a fixed degree-day base, i.e., a 65 °F heating degree-day base. Below we summarize the pros and cons of weather normalizing.

Reasons to use weather normalized consumption:

- Monthly modeling provides a much more accurate accounting of the effect of weather on consumption than does an annual value. While the monthly degree-day model omits some of the detailed effects of other weather variables and of differing patterns at extreme high or low temperature, these effects on annual consumption are likely to be relatively minor. Aggregate

normalized annual consumption is estimated with a high degree of accuracy by variable degree-day models.

- If the number of years in the regression is small, there is a high potential for unrelated factors to be confounded with weather variation just by chance, resulting in coefficients that are not meaningful.

Reasons not to weather normalize:

- Weather normalization means that the dependent variable is the predicted consumption for that year, at normal weather. The prediction error becomes an additional error source in the dependent variable.
- Consumption in each year depends on the combined effects of weather and other factors, and these effects should be jointly estimated.
- The weather normalization models are themselves imperfect.

On balance, our assessment is that weather normalization as a separate step is a good idea, particularly for relatively short time series and it is worth exploring whether the model structure can be improved. However, improvements to the weather modeling are likely to have only minor effects on the overall model.

4.2 TREATMENT OF LAG EFFECTS

Our study reviewed the use of lagged terms used in top-down models in both the energy sector and other research applications including macroeconomic forecasting, finance, public policy analysis, and environmental and health studies. In this review, the team investigated the following research questions regarding the inclusion of lagged effects:

How should the top-down models account for the cumulative effects of energy efficiency programs over time? Are cumulative lagged terms, individual lagged terms, or indexed lagged terms preferred?

With respect to measuring savings from energy efficiency programs, there are two general reasons for using lag effects in the model:

1. Energy efficiency measures installed in a specific year generate savings in successive years
2. Economic variables from one period may have continuing effects in later periods.

The first issue was addressed primarily through the interviews while the second issue was examined as part of the literature review.

The Team reviewed 14 studies related to models with lagged variables. Table 4-2 shows the studies reviewed and key research issue addressed in each study.

Table 4-2. Articles Addressing Lagged Variables

Author	Study	Model Specification	Stats Technique	Energy	Macro economics	Finance	Public Policy	Environmental	Health

Author	Study	Model Specification	Stats Technique	Energy	Macro economics	Finance	Public Policy	Environmental	Health
Jorgensen, Jason B.	“Modelling and Forecasting Electricity Consumption in the U.S. Mountain Region”, 2012, George Washington University	X		X					
Lu, MJ, CYCL Lawell, S. Chen	“The Effects of Energy Policies on Energy Consumption in China”, 2015, University of California- Davis.	X		X					
Ozark, Ilhan	“Measuring the Impact of Energy Consumption and Air Quality Indicators on Climate Change: Evidence from the Panel of UNFCC Classified Countries”, 2015, Environmental Science and Pollution Research.		X	X					
Metcalf, Gilbert E.	“Energy Conservation in the United States: Understanding its Role in Climate Change”, 2006, National Bureau of Economic Research.	X	X	X				X	
Metcalf, Gilbert E.	“An Empirical Analysis of Energy Intensity and its Determinants at the State Level”, 2008, MIT Joint Program on the Science and Policy of Global Change.	X		X					
Hassett, Kevin A., and Gilbert E. Metcalf	“Energy Tax Credits and Residential Conservation Investment: Evidence from Panel Data”, 1995, Journal of Public Economics.	X		X			X		
Kaufman, Robert R., and Alex Segura-Ubiergo	“Globalization, Domestic Politics and Social Spending in Latin America: A Time-series Cross-section Analysis”, 2001, World Politics.	X			X		X		
Campbell, John Y., Andrew W. Lo and Archie Craig MacKinlay	“The Econometrics of Financial Markets”, 1997, Princeton University Press.	X				X			

Author	Study	Model Specification	Stats Technique	Energy	Macro economics	Finance	Public Policy	Environmental	Health
Achen, C.	“Why Lagged Dependent Variables Can Reduce the Explanatory Power of Independent Variables”, 2001.	X	X		X				
Dimitropoulos, J. Hunt, LC and Judge	“Estimating Underlying Energy Demand Trends Using UK Annual Data”, 2005, SEEDS.	X		X					
Scholtens, Bert, and Veldhuis Rineke	“How Does the Development of the Financial Industry Advance Renewable Energy? A Panel Regression of 198 Countries Over Three Decades”, 2015, Beiträge Zur Jahrestagung Des Vereins Für Socialpolitik.	X				X			
Moffitt, Robert	“Identification and Estimation of Dynamic Models with a Time Series of Repeated Cross Sections”, 1993, Journal of Econometrics.	X							X
Bhargava, Alok	“Econometrics, statistics, and computational approaches in food and health science”, 2006, World Scientific.		X						X
Bhaskaran, Krishnan, Antonio Gasparrini, Shakoor Haiat, Liam Smeeth and Ben Armstrong	“Time series regression studies in environmental epidemiology”, 2013, Oxford University Press.		X					X	X

In the 2015 top-down modeling work conducted in Massachusetts using muni and PA data, models were fitted with several years of lagged program activity variables. The coefficients of the program activity variable lagged up to six years and were found to be statistically significant, but there was no consistent pattern in the coefficients across the different lags. This result suggests that obtaining separate meaningful lags may be difficult because of the high correlation among the activity variables of different lags.

The literature review and interviews both point to the importance of understanding the causal relationship between the dependent variable and the independent variables in the model (e.g., program spending, level of investment, fuel prices) when specifying all lag structures. For the program activity variable, our review and interviews considered three broad approaches to dealing with the effect of prior year program activity on a later year's consumption.

1. Estimate separate lags and test for how many lags should be included, similar to the prior work with the muni-PA model
2. Use a single variable representing cumulative activity, such as cumulative spending
3. Use a single variable representing cumulative surviving savings, based on ex ante or ex post estimates together with agreed EULs.

Both the literature review and interviews indicate that multicollinearity is a problem when estimating individual lags (Achen 2001, Kaufman Segura-Ubiergo 2001, Keefe and Kelley 2006). The interviewees agreed that the use of surviving savings imposes effective useful life (EUL) assumptions, but some still felt this is making better use of available information than trying to get separate lag effects. At least one commenter felt strongly that if separate lags are estimated, tests should be done to determine the lag period.

One interviewee favored an indexing method, where the program activity variable in the model is the difference in spending or estimated savings (or expenditures) compared to the first year in the model. However, another interviewer felt strongly that this approach was limiting because it imposes an untestable lag structure on the model.

A suggestion from both the literature and interviewees is to take first differences as a means of mitigating lag effects of all kinds and reducing sensitivity to lag specification (Kaufman Segura-Ubiergo 2001). That is, regress year-to-year change in consumption on year-to-year change in the predictors. One interviewee suggested that a good test for the model specification and robustness of the result is that the original and first-difference models produce similar coefficients. The limitation of differencing is that taking the first difference further limits the length of time series available for modeling by one period.

Including a lagged predictor variable when in fact there is not a causal relationship between the dependent variables and that predictor in prior periods can also be problematic. Achen showed that when a lagged dependent variable is included in a model when there truly is no structural relationship, it will produce what appears to be a well-fitting model but with insignificant and counter-intuitive coefficients. In these cases, it is important to consider the difference between a predictive model and an explanatory model. In general, the most recent observation can be a strong predictor of the current. However, if the goal is to decompose consumption into various drivers to isolate the effect of a particular driver, inclusion of the lagged dependent variable as a predictor will dampen the effects of the other terms; the predictive power may be improved but the explanatory power is diminished.

Metcalf (2006) looked at the relationship between energy intensity and carbon impacts. He estimated a series of regressions testing various lagged electricity prices. He found that long run price effects had a larger impact on consumption and energy intensity than short-term price fluctuations.

As with the issues discussed with the other research questions of interest, the literature and industry experts do not point to a single framework for the inclusion of lagged variables. This finding is consistent across various types of studies including energy, climate change, health policy, etc. However, there is a

common recommendation amongst these studies: first analyze the causal relationship between periods and then explore various model specifications to ensure an overall well-fitting model with logical results.

4.3 RECESSION

A recession is a decline in economic activity spanning several months, or in the case of the most recent recession, years. The most recent recession officially lasted 19 months and spanned December 2007–June 2009, resulting in the coin “Great Recession.”

The Bureau of Economic Analysis (BEA) of the U.S. Department of Commerce defines the official beginning and end of a recession. The Great Recession exhibited unique patterns beyond its length and spawned a large body of research aimed at understanding its true effects, and why it challenged traditional methods for calculating standard economic indicators (Lytras and Bell, 2015) (Cooper, 2012) (Ng and Wright, 2013). Our team thus wanted to better understand how to model recession effects, responding to the research question which asks:

How should models determine the impacts on energy usage for the recession years (2007-2009)?

The Team reviewed seven studies which addressed recessionary effects. Table 4-3 shows the studies reviewed and key research issue addressed in each study.

Table 4-3. Articles Addressing Recessionary Effects

Author	Study	Model Specification	Stats Technique	Energy	Macro economics
Lytras, Demetra and Bell, William R.	“Modeling Recession Effects and the Consequences on Seasonal Adjustment.” U.S. Census Bureau, 2013.	X	X		X
Cooper, Daniel.	“U.S. Household Deleveraging: What Do the Aggregate and Household-Level Data Tell Us?” Public Policy Briefs, Federal Reserve Bank of Boston. 12-2, (2012).	X	X	X	X
Ng, Serena, and Jonathan H. Wright.	Facts and challenges from the Great Recession for forecasting and macroeconomic modeling (Cambridge, Mass: National Bureau of Economic Research, 2013). http://papers.nber.org/papers/w19469 .	X	X		X
Nelson, Hal T., David von Hippel, Tom Peterson, and Roman Gaagulagian.	“The Great Recession or progressive energy policies? Explaining the decline in US greenhouse gas emissions forecasts”. Journal of Environmental Planning and Management. 59 (3): 480-500 (2015).	X	X	X	X
Aguiar, Mark, Hurst, Eric, Karabarbounis, Loukas	“Time Use During the Great Recession; Identifying the relationship of household spending and labor supply,” American Economic Review 103 (5): 1664–1696, 2013.	X	X		X
Hunt, Lester C., and Yasushi Ninomiya.	“Unravelling Trends and Seasonality: A Structural Time Series Analysis of Transport Oil Demand in the UK and Japan”. The Energy Journal. 24, 2003.	X	X	X	X

Author	Study	Model Specification	Stats Technique	Energy	Macro economics
Poudyal, Neelam C., Bamadev Paudel, and Michael A. Tarrant	“A time series analysis of the impact of recession on national park visitation in the United States,” <i>Tourism Management</i> . 35: 181-189, 2013.	X	X		X

4.3.1 Overview and context

The standard measure of economic activity is Gross Domestic Product (GDP). GDP is composed of four primary indicators: income, wholesale-retail sales, employment, and industrial production. The indicators of GDP given the greatest weight are real personal income and employment.

A recession begins at the low point of a prolonged fall (contraction) in economic activity and ends as the economic activity increases enough to reverse out of a trough toward a new peak (expansion). Recessions have been rare over the last several decades, and were short when they have occurred. However, the Great Recession was both long and deep (Ng and Wright 2013). Further, recovery was delayed and the effects of the recession continued well beyond 2009 (Ireland 2010). The Massachusetts Budget and Policy Center reported in 2016 that “during the current economic recovery (since 2009), wages for low and middle wage workers are still lower than they were at the end of the recession.”⁵

4.3.2 Recession and the top-down model

Given its dramatic impact on household spending and income (and other behaviors), we find that recession should be accounted for in a top-down model of energy consumption. The literature on the “Great Recession” agrees on the following:

- Recessionary effects preceded and extended well beyond the official dates of 2007–2009
- Traditional economic indicators did not follow patterns of post-war recessions
- Indicators of personal income and employment status were as significant as commercial and industrial production
- Changes in household behavior and energy consumption were significant at the household level (this effect is described below in more detail).

“The Great Recession or progressive energy policies?” was one of the few applied studies that estimated the recessionary effects to forecasts of greenhouse gas emissions. When examining the 23 percent drop in 2030 forecasts of US greenhouse gas emissions from 2007–2011, Nelson (2015) found that the Great Recession was responsible for approximately 67 percent of the 2008–2009 decline in emissions due to the severe decline in GDP growth. The authors included the effects of the American Recovery and

⁵ The Massachusetts Budget and Policy Center, *State of Working Massachusetts 2016*, <http://www.massbudget.org/reports/swma/>.

Reinvestment Act of 2009 and ACEEE's estimates of energy savings from the Energy Independence and Security Act of 2007.

To identify the relationship between household spending and labor supply, Aguiar, Hurst, and Karabarbounis (2013) used data from American Time Use Survey between 2003 and 2010 (the first time such granular data was available) to show that home production absorbed roughly 30 percent of foregone market work hours. Home leisure (sleeping and TV) absorbed roughly 50 percent of foregone market work hours and the remaining time was spent on shopping, childcare, education, and health activities. This provides evidence that daily in-home activity increased during the recession and may have increased household energy consumption during the same period.

The Hunt and Nimomiya study described earlier, "Unravelling Trends and Seasonality," also considered recession when estimating energy demand and used a dummy variable for recession but did not go into an in-depth discussion on the topic.

In a 2013 study on the impact of recession on leisure, Poudyal, Bamadev and Tarrant (2013) identified which measures of recession were the best indicators of types of visits to National Parks (i.e., short vs. long) to inform future planning. The impacts of recession were modeled using both macroeconomic data from secondary sources and survey data that measured consumer confidence. The macroeconomic variables with the highest explanatory power were unemployment (income), savings (net worth), consumer confidence, and the rate of expected inflation (which aligns with Cooper, 2012). The authors concluded that, although the macroeconomic analysis using secondary data (top-down) showed a negative correlation with the demand for park visitation, the survey data (bottom up) had a greater explanatory power in modeling the impact of recession on park visits. In this case, aggregate panel regressions give broad results but were not as precise as bottom-up estimates.

4.3.3 Treating recession in the top-down model

The literature is unanimous in declaring that the Great Recession had profound and prolonged impacts on all sectors of the economy. It is also clear that declaring firm start and end dates of the recession for modeling purposes may not account for effects that preceded and/or extended beyond the official 2007–2009 recessionary period.

One way to solve for this is to account for recessionary impacts as they occur and to incorporate the most significant household economic indicators, real personal income, and employment into the top-down model. Current and historical values of these (and all other) indicators are readily available from the Federal Reserve Bank of Boston, and are reported at the state and metropolitan level.⁶ This is especially useful for the current model since there is evidence that Massachusetts wages for the lowest 30 percent of earners may have only recently returned to pre-recession levels.

4.4 FIXED EFFECTS

Fixed effects terms in panel regressions such as ours are dummy variables for geographic unit and for time (year). The panel regressions reviewed in the literature typically included geographic and time fixed effects.

⁶ Data is available at both the state and metropolitan level (Barnstable, Boston-Cambridge-Nashua, Leominster-Gardner, New Bedford, Pittsfield, Springfield, Worcester) from the Federal Reserve Bank of Boston, New England Economic Indicators Database, <https://www.bostonfed.org/data/data-items/new-england-economic-indicators-database.aspx>.

The team therefore wanted to better understand how to handle fixed effects and addressed the following research question:

What kinds of fixed effects terms, if any, should be included in the top-down models?

Practitioners generally agreed that these terms are important. However, some observed that inclusion of these dummy variables, particularly time fixed effects, can mask an effect of interest. One alternative identified is to use a trend term, which allows for a consistent change over time but does not allow time to have its own term. There are also formal tests available for the need to include the fixed effect. Taking first differences can also mitigate the need for trend terms or other fixed effects terms. As such, to address fixed effects in top-down modeling the following should be considered:

1. Include fixed effect terms, particularly for geographic units
2. Explore specifications that omit time fixed effects, or that 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.

4.5 ENDOGENEITY

Endogeneity is a broad class of problems in econometric regression. The technical issue is that the regression estimates are biased if a predictor variable is correlated with the unexplainable portion of the dependent variable. In considering how endogeneity may impact our modeling results, the team investigated the following:

What concerns are there related to key endogeneity problems in our modeling approach, and how should they be addressed? These problems arise because an independent variable (program activity or energy price) is partially determined by the dependent variable (consumption) itself, which builds a bias into the regression if not addressed.

A general structure that creates this problem is one where the program or policy variable is itself a function of consumption, while consumption is a function of the program activity or other policy variable. Horowitz (2012) addresses this feedback loop as a set of simultaneous equations, for which he uses two-stage least squares in the estimation.

We focused our review on two specific sources of endogeneity:

1. The electricity and gas price variables we have used are calculated as the ratio of utility revenues to the quantity of energy sold. This essentially puts the dependent variable, aggregate consumption, on the right-hand side of the regression equation as the denominator of the calculated price.
2. Program spending has been used as the variable representing the level of program activity, the key variable used to extract the net program effect from the regression. Since spending levels are set as a percentage of utility revenues, which in turn depend on consumption, use of program spending as a predictor variable again builds into the regression equation a correlation between a predictor and unexplained variation in the dependent variable.

4.5.1 Electricity and gas prices

Practitioners indicated that prices tend not to vary greatly, thus this issue is not worth a lot of attention. Suggestions included the following:

- Use a generic price rather than one that is directly related to the dependent variable for a geography and time, such as by using service territory level price when observations are at finer geographic units
- Use the marginal tariff price if available
- The actual driver of consumption is more related to long term price expectation than to current average price, which is not directly observable by consumers.

Based on the interview results, possible approaches to addressing endogeneity include:

1. Assume endogeneity is minor: several experts interviewed indicated that, while endogeneity is a concern, its impact on savings estimates are likely minor and their research does not address endogeneity concerns. One approach to handling endogeneity is to simply note it exists, but not attempt to fix the problem.
2. If we do analysis with geographic units 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. Construct an approximate marginal price: for this approach, we would model the ratio of revenue to consumption ratio versus consumption changes. However, it would be necessary to construct this price for a different year than the observation year, or else it will create a different source of endogeneity.

4.5.2 Program spending

This specific problem was not addressed in the literature we reviewed and had not been considered by the interviewees. Some suggestions that emerged from our conversations are below.

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 in any case.
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 type (such as lighting or non-lighting). Each of these will be much less correlated with the total consumption than would total spending.
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.

4.6 ERROR BOUNDS

The team also considered the following question regarding error bounds: “*What are the considerations for determining error bounds from the models’ results?*” We found there are two issues when constructing error bounds for estimated savings from a top-down model:

1. Whether the regression standard error is correctly calculated. Several interviewees emphasized the importance of using estimation methods that calculate standard errors robustly. One reason this is needed is to address the potential for unequal levels of unexplained variation across the different observations (model heteroscedasticity).
2. Panel models tend to have substantial serial correlation across observations. If this is not addressed, the standard errors may be substantially understated. This will mean that the true uncertainty of the results will be much worse than what the calculated standard errors and t-statistics indicate.

4.7 CONSTRUCTING PORTFOLIO SAVINGS

The team also further investigated how to construct portfolio savings, specifically investigating “*What are the most appropriate methods for constructing portfolio-level savings estimates?*” The problem of how to use the results of a top-down analysis to construct savings estimates is not addressed by literature outside of this specific application. For those studies that have used top-down methods for net savings estimation, results were reported as total savings over the period of the study. This is appropriate because the top-down coefficients represent factors that apply on average over the observations included in the analysis. Details on how the savings estimates were constructed were not given.

To compare top-down and bottom-up analysis, it is essential to set up apples-to-apples alignments at two stages. First, to estimate savings from the top-down model, the coefficient must be applied in a way that is consistent with how it appeared in the estimated model. Second, top-down savings must be compared with bottom-up estimates that cover the same period and the same cumulative effects.

The specifics of how this is done depends on the model form, including if any terms appear in log form or with other nonlinearity, how lag program savings is incorporated, and the time frame of the analysis. Our primary recommendation is to be transparent in how the top-down and bottom-up savings are calculated, what period and what cumulative effects they represent, and why these are comparable.

4.8 SELF-SELECTION

One of the limitations identified in our original methods review that top-down models using geographic units finer than service territory, such as town, are subject to self-selection bias. That is, program activity levels in each geographic area are determined by choices of the customers in that area, and customers who are more inclined to participate may have different consumption levels apart from the program compared to customers who participate less. As such, we also researched whether *there are ways to mitigate self-selection bias in the modeling*.

Some of our interviewees identified that this is a challenge, and that it becomes worse as the geographic unit becomes finer. This was also an observation from the 2015 study.

In a current study by members of the evaluation team, we are exploring self-selection correction methods for net savings analysis based on regressions with individual customers’ change in consumption as the

dependent variable. This study led us to identify an improved approach that we will consider adapting to the panel data regressions for top-down modeling.

4.9 SPILLOVER

Another problem that arises with geographic units finer than service territory is a spillover effect across geographic units and we thus considered whether *there are ways to mitigate bias due to cross-geography spillover in the modeling*. This issue was not explored in our literature review or interviews; instead we proposed to 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. As with the lag terms, we may find that there is too much collinearity to identify these effects separately.



5. CONCLUSIONS AND GUIDELINES

The literature review and interviews all support the concept that panel regressions are a fairly blunt tool that can be used to glean broad qualitative results, but may not provide definitive quantitative estimates. The literature and interviews suggest use of experimental design or natural experiments as being highly preferable as key concerns include the difficulty of determining the right model specification, and the limited explanatory variables available. Some of the experts interviewed, and the literature, emphasized the importance of incorporating available information on the structural drivers of the dependent variable and of the variables being controlled for—in our case, often physical relationships as well as savings estimates. Several suggested trying the analysis multiple ways and then explaining why one approach should be preferred.

The specific methods that can be used for top-down analysis are nearly as broad as the field of econometrics itself. There is no instruction manual for procedures that can be followed. Most approaches that have been considered have some merit and may be the best for some circumstances. Nevertheless, there are some general guidelines that emerge for our top-down research. These are discussed below, organized by research question.

5.1.1 Weather modeling

1. Two approaches to weather modeling are:
 - 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.
 - 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.

2. Use robust standard error estimation to account for varying variability, whether or not weather normalization is done as a first step.
3. If practical, run the analysis with and without first weather normalizing. Compare the results and understand why the differences are there if they are. The likely cause of differences in a short time series is that annual weather would be confounded with some other effect not otherwise captured in the model.

5.1.2 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 is a suggestion that is useful to address several issues.

5.1.3 Recession

1. At a minimum, include a dummy for the recession period and test whether it is significant.

2. Test including years before and after the formal definition of the recession period.
3. Explore including variables that account explicitly for what is 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.

5.1.4 Fixed effects

1. Include fixed effect terms, particularly for geographic units.
2. Explore specifications that omit time fixed effects, or that 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.

5.1.5 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 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. Possibly construct 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.

5.1.6 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 will 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.

5.1.7 Error bounds

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.

5.1.8 Constructing portfolio savings

1. Ensure that savings from the top-down model are estimated consistently with how each savings coefficient appeared 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.

5.1.9 Self-selection

Explore adapting the recent advances in self-selection correction for individual customer regressions to top-down methods.

5.1.10 Spillover

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.



APPENDIX A: LIST OF RELEVANT LITERATURE

Title and Abstract
Lags
The effects of energy policies on energy consumption in China MJ Lu, CYCL Lawell, S Chen – 2015 – des.ucdavis.edu
Ozturk, Ilhan. 2015. “Measuring the impact of energy consumption and air quality indicators on climate change: evidence from the panel of UNFCCC classified countries”. <i>Environmental Science and Pollution Research</i> . 22 (20): 15459–15468.
Scholtens, Bert, and Veldhuis, Rineke. 2015. How does the development of the financial industry advance renewable energy? A panel regression study of 198 countries over three decades. <i>Beiträge Zur Jahrestagung Des Vereins Für Socialpolitik 2015</i> :
Achen, C. Why lagged dependent variables can reduce the explanatory power of independent variables, November 2, 2001
Gilbert E. Metcalf see these studies first Metcalf, Gilbert E. 2006. <i>Energy conservation in the United States: understanding its role in climate policy.</i> Cambridge, Mass: National Bureau of Economic Research.
Metcalf, Gilbert E. 2008. <i>An empirical analysis of energy intensity and its determinants at the state level.</i> Cambridge, MA: MIT Joint Program on the Science and Policy of Global Change.
Hassett, Kevin A., and Gilbert E. Metcalf. 1995. “Energy tax credits and residential conservation investment: evidence from panel data”. <i>Journal of Public Economics</i> . 57 (02): 201–217.
Kaufman. Robert R., and Alex Segura-Ubierno. 2001. “Globalization, domestic politics, and social spending in Latin America: a time-series cross-section analysis, 1973–97”. <i>World Politics</i> . 53 (4).
Anderson, T. W., and Cheng Hsiao. 1981. <i>Formulation and estimation of dynamic models using panel data.</i> Toronto: Institute for Policy Analysis, University of Toronto.
Bhargava, Alok. 2006. <i>Econometrics, statistics, and computational approaches in food and health sciences.</i> New Jersey: World Scientific.
Achen, Christopher H. “Why lagged dependent variables can suppress the explanatory power of other independent variables.” <i>Ann Arbor 1001</i> , no. 2000 (2000): 48106–1248.
Keele, Luke and Kelley, Nathan J. <i>Dynamic Models for Dynamic Theories: The Ins and Outs of Lagged Dependent Variables</i> , <i>Political Analysis</i> (Spring 2006) 14 (2): 186–205. Doi: 10.1093/pan/mpj00
Recession
Lytras, Demetra and Bell, William R. n.d. <i>Modeling Recession Effects and the Consequences on Seasonal Adjustment.</i> U.S. Census Bureau, 2013.
Aguiar, Mark, Hurst, Eric, Karabarbounis, Loukas, “Time Use During the Great Recession; Identifying the relationship of household spending and labor supply,” <i>American Economic Review</i> 103 (5): 1664–1696, 2013”
Cooper, Daniel. 2012. “U.S. Household Deleveraging: What Do the Aggregate and Household-level Data Tell Us?” <i>Public Policy Briefs</i> , Federal Reserve Bank of Boston. 12–2.
Campbell, John Y., Andrew W. Lo, and Archie Craig MacKinlay. 1997. <i>The econometrics of financial markets.</i> Princeton, N.J.: Princeton University Press.

Title and Abstract
Poudyal, Neelam C., Bamadev Paudel, and Michael A. Tarrant. 2013. "A time series analysis of the impact of recession on national park visitation in the United States," <i>Tourism Management</i> . 35 (4): 181–189.
Nelson, Hal T., David von Hippel, Tom Peterson, and Roman Gaagulagian. 2015. "The Great Recession or progressive energy policies? Explaining the decline in US greenhouse gas emissions forecasts". <i>Journal of Environmental Planning and Management</i> . 59 (3): 480–500.
Declercq, Bruno, Erik Delarue, and William D'haeseleer. 2011. "Impact of the economic recession on the European power sectors CO2 emissions". <i>Energy Policy</i> . 39 (3): 1677–1686
Dejuan, Oscar, Eladio Febrero, and Maria Cristina Marcuzzo. 2011. <i>The First Great Recession of the 21st Century: Competing Explanations</i> . Cheltenham: Edward Elgar Pub.
Ireland, Peter N. 2010. <i>A New Keynesian perspective on the Great Recession</i> . Cambridge, Mass: National Bureau of Economic Research.
Kocherlakota, Narayana Rao. 2009. <i>Modern macroeconomic models as tools for economic policy</i> (Minneapolis, MN: Federal Reserve Bank of Minneapolis, 2009).
Ng, Serena, and Jonathan H. Wright. 2013. <i>Facts and challenges from the Great Recession for forecasting and macroeconomic modeling</i> (Cambridge, Mass: National Bureau of Economic Research).
Weather
Elkhafif, Mahmoud A.T. 1996. "An iterative approach for weather-correcting energy consumption data". <i>Energy Economics</i> . 18 (3): 221–230.
Won, S. J., X. H. Wang, and H. E. Warren. 2016. "Climate normals and weather normalization for utility regulation". <i>Energy Economics</i> . 54: 405–416.
Hirshleifer, D. and Shumway, T. (2003), <i>Good Day Sunshine: Stock Returns and the Weather</i> . <i>The Journal of Finance</i> , 58: 1009–1032. doi:10.1111/1540-6261.00556
Sailor, David J., and Chittaranjan Vasireddy. 2006. "Correcting aggregate energy consumption data to account for variability in local weather". <i>Environmental Modelling and Software</i> . 21 (5): 733–738.
Aroonruengsawat A., Auffhammer M., and Sanstad A.H. 2012. "The impact of state level building codes on residential electricity consumption". <i>Energy Journal</i> . 33 (1): 31–52.
Auffhammer, Maximilian, Kellogg, R, <i>Clearing the Air? The Effects of Gasoline Content Regulation on Air Quality</i> . 2011. <i>American Economic Review</i> 2687–2722.
Auffhammer, Maximilian. 2013. <i>Using weather data and climate model output in economic analyses of climate change</i> . Cambridge, Mass: National Bureau of Economic Research. http://papers.nber.org/papers/w19087 .
"Levinson, A. "How Much Energy Do Building Energy Codes Save? Evidence from California Houses," <i>American Economic Review</i> 2016, 106(10): 2867–2894
Filippini, Massimo, and Lester C. Hunt. 2015. "Measurement of energy efficiency based on economic foundations". <i>Energy Economics: Supplement 1</i> . 52: S5–S16.
Alberini, Anna, and Charles Towe. 2015. <i>Information v. energy efficiency incentives evidence from residential electricity consumption in Maryland, Zürich: CER–ETH</i>
Filippini, Massimo, and Lester C. Hunt. 2016. "Measuring persistent and transient energy efficiency in the US". <i>Energy Efficiency</i> . 9 (3): 663–675.

Title and Abstract
From Filippini, Massimo, and Lester C. Hunt. 2016. "Measuring persistent and transient energy efficiency in the US". <i>Energy Efficiency</i> . 9 (3): 663–675.
Filippini, Massimo, Lester C. Hunt, and Jelena Zorić. 2014. "Impact of energy policy instruments on the estimated level of underlying energy efficiency in the EU residential sector". <i>Energy Policy</i> . 69 (2014): 73–81.
Hunt, Lester C., and Yasushi Ninomiya. 2003. "Unravelling Trends and Seasonality: A Structural Time Series Analysis of Transport Oil Demand in the UK and Japan". <i>The Energy Journal</i> . 24 (3).
Filippini, Massimo, and Lester C. Hunt. 2012. "US residential energy demand and energy efficiency: A stochastic demand frontier approach". <i>Energy Economics</i> . 34 (5): 1484–1491.
Horowitz, Marvin J., and Paolo Bertoldi. 2015. "A harmonized calculation model for transforming EU bottom-up energy efficiency indicators into empirical estimates of policy impacts". <i>Energy Economics</i> . 51 (2): 135–148.
Horowitz, Marvin J. 2014. "Purchased energy and policy impacts in the US manufacturing sector". <i>Energy Efficiency</i> . 7 (1): 65–77.
David S. Loughran and Jonathan Kulick. Demand-Side Management and Energy Efficiency in the United States, <i>he Energy Journal</i> , Vol. 25, No. 1 (2004), pp. 19–43
Deschênes, Olivier, and Michael Greenstone. 2007. <i>Climate change, mortality, and adaptation: evidence from annual fluctuations in weather in the U.S.</i> Cambridge, MA: National Bureau of Economic Research.
Slini, T., E. Giama, and A.M. Papadopoulos. 2014. "The impact of economic recession on domestic energy consumption". <i>International Journal of Sustainable Energy</i> . 34 (3–4): 259–270.
Matt Kotchen, Grant Jacobsen. "Do Building Codes Save Energy? Evidence from Residential Billing Data in Florida," <i>The Review of Economics and Statistics</i> , 95 (2013) 34–49. Kotchen, Matthew J. "Longer-run evidence on whether building energy codes reduce residential energy consumption." (2016). <i>Journal of the Association of Environmental and Resource Economists</i> . This updates 2013 study and addresses weaknesses.
Braga AL, A Zanobetti, and J Schwartz. 2002. "The effect of weather on respiratory and cardiovascular deaths in 12 U.S. cities". <i>Environmental Health Perspectives</i> . 110 (9): 859–63.
Fels, Margaret F., and Miriam L. Goldberg. "Using the scorekeeping approach to monitor aggregate energy conservation," <i>Energy and Buildings</i> . 9 (1986): 161–168.



APPENDIX B: NAME, AFFILIATION, AND RESEARCH AREA OF INTERVIEWEES

Name	Affiliation	Research Areas
Max Aufhammer	Professor of International Sustainable Development and Associate Dean, Division of Social Sciences at U.C. Berkeley (CA).	Environmental and resource economics, energy economics and applied econometrics. Lead author for the Intergovernmental Panel on Climate Change (IPCC)
Arik Levenson	Professor, Department of Economics, Georgetown University (Washington, DC.)	Environmental Economics, Public Finance
Marvin Horowitz	President, Demand Research, LLC	Econometrics and impact evaluation of energy efficiency programs
Anna Alberini	Professor, Department of Agricultural and Resource Economics, University of Maryland, (College Park). Research Affiliate, Centre for Energy Policy and Economics, ETH (Zürich)	Natural resource and environmental economics, energy economics, applied econometrics and statistics
Dylan Small	Professor, Department of Statistics, The Wharton School, University of Pennsylvania.	Causal inference, design and analysis of observational studies and experiments, applications of statistics to public health, medicine and public policy
Greg Nemet	Associate Professor at the University of Wisconsin–Madison, La Follette School of Public Affairs and the Nelson Institute's Center for Sustainability and the Global Environment.	Global energy systems analysis, governance of global energy problems, and international environmental policy
Michael Cohen	Senior Program Officer, National Academies of Science, Engineering and Medicine, Committee on National Statistics (Washington DC.).	Research methodologies, statistical approaches
Dan Violette	Managing Director, Energy Practice, Navigant Consulting	Energy, strategy and technology



APPENDIX C: TOP-DOWN MODELING INTERVIEW GUIDE

Interview Guide

This interview guide is intended to provide a conversational outline and talking points for interviewers to obtain information from respondents that will help inform the methodological review for top-down modeling. These interviews are intended to obtain expert opinions about existing studies that will inform the development of top-down research in Massachusetts. The interview is designed to probe respondents for examples of econometric modeling both in and outside the energy industry. We are interested in how researchers handled the methodological and data concerns listed below when constructing macro-economic, time-series cross-sectional models based on aggregate geographic data (i.e. town, county, or state level data).

The scope of work specified the following lines of questioning for the interviews:

Methodological considerations:

- How should weather variation (weather-normalized annual data vs. weather terms in the econometric model) be addressed in top-down models?
- How should models account for cumulative program effects of programmatic activity over time, such as, whether to use cumulative lagged terms, individual lagged terms, or indexed lagged terms?
- What are the most appropriate methods for constructing portfolio-level savings estimates and error bounds from the models' results?
- How should models determine the impacts on energy usage for the recession years (2008-2010)?

Data-related considerations:

- We have further discussed with the PAs whether consumption data are available at more detailed levels of geographic resolution compared to the previous effort, which used utility service territory as the geography to estimate program effects. Eversource and Cape Light Compact provided data at the town-level, so the team will investigate what impact this data will have on the PA and Muni-models.
- What is the appropriate number of years before the team should attempt re-estimation of the PA data models? Does utilizing the town-level data impact the number of years needed in the model?
- Which modeling approaches are most appropriate with more geographic-based variation? How robust are estimated effects to alternative model specifications? What are the apparent advantages and disadvantages of alternative specification choices in terms of model precision, model stability, bias potential, and interpretability?

This document is *not a script*. Interviewers will be econometricians familiar with the topics of discussion and the likely eventual analyses they will perform. In addition, the nature of these conversations is open-ended. Thus, interviewers are expected to probe and come up with additional questions during the interview, rather than depend on this guide to enumerate every question they need to ask.

Pre-interview Prep

Purpose: Do some background reading before you get on the phone so you can be more efficient with the time you have with the respondent.

1. Who is the individual you will be interviewing?
 - a. What is their position?
 - b. What is their relationship to top-down modeling?
 - c. Are there any studies we know of already that they authored?

Study name	Date	Brief description

Introduction

Purpose: Break the ice, introduce yourself, introduce what you want to talk about.

1. **Introduce yourself.** Tell them your name and that you're calling from DNV GL on behalf of the Massachusetts electricity and gas providers.
2. **Introduce the project.** We were hired by the Massachusetts Energy Efficiency Program Administrators to explore the use of top-down modeling in evaluating the energy savings of a portfolio of energy efficiency programs in Massachusetts. By Top-down modeling, I am referring to:

Macro-economic/econometric techniques that use aggregate consumption data across geographic units and time to estimate program impacts across all energy-efficiency programs in a given region or service territory. This contrasts with bottom-up techniques that run separate studies for each program or measure/end-use (i.e. widget) within a program. Top-down models attempt to measure changes in energy consumption over time that are attributable to programmatic interventions. The goal of this type of modeling is to isolate the effect of program portfolio activity from other natural changes and policy variables.

We are interested in exploring the use of Top-down modeling of energy efficiency savings because we think it may be a more cost-effective, and accurate way to measure savings since there are so

many individual energy efficiency programs, and their effects can be interactive and often have spillover and market effects that reach many more customers than those reached directly through the program. For this interview, we are looking at the advantages and disadvantages econometric methods have had in evaluating energy policy and in other fields. We are considering challenges associated with using econometric modeling of aggregate data to analyze the impact of public policies. We are considering two types of issues:

- a. **Methodological considerations:** How to account for different exogenous and endogenous factors when structuring models and the aggregate variables including: Weather variation, cumulative programmatic impacts, accommodating economic anomalies I (i.e. recessions), and how to construct portfolio-level estimates and error bounds from the models' results.
- b. **Data-related considerations:** Determining the conditions where the data are likely to be adequate to support the savings estimation (e.g. the length of time series needed (with or without lags), the size of effects that can be detected; key steps required to clean and align data series; and identifying the strengths and weaknesses of constructing models at different levels of geographic aggregation).

We plan to take insights from this interview and others to inform how we develop our models and gather the necessary data.

Role, Responsibility, and Experience with Top-down modeling

Can we start with a little information about you?

- 1.1. What are your primary areas of research?
- 1.2. Have you used employed top-down (macroeconomic) models in this research?
 - a. For which types of policies, what industries, and at what level of society did your evaluations apply?
 - b. What was the focus of the biggest/most interesting 2-3 macroeconomic studies you were involved with?
 - i. Which Sectors/Industries were the focus of these studies (e.g. energy policy, education policy, health policy, fiscal policy, other)?
 - ii. Describe the specific policies that were being evaluated.
 - iii. What was your typical role in the studies?
 - iv. How were the results used? By whom?

Now I'd to talk about these studies in a bit more detail. For each of the studies you just identified please answer the following:

- 1.3. At what level of geographic aggregation was the study conducted? (block-group, town, county, utility, state, other?)

- a. What were the strengths and weaknesses of constructing models at this level of geographic aggregation?
 - b. Which modeling approach was most appropriate for this level of geographic resolution? How would that change with a finer level of geographic resolution?
 - c. How did the level of geographic resolution impact estimates of the program effects?
 - i. What did you do to ensure that the model specification is meaningful, so that the key results are neither fortuitous nor biased by researcher expectations?
 - d. Do you have references that discuss the appropriate level of data aggregation?
- 1.4. What are key explanatory variables to include and how are they typically constructed given standard economic data available? (e.g. is it possible to construct price variables that are not simply the ratio of consumption to expenditures, which makes the price variable correlated with the consumption residual?)
- 1.5. How long a time series did you use and what was/is the time interval?
- 1.6. What is your general assessment of the overall strengths and limitations of this approach in general in terms of policy evaluation?
- a. Under what conditions does it work well?
 - b. What are the key constraints to the approach?
 - c. When you specify a model for this type of analysis, are there certain elements that are always important to include or to avoid?
 - What are the implications for precision, stability, bias, interpretability if you do/don't do those things?

Methodological Considerations

Next, I'd like to ask you about your experience with some specific methodological considerations.

- 1.7. In your experience, what are general conditions needed to have a successful model? [If necessary, probe for the following] Do you have any guidance regarding:
- a. Length of time series
 - b. Diversity of program activity level across units of observation (time-geography combinations)
 - c. Ability to establish the counterfactual (no-program) scenario
 - d. The relationship between program activity and program impact between observational units
 - e. Effect of one area on another (cross-area spillover)
 - f. The ability to account for the lag structure of program impacts

Next, I'd like to discuss model specification testing/validation.

- 1.8. In bottom-up energy efficiency evaluation work, we typically are dealing with the evaluation of individual programs. We then construct portfolio level savings by adding up the pieces from the

individual programs in each year. However, with the top-down approach, savings are estimated in aggregate across all years over which the model was run.

One way or another you have some cumulative program effect for each of the years used to fit the model, and we need to construct an estimate that's comparable to the first-year or life-time savings estimates we have from bottom up. Translating the model coefficients into savings comparable to bottom-up values is not trivial.

- a. Do you have experience constructing portfolio level impacts/savings across a range of programs over time using a top-down approach?
- b. Can you discuss how you went about constructing the estimate and what it represented?
- c. Can you discuss how you went about constructing error bounds from the models' results?
- d. What were the key challenges doing this? How did you handle them?
- e. Do you have references to these studies and/or resources you used to inform your decisions?

1.9. Not addressing endogeneity is one of the ways you get spurious correlations all over the place from time series regressions. Can you discuss how you've handled assumptions concerning endogeneity in your models?

- a. Is it best to always assume that everything is potentially endogenous with the dependent variable (e.g. consumption), and use instrumental variables or other econometric technique to address that, or
- b. Are the situations where that can be assumed to be minor?
- c. Do you have references to studies where the treatment of endogeneity was discussed that you can provide?

1.10. An issue that arises in models of energy policy is how to distinguish changes in consumption resulting from policy, economic, and other variables from those that result from changes in weather conditions over time, measured by degree-days. To remove the impacts of weather from the remaining measured impacts, including the desired policy impacts, we can use one of two general approaches:

- Use nominal annual consumption as our dependent variable and account for weather by including annual weather term(s) in a TSXS model,
- Use monthly consumption data to weather normalize each year and then use normalized annual consumption as the dependent variable.

Do you have experience addressing this type of weather normalization in your past work with top-down models?

- a. What are the pros and cons of each approach as you see them?
- b. Which approach did you use? Why did you choose that technique?
- c. Do you have any literature that discusses the pros and cons of different techniques for dealing with the impact of weather in top-down models?

1.11. Another modeling complication we need to address is that program activities in one year affect program impacts in future years (i.e., program impacts persist from year to year.). There are a few ways to address this:

- Use explicit lag variables (i.e., 1-year, 2-year, 3-year lag in programmatic expenditures) – This approach tells us about persistence rates, though usually the set of lag coefficients don't look like a physical decay pattern.
- Use cumulative lag variables (i.e., sum of expenditures over the past 1-year, 2-years, 3-years). In this case, a four-year lag would only enter as a single variable for the sum of expenditures over the past 4 years, rather than separate variables for each past year.
- Use cumulative surviving savings (based on an ex ante or bottom-up ex post estimate). In this case, we're building our assumptions about persistence rates into the program activity variable, but we have fewer coefficients confusing us and each other.

How should models account for cumulative program effects of programmatic activity over time, such as, whether to use cumulative lagged terms, individual lagged terms, or some other approach? Do you have experience addressing these types of cumulative impacts in your past work with top-down models?

- a. What are the pros and cons of each approach as you see them?
- b. Which approach did you use? Why did you choose that technique?
- c. Do you have any literature that discusses the pros and cons of different techniques for dealing with lagged program impacts in top-down models?

1.12. Next, I'd like to ask you about your experience with using fixed effects terms. If we include fixed effects for year and for geography, we eliminate a lot of serial correlation, though not all. We also eliminate any opportunity to identify a general trend if we want to.

- a. Do you have experience addressing fixed effects with top-down models?
- b. What are the pros and cons of each approach as you see them?
- c. Which approach did you use? Why did you choose that technique?

1.13. One concern we encountered was how to address the impacts on energy usage for the recession years (2008-2010). Do you have experience with accounting for recession periods in time-series cross sectional models?

- a. What did you do?
- b. Do you have studies that you can reference for how to handle recession periods?

Summary and Additional Contacts

Purpose: Gather contact information for anyone else they could suggest that it would be helpful for us to speak to.

1.14. Do you have other information on your use of econometric models you'd like to share that we have not discussed today?

- 1.15. Do you have any suggestions for other firms we could contact that might be able to provide the same kind of information?