Macro Consumption Metrics White Paper

by

Marvin J. Horowitz, Ph.D.
Demand Research, LLC

August 15, 2011

Presented to the
California Public Utilities Commission

Table of Contents

Executive Summary

Demand Research's Macro Consumption Metrics (MCM) white paper contains detailed literature reviews and discussions of two top-down methodologies for analyzing aggregate energy consumption. These are econometric energy demand modeling and energy intensity index decomposition. Set in the context of the CPUC's energy efficiency program EM&V goals, *Demand Research* recommends that econometric energy demand modeling be the focus of the MCM pilot project. Towards this end *Demand Research* proposes to build a unique, geographically-granular database and to complete up to three studies to explore, test, and assess the viability of making MCM policy evaluations a component of the CPUC's future EM&V activities.

D.10-10-033 (CPUC, 2010-A) orders that MCM studies be looked into for investigating the total, long-term energy savings resulting from California's energy efficiency programs. It references Horowitz (2011) as the kind of MCM study that it is interested in pursuing. In it, residential electricity sales are analyzed for four major utilities in the state of California. The findings indicate that for 2006 and 2007 the long run energy savings from energy efficiency policies averaged about 4.8 percent per year, and the relative precision of the counterfactual energy consumption estimate for each utility roughly averaged plus or minus 12 percent at the 95 percent confidence level. These findings were achieved using an econometric model containing panel data, that is, sixteen continuous years of data for each of the four utilities.

To further demonstrate the applicability of MCM policy evaluations to CPUC's EM&V efforts, *Demand Research* provides three examples of MCM studies, complete with descriptions of how policy counterfactuals are constructed, that are feasible to implement in the near future in the manufacturing, residential, and commercial sectors. These examples supplement the proofs of concept published in Horowitz (2004), Horowitz (2007), and Horowitz (2011). *Demand Research's* value proposition to the CPUC is:

- an updatable, permanent, state-of-the-art, geographic database that permits the CPUC to monitor the long-term impacts of energy efficiency policies on state and local energy consumption and environmental protection; and,
- an MCM policy evaluation system that can lead to substantial reductions in the size, and cost, of the CPUC's bottom-up EM&V activities by helping to refocus bottom-up evaluations on program

management and contractor performance auditing rather than first-year measure and program energy savings.

Demand Research's MCM pilot project proposal, which constitutes the last half of this white paper, offers the following features:

- use of public access data that come from state, local, and federal agencies and are free to the public, are geographically-based, and can be divided as finely as the the census tract or zip code levels;
- compilation, merging, and joining of all the selected datasets in a single database platform that permits standardized data manipulation and variable creation; and,
- reduced form econometric models of energy demand using either time series or panel data; each demand model is for a single fuel, and a single customer type, market segment, or economic sector.

Although the scope of *Demand Research's* MCM pilot project proposal is ambitious, *Demand Research* believes that only a project of this breadth and depth is capable of demonstrating the practical value of MCM policy evaluations to the CPUC. Moreover, there are critical elements of this pilot project that make it realistic to think that it can be successfully completed, on time and on budget. For one, once the residential sector MCM database is created, the marginal effort in time and cost to expand the MCM database to include data for the commercial, industrial, and agricultural sectors is relatively modest. This is because similar datasets are involved in all sectors, and similar data processing routines are needed for many of the variables across sectors. Another element that is in this project's favor is that once the MCM database is in place there are declining marginal cost and time requirements for performing more than one policy evaluation. It is due to this phased cost/effort structure that *Demand Research* believes that this pilot project's scope is both reasonable and achievable.

Finally, it is worth noting that it is because of these declining marginal costs that *Demand Research* refers to the final product of this pilot project as an *MCM policy evaluation system.* If the CPUC assesses, based on this pilot project, that MCM studies should be a component of its next EM&V cycle, a database system and modeling protocol will be in place for conducting multiple, inexpensive policy evaluations simultaneously. These policy evaluations can address a variety of subjects and can be focused on small market segments and small areas as well as on broad economic sectors and whole utility service territories.

1. Introduction

Demand Research's Macro Consumption Metric (MCM) white paper is organized such that the first part (sections 2 through 6) contains discussions of the CPUC's EM&V goals (D.10-10-033, CPUC 2010-A) and of two top-down methodologies for analyzing aggregate energy consumption. The second part (sections 7 through 9) contains *Demand Research's* proposal for the MCM pilot project, including tasks, a timeline, and project costs.

In this white paper, the terms *macro* or *aggregate* consumption are interchangeable with *market demand*. Market demand, as defined in economics, is the sum of a group of individual consumer's demands for a specific product at a specific time. It can be an entire country's single day demand for a general commodity, such as the U.S. demand for primary energy, or it can be a smaller group of consumers' annual demand for a specific product, such as Bay Area residents' demand for electricity. As such, *energy demand*, as in energy market demand, is used liberally to refer to the demand for either electricity or natural gas of specific consumer segments or rate classes in specific geographic locations. Depending on the policy topic, the geographic boundaries of a market may be as small as census tracts or as large as the state as a whole, and the designated time period could be months or years.

Another term, *policy*, as in *energy efficiency policy* and *policy evaluation*, is also applied liberally. It refers to one or more energy efficiency programs operating simultaneously in a geographic area that affect a specific group of consumers. An energy efficiency *program* is singular and refers to anything from a mandatory statewide energy efficiency building standard, to a financial incentive offering, to an educational program. Given the nature of top-down studies, it would be a misnomer to refer to MCM studies as *program* evaluations since most consumers in California have been subject to, or eligible for, more than one energy efficiency program in the recent past. The term *policy* also differs from the term *portfolio,* which is usually taken to mean a single organization's entire collection of energy efficiency programs.

As part of the discussion of the CPUC's EM&V goals, simple comparisons are offered of the capabilities of bottom-up and top-down evaluation methods to address various factors that relate to energy efficiency policy impacts. The purpose of these comparisons is to emphasize that there is little overlap in the kinds of policy impact information that different methods produce. It foreshadows the conclusion in this white paper that only by developing policy-focused econometric energy demand models will the CPUC be able to measure current and future high-level policy achievements.

The technical portion of this paper describes applications of econometric energy demand modeling and energy indicators to energy efficiency policy evaluation. Both methodologies have extensive histories. However, econometric modeling is a vast discipline that, unlike energy indicators, cannot be described by a handful of equations. To provide an in-depth treatment of econometric modeling and demonstrate its application to the CPUC's EM&V plans, three examples are provided of MCM evaluations that might be implemented in the manufacturing, residential buildings, and commercial buildings sectors. These examples supplement the proofs of concept published in Horowitz (2004), Horowitz (2007), and Horowitz (2011) in which energy efficiency policies were found to produce substantial long-term net energy savings. Complete with descriptions of how policy counterfactuals are constructed, the three policy evaluation examples offered in this white paper show the applicability of aggregate energy demand models to the CPUC's EM&V goals..

Energy indicators are another form of top-down analysis that have gained a large following throughout many U.S. organizations, the European Union, and the international community. After reviewing the literature on this method, a theoretical treatment and empirical example of index decomposition is provided based on the well-known Fisher Ideal index. The treatment shows that the so-called *efficiency index*, which is the most statistically-rigorous of all energy indicators, cannot separate policy effects from market effects. The final section of the first part this white paper recommends that the CPUC avoid using energy indicators as a policy evaluation method and instead focus on geographically-based aggregate energy demand modeling.

The second part of this white paper is a detailed proposal for implementing *Demand Research's* plan for an MCM pilot project in the twelve months beginning in September, 2011. The two primary achievements of *Demand Research's* MCM pilot project will be the creation of a one-of-a-kind policy evaluation database and the completion of one or more policy evaluations.

2. CPUC's EM&V Goals and MCM Evaluations

Prior to examining the top-down methodologies it is instructive to review the CPUC's recent energy efficiency program EM&V goals and the contexts in which they were formed. For evaluating the 2006-2008 cycle of energy efficiency programs the CPUC's Energy Division was directed to pursue the following goals (CPUC, 2010-B):

- verify the costs and installations of energy efficiency program activities;
- update the ex-ante parameters used to estimate program savings and benefits; and,
- publish reports that calculate earnings the utilities are eligible to claim.

These goals were intended to fit the needs of the CPUC's risk reward incentive mechanism, a framework for compensating investor owned utilities (IOUs) for implementing energy efficiency programs. As part of its effort to accomplish these EM&V goals, bottom-up energy efficiency program evaluations were conducted between 2007 and 2010. In all, the CPUC undertook eleven large-scale evaluations of resource acquisition programs at a total cost of approximately $69 million, with individual studies ranging from a low of $1.7 million for the evaluation of combined agricultural and food processing programs to a high of $18.7 million for the evaluation of combined residential retrofit and lighting programs. Much of the evaluation research involved energy efficiency measure auditing and monitoring.

The generous amounts of money allotted for evaluating the 2006-2008 program cycle, and now the 2010-2012 program cycle, are the direct result of following the CPUC's goals, because all large-scale *in situ* measure-based studies require extensive, and expensive, primary data collection. However, two developments in 2008 reshaped the CPUC's EM&V goals. The first was Assembly Bill (AB) 32 which chose energy efficiency as California's primary resource for achieving 15 percent of its mandated GHG reduction target, and the second was the CPUC's adoption of a long term energy efficiency strategic plan that directs energy efficiency programs away from promoting short-term energy savings and towards more comprehensive, long-term savings (CPUC, 2010-A). These events produced a domino effect. With new policy goals came a need for new programs, and with new programs and new goals came a need for new approaches to program evaluation. In addition to supporting the CPUC's risk reward incentive mechanism, the CPUC recognized that new genres of evaluation studies were needed to quantify the permanent, long-term effects of overlapping public programs.

The CPUC's decision to expand the scope of EM&V activities to include an MCM pilot project is one avenue for addressing these new policy goals. According to D.10-10-033:

> Macro Consumption Metrics may allow the Commission to accurately measure the impact of the Commission's energy efficiency efforts on overall energy consumption and provide a more direct account of aggregate reductions in GHG emissions. As an added benefit, because Macro Consumption Metrics rely on existing energy usage data and relatively simple statistical analysis, they can be developed quickly and at a reasonable cost to ratepayers. These metrics offer substantial benefits to the Commission's EM&V portfolio and can be delivered quickly at a reasonable cost. We therefore support the examination of Macro Consumption Metrics to assess the aggregate impact of the 2013-2015 energy efficiency programs on energy consumption. (p. 32)

The CPUC decision calls for the MCM pilot project to explore, test, and assess the viability of MCM evaluations. This opens the door to shifting EM&V away from costly verification of the factors associated with first-year measure savings and towards less expensive verification of aggregate long-term energy savings. The CPUC decision does not suggest that top-down studies should replace bottom-up studies, or reconcile them, or produce similar measure and program-level information that bottom-up studies produce.

To elaborate on how MCM policy evaluations are different from bottom-up EM&V methodologies, Table 1 lists fifteen factors that affect energy savings estimates or that are valuable for impact evaluations to take into consideration. These factors are not exhaustive or mutually exclusive. For the purposes of these comparisons no differential importance is given to any of these issues or to the order in which they are listed. In reading over these factors it is useful to begin each with the question, "*does the method*…," and to end each with "…*when estimating long-term energy savings*."

To the right of the checklist are the scores for all sixteen factors for four different EM&V methods. Two are bottom-up methods, i.e., measure monitoring and billing data analysis, and two are top-down methods, i.e., energy indicators and energy demand models. The scores are in the form of the answers *yes*, *no*, or *maybe*. The first bottom-up method, *monitoring*, is meant in the most general sense to represent all impact evaluations that focus on engineering algorithms and engineering models. These include run-time monitoring or data logging and spot electricity metering, as well as verification of measure retention and operationality via visual inspection or telephone and mail surveys. Building simulation modeling using participant data and all other modes of data collection and analysis designed to produce *in situ* measure installation estimates are also included in the category of *monitoring*.

On the other hand, *billing data* is meant in the most general sense as any kind of impact evaluation that uses short-term customer billing data (two years or less in the post-treatment period) to measure changes in billed energy consumption. Included in the broad definition of these two methods is the assumption that some type of customer survey is conducted to assess free ridership rates in whatever chosen way it is defined. Estimates from surveys of spillover are less well-accepted than those of free ridership, thus the entry of *maybe* for this factor in the two bottom-up methods.

Table 1: Comparisons of Impact Evaluation Methods

	Policy Impact Methodology Check-List	***Bottom- Up Methods***		***Top-Down Methods***	
Factor	***"Does the method…when estimating long-term energy savings."***	**Monitoring**	**Billing**	**Indicators**	**Models**
1.	control for the effects of changes in energy prices	no	no	no	yes
2.	control for the effects of changes in income, profits, or net wealth	no	maybe	no	yes
3.	control for changes in macroeconomic conditions or business cycles	no	no	maybe	yes
4.	control for changes in economic structure or activity-levels	no	no	yes	yes
5.	control for consumer learning, adaptation, and changes in preferences	no	no	no	yes
6.	control for policy gaming (free ridership)	yes	yes	no	yes
7.	control for policy diffusion (spillover)	maybe	maybe	no	yes
8.	control for autonomous technology trends	no	no	no	yes
9.	control for changes in the number of installed efficiency measures	yes	no	no	no
10.	control for changes in post-installation hours of use estimates	yes	no	no	no
11.	produce revised savings estimates for individual programs	yes	maybe	no	maybe
12.	produce revised estimates of 1st year gross and net savings	yes	yes	no	no
13.	produce sufficient information for determining financial incentives	yes	yes	no	maybe
14.	produce a complete estimate of the standard error of net impacts	no	yes	no	yes
15.	produce empirical evidence of policy-related changes in energy consumption and thereby provide evidence of policy-related changes in GHG emissions	no	no	no	yes

In these comparisons, study quality issues are set aside. Rather, the comparisons try to capture the capability of the methods, at their best, to deal with factors related to energy efficiency policy impacts. For example, as can be seen with the indicators method which includes everything from univariate statistics to energy intensity index decomposition, even at its most rigorous its can only be sure to address activity or structural effects. Since activity is often related to macroeconomic conditions and business cycles, under certain circumstances the energy indicators method may address this factor, too, hence the entry of *maybe* for this factor.

The point of this exercise is to show the fundamental differences and similarities in the kinds of issues that each method is capable of taking into account. From this exercise several conclusions are apparent. First, bottom-up monitoring and top-down energy demand models for the most part address different sets of issues. Second, short-term billing analysis addresses some of the issues that monitoring addresses and some of the issues that energy demand modeling addresses. However, it also fails to address some issues that the others do address. Third, the CPUC does not, as yet, have any significant EM&V studies that focus on the issues that top-down methods address. The ultimate goal of the MCM pilot project can thus be viewed as an effort to help determine whether it is feasible to develop an alternative MCM policy evaluation system that adequately takes into consideration the factors that bottom-up methodologies do not address.

This comparison also shows that one way to think of an MCM energy demand model is as a whole building billing data analysis writ large. Like billing data analysis, energy demand models compare pre and post energy consumption, or treatment and control group energy consumption, or a combination of both. Exactly which variables are contained in the energy demand model, exactly what functional form the model takes and what estimator is used, exactly what dates are used for pre-post period cutoffs and what sampling plan is used for determining who is in or out of the treatment and control groups – all of these are research design decisions depend on the particular policies being evaluated and the questions being asked.

What energy demand modeling lacks is the same thing that billing analysis lacks, which is the measure and program-level specificity offered by monitoring, metering, auditing, and building simulation. Whole building billing data generally cannot be used to measure end use or equipment energy consumption, nor the influence of, or participation in, more than one energy efficiency program. In the same vein, energy demand modeling cannot be used to estimate equipment retention rates, hours of use, or any other measure-related relationships. Nor can it definitively distinguishing between the impacts of two or more overlapping energy efficiency programs, policies, or implemented energy efficient technologies. Metaphorically, energy demand models paint landscapes, not portraits.

Nevertheless, it appears as if energy demand models should be an integral part of the CPUC's future EM&V framework. Until now, all of the CPUC's EM&V efforts have focused on painting portraits. These serve the needs of policymakers and program managers for feedback and accountability, and the needs of stakeholders for documentation of program activities. However, top-down evaluations are now needed to estimate the impacts of energy efficiency policies on total energy consumption and environmental protection.

To put the CPUC's current EM&V activities into perspective relative to MCM policy evaluations it is instructive to discuss MCM in light of five major issues. These are: policy impact uncertainty, EM&V cost savings, synergies with market effects and behavioral studies, synergies with CEC load forecasting, and attribution capabilities.

- **Policy Impact Uncertainty**

All bottom-up methods for estimating program-related "energy savings" revolve around estimating first-year savings for new participants (with or without the aid of a comparison group). The precision of these energy savings estimates is necessarily partial because standard errors are not calculated for all of the relevant variables that make up energy savings. For example, in monitoring studies the standard errors associated with pre-measure installation hours of use are not known because monitoring is not done for the pre-installation hours of use. As a result, it is *assumed* that pre-installation conditions are known, i.e., that the standard error of the estimate of pre-installation hours of use is zero. The same is true for many other variables that go into bottom-up studies; their estimates are assumed to be perfectly precise when, in fact, they are far from precise.

Not only is there more uncertainty in bottom-up, first-year estimates of program energy savings than is generally acknowledged, but there is additional uncertainty that goes unacknowledged when using these energy savings estimates for; (a) calculating lifetime program energy savings; (b) calculating the combined effects of all new energy savings from all programs in a given year; and (c) calculating the cumulative effects of energy efficiency programs over two or more years.

The potential magnitude of the unacknowledged uncertainty in first-year, bottom-up estimates of energy savings can be shown mathematically. However, a brief look at what has happened in California in the past few years is sufficient for illustrating the level of uncertainty in current EM&V savings estimates. In the past three to four years, California's economic crisis has caused innumerable personal and business bankruptcies and an unprecedented flood of mortgage defaults and real estate foreclosures.

As shown in Figure 1 and Figure 2, residential housing foreclosures and unemployment rates rose dramatically since 2007.

Figure 1: California Annual Residential Foreclosure Rate (1998-2010)

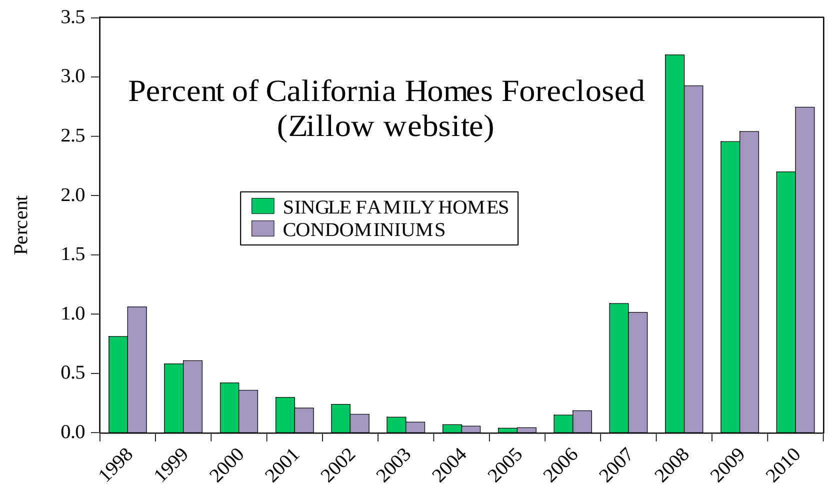
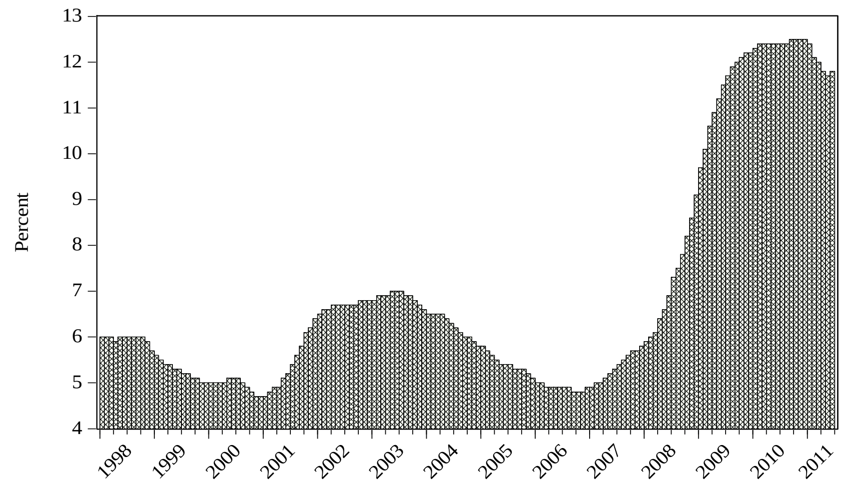


Figure 3: California Monthly Unemployment Rate (1998-2010)



As a result, key drivers of residential and commercial sector energy savings, such as run times, usage settings, and installation rates that were verified by bottom-up evaluations in the past evaluation cycle, and continue to be verified at present, are likely to be changing unexpectedly. Hence the findings of impact evaluations such as The Cadmus Group et al., (2010), in which one of the major goals was to revise estimates of new program savings for residential retrofit high impact measures by estimating ex-post measure installation rates and run times for several energy efficient appliances and equipment types, are now likely to be unreliable. Furthermore, the actualized energy savings from these measures is likely to continue to change unexpectedly for the next few years as the economic crisis passes and a new economic recovery takes place. On top of all of this, survey estimates of free ridership and other consumer actions – all of which are based on consumers' *stated expectations* – are likely to be entirely different from what they were a few years ago when economic conditions were more stable. These, too, are likely to change yet again when a new recovery begins.

Policy impact uncertainty can be reduced considerably by using aggregate, time series total energy consumption data to estimate long-term energy savings. Indeed, D.10-10-033 references Horowitz (2011) as the kind of MCM study that it is interested in pursuing. As a part of this study, residential electricity sales are analyzed for four major utilities in the state of California. The findings indicate that for 2006 and 2007 the long run energy savings from energy efficiency policies averaged about 4.8 percent per year, and the relative precision of the long-run energy savings estimates for each utility roughly averaged plus or minus 12 percent at the 95 percent confidence level. This high level of accuracy was achieved using an econometric model containing panel data, that is, sixteen continuous years of data for each of the four utilities.

- **EM&V Cost Savings**

Relatively large costs have been incurred using bottom-up EM&V methodologies to meet the CPUC's needs of measurement and verification of savings resulting from energy efficiency measures, programs, and portfolios. These costs have been incurred not because of the time it takes to perform statistical analyses, but rather because bottom-up studies are used *for the purpose of estimating first-year savings*. This requires large customer and measure samples, on-site visits by engineers and technicians, and other expensive primary data collection efforts. However, D.10-10-033 indicates that EM&V should now reflect a reasonable balance between evaluation costs and their benefits:

> Measurement and verification of savings resulting from energy efficiency measures, programs, and portfolios serves the fundamental purpose of developing estimates of reliable load impacts delivered through ratepayer-funded efficiency efforts. Measurement and verification work should reflect a reasonable balance of accuracy and precision, cost, and certainty, and be designed for incorporation in procurement planning activities. (Appendix A)

As already noted, bottom-up EM&V methods focus on estimating first-year savings. The most costly of these methods, monitoring and metering studies, rely on short-term, *in situ* measurements of only a few of the many parameters that drive estimated savings in engineering models and algorithms. Less costly but still expensive methods, such as pre-post billing data analysis (for participants and matching non-participants) focus on whole building energy use rather than measure-specific energy use. These, also, are limited to estimating first year savings and rely heavily on surveys for collecting consumer characteristics data and variables that help address self-selection and free ridership. The survey data for these evaluations are of a short-term and changeable nature, too.

Given the long-term energy savings uncertainty inherent in the findings from these types of impact evaluations, a better balance between evaluation costs and their benefits can be found in using inexpensive MCM evaluations to estimate aggregate, long-term policy impacts, and using bottom-up EM&V activities as program quality control and program management tools.

- **Synergies with CEC Load Forecasting**

MCM policy studies are, in reality, a particular form of load forecasting. This is because estimating policy impacts requires estimating a *counterfactual,* and a counterfactual is a “forecast” of what would have happened in the absence of policies. In other words, a counterfactual is what load forecasters call the business-as-usual or baseline condition, albeit one that has been focused on revealing policy impacts. Since MCM studies produce estimates of business-as-usual conditions, these policy-focused estimates can be compared to those made via conventional load forecasting models (such as the CEC’s technology-driven load forecasts) that do not explicitly recognize policy impacts other than as an afterthought.

Comparisons of MCM forecasts with conventional load forecasts can yield important new information and insights into the trends in energy consumption. For example, they can reveal that energy price elasticities or energy income elasticities have changed over time and need to be revised in the conventional models. Most load forecast models have limited tools for studying policy impacts because energy demand relationships are fixed in advance of load estimation. MCM studies, on the other hand, analyze the actual, revealed relationships among the drivers of energy consumption. By combining the empirical information derived from the MCM models with the highly-structured, and often dated, assumptions embodied in conventional load forecasting models, better load forecasts may be produced. Moreover, better estimates of long-run policy impacts are obtainable.

- **Synergies with Market Effects and Behavioral Studies**

Not only can an MCM database be used for estimating aggregate energy savings for groups of customers in geographic areas, but it can also be used as the basis for investigating market effects and behavioral studies. This is because through the MCM database various locations can be identified where for example, energy consumption may be falling more rapidly than in other locations in the state. This information can allow evaluators to target these locations for further study. Much more about these synergies will be known once MCM studies begin.

- **Attribution Capabilities**

MCM policy evaluations, as shown in Horowitz (2007), are capable of estimating policy-related energy saving (and GHG emission reductions) by fuel and economic sector, e.g., residential, commercial, and industrial/agricultural. Also, as shown in Horowitz (2011), MCM studies are capable of estimating policy-related savings by utility service territory. Given smaller geographic units, such as counties and census tracts, there is no reason why the same MCM approach used in these studies cannot be applied. Further, given the ability from utility billing data to identify and isolate different building types (e.g., single versus multi-family, or office buildings versus other commercial buildings) it should also be possible to estimate long-run policy impacts by building type. The information contained in utility billing records might also be able to be used to create cohorts of buildings that were built under different versions of Title 24 building codes. This could enable the impacts of building codes to be estimated using the MCM approach.

The capability of using the MCM approach for smaller and smaller levels of program attribution is something that can only be discovered once an MCM database is constructed and the potential information contained in the IOUs' billing data files is ascertained. However, regardless of the level of attribution that MCM evaluations can attain, it is expected that the CPUC will continue to rely on conventional, bottom-up, EM&V studies to know which programs are performing better or worse than planned. One of the benefits of MCM is that it is not a replacement for bottom-up studies, but rather a supplement that frees conventional EM&V studies to concentrate on what they do best, namely, assist in program planning, program management, program quality control, and contractor performance auditing.

In summary, just as a painted landscape is not an alternative to a portrait, MCM studies using energy demand modeling are not alternatives to CPUC's EM&V bottom-up evaluations. However, neither are bottom-up evaluations alternatives to aggregate energy demand modeling for meeting all of the CPUC's current needs.

3. Econometric Energy Demand Models

3.1 Literature Review

Econometric energy demand models are not to be confused with engineering-based energy demand models in which thermodynamic principles are used to analyze energy consumption. Econometric models are probabilistic and are concerned with the economic behavior of consumers. Like many other subjects in which econometric techniques are used to analyze time series, cross section, or panel (time series cross section) data, empirically-based energy demand studies have been popular for more than half a century. In most of these studies market demand is analyzed in a partial-equilibrium framework with reduced-form equations taking historical energy prices, incomes, weather, and other trend variables as fundamental drivers.

An econometric energy demand model is a regression equation, or group of equations, that takes either energy consumption, energy expenditures, or an energy ratio as their outcome or dependent variable. Long before there were such things as voluntary and mandated energy efficiency programs, economists had already produced scores of studies using energy market demand models. The majority were focused on electricity and natural gas demand. A few, but not many, used individual consumer microdata data, but most employed aggregate data at the residential, commercial, or industrial sector levels for consumers located in local communities, cities, counties, or states. Perhaps as a result of the disarray of the electric utility industry in the 1970s and the need for better information than had previously been used for power planning, many market demand studies were published in the decades of the 70s and 80s. Surveys of the early generation of energy demand studies include Hartman (1979), Bohi (1981), Bohi and Zimmerman (1984), Watkins (1992), Dahl (1993), Kneese and Sweeney (1993), Berndt (1996), and Fisher (1999).

Given the expense and practical difficulty of repeatedly collecting primary data for large samples of cross sectional observations, it is no surprise that from the early days of energy consumption studies to now, a majority of energy demand studies employ aggregate rather than individual consumer data in their models. Aggregate data are, of course, the microdata from individual cross sections that are summed or averaged across many consumers of a single population, sub-population, or group of interest. These data are usually collected by government agencies that are bound by confidentiality laws to keep the microdata secret. Because the agencies that collect these data do so on a permanent basis, aggregate data tend to be of consistent quality and scope; moreover, they are usually free to the public. For the most part, with the exception of official censuses and other specialized surveys collected by the federal government, these data are collected at fixed frequencies that are no longer than a year in length. This makes them usable for time series, cross section, or panel studies.

Given the public availability and high quality of aggregate energy data, from the earliest, classic econometric studies of electricity and natural gas market demand of Fisher and Kaysen (1962) and Balestra and Nerlove (1966), to the more recent studies of Bernstein et al. (2003) and Horowitz (2011), aggregate data have remained the single most cost-effective and timely resource for studying energy demand. What is different between the earlier energy demand studies and the more recent studies is not only the range of aggregate data that is currently available, or the new research software and hardware that is available, but the subject matter of the studies.

To a large degree the older studies struggled with strategies for accurately estimating the impacts of market factors, specifically, energy prices, consumer incomes, and appliance and equipment holdings on energy demand. For example, Fisher and Kaysen (1962) and Houthakker and Taylor (1970) focus on equipment stock and utilization rates, and Taylor (1975) and Berndt (1984) focus on the desirability of using marginal rather than average electricity prices for estimating short and long-run energy price elasticities. Many of these technical issues have since been resolved. Flow or partial adjustment models have become an accepted method for handling stock turnover investigations, and average prices have been accepted as the most practical variable for most model specifications. Other notable early studies that estimate energy demand models include Halverson (1975), Nelson (1975), Murray et al. (1977), Beierlein et al. (1980), Blattenberger et al. (1983), and Taylor et al. (1984). The underlying purpose of most of these studies was to provide information for long-term load forecasting.

While aggregate energy demand studies like these older ones continue to be produced and to add valuable contributions to the energy and environmental literature, e.g., Considine (2000) and Kamerschen and Porter (2004), the focus of several more recent econometric studies of aggregate demand is the impact of public programs and government policies on energy consumption. This interest is inspired by the rapid growth, especially since the late 1990s, in publicly-sponsored energy efficiency programs and policies. Up till now the findings of these studies have only been used to inform energy and environmental policies. However, there is no reason why in the future they cannot also be used to inform load forecasting and infrastructure planning.

Four econometric studies, Parfomak and Lave (1996), Loughran and Kulick (2004), Auffhammer et al. (2008), and Arimura et al. (2009) investigate the impact of electric utility demand side management (DSM) programs on electricity demand using retail electricity sales data. The former uses aggregate panel data for the commercial and industrial sector for 39 electric utilities in ten states for the years 1970 to 1993 to estimate a fixed effects, weighted least squares energy demand model. Controlling for market factors, this study confirms that 99 percent of the DSM savings reported by electric utilities for programs begun in the 1980s and early 1990s were realized.

Alternatively, the three latter studies use the combined retails sales data of the residential, commercial, and industrial sectors for electric utilities, along with their reported total annual DSM program expenditures since the early 1990s, to estimate DSM program cost-effectiveness. Controlling for market factors and employing a fixed effects panel model, using their best model Loughran and Kulick (2004) find an average cost to utilities for a saved kilowatthour of approximately 6.5 cents (2000$). Auffhammer et al. (2008) and Arimura et al. (2009) are follow-on studies that use the same EIA-861 data as Loughran and Kulick (2004). The first of these studies addresses bias in the Loughran and Kulick standard errors of the cost–effectiveness estimate, and the second updates the Loughran and Kulick study with a non-linear panel model containing somewhat different independent variables.

Not surprisingly, the findings of all three of these studies are not substantially different from each other and are somewhat disappointing to policymakers. One possible cause for these results is a poorly-measured policy variable. This flaw is the source of what Hausman (2001) refers to as the "Iron Law of Econometrics." In the context of policy-focused energy demand models it produces policy impacts that are downward biased in magnitude, towards zero. Other policy-focused energy demand model flaws, such as misspecified models or coarse research designs, will also tend to produce underestimates of policy impacts. This may be why Kavalec (2011) in a recent modeling experiment for the CEC finds little energy savings from IOU programs.

Bernstein et. al, (2003) investigates state-level total primary energy consumption, i.e., the sum, in Btu, of all fuel consumption combined. After calculating energy intensity for each state in the residential, commercial, industrial, and transportation sectors, fixed effects panel models are estimated to control for market factors, state-specific unmeasured factors, and year-specific unmeasured factors. From each of these models, the average annual percent change in a state's residuals is used as a state ranking device. According to the authors:

> By definition, we do not know what is contained in this "residual energy intensity" $\hat{e}_{it}$, but we assume that it partly represents unobserved differences in energy-related policy both across and within states over time. As we discussed earlier, different states have pursued different energy policies over the period covered by our data (1977–1999). For example, the states' building codes differ, and the amounts that states spend on DSM and other programs differ, all of which may affect energy use in the states. We do not claim that the residual energy intensity captures only state-specific policy, but rather that policy is likely to be an important component of the residual energy intensity; therefore, the residual energy intensity may contain useful information about the role of policy in lowering energy intensity. (p. 21)

In other words, a state's residuals are employed as policy counterfactuals that implicitly reveal how actual energy intensity has varied from expected energy intensity.

Horowitz (2004) uses a state-level commercial sector panel model to study the effects of both electric utility DSM programs and a broader set of energy efficiency programs known as market transformation programs – which include regional and federal programs as well as codes and standards programs – on commercial sector electricity intensity across 42 states from 1989 to 2001. The analysis separates market effects from public program effects, finding that DSM programs were responsible for reducing commercial sector electricity intensity in 2001 by 1.9 percent relative to the 1989 level and that market transformation programs were responsible for reducing electricity intensity in this sector by 5.8 percent relative to the 1989 level. In 2001, it is estimated that the combined effects of these public programs reduced commercial sector retail electricity sales by 77.1 million MWh, representing about 2.3 percent of total U.S. retail electricity sales.

Horowitz (2007) analyzes changes in national electricity demand and estimates the impacts of electricity energy efficiency policies in the commercial, industrial, and residential sectors from 1992 to 2003 using the 48 mainland states and separate panel models for each sector. This study finds that those states that have at least moderate commitment to energy efficiency programs reduced electricity intensity relative to what it would have been with weak program commitment; in the residential sector by 4.4 percent, in the commercial sector by 8.1 percent, and in the industrial sector by11.8 percent. Moreover, the evidence in this paper indicates that energy efficiency programs in all three sectors of the U.S. economy have transformed electricity demand with respect to at least three key economic variables: electricity price, income as measured by per capita income or gross state product, and technology growth.

Metcalf (2008), employs an econometric approach that takes advantage of energy intensity index decomposition, the subject of the second top-down methodology discussed in this white paper. Like the previous policy-oriented studies, a panel model is specified, this time using the 32 years from 1970 to 2001, and 46 of the 48 mainland states. With the four main sectors of the economy, that is, residential, commercial, industrial, and transportation, total energy consumption, in primary Btu, is used to produce a total energy intensity index, an activity index, and an efficiency index for each state in each year. Three separate regression models are then estimated with each index being a different dependent variable, and comparisons, especially of the price and income coefficients, are made between them. The study finds that rising per capita income and higher energy prices helped lower energy intensity, implying that market factor are the main drivers of energy efficiency in the economy. Based on these finding the author concludes that the Bush Administration's 2002 goal of achieving an 18 percent reduction in carbon intensity by 2011 would be met without the need for significant policy intervention.

Like the former study, Sheppard et al. (2010) also combines energy intensity indicators with panel modeling. However, its main emphasis is the energy intensity indicator, not the models. The study period spans 30 years and uses a panel fixed effects model only for the purposes of weather-adjusting energy intensity ratios and testing the sensitivity of the changes in the ratios to energy prices and disposable income. The goal of the study is to rank states based on improvements in their residential sector energy intensity ratios, where total primary energy consumption is the numerator of the ratio and population is the denominator. The study implies that these ranking reflect the impacts of energy efficiency policies. However, there is no rigorous control from market factors that can support this idea.

Horowitz (2011) presents four examples of how different types of aggregate data, i.e. annual and monthly and state and utility level, along with different variations of time series and panel models, can be used to measure or verify the total, long-term energy savings that come from electricity efficiency policies. This study fundamentally differs from Parfomak and Lave (1996), Loughran and Kulick (2004), and Horowitz (2004) in that a target variable explicitly representing policy activity is *not* included in the energy demand model specification. Like in Horowitz (2007), Horowitz (2011) demonstrates that a policy-focused energy demand model does not require inclusion of an explicit policy-related variable if the evaluation research design and energy demand model are properly coordinated. This innovation offers a valuable lesson for future policy evaluation studies because longitudinal, interval-level policy-related variables often either do not exist or are flawed with unacceptably large errors of measurement.

Table 2, based on studies the studies of Horowitz, shows three ways that energy demand models can be designed to produce counterfactual estimators. Horowitz (2004) uses the target variable approach in which two different indicators are used as independent variables to represent energy efficiency policy savings: the first is an estimate of cumulative DSM savings in the commercial sector, and the second is an index of market transformation impacts based on national shipments of energy efficient lighting equipment. In Horowitz (2007) the difference between pre and post-treatment simulated and actual use is calculated for states with strong commitments to energy efficiency policies. Lastly, in Horowitz (2011) out-of-sample model residuals, that is, the difference between actual and forecasted energy consumption in the policy period being studied, are used to quantify policy impacts.

Table 2: Energy Demand Model Policy Impact Estimators

MCM Study	**Analysis Approach**	**Impact Estimator**
Horowitz (2004)	Target Variable	a) DSM Savings b) Market Trans. Index
Horowitz (2007)	Simulation	Difference-in- Differences
Horowitz (2011)	Policy Residual	Load Forecast

Although more impact estimators may yet be devised, the different impact estimators used in these three studies represent all of the estimators that are currently found in the econometric literature. There are pros and cons to each of them, and each must be judged, and deployed, based on how well they address the policy evaluation question being posed. Like all statistical techniques, each requires the acceptance of strong assumptions whose reasonableness depends entirely on the individual research contexts in which they are used.

3.2 Econometrically-Derived Policy Counterfactuals

To reiterate, the single most important question for any policy evaluation methodology, be it bottom-up or top-down, is the nature of the *policy counterfactual* it can produce, a policy counterfactual being an estimate of what energy demand would have been in the absence of the policy. As has been shown, there are many ways to produce policy counterfactuals. In general, the more non-policy related factors a methodology can take into account, the more reliable the counterfactual. This is because the counterfactual is the business-as-usual energy demand. By comparing business-as-usual demand to actual policy-influenced demand, the magnitude of the policy influence can be deduced.

Fundamentally, the econometric approach to policy evaluation relies on variations in governmental policies and market factors – be they over time, over cross sections, or over both -- to explain changes in energy demand trends and thereby infer what would have occurred in the absence of policies. Econometric modeling is unique as a statistical methodology in that it is able to simultaneously control for multiple factors that affect energy consumption. However, powerful statistical techniques alone are not sufficient for analyzing a topic as complex as energy efficiency policy impacts. Energy efficiency policies influence market behavior slowly through many different pathways; they do not have the immediate and dramatic effects sought by social engineering policies. Thus, detecting policy impacts is a difficult task that not only requires powerful statistical techniques but felicitous research designs that are sensitive to how policy environments and market conditions interact.

Given a well-specified energy demand model and an appropriate research design, a model-based policy counterfactual will have the following qualities:

- by containing a variable that represents activity levels (either as the denominator of the dependent variable or as a separate independent variable), the model will produce a counterfactual in which changes in activity or economic structure are controlled for;
- since energy prices, consumer incomes, and technological factors drive capital investments, controlling for these factors in the model will produce a counterfactual that is net of free ridership;
- since aggregate energy demand is the outcome variable of the model, the effects of policy spillover, rebound, and equipment interactions will be absorbed into the counterfactual;
- when the energy demand model has a time series component, the counterfactual will reflect changes in measure retention, behavior persistence, and other factors related to energy savings.

The empirical analyses in Horowitz (2011) illustrate the value of combining statistical expertise with research design expertise. In this study, four different policy evaluation research designs are implemented for policy evaluations in the residential sector. They are a state-level annual time series econometric model estimated for a pre-treatment period from which annual treatment period state energy use was forecast; a state-level monthly time series econometric model for a pre-treatment period from which monthly treatment period state energy use was forecast; a four-utility annual time series cross section model for a pre-treatment period from which annual treatment period energy use for each of the four utilities was forecast; and lastly, a three-utility annual time series cross section model for a pre-treatment period from which annual treatment period energy use was simulated for 27 same-state public utilities. From all four of these sample evaluations, counterfactuals and aggregate energy savings were estimated.

For the remainder of this section, three new examples will be provided of realistic applications of energy demand modeling to impact evaluation topics in California. They were selected to illustrate how the energy demand modeling approach can be customized to the different economic sectors of California and to the different layers of programs in California. These three policy evaluation examples are:

1. a manufacturing sector study using a time series cross section model is readily adaptable to one or more industries or sub-industries, be it agriculture, resource extraction, or manufacturing;
2. a cohort analysis designed to estimate the aggregate energy savings from statewide Title 24 building energy efficiency standards for residential homes;
3. a utility-level analysis designed to estimate the aggregate energy savings for commercial buildings, individual building types, and identifiable program participants and non-participants.

The actual collection and compilation of the data needed to implement these policy evaluations are the subject of *Demand Research's* MCM pilot project proposal in the second part of this white paper. It is expected that an the residential portion of the MCM database will be operational within the first four months of the MCM pilot project. For the present discussions, it is assumed that a database platform with the following features is in place:

a. an updatable geocoding system that permits publically-available historical data for as far back as 1990 to be joined down from the 58 counties in California to the approximately 6,000 five-digit zip codes and approximately 7,000 census tracts in California;
b. weather stations are joined at the five-digit zip code level so that every part of California is designated its best possible historical heating and cooling degree data;
c. all public and investor-owned utility service territories are joined to the five-digit zip codes;
d. upon request from the CPUC, the investor-owned utilities will provide historical billing and pricing data for specific rate classes, geographic areas, NAICS categories or other identifiable groups of customers (hopefully, the public utilities will cooperate with these data requests, too);
e. data specifications will be clearly communicated to the utilities and *Demand Research* will process these data for use in the MCM database; and,
f. the appropriate parties, be they program implementers or CPUC evaluation contractors or others, will provide information to identify program participants by the location, type of program, and year of participation.

The three policy evaluation examples below are intended to be general enough to describe the use of econometric modeling in the residential, commercial, industrial, and agricultural sectors as well as to describe the different kinds of policy questions that can be answered. The research designs themselves are governed by the specifics of different policies, e.g., the targeted consumers, geographic areas, and energy efficiency goods and services. As suggested by these three examples, with the flexibility offered by econometric models and the appropriately-stocked geographically-disaggregated database, there are innumerable MCM policy evaluations that can be performed annually.

3.3 MCM Evaluation Examples

3.3.1 Manufacturing Sector Study

The fundamental question for a policy evaluation methodology is the nature of the policy counterfactuals that it can produce. A counterfactual that is only valid for first-year participants is not the same as one that is valid for all participants new and old. Likewise, a counterfactual that only controls for the effects of weather on energy demand is not the same as one that controls for economic factors in addition to weather. An example of a policy counterfactual function that takes into consideration all of the major factors driving aggregate energy consumption is: :

$$C_{it} = f(M_{it}, G_{it}, P_{it}, N_{it}, L_{it}, V_{it})$$

In this function, C is an energy demand model-derived counterfactual at time t for a group of manufacturing industries, and i represents the individual manufacturing industries in the group. This counterfactual quantifies the level of aggregate energy use that would have occurred in year t after controlling for market and non-market drivers of energy demand. The market drivers include, but are not limited to, GWH consumption, energy expenditures, or one of many other variables indicative of an industry's energy consumption represented by *M;* data related to industry activity level such as industry value added (GDP) or units shipped represented by *G;* energy prices represented by *P*; one or more final or intermediate factors of production, represented collectively by *N*; and perhaps geographic, regulatory, technological or other factors represented by *L*. A non-market or policy-related driver variable may be represented by *V*. Once the counterfactuals are estimated for each industry, the total energy efficiency policy impact at time *t* can be calculated as the difference between the sum of the counterfactuals for each industry and the sum of the actual energy use of each industry:

$$\text{Policy Impact}_{\ t} = \sum C_{it} - \sum M_{it}$$

The general approaches to evaluating the impacts of energy efficiency policies in the manufacturing sector are similar to those that can be used for evaluating policy impacts in the industrial sector as a whole, as well as in individual industries and separate sub-sectors such as agriculture, resource extraction, durable goods, and non-durable goods.

Due to the relatively small number of companies in many industries and the wide mix of activities between industries, there likely to be many occasions in which MCM evaluations of industrial and agricultural sector energy efficiency policies are not amenable to geographic disaggregation below the state level. As such, it is instructive to offer an example of how a state-level MCM evaluation might take shape. With energy consumption data provided by utilities at the NAICS-level and annual state-specific industry data available from government agencies, a logical choice for a manufacturing sector analysis is to begin at the three-digit NAICS level. For most federal data series NAICS classifications took over from SIC classifications beginning in 1997, so this is generally the earliest year when an industrial sector analysis should begin. Particularly in the manufacturing sector, but also in the natural resource and commercial sectors, this does not necessarily pose a handicap. Arguably, U.S. industries have changed enough since the mid-1990s that going back any further than this year creates more problems than are solved by adding extra data to the analyses.

To provide a sense of the data processing that is needed for performing this study, note that there are 21 (actually, 22 if the two transportation sub-categories are counted) three-digit NAICS industries that make up the durable and non-durable goods manufacturing sector that are commonly used by government agencies for reporting. The leftmost columns of Table 3 list these 22 industries. For standardization with other federal datasets such as the national industry accounts, these 22 industries must be collapsed into their lowest common units, which are the 18 industries found on the right of Table 3. For example, NAICS 311 (food) must be combined with NAICS 312 (beverage and tobacco products). Only when this is done can industry-level price deflators be joined with the economic data from other federal government datasets.

Any number of different kinds of manufacturing sector energy demand econometric models could be specified and estimated depending on the available data and the assumptions made about the relationships among the variables. The estimated models could then be used to measure the marginal impacts of each independent variable on the dependent variable, to test hypotheses, and to make forecasts or construct counterfactuals. In the absence of public policy impacts there would be no need to construct a policy counterfactual, and the demand function above would simply show the contents of a reduced-form energy demand analysis.

Table 3: Dataset Standardization Scheme for 3-Digit NAICS

NAICS	22 Categories (depending on data source)	NAICS	18 Industries (after standardization)
311	Food	311, 312	Food, beverage plus tobacco
312	Beverage and tobacco		
313	Textile mills	313, 314	Textile mills plus products
314	Textile product mills		
315	Apparel	315, 316	Apparel, plus leather, allied
316	Leather and allied		
321	Wood	321	Wood
322	Paper	322	Paper
323	Printing	323	Printing
324	Petroleum, coal	324	Petroleum, coal
325	Chemical	325	Chemical
326	Plastics, rubber	326	Plastics, rubber
327	Nonmetallic mineral	327	Nonmetallic mineral
331	Primary metals	331	Primary metals
332	Fabricated metal	332	Fabricated metal
333	Machinery	333	Machinery
334	Computer, electronics	334	Computer, electronics
335	Elect. equip., appliances	335	Elect. equip., appliances
366	Transportation	336	Transportation
3361-3363	Motor vehicles, parts		
3364-3366,3369	Other transport. equip.		
337	Furniture and related	337	Furniture
339	Misc. manufacturing	339	Miscellaneous

Explicit inclusion in an energy demand model of a policy-related variable or instrument that varies over industries (the term V in the full energy demand function above) is the most direct way of estimating public policy impacts. This would constitute the “target variable” impact estimator approach. However, it is frequently the case that a useful instrument cannot be found. The alternative is then to configure the underlying research design so that the energy demand model allow for inferences about policy impacts. For example, there may be industries that receive different levels of treatments which, in effect, can be used to specify a treatment-control group natural experiment. Likewise, the measurements of a relevant variable for the same subjects for the periods before and after policy implementation may provide sufficient contrast for estimating policy impacts. Various schemes such as these are sometimes referred to as interrupted time series designs or regression discontinuity designs. In the end, the research design that is chosen must depend on practical issues associated with the available data, the policies under study question, and the research questions that are posed.

The availability of U.S. Census Annual Survey of Manufacturers (ASM) data offers an opportunity to construct such a policy counterfactual by specifying a panel econometric model of energy demand for California manufacturing industries in one time span, and then using it to predict or simulate what total energy demand would in another time span. An example of a fully-specified panel model for the ten years of 1997 through 2006 could be:

$$E_{it} = \sum_{i=1}^{18} b_i N_i + \sum_{t=10}^{10} b_t T_t + b_1 PRICE_{it} + b_2 OTHPRICE_{it} + b_3 VALUE_{it} + b_4 CAP_{it}$$
$$+ b_5 HOURS_{it} + b_6 EMPLOY_{it} + b_7 PAYROLL_{it} + b_8 MATERIAL_{it} + \varepsilon_{it}$$

where *i* represents an individual NAICS industry in California and *t* represents individual years. The definition of the dependent variable, *E,* in this equation is *MWh*, but this is only one of many ways of expressing energy demand. The other definitions of the terms in the equation are as follows:

PRICE	=	price per purchased MWh
OTHPRICE	=	price per alternative fuel(s)
VALUE	=	industry value added (aka, industry GDP)
CAP	=	total capital expenditures
HOURS	=	total production hours
EMPLOY	=	total number of industry employees
PAYROLL	=	total industry payroll
MATERIAL	=	total expenditures on materials
N	=	18 individual NAICS industries
T	=	10 years
b_i	-	industry fixed effects coefficients
b_t	=	time fixed effects coefficients
b_1 - b_8	=	continuous variable coefficients
ε_{it}	=	error term.

Once the model specification and functional form are finalized, estimates of the impact of energy efficiency policies on *MWh* demand can be produced by forecasting counterfactual energy use for the out-of-sample years, i.e., 2007, 2008, and 2009. This constitutes the “policy residual” impact estimator approach. Forecasting, projecting, and simulation are terms that are closely related to each other and are often used synonymously. To the degree that there are differences, it is because the former terms connote attempts to predict future behavior using past behavior, while simulation is broader and connotes attempts to predict past, current, or future behavior given the behavior of a separate but similar set of subjects. Subtraction of the values of the counterfactual from actual energy use in each year produces a remainder that is interpretable as a *policy impact*.

As already noted, this model will produce estimates of total policy impacts for the entire state and for all manufacturing sector energy efficiency policies combined. In and of itself this is valuable high-level information. However, there are many ways in which it can be made even more valuable. For example, it may be that the model can be broken into separate models for different groups of industries. Indeed, smaller models can even be produced for each individual industry by extending the data backwards and forwards for a few more years. For disaggregating the impacts of different manufacturing sector energy efficiency policies, it may be useful to produce a logic model and a policy timeline, and to specify the model(s) and interpret various findings accordingly. Ultimately, the scope and character of the industrial and agricultural energy demand models and the analyses depend on data availability, but also on the history of the energy efficiency policies in these industries and on the kinds of questions for which answers are sought.

3.3.2 Residential Building Energy Efficiency Standards Study

The prior example described an energy demand model estimated at the highest level of geographic aggregation, the state of California. The next example of an MCM impact evaluation describes an energy demand model that is meant to be estimated at a very low level of geographic aggregation, the census tract level. This evaluation focuses on residential building standards. As with all MCM investigations, many types of statistical analyses will be possible once a geographic database containing basic economic and social data is operational. In this example, begin with a function:

$$Y_{tk} = f(P_{tk}, G_{tk}, F_{tk}, W_{tk}, S_{tk}).$$

where the subscript t represents a given year and k represent a given geographic unit, in this case, census tracts. Like in all energy demand models, in this function Y could be the total level of energy use or an energy intensity ratio of some kind, such as energy use per house. Determinants of energy demand or intensity may include and of course are not limited to electricity prices as represented by *P,* and perhaps also the price of the nearest substitutable fuel, such natural gas prices; absolute or relative incomes as represented by *G*; socio-demographic conditions represented by *F*; climatic conditions as represented by *W*; and one or more variables related to changes in equipment stock as represented by *S*.

Generally speaking, utilities maintain information not just on customer turnover on existing meters, but on new service hook-ups. With this information it is possible to create vintage cohorts of houses that were built and occupied in different years. For example, the billing data for houses built after 2005 should be separable from the billing data for houses built through 2005. Likewise the billing data for houses built prior to 2002. Once the billing data for pre-defined vintage cohorts are isolated, it should be possible to organize them by the census tracts.

There are roughly 7,000 census tracts in California, and many of them may have few residences or may have some other feature that might exclude them from the study sample. Nevertheless, the cross section sample may be quite large. Suppose an evaluation is focused on three housing vintage cohorts, the first one composed of houses built between 1997 and the Title 24 update of 2001, the second one of houses built between 2002 and the Title 24 update of 2005, and the third of houses built from 2006 to 2011. All things being equal, if the building standards were followed and there was a high compliance rate, there should be less average energy consumption per house in each cohort. Of course, all things are not equal, and this is why an energy demand model is needed to control for residential energy prices, household size, and so forth.

The intention in creating the geographic database in the MCM pilot project is to have at hand the variables necessary to control for all the major factors that drive energy demand, thereby enabling valid policy counterfactuals to be constructed. Table 4 contains a selection of variables that are available at different geographic levels from the U.S. Census (Census) and the U.S. Department of Commerce Bureau of Economic Analysis (BEA). Using the appropriate geocoding, the decennial census tract level data can be linked to the country-level data and weighted, for example by population, employment, or personal income, so that annual estimates are available, by census tract, for all of the desired variables.

Table 4: Selected Residential Sector Socio-Economic Variables and Sources

Geographic Level	Variable	Frequency	Source
County	Population	Annual	BEA
	Personal Income	Annual	BEA
	Total Employment	Annual	BEA
	Median Age	Annual, Tri-annual	Census
	Education	Annual, Tri-annual	Census
	Persons per Household	Annual, Tri-annual	Census
Census Tract	All of the Above	Decennial	Census

There are many ways of estimating policy counterfactual for this evaluation once the data are in place. For example, a panel model can be estimated for the oldest cohort of houses. By definition, the model captures aggregate residential sector energy-related behavior, such as the how energy demand changes when energy prices, incomes, and heating and cooling requirements change. A simulation of energy demand for the residents of the newer houses using the model based on the behavior of the oldest house cohort can show how, all things being equal, the older houses perform vis-à-vis energy demand. A policy counterfactual such as this one might be derived from the following equation:

$$\hat{E}_{t,k,New} = b_{o,Old} + b_{1,Old} P_{t,k,New} + b_{2,Old} G_{t,k,New} + b_{3,Old} F_{t,k,New} + b_{4,Old} W_{t,k,New} + b_{5,Old} S_{t,k,New}$$

where the subscript *New* represent the houses built under the 2001 Title 24 revised standards and the subscript *Old* represents a cohort of older house built under a different standard. The total simulated energy use for the new houses is then:

$$C_{New} = \sum \hat{E}_{t,k,New}$$

which can be compared to actual energy use for these homes in the *t* year(s) being chosen for evaluation. Likewise, the most recent cohort of houses built to the 2005 Title 24 revised standards can be compared to the cohort of old houses, as well as to the cohort of houses built under the 2001 revised standards.

Many different model specifications and counterfactuals may thus be developed given several years of data, several cohorts, and many cross section (census tract) observations. Of course, the interpretations of the policy impact findings are not without complications given that there are many other energy efficiency programs and policies that overlap the revised building standards. At least some of these complications can be sorted out by carefully choosing research designs and model specifications that isolate the impacts of specific programs.

3.3.3 Commercial Buildings Study

A final MCM evaluation example involves statewide commercial buildings programs. In the manufacturing sector example a panel energy demand model was proposed for the state as a whole, with the cross sections being industries rather than locations. In the residential sector example, the building standards evaluation research design, the cross sections of the panel model are the state's census tracts, with the focus being different cohorts of residential homes. Now, in this third example, the geographic focus is neither the state as a whole nor the state broken into thousands of census tracts, but rather different utility service territories within the state.

The evaluation may begin with an analysis of the energy consumption of the entire commercial building sector by first estimating a panel model consisting of data for a selected group of public and private utilities in California, going back as far as 1995 and extending up until the present. Given that the MCM database is geocoded so that the variables can be compiled from the census tracts to the counties to the utility service territories, the commercial building sector models will have all the major independent variables necessary to explain aggregate energy demand. Many alternative counterfactuals can be produced with this data scheme, from simulated energy use for one sub-sample of utilities based on the behavioral model of the other sub-sample to a forecast of energy demand for all the utilities in the out-of-sample years.

With refinements to this basic research design it may be possible to differentiate the impacts of different programs. For example, with aggregated billing data, by county, for commercial buildings that participated in either the Continuous Energy Improvement (CEI) program or the nonresidential audit program, or both, in a given window of time, say between 2006 and 2008, it would be possible to compare their pre and post-participation usage. Also, the aggregate energy consumption of buildings that were eligible non-participants can be compared to aggregate participant usage.

Another research design and energy demand model that may be possible to be applied to the commercial building sector is an evaluation by building type or function. Assume that an IOU commercial sector energy efficiency program began in 2003 and was targeted specifically to the health services industry. An energy demand model may be specified such that the dependent variable is energy intensity *EI,* that is, aggregate health services building electricity use *E* divided by the total personal income for employees in this industry, *M*, and the independent variables are average retail price of electricity *P*; average retail price of natural gas *N*; local demographic characteristics *D,* local economic conditions captured by *G*; and climatic conditions *H*. In addition, the model contains a general time trend related to equipment stock *O* that has relevance for health services-related buildings. The general function of the energy demand model is:

$$EI_{t,s,R} = f(P_{t,s,R}, N_{t,s,R}, D_{t,s,R} G_{t,s,R}, H_{t,s,R}, O_{t,R})$$

in which subscript *t* represents a given year; subscript *s* represents a given service territory; and, subscript *R* represents the presence or absence of the health services-specific energy efficiency policy. This split-case function asserts that *R* influences each of the behavioral relationships associated with electricity intensity. In other words, *R* is a transformative agent. Additionally, the years over which the models are estimated are divided into two periods, a base period and a treatment period. The former consists of the years 1995 through 2002 and the latter consists of the years 2003 through 2010.

To estimate the annual energy savings due to the IOU health services program a simulation is performed to infer what the changes in energy use might have been for buildings located in one set of service territories, *W*, had they consumed energy like the buildings located in another set of service territories, *U*. First, an econometric model is estimated based on the specification:

$$EI_s^U = \beta_0^U + \sum_{j=1}^{n} \beta_j^U X_s^U + e_s^U$$

where the superscript *U* represents the control service territories and *j* represents the individual independent variables. In this specification, time series subscripts are suppressed, the β_j's are the coefficients associated with each of the *X* independent variables, and the *e* is an independent error term. This energy demand model is estimated separately for the base and treatment periods and then, using their respective coefficients, the energy demand for the *W*-located buildings is simulated. This is done by inserting the annual values of the independent variables for the *W*-located buildings into the period-specific energy demand models. The absolute level of energy consumption, *E*, for the *W*-located buildings can then be calculated by multiplying *EI* by its numerator, personal income, *M*. The formulas for these calculations are:

$$EI^{W}_{Simulated,t,s} = \beta^{U}_{0} + \sum\nolimits_{j=1}^{n} \beta^{U}_{j} X^{W}_{s,t}$$
$$E^{W}_{Simulated,t,s} = EI^{W}_{Simulated,t,s} \times G^{W}_{t,s}$$

In this way, the energy demand model produces estimates of the aggregate levels of electricity use of the health services buildings in the *W* utility service territories in any given year had their demand behavior been like that of the buildings in the *U* utility service territories.

Based on the aggregation of the individual utility simulations, total health services buildings energy savings for a chosen evaluation year in the IOU service territories, that is, *Health Services Savings*$_{t*}$, can be calculated using a difference-in differences estimator. One such estimator might be:

$$Health\ Services\ Savings_{t*} = C^{W}_{t*} - TE^{W}_{t*} = \left[\left(\frac{\sum \hat{E}^{W}_{t*}}{\sum \hat{E}^{W}_{tbase}} \right) \times \sum E^{W}_{tbase} \right] - \sum E^{W}_{t*}$$

where:

C^W	=	total simulated energy use for the *W*-located buildings in chosen evaluation year *t** based on the energy consumption behavior in the *U*-located buildings
TE^W	=	total actual energy use for *W*-located buildings in the chosen evaluation year
$\hat{E}^{W}_{t*}$	=	total simulated energy use for the *W*-located buildings in the chosen evaluation year
$\hat{E}^{W}_{tbase}$	=	total simulated energy use for the *W*-located buildings in the last base period year
E^{W}_{t*}	=	actual energy use for *W*-located buildings in the chosen evaluation year
E^{W}_{tbase}	=	actual energy use for *W*-located buildings in the last base period year

This policy impact estimator takes as its counterfactual the percentage change in the simulated energy use between the last pre-treatment year and the chosen evaluation year and multiplies this value by the actual energy use in the chosen evaluation year. The result is an estimate of what total energy use for health services industry buildings in the *W* service territories would have been in the evaluation year had energy demand been similar to the energy demand of *U*-located buildings. Since all major energy demand market factors have been controlled for in the energy demand model, the difference between the counterfactual value and the actual value for the *W*-located buildings is attributable to the *W* service territories' energy efficiency policies.

This ends the discussions of how econometric models, coupled with innovative and felicitous research designs, can be devised for any sector of the economy to estimate the impact of energy efficiency policies on aggregate energy consumption. Perhaps the most important lesson to be learned from these examples is that carefully devised plans are needed for detecting the elusive effects of energy efficiency policies on aggregate energy consumption. These plans must include geographically-disaggregated data and sector-disaggregated data as well as appropriate research designs and model specifications. Without all of these features coming together, few policy evaluations can be performed, and those that are performed will find it challenging to uncover policy impacts, real though they may be.

4. Energy Indicators

4.1 Literature Review

The term *energy indicator* can refer to a single number, but usually refers to a ratio in which the numerator is energy use or energy expenditures and the denominator is either a physical or financial quantity that is closely related or co-moves with the numerator. In this form, an energy indicator is typically referred to as *energy intensity* or an energy intensity ratio. Examples of energy intensity ratios are Btu per square foot, electricity per ton of aluminum, and energy costs per value added. Table 5 offers a glossary of terms related to energy indicators.

Table 6 provides a list of selected energy indicators maintained by ODYSSEE (2011), an organization located in France and funded by the European Commission, that tracks energy efficiency trends for the 27 European Union (EU) member states. This database contains a detailed hierarchy of energy indicators that includes time series of unit energy consumption values for such things as the major appliances in the residential sector.

Table 5: Glossary of Energy Indicator Terms

Term	Definition
Energy indicator	Any statistic that provides information related to energy consumption. Typically this statistics is composed of a time series, energy-related variable. However, any variable that is highly-correlated with energy use, such as the number of personal computers in a building, may also be considered an energy indicator.
Energy intensity	A ratio in which an energy-related variable such as energy use or energy costs is in the numerator, and the denominator of the ratio is a variable that is closely related to the numerator.
Energy intensity index	a statistic created by combining two or more energy intensity ratios; these indexes may be simple (an arithmetic average) or complex, depending on how they are constructed and how they weighted.
Activity index	a statistic created by rearranging the terms of certain types of energy intensity indexes in such a way that the effects of the relative sizes of the denominators of each ratio is quantified.
Efficiency index	a statistic created by rearranging the terms of a certain type of energy intensity index in such a way that the effects of the relative sizes of the denominators is removed from the energy intensity index.

Table 6: Selected ODYSSEE Energy Indicators

Industrial Sector	Numerator	Denominator
Steel	total energy use	ton of product
Glass	total energy use	ton of product
Food	total energy use	industrial production index
Textile	total energy use	industrial production index
Residential Sector		
Heating	unit consumption at normal weather	square meter
Cooking	unit consumption	dwelling
Refrigerators	kWh	unit
Televisions	kWh	unit

By itself and without any supplemental information, the change in an energy intensity indicator may yield useful information provided the numerator and denominator are so specific that there can be only one reason why the indicator has changed. For example, if a steel plant does nothing to its production process other than replace an old piece of equipment with one that does the same job using less energy, then a decrease in the ratio of energy input to steel output can confidently be attributed to the equipment replacement. However, suppose the plant also increases its steel output. An economist or engineer may now ask whether the decline was wholly due to the equipment replacement, or possibly due, at least in part, to production economies of scale. The problem is not that the energy indicator is not exact, but rather than its interpretation is cloudy. Production curve information must be introduced to determine the points at which the marginal product of energy increases, decreases, and stays the same. To properly assign causality to the improvement in energy intensity, an *efficiency* effect and an *activity* effect must be separated.

It is universally acknowledged that simple energy indicators like the one just described do not, except in the most limited of circumstances, yield insights into the impacts of public policies. Being difficult to interpret, their primary value is in provoking discussion and encouraging further research. One avenue for further research is energy intensity index decomposition. This method produces a statistically-refined energy indicator and as a result many governments and international organizations have adopted this approach not only for tracking changes in energy demand, but for tracking national changes in the use of many different natural resources. The benefit of index decomposition is that it directly links resource use to specific activities or physical objects, and then controls for changes in the volume of these activities or physical objects to determine how resource productivity has changed over time (Schipper, et al. 2001).

The index decomposition methodology has two attractive features; it has few data requirements and it is computed with a few basic mathematical identities. This makes it possible for index decomposition to be used everywhere in the world for objective, standardized, inter-resources, inter-period, and inter-jurisdictional comparisons. As stated earlier, due to its growing popularity a full discussion of this approach will serve CPUC's needs to be informed about why this method does not address its EM&V needs.

Index composition and decomposition is a methodology that has been used for over a century for separating an expenditure index into price and quantity indexes and for converting current prices into real or constant prices. However, its rigorous use in the area of energy consumption does not go back further than the 1980s. Boyd et al. (1987) is one of the first theoretically rigorous energy index decomposition analyses. In this study, the time periods covered are the pre-oil embargo years of 1967 – 1974 and the post oil-embargo years of 1974 – 1981, the cross sections are various levels of aggregation of industries in the manufacturing sector, and two energy intensity ratios are used, one consisting of electricity consumption in the numerator, the other of purchased fossil fuel. For both ratios, the denominator is real value added. The main purpose of the study was to measure structural shifts in the economy and to see whether or not shifts in energy intensive industries occurred after the 1974 oil embargo. The study explicitly avoids speculation as to the causes of the shift.

Judging by the volume of studies alone, index decomposition must be taken seriously as a top-down approach for studying energy consumption and energy efficiency trends. According to an extensive literature review by Ang and Zhang (2000) a total of 51 energy index studies were published between the 1970s and 1995, and another 73 were published between 1995 and 2000. In an updated count, Ang and Liu (2007) refers to a total of 172 energy index decomposition studies published between 1978 and 2003. Explaining the surge, the study notes:

> Arising from the Kyoto Protocol and the growing concern about world climate change and sustainable development, many countries have been taking steps to reduce greenhouse gas emissions. A common reduction strategy is through taking measures to increase energy efficiency. To evaluate performance, it is necessary to track energy efficiency change and to assess fulfillment of energy efficiency improvement targets on a regular basis and in a rigorous manner. Index decomposition analysis has been used in a number of countries and international organizations to serve this purpose. (p. 1431)

According to *Energy Economics* editors Tol and Weyant (2006), starting in 1988 eight out of ten of their journal's most cited papers are studies of energy index methods and their applications. Not only have papers on this subject been a staple in the International Association for Energy Economics publication *The Energy Journal,* but in 1997 another academic journal, *Energy Policy*, devoted an entire issue to energy and CO_2 emissions index-related papers (Schipper and Haas, 1997).

The contents of many of the index decomposition studies are index number theory and the advantages and disadvantages of using one version of an index formula over another. The two seminal works on which these theoretical studies are based are Fisher (1921) and Diewert (1976). At present, there is widespread agreement that several versions of index decomposition produce comparable results and satisfy all the necessary criteria for theoretical rigor and practical usefulness. For example, Boyd and Roop (2004) provide a mathematical proof that the Fisher Ideal index, like the Divisia index, produces an exact decomposition with no remaining residual term. Nevertheless, because of their spread internationally, controversies exist over the best way to standardize index decomposition efforts across nations, e. g., Ang et al. (2010) and Cahill and O'Gallachoir (2010). Many other decomposition studies are focused on specific industries or sectors of the economy, or specific countries, or other natural resources in addition to energy, e.g., Cornillie and Frankhauser (2004), Liao et al. (2007), and Vera and Langlois (2007).

Of relevance to this white paper, many index decomposition efforts attempt to directly link changes in the calculated energy efficiency index to the efficacy of energy efficiency policies. For example, Sun (1999) uses index decomposition to analyze CO_2 emissions from 1960 to 1995 for the 24 original OECD members, distinguishing between a total international GDP effect, a fuel switching effect, a national GDP share effect, and an energy intensity effect. Finding that energy intensity has mostly declined from 1973 to 1993, the study concludes, "This reveals that, after the first oil price shock, policy makers in OECD countries have included improving energy efficiency as part of their economic strategy, and succeeded in that strategy."

In 2001, a key recommendation of the U. S. National Energy Policy, was to "…improve the energy intensity of the U.S. economy as measured by the amount of energy required for each dollar of economic productivity" (NEPDG, 2001). In 2002, the Bush Administration called for an eighteen percent reduction in carbon intensity (carbon emissions per dollar GDP) by the end of the decade. These goals were couched in terms of indicators, i.e., energy and carbon intensity, not consumption levels. In support of these goals, EERE (2011) developed a website that came online in 2006 containing energy indicators and efficiency indexes for each of the major economic sectors of the U.S economy.

The linkage of index decomposition findings to energy efficiency policy was further tightened when an energy efficiency index approach, ODEX, produced by ODYSSEE, was positioned to become one of the EU's main methods for evaluating national energy efficiency progress in the member states. The status of this approach is documented in Thomas, et al. (2007), a report supporting the EU directive on energy end use efficiency and energy services (European Commission, 2006). According to Bosseboeuf, et al. (2005), ODEX was conceived, "in order to meet the political need for monitoring energy efficiency and to have an easily understandable, workable, and comparable indicator depicting the energy efficiency progress in EU member states."

Enthusiasm for energy efficiency indexes has been expressed by many leading scientific and economic international organizations. The International Atomic Energy Agency (IAEA), in collaboration with the United Nations, the International Energy Agency (IEA), Eurostat, and the European Environment Agency, proposed that energy indexes and related analyses be used for promoting sustainable development in emerging economies (IAEA, 2005). The World Bank and the IEA reinforced this recommendation by producing, through LBNL, a methodological guide to the construction of energy efficiency indicator geared to emerging economies (de la Rue du Can et al., 2010).

One of the attractions of index decomposition is that the denominators of most forms of energy intensity indexes, unlike the numerators, need not be of the same activity metric. For example, Metcalf (2008) creates national and state-level indexes using the four major economic sectors of the economy. In this study, the numerators of the energy intensity ratios for each sector are total primary energy use in Btu. However, the denominators for the commercial and industrial sectors are each sector's earnings by place of work, whereas the residential sector denominator is personal income and the transportation sector denominator is vehicle miles traveled. ODYSSEE (2011), EERE (2011), and Natural Resources Canada's (NRCAN) Office of Energy (2011) also produce energy indicators and energy intensity indexes for different sectors of the economy that are made up of sector-specific energy intensity ratios, each with a variety of different kinds of data in the numerators and denominators.

Another attraction is that an energy intensity index can be made up of an unlimited number of individual energy intensity ratios. For example, Huntington (2010) uses all 65 three-digit NAICS industries that together comprise the entire U.S. economy in a single index decomposition exercise. In this study, the numerators and denominators of each industry use the same data sources and the same metrics. By breaking the industries into different sub-categories, this study cautions that different levels of industry aggregation produce difference estimates of the effects of structural change on national energy expenditures.

Despite the attractions of energy indicators and their widespread use, warnings are prevalent of the dangers in using them for policy evaluation. Freeman et al. (1997) identify several problems that can arise in the construction of indexes in the industrial sector, including changes in findings that can occur when using alternative measures of output, or different levels of industry aggregation, or when price and quality changes are mismeasured. Golove and Shipper (1997) note that energy indexes have shortcomings as measurements of policy impacts and Boyd and Laitner (2001) caution that measured historical trends do not capture policy impacts. Thomas (2005) asserts that absent a hypothetical baseline, energy efficiency indicators cannot distinguish the impacts of energy efficiency policies from stochastic variations in weather, economic growth, and other related factors. Using ODEX as an example of an index decomposition method, Horowitz (2008) demonstrates via a Monte Carlo simulation that there is a low probability that energy efficiency indexes will produce similar energy efficiency policy evaluation findings to those produced by econometric energy demand modeling.

4.2 Fisher Ideal Index Decomposition

Although there is more than enough published evidence to show that energy indicators should be dismissed as inappropriate for the CPUC's EM&V needs, their prevalence demands that the CPUC be provided proof of their inability to shed light on policy-related impacts. To do so, the equations that decompose the Fisher Ideal index and lead to the construction of an index-based counterfactual are documented. Also, to help illustrate the meaning behind the equations, a simple, three-industry, two period index is built. To finalize the proof, the three-industry index is calculated for thirteen periods and decomposed into activity and efficiency effects. As should become clear by the end of this discussion, the efficiency effects are not specifically market effects or policy effects; indeed, there is no precise way to define them.

The Fisher Ideal index is widely used by U.S federal agencies. Although there are many different mathematical formula that can be used for index decomposition, such as the Divisia and Tornqvist, Diewert (1978) proved that certain types of indexes, among them the Fisher Ideal, have the same desirable properties. Each can be applied with relatively simple formulae, and for most practical purposes each produces energy index decomposition results that are similar. These techniques are not statistical in the usual sense. Unlike probabilistic statistics, energy index statistics do not admit to uncertainty. They take all their values to be non-random, perfectly accurate measurements.

As will be shown, energy indicators in their most complex form, energy intensity indexes, can only produce counterfactuals that differentiate between the energy intensity changes due to changes in activity levels, i.e. structural effects, and energy intensity changes *not due* to changes in activity levels. Decomposition can be seen as the ultimate stage of analysis after two or more energy intensity ratios are combined into an energy intensity index. Using manufacturing sector data as an example, the energy intensity for a given three-digit NAICS industry, i, in a given time period, t, can be defined as:

$$e_{it} = m_{it} / g_{it}$$

where m could be GWH consumption, energy expenditures, or one of many other variables related to an industry's energy consumption in a particular period, and g could be financial data such as industry value added (GDP) or value of shipments, or it can be physical data such as number of new units produced, shipped, sold, or installed.

Table 7 contains national values for three manufacturing industries, NAICS 331(food), NAICS 324 (chemicals), and NAICS 335 (computers). These were selected because, in terms of state GDP, they represent a large portion of California's manufacturing sector. The source of the GWH consumption data is the U.S. Census ASM and the source of the industry GDP data is the BEA national industry accounts. All dollar values are deflated using BEA's industry-specific implicit price deflator at base year 2005.

Table 7: Energy Indicator Data for Three National Industries (2008, 2009)

MANUFACTURING INDUSTRY	NAICS (i)	YR (t)	GWH (m)	Real GDP (bill.) (g)	INTENSITY (e)	ACTIVITY SHARE (s)
Food	311	2008	98,092	179	549.5	0.31
Chemicals	324	2008	50,675	116	437.8	0.20
Computers	335	2008	33,539	284	118.0	0.49
Food	311	2009	91,634	175	522.3	0.29
Chemicals	324	2009	48,083	128	375.0	0.21
Computers	334	2009	31,681	294	107.8	0.49

The most elementary analysis that can be done with these data is to calculate the change in each industry's GWH use. These calculations show an absolute decrease of about 6,500 GWH for NAICS 331, a decrease of about 2,600 GWH for NAICS 324, and a decrease of about 1,900 GWH for NAICS 334. On a percentage basis, these are declines in GWH consumption of 7 percent, 5 percent, and 6 percent, respectively. Although these estimates are *indicative* of changes in GWH, they leave many questions unanswered. One question that immediately arises is how these changes relate to business activity in each industry. The quickest way to address this question is to calculate the energy intensity for each industry. As shown in the energy intensity formula above and as reported in the next to last column in Table 7, in 2009 energy intensity decreased for all three industries. The energy intensities answer the question of how changes in individual industry activities relate to changes in individual industry energy use, but do not answer the broader question of how changes in overall business activity affected changes in overall energy use.

The final column of Table 7 shows the fraction of the business activity (GDP) that each industry is responsible for in this three-industry economy, with the fractions summing to unity for each year. This fraction is labeled *s* and is calculated as:

$$s_{it} = g_{it} / \sum_i g_{it}$$

Note that the activity shares of each industry hardly changed from 2008 to 2009. However, over time they are likely to have changed much more, just as putting more industries into this example would result in different activity shares.

Using the activity share information and the energy intensity data for these two years, an energy intensity *index* can be formed. The Fisher index, F^{Total} , is calculated by taking the earliest period as the base period, *0*, and the next adjacent period, *t*, such that:

$$F_t^{Total} = \sum_i e_{it} s_{it} / \sum_i e_{io} s_{io}$$

Based on these calculations and multiplying by 100, the Fisher energy intensity index for period *t* is interpretable as the percentage change in overall or total energy intensity from the base period *0*.

With nothing but these data only, the Fisher energy intensity index is decomposable into two separate indexes, an *activity* index and an *efficiency* index. This is done by first calculating Laspeyres and Paasche activity indexes, L^{act} and P^{act}, and then parallel efficiency indexes, L^{eff} and P^{eff}. The formulae for the efficiency indexes are:

$$L_t^{act} = \sum_i e_{i0} s_{it} / \sum_i e_{io} s_{io}$$
$$P_t^{act} = \sum_i e_{it} s_{it} / \sum_i e_{it} s_{i0}$$
$$L_t^{eff} = \sum_i e_{it} s_{i0} / \sum_i e_{io} s_{io}$$
$$P_t^{eff} = \sum_i e_{it} s_{it} / \sum_i e_{io} s_{it}$$

Similar to the formula for the calculation of F^{Total}, it is notable that each term involves multiplying each industry's energy intensity by either its own period, or the alternative period's, activity share. It is these activity shares, used as weights, that permits index decomposition to take place.

The Fisher activity index, F^{act}, is calculated as the geometric mean of L^{act} and P^{act}, and the Fisher efficiency index, F^{eff}, is calculated as the geometric mean of L^{eff} and P^{ef}:

$$F_t^{act} = \sqrt{L_t^{act} P_t^{act}}$$
$$F_t^{eff} = \sqrt{L_t^{eff} P_t^{eff}}$$

Multiplying the Fisher activity and efficiency indexes together produces the Fisher energy intensity index. This brings the calculations full circle; F^{Total} exactly decomposes and leaves no residual:

$$F_t^{Total} = \sqrt{L_t^{act} P_t^{act}} \times \sqrt{L_t^{eff} P_t^{eff}} = \sum_i e_{it} s_{it} / \sum_i e_{io} s_{io}$$

Finally, to produce a time series from each year-over-year index, each current year's index is chained to all the previous years' values by multiplying it by all the past period indexes. The chained Fisher efficiency index, where *0* is the reference year and *t-n* represents the next year, is calculated as:

$$F_t^{chained-eff} = F_{t,t-1}^{eff} \times F_{t-1,t-2}^{eff} \times \ \ldots \ \times F_{t-n,0}^{eff}$$

This produces a rolling rather than fixed base because the industry weights in the time series move as the mix of industries move from one period to the next. Rebasing the chained indexes to any reference year is straightforward, as are interpretations of the percentage changes in the indexes from any one period to any other.

Using the index decomposition methodology, the counterfactual in any period is defined as the energy consumption that would have occurred in time *t* had there been no change in energy intensity from time *0* to time *t*. This index-based counterfactual, C^{Index}, is calculated as:

$$C_t^{Index} = \sum g_{it} \times \frac{\sum m_{io}}{\sum g_{io}}$$

which shows, for example, that to solve for the aggregate energy use that would have occurred in 2009 had energy intensity remained constant, aggregate energy intensity in 2008 is multiplied by aggregate activity in 2009. It also clearly shows that unlike the energy demand model counterfactual that can include any number of control for any number of variables, the index counterfactual is based on a function whose only inputs are an energy-related variable and an activity related variable:

$$C_t^{Index} = f(m_{it}, g_{it})$$

Using the index counterfactual, the total change in aggregate energy use due to changes in energy intensity is calculated as:

$$\text{Total Change in Energy Use}_t = C_t^{Index} - \sum m_{it}$$

However, this counterfactual, and hence the total change in energy use, contains the effects of changes both in activity and in efficiency. To separate them, following Metcalf (2008) the portions of the changes in total energy use that are attributable to activity and efficiency can be derived as:

$$\text{Attribution to Activity} = \left(\frac{\ln(F_t^{act})}{\ln(F_t^{Total})} \right)$$

$$\text{Attribution to Efficiency} = \left(\frac{\ln(F_t^{eff})}{\ln(F_t^{Total})} \right)$$

Then, efficiency impacts, i.e., non-activity related impacts, can be calculated by subtracting actual aggregate energy use from the counterfactual and multiplying by the percent of the energy intensity index that is attributable to efficiency:

$$\text{Efficiency Impact}_t = (C_t^{Index} - \sum m_{it}) \times \left(\frac{\ln(F_t^{eff})}{\ln(F_t^{Total})} \right)$$

Using the three-industry economy example and the data from all the years since 1997, the findings of all three indexes are displayed in Table 8. As can be seen, in aggregate there was a 7.4 percent decrease in overall activity from 1997 to 2009 and controlling for activity there was a 42.5 percent decrease in energy intensity, i.e., an efficiency index value of 57.5 percent. Translated into GWH, the total change in energy intensity results in an aggregate decrease of 150,620 GWH for this economy in 2009, with approximately 12 percent of that amount due to decreased activity and the remaining 88 percent, or 132,133 GWH, due to all other factors that are collectively referred to as *efficiency*.

Table 8: Index Decomposition Findings, Three-Industry Economy

2009 Index	**Fisher**
Activity	92.6
Efficiency	57.5
Total Intensity	53.2
Attribution	**GWH**
2009 Actual	171,398
2009 Counterfactual	322,018
Total Difference	150,620
Attribution to Activity	18,487
Attribution to Efficiency	132,133
% Efficiency Attribution	88%

Figure 3 displays all three indexes graphically, beginning with 1997 and ending in 2009. In Figure 4, the changes in GWH due to efficiency and activity are displayed. These estimates are based on the fraction of the differences between the counterfactual and actual aggregate GWH each year that is attributable to these two factors based on their respective indexes.

Several features of the index decomposition methodology are now worth restating. First, as the method is non-stochastic, there are no confidence intervals associated with any of these statistics. Second, as all the findings are derived from pairing a single variable representing activity with energy consumption, efficiency-attributed changes are nothing more than what remains after activity-attributed effects are removed. Third, all of the efficiency-attributed changes could be due to market factors (excluding activity levels) and none to public policies, or vice versa. Fourth, because this method cannot be pushed further, only in conjunction with econometric modeling can the energy intensities be used to disentangle policy impacts from market impacts. In conclusion, if the CPUC were to encourage the use of energy indicators it would be undermine the importance of energy efficiency policy EM&V, whose very purpose is to disentangle the impacts of public policies on energy consumption from the impacts of market factors on energy consumption.

Figure 3: Energy Indexes: Three-Industry Manufacturing Sector (1997-2009)

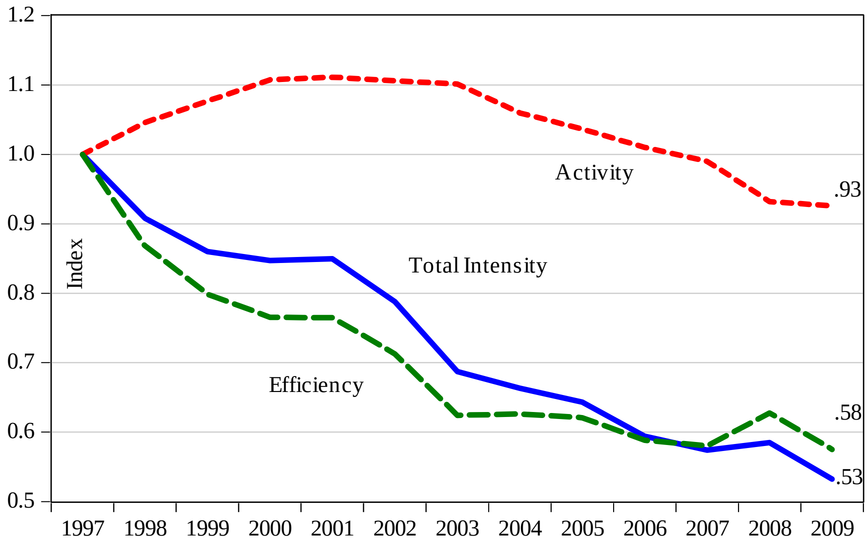
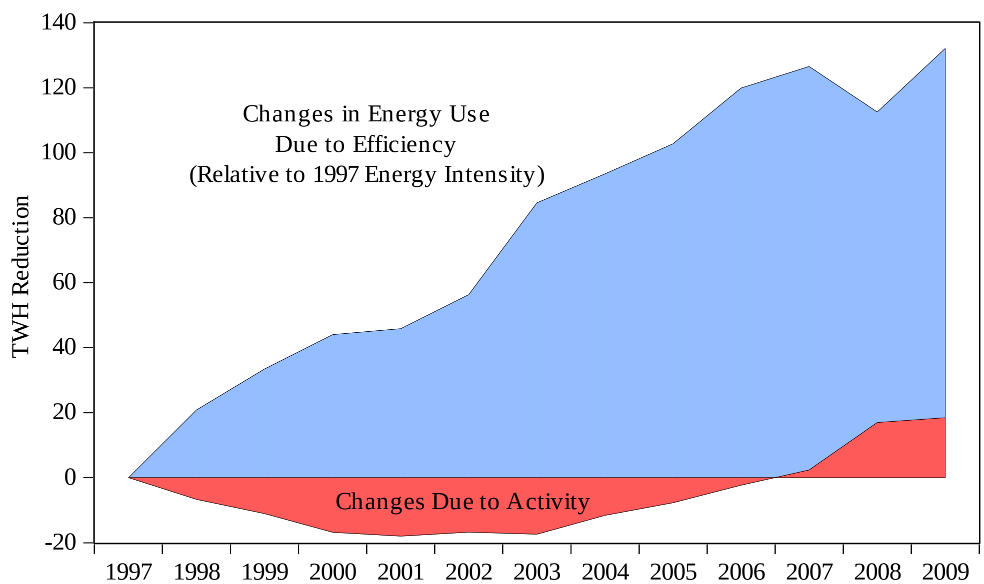


Figure 4: Changes in Energy Use: Three-Industry Manufacturing Sector (1997-2009)



5. MCM Pilot Project Recommendations

The literature reviews above were intended to provide historical and current information on the two top-down methodologies. As has been described, both approaches have been used in recent years for studying aggregate energy consumption. Furthermore, each has been applied to many different kinds of data and many different situations. The technical discussions that followed each literature review then addressed the extent to which each method can help address the CPUC's needs by describing the character of each method's policy counterfactuals are how they are produced. Based on the CPUC's regulatory environment and its stated EM&V needs for the 2013-2015 energy efficiency program cycle, *Demand Research* offers the following recommendations for the MCM pilot project:

1) the CPUC should use the MCM pilot project to develop an MCM policy evaluation database that offers geographic area granularity down to the census tract and five-digit zip code levels for the residential, commercial, industrial/agricultural sectors; and,
2) the CPUC should use the MCM pilot project to conduct one or more policy evaluations using econometric energy demand models.

In the proposal that follows, *Demand Research* describes all the steps that will be undertaken to implement these two recommendation.

6. References

Ang, B. W. and Liu, Na (2007). "Energy Decomposition Analysis: IEA Model Versus Other Methods." *Energy Policy*, Vol. 35, No. 3, pp. 1426-1432

Ang, B. W., Mu, A. R., and P. Zhou (2010). "Accounting Framework for Tracking Energy Efficiency Trends." *Energy Economics*, Vol. 32, pp. 1209-1219.

Ang, B. W. and Zhang, F. Q. (2000). "A Survey of Index Decomposition Analysis in Energy and Environmental Studies." *Energy Economics*, 25, pp. 1149-1176.

Arimura, Toshi H., Richard G. Newell, and Karen Palmer (2009). *Cost-effectiveness of Electricity Energy Efficiency Programs.* Resources for the Future. Discussion Paper.

Auffhammer, Maximilian, Carl Blumstein, and Meredith Fowlie (2008). "Demand Side Management and Energy Efficiency Revisited." *The Energy Journal* (29) 3:91-104.

Balestra, Pietro and Marc Nerlove (1966). "Pooling Cross Section and Time Series Data in the Estimation of a Dynamic Model: The Demand for Natural Gas." *Econometrica*, Vo. 34, No. 3, pp. 585-611.

BEA (U.S. Bureau of Economic Analysis) (2011). U. S. Department of Commerce, Regional Economic Accounts, http://www.bea.gov/regional.

Beierlein, James G., James W. Dunn, and James C. McConnon, Jr. (1980). "The Demand for Electricity and Natural Gas in the Northeastern United States." *The Review of Economics and Statistics*, Vol. 62, pp. 403-408.

Berndt, Ernst R. (1984). "Modeling the Aggregate Demand for Electricity: Simplicity vs. Virtuosity," in *Advances in the Economics of Energy and Resources,* Vol.5, pp. 141-152, JAI Press.

Berndt, Ernst R. (1996). "The Demand for Electricity: Structural and Time Series Approaches," in *The Practice of Econometrics,* Addison-Wesley Publishing Company, MA.

Bernstein, Mark, K., Kateryna Fonkych, Sam Loeb, and David Loughran (2003). *State-Level Changes in Energy Intensity and Their National Implications*, Rand, MR-1616-DOE, Santa Monica, CA.

Blattenberger, Gail R., Lester D. Taylor, and Robert Rennback (1983). "Natural Gas Availability and the Residential Demand for Energy." *The Energy Journal*, Vol. 4, No. 1, pp. 23-45.

Bohi, Douglas R. (1981). *Analyzing Demand Behavior*. Johns Hopkins University Press for Resources for the Future, Baltimore, MD.

Bohi, Douglas R. and Mary Beth Zimmerman (1984). "An Update on Econometric Studies of Energy Demand Behavior." *Annual Review of Energy*, Vol. 9, pp. 105-154.

Bosseboeuf, Didier, B. Lapillonne, and W. Eichhammer (2005). "Measuring Energy Efficiency Progress in the EU: The Energy Efficiency Index ODEX." *Proceedings of the 2005 European Council for an Energy Efficient Economy Summer Study,* Stockholm, Sweden, pp. 1127-1135.

Boyd, Gale, J. F. McDonald, M. Ross, and D. A. Hanson (1987). "Separating the Changing Composition of U.S. Manufacturing Production for Energy Efficiency Improvements: A Divisia Index Approach." *The Energy Journal*, Vol. 8, No. 2, pp. 77-95.

Boyd, Gale and Laitner, John A. "Skip" (2001). "Recent Trends in the U.S. Energy Intensity: An Index Number Analysis." *International Association for Energy Economics Newsletter*, Volume 10, 2nd quarter, pp. 4-9.

Boyd, Gale, and Joseph M. Roop, (2004). "A Note on the Fisher Ideal Index Decomposition for Structural Change in Energy Intensity." *The Energy Journal*, Vol. 25, No. 1, pp. 87-101.

Cahill, Caiman J. and Brian P. O'Gallachoir (2010). "Monitoring Energy Efficiency Trends in European Industry: Which Top-Down Method Should Be Used?" *Energy Policy*, Vol. 38, pp. 6910-6918.

Considine, Timothy J. (2000). "The Impacts of Weather Variations on Energy Demand and Carbon Emissions." *Resource and Energy Economics*, Vol. 22, pp. 295-314.

Cornillie, Jan and Frankhauser, Samuel (2004). "The Energy Intensity of Transition Countries." *Energy Economics*, 26, pp. 283-295.

CPUC (California Public Utilities Commission) (2010-A). *Decision on Evaluation, Measurement, and Verification of California Utility Energy Efficiency Programs*. Decision 10-10-33 (Rulemaking 09-11-014). October, San Francisco, CA.

CPUC (California Public Utilities Commission) (2010-B). *2006-2008 Energy Efficiency Program Evaluation Report.* Prepared by the Energy Division. July, San Francisco, CA.

Dahl, Carol (1993). *A Survey of Energy Demand Elasticities in Support of the Development of the NEMS*. Washington, DC: U.S. Department of Energy.

de la Rue du Can, Stephane, Jayant Sathaye, Lynn Price, and Michael McNeil, (2010). *Energy Efficiency Indicators Methodology Booklet*, Prepared for the World Bank and the International Energy Agency, LBNL under contract No. DE-AC02-05CH11231

Diewert, W. E. (1976). "Exact and Superlative Index Numbers." *Journal of Econometrics*, Vol. 4, No. 2, pp. 115-145.

EERE (U. S. Department of Energy, Energy Efficiency and Renewable Energy) (2011). *Energy Intensity Indicator Website*. http://www1.eere.energy.gov/ba/pba/ intensityindicators

EMEEES (2008). *Harmonised Methods for Evaluating Energy End-Use Efficiency and Energy Services – The European Commission's project on evaluation and monitoring for the end-use directive on energy end use and energy services, Final Conference*, Brussels, BE.

European Commission (2006). Directive 2006/32/EC of the European Parliament and of the Council of 5 April 2006 on energy end use efficiency and energy services (ESD), Official Journal of the European Union 27.4.2006, L 114/64.

Fisher, Irving (1921). "The Best Form of Index Number." *Quarterly Publications of the American Statistical Association*, Nov. 17, No.133, pp.53-537.

Fisher, Franklin M. and Carl Kaysen (1962). *A Study in Econometrics: The Demand for Electricity in the United States.* Amsterdam, North Holland Publishing Co.

Fisher, Franklin M. (1999). *Microeconomics: Essays in Theory and Applications*. Cambridge University Press. Edited by Maarten-Pieter Schinkel. Cambridge, United Kingdom.

Freeman, Scott L., Mark J. Niefer, and Joseph M. Roop (1997). "Measuring Industrial Energy Intensity." *Energy Policy*, Vol. 25, Issues (7-9), pp. 703-714.

Golove, William and Schipper, Lee (1997). "Restraining Carbon Emissions: Measuring Energy Use and Efficiency in the USA." *Energy Policy*, Vol. 25, Issues (7-9), pp. 803-812.

Halvorsen, Robert (1975). "Demand for Electric Energy in the United States." *Southern Economic Journal*, Vol. 42, No. 4, pp. 610-625.

Hartman, Raymond S. (1979). "Frontiers in Energy Demand Modeling." *Annual Review of Energy*, Vol. 4, pp. 433-466.

Hausman, Jerry (2001). "Mismeasured Variables in Econometric Analysis: Problems from the Right and Problems from the Left." *The Journal of Economic Perspectives*, Vol. 15, No. 4, pp. 57-67.

Horowitz, Marvin J. (2004). " Electricity Intensity in the Commercial Sector: Market and Public Program Effects." *The Energy Journal*, Vol. 25, No. 2, pp. 115-137.

Horowitz, Marvin J. (2007). "Changes in Electricity Demand in the United States from the 1970s to 2003." *The Energy Journal*, Vol. 28, No. 3, pp. 93-119.

Horowitz, Marvin J. (2008). "The Trouble with Energy Intensity Indexes: La aritmetic non e opinione." *Energy Efficiency*, Vol. 1, No. 3, pp. 199-210, August 2008.

Horowitz, Marvin J. (2011). "Measuring the Savings from Energy Efficiency Policies: A Step Beyond Program Evaluation." *Energy Efficiency*, Vol.4, No.1, pp. 43-66.

Houthakker, Henrik S. and Lester D. Taylor (1970). *Consumer Demand in the United States*. 2nd Ed. Cambridge: Harvard University Press.

Huntington, Hilliard G. (2010). "Structural Change and U. S. Energy Use: Recent Patterns." *The Energy Journal*, Vol. 31, No. 3, pp. 25-39.

IAEA (International Atomic Energy Agency) (2005). *Energy Indicators for Sustainable Development: Guidelines and Methodology.* Vienna, Austria.

Kamerschen, David R. and David V. Porter (2004). "The Demand for Residential, Industrial, and Total Electricity, 1973-1998." *Energy Economics*, Vol. 26, pp. 87-100.

Kavalec, Chris (2011). "Presentation to the 2011 IEPR Staff Workshop on Historical Energy Efficiency Estimate and Update to the 2009 California Energy Demand Forecast." California Energy Commission.

Kneese, Allan V. and James B. Sweeney (1993). *Handbook of Natural Resources and Energy Economics*. Elsevier, Ltd., Amsterdam, The Netherlands.

Liao, Hua, Ying Fan, and Yi-Ming Wei (2007). "What Induced China's energy Intensity to Fluctuate: 1997-2006." *Energy Policy*, Vol. 35, pp. 4640-4649.

Loughran, David S. and Jonathan Kulick (2004). "Demand Side Management and Energy Efficiency in the United States." *The Energy Journal*, Vol. 25, No. 1, pp. 19-43.

Metcalf, Gilbert (2008). "An Empirical Analysis of Energy Intensity and Its Determinants as the State Level." *The Energy Journal*, Vol.29, No. 3, pp 1 – 27.

Murray, Michael P., R. Spann, L. Pulley, and E. Beauvais (1977). "The Demand for Electricity in Virginia." *The Review of Economics and Statistics*, Vol. 59, No. 5, pp. 585-560.

Nelson, Jon P. (1975). "The Demand for Space Heating Energy." *The Review of Economics and Statistics*, Vol. 57, No. 5, pp. 508-512

NEPDG (National Energy Policy Development Group) (2001). *National Energy Policy: Reliable, Affordable, and Environmentally Sound Energy for America's Future.* Washington D.C.ODYSSEE (2011). http://www.odyssee-indicators.org/Indicators

NRCAN (Natural Resources Canada) Office of Energy (2011). *Energy Efficiency Trends Analysis Tables.* http://www.oee.nrcan.gc.ca/corporate/statistics/neud

Parfomak, Paul W. and Lester B. Lave (1996). "How Many Kilowatts are in a Negawatt? Verifying *Ex Post* Estimates of Utility Conservation Impacts at the Regional Level." *The Energy Journal*, Vol. 17, No. 4. pp. 59-87.

Schipper, Lee and Fridtjof Unander, Scott Murtishaw, and Mike Ting (2001). "Indicators of Energy Use and Carbon Emissions: Explaining the Energy Economy link." *Annual Review of Energy and the Environment*, Vol. 26, pp. 49-81.

Schipper, Lee and Haas, Reinhard (editors) (1997). "Cross-country comparisons of indicators of energy use, energy efficiency and CO_2 emissions." *Energy Policy*, Vol. 25, Issues (7-9), pp. 639-831.

Sheppard, Colin, Charles Chamberlin and Arne Jacobson (2010). *Exploring Strategies for Implementing a Performance based State Efficiency Program: State Energy Consumption Metrics – Residential Sector Analyses*. Final Report. Natural Resources Defense Council and Schatz Energy Research Center, Humboldt State University.

Sun J. W. (1999). "Decomposition of Aggregate CO_2 Emissions in the OECD: 1960-95." *The Energy Journal*, Vo. 20, No. 3, 1999 pp. 147-156.

Taylor, Lester D. (1975). "The Demand for Electricity: A Survey." *Bell Journal of Economics and Management Science,* 6(1): pp. 74-110

Taylor, Lester D., Blattenberger, G. R., and R. K. Rennback (1984). "Residential Energy Demand in the United States: Introduction and Overview of Alternative Models," in *Advances in the Economics of Energy and Resources*, Vol. 5, pp. 85-105.

The Cadmus Group, Itron Inc., Jai J. Mitchell Analytics, Kema, PA Consulting Group, Summit Blue, LLC (2010). *Residential Retrofit High Impact Measure Evaluation* . Prepared for the California Public Utilities Commission Energy Division, San Francisco, CA.

Thomas, Stephan (2005). "The Quantitative Targets for Energy Savings Proposed by the Draft Directive on Energy End-Use Efficiency and Energy Services – How to Define, How to Monitor Them." Wuppertal Institute for Climate, Environment, and Energy, DE.

Tol, Richard S. J. and Weyant, John P. (2006). "Energy Economics' Most Influential Papers." *Energy Economics,* 28, pp. 405-409.

Vera, Ivan and Lucille Langlois (2007). "Energy Indicators for Sustainable Development." *Energy*, Vol. 32, pp. 875-882.

Watkins, G. C. (1992). "The Economic Analysis of Energy Demand: Perspectives of a Practitioner," in *Energy Demand: Evidence and Expectations*, edited by David Hawdon, Surrey University Press, London

7. Demand Research's MCM Pilot Project Proposal

Demand Research's value proposition to the CPUC is an updatable, permanent, state-of-the-art, geographic database that permits the CPUC to monitor the long-term impacts of energy efficiency policies on state and local energy consumption and environmental protection, and, an MCM policy evaluation system that can lead to substantial reductions in the size, and cost, of the CPUC's bottom-up EM&V activities. *Demand Research's* MCM pilot project proposal offers the following features:

- use of public access data that come from state, local, and federal agencies and are free to the public, are geographically-based, and can be divided as finely as the the census tract or zip code levels;
- compilation, merging, and joining of all the selected datasets in a single database platform that permits standardized data manipulation and variable creation; and,
- reduced form econometric models of energy demand using either time series or panel data; each demand model is for a single fuel, and a single customer type, market segment, or economic sector.

As described below, the scope of *Demand Research's* MCM pilot project proposal is ambitious. However, *Demand Research* believes that only a project of this breadth and depth is capable of demonstrating the practical value of MCM policy evaluations to the CPUC. Moreover, there are critical elements of this pilot project that make it realistic to think that it can be successfully completed, on time and on budget. For one, once the residential sector MCM database is created, the marginal effort in time and cost to expand the MCM database to include data for the commercial, industrial, and agricultural sectors is relatively modest. This is because similar datasets are involved in all sectors, and similar data processing routines are needed for many of the variables across sectors. Another element that is in this project's favor is that once the MCM database is in place there are declining marginal cost and time requirements for performing more than one policy evaluation. It is due to this phased cost/effort structure that *Demand Research* believes that this pilot project's scope is both reasonable and achievable.

Also, because there are declining marginal costs for conducting MCM evaluations, *Demand Research* refers to the final product of this pilot project as an *MCM policy evaluation system.* If the CPUC assesses, based on this pilot project, that MCM studies should be a component of its next EM&V cycle, a database system and modeling protocol will be in place for conducting multiple, inexpensive policy evaluations simultaneously. These policy evaluations can address a variety of subjects and can be focused on small market segments and small areas as well as on broad economic sectors and whole utility service territories.

7.1 Build Geographically-Granular Database

The only way that the recommended MCM energy demand modeling approach can be fully developed is by creating a new and unique database for California. This database will be designed so that numerous policy evaluations can be worked-on and updated every year for the foreseeable future. Creating such a database does not cost very much more than if serial policy evaluation projects were funded by the CPUC because every discreet study requires a significant amount of effort simply to process datasets and create variables. Going through these steps piecemeal is cost inefficient and time inefficient, forcing policy evaluators to reinvent the wheel every time a new project is undertaken. Also, because data cleaning and processing for serial policy evaluations is often done by different individuals on a first-time only-time basis, individual study-specific databases can be prone to unintentional errors and inadequate documentation.

The idea behind *Demand Research's* energy demand modeling database is that it is a permanent EM&V resource that will be used by the CPUC for the foreseeable future to study the impacts of energy efficiency policies and programs on energy consumption and environmental protection in the state of California. Once the investment is made to build this unique database, MCM evaluations will be capable of being performed for each sector of the economy and each identifiable customer segment at the state level, the utility service territory level, and various regional and local levels. Moreover, it will allow the CPUC to undertake multiple policy evaluations simultaneously and to continually update permanent policy-related evaluations at low marginal costs. *Demand Research's* database will also be able to translate energy efficiency policy impacts into environmental accomplishments. This can be done with built-in database routines as uncomplicated as applying geographically-anchored environmental impact constants to the estimated energy impacts, or by complex algorithms and environmental models.

- **Datasets**

For the reasons described in the technical discussions, very little can be learned about policy impacts from energy consumption data unless they are tied to geography-specific time series and/or cross section economic, social, demographic, and weather data. Datasets with all this information exists; however, they come from many different government agencies. Table 9 contains a list of the main federal agency datasets that contain information for California or contain national-level information that could be useful for California-based policy evaluations. At present it appears that most if not all State of California departments and agencies draw from these federal datasets for acquiring historical economic and demographic data about the state.

There are additional datasets that are collected in California and that are specific to California consumers and businesses that may also be useful for the MCM pilot project. For example, the DEER database may come in useful, as might RASS database. Data from prior program evaluations are necessary, too. All relevant datasets will be looked into on a case-by-case basis to see if they can be of use for MCM policy evaluations. Table 9 provides a listing a publically-accessible selected datasets from the federal government and the state of California. As the MCM pilot project proceeds it is hoped that additional useful datasets that are free to the public may be discovered.

Table 9: Selected Federal and California and Datasets for MCM Pilot Project

Federal Data Sources	**Dataset**
Energy Information Administration (U.S. DOE)	State energy data, all fuels, by sector Utility electricity sales, by sector
Bureau of Economic Analysis (U.S. Dept. of Commerce)	REIS state/country accounts Industry GDP by state GDP, quantity indexes, price deflators
Census Bureau (U.S. Dept. of Commerce)	Annual Survey of Manufacturers Annual Community Survey Decennial population and housing census Quinquennial economic survey
Nat. Climatic Data Center (U.S. Dept. of Commerce)	Local HDD and CDD
Bureau of Labor Statistics (U.S. Dept. of Labor)	Consumer/producer price indexes Local area unemployment
Economic Research Service (U.S. Dept. of Agriculture)	Farm characteristics
Federal Reserve Board	National financial and industry time trends
California Data Sources	
California Energy Commissions (CEC)	Energy Consumption Database (ECDMS) Databases used for load forecasting/ special studies Residential Appliance Saturation Survey DEER database
California Public Utilities Commission (CPUC)	Documents and records (e.g., EEGA)
Other state and local agencies and organizations	Past and present EM&V studies Additional social, economic, and demographic data

- **Database and Analysis Platforms**

There are two software products that are needed to create the MCM policy evaluation system, a database platform and a statistical analysis platform. The database platform must be capable of housing and updating many different datasets and of expanding as more datasets are found or become available. It must also be capable of easily merging, concatenating, and otherwise joining the observations from different datasets without having to go directly to each separate raw dataset. The ease with which this can be done is critical to the database because datasets from different organizations, and even from the same organization but different divisions, can be organized and formatted in dissimilar ways. In addition to geographic differences, data frequencies differ. Some data are produced monthly, while others are produced quarterly, annually, or over multiple years.

To accomplish all of the anticipated database tasks as well as to work in a platform that has the potential to handle any new tasks that may arise, a few of the features that the data platform needs are:

- the ability to add new datasets, additional years of data, and new data sources;
- precisely defined and documented data storage;
- a user-friendly shell that enables data to be accessed efficiently and that can be updated as datasets grow;
- the ability to perform calculations and store the routines so that they do not have to be re-written each time calculations need to be redone;
- the ability to produce publication-quality tables and graphics; and,
- the ability to output customized datasets to statistical platforms.

This is not a complete list of all the tasks the data platform will need to perform but it does provides a sense of its necessary capabilities. Also, it is meant to show why spreadsheet programs like Excel and simple relational databases such as Access are not suitable for the kind of database that is being proposed for the MCM pilot project.

It will remain for the MCM pilot project staff to develop this database resource to its fullest potential. The ultimate goal will be to deliver a well-stocked database platform that is capable of fast, efficient, and reliable use. This will allow energy efficiency policy research projects to focus on analytical issues and to avoid spending much time repeating data collection, cleaning, and processing tasks. At present it appears that licensing IBM's SPSS is likely to be the most cost-effective and appropriate product for the pilot project.

A statistical analysis platform must be chosen for doing the actual policy evaluation research. Choosing this platform is independent of the choice of database platform and will largely be left up the pilot project researchers. SPSS, Stata, SAS, EViews and many other software products all have the statistical analysis capabilities and modules for running econometric models. A major consideration in selecting and using one or more statistical platforms will not only be the comfort levels of the analysts, but the fees or licensing costs associated with using each of them.

- **Joining, Merging, Attaching, Aggregating, Disaggregating**

Before MCM policy evaluations can begin, all of the disparate datasets must be joined geographically and temporally so that analytical datasets can be created. This means that the database platform must have an interface that allows users to specify and download variables that are available at the census tract level (n~7,000), the zip code (n~6,000), the county level (n=58), the utility service territory level (n>50), and the state level. For load forecasting purposes the CEC also defines 16 climate zones and 8 planning zones. Given their boundaries (e.g., counties or census tracts), the MCM database can also be designed to output variables for these geographic entities.

Geographically overlaying all of the datasets is done by linking locations. Each county has identifiable census tracts, zip codes and zctas (Census-developed links between zip codes and census tracts). Once overlaid, weighting schemes must be created for transferring economic and demographic variables between boundaries.

Weather station data can be linked in a variety of way. However, since the intention of the MCM pilot project is to get down to the census tract or zip code levels, local weather stations will be assigned to zip codes based on the area surrounding their geographical coordinates, that is, latitude and longitude. Although there is no publically-available list of zip codes mapped to latitude and longitude, this information is available from commercial data services such as *ZIPCodeWorld*. Matching weather stations to zip codes can then be done on the basis of proximity using various mathematical criteria such as the great-circle formula. Given the complexity of California's geography, this process will not be fully automated. Some manual inspection of the weather station assignments will be necessary in zip codes suspected of having have multiple micro-climates.

Joining zip codes to individual utility service territories can also follow a similar approach. Vendors such as *Ventyx* produce lists of the zip codes within each utility service territory. In addition to zip code-to-county linkages, *Ventyx* also estimates the percent of a zip code's area that overlaps into a different service territory. For merging billing data, the utilities will be given exact information on the needed zip codes and census tracts.

Once the MCM database is programmed with all the geographical joining points, upward aggregation requires arithmetic procedures. For example, census tract populations can be summed to produce population estimates at the county level, the utility service territory level, or the state level. However, not all variables can be summed this way from lower to higher levels of aggregation. Census tract heating degree days is a good example. If the research involves the residential sector, then census tract heating degree days will be weighted by census tract population, and then these products summed to create population-weighted heating degree days at the county level. If the heating degree days throughout the country do not vary, then the population-weighted heating degree days will be identical to average heating degree days. However, since California has many micro-climates, population-weighted heating degree days is likely to be more representative of country-level heating requirements than the unweighted average.

More challenging than upward aggregation is downward disaggregation. Suppose annual population data is only available at the county level and that census tract population estimates are desired. Here, an allocation method must devised. A simple method would be to take the census tract populations for 1990, 2000, and 2010 from the decennial censuses, and then to impute census tract populations for the intervening years based on proportional changes between the decades. When there are more complex variables and proportional allocation may not be appropriate, other imputation techniques may be called for. These could involve the use of linear or non-linear regression models or trending techniques, or even expert judgment based on external evidence. For example, county manufacturing employment may have grown exponentially from one decade to the next due to the opening of car assembly plant in a particular zip code. In this case, proportional allocation would fail to capture this phenomena accurately, and a better method would be to assign the appropriate amount of employment to the specific census tract in the specific year in which the plant began operating.

In the end, upward and downward allocation methods will have to be determined on a case-by-case bases depending on the variable in question. All allocation methods with be fully justified and documented as part of the completed MCM database. Despite the potential inaccuracies, imputations are commonplace in scientific studies and yield valuable research data.

A final, related data-joining issue is that of intra-year data. Some variables are only available for periods shorter than a year. For example, heating degree days may only be available in monthly format. For these higher frequency data, algorithms will be put in place in the database to aggregate them to the quarterly, seasonal, or annual frequencies, as needed.

7.2 Process Utility Billing Data and Other Energy Sales Datasets

The prerequisite for the MCM energy demand modeling system is a database whose foundation is, first and foremost, energy consumption data. The most readily available energy consumption data is utility-level sales data for the major sectors of the economy. One source for these data is the CEC's Energy Consumption Database (ECDMS). This dataset contains energy sales data from all utilities in California, public and private, going back to 1990.

Another source for utility-level energy sales data is the Energy Information Administration (EIA). The data referred to as EIA-861 also goes back to 1990 and covers the major economic sectors. However, it differs from ECDMS data in several ways that will need to be carefully assessed. For example, EIA-861 and ECDMS define the commercial and industrial sectors by different SIC/NAICS codes.

Other than these sector level, utility-level energy consumption datasets, the utilities themselves must be relied on to deliver geographic-level and customer segment-level billing data to the MCM pilot project. This is clearly recognized by D.10-10-033 in which it is ordered that:

> Pacific Gas and Electric Company, Southern California Edison Company, San Diego Gas & Electric Company, and Southern California Gas Company shall cooperate fully with Energy Division's efforts to expedite the Total Energy Consumption Pilot and shall timely (sic) provide any energy usage data Energy Division deems necessary. (p.45)

As such, once one or more of the topics of individual MCM pilot project policy evaluation studies are decided on, the billing data that fit the needs of the evaluation(s) will be specified. At present it is assumed that all of the utilities will provide the specified billing data within 30 days of a request. For planning and costing purposes, it is assumed that the utilities will only provide the billing data at the individual account level, not at the level of aggregation that the policy evaluations may need. The utilities will be requested to provide the account number, rate class, billing dates and respective energy consumption, and census tract and zip code locations for all customers in the policy evaluation research design. Also, for now it is assumed that the utilities can provide these data for at least 10 years back, that is, from 2002 to the present. However, there may be policy evaluation topics for which less historical data are needed and others for which more historical data are need. Issues related to historical data will have to be addressed on a case-by-case basis for each MCM study.

Aggregating the energy sales data from individual accounts involves careful processing. *Demand Research* will aggregate the billing data for all customers by calendar month using a standardized procedure for start-of-month and end-of-month allocations, performing and documenting quality checks to ensure that outliers and anomalies are not present and that the final datasets are as planned. *Demand Research* recognizes that utility customer rate class identifications, and NAICS classifications, may be faulty. All such data issues will be documented, and where possible fixes to the data issues will be sought.

7.3 Select MCM Evaluation Topics

As the technical discussions have suggested, given the geographic database that will be built, there are a many ways in which energy demand models can be focused on estimating the impacts of energy efficiency policies on aggregate energy consumption. For the MCM pilot project it will be necessary to limit the number of studies that can be undertaken. As such, priority will be given to those topics that are of strategic interest to policymakers, such as the impacts of residential lighting programs and residential building codes and standards. However, priority must also be given to those topics that have the best chance of being completed within the project timeframe.

Selecting the best topics to be studied in the MCM pilot project will depend on several practical factors. First, it will be necessary to assess the progress of the MCM database to determine what data are available for use and in what timeframe. A number of criteria can then guide the selection process. These may include:

- the ability of the topic to address high-level, strategic policy goals
- the value of the topic in helping the CPUC plan for the next cycle of EM&V studies
- the clarity with which the topic can investigate individual energy efficiency programs
- the clarity with which the topic can investigate policy impacts in specific locations
- the generalization of the research design/model to different economic sectors
- the specificity of the research design/model to an individual sector
- the probability of unambiguous and clear findings emerging from the study

- the type of methodological lessons that can be learned from the topic
- the level of expected confidence in the results that the topic may yield.

Once a short list of evaluation topics is compiled, preliminary analyses will be done to determine which topics are viable and which will need to be put off for the future. At present it is assumed that up to three topics will be chosen for study.

7.4 Compile Energy Efficiency Program/Policy Histories

MCM policy evaluation studies cannot be planned and executed without good information on where and when programs have been operating in California. This is recognized in D.10-10-033:

> Key activities to be initiated as part of the near-term pilot (and, where successful, incorporated into any permanent EM&V activities) include compilation of a historical record of program impacts, exploration of methodologies for developing enhanced saturation studies that leverage advanced metering data with onsite information, and longitudinal analyses to study energy use and energy efficiency in buildings over time. (p. 33)

Demand Research's focus will be on compiling an historical record for those programs and policies that are relevant to the policy evaluation topics selected for the MCM pilot project. These histories serve several indispensible purpose:

a) they are essential for specifying the optimal research design, regression model, and counterfactual for an MCM evaluation;
b) they guide the selection or construction of target variables and their integration into the evaluation;
c) they permit benchmarking of MCM findings with bottom-up findings (gross, adjusted-gross, net, etc.) and CEC load forecasts;
d) they guide program, or measure, attribution analyses; and
e) they are essential for interpreting the statistical findings that emerge from the MCM energy demand models.

There are several sources of information for compiling a region and sub-region-based history of major California energy efficiency programs. In addition to the fact that there may be one or more studies that have already compiled much of this information in stand-alone reports, there is the CALMAC library and other records and documents of energy efficiency programs residing at the CPUC and the CEC. The utilities, too, may have reports and documents that are specific to their historical programs. The primary information that will be sought will be:

- the sector(s) and distinct consumer populations targeted by the program(s)
- the geographical coverage of the program(s)
- the start and end dates of the programs(s)
- gross, adjusted gross, and/or net annual and cumulative energy savings of the program(s)

To the extent possible, this information will be supplemented by information regarding other federal, local, and non-IOU programs that overlapped the major energy efficiency programs.

7.5 Test and Assess the Viability of MCM Evaluations

At the very outset of the MCM pilot project *Demand Research* will begin discussing with CPUC staff the topics that might best be pursued for assessing how MCM studies may meet the CPUC's EM&V needs. This will help guide the initial database development and will permit work on the first agreed-upon evaluation topic to begin by the second or third month of the pilot project. Since MCM evaluations cannot be conducted until the appropriate data are available in the MCM database, the timing of the evaluations will have to be coordinated with the timing of the database. At present, *Demand Research* expects to meet all of its deadlines for database construction. Again, *Demand Research* assumes that there is enough time to conduct evaluations on up to three topics, and that energy consumption data will be provided by the IOUs in the manner that *Demand Research* specifies in no longer than 30 days after a data request is made.

Quality control is just as important for developing econometric models as it is for developing analyzable, error-free data. However, unlike data quality control, the quality control for multivariate models cannot be automated with programming instructions. Reliable and informative econometric models are created by applying both economic theory and econometric theory to a research subject. Economic theory offers general guidance on which variables and functional forms may be most appropriate for the problem at hand, but rarely offers specific guidance on the expected signs of most coefficients, their magnitudes, or the magnitudes of the standard errors and goodness-of-fits of the models. Moreover, economic theory is largely silent about the actual statistical qualities of the data that go into the models or the statistical qualities of the models themselves. More so in the social sciences than in the physical sciences, *there is no such thing as only one true model*.

Econometric theory, like economic theory, does not provide absolute standards by which to judge an econometric model, either. However, it does provides a host of diagnostic tools and statistical procedures for helping to determine if a statistical model possess desirable technical qualities. Desirable technical qualities for empirical models fall into three categories; consistency, unbiasedness, and efficiency. Empirical researchers have only limited control over these qualities, not only because of data limitations, but because there few absolute standards by which to judge these qualities or determine what the optimal tradeoff may be between them. Thus, every estimated model is a compromise between competing qualities. Contrary to the old saying, econometric practitioners know that sometimes it is better to be precisely wrong than it is to be approximately right. Approximately right could mean that a model is unstable and that little confidence can be placed in its results, while precisely wrong could mean that a model's estimates are highly reliable and close enough to the truth to be of great usefulness.

Despite the lack of absolute standards by which to judge econometric models, there are many generally-accepted professional practices that scientists use to test models and judge their merits. The number and scope of these analyses and diagnostics fill volumes and new ones are continually being invented. It thus serves no purpose to list a handful of them or to commit the MCM pilot project to using one set of model diagnostics versus another.

Statistical testing is but one form of assurance that the findings of energy demand models are valid. *Demand Research* views test of statistical significance as crucial to model specification and testing. However, the focus of MCM energy demand models must be on producing the best counterfactuals possible, not on ensuring that every sub-hypothesis is in conformance with a priori expectations or orthodox interpretations. As such, *Demand Research* relies on tests of statistical significance, but does not maintain a slavish devotion to them. Indeed, a second form of model development and validation testing can be of equal or greater importance. This is *expert judgment*. One of the great mathematicians of the 20ieth century, John von Neumann, once put together a dataset that showed that there was a strong relationship between an area's birthrate and its stork population. The point of the example was not merely to show that regression models can be deceiving, but to show that regression models must be subject to tests of reasonableness. To develop MCM energy demand models that produce valid policy findings it is important to understand the history of the policies being investigated and the set of circumstances they operate under.

Of the many concerns that may arise in the development of MCM policy models is the ability of the models to detect small changes in aggregate energy consumption. This is a legitimate worry and thus early efforts must be put into determining the appropriateness of a research design for given policy topics. When an investigation is focused on the combined impacts of many programs that have gone on for some time, it may be reasonable to suppose that energy demand models will be sensitive enough to detect policy-related changes. For example, IOUs in California reported that in the past decade all of their residential energy efficiency programs combined reduced residential energy consumption by more than 10 percent from what it might otherwise have been. This is a substantial amount of savings and an amount that should be detectable with the proper energy demand model and research design. On the other hand, if the actual energy savings was a fraction of this amount, it is unlikely that an MCM policy model could detect this small change with any confidence.

It is also the case, as noted in D.10-10-033, that there is concern over omitted variables, or measurement error, in the MCM model specifications. The CPUC decision quotes PG&E as asserting that, "the inherent limitation of such a metric is that factors outside of the energy efficiency arena could skew the perceived effect of energy efficiency programs themselves." The CPUC decision also notes that, "Other parties agree that it will be difficult to control for factors beyond energy efficiency policy when assessing the impact of energy efficiency programs on energy consumption."

Addressing these kinds concerns requires an eclectic approach in which a wide variety of expert evidence is brought to bear on the subject being investigated, from the record of bottom-up impact evaluations of energy efficiency programs to familiarity with the related public policies that may have affected the energy consumption levels that are being investigated. As such, efforts will be made as part of the MCM pilot project to validate the MCM energy demand models by referring to outside evidence and to the advice of policy experts.

In the end, when all is said and done it is of utmost importance that all MCM energy demand models and policy evaluations *meet the the current standards of social science research as judged by academic journals with double blind peer review*. This commitment should reassure the CPUC that the results of *Demand Research's* MCM pilot project will be scientifically defensible and worthy of use for policymaking.

8. MCM Pilot Project Tasks and Timeline

Task	Activities	2011 Sept.	Oct	Nov	Dec	2012 Jan.	Feb	Mar	Apr	May	June	July	Aug
1	**Residential**												
1.1	Data Collection												
1.2	Data Processing												
1.3	Model Development												
1.4	Data, Model Review												
1.5	Final Model/Report												
2	**Commercial**												
2.1	Data Collection												
2.2	Data Processing												
2.3	Model Development												
2.4	Data, Model Review												
2.5	Final Model/Report												
3	**Industrial/Ag**												
3.1	Data Collection												
3.2	Data Processing												
3.3	Model Development												
3.4	Data, Model Review												
3.5	Final Model/Report												

Note: A majority of the activities needed to develop the residential database and perform the analyses are generic and overlap those required for the commercial and industrial/agricultural sectors, hence the lightly shaded areas for these sectors. Due to the fixed costs involved, the marginal labor hours/costs associated with completing the databases and analyses for these two other sectors are substantially lower than for the residential sector.